



TECHNICAL UNIVERSITY OF CRETE
Electronic and Computer Engineering Department

**Power Quality Monitoring in Isolated Power Grids,
Using Advanced Neuro-Fuzzy Wavelet Based Techniques**

by

Epameinondas Ellinakis

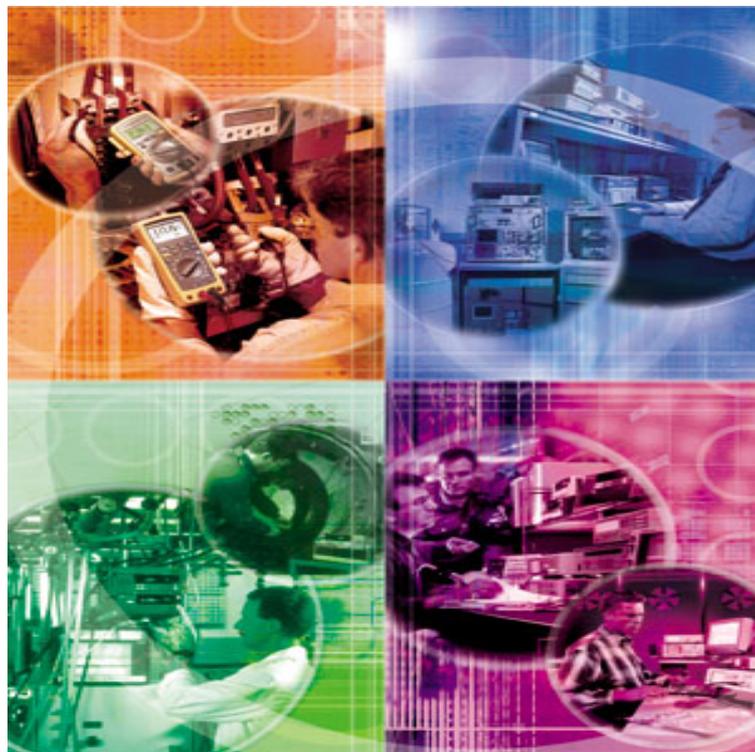
Electrical Engineer, Aristotelian University of Thessaloniki

Examination Committee:

Prof. George Staurakakis (Supervisor)

Prof. Michalis Zervakis

Prof. Konstantinos Kalaitzakis



Chania, February 2010

1 INTRODUCTION

The late epoch of the 20th century has viewed an enormous increase in electronic sensitive equipment due to the huge growth in communication, computer and automation industries. These new generations of electronic equipment are becoming progressively sensitive to power quality disturbances. Any variation in the supply voltage magnitude or frequency may have detrimental effects on the equipment. Electric utilities must assess the present value of the power before taking any quality improvement actions. Therefore, monitoring and detection of power quality disturbances has become a significant issue.

The increased concern for power quality has resulted in significant advances in monitoring algorithms, techniques and equipment that can be used to characterize disturbances and power quality variations. Analysis tools can present the power quality information as individual events (disturbance waveforms), trends, or statistical summaries. By comparing events with libraries of typical power quality variation characteristics and correlating with system events (e.g., capacitor switching), causes of the variations can be determined. In the same manner, the measured data should be correlated with impacts to help characterize the sensitivity of end use equipment to power quality variations. This will help identify equipment that requires power conditioning and provide specifications for the protection that can be developed based on the power quality variation characteristics.

The main purpose of this thesis is to find, present, test and evaluate three relatively new algorithms that are being used for power quality monitoring and classification. These are: firstly the “Adaline” which is a neural network structure that detects an anomaly in the signal of the voltage/current of a Power System whenever it occurs, secondly a “Wavelet-Based PNN, Probabilistic Neural Network”, which is also a neural network structure that makes use of the Wavelet Transform, which not only detects a disturbance when it occurs but it also classifies it, and finally a “Wavelet-Based ANFIS, Adaptive Neuro-Fuzzy Inference System” a fuzzy network structure that also uses the Wavelet Transform, and is also capable of power quality disturbance classification. The algorithms were implemented in the Matlab v6.5.1 platform and tested with data that have been acquired from the Power Supply Station in Katsampas at Heraklion, Crete. Those data were acquired via the Series 5500 Dual-

Node, a Power Quality Monitoring Device that was installed at the capacitors' 13.8kV busbar of the facility. The results of the algorithms' use were evaluated, compared, and some final conclusions came up, regarding their effectiveness and their flexibility.

This thesis is mainly referring to electrician engineers that are already dealing with Power Quality Disturbances and are in the pursuit of better methods, algorithms and devices for monitoring in order to improve the security level of their devices or facilities. However, because of the importance of Power Quality, in general, it was found crucial for the medium reader, with some basic knowledge of electricity, to be introduced to the concept of Power Quality and its monitoring. Therefore, in order for the average reader, as well, to fully comprehend where, how and why those algorithms are utilised, an introduction was made, regarding Power Quality including the instruments and the techniques used for its monitoring.

As a result the final structure of the thesis came up as follows. In Section 2 a very brief overview of Power Systems is written as well as its association with the consumers, which is, if not unknown, at least misunderstood! Section 3 introduces Power Quality. It explains what Power Quality really is and describes and analyses the disturbances that occur in a Power System, which, by the way are the only factors that establish the Power Quality of a System. Section 3 also explains the cause of their occurrence, and helps to realise how a classification of those phenomena can be made based on their features. In Section 4 the need for monitoring Power Quality is discussed by presenting the observed effects of the disturbances on the operation of various types of equipment, as well as the tolerances of the equipment upon those phenomena. Section 5 presents by category, based on their field of use and specifications, briefly some of the Power Quality Monitoring Devices that are usually utilized. In this presentation, the device that was used for the acquisition of the data used on this thesis, from the Power Supply Station in Heraklion, the Series 5500 DualNode and its software, Signature System, is highlighted. Section 6 presents a state of the art on the classification of power quality events. Initially, basic tools and techniques that have been employed during the last few years are presented. Consequently, there is a rather detailed introduction to the scientific fields, Neural Networks, Fuzzy Networks and Wavelet Transform, that the power quality classification techniques used in this thesis are taken from. Section 7 demonstrates the three algorithms that were used in this thesis, for power quality monitoring and classification. Firstly, the algorithms were theoretically-mathematically presented, afterwards they

were implemented in the Matlab v6.5.1. platform, where they were simulated with laboratory created power quality disturbances, once again in Matlab v6.5.1. and finally they were tested with real disturbances, taken from the Power Supply Station, via the Series 5500 DualNode and its software, Signature System, as mentioned earlier. In section 8 the conclusions are written and a comparison between the methods is made in order to highlight the advantages and the disadvantages of the algorithms. Section 9 discusses the future work that has to be done in order to get a better understanding of voltages and currents behavior under different event conditions, and a better understanding of relations between voltage and currents in different voltage levels and its propagation in the network as well. Finally, section 10 presents the references of the thesis.

2 POWER SYSTEMS IN GENERAL

The aim of the electric power system is to generate electrical energy and to deliver this energy to the end-user equipment at an acceptable voltage. The constraint that was traditionally mentioned is that the technical aim should be achieved for reasonable costs. The optimal level of investment was to be obtained by means of a trade-off between reliability and costs. A recurring argument with industrial customers concerned the definition of reliability: should it include only long interruptions or short interruptions and even voltage dips as well. The term power quality came in use referring to the other characteristics of the supply voltage (i.e. other than long interruptions). But, immediately, the first confusion started as utilities included the disturbances generated by the customers in the term ‘power quality’. This difference in emphasis will be discussed in more detail below. The main complaint of domestic customers concerned the costs which were perceived too high, especially where cross-subsidising was used to keep prices low for industrial or agricultural customers. This classical model of the power system, as it can be found in many textbooks, is found in fig. (1).

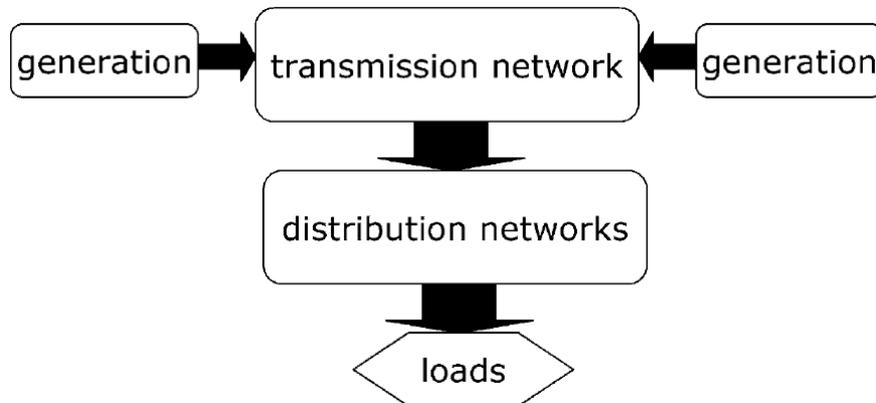


Figure 1. Classical Model Of The Power System.

The customers are traditionally referred to as loads. Various developments have led to a different view at the power system. These developments are strongly interrelated, but the three main ones are:

- The deregulation of the electricity industry makes that there is no longer one single system but a number of independent companies with customers.

- Electricity customers have become more aware of their rights and demand low-cost electricity of high reliability and quality, where the priorities are different for different (types of) customers. Customers are certainly no longer willing to accept their position as merely one parameter in a global optimisation.
- Generation of electricity is shifting away from large power stations connected to the transmission system towards smaller units connected at lower voltage levels. Examples are combined-heat-and-power and renewable sources of energy like sun and wind.

Because of this the power system can no longer be seen as one entity but as an electricity network with customers. This new model is shown in fig. (2).

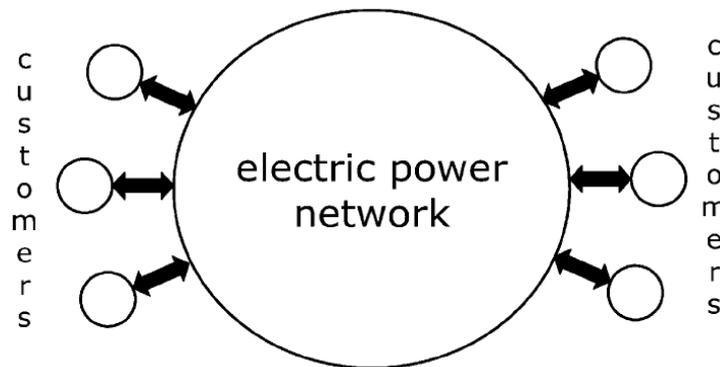


Figure 2 Modern Model Of The Power System

Note that the physical structure of the power system/network has not changed, it is only the way of viewing it that has changed. In fig. (2) the electric power network connects some or many customers. Customers may generate or consume electrical energy, or even both albeit at different moments in time. Different customers have different demands on voltage magnitude, frequency, waveform, etc. Different customers have different patterns of current variation, fluctuation and distortion, thus polluting the voltage for other customers in different ways. The power network in fig. (2) could be a transmission network, a distribution network, an industrial network, or any other network owned by one single company. For a transmission network, the customers are, e.g., generator stations, distribution networks, large industrial customers (who could be generating or consuming electricity at different times, based on the electricity price at that moment), and other transmission networks. For a distribution network, the customers are currently mainly end-users that only consume electricity, but also the transmission network and smaller generator stations are customers. Note

that all customers are equal, even though some may be producing energy while others are consuming it. The aim of the network company is only to transport the energy, or in economic terms: to enable transactions between customers. The technical aim of the power network becomes one of allowing the transport of electrical energy between the different customers, guaranteeing an acceptable voltage and allowing the currents taken by the customers. Note that this same model also becomes attractive when considering the integration of renewable or other environmentally friendly sources of energy into the power system. The power network is no longer the boundary condition that limits e.g. the amount of wind power that can be produced at a certain location. Instead the power network's task becomes to enable the transport of the amount of wind power that is produced. With an ideal network each customer should perceive the electricity supply as an ideal voltage source with zero impedance, which means that whatever the current is, the voltage should be constant. As always, reality is not ideal. Power quality is all about this deviation between reality and ideal.

Now that we have made a wide introduction into the term of Power Quality (PQ) as well as the reasons that engendered it let us take a closer, more scientific, look hereupon.

3 POWER QUALITY

The term power quality refers to a wide variety of electromagnetic phenomena that characterize the voltage and the current at a given time and at a given location on the power system. The purpose of this clause is to present concise definitions of terms that convey the basic concepts of power quality and moreover power quality monitoring. This is done by providing technical descriptions and examples of the principal electromagnetic phenomena causing power quality problems. Electromagnetic disturbances are caused by the increasing application of electronic equipment, of any size or function. Accompanying the increase in operation problems have been a variety of attempts to describe the phenomena. Unfortunately, different segments of the electronics community have utilized different terminologies to describe electromagnetic events. This clause expands the terminology that will be used in the power quality community to describe these common events and also offers explanations as to why commonly used terminology in other communities will not be used in power quality discussions.

3.1 Electromagnetic Compatibility

This document uses the electromagnetic compatibility approach to describing power quality phenomena. The electromagnetic compatibility approach has been accepted by the international community in IEC standards produced by IEC Technical Committee 77. Reference (4) provides an excellent overview of the electromagnetic compatibility concept and associated IEC documents.

3.2 General Classification Of Phenomena

The IEC classifies electromagnetic phenomena into several groups as shown in Table 1 (5). The IEC standard addresses the conducted electrical parameters shown in table 1. The terms *high-* and *low-frequency* are not defined in terms of a specific frequency range, but instead are intended to indicate the relative difference in principal frequency content of the phenomena listed in these categories. This practice also contains a few additional terms related to the IEC terminology. The term *sag* is used in the power quality community as a synonym to the IEC term *dip*. The category *short duration variations* is used to refer to voltage dips and short interruptions. The term *swell* is introduced as an inverse to *sag (dip)*. The category *long duration variation*

has been added to deal with ANSI C84.1-1989 (6) limits. The category *noise* has been added to deal with broad-band conducted phenomena. The category *waveform distortion* is used as a container category for the *IEC harmonics*, *interharmonics*, and *dc in ac networks* phenomena as well as an additional phenomenon from IEEE Std 519-1992 (7) called *notching*. Table 2 shows the categorization of electromagnetic phenomena used for the power quality community.

Conducted low-frequency phenomena	Harmonics, interharmonics
	Signal systems (power line carrier)
	Voltage fluctuations
	Voltage dips and interruptions
	Voltage imbalance
	Power-frequency variations
	Induced low-frequency voltages
	DC in ac networks
Radiated low-frequency phenomena	Magnetic fields
	Electric fields
Conducted high-frequency phenomena	Induced continuous wave voltages or currents
	Unidirectional transients
	Oscillatory transients
Radiated high-frequency phenomena	Magnetic fields
	Electric fields
	Electromagnetic fields
	Continuous waves
	Transients
Electrostatic discharge phenomena	—
Nuclear electromagnetic pulse	—

Table 1. Principal Phenomena Causing Electromagnetic Disturbances As Classified By The IEC

Categories	Typical spectral content	Typical duration	Typical voltage magnitude
1.0 Transients			
1.1 Impulsive			
1.1.1 Nanosecond	5ns rise	<50ns	
1.1.2 Microsecond	1 μ s rise	50ns-1ms	
1.1.3 Millisecond	0,1ms rise	>1ms	
1.2 Oscillatory			
1.2.1 Low frequency	<5kHz	0,3-50ms	0-4pu
1.2.2 Medium frequency	5-500kHz	20 μ s	0-8pu
1.2.3 High frequency	0,5-5MHz	5 μ s	0-4pu
2.0 Short duration variations			
2.1 Instantaneous			
2.1.1 Sag		0,5-30 cycles	0,1-0,9pu
2.1.2 Swell		0,5-30 cycles	1,1-1,8pu
2.2 Momentary			
2.2.1 Interruption		0,5cycles-3s	<0,1pu
2.2.2 Sag		30cycles-3s	0,1-0,9pu
2.2.3 Swell		30cycles-3s	1,1-1,4pu
2.3 Temporary			
2.3.1 Interruption		3s-1min	<0,1pu
2.3.2 Sag		3s-1min	0,1-0,9pu
2.3.3 Swell		3s-1min	1,1-1,2pu
3.0 Long duration variations			
3.1 Interruption, sustained		>1min	0,0pu
3.2 Undervoltages		>1min	0,8-0,9pu
3.3 Overvoltages		>1min	1,1-1,2pu
4.0 Voltage imbalance		steady state	0,5-2%
5.0 Waveform distortion			
5.1 DC offset		steady state	0-0,1%
5.2 Harmonics		steady state	0-20%
5.3 Interharmonics		steady state	0-2%

5.4 Notching		steady state	
5.5 Noise	broad-band	steady state	0-1%
6.0 Voltage fluctuations	<25Hz	intermittent	0,1-7%
7.0 Power frequency variations		<10sec	

Table 2. Categories And Typical Characteristics Of Power System Electromagnetic Phenomena

The phenomena listed in table 1 can be described further by listing appropriate attributes. For steady-state phenomena, the following attributes can be used (8):

- Amplitude
- Frequency
- Spectrum
- Modulation
- Source impedance
- Notch depth
- Notch area

For non-steady state phenomena, other attributes may be required (8):

- Rate of rise
- Amplitude
- Duration
- Spectrum
- Frequency
- Rate of occurrence
- Energy potential
- Source impedance

Table 1 provides information regarding typical spectral content, duration, and magnitude where appropriate for each category of electromagnetic phenomena (8), (9), (10). The categories of table 2, when used with the attributes mentioned above, provide a means to clearly describe an electromagnetic disturbance. The categories and their descriptions are important in order to be able to classify measurement results and to describe electromagnetic phenomena that can cause power quality problems. The remainder of this clause will discuss each category in detail.

3.3 Detailed Descriptions Of Phenomena

This paragraph provides more detailed descriptions for each of the power quality variation categories presented in Table 2. These descriptions provide some history regarding the terms currently in use for each category. Typical causes of electromagnetic phenomena in each category are introduced, and are expanded later on. One of the main reasons for developing the different categories of electromagnetic phenomena is that there are different ways to solve power quality problems depending on the particular variation that is of concern. The different solutions available are discussed for each category. There are also different requirements for characterizing the phenomena using measurements. It is important to be able to classify events and electro-magnetic phenomena for analysis purposes. The measurement requirements for each category of electro-magnetic phenomenon are discussed.

3.3.1 Transients

The term transient has been used in the analysis of power system variations for a long time. Its name immediately conjures up the notion of an event that is undesirable but momentary in nature. The IEEE Std 100-1992 definition of transient reflects this understanding. The primary definition uses the word rapid and talks of frequencies up to 3 MHz when defining transient in the context of evaluating cable systems in substations. The notion of a damped oscillatory transient due to a RLC network is also mentioned. This is the type of phenomena that most power engineers think of when they hear the word transient.

Other definitions in IEEE Std 100-1992 are broader in scope and simply state that a transient is “that part of the change in a variable that disappears during transition from one steady-state operating condition to another”. Unfortunately, this definition could be used to describe just about anything unusual that happens on the power system.

Another word used in current IEEE standards that is synonymous with transient is surge. IEEE Std 100-1992 defines a surge as “a transient wave of current, potential, or power in an electric circuit”. The IEEE C62 Collection (11) uses the terms surge, switching surge, and transient to describe the same types of phenomena. For the purposes of this document, surge will not be used to describe transient electromagnetic phenomena. Since IEEE Std 100-1992 uses the term transient to define surge, this limitation should not cause conflicts.

Broadly speaking, transients can be classified into two categories: impulsive and oscillatory. These terms reflect the waveshape of a current or voltage transient.

3.3.1.1 Impulsive Transient

An impulsive transient is a sudden, nonpower frequency change in the steady-state condition of voltage, current, or both, that is unidirectional in polarity (primarily either positive or negative). Impulsive transients are normally characterized by their rise and decay times. These phenomena can also be described by their spectral content. For example, a 1.2/50 ms 2000 V impulsive transient rises to its peak value of 2000 V in 1.2 ms, and then decays to half its peak value in 50 ms (11).

The most common cause of impulsive transients is lightning. fig. (3) illustrates a typical current impulsive transient caused by lightning.

Due to the high frequencies involved, impulsive transients are damped quickly by resistive circuit components and are not conducted far from their source. There can be significant differences in the transient characteristic from one location within a building to another. Impulsive transients can excite power system resonance circuits and produce the following type of disturbance: oscillatory transients.

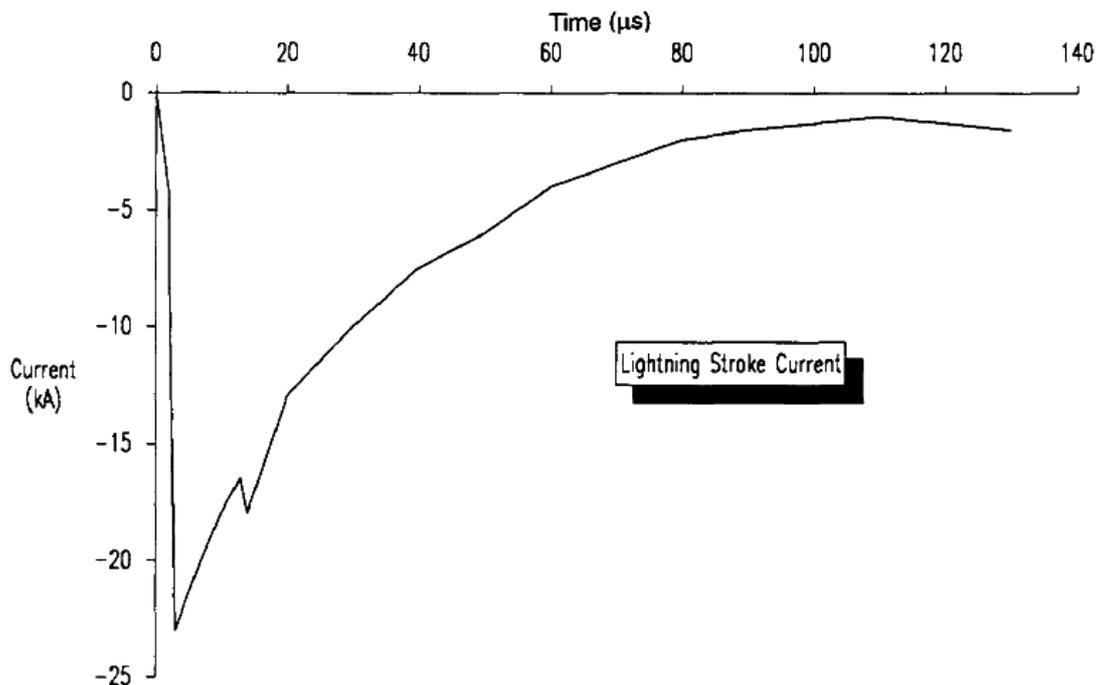


Figure 3. Lightning Stroke Current That Can Result In Impulsive Transients On The Power System

3.3.1.2 Oscillatory Transient

An oscillatory transient consists of a voltage or current whose instantaneous value changes polarity rapidly. It is described by its spectral content (predominant frequency), duration, and magnitude. The spectral content subclasses defined in table 2 are high, medium, and low frequency. The frequency ranges for these classifications are chosen to coincide with common types of power system oscillatory transient phenomena.

As with impulsive transients, oscillatory transients can be measured with or without the fundamental frequency component included. When characterizing the transient, it is important to indicate the magnitude with and without the fundamental component.

Oscillatory transients with a primary frequency component greater than 500 kHz and a typical duration measured in microseconds (or several cycles of the principal frequency) are considered high-frequency oscillatory transients. These transients are almost always due to some type of switching event. High-frequency oscillatory transients are often the result of a local system response to an impulsive transient.

Power electronic devices produce oscillatory voltage transients as a result of commutation and RLC snubber circuits. The transients can be in the high kilohertz range, last a few cycles of their fundamental frequency, and have repetition rates of several times per 60 Hz cycle (depending on the pulse number of the device) and magnitudes of 0.1 pu (less the 60 Hz component).

A transient with a primary frequency component between 5 and 500 kHz with duration measured in the tens of microseconds (or several cycles of the principal frequency) is termed a medium-frequency transient.

Back-to-back capacitor energization results in oscillatory transient currents in the tens of kilohertz. This phenomenon occurs when a capacitor bank is energized in close electrical proximity to a capacitor bank already in service. The energized bank sees the de-energized bank as a low impedance path (limited only by the inductance of the bus to which the banks are connected, typically small). Fig. (4) illustrates the resulting current transient due to back-to-back capacitor switching. Cable switching results in oscillatory voltage transients in the same frequency range. Medium-frequency transients can also be the result of a system response to an impulsive transient.

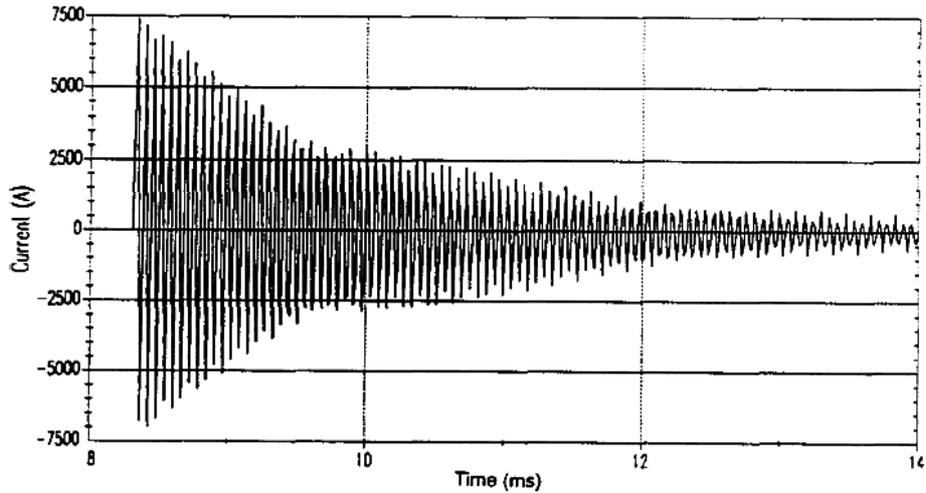


Figure 4. Oscillatory Transient Caused By Back-To-Back Capacitor Switching

A transient with a primary frequency component less than 5 kHz, and a duration from 0.3 to 50 ms, is considered a low-frequency transient.

This category of phenomena is frequently encountered on subtransmission and distribution systems and is caused by many types of events, primarily capacitor bank energization. The resulting voltage waveshape is very familiar to power system engineers and can be readily classified using the attributes discussed so far. Capacitor bank energization typically results in an oscillatory voltage transient with a primary frequency between 300 and 900 Hz. The transient has a peak magnitude that can approach 2.0 pu, but is typically 1.3-1.5 pu lasting between 0.5 and 3 cycles, depending on the system damping, see fig. (5) below.

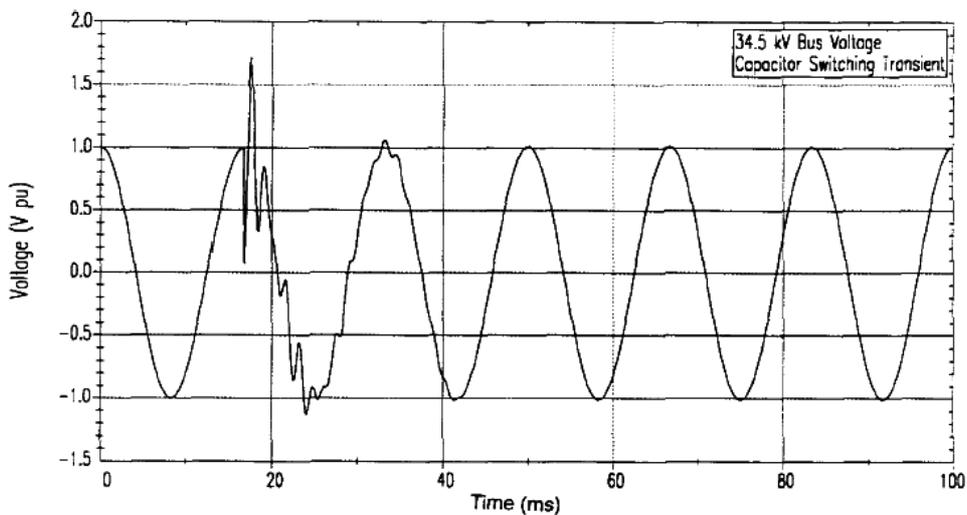


Figure 5. Low Frequency Oscillatory Transient Caused By Capacitor-Bank Energization

Oscillatory transients with principal frequencies less than 300 Hz can also be found on the distribution system. These are generally associated with ferroresonance and transformer energization, see fig. (6). Transients involving series capacitors could also fall into this category. They occur when the system resonance results in magnification of low-frequency components in the transformer inrush current (second, third harmonic) or when unusual conditions result in ferroresonance. IEEE Std C62.41-1991 (11) describes surge waveforms deemed to represent the environment in which electrical equipment and surge protective devices will be expected to operate. Reference (11) covers the origin of surge (transient) voltages, rate of occurrence and voltage levels in unprotected circuits, waveshapes of representative surge voltages, energy, and source impedance.

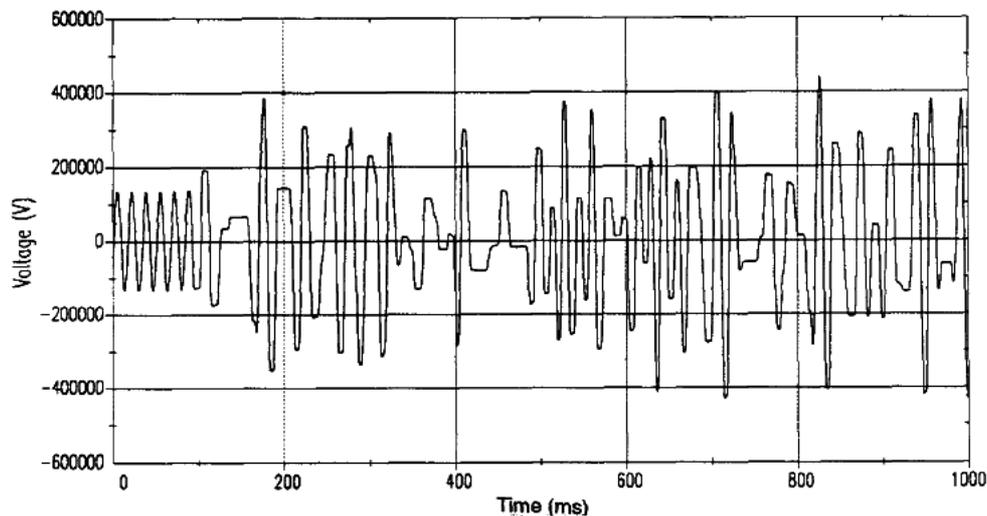


Figure 6. Low-Frequency Oscillatory Transient Caused By Ferroresonance Of An Unloaded Transformer

3.3.2 Short-Duration Variations

This category encompasses the IEC category of voltage dips and short interruptions as well as the antithesis of dip or swell. Each type of variation can be designated as instantaneous, momentary, or temporary, depending on its duration as defined in Table 2.

Short-duration voltage variations are almost always caused by fault conditions, the energization of large loads that require high starting currents, or intermittent loose connections in power wiring. Depending on the fault location and the system conditions, the fault can cause either temporary voltage rises (swells) or voltage drops (sags), or a complete loss of voltage (interruptions). The fault condition can be close

to or remote from the point of interest. In either case, the impact on the voltage during the actual fault condition is a short-duration variation. Changes in current which fall into the duration and magnitude categories are also included in short-duration variations.

3.3.2.1 Interruption

An interruption occurs when the supply voltage or load current decreases to less than 0.1 pu for a period of time not exceeding 1 min.

Interruptions can be the result of power system faults, equipment failures, and control malfunctions. The interruptions are measured by their duration since the voltage magnitude is always less than 10% of nominal. The duration of an interruption due to a fault on the utility system is determined by utility protective devices and the particular event that is causing the fault. The duration of an interruption due to equipment malfunctions or loose connections can be irregular.

Some interruptions may be preceded by a voltage sag when these interruptions are due to faults on the source system. The voltage sag occurs between the time a fault initiates and the protective device operates. On the faulted feeder, loads will experience a voltage sag followed immediately by an interruption. The duration of the interruption will depend on the reclosing capability of the protective device. Instantaneous reclosing generally will limit the interruption caused by a non-permanent fault to less than 30 cycles. Delayed reclosing of the protective device may cause a momentary or temporary interruption.

Fig. (7) shows a momentary interruption during which voltage drops for about 2.3 s. Note from the wave-shape plot of this event that the instantaneous voltage may not drop to zero immediately upon interruption of the source voltage. This residual voltage is due to the back-emf effect of induction motors on the interrupted circuit.

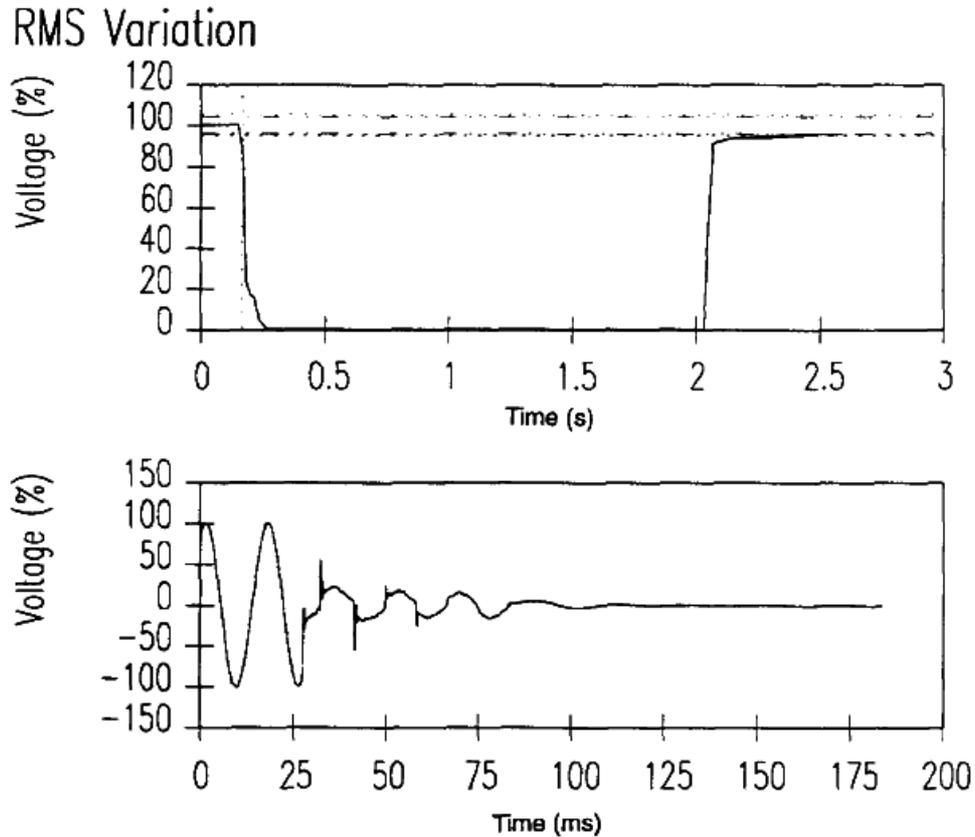


Figure 7. Momentary Interruption Due To A Fault And Subsequent Recloser Operation

3.3.2.2 Sags (Dips)

Terminology used to describe the magnitude of a voltage sag is often confusing. The recommended usage is “a sag to 20%” which means that the line voltage is reduced down to 20% of the normal value, not reduced by 20%. Using the preposition “of” (as in “a sag of 20%” or implied in “a 20% sag”) is deprecated. This preference is consistent with IEC practice, and with most disturbance analyzers that also report remaining voltage. Just as an unspecified voltage designation is accepted to mean line-to-line potential, so an unspecified sag magnitude will refer to the remaining voltage. Where possible, the nominal or base voltage and the remaining voltage should be specified.

Voltage sags are usually associated with system faults but can also be caused by switching of heavy loads or starting of large motors. Fig. (8) shows a typical voltage sag that can be associated with a single line-to-ground (SLG) fault. Also, a fault on a parallel feeder circuit will result in a voltage drop at the substation bus that affects all of the other feeders until the fault is cleared. Typical fault clearing times

range from 3 to 30cycles, depending on the fault current magnitude and the type of overcurrent detection and interruption.

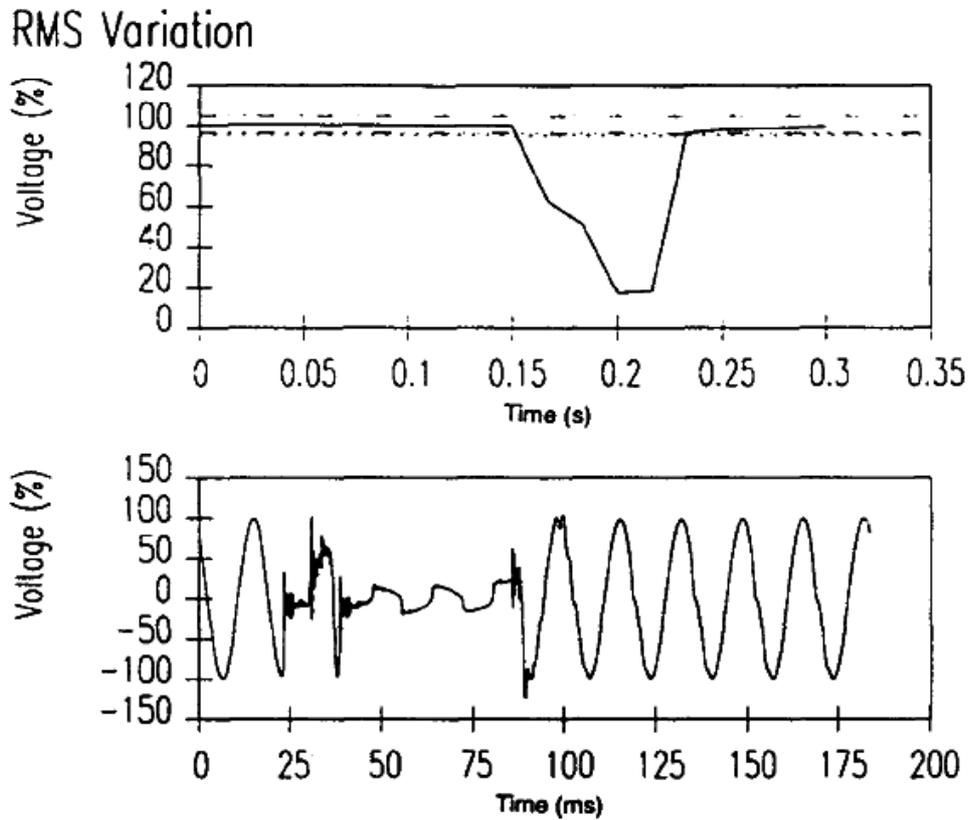


Figure 8. Instantaneous Voltage Sag Caused By A SLG Fault

Voltage sags can also be caused by large load changes or motor starting. An induction motor will draw six to ten times its full load current during starting. This lagging current causes a voltage drop across the impedance of the system. If the current magnitude is large relative to the system available fault current, the resulting voltage sag can be significant. Fig. (9) illustrates the effect of a large motor starting.

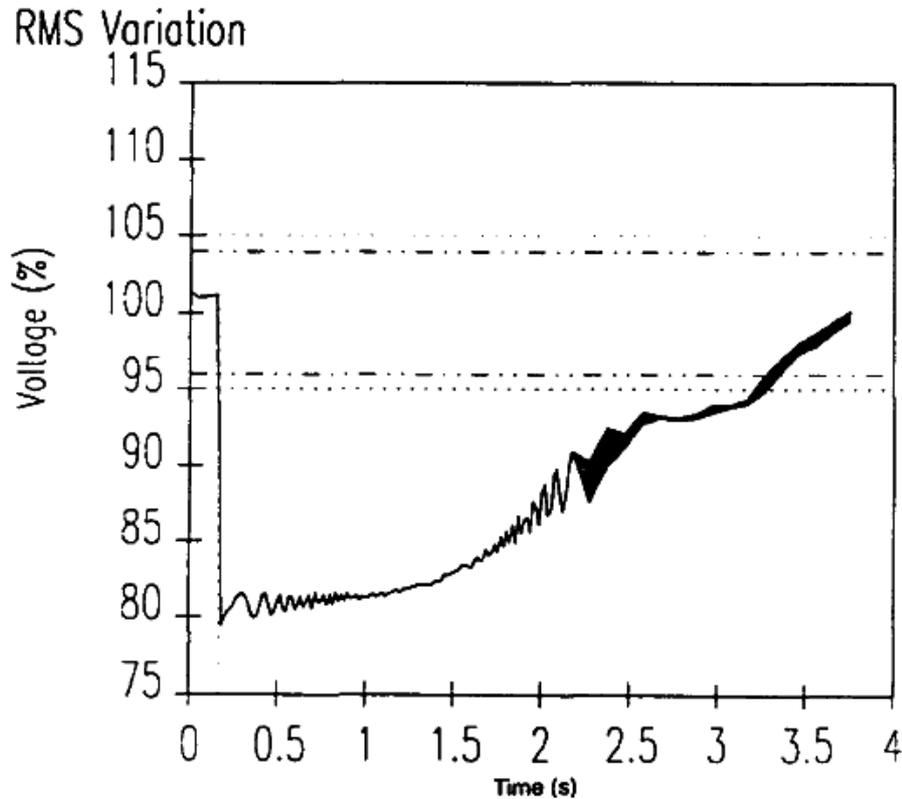


Figure 9. Temporary Voltage Sag Caused By Motor Starting

The term sag has been used in the power quality community for many years to describe a specific type of power quality disturbance: a short duration voltage decrease. Clearly, the notion is directly borrowed from the literal definition of the word sag. The IEC definition for this phenomenon is dip. The two terms are considered interchangeable, with sag being preferred in the US power quality community.

Previously, the duration of sag events has not been clearly defined. Typical sag duration defined in some publications ranges from 2 ms (about 1/8 of a cycle) to a couple of minutes. Undervoltages that last less than 1/2 cycle cannot be characterized effectively as a change in the rms value of the fundamental frequency value. Therefore, these events are considered transients; see IEC 1000-2-1 (1990). Undervoltages that last longer than 1 min can typically be controlled by voltage regulation equipment and may be associated with a wide variety of causes other than system faults. Therefore, these are classified as long duration variations in 2.3.3.

Sag durations are subdivided here into three categories:

- Instantaneous
- Momentary and
- Temporary

Which coincide with the three categories of interruptions and swells. These durations are intended to correlate with typical protective device operation times as well as duration divisions recommended by international technical organizations (9).

3.3.2.3 Swells

A swell is defined as an increase in rms voltage or current at the power frequency for durations from 0.5 cycles to 1 min. Typical magnitudes are between 1.1 and 1.8 pu. Swell magnitude is also described by its remaining voltage, in this case, always greater than 1.0.

As with sags, swells are usually associated with system fault conditions, but they are much less common than voltage sags. A swell can occur due to a single line-to-ground fault on the system resulting in a temporary voltage rise on the unfaulted phases. Swells can also be caused by switching off a large load or switching on a large capacitor bank. Fig. (10) illustrates a voltage swell caused by a SLG fault.

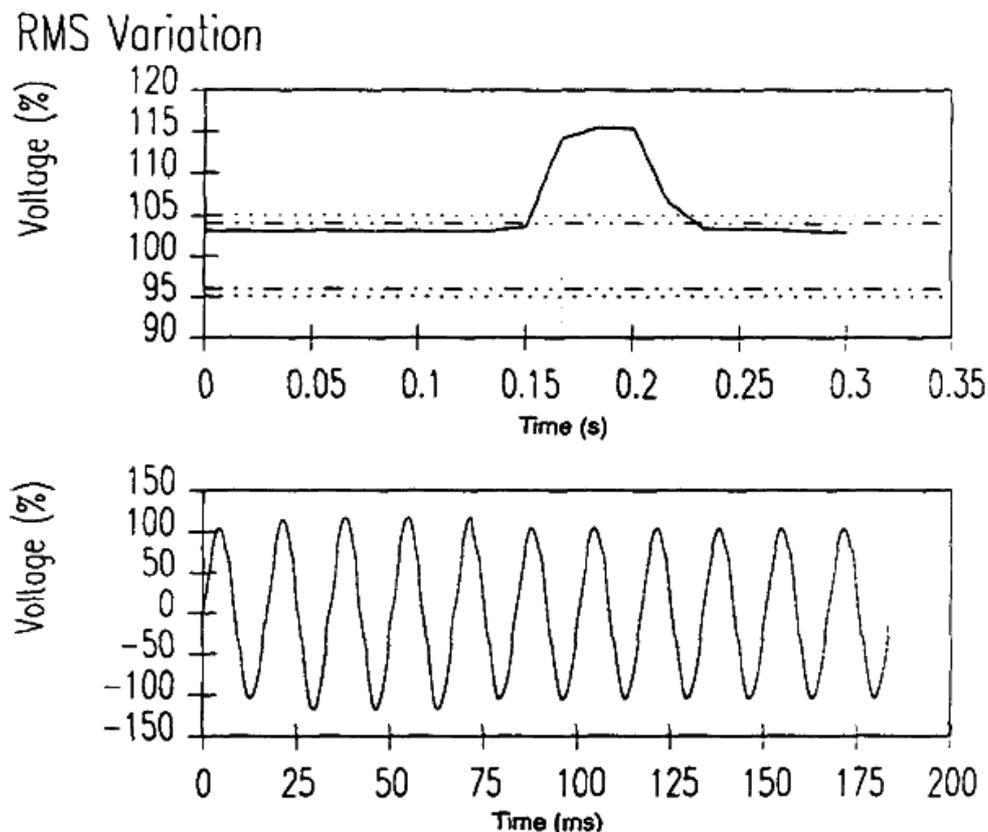


Figure 10. Instantaneous Voltage Swell Caused By A SLG Fault

Swells are characterized by their magnitude (rms value) and duration. The severity of a voltage swell during a fault condition is a function of the fault location,

system impedance, and grounding. On an ungrounded system, the line-to-ground voltages on the ungrounded phases will be 1.73 pu during a line-to-ground fault condition. Close to the substation on a grounded system, there will be no voltage rise on the unfaulted phases because the substation transformer is usually connected delta-wye, providing a low impedance zero-sequence path for the fault current.

In some publications, the term momentary overvoltage is used as a synonym for the term swell. A formal definition of swell in IEEE Std C62.41-1991 is “A momentary increase in the power-frequency voltage delivered by the mains, outside of the normal tolerances, with a duration of more than one cycle and less than a few seconds (11)”. This definition is not preferred by the power quality community.

3.3.3 Long Duration Variations

Long duration variations encompass rms deviations at power frequencies for longer than 1 min. The steady-state voltage tolerances expected on a power system are specified in (3). These magnitudes are reflected in Table 2. Long duration variations are considered to be present when the ANSI limits are exceeded for greater than 1 min.

Long duration variations can be either overvoltages or undervoltages, depending on the cause of the variation. Overvoltages and undervoltages generally are not the result of system faults. They are caused by load variations on the system and system switching operations. These variations are characterized by plots of rms voltage versus time.

3.3.3.1 Overvoltage

Overvoltages can be the result of load switching (e.g., switching off a large load), or variations in the reactive compensation on the system (e.g., switching on a capacitor bank). Poor system voltage regulation capabilities or controls result in overvoltages. Incorrect tap settings on transformers can also result in system overvoltages.

3.3.3.2 Undervoltage

Undervoltages are the result of the events that are the reverse of the events that cause overvoltages. A load switching on, or a capacitor bank switching off, can cause an undervoltage until voltage regulation equipment on the system can bring the voltage back to within tolerances. Overloaded circuits can result in under-voltages also.

The term brownout is sometimes used to describe sustained periods of low power-frequency voltage initiated as a specific dispatch strategy to reduce power delivery. The type of disturbance described by brownout is basically the same as that described by the term undervoltage defined here. Because there is no formal definition for the term brownout, and because the term is not as clear as the term undervoltage when trying to characterize a disturbance, the term brownout should be avoided in future power quality activities in order to avoid confusion.

3.3.3.3 Sustained Interruptions

The decrease to zero of the supply voltage for a period of time in excess of 1 min is considered a sustained interruption. Voltage interruptions longer than 1 min are often permanent in nature and require manual intervention for restoration. Sustained interruptions are a specific power system phenomena and have no relation to the usage of the term outage. Outage, as defined in IEEE Std 100-1992, does not refer to a specific phenomenon, but rather to the state of a component in a system that has failed to function as expected. Also, use of the term interruption in the context of power quality monitoring has no relation to reliability or other continuity of service statistics.

3.3.4 Voltage Imbalance

Voltage imbalance (or unbalance) is defined as the ratio of the negative or zero sequence component to the positive sequence component. The negative or zero sequence voltages in a power system generally result from unbalanced loads causing negative or zero sequence currents to flow. Fig. (11) shows an example of a one-week trend of imbalance measured at one point on a residential feeder.

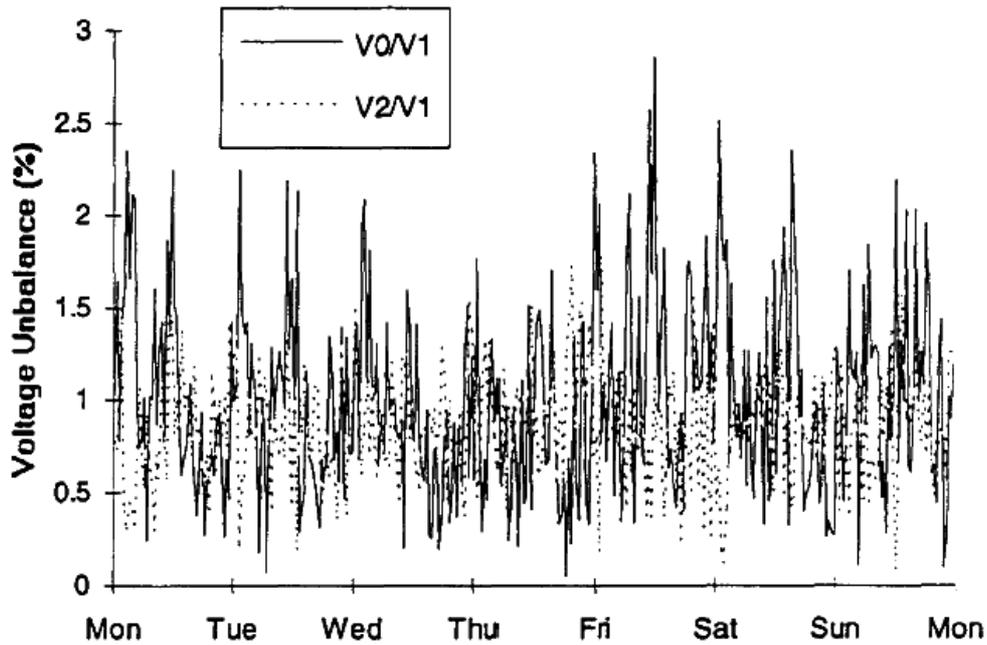


Figure 11. Imbalance Trend For A Residential Feeder

Imbalance can be estimated as the maximum deviation from the average of the three-phase voltages or currents, divided by the average of the three-phase voltages or currents, expressed in percent. In equation form:

$$\text{voltage imbalance} = 100 \times \frac{\text{max deviation from average voltage}}{\text{average voltage}}$$

For example, with phase-to-phase voltage readings of 230, 232, and 225, the average is 229. The maximum deviation from the average among the three readings is 4. The percent imbalance is:

$$\text{voltage imbalance} = 100 \times \frac{4}{229} = 1.7\%$$

The primary source of voltage imbalance less than 2% is unbalanced single phase loads on a three-phase circuit. Voltage imbalance can also be the result of capacitor bank anomalies, such as a blown fuse on one phase of a three-phase bank. Severe voltage imbalance (greater than 5%) can result from single-phasing conditions.

3.3.5 Waveform Distortion

Waveform distortion is a steady-state deviation from an ideal sine wave of power frequency principally characterized by the spectral content of the deviation.

There are five primary types of waveform distortion as follows:

- a) DC offset

- b) Harmonics
- c) Interharmonics
- d) Notching
- e) Noise

Each of these will be discussed separately.

3.3.5.1 DC Offset

The presence of a dc voltage or current in an ac power system is termed dc offset. This phenomenon can occur as the result of a geomagnetic disturbance or due to the effect of half-wave rectification. Incandescent light bulb life extenders, for example, may consist of diodes that reduce the rms voltage supplied to the lightbulb by half-wave rectification. Direct current in alternating current networks can be detrimental due to an increase in transformer saturation, additional stressing of insulation, and other adverse effects.

3.3.5.2 Harmonics

Harmonics are sinusoidal voltages or currents having frequencies that are integer multiples of the frequency at which the supply system is designed to operate (termed the fundamental frequency; usually 50 Hz or 60 Hz), see IEC 1000-2-1 (1990). Harmonics combine with the fundamental voltage, or current, and produce waveform distortion. Harmonic distortion exists due to the nonlinear characteristics of devices and loads on the power system.

These devices can usually be modelled as current sources that inject harmonic currents into the power system. Voltage distortion results as these currents cause nonlinear voltage drops across the system impedance. Harmonic distortion is a growing concern for many customers and for the overall power system due to increasing application of power electronics equipment.

Harmonic distortion levels can be characterized by the complete harmonic spectrum with magnitudes and phase angles of each individual harmonic component. It is also common to use a single quantity, the total harmonic distortion, as a measure of the magnitude of harmonic distortion.

Harmonic currents result from the normal operation of nonlinear devices on the power system. Fig. (12) illustrates the waveform and harmonic spectrum for a typical adjustable speed drive input current. Current distortion levels can be character-

ized by a total harmonic distortion, as described above, but this can often be misleading. For instance, many adjustable speed drives will exhibit high total harmonic distortion values for the input current when they are operating at very light loads. This is not a significant concern because the magnitude of harmonic current is low, even though its relative distortion is high.

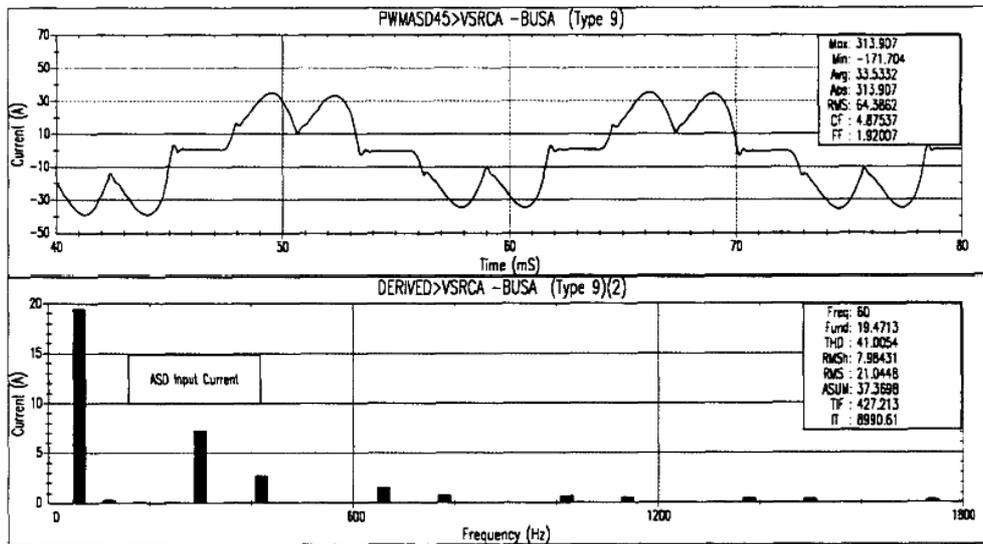


Figure 12. Current Waveform And Harmonic Spectrum For An ASD Input Current

To handle this concern for characterizing harmonic currents in a consistent fashion, IEEE Std 519-1992 (7) defines another term, the total demand distortion. This term is the same as the total harmonic distortion except that the distortion is expressed as a percent of some rated load current rather than as a percent of the fundamental current magnitude. Guidelines for harmonic current and voltage distortion levels on distribution and transmission circuits are provided in (7).

3.3.5.3 Interharmonics

Interharmonics can be found in networks of all voltage classes. They can appear as discrete frequencies or as a wide-band spectrum. The main sources of interharmonic waveform distortion are static frequency converters, cyclo-converters, induction motors, and arcing devices. Power-line carrier signals can also be considered as interharmonics.

The effects of interharmonics are not well known, but have been shown to affect power line carrier signaling, and induce visual flicker in display devices such as CRTs. IEC 1000-2-1 (1990) places background noise phenomenon in the interharmonic category.

3.3.5.4 Notching

Notching is a periodic voltage disturbance caused by the normal operation of power electronics devices when current is commutated from one phase to another.

Voltage notching represents a special case that falls between transients and harmonic distortion. Since notching occurs continuously (steady state), it can be characterized through the harmonic spectrum of the affected voltage. However, the frequency components associated with notching can be quite high and may not be readily characterized with measurement equipment normally used for harmonic analysis.

Three-phase converters that produce continuous dc current are the most important cause of voltage notching, fig. (13). The notches occur when the current commutates from one phase to another. During this period, there is a momentary short circuit between two phases. The severity of the notch at any point in the system is determined by the source inductance and the isolating inductance between the converter and the point being monitored. Notching is described in detail in IEEE Std 519-1992 (7).

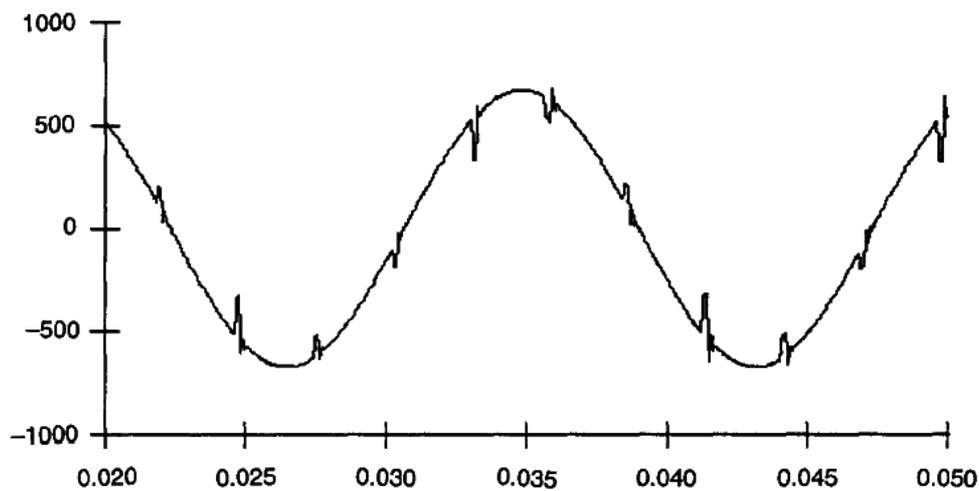


Figure 13. Example Of Voltage Notching Caused By Converter Operation

3.3.5.5 Noise

Noise is unwanted electrical signals with broadband spectral content lower than 200 kHz superimposed upon the power system voltage or current in phase conductors, or found on neutral conductors or signal lines. Noise in power systems can be caused by power electronic devices, control circuits, arcing equipment, loads with solid-state rectifiers, and switching power supplies. Noise problems are often exacer-

bated by improper grounding. Basically, noise consists of any unwanted distortion of the power signal that cannot be classified as harmonic distortion or transients.

The frequency range and magnitude level of noise depend on the source, which produces the noise and the system characteristics. A typical magnitude of noise is less than 1% of the voltage magnitude. Noise disturbs electronic devices such as microcomputer and programmable controllers. The problem can be mitigated by using filters, isolation transformers, and some line conditioners.

3.3.6 Voltage Fluctuations

Voltage fluctuations are systematic variations of the voltage envelope or a series of random voltage changes, the magnitude of which does not normally exceed the voltage ranges specified by (3) of 0.95 - 1.05 pu.

IEC 555-3, which has been revised as IEC 1000-3-3 (1994) (12) defines various types of voltage fluctuations. The reader is referred to this document for a detailed breakdown of these types. The remainder of this discussion on voltage fluctuations will concentrate on the IEC 1000-3-3 (1994) Type (d) voltage fluctuations. This type is characterized as a series of random or continuous voltage fluctuations.

Any load that has significant current variations, especially in the reactive component, can cause voltage fluctuations. Loads that exhibit continuous, rapid variations in load current magnitude can cause voltage variations erroneously referred to as flicker. The term flicker is derived from the impact of the voltage fluctuation on lighting intensity. Voltage fluctuation is the response of the power system to the varying load and light flicker is the response of the lighting system as observed by the human eye. The power system, the lighting system, and the human response are all variables. Even though there is a clear distinction between these terms -cause and effect- they are often confused to the point that the term “voltage flicker” is used in some documents. Such incorrect usage should be avoided.

Arc furnaces are the most common cause of voltage fluctuations on the transmission and distribution system. Voltage fluctuations are defined by their rms magnitude expressed as a percent of the fundamental. Lighting flicker is measured with respect to the sensitivity of the human eye. An example of a voltage waveform that produces flicker is shown in fig. (14).

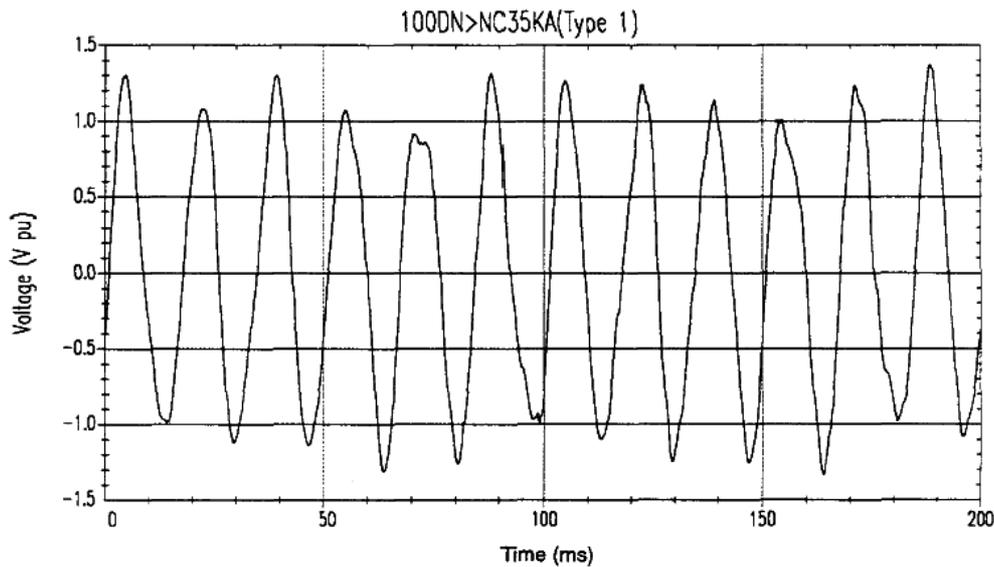


Figure 14. Example Of Voltage Fluctuations Caused By Arc Furnace Operation

Voltage fluctuations generally appear as a modulation of the fundamental frequency (similar to amplitude modulation of an am radio signal). Therefore, it is easiest to define a magnitude for the voltage fluctuation as the rms magnitude of the modulation signal. This can be obtained by demodulating the waveform to remove the fundamental frequency and then measuring the magnitude of the modulation components. Typically, magnitudes as low as 0.5% can result in perceptible light flicker if the frequencies are in the range of 6-8 Hz.

3.3.7 Power Frequency Variations

The power system frequency is directly related to the rotational speed of the generators on the system. At any instant, the frequency depends on the balance between the load and the capacity of the available generation. When this dynamic balance changes, small changes in frequency occur. The size of the frequency shift and its duration depends on the load characteristics and the response of the generation system to load changes.

Frequency variations that go outside of accepted limits for normal steady-state operation of the power system are normally caused by faults on the bulk power transmission system, a large block of load being disconnected, or a large source of generation going off-line.

Frequency variations that affect the operation of rotating machinery, or processes that derive their timing from the power frequency (clocks), are rare on modern interconnected power systems. Frequency variations of consequence are much more

likely to occur when such equipment is powered by a generator isolated from the utility system. In such cases, governor response to abrupt load changes may not be adequate to regulate within the narrow bandwidth required by frequency sensitive equipment.

4 MONITORING OBJECTIVES

Power quality monitoring is necessary to characterize electromagnetic phenomena at a particular location on an electric power circuit. In some cases, the objective of the monitoring is to diagnose incompatibilities between the electric power source and the load. In others, it is to evaluate the electrical environment at a particular location to refine modelling techniques or to develop a power quality baseline. In still others, monitoring may be used to predict future performance of load equipment or power quality mitigating devices. In any event, the most important task in any monitoring project is to define clearly the objectives of monitoring.

The objectives of monitoring for a particular project will determine the choice of monitoring equipment, the method of collecting data, the triggering thresholds needed, the data analysis technique to employ, and the overall level of effort required of the project. The objective may be as simple as verifying steady-state voltage regulation at a service entrance, or may be as complex as analyzing the harmonic current flows within a distribution network. The resulting data need only meet the objectives of the monitoring task in order for the monitoring to be successful.

The procedure for defining monitoring objectives differs by the type of study. For diagnostic monitoring to solve shutdown problems with sensitive equipment, the objective may be to capture out-of-tolerance events of certain types. Evaluative or predictive monitoring may require collection of several voltage and current parameters in order to characterize the existing level of power quality.

Measurement of electromagnetic phenomena includes both time and frequency domain conducted parameters, which may take the form of overvoltages and undervoltages, interruptions, sags and swells, transients, phase imbalance, frequency aberrations, and harmonic distortion. Non-conducted environmental factors can also have an effect on load equipment, although these types of disturbances are not considered in this document. Such factors include temperature, humidity, electromagnetic interference (EMI), and radio frequency interference (RFI).

4.1 Need For Monitoring Power Quality

There are several important reasons to monitor power quality. The primary reason underpinning all others is economic, particularly if critical process loads are being adversely affected by electromagnetic phenomena. Effects on equipment and

process operations can include misoperation, damage, process disruption, and other such anomalies. Such disruptions are costly since a profit-based operation is interrupted unexpectedly and must be restored to continue production. In addition, equipment damage and subsequent repair cost both money and time. Product damage can also result from electromagnetic phenomena requiring that the damaged product either be recycled or discarded, both of which are economic issues.

In addition to resolving equipment disruptions, a database of equipment tolerances and sensitivity can be developed from monitored data. Such a database can provide a basis for developing equipment compatibility specifications and guidelines for future equipment enhancements. In addition, a database of the causes for recorded disturbances can be used to make system improvements. Finally, equipment compatibility problems can create safety hazards resulting from equipment misoperation or failure.

Problems related to equipment misoperation can only be assessed if customer disturbance reports are kept. These logs describe the event inside the facility, the type of equipment that was affected, how it was affected, the weather conditions, and the losses incurred.

4.2 Equipment Tolerances And Effects Of Disturbances On Equipment

The tolerance of various equipment needs to be considered in power quality monitoring. A specific type of equipment, such as an ASD (Adjustable Speed Drive), may be sensitive to an overvoltage or undervoltage condition, for example, while there may also be a significant variation to the same phenomena between ASDs built by other manufacturers. Power quality monitoring should attempt to characterize individual process equipment by matching monitoring results with reported equipment problems. This characterization of individual loads will show which equipment needs protection, and the level of protection required.

4.3 Equipment Types

Although there may be a wide variety in the response of specific equipment types manufactured by different companies, there may be some similarity in the response of certain types of equipment to specific disturbance parameters. In any case, it is useful to consider certain specific equipment types or groupings in terms of their immunity to power quality disturbances.

4.4 Effect On Equipment By Phenomena Type

Clause 2.3 defines seven major categories of electromagnetic phenomena. The following subclauses describe the observed effects of these phenomena on the operation of various types of equipment.

4.4.1 Transients

Transient voltages caused by lightning or switching operations can result in degradation or immediate dielectric failure in all classes of equipment. High magnitude and fast rise time contribute to insulation breakdown in electrical equipment like rotating machinery, transformers, capacitors, cables, CTs, PTs, and switchgear. Repeated lower magnitude application of transients to these equipment type cause slow degradation and eventual insulation failure, decreasing equipment mean time between failure (MTBF). In electronic equipment, power supply component failures can result from a single transient of relatively modest magnitude. Transients can also cause nuisance tripping of adjustable speed drives due to the dc link overvoltage protection circuitry.

4.4.2 Short Duration Variations

The most prevalent problem associated with interruptions, sags, and swells is equipment shutdown. In many industries with critical process loads, even instantaneous short duration phenomena can cause process shut-downs requiring hours to restart. In these facilities, the effect on the process is the same for a short duration variation as for long duration phenomena.

Monitoring is important because it is often difficult to determine from the observable effects on customer equipment which electromagnetic phenomena caused the disruption. Further, solution alternatives are much different if the equipment is being affected by sags, for instance, rather than by interruptions.

4.4.2.1 Interruptions

Even instantaneous interruptions may affect electronic and lighting equipment causing misoperation or shut-down. Electronic equipment includes power and electronic controllers, computers, and the electronic controls for rotating machinery. Momentary and temporary interruptions will almost always cause equipment to stop op-

erating, and may cause drop-out of induction motor contactors. In some cases, interruptions may damage electronic soft-start equipment.

4.4.2.2 Sags

Short duration sags, in particular, cause numerous process disruptions. Often, the sag is sensed by electronic process controllers equipped with fault-detection circuitry, which initiates shutdown of other, less-sensitive loads. A common solution to this problem is to serve the electronic controller with a constant-voltage transformer, or other mitigating device, to provide adequate voltage to the controller during a sag. The application challenge is to maintain the electronic controller during sags that will not damage process equipment protected by the fault circuitry, while simultaneously reducing nuisance shutdowns.

Electronic devices with battery backup should be unaffected by short duration reductions in voltage. Equipment such as transformers, cable, bus, switchgear, CTs and PTs should not incur damage or malfunction due to short duration sags. A slight speed change of induction machinery and a slight reduction in output from a capacitor bank can occur during a sag. The visible light output of some lighting devices may be reduced briefly during a sag.

4.4.2.3 Swells

An increase in voltage applied to equipment above its nominal rating may cause failure of the components depending upon the frequency of occurrence. Electronic devices, including adjustable speed drives, computers, and electronic controllers, may show immediate failure modes during these conditions. However, transformers, cable, bus, switchgear, CTs, PTs, and rotating machinery may suffer reduced equipment life overtime. A temporary increase in voltage on some protective relays may result in unwanted operations while others will not be affected. Frequent voltage swells on a capacitor bank can cause the individual cans to bulge while output is increased from the bank. The visible light output from some lighting devices may be increased during a temporary swell. Clamping type surge protective devices (e.g., varistors or silicon avalanche diodes) may be destroyed by swells exceeding their MCOV (Maximum Continuous Operating Voltage) rating.

4.4.3 Long Duration Variations

Variations in supply voltage lasting longer than 1 min can cause equipment problems. Overvoltage and undervoltage problems are less likely to occur on utility feeders, as most utilities strive to maintain $\pm 5\%$ voltage regulation. Overvoltage and undervoltage problems can occur, however, due to overloaded feeders, incorrect tap settings on transformers, blown fuses on capacitor banks, and capacitor banks in service during light load conditions. Sustained interruptions can result from a variety of causes, including tripped breakers, blown fuses, utility feeder lockouts, and failed circuit components.

4.4.3.1 Sustained Interruptions

The effect of a sustained interruption is equipment shutdown, except for those loads protected by UPS systems, or other forms of energy storage devices.

4.4.3.2 Undervoltages

Undervoltages in excess of 1 min can also cause equipment to malfunction. Motor controllers can drop out during undervoltage conditions. The dropout voltage of motor controllers is typically 70-80% of nominal voltage. Long duration undervoltages cause an increased heating loss in induction motors due to increased motor current. Speed changes are possible for induction machinery during undervoltage conditions. Electronic devices such as computers and electronic controllers may stop operating during this condition. Undervoltage conditions on capacitor banks result in a reduction of output of the bank, since var output is proportional to the square of the applied voltage. Generally, undervoltage conditions on transformers, cable, bus, switchgear, CTs, PTs, metering devices, and transducers do not cause problems for the equipment. The visible light output from some lighting devices may be reduced during undervoltage conditions.

4.4.3.3 Overvoltages

Overvoltages may cause equipment failure. Electronic devices may experience immediate failure during the overvoltage conditions; however, transformers, cable, bus, switchgear, CTs, PTs, and rotating machinery do not generally show immediate failure. Sustained overvoltage on transformers, cable, bus, switchgear, CTs, PTs and rotating machinery can result in loss of equipment life. An overvoltage condition on

some protective relays may result in unwanted operations while others will not be affected. A sign of frequent overvoltage conditions on a capacitor bank is the bulge of individual cans. The var output of a capacitor will increase with the square of the voltage during an overvoltage condition. The visible light output from some lighting devices may be increased during overvoltage conditions.

4.4.4 Voltage Imbalance

In general, utility supply voltage is maintained at a relatively low level of phase imbalance since even a low level of imbalance can cause a significant power supply ripple and heating effects on the generation, transmission, and distribution system equipment. Voltage imbalance more commonly emerges in individual customer loads due to phase load imbalances, especially where large, single-phase power loads are used, such as single-phase arc furnaces. In these cases, overheating of customer motors and transformers can readily occur if the imbalance is not corrected. Phase current imbalance to three-phase induction motors varies almost as the cube of the voltage imbalance applied to the motor terminals. The effects on other types of equipment are much less pronounced, although significant imbalance can cause loading problems on current-carrying equipment such as bus ducts. Desirable levels of imbalance are less than 1% at all voltage levels to reduce possible heating effects to low levels.

Utility supply voltages are typically maintained at less than 1%, although 2% is not uncommon. Voltage imbalance of greater than 2% should be reduced, where possible, by balancing single-phase loads as phase current imbalance is usually the cause. Voltage imbalance greater than 2% may indicate a blown fuse on one phase of a three-phase capacitor bank. Voltage imbalance greater than 5% can be caused by single-phasing conditions, during which one phase of a three-phase circuit is missing or de-energized. Phase monitors are often required to protect three-phase motors from the adverse affects of single phasing.

4.4.5 Waveform Distortion

Harmonic current injection from customer loads into the utility supply system can cause harmonic voltage distortion to appear on the utility system supply voltage. This harmonic current and voltage distortion can cause overheating of rotating equipment, transformers, and current-carrying conductors, premature failure or opera-

tion of protective devices (such as fuses), harmonic resonance conditions on the customer's electric power system, which can further deteriorate electrical system operation, and metering inaccuracies. Harmonic voltage distortion on a utility system can cause the same problems to a customer's equipment and can cause overheating of utility transformers, power-carrying conductors, and other power equipment. Typical harmonic current limits for customers and harmonic voltage limits for utility supply voltage that customers and utilities in general should attempt to operate within in order to minimize the effects of harmonic distortion on the supply and end-user systems is outlined in (7).

4.4.6 Voltage Fluctuations

Fluctuations in the supply voltage are most often manifested in nuisance variations in light output from incandescent and discharge lighting sources. A sudden voltage decrease of less than 1/2% can cause a noticeable reduction in light output of an incandescent lamp and a less noticeable reduction in light output of gaseous discharge lighting equipment. Voltage fluctuations less than 7% in magnitude have little effect on other types of customer loads (13).

4.4.7 Power-Frequency Variations

In general, utilities maintain very close control of the power system frequency. Slight variations in frequency on an electric system can cause severe damage to generator and turbine shafts due to the subsequent large torques developed. In addition, cascading system separations can result with even slight deviations in frequency since electric systems are closely connected and operate in synchronism. Frequency variations are more common on customer-owned generation equipment systems. Generator over-speed can result in a frequency increase on small systems operating independent of utility sources.

Frequency synchronization errors can sometimes occur on a customer feeder that serves large rectifier loads. These loads can cause voltage notching severe enough to register extra zero crossing events on electronic loads that count zero crossings of the ac voltage to obtain frequency. While these events are recorded as frequency errors by electronic controllers, the fundamental frequency has not changed.

5 MEASUREMENT INSTRUMENTS

Instruments used to monitor electromagnetic phenomena can be as simple as an analog voltmeter to an instrument as sophisticated as a spectrum analyzer. Selecting and using the correct type of monitor requires the user to understand the capabilities and limitations of the instrument, its responses to power system variations, and the specific objectives of the analysis. This clause will focus on the capabilities and limitations of various monitoring equipment.

Instrument features required are dependent on the monitoring location and objectives. If assessing power quality at the service entrance, for example, the emphasis may be only on long-term steady-state conditions and utility-transmitted anomalies. The level of detail required – rms voltage stripcharts or high-speed waveform captures – is indicated by the type of phenomena likely to be causing problems.

As can anyone understand there is a very wide range of options regarding the final choice of the proper equipment that can satisfy one's needs to monitor his facility's power quality. It is not the purpose of this paper to numerate all the available instruments regarding the monitoring of power quality, however a small review, including some examples, is useful in order for the reader to acquire a general idea. Therefore some pictures and specifications of the most representative equipment have been gathered and are being presented later on.

Basically there are three main categories of power quality monitoring equipment:

- handheld devices
- portable devices and
- fixed systems

All the instruments presented below are manufactured from the same company, Dranetz-BMI. Of course this has no significance rather than to preserve a continuity among the instruments. More information regarding power quality monitoring equipment from other worthy manufacturing companies is available (27-45).

5.1 Handheld Devices

5.1.1 Analyst 2060

This clamp-on power quality and harmonics meter from Dranetz-BMI and LEM combines the features of a power meter, scope and recorder and moreover has a very good shielding that makes it suitable for use in areas with noisy, fast-switching power electronics (22).



Figure 15. Analyst 2060

Specifications for the Analyst 2060

Voltage Measurements	<ul style="list-style-type: none">• 600V AC RMS or DC between input terminals.
Current Measurements	<ul style="list-style-type: none">• 600V AC RMS or DC between uninsulated conductor and ground
Range	<ul style="list-style-type: none">• Voltage: 4V, 40V, 400V, 750V• Current: 40A, 400A, 2000A• Watts: 4kW, 40kW, 400kW, 1200kW• VA: 4kVA, 40kVA, 400kVA, 1200kVA• VAR: 4kVAR, 40kVAR, 400kVAR, 850kVAR• PF: 0.3 cap to 0.3 ind• kWhr: 4, 40, 400, 4000, 40,000
Environment	<ul style="list-style-type: none">• 0 - 50C (operating)• -20 - 60C (storage)
Power	<ul style="list-style-type: none">• 6 x AA Alkaline MN1500, LR6• Typical 24 Hr operation
Certifications	ISO 9001

5.1.2 Analyst Q70

The single phase Analyst Q70 power quality analyzer from Dranetz-BMI and LEM comes equipped with Dran-View software for enhanced data presentation and analysis. The Analyst Q70 combines the functionality of an EN50160-compliant power quality meter, harmonic spectrum analyzer, oscilloscope and data logger in a single handheld instrument (23).



Figure 16. Analyst Q70

Specifications for the Analyst Q70

Voltage Measurements	<ul style="list-style-type: none"> • 600V AC RMS or DC between live terminals, or live and ground.
Current Measurements	<ul style="list-style-type: none"> • 600V AC RMS or DC between live terminals
Range	<ul style="list-style-type: none"> • Current: 30A, 300A, 3000A • Voltage: 115/230/480 • Watts: 3.4kW, 1,440kW • VA: 3.4kVA, 1440kA • VAR: 3.4kVAR, 1440kAR • PF: 0.3 cap to 0.3ind • WHr: 0 to 10TWHr
Environment	<ul style="list-style-type: none"> • 0C to 50C (operating) • -20C to 60C (storage)
Certifications	ISO9001

5.1.3 PowerXplorer PX5-400

The PowerXplorer integrates the most advanced feature set available in a power monitoring instrument, with an easy-to-navigate, color graphical user interface. With high-speed sampling and data capture (1 microsecond/channel), this 8-channel workhorse simultaneously captures and characterizes thousands of parameters, using a range of standard and customizable operating modes. The unique measurement capabilities of the PowerXplorer include capture of low-medium-high frequency transients through peak, waveshape, rms duration and adaptive high-speed sampling, as well as power measurements to clearly characterize non-sinusoidal and unbalanced systems (21).



Figure 17. PowerXplorer PX5-400

Specifications for the PowerXplorer PX5

Measured Parameters	<ul style="list-style-type: none"> • (4) differential inputs, 1-600 Vrms, AC/DC, 0.1% rdg, 256 samples/cycle, 16 bit ADC • (4) inputs with CTs 0.1-6000 Arms CT-dependent, AC/DC, 0.1%rdg + CTs, 256 samples/cycle, 16 bit ADC • 1 MHz High Speed Sampling, 14 bit ADC, 1%FS • Frequency Range, 10m Hz resolution, 45-65 Hz or 380-420 Hz • Phase Lock Loop - Generator tracking • Phase Lock Loop - Standard PQ mode
Power Quality Triggers	<ul style="list-style-type: none"> • Cycle-by-cycle analysis • 256 samples/cycle; 1/2 cycle RMS steps (1) • L-L, L-N, N-G RMS Variations:

	<p>Sags/swells/interruptions)</p> <ul style="list-style-type: none"> • RMS Recordings V&I (32 pre-fault, 10K post-fault cycles) • Waveshape Recordings (2/6/2 cycles) • Low and Medium Frequent Transients - V&I • High Frequency Transients - V&I, 3% FS trigger (1) • Harmonics Summary Parameters • Cross trigger V & I channels • RMS Event Characterization (IEEE or IEC) • Transient Event Characterization (1)
Distortion/Power/Energy	<ul style="list-style-type: none"> • W, VA, VAR, TPF, DPF, Demand, Energy, etc. • IEEE 1459 Parameters of distorted and unbalanced • Harmonics/Interharmonics per IEC 1000-4-7 • THD/Harmonic Spectrum (V,I,W) to 63rd • TID /Interharmonic Spectrum (V,I) to 63rd • Flicker per IEC 1000-4-15 (Pst,Plt,Sliding Plt) • Crest Factor, K Factor, Transformer Derating Factor, Telephone Interference Factor • Unbalance (max. rms deviation) & sequencing components • 5 User Spec Harmonics or Signaling Frequency • Vector/Arithmetic/Coincident Parameters
General Specifications	<ul style="list-style-type: none"> • Size (HxWxD): 12" x 2.5" x 8" (30cm x 6.4cm x 20.3 cm) • Weight: 4.2 pounds (1.9 kg) • Operating Temperature: 0 -50 (32 to 122) • Storage Temperature: -20 to 55 (4 to 131 F) • Humidity: 10 to 90% non-condensing • System Time Clock-Crystal controlled-1 second resolution • Charger /Battery Eliminator: 90-264 VAC 47-63 Hz • Display: LCD color touch screen

	<ul style="list-style-type: none"> • Memory options (must have one): 32M-128M removable compact flashcard
--	--

5.1.4 PowerGuide 4400

The PowerGuide 4400 is equipped with 8 independent channels. Automated setups provide instant detection of circuits and configurations, ensuring that the instrument is ready to successfully collect data. Users can select the length and mode of data collection, including troubleshooting, data logging, power quality surveys, energy and load balancing (26).



Figure 18. PowerGuide 4400

Specifications for the PowerGuide 4400

Measured Parameters	<ul style="list-style-type: none"> • (4) differential inputs, 1-600V rms, AC/DC, 0.1% rdg , 256 samples/cycle, 16bit ADC • (4) inputs with CTs 0.1-3000A rms, AC/DC, 256s/c, 0.1% rdg + CTs, 16 bit ADC • Frequency range, 10mHz resolution, 45-65 Hz • Phase Lock Loop - Standard PQ mode
Monitoring/Compliance Modes	<ul style="list-style-type: none"> • IEEE 1159 • IEC 61000-4-30 • EN50160

	<ul style="list-style-type: none"> • Current Inrush / Energization • Voltage Fault recording • Quality of Supply (EN50160 or annunciator) • Long Term Monitoring • Continuous Data Logging w/min/max/avg
Power Quality Triggers	<ul style="list-style-type: none"> • Cycle-by-cycle analysis • 256 samples/cycle; 1/2 cycle RMS steps • L-L, L-N, N-G RMS Variations: Sags/swells/interruptions) • RMS Recordings V& I (x/x/x cycles) • Waveform Recordings (2/6/2 cycles) • Low and Medium Freq Transients - V & I • Harmonics Summary Parameters • Cross trigger V & I channels • RMS Event Characterization (IEEE or IEC)
Distortion/Power/Energy	<ul style="list-style-type: none"> • W, VA, VAR, TPF, DPF, Demand, Energy • Harmonics/Interharmonics per IEC 1000-4-7 • THD/Harmonic Spectrum (V,I,W) to 63rd • TID/Interharm. Spectrum (V,I) to 63rd • Flicker per IEC 1000-4-15 (Pst,Plt,Sliding Plt) • CF, KF, TDF, TIF • Unbalance (rms dev. & sequencing components)
General Specifications	<ul style="list-style-type: none"> • Size (HxWxD): 12" x 2.5" x 8" (30cm x 6.4cm x 20.3 cm) • Weight: 3.8 pounds(1.8 kg)

	<ul style="list-style-type: none"> • Operating Temperature: 0 -50 (32 to 122) • Storage Temperature: -20 to 55 (4 to 131 F) • Humidity: 10 to 90% non-condensing • System Time Clock: Crystal controlled, 1 second resolution • Charger /Battery Eliminator: 90-264 VAC 47-63 Hz • Display: LCD color touch screen • Memory options (must have one): 32M-128M removable compact flashcard
--	---

5.1.5 Power Platform PP-4300

The Power Platform PP-4300 has four differential voltage channels and four independent current channels, measures, analyzes and records power quality, harmonics and energy data simultaneously and continuously. Unique to the PP4300 are interchangeable TASKCards that expand the capabilities of the instrument within the same mainframe. TASKCards can transform the PP4300 from a single-phase troubleshooting tool into a full three-phase instrument for advanced power quality management or specific analysis such as inrush or fault recording (25).



Figure 19. Power Platform PP-4300

Specifications for the Power Platform PP-4300

Voltage Measurements	<ul style="list-style-type: none"> • 4 fully differentiated channels • 100-600Vrms; user selected 0.5-20Vrms on one channel • Accuracy: $\pm 1\%$ reading $\pm 0.05\%$ full scale
----------------------	--

Voltage Transients	<ul style="list-style-type: none"> • 50-1000Vpk; user selected • 1-30Vpk on one channel • 1 microsecond minimum duration • Accuracy: $\pm 10\%$ reading $\pm 1\%$ full scale (requires Task-Card PQLite)
Current Measurement	<ul style="list-style-type: none"> • 4 fully independent current channels • 10 - 200% of full-scale current probe rating • Accuracy: $\pm 1\%$ reading $\pm 0.05\%$ full scale • (at fundamental, plus current probe accuracy)
Current Transients	<ul style="list-style-type: none"> • 10-300% CT full scale except Chan D 2-200% CT full scale • 1 microsecond minimum duration • Accuracy: $\pm 10\%$ reading $\pm 1\%$ full scale plus probe • NOTE: Requires TASKCard PQLite H-T, PQLite H-T-M or H-T-E-M
Frequency	<ul style="list-style-type: none"> • Fundamental range 35 - 60 Hz (For frequencies outside the 50-60Hz range, contact the factory before ordering) • Accuracy $\pm 0.2\%$ of reading
Update Rates	<ul style="list-style-type: none"> • All parameters updated once per second (Harmonic-based parameters updated every 5 seconds)
Environment	<ul style="list-style-type: none"> • 41°F to 113°F • +5°C to +45°C • Humidity 10% - 90% non-condensing
Battery	<ul style="list-style-type: none"> • 2 hours operation • 3 hours full recharge (continuous operation from battery eliminator)
Certifications	CE, FCC, ISO-9001

5.2 Portable Devices

5.2.1 658 Power Quality Analyzer

The 658 can be used to monitor power disturbances and harmonics, as well as other sources of disruption like temperature, humidity, and radiated RF noise, making it the ideal tool for field service and site surveys (29).



Figure 20. 658 Power Quality Analyzer

Specifications for the 658 Power Quality Analyzer

Voltage & Current Measurement	4 fully differential channels One AC/DC voltage channel and 3 voltage/current channels 0-600Vrms; 0-3000A, depending on current probe selected Accuracy: $\pm 1\%$ reading $\pm 0.2\%$ full scale
Impulse Measurement	2.4 to 6144 Vpeak 2.4 to 6000 Amps pk 1 microsecond minimum duration
Sampling Rate	7.2kHz/channel for RMS 1.85MHz/channel for impulse
Frequency	Fundamental range 45 - 65 Hz Optional Model 658-400 45-65Hz or 310-445 Hz
Optional Inputs	8 independent differential channels Configurable as voltage or current 0-10Vdc, 0-20mA DC Sampled at 12.5 Hz
Battery	Standard, 5 minutes operation typical
PC Software	Dran-View

Certifications	UL, CE, FCC, ISO-9100 (CE certification for model 658E, UL listing for model 658)
----------------	--

5.2.2 Power Platform PP1

The PP1 has the capability of changing its functions by inserting a different Dranetz-BMI TASKCard each time. TASKCards include PQPlus, 8000 (Energy/Harmonics), Inrush, and TASKCard Flicker (28).



Figure 21. Power Platform PP1

Specifications for the Power Platform PP1

Voltage Measurements	4 fully differential channels 10-600V _{rms} Accuracy: $\pm 1\%$ reading $\pm 0.05\%$ full scale
Voltage Transients	50-6000V _{pk} 1 microsecond minimum duration Accuracy: $\pm 10\%$ reading $\pm 1\%$ full scale Requires TASKCard PQPlus
Current Measurements	4 fully independent current channels 10 - 200% of full-scale current probe rating Accuracy: $\pm 1\%$ reading $\pm 0.05\%$ full scale (at fundamental, plus current probe accuracy)
Current Transients	10-300% CT full scale except Chan D 2-200% CT full scale 1 microsecond minimum duration

	Accuracy: $\pm 10\%$ reading $\pm 1\%$ full scale plus probe Requires TASKCard PQPlus
Frequency	Fundamental range 30 - 450 Hz Accuracy $\pm 0.2\%$ of reading
Update Rates	All parameters updated once per second (Harmonic-based parameters updated every 5 seconds)
PC Software	Dranview

5.3 Fixed Systems

5.3.1 3100 PQ Pager

The 3100 PQNode is a power quality tool for utility key account programs. It has the ability to communicate directly to account executives, managers, and customers through its built-in voice-mail-like interface. It can also send messages directly to up to 4 pagers thereby bypassing the traditional bottleneck of a utility master station (17).

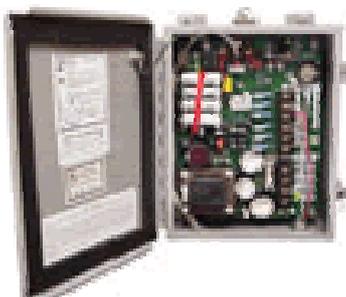


Figure 22. 3100 PQ Pager

Specifications for the 3100 PQ Pager

Voltage Inputs	Phase A, B, C, Neutral
Power Types Supported	Single phase, split single phase, 3-phase 4-wire wye, 3-phase 3-wire delta
Measurement Range	0-150 Vrms, ± 250 Vpk or 0-600 Vrms, ± 1000 Vpk (user selected)
Accuracy	$\pm 1.0\%$ of reading ± 0.5 digit
Sampling	Continuous; 32 samples/cycle 50/60 Hz
Events Recorded	Sags, swells, interruption, restored, transient, contact closure(two), min/average/max RMS strip chart Event signatures

	Pre- and post-samples recorded: 128 RMS samples for RMS sags, swells 4 cycles of waveform data for transients
Contact Closure Input	Two inputs for dry contacts; NO or NC (user selected)
Event Memory	Sufficient for at least 32 events, including event signatures, plus at least 30 days of strip chart
Operating Environment	NEMA 4X, -40°F to 122°F -40°C to +50°C
Instrument Power	110V-240V or 220V-480V, user selected 50/60Hz Internal automatically charged 15-minute batteries Memory contents preserved for 10 years
Certifications	UL, FCC, CE, and ISO-9001

5.3.2 7100/7100S

The 7100 PQNode can be permanently installed at a site, or transported from site to site as a portable instrument. 7100 PQNodes can be configured to monitor power quality, power flow, or harmonics, depending on site-specific needs (18).



Figure 23. 7100/7100S

Specifications for the 7100/7100S

Voltage inputs	Phase A, B, C, Neutral, and Ground (supports single phase, split single phase, 3-phase 4-wire wye, 3-phase 3-wire delta)
Current inputs	Phase A, B, C, and Neutral using optional current probes
Voltage measurements	0 - 600 Vrms ($\pm 1,000$ Vpk), 50/60 Hz Accuracy: $\pm 0.5\%$ reading $\pm 0.35\%$ full scale
Voltage transients	100 - 1500 Vpk Peak detection: IEEE 587 type A and B Accuracy: $\pm 5\%$ reading $\pm 5\%$ full scale

Current measurements	4 fully independent current channels 0 - 3000 Amps rms (depends on probe selected) Accuracy: $\pm 1\%$ reading $\pm 1\%$ full scale
Sampling rate	128 samples per cycle, continuous on all voltage and current channels
Harmonic measurements	Through 49th harmonic at 50/60 Hz
User-Selected Modes	Power quality: snapshots,rms strip charts, sags, swells, impulses, waveshape faults, kW and kVA demand, harmonic snapshots to the 49th, kWh Power flow: Vrms, Irms, W, VA, kW and kVA demand,kWh,kVAR,PF,dPF, Vthd,Ithd Harmonics triggering: Vthd, Ithd, Vn and In stripcharts, triggering on individual harmonics to the 49th,1-cycle or 4-cycle FFT's
Environment	Rain-,dust-,ice-resistant enclosure -4°F to 113°F (-20°C to +45°C) Humidity 0% - 95% non-condensing
Battery	10 minute UPS built in External battery supported
Software	PES

5.4 Series 5500 DualNode

Now that an adequate number of power quality monitoring equipment has been presented and the reader has gained a sufficient knowledge it is time to present the equipment that has been used on this paper in order to acquire the experimental data.

5.4.1 Signature System

The Signature System is a new vision for distributed electric power information systems. It is based on a few simple premises:

- ⇒ Capture all data.
- ⇒ Convert the data to information.
- ⇒ Manage the information while saving the data.
- ⇒ Move the information to those who need it, when they need it.
- ⇒ Share the information.
- ⇒ Provide answers, not just data.
- ⇒ Eliminate installed software.
- ⇒ Use the Internet.

A typical Signature System is built from several DataNodes, plus one or more InfoNodes equipped with a selection of Answer Modules. Large Signature Systems may also include NodeLink or NodeCenter, a powerful suite of server-based management analysis tools for enterprise-wide systems.

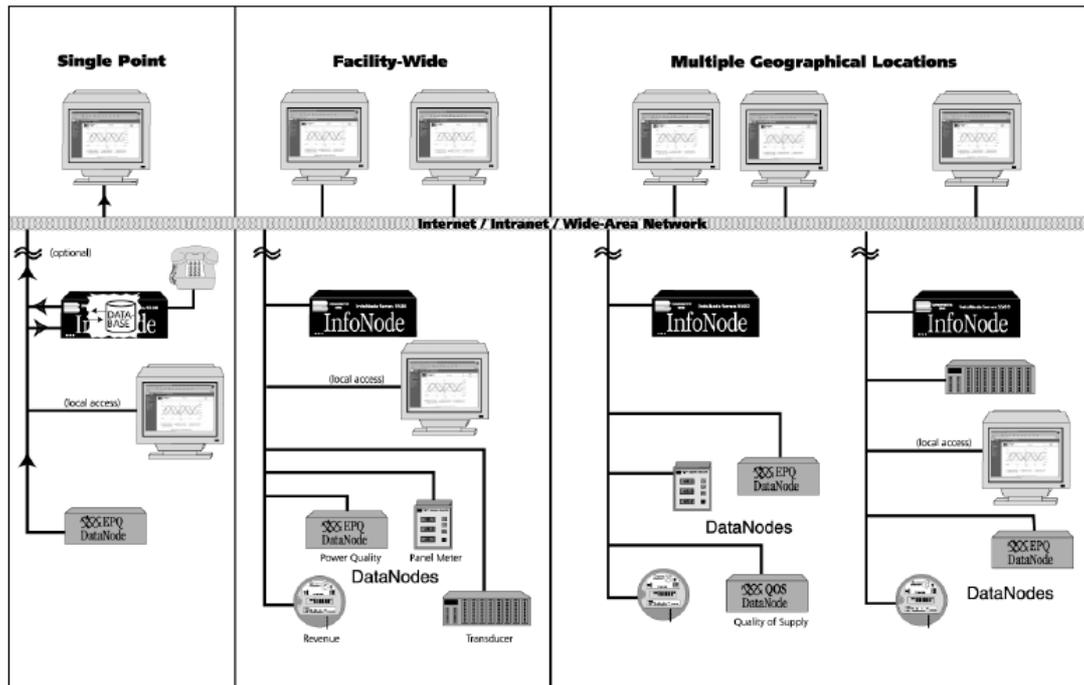


Figure 24. Signature System Architecture: A Conceptual Illustration

DataNodes gather readings from circuits and processes. Inexpensive, small and easy to install, DataNodes have the intelligence to convert raw readings into useful data. They communicate their data to InfoNodes through RS-485 or Ethernet links.

InfoNodes gather DataNode data, convert the data to information, and manage and communicate the information. They form the central component of the whole Signature System. InfoNodes are equipped with Answer Modules which convert information into application-specific answers using patented and proprietary expertise developed by Electrotek Concepts, Electric Power Research Institute (EPRI) and Dra-netz-BMI. These plug-in application-specific answers will cover applications from identifying power disturbance origins, to reporting based on evolving standards, to predicting maintenance schedules at substations.

5.4.2 InfoNode: The Central Component

The InfoNode is the central component of the SignatureSystem, “the System that Learns from the Past, to Inform You in the Present, and Prevents Problems in the Future.” The Signature System consists of one or more DataNodes (the data acquisition modules), connected to an InfoNode (the data storage and analysis module). Optionally, the data and information from multiple InfoNodes can be combined at the enterprise level using the NodeLink or NodeCenter software. InfoNodes can be ac-

cessed through either LAN or modem connections. The InfoNode provides the user interface through a self-contained web server. This frees the user from having to load software onto the user's PC or laptop. Access is possible from anywhere in the world, through the Intranet, Internet, or via a modem, with only a standard web browser (Microsoft Internet Explorer V5.5 or newer or Netscape Navigator V4.75) necessary. Access time is dependent primarily on the communication media, with a direct network connection being the fastest. The InfoNode can also provide GPS time synchronization to the DataNodes with the optional GPS module. Some of the most important options are the software Answer Modules. These options can be part of the initial purchase or easily added later. Data from one or more DataNodes is analyzed to provide such answers as the direction of the PF cap switching transient (upstream or downstream), sag directivity, location of faults on radial feeders, reliability-benchmark indices for power quality, and different characterizations of data, such as QOS (Quality of Supply), IEEE 1159, EPRI DPQ.

5.4.3 Signature System InfoNode Graphical User Interface

The InfoNode user interface consists of a series of tab pages. The pages are labeled as follows: Home, Views, Reports, Real-time, and Setup. Each tab page has its own tree directory located in the left window pane. The tree can be expanded or collapsed. Click on the plus (+) sign to expand the tree and show more options available. Click on the minus (-) sign to collapse the tree one level backward. A folder is empty if it does not display a plus or minus sign. All detailed tab page information is found in the right window pane. The InfoNode system provides a direct, no-fuss interface which displays information called out in tab, hyperlink and button format. Each tab is provided with a Help option to provide users with immediate, onscreen assistance. Below is a sample screen showing the five main tab pages of the Signature System InfoNode.

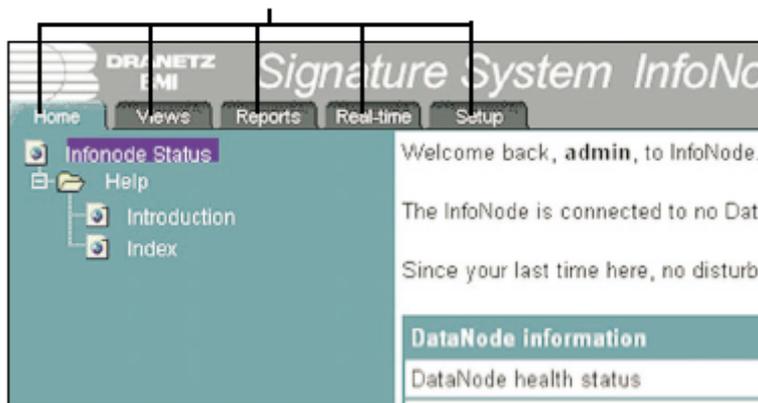


Figure 25. Main menu tabs of the InfoNode System

The Home page provides basic status information about the InfoNode and DataNodes connected, along with easy access to the first, last, and most recent events in memory.

The Views page provides access to three interactive sections: the QOS (Quality of Supply) Status, Timeline, and Smart Views. The QOS module is designed to monitor and report quality of supply compliance as specified by European Standard EN50160. QOS Status will appear in InfoNode systems that have QOS data acquisition modules (5560 DataNode) in it. The Timeline is a two pane browser, with the timeplot of selected parameters and channels in the top pane, and the event list and details (wavershapes) in the lower pane. The Smart Views include: 3D RMS Mag/Dur (Magnitude/ Duration), RMS Mag/Dur, Smart Trends, Event Summary, RMS Variations, Snapshots, and Transients.

The Reports page is used to generate reports formatted for direct printing, through Smart Reports and Standard Reports. Smart Reports have pre-selected output formats and include: DataNode Summary, Voltage Quality, Energy & Demand, Event Summary, and Top 10 Events. Standard Reports have output formats that can be customized by the user and include: Event Summaries, Top10 Events, Event Statistics, Quality of Supply, Waveform Distortion, Energy & Demand, and InfoNode Summary. The Answer Module is a customized facility which enables you to identify the source, cause and time of faults or disturbances like sags and swells. The system is able to record and document the source of the problem, whether coming from inside your facility or in the supply from your power supplier.

The Real-time page displays real time metered data in one of three formats: Meter Panel, Meter Dial and Scope Mode. Meter Panel shows a textual list of metered

parameters for the selected DataNode. Parameters displayed are those configured for logging and trending. Meter Dial shows the same information as Meter Panel but in an analog meter dial format. Scope Mode shows real time waveforms for all enabled channels in an oscilloscope type of display. Note that Scope Mode is not available for all DataNode types.

The Setup page allows the user to configure both the InfoNode and any DataNodes connected to it. Additional users and their access permissions and passwords are programmed on this page. Additional DataNodes connected to the InfoNode are also set up on this page. Other parameters which you can view and/or customize (depending on your user access privilege) are: Notifications, Communications, Answer Module, and DataNodes.

5.4.4 The EPQ DataNode Series

Signature System, as mentioned earlier, consists of one or more DataNodes (the data acquisition modules) connected to an InfoNode (the data storage, analysis module and web server). The EPQ DataNode models are designed with comprehensive and PQ-optimized data acquisition capabilities for power quality related disturbances or events. The trigger and capture mechanisms include RMS variations, peak transients, waveshapes, steady-state parameter limits, sensitivity changes, harmonics and more. From microsecond transients on voltage and current to cold load pick-up after a sustained interruption, the EPQ DataNode will provide the data and information needed to help determine the cause and severity of the event. In addition, the EPQ DataNodes can be set up to collect and trend and trigger on values from a list of hundreds of parameters, including power and energy related parameters. EPQ DataNodes are available in a wide variety of functions and configurations.

5.4.5 Series 5500 DualNode: the absolute combination

Summarizing, the Series 5500 DualNode is the result of the mixture of everything written earlier. Thus it performs the combined functions of the InfoNode and the EPQ DataNode. It monitors and acquires power quality data, then converts and manages data into information. The 5590 powerhouse does powerquality monitoring and energy management combined with analysis, notification, and web server functions all in one. The InfoNode provides the user interface for the 5590 through a self-contained web server.



Figure 26. A 5590 DualNode



Figure 27. Voltage Modules



Figure 28. Current Modules

Features and Highlights of the Series 5500 DualNode:

- No specialized software or hardware necessary. The 5590 DualNode visualization and analysis tools, as well as standard or customized reports, are integrated into the InfoNode tab pages accessible through a web browser.
- Advanced triggering and database management capture all critical information – from submicrosecond transients to long duration interruptions.
- Adaptive sampling techniques provide min/max/avg RMS values, updated every half cycle, along with up to 22 continuous waveform cycles of event data.
- Harmonic, interharmonic, and power parameters are calculated using internationally recognized standard algorithms.
- An internal UPS keeps the 5590 collecting critical data, even when power is interrupted due to fault conditions.
- Options such as GPS time synch and cross triggering can expand the functionality of the 5590 into a larger scale system.
- Data can be permanently archived or uploaded via NodeLink for additional analysis in DranView.

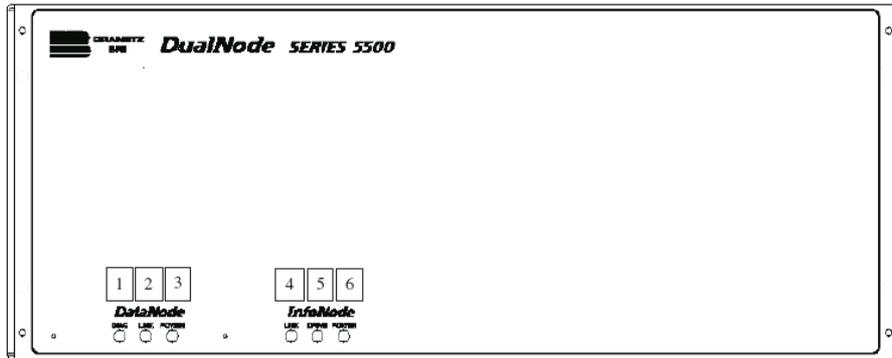


Figure 29. DualNode front panel

DataNode

1. DIAG- Indicator lamp will be on during start-up and periodic health-check diagnostics.
2. LINK- Indicator lamp will flash when the unit is responding to network requests.(Except when connected via AUI port.)
3. POWER- Indicator lamp will flash in a heartbeat fashion when the unit is operating normally.

Note: All three lamps will flash simultaneously when the unit is in Administrator mode.

InfoNode

4. LINK- Indicator lamp will flash when the unit is responding to network requests.
5. DRIVE- Indicator lamp will flash when the unit's database is being read or written to.
6. POWER- Indicator lamp will flash in a heartbeat fashion when the unit is operating normally.

Note: All three lamps will flash simultaneously when the unit is in Administrator mode.

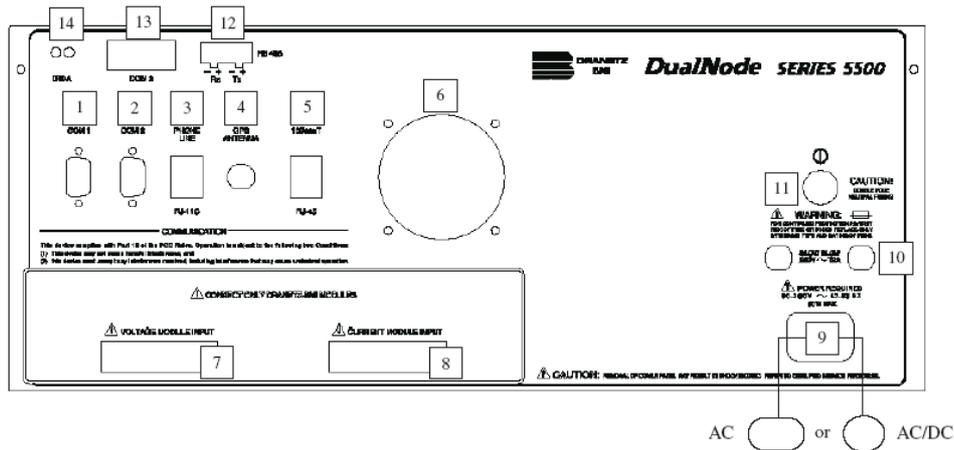


Figure 30. DualNode rear panel

1. COM 1- Serial communications port 1. Use for connection to set InfoNode IP.
2. COM 2- Serial communications port 2.
3. Phone Line (RJ-11C) - Allows modem communication via telephone line if optional modem is installed.
4. GPS Antenna - Allows connection of GPS antenna if optional GPS is installed.
5. 10BaseT (RJ-45)- Allows connection to Ethernet.
6. Cooling fan- Runs continuously while unit is on.
7. Voltage Pod- Allows connection to voltage pod via interface cable.
8. Current Pod- Allows connection to current pod via interface cable or optional CT adapter (BNCTO55,TRTO55).
9. Line power.
 1. AC only power version – 90 – 250V ac, 47-63 Hz.
 2. AC/DC power version - 90– 250V ac, 47-63 Hz.
-105-125V dc, 60W max.
10. Fuses- Slow blow, 250V ac T2A.
11. Power switch - Press to turn unit power on or off. Power indicator lamp on front panel will glow while unit is on.

Note: Power indicator lamp will glow for approximately 5 seconds after unit is powered off.

12. RS 485- Not activated at this time.

13. COM 3- Serial communications port 3. Use for connection to set DataNode IP.

14. IRDA- Not activated at this time.

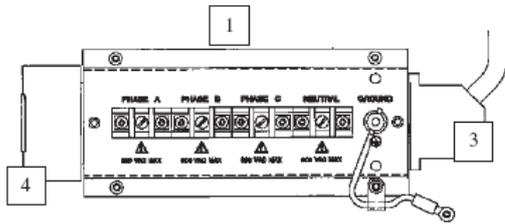


Figure 31. Voltage input pod

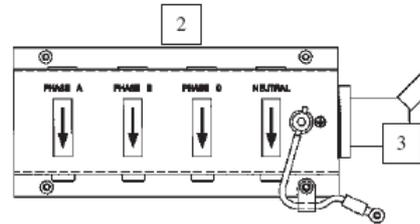


Figure 32. Current input pod

1. Voltage Pod- Accepts four 5 - 600 Vrms (AC or DC), ± 1000 Vpk phase A, B, C voltage, plus neutral and ground. Neutral to ground voltage range: 0.5 - 20 Vrms (AC or DC).
2. Current Pod- Accepts four 0.01-5 Arms and up to #10 AWG wires. The current tube diameter is 0.215" (5.461mm). Measurement range allows 25 Apk with 5532A Pod, 100 Apk with 5533A Pod.
3. Data cables- Enables connection of measurement pods to the DataNode. Cable length is 3' (0.9m).
4. Power type card – Display and set power measurement type: Phase-to-Phase or Phase-to-Neutral. The power type showing should be the power type in use. Front and back of card shown below.

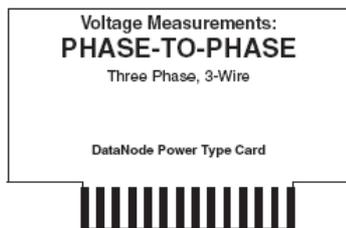


Figure 33. Front of power type card

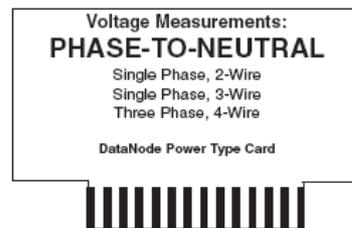


Figure 34. Back of power type card

Specifications for the 5500Series DualNode

Configurations	External CT and voltage pods; 1A/5A current with 5x overcurrent.
Measurements	160 parameters, including tru 1/2 cycle RMS sags & swells, interruptions, microsecond transients, kVA, KW, True PF, DPF, KVAR, kWhr, kVAR and other power related parameters, TIF, K-factor, THD, individ-

	ual harmonics through 50th, 45-675 Hz phase locked sampling, Up to 22 cycles of waveform per event.
Voltages	4 channels, accuracy +/- 0.1% of reading, +/- 0.1% FS.
Currents	4 channels, accuracy +/- 0.1% of reading, +/- 0.1% FS.
Instrument Power	90-250Vac, 50/60Hz; optional 105-150Vdc; built-in UPS with 4-year battery life.
Enclosure/Environments	17"w x 7"h x 8"d. Rack, desktop, wall mount; 0-60 deg C standard.
Communications	Access through Internet, Intranet, dial-up or wireless telephone line.
Additional Features	Notification (e-mail, pager, contact closure); remote firmware update; cross triggering, 10 msec accuracy w/optional GPS, AnswerModules, including Sag Direction, Radial Fault locator, Capacitor Switching Transient w/Directivity, Energy Usage and Expense Reporting.
Certifications and design standards	CE, ISO9001, EMC Directive (89/366/EEC), IEC 61000-4-7, IEC 61000-4-15, EN61010-1 (1993), EN61010-1/A2.

6 CLASSIFICATION METHODS

Having read the above information regarding the equipment that is being used nowadays, someone can easily understand that the time when power quality monitoring equipment just recorded data and took pictures of raw waveforms has gone. Various techniques in signal processing for automatic classification of power quality events have been implemented and evolved, thus making the power quality monitoring systems able to identify and classify events automatically in order to solve problems in electrical network. The purpose of this clause is to present the state of the art in power quality monitoring and classification techniques.

To have a general idea of the state of the art in power quality events classification, let's take a look at the evolution of power quality monitoring in terms of technology and users. In table 3, a time line has been plotted (The first 30 years of the diagram has been extracted from (49)).

TECHNOLOGY

-Voltmeter	-Oscilloscopes	-Digital signal Processing	-Pattern recognition
-Paper	-Graphics	-Computer	-Data Mining
-Tape		-Mass Storage	-Decision Making
		-Communications	-Networking – Internet
70's 80's 90's 2000's			
-Field Service Engineer	-Field Service Engineer	-Utility Companies	-Regulatory Agencies
	-Power Quality Groups	-Industrial / Plant / Facilities	-ISO
		-Engineers / Consultants	-IPP

USERS

Table 3. Time Line of Power Quality monitoring equipment Evolution

In the 90's, the technology applied in classification tended to merge power engineering knowledge mainly with signal processing techniques. Recently, pattern rec-

ognition, data mining, decision-making and networking were incorporated as new technologies for automatic classification. This entire advance tends to process raw data and extract information to obtain knowledge in order to make decisions (49) and (50).

Moreover, users of power quality events classification have spread, from a few field service engineers in the 70's to hundreds of people in the 2000's, in power utilities, consultant companies and governmental agencies, working to include power quality indexes in power system economic performance.

6.1 State Of The Art

A general scheme for automatic classification systems is defined in fig. (35), where its application for power quality events' classification is feasible.

Block I represents a pre-processing stage. In this block estimation of the signal components is performed. Then, an algorithm for signal segmentation in different stages is applied, e.g. pre-event, during-event or transition and post-event stages.

Block II represents a feature extraction stage. Feature extraction can be done through Wavelets or Kalman filter. Wavelets are mainly used to quantify features for different types of power system events. However, researchers do not fully agree with wavelets universal use, due to features obtained with wavelets are highly dependent on the type of mother wavelet chosen. Most of the work was done with simulated data, which is another drawback. Therefore, type event features and its extraction procedure using wavelets are still subjective.

Block III represents the classification stage based on defined rules, e.g. knowledge based expert systems or any logic to discriminate different types of events.

Finally, block IV represents the decision making stage. In this stage the type event is assigned to the current event. In many proposed algorithms, blocks III and IV are merged in one process held by neural networks, fuzzy logic, Bayesian or pattern recognition techniques.

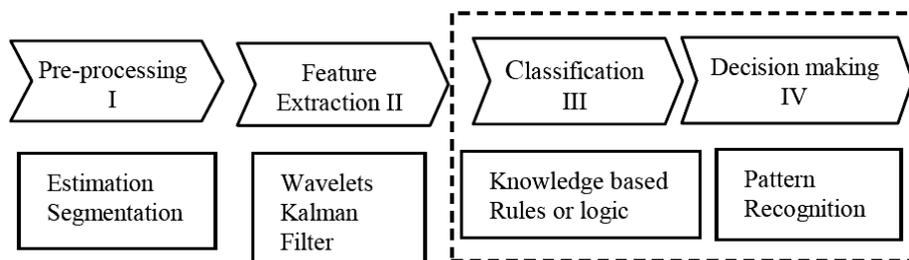


Figure 35. Automatic Classification Scheme

The next section will show a more detailed view of the state of the art developments on the automatic classification of power quality events.

6.1.1 State of the Art on the Classification of Power Quality Events

Expert systems were proposed to identify, classify and diagnose power systems events successfully for a limited number of events (62), e.g. an expert system for classification and analysis of voltage dips using Kalman filter for estimation of the amplitude has been shown in (63). Rules based expert system are highly dependant on if ... then clauses. If many event types or features were analyzed, the expert system would become more complicated and the risks of losing selectivity would increase (ambiguity). Another drawback is that these systems are not always portable due to settings depend mostly on the designer or operator of the systems for a particular set of events.

Under the thesis that energy contents of the non-fundamental components in a signal change depending on the type of event, wavelets are widely applied for detection, quantification and classification of a variety of power quality disturbances, e.g. harmonics and transients. In (64), (65) and (66) automatic classification systems based on wavelets feature extraction are proposed. However, feature extraction based in wavelets transform has some disadvantages, which will be shown later on.

In recent years, more sophisticated algorithms for automatic classification were proposed. An on-line power quality disturbance detector was proposed in (67), which was based in wavelet feature extraction. The novel idea of this detector was the application of a Bayesian classifier. This algorithm analyses the missing voltage, which is decomposed using wavelets, and extracts features regarding the energy contents of the scaled signals with respect to the error signal, which change upon the event type. Then, classifies this energy features using Bayesian approach. A drawback of using Baye's formula is that 'a priori' probability density function (pdf) of each event must be known in advance. Although this kind of classification algorithms works well for Gaussian pdfs, its performance in power quality events is observed due to the non-Gaussian nature of the event's pdfs.

It was suggested before that for voltage sag or swells, time domain analyses has shown better results and for automatic classification of fast changes like capacitor switching is better to apply frequency methods, another novel algorithm is shown in (68), even though, its feature extraction stage is wavelet based.

Up to here, all algorithms were applied just to voltage waveforms in a single network node. Finally, another novel algorithm using voltages and currents and applying non-supervised classification techniques is shown in (69), e.g. clustering and K-nearest neighbor.

6.1.2 Basic tools for signal components estimation

Although root mean square (rms) is not an inherent signal processing technique, is the most used tool. Rms gives a good approximation of the fundamental frequency amplitude profile of a waveform. A great advantage of this algorithm is its simplicity, speed of calculation and less requirement of memory, because rms can be stored periodically instead of per sample (51). However, its dependency of window length is considered as a disadvantage. One cycle window length will give better results in terms of profile than a half cycle window. Moreover, rms does not distinguish fundamental frequency, harmonics or noise components. On the other hand, rms voltage profiles are used for event analysis and automatic classification as proposed in (52). A great quantity of work has been focused in the estimation of amplitude and phase of the fundamental frequency as well as its related harmonics. A primary tool for estimation of fundamental amplitude of a signal is the discrete fourier transform (DFT) or its computationally efficient implementation called fast fourier transform (FFT). FFT transforms the signal from time domain to the frequency domain. Its fast computation is considered as an advantage. With this tool is possible to have an estimation of the fundamental amplitude and its harmonics with a reasonable approximation. However, window dependency resolution is a disadvantage. e.g. longer the sampling window better the frequency resolution. FFT performs well for estimation of periodic signals in stationary state; however it doesn't perform well for detection of suddenly or fast changes in waveform e.g. transients or voltages dips. In some cases, results of the estimation can be improved with windowing or filtering, e.g. hanning window, hamming window, low pass filter or high pass filter.

A combination of quadrature mirror filters (QMF) arranged in binary trees is called filter banks. Filter banks have been used to study in more detail a specific sub band of the frequency spectrum. This technique was used in different applications to detect rapid changes in the waveform or for estimation of specific sub-band components, e.g. harmonic contents between 500 to 1000 Hz, capacitor switching or transients.

6.1.3 More Advance tools for signal components estimation

A well-known technique is the so-called Kalman Filter. This technique is defined as a state space model for tracking amplitude and phase of fundamental frequency and its harmonics in real time under noisy environment, which was proposed in (53). Since then, many applications have come up, frequency estimation under distorted signals (54), detection of harmonics sources and optimal localization of power quality monitors (55).

In 1994, the use of wavelets was proposed to study power systems non-stationary harmonics distortion (56). This technique is used to decompose the signal in different frequency sub-bands and study separately its characteristics. As described in (57) and (58), wavelets performs better with non-periodic signals that contains short duration impulse components as is typical in power systems transients. Many different types of wavelets have been applied to power systems events, from those, Daubechies, Dyadic, Coiflets, Morlet and Symlets wavelets were found more suitable for power systems studies. Furthermore, Wavelets based techniques were proposed for detection and measuring of power systems disturbances (59) and (60).

Finally, the short time fourier transform (STFT) is commonly known as a sliding window version of the FFT, which has shown better results in terms of resolution and frequency selectivity. However, STFT has a fixed frequency resolution for all frequencies, and has shown be more suitable for harmonic analysis of voltage disturbances than binary tree filters or wavelets when is applied to study voltage dip (61).

Let us now take a closer look at some of the most commonly used tools – techniques used on power quality monitoring which are those that will be used in this paper.

6.2 Most Used Techniques

This clause attempts to give a glance at the fundamental signal processing tools and techniques used in power quality events classification in order to prepare the reader for the techniques that will be used later on in this paper.

6.2.1 Neural Networks

The purpose of this clause is not to present Neural Networks in general, besides an action like that would demand a whole new paper if not an entire book! This

section will present only the sectors of the Neural Networks that will be used for the purposes of this paper and only. Those are the following two:

- ADALINE (Adaptive Linear Neuron Networks)
- PNN (Probabilistic Neural Networks)

6.2.1.1 Adaptive Linear Neuron Networks (ADALINE)

The ADALINE networks discussed in this section are similar to the, widely known and therefore not worth mentioning, perceptron, but their transfer function is linear rather than hard-limiting. This allows their outputs to take on any value, whereas the perceptron output is limited to either 0 or 1. Both the ADALINE and the perceptron can only solve linearly separable problems. However, here we will make use of the LMS (Least Mean Squares) learning rule, which is much more powerful than the perceptron learning rule. The LMS or Widrow-Hoff learning rule minimizes the mean square error and, thus, moves the decision boundaries as far as it can from the training patterns. In this clause, we design an adaptive linear system that responds to changes in its environment as it is operating. Linear networks that are adjusted at each time step based on new input and target vectors can find weights and biases that minimize the network's sum-squared error for recent input and target vectors. Networks of this sort are often used in error cancellation, signal processing, and control systems. The pioneering work in this field was done by Widrow and Hoff, who gave the name ADALINE to adaptive linear elements (70). We also consider the adaptive training of self organizing and competitive networks in this section.

6.2.1.1.1 Linear Neuron Model

A linear neuron with R inputs is shown below.

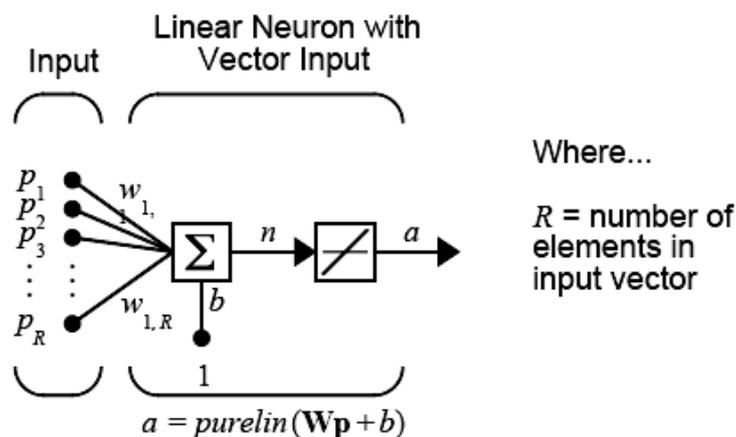


Figure 36. Linear Neuron

This network has the same basic structure as the perceptron. The only difference is that the linear neuron uses a linear transfer function, which we name purelin.

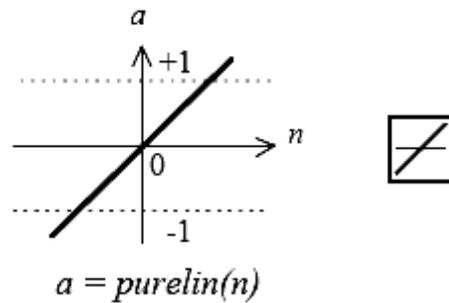


Figure 37. Linear Transfer Function

The linear transfer function calculates the neuron's output by simply returning the value passed to it.

$$\mathbf{a} = \text{purelin}(\mathbf{n}) = \text{purelin}(\mathbf{Wp} + \mathbf{b}) = \mathbf{Wp} + \mathbf{b} \quad \text{Equation 1}$$

This neuron can be trained to learn an affine function of its inputs, or to find a linear approximation to a nonlinear function. A linear network cannot, of course, be made to perform a nonlinear computation.

6.2.1.1.2 Adaptive Linear Network Architecture

The ADALINE network shown below has one layer of S neurons connected to R inputs through a matrix of weights W .

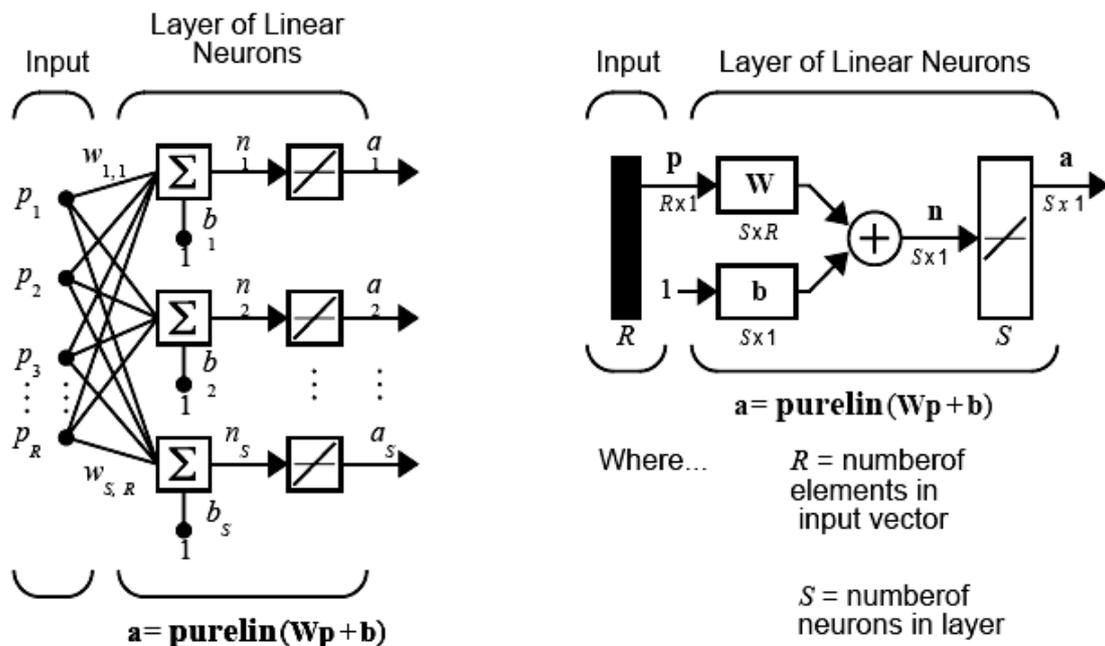


Figure 38. ADALINE Network

This network is sometimes called a MADALINE for Many ADALINES. Note that the figure on the right defines an S-length output vector a .

The Widrow-Hoff rule can only train single-layer linear networks. This is not much of a disadvantage, however, as single-layer linear networks are just as capable as multilayer linear networks. For every multilayer linear network, there is an equivalent single-layer linear network.

6.2.1.1.3 Single ADALINE

Consider a single ADALINE with two inputs. The diagram for this network is shown below.

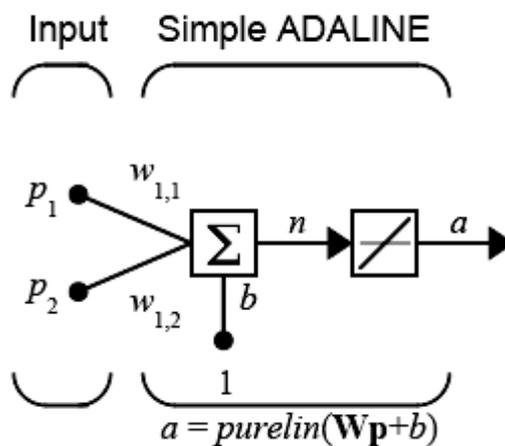


Figure 39. Single ADALINE

The weight matrix W in this case has only one row. The network output is:

$$\mathbf{a} = \text{purelin}(n) = \text{purelin}(W\mathbf{p} + b) = W\mathbf{p} + b \quad \text{or} \quad \text{Equation 2}$$

$$\mathbf{a} = w_{1,1}p_1 + w_{1,2}p_2 + b \quad \text{Equation 3}$$

Like the perceptron, the ADALINE has a decision boundary that is determined by the input vectors for which the net input n is zero. For $n = 0$ the equation $W\mathbf{p} + b = 0$ specifies such a decision boundary as shown below (71).

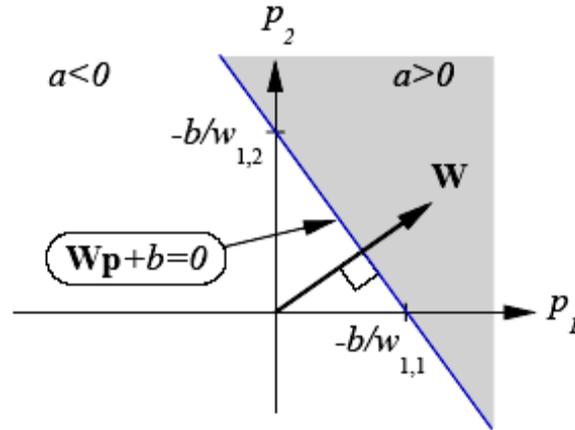


Figure 40. ADALINE's Decision Boundary

Input vectors in the upper right gray area lead to an output greater than 0. Input vectors in the lower left white area lead to an output less than 0. Thus, the ADALINE can be used to classify objects into two categories.

6.2.1.1.4 Mean Square Error

Like the perceptron learning rule, the least mean square error (LMS) algorithm is an example of supervised training, in which the learning rule is provided with a set of examples of desired network behavior.

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\}$$

Here p_q is an input to the network, and t_q is the corresponding target output. As each input is applied to the network, the network output is compared to the target. The error is calculated as the difference between the target output and the network output. We want to minimize the average of the sum of these errors.

$$\text{mse} = \frac{1}{Q} \sum_{k=1}^Q e(k)^2 = \frac{1}{Q} \sum_{k=1}^Q (t(k) - a(k))^2 \quad \text{Equation 4}$$

The LMS algorithm adjusts the weights and biases of the ADALINE so as to minimize this mean square error. Fortunately, the mean square error performance index for the ADALINE network is a quadratic function. Thus, the performance index will either have one global minimum, a weak minimum, or no minimum, depending on the characteristics of the input vectors. Specifically, the characteristics of the input vectors determine whether or not a unique solution exists. More about this topic are available in (71).

6.2.1.1.5 LMS Algorithm

Adaptive networks will use the LMS algorithm, shown in eqs. (5) & (6), or Widrow-Hoff learning algorithm based on an approximate steepest descent procedure. Here again, adaptive linear networks are trained on examples of correct behavior.

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2ae(k)\mathbf{pT}(k) \quad \text{Equation 5}$$

$$\mathbf{b}(k+1) = \mathbf{b}(k) + 2ae(k) \quad \text{Equation 6}$$

6.2.1.1.6 Adaptive Filtering

The ADALINE network, much like the perceptron, can only solve linearly separable problems. Nevertheless, the ADALINE has been and is today one of the most widely used neural networks found in practical applications. Adaptive filtering is one of its major application areas.

A new component is needed, the tapped delay line, to make full use of the ADALINE network. Such a delay line is shown below.

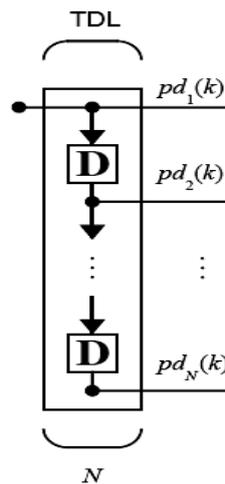


Figure 41. Tapped Delay Line

There the input signal enters from the left, and passes through N-1 delays. The output of the tapped delay line (TDL) is an N-dimensional vector, made up of the input signal at the current time, the previous input signal, etc.

A tapped delay line can be combined with an ADALINE network to create the adaptive filter shown below.

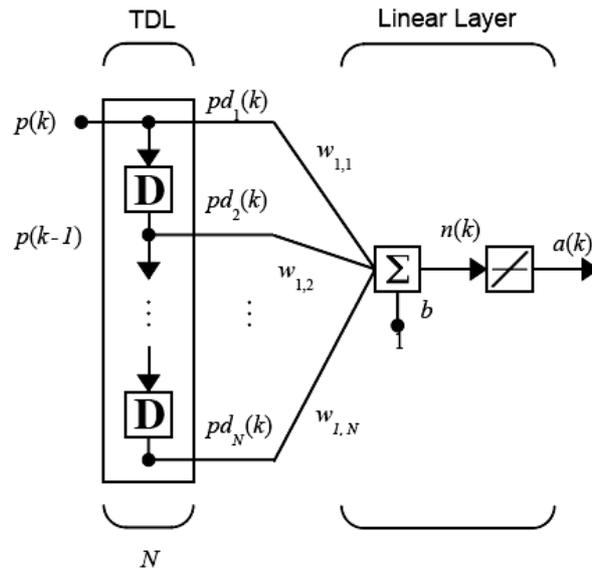


Figure 42. Adaptive Filter

The output of the filter is given by

$$a(k) = \text{purelin}(Wp + b) = \sum_{i=1}^R w_{1,i} a(k - i + 1) + b \quad \text{Equation 7}$$

The network shown above is referred to the digital signal processing field as a finite impulse response (FIR) filter (72).

6.2.1.1.7 Multiple Neuron Adaptive Filters

Sometimes the use of more than one neuron in an adaptive system is wanted, so some additional notation is needed. A tapped delay line can be used with S linear neurons as shown below.

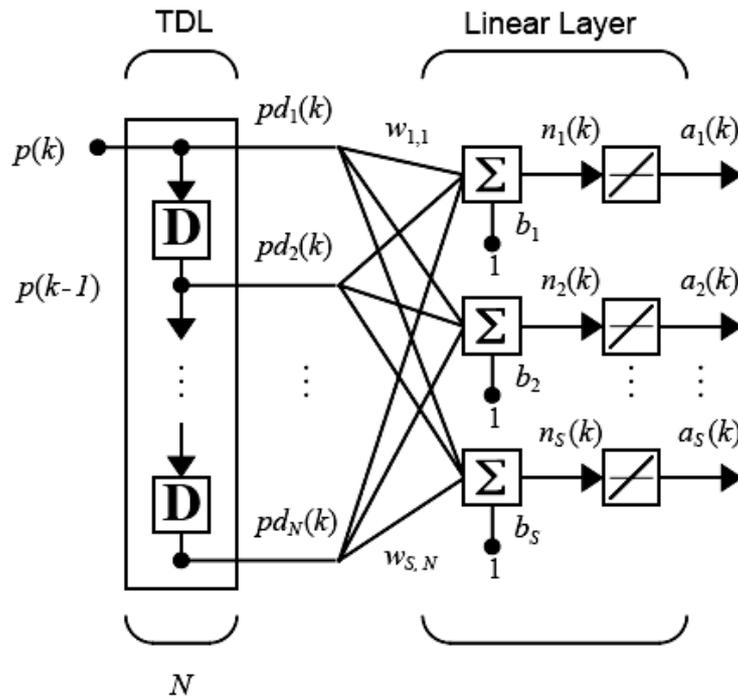


Figure 43. Multiple Neuron Adaptive Filter

Alternatively, the same network can be shown in abbreviated form.

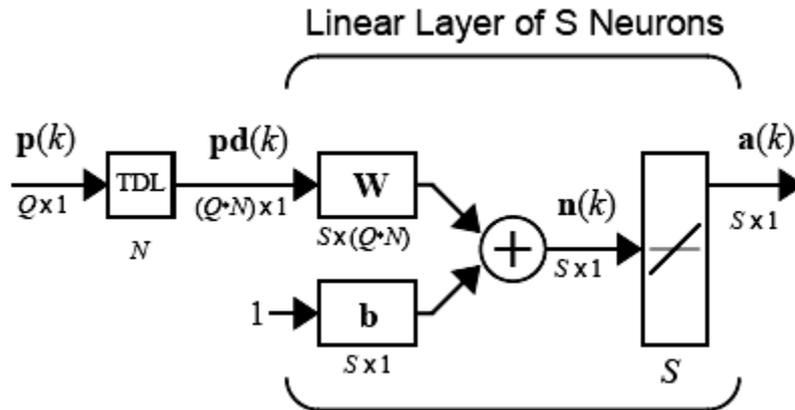


Figure 44. Abbreviated Form Of A Multiple Neuron Adaptive Filter

If more of the detail of the tapped delay line is needed to be shown and there are not too many delays, we can use the following notation.

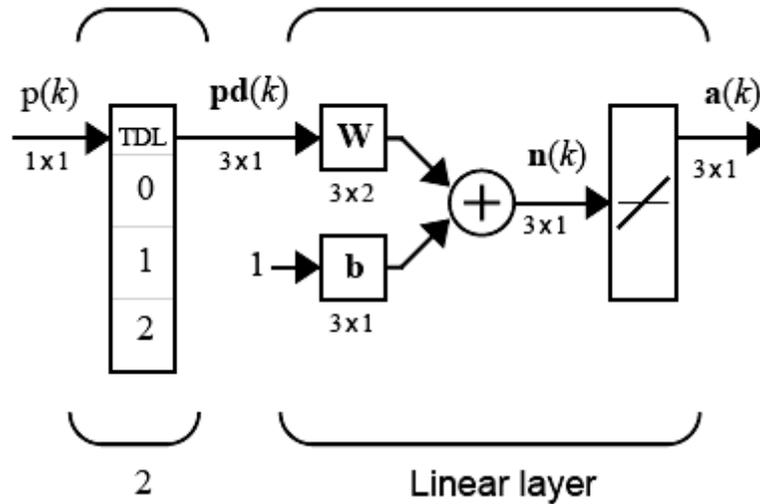


Figure 45. Abbreviated Notation

Here there is a tapped delay line that sends the current signal, the previous signal, and the signal delayed before that to the weight matrix. We could have a longer list, and some delay values could be omitted if desired. The only requirement is that the delays are shown in increasing order as they go from top to bottom

6.2.1.1.8 Summary

The ADALINE (Adaptive Linear Neuron networks) networks discussed in these clauses are similar to the perceptron, but their transfer function is linear rather than hard-limiting. They make use of the LMS (Least Mean Squares) learning rule, which is much more powerful than the perceptron learning rule. The LMS or Widrow-Hoff learning rule minimizes the mean square error and, thus, moves the decision boundaries as far as it can from the training patterns. Thus an adaptive linear system that responds to changes in its environment as it is operating can be designed. Linear networks that are adjusted at each time step based on new input and target vectors can find weights and biases that minimize the network's sum-squared error for recent input and target vectors.

Adaptive linear filters have many practical applications such as noise cancellation, signal processing, and prediction in control and communication systems.

6.2.1.2 Probabilistic Neural Networks (PNN)

Probabilistic Neural Networks are a variant of Radial Basis Networks, thus in order to understand PNN there must first be a small introduction to Radial Basis Networks. Radial Basis Networks may require more neurons than standard feed-forward backpropagation networks, but they can often be designed in a fraction of the time that takes to train standard feed-forward networks. They work best when many training vectors are available (73).

6.2.1.2.1 Radial Basis Networks

A radial basis network with R inputs is represented in the following picture.

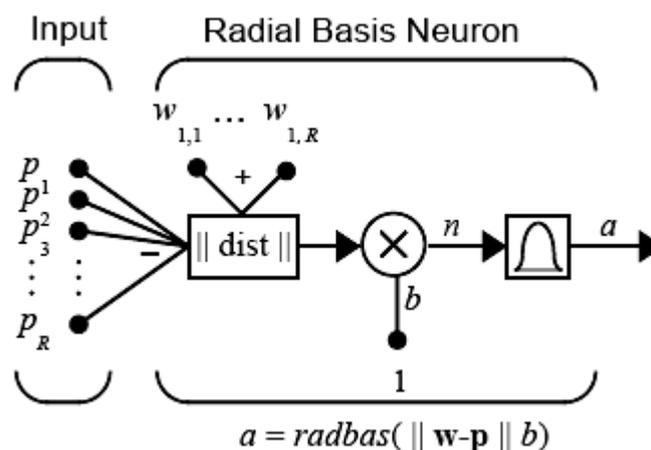


Figure 46. Radial Basis Network

Notice that the expression for the net input of a radbas neuron is different from that of the most commonly used neurons. Here the net input to the radbas transfer function is the vector distance between its weight vector w and the input vector p , multiplied by the bias b . (The $\| \text{dist} \|$ box in this figure accepts the input vector p and the single row input weight matrix, and produces the dot product of the two.)

The transfer function and its plot, for a radial basis neuron are respectively:

$$\text{radbas}(n) = e^{-n^2} \quad \text{Equation 8}$$

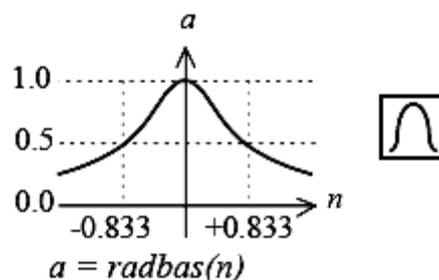


Figure 47. Transfer Function For A Radial Basis Neuron

The radial basis function has a maximum of 1 when its input is 0. As the distance between w and p decreases, the output increases. Thus, a radial basis neuron acts as a detector that produces 1 whenever the input p is identical to its weight vector w .

The bias b allows the sensitivity of the radbas neuron to be adjusted. For example, if a neuron had a bias of 0.1 it would output 0.5 for any input vector p at vector distance of 8.326 ($0.8326/b$) from its weight vector w .

Radial basis networks consist of two layers: a hidden radial basis layer of S^1 neurons, and an output linear layer of S^2 neurons.

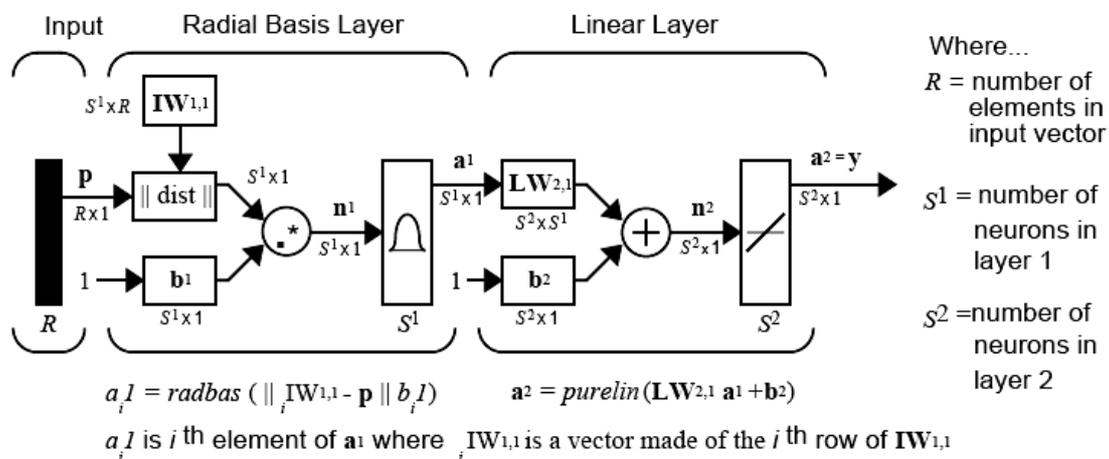


Figure 48. Radial Basis Network Architecture

The $\|\text{dist}\|$ box in this figure accepts the input vector p and the input weight matrix $\mathbf{IW}_{1,1}$, and produces a vector having S^1 elements. The elements are the distances between the input vector and vectors $\mathbf{IW}_{1,1}^i$, formed from the rows of the input weight matrix. The bias vector \mathbf{b}^1 and the output of $\|\text{dist}\|$ are combined with element-by-element multiplication.

We can understand how this network behaves by following an input vector p through the network to the output \mathbf{a}^2 . If we present an input vector to such a network, each neuron in the radial basis layer will output a value according to how close the input vector is to each neuron's weight vector. Thus, radial basis neurons with weight vectors quite different from the input vector p have outputs near zero. These small outputs have only a negligible effect on the linear output neurons. In contrast, a radial basis neuron with a weight vector close to the input vector p produces a value near 1. If a neuron has an output of 1 its output weights in the second layer pass their values to the linear neurons in the second layer. In fact, if only one radial basis neuron had an

output of 1, and all others had outputs of 0's (or very close to 0), the output of the linear layer would be the active neuron's output weights. This would, however, be an extreme case. Typically several neurons are always firing, to varying degrees.

Now let us look in detail at how the first layer operates. Each neuron's weighted input is the distance between the input vector and its weight vector. Each neuron's net input is the element-by-element product of its weighted input with its bias. Each neuron's output is its net input passed through radbas. If a neuron's weight vector is equal to the input vector (transposed), its weighted input is 0, its net input is 0, and its output is 1. If a neuron's weight vector is a distance of spread from the input vector, its weighted input is spread, its net input is $\sqrt{-\log(.5)}$ (or 0.8326), therefore its output is 0.5. For the exact design of a radial basis network more specific information are available in (75).

6.2.1.2.2 Probabilistic Neural Networks (PNN)

Probabilistic neural networks can be used for classification problems. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes. The architecture for this system is shown below.

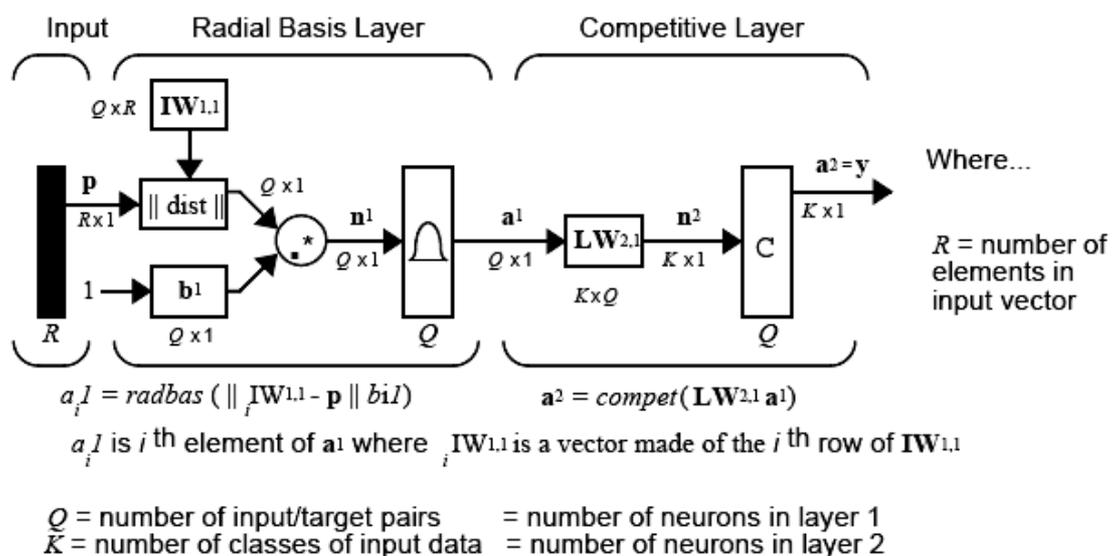


Figure 49. Architecture Of A Probabilistic Neural Network

It is assumed that there are Q input vector/target vector pairs. Each target vector has K elements. One of these elements is 1 and the rest is 0. Thus, each input vector is associated with one of K classes.

The first-layer input weights, $IW^{1,1}$ are set to the transpose of the matrix formed from the Q training pairs, P^1 . When an input is presented the $\|dist\|$ box produces a vector whose elements indicate how close the input is to the vectors of the training set. These elements are multiplied, element by element, by the bias and sent the radbas transfer function. An input vector close to a training vector is represented by a number close to 1 in the output vector a^1 . If an input is close to several training vectors of a single class, it is represented by several elements of a^1 that are close to 1.

The second-layer weights, $LW^{1,2}$, are set to the matrix T of target vectors. Each vector has a 1 only in the row associated with that particular class of input, and 0's elsewhere. The multiplication Ta^1 sums the elements of a^1 due to each of the K input classes. Finally, the second-layer transfer function, compete, produces a 1 corresponding to the largest element of n^2 , and 0's elsewhere. Thus, the network has classified the input vector into a specific one of K classes because that class had the maximum probability of being correct.

6.2.1.2.3 Summary

Radial basis networks can be designed very quickly in two different ways. The first design method finds an exact solution, by creating radial basis networks with as many radial basis neurons as there are input vectors in the training data. The second method finds the smallest network that can solve the problem within a given error goal. Typically, far fewer neurons are required by making use of the second method. However, because the number of radial basis neurons is proportional to the size of the input space, and the complexity of the problem, radial basis networks can still be larger than backpropagation networks.

Probabilistic neural networks (PNN) can be used for classification problems. Their design is straightforward and does not depend on training. A PNN is guaranteed to converge to a Bayesian classifier providing it is given enough training data. These networks generalize well. PNN have many advantages, but they suffer from one major disadvantage. They are slower to operate because they use more computation than other kinds of networks to do their function approximation or classification.

6.2.2 Fuzzy Networks

Acting in an exactly similar way as with Neural Networks, the purpose of this clause is to present only the sector of Fuzzy Networks which is being used for the purposes of this paper and not the presentation of Fuzzy Networks in general. To be more precise, this paper makes use of the Fuzzy Networks by means of the ANFIS.

6.2.2.1 Adaptive Neuro – Fuzzy Inference System ANFIS

The acronym ANFIS derives its name from Adaptive Neuro-Fuzzy Inference System. The basic idea behind these neuro-adaptive learning techniques is very simple. These techniques provide a method for the fuzzy modeling procedure to learn information about a data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. This learning method works similarly to that of neural networks.

Using a given input/output data set, a fuzzy inference system (FIS) is constructed and its membership function parameters are tuned (adjusted) using either a back propagation algorithm alone, or in combination with a method of least squares type, this is called the hybrid method. This allows fuzzy systems to learn from the data they are modeling.

6.2.2.1.1 FIS Structure and Parameter Adjustment

A network-type structure similar to that of a neural network, which maps inputs through input membership functions and associated parameters, and then through output membership functions and associated parameters to outputs, can be used to interpret the input/output map. The parameters associated with the membership functions will change through the learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization routines could be applied in order to adjust the parameters so as to reduce some error measure (usually defined by the sum of the squared difference between actual and desired outputs).

6.2.2.1.2 Familiarity Breeds Validation; Knowledge of the Data

The modeling approach used is similar to many system identification techniques. First, a parameterized model structure (relating inputs to membership functions to rules to outputs to membership functions, and so on) is hypothesized. Next, input/output data are collected in a form that will be usable by the network for training. Then the FIS model can be trained to emulate the training data presented to it by modifying the membership function parameters according to a chosen error criterion. In general, this type of modeling works well if the training data presented to the ANFIS for training (estimating) membership function parameters is fully representative of the features of the data that the trained FIS is intended to model. This is not always the case, however. In some cases, data is collected using noisy measurements, and the training data cannot be representative of all the features of the data that will be presented to the model. This is where model validation comes into play.

Model validation is the process by which the input vectors from input/output data sets on which the FIS was not trained, are presented to the trained FIS model, to see how well the FIS model predicts the corresponding data set output values. This is accomplished with the ANFIS Editor GUI using the so-called testing data set. You can also use another type of data set for model validation in ANFIS. This other type of validation data set is referred to as the checking data set and this set is used to control the potential for the model overfitting the data. When checking data is presented to ANFIS as well as training data, the FIS model is selected to have parameters associated with the minimum checking data model error. One problem with model validation for models constructed using adaptive techniques is selecting a data set that is both representative of the data the trained model is intended to emulate, yet sufficiently distinct from the training data set so as not to render the validation process trivial. If a large amount of data is collected, hopefully this data contains all the necessary representative features, so the process of selecting a data set for checking or testing purposes is made easier. However, if noisy measurements are expected to be presenting to the model, it's possible the training data set does not include all of the representative features that are wanted to model.

The basic idea behind using a checking data set for model validation is that after a certain point in the training, the model begins overfitting the training data set. In principle, the model error for the checking data set tends to decrease as the training

takes place up to the point that overfitting begins, and then the model error for the checking data suddenly increases.

6.2.2.1.3 Constraints of ANFIS

ANFIS is much more complex than the commonly used fuzzy inference systems, and is not available for all of the fuzzy inference system options. Specifically, ANFIS only supports Sugeno-type systems, and these must have the following properties:

- Be first or zeroth order Sugeno-type systems.
- Have a single output, obtained using weighted average defuzzification. All output membership functions must be the same type and either be linear or constant.
- Have no rule sharing. Different rules cannot share the same output membership function, namely the number of output membership functions must be equal to the number of rules.
- Have unity weight for each rule.

An error occurs if your FIS structure does not comply with these constraints.

Moreover, ANFIS cannot accept all the customization options that basic fuzzy inference allows. That is, you cannot make your own membership functions and defuzzification functions; you must use the ones provided.

6.2.3 WAVELET TRANSFORM

The wavelet transform is a relatively new concept, which is why there will be a comparative wide presentation on it. Later on, basic principles underlying the wavelet theory are given. The proofs of the theorems and related equations will not be given in due to the simple assumption that the intended readers of this paper do not need them at this time. However, there are some related references for further and in-depth information. Let's begin with a small introduction regarding the Wavelet Transform and then take a closer look.

The Wavelet transform provides the time-frequency representation. (There are other transforms which give this information too, such as short time Fourier transform, Wigner distributions, etc.)

Often times a particular spectral component occurring at any instant can be of particular interest. In these cases it may be very beneficial to know the time intervals

these particular spectral components occur. For example, in EEGs, the latency of an event-related potential is of particular interest (Event-related potential is the response of the brain to a specific stimulus like flash-light, the latency of this response is the amount of time elapsed between the onset of the stimulus and the response).

Wavelet transform is capable of providing the time and frequency information simultaneously, hence giving a time-frequency representation of the signal.

To make a real long story short, we pass the time-domain signal from various highpass and low pass filters, which filters out either high frequency or low frequency portions of the signal. This procedure is repeated, every time some portion of the signal corresponding to some frequencies being removed from the signal.

Here is how this works: Suppose we have a signal which has frequencies up to 1000 Hz. In the first stage we split up the signal in to two parts by passing the signal from a highpass and a lowpass filter (filters should satisfy some certain conditions, so-called admissibility condition) which results in two different versions of the same signal: portion of the signal corresponding to 0-500 Hz (low pass portion), and 500-1000 Hz (high pass portion).

Then, we take either portion (usually low pass portion) or both, and do the same thing again. This operation is called decomposition.

Assuming that we have taken the lowpass portion, we now have 3 sets of data, each corresponding to the same signal at frequencies 0-250 Hz, 250-500 Hz, 500-1000 Hz.

Then we take the lowpass portion again and pass it through low and high pass filters; we now have 4 sets of signals corresponding to 0-125 Hz, 125-250 Hz, 250-500 Hz, and 500-1000 Hz. We continue like this until we have decomposed the signal to a pre-defined certain level. Then we have a bunch of signals, which actually represent the same signal, but all corresponding to different frequency bands. We know which signal corresponds to which frequency band, and if we put all of them together and plot them on a 3-D graph, we will have time in one axis, frequency in the second and amplitude in the third axis. This will show us which frequencies exist at which time.

There is an issue, called "uncertainty principle", which states that, we cannot exactly know what frequency exists at what time instance, but we can only know what frequency bands exist at what time intervals. A brief explanation of this is written below.

noted at this time that the scale is inverse of frequency. That is, high scales correspond to low frequencies, and low scales correspond to high frequencies. Consequently, the little peak in the plot corresponds to the high frequency components in the signal, and the large peak corresponds to low frequency components (which appear before the high frequency components in time) in the signal.

There might be a small concern regarding the frequency resolution shown in the plot, since it shows good frequency resolution at high frequencies. Note however that, it is the good scale resolution that looks good at high frequencies (low scales), and good scale resolution means poor frequency resolution and vice versa. More about this will be written on the next clauses. After this small introduction let's take a closer look on the wavelet transform.

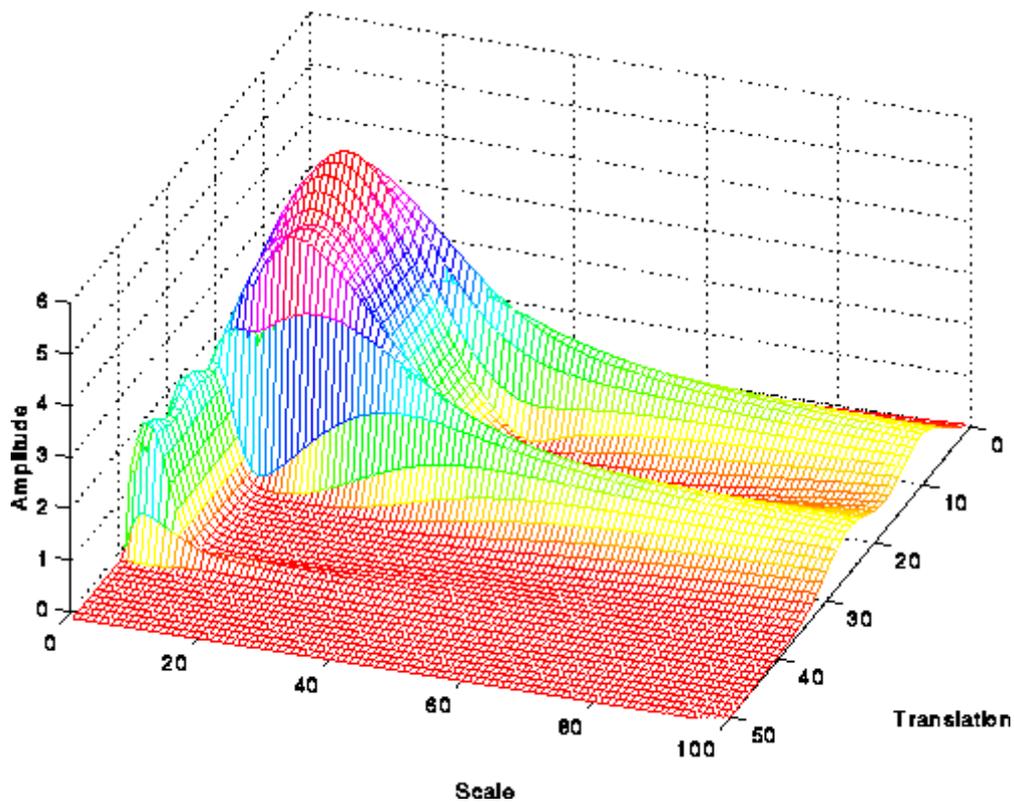


Figure 51. The Continuous Wavelet Transform Of The Above Signal

The continuous wavelet transform was developed as an alternative approach to the short time Fourier transform to overcome the resolution problem. The wavelet analysis is done in a similar way to the STFT analysis, in the sense that the signal is multiplied with a function, {it the wavelet}, similar to the window function in the STFT, and the transform is computed separately for different segments of the time-

domain signal. However, there are two main differences between the STFT and the CWT:

1. The Fourier transforms of the windowed signals are not taken, and therefore single peak will be seen corresponding to a sinusoid, i.e., negative frequencies are not computed.
2. The width of the window is changed as the transform is computed for every single spectral component, which is probably the most significant characteristic of the wavelet transform.

The continuous wavelet transform is defined as follows:

$$\text{CWT}_X^\Psi(\tau, s) = \Psi_X^\Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi\left(\frac{t-\tau}{s}\right) dt \quad \text{Equation 9}$$

As seen in the above equation, the transformed signal is a function of two variables, tau and s, the translation and scale parameters, respectively, psi(t) is the transforming function, and it is called the mother wavelet. The term mother wavelet gets its name due to two important properties of the wavelet analysis as explained below:

The term wavelet means a small wave. The smallness refers to the condition that this (window) function is of finite length (compactly supported). The wave refers to the condition that this function is oscillatory. The term mother implies that the functions with different region of support that are used in the transformation process are derived from one main function, the mother wavelet. In other words, the mother wavelet is a prototype for generating the other window functions.

The term translation is used in the same sense as it was used in the STFT; it is related to the location of the window, as the window is shifted through the signal. This term, obviously, corresponds to time information in the transform domain. However, we do not have a frequency parameter, as we had before for the STFT. Instead, we have scale parameter which is defined as “1/frequency”. The term frequency is reserved for the STFT.

The parameter scale in the wavelet analysis is similar to the scale used in maps. As in the case of maps, high scales correspond to a non-detailed global view (of the signal), and low scales correspond to a detailed view. Similarly, in terms of frequency, low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signal), whereas high frequencies (low scales) corre-

spond to a detailed information of a hidden pattern in the signal (that usually lasts a relatively short time). Cosine signals corresponding to various scales are given as examples in the following figure .

Fortunately in practical applications, low scales (high frequencies) do not last for the entire duration of the signal, unlike those shown in fig. (52), but they usually appear from time to time as short bursts, or spikes. High scales (low frequencies) usually last for the entire duration of the signal.

Scaling, as a mathematical operation, either dilates or compresses a signal. Larger scales correspond to dilated (or stretched out) signals and small scales correspond to compressed signals. All of the signals given in the figure are derived from the same cosine signal, i.e., they are dilated or compressed versions of the same function. In the following figures, $s=0.05$ is the smallest scale, and $s=1$ is the largest scale.

In terms of mathematical functions, if $f(t)$ is a given function $f(st)$ corresponds to a contracted (compressed) version of $f(t)$ if $s > 1$ and to an expanded (dilated) version of $f(t)$ if $s < 1$.

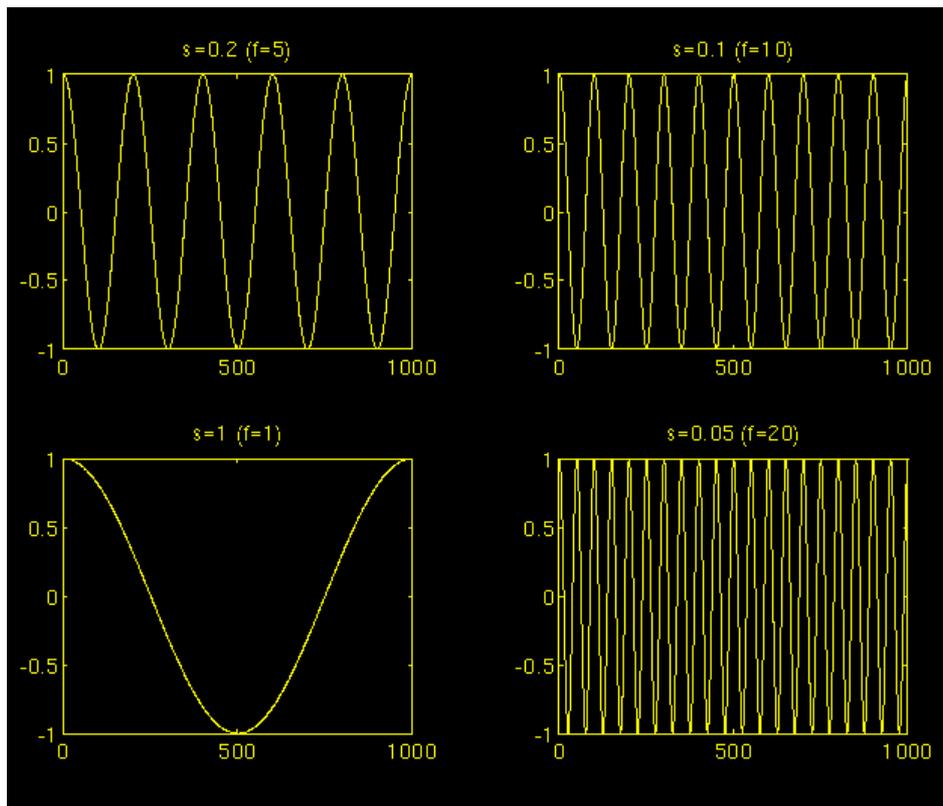


Figure 52. Cosine signals corresponding to various scales

However, in the definition of the wavelet transform, the scaling term is used in the denominator, and therefore, the opposite of the above statements holds, i.e., scales

$s > 1$ dilates the signals whereas scales $s < 1$, compresses the signal. This interpretation of scale will be used throughout this paragraph.

6.2.3.1 Computation Of The Cwt

Interpretation of the above equation will be explained in this clause. Let $x(t)$ is the signal to be analyzed. The mother wavelet is chosen to serve as a prototype for all windows in the process. All the windows that are used are the dilated (or compressed) and shifted versions of the mother wavelet. There are a number of functions that are used for this purpose. The Morlet wavelet and the Mexican hat function are two candidates, and they are used for the wavelet analysis of the examples which are presented later in this chapter.

Once the mother wavelet is chosen the computation starts with $s=1$ and the continuous wavelet transform is computed for all values of s , smaller and larger than “1”. However, depending on the signal, a complete transform is usually not necessary. For all practical purposes, the signals are bandlimited, and therefore, computation of the transform for a limited interval of scales is usually adequate. In this study, some finite interval of values for s were used, as will be described later in this chapter.

For convenience, the procedure will be started from scale $s=1$ and will continue for the increasing values of s , i.e., the analysis will start from high frequencies and proceed towards low frequencies. This first value of s will correspond to the most compressed wavelet. As the value of s is increased, the wavelet will dilate.

The wavelet is placed at the beginning of the signal at the point which corresponds to $\text{time}=0$. The wavelet function at scale “1” is multiplied by the signal and then integrated over all times. The result of the integration is then multiplied by the constant number $1/\sqrt{s}$. This multiplication is for energy normalization purposes so that the transformed signal will have the same energy at every scale. The final result is the value of the transformation, i.e., the value of the continuous wavelet transform at time zero and scale $s=1$. In other words, it is the value that corresponds to the point $\tau=0, s=1$ in the time-scale plane.

The wavelet at scale $s=1$ is then shifted towards the right by τ amount to the location $t=\tau$, and the above equation is computed to get the transform value at $t=\tau, s=1$ in the time-frequency plane.

This procedure is repeated until the wavelet reaches the end of the signal. One row of points on the time-scale plane for the scale $s=1$ is now completed.

Then, s is increased by a small value. Note that, this is a continuous transform, and therefore, both τ and s must be incremented continuously. However, if this transform needs to be computed by a computer, then both parameters are increased by a sufficiently small step size. This corresponds to sampling the time-scale plane.

The above procedure is repeated for every value of s . Every computation for a given value of s fills the corresponding single row of the time-scale plane. When the process is completed for all desired values of s , the CWT of the signal has been calculated. Fig. (53) illustrates the entire process step by step.

In fig. (53), the signal and the wavelet function are shown for four different values of τ . The scale value is 1, corresponding to the lowest scale, or highest frequency. Note how compact it is (the blue window). It should be as narrow as the highest frequency component that exists in the signal. Four distinct locations of the wavelet function are shown in the figure at $\tau=2$, $\tau=40$, $\tau=90$, and $\tau=140$. At every location, it is multiplied by the signal. Obviously, the product is nonzero only where the signal falls in the region of support of the wavelet, and it is zero elsewhere. By shifting the wavelet in time, the signal is localized in time, and by changing the value of s , the signal is localized in scale (frequency).

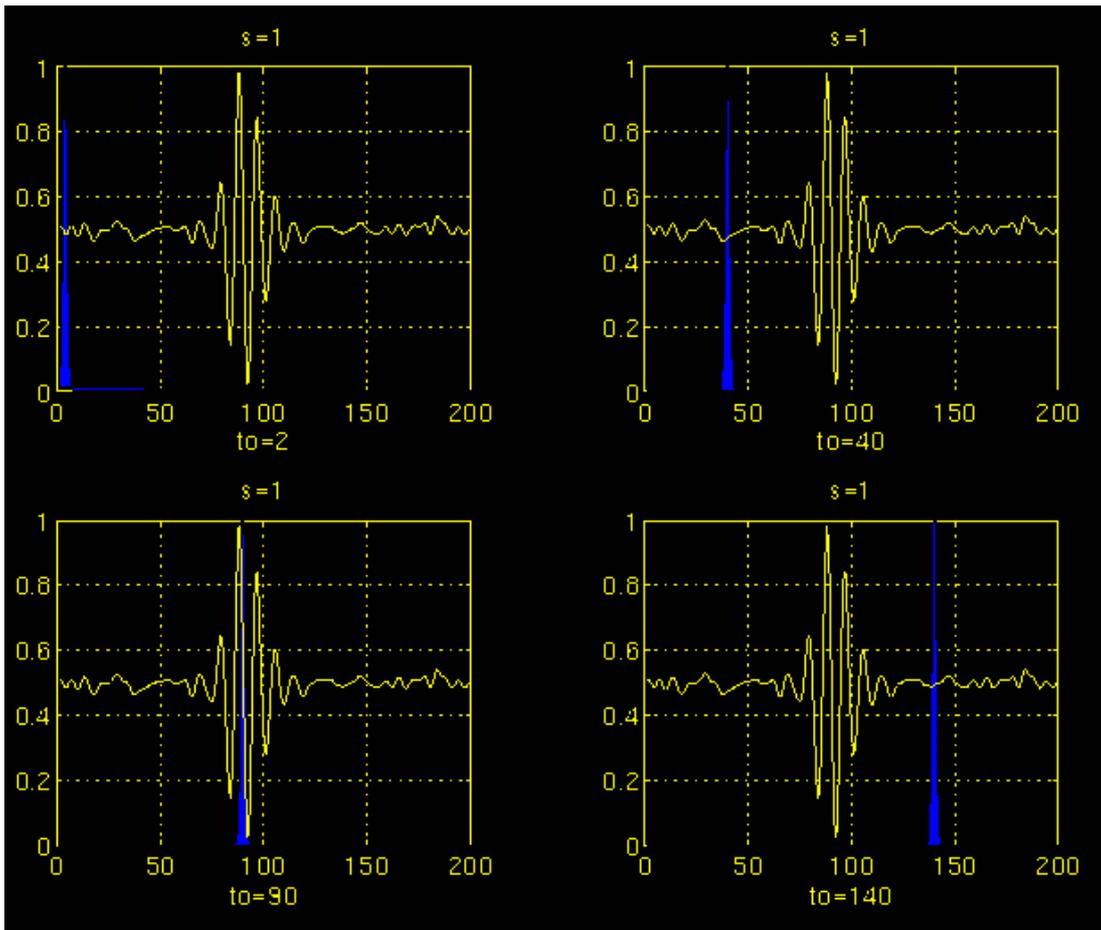


Figure 53. Calculation of a signal's CWT ($s=1$)

If the signal has a spectral component that corresponds to the current value of s (which is 1 in this case), the product of the wavelet with the signal at the location where this spectral component exists gives a relatively large value. If the spectral component that corresponds to the current value of s is not present in the signal, the product value will be relatively small, or zero. The signal in fig. (53) has spectral components comparable to the window's width at $s=1$ around $t=100$ ms.

The continuous wavelet transform of the signal in fig. (53) will yield large values for low scales around time 100 ms, and small values elsewhere. For high scales, on the other hand, the continuous wavelet transform will give large values for almost the entire duration of the signal, since low frequencies exist at all times.

Fig. (54) and (55) illustrate the same process for the scales $s=5$ and $s=20$, respectively. Note how the window width changes with increasing scale (decreasing frequency). As the window width increases, the transform starts picking up the lower frequency components.

As a result, for every scale and for every time (interval), one point of the time-scale plane is computed. The computations at one scale construct the rows of the time-scale plane, and the computations at different scales construct the columns of the time-scale plane.

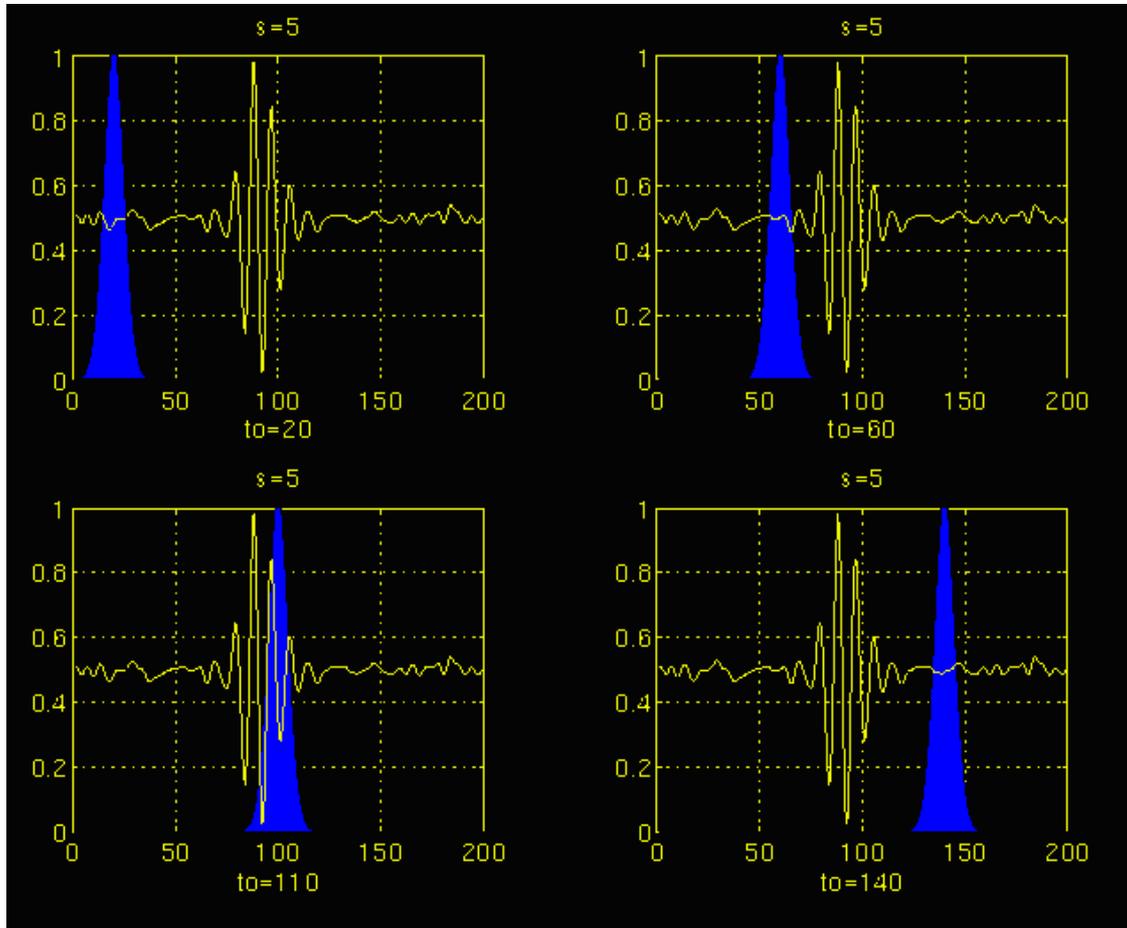


Figure 54. Calculation of a signal's CWT ($s=5$)

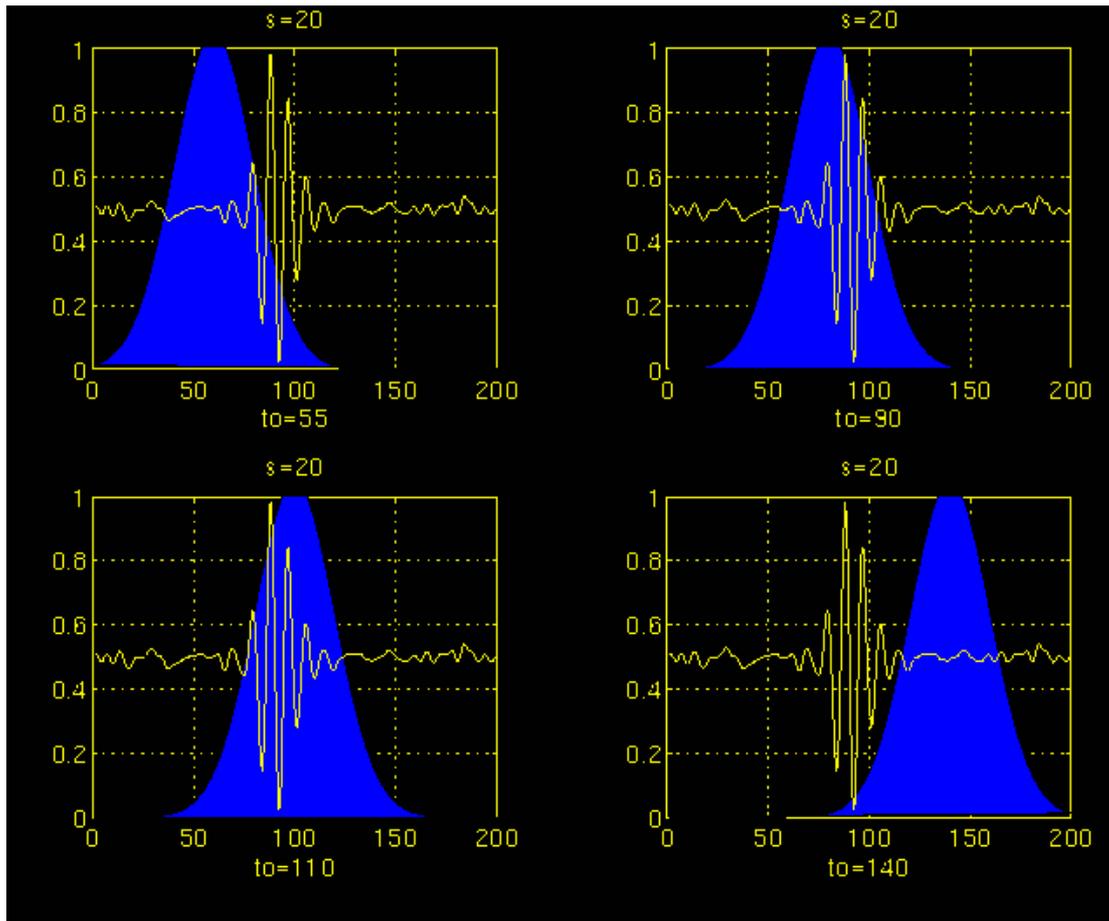


Figure 55. Calculation of a signal's CWT ($s=20$)

Now, let's take a look at an example, and see how the wavelet transform really looks like. Consider the non-stationary signal in fig. (56). As stated on the figure, the signal is composed of four frequency components at 30 Hz, 20 Hz, 10 Hz and 5 Hz.

Fig. (57) is the continuous wavelet transform (CWT) of this signal. Note that the axes are translation and scale, not time and frequency. However, translation is strictly related to time, since it indicates where the mother wavelet is located. The translation of the mother wavelet can be thought of as the time elapsed since $t=0$. The scale, however, has a whole different story. Remember that the scale parameter s in eq. (9) is actually inverse of frequency. In other words, whatever we said about the properties of the wavelet transform regarding the frequency resolution, inverse of it will appear on the figures showing the WT of the time-domain signal.

Note that in fig. (40) that smaller scales correspond to higher frequencies, i.e., frequency decreases as scale increases, therefore, that portion of the graph with scales around zero, actually correspond to highest frequencies in the analysis, and that with high scales correspond to lowest frequencies. Remember that the signal had 30 Hz

(highest frequency) components first, and this appears at the lowest scale at a translations of 0 to 30. Then comes the 20 Hz component, second highest frequency, and so on. The 5 Hz component appears at the end of the translation axis (as expected), and at higher scales (lower frequencies) again as expected.

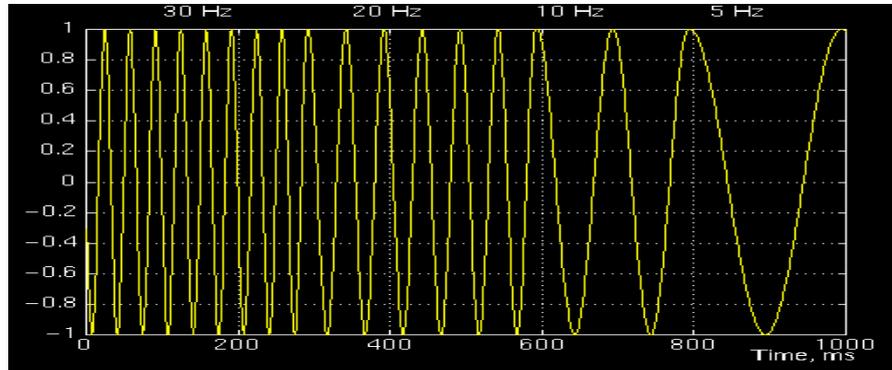


Figure 56. A Non-Stationary Signal

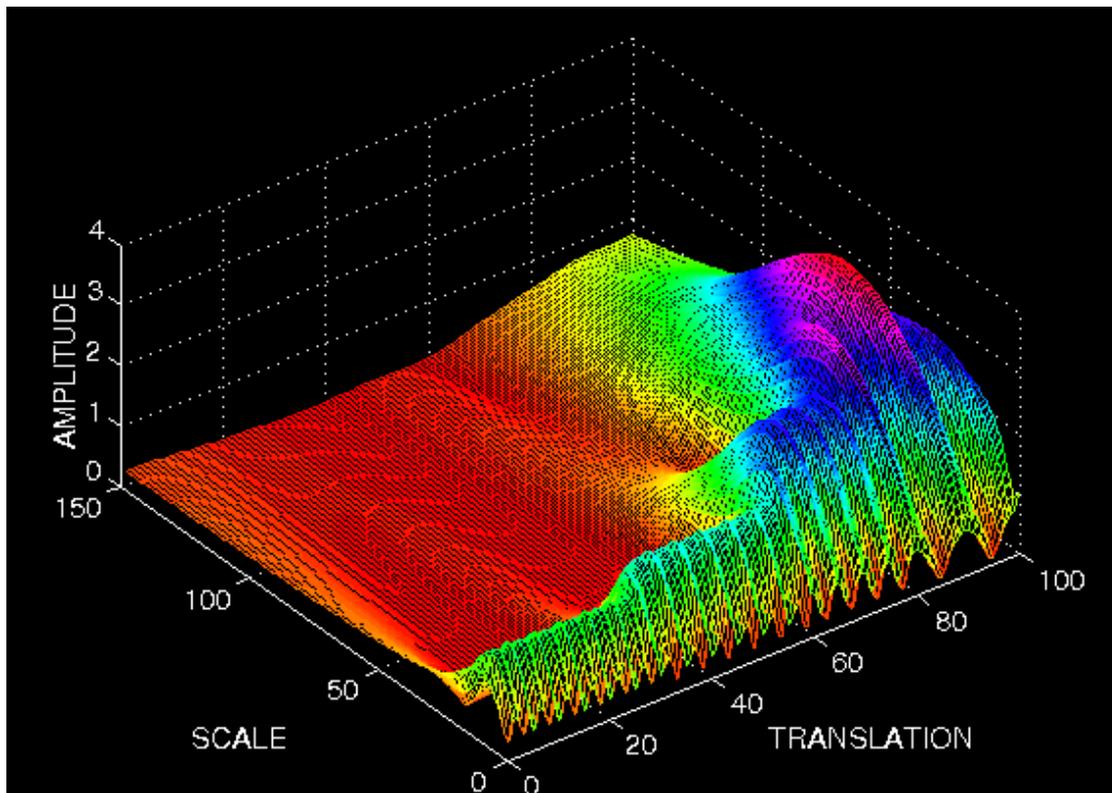


Figure 57. Continuous Wavelet Transform (CWT) Of The Signal

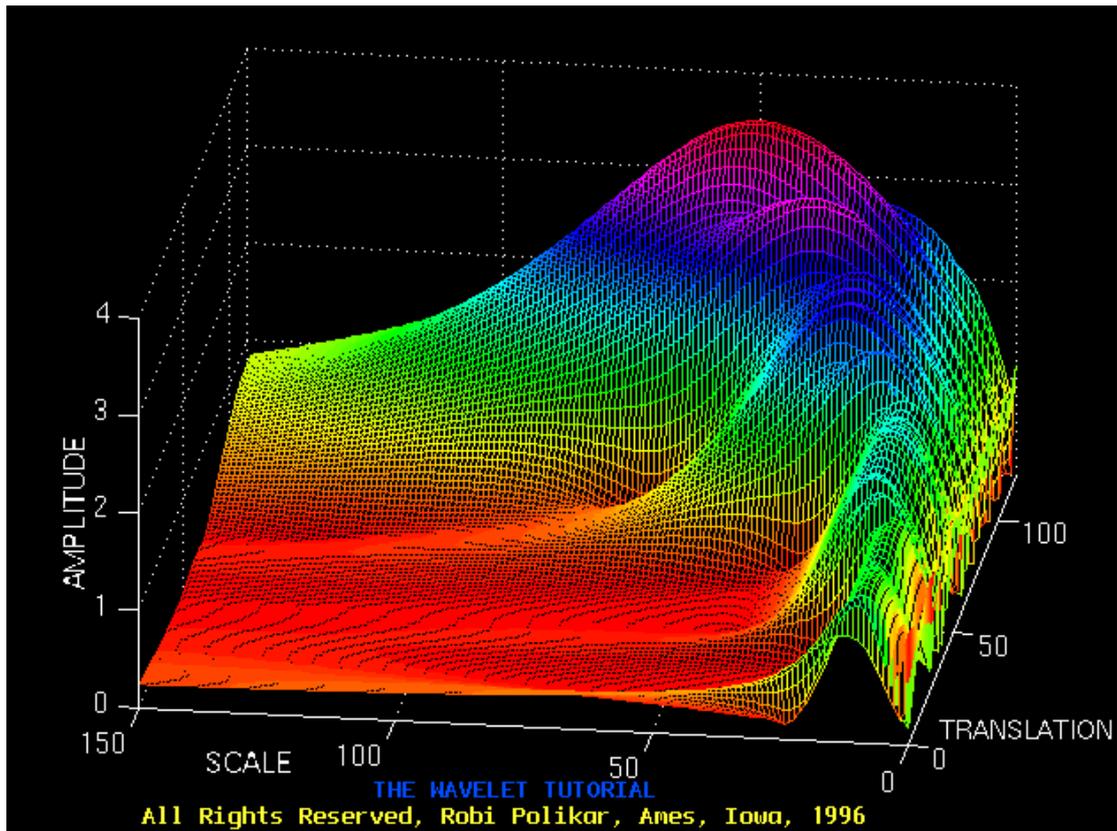


Figure 58. Continuous Wavelet Transform (CWT) Of The Signal from another angle

Now, recall these resolution properties: Unlike the STFT which has a constant resolution at all times and frequencies, the WT has a good time and poor frequency resolution at high frequencies, and good frequency and poor time resolution at low frequencies. Fig. (58) shows the same WT in fig. (57) from another angle to better illustrate the resolution properties: In fig. (58), lower scales (higher frequencies) have better scale resolution (narrower in scale, which means that it is less ambiguous what the exact value of the scale) which correspond to poorer frequency resolution. Similarly, higher scales have scale frequency resolution (wider support in scale, which means it is more ambiguous what the exact value of the scale is), which correspond to better frequency resolution of lower frequencies.

The axes in figs. (57) and (58) are normalized and should be evaluated accordingly. Roughly speaking the 100 points in the translation axis correspond to 1000 ms, and the 150 points on the scale axis correspond to a frequency band of 40 Hz (the numbers on the translation and scale axis do not correspond to seconds and Hz, respectively, they are just the number of samples in the computation).

6.2.3.2 Time And Frequency Resolutions

In this section we will take a closer look at the resolution properties of the wavelet transform. Remember that the resolution problem was the main reason why we switched from STFT to WT.

The illustration in fig. (59) is commonly used to explain how time and frequency resolutions should be interpreted. Every box in fig. (59) corresponds to a value of the wavelet transform in the time-frequency plane. Note that boxes have a certain non-zero area, which implies that the value of a particular point in the time-frequency plane cannot be known. All the points in the time-frequency plane that falls into a box is represented by one value of the WT.

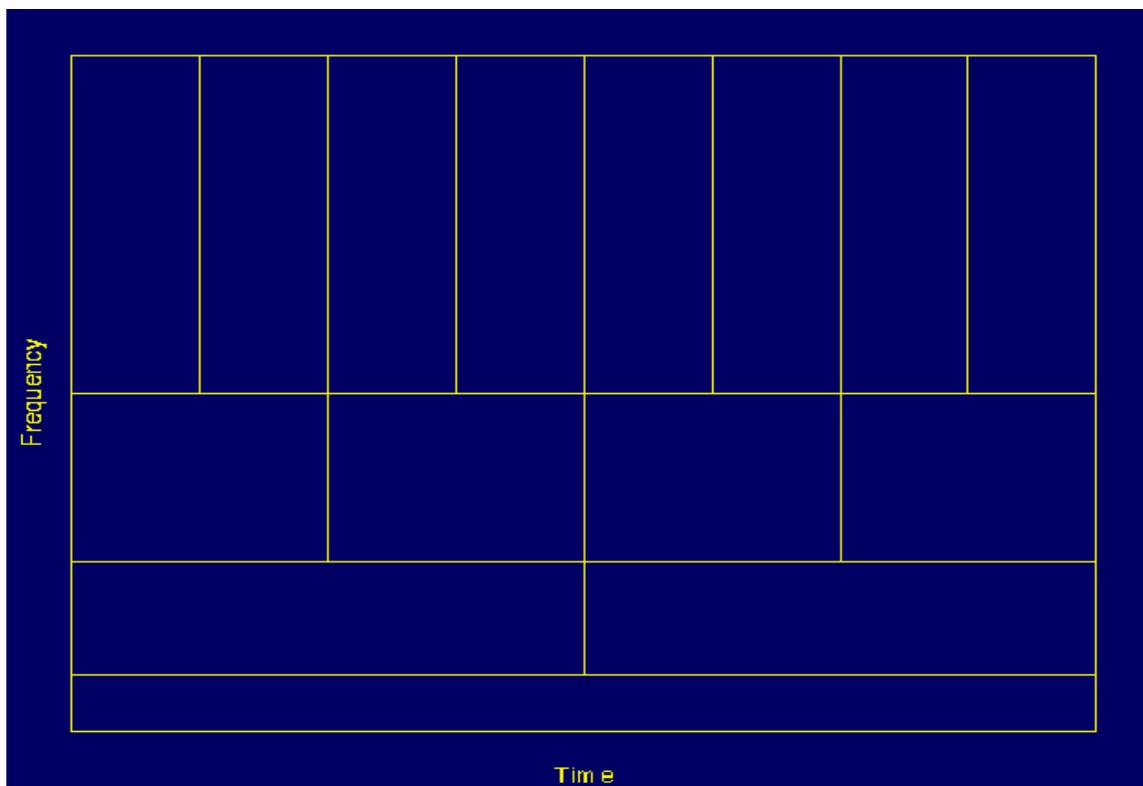


Figure 59. Time And Frequency Resolutions On WT

Let's take a closer look at fig. (59): First thing to notice is that although the widths and heights of the boxes change, the area is constant. That is each box represents an equal portion of the time-frequency plane, but giving different proportions to time and frequency. Note that at low frequencies, the height of the boxes are shorter (which corresponds to better frequency resolutions, since there is less ambiguity regarding the value of the exact frequency), but their widths are longer (which correspond to poor time resolution, since there is more ambiguity regarding the value of the

exact time). At higher frequencies the width of the boxes decreases, i.e., the time resolution gets better, and the heights of the boxes increase, i.e., the frequency resolution gets poorer.

Before concluding this clause, it is worthwhile to mention how the partition looks like in the case of STFT. Recall that in STFT the time and frequency resolutions are determined by the width of the analysis window, which is selected once for the entire analysis, i.e., both time and frequency resolutions are constant. Therefore the time-frequency plane consists of squares in the STFT case.

Regardless of the dimensions of the boxes, the areas of all boxes, both in STFT and WT, are the same and determined by Heisenberg's inequality. As a summary, the area of a box is fixed for each window function (STFT) or mother wavelet (CWT), whereas different windows or mother wavelets can result in different areas. However, all areas are lower bounded by $1/4 \cdot \pi$. That is, we cannot reduce the areas of the boxes as much as we want due to the Heisenberg's uncertainty principle. On the other hand, for a given mother wavelet the dimensions of the boxes can be changed, while keeping the area the same. This is exactly what wavelet transform does.

6.2.3.3 The Wavelet Theory: A Mathematical Approach

This clause describes the main idea of wavelet analysis theory, which can also be considered to be the underlying concept of most of the signal analysis techniques. The FT defined by Fourier use basis functions to analyze and reconstruct a function. Every vector in a vector space can be written as a linear combination of the basis vectors in that vector space, i.e., by multiplying the vectors by some constant numbers, and then by taking the summation of the products. The analysis of the signal involves the estimation of these constant numbers (transform coefficients, or Fourier coefficients, wavelet coefficients, etc). The synthesis, or the reconstruction, corresponds to computing the linear combination equation.

All the definitions and theorems related to this subject can be found in Keiser's book, *A Friendly Guide to Wavelets* but an introductory level knowledge of how basis functions work is necessary to understand the underlying principles of the wavelet theory. Therefore, this information will be presented in this section.

6.2.3.4 Basis Vectors

A basis of a vector space V is a set of linearly independent vectors, such that any vector v in V can be written as a linear combination of these basis vectors. There may be more than one basis for a vector space. However, all of them have the same number of vectors, and this number is known as the dimension of the vector space. For example in two-dimensional space, the basis will have two vectors.

$$v = \sum_k u^k \cdot b_k \quad \text{Equation 10}$$

Eq. (10) shows how any vector v can be written as a linear combination of the basis vectors b_k and the corresponding coefficients u^k .

This concept, given in terms of vectors, can easily be generalized to functions, by replacing the basis vectors b_k with basis functions $\phi_k(t)$, and the vector v with a function $f(t)$. Eq. (10) then becomes

$$f(t) = \sum_k \mu_k \phi_k(t) \quad \text{Equation 11}$$

The complex exponential (sines and cosines) functions are the basis functions for the FT. Furthermore, they are orthogonal functions, which provide some desirable properties for reconstruction.

Let $f(t)$ and $g(t)$ be two functions in $L^2 [a,b]$. ($L^2 [a,b]$ denotes the set of square integrable functions in the interval $[a,b]$). The inner product of two functions is defined by Eq. (12):

$$\langle f(t), g(t) \rangle = \int_a^b f(t) \cdot g^*(t) dt \quad \text{Equation 12}$$

According to the above definition of the inner product, the CWT can be thought of as the inner product of the test signal with the basis functions $\Psi_{\tau,s}(t)$:

$$CWT_x^\Psi(\tau, s) = \Psi_x^\Psi(\tau, s) = \int x(t) \cdot \psi_{\tau,s}(t) dt \quad \text{Equation 13}$$

where,

$$\psi_{\tau,s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad \text{Equation 14}$$

This definition of the CWT shows that the wavelet analysis is a measure of similarity between the basis functions (wavelets) and the signal itself. Here the similarity is in the sense of similar frequency content. The calculated CWT coefficients refer to the closeness of the signal to the wavelet at the current scale .

This further clarifies the previous discussion on the correlation of the signal with the wavelet at a certain scale. If the signal has a major component of the frequency corresponding to the current scale, then the wavelet (the basis function) at the current scale will be similar or close to the signal at the particular location where this frequency component occurs. Therefore, the CWT coefficient computed at this point in the time-scale plane will be a relatively large number.

6.2.3.5 Inner Products, Orthogonality, and Orthonormality

Two vectors v , w are said to be orthogonal if their inner product equals zero:

$$\langle v, w \rangle = \sum_n v_n \cdot w_n^* = 0 \quad \text{Equation 15}$$

Similarly, two functions f and g are said to be orthogonal to each other if their inner product is zero:

$$\langle f(t), g(t) \rangle = \int_a^b f(t) \cdot g^*(t) dt = 0 \quad \text{Equation 16}$$

A set of vectors $\{v_1, v_2, \dots, v_n\}$ is said to be orthonormal, if they are pairwise orthogonal to each other, and all have length “1”. This can be expressed as:

$$\langle v_m, v_n \rangle = \delta_{mn} \quad \text{Equation 17}$$

Similarly, a set of functions $\{\phi_k(t)\}$, $k=1,2,3,\dots$, is said to be orthonormal if

$$\int_a^b \phi_k(t) \cdot \phi_l^*(t) dt = 0 \quad k \neq l \quad (\text{orthogonality condition}) \quad \text{Equation 18}$$

and

$$\int_a^b \{|\phi_k(t)|\}^2 dt = 1 \quad \text{Equation 19}$$

or equivalently

$$\int_a^b \phi_k(t) \cdot \phi_l^*(t) dt = \delta_{kl} \quad \text{Equation 20}$$

where, δ_{kl} is the Kronecker delta function, defined as:

$$\delta_{kl} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases} \quad \text{Equation 21}$$

As stated above, there may be more than one set of basis functions (or vectors). Among them, the orthonormal basis functions (or vectors) are of particular importance because of the nice properties they provide in finding these analysis coeffi-

cients. The orthonormal bases allow computation of these coefficients in a very simple and straightforward way using the orthonormality property.

For orthonormal bases, the coefficients, μ_k , can be calculated as

$$\mu_k = \langle f, \phi_k \rangle = \int f(t) \phi_k^*(t) dt \quad \text{Equation 22}$$

and the function $f(t)$ can then be reconstructed by eq. (11) by substituting the μ_k coefficients. This yields to

$$f(t) = \sum_k \mu_k \phi_k(t) = \sum_k \langle f, \phi_k \rangle \phi_k(t) \quad \text{Equation 23}$$

Orthonormal bases may not be available for every type of application where a generalized version, biorthogonal bases can be used. The term “biorthogonal” refers to two different bases which are orthogonal to each other, but each do not form an orthogonal set.

In some applications, however, biorthogonal bases also may not be available in which case frames can be used. Frames constitute an important part of wavelet theory, and interested readers are referred to Kaiser's book mentioned earlier.

Some examples of continuous wavelet transform are presented next. The figures given in the examples were generated by a program written to compute the CWT.

Before we close this clause, it would be helpful to include two mother wavelets commonly used in wavelet analysis. The Mexican Hat wavelet is defined as the second derivative of the Gaussian function:

$$w(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \quad \text{Equation 24}$$

which is

$$\psi(t) = \frac{1}{\sqrt{2\pi}\sigma^3} \left(e^{-\frac{t^2}{2\sigma^2}} \cdot \left(\frac{t^2}{\sigma^2} - 1 \right) \right) \quad \text{Equation 25}$$

The Morlet wavelet is defined as

$$w(t) = e^{i\alpha t} \cdot e^{-\frac{t^2}{2\sigma}} \quad \text{Equation 26}$$

where α is a modulation parameter, and σ is the scaling parameter that affects the width of the window.

6.2.3.6 Examples

All of the examples that are given below correspond to real-life non-stationary signals. These signals are drawn from a database signals that includes event related potentials of normal people, and patients with Alzheimer's disease. Since these are not test signals like simple sinusoids, it is not as easy to interpret them. They are shown here only to give an idea of how real-life CWTs look like.

The following signal shown in fig. (60) belongs to a normal person and the following is its CWT. The numbers on the axes are of no importance to us. Those numbers simply show that the CWT was computed at 350 translation and 60 scale locations on the translation-scale plane. The important point to note here is the fact that the computation is not a true continuous WT, as it is apparent from the computation at finite number of locations. This is only a discretized version of the CWT, which is explained later on this page. Note, however, that this is NOT discrete wavelet transform (DWT) which is the topic of the clause that follows next.

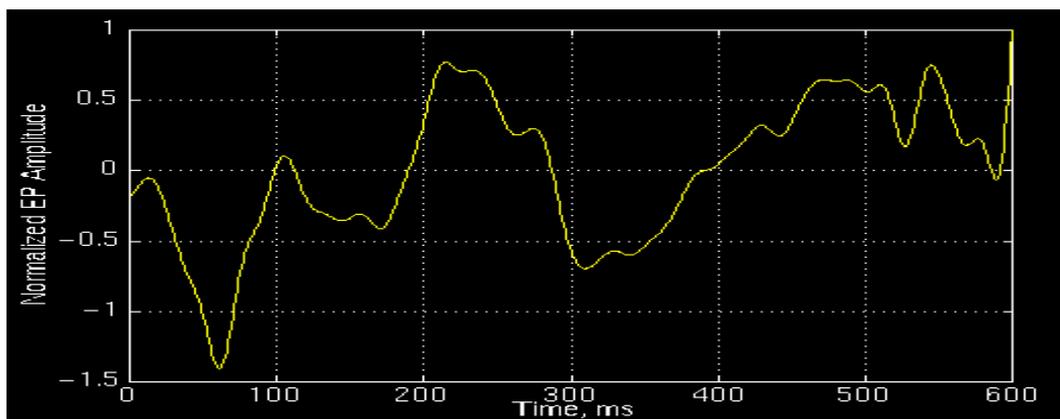


Figure 60. Event Related Potential Of Normal People

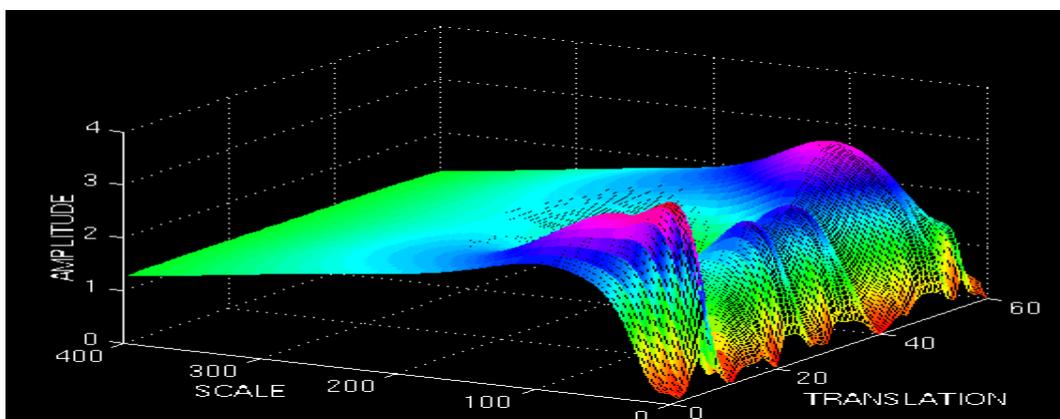


Figure 61. CWT Of An Event Related Potential Of Normal People

and the fig. (62) plots the same transform from a different angle for better visualization.

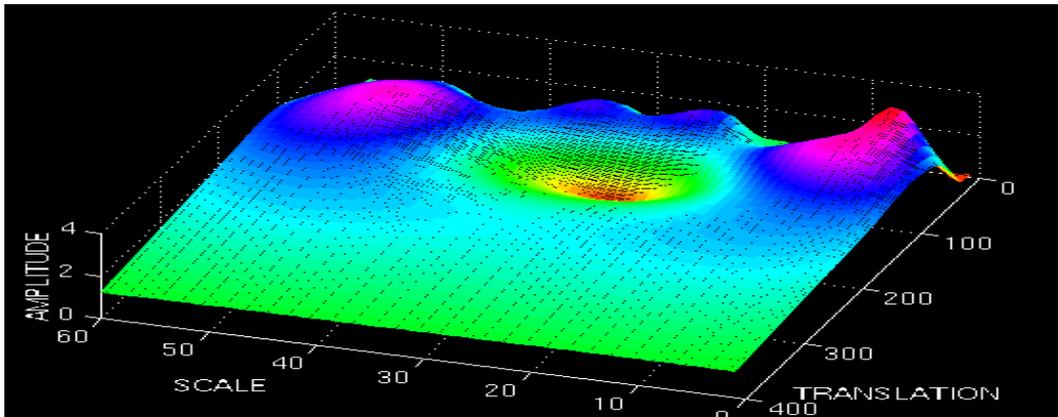


Figure 62. CWT Of An Event Related Potential Of Normal People from another angle

Fig. (63) plots an event related potential of a patient diagnosed with Alzheimer's disease

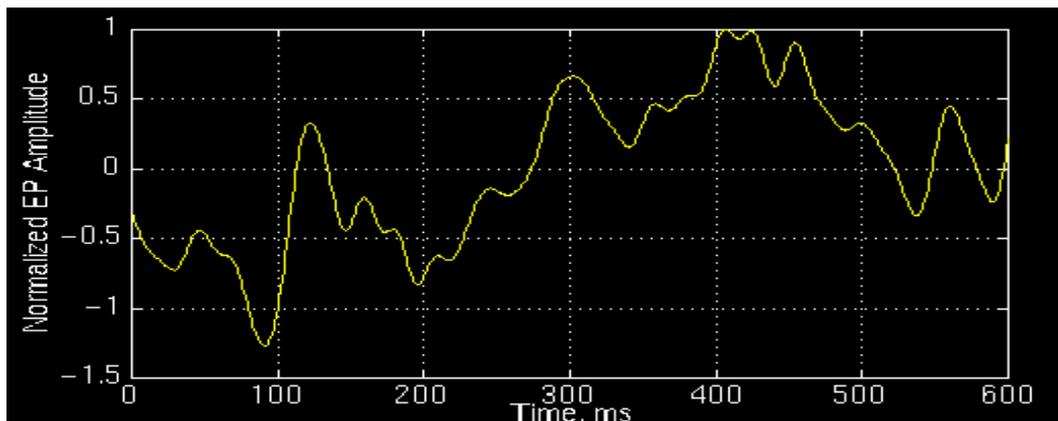


Figure 63. Event Related Potential Of A Patient Diagnosed With Alzheimer's Disease

and fig. (64) and (65) illustrates its CWT from two different angles:

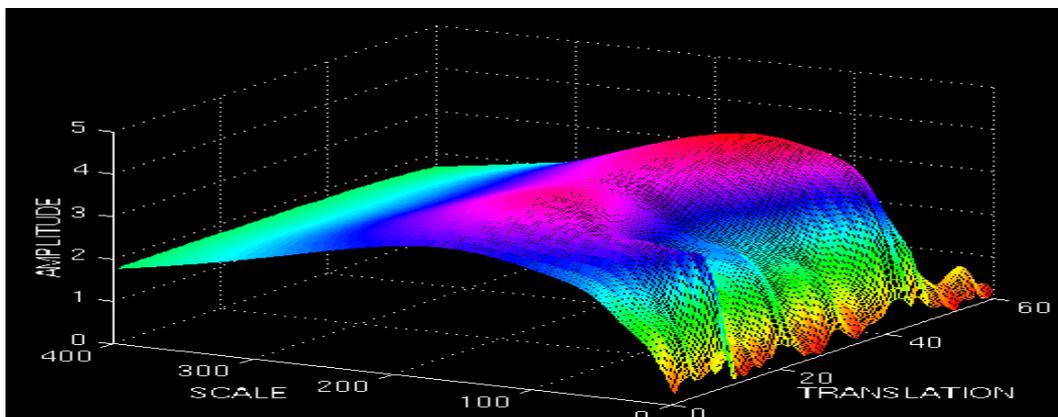


Figure 64. CWT Of An Event Related Potential Of A Patient Diagnosed With Alzheimer's Disease

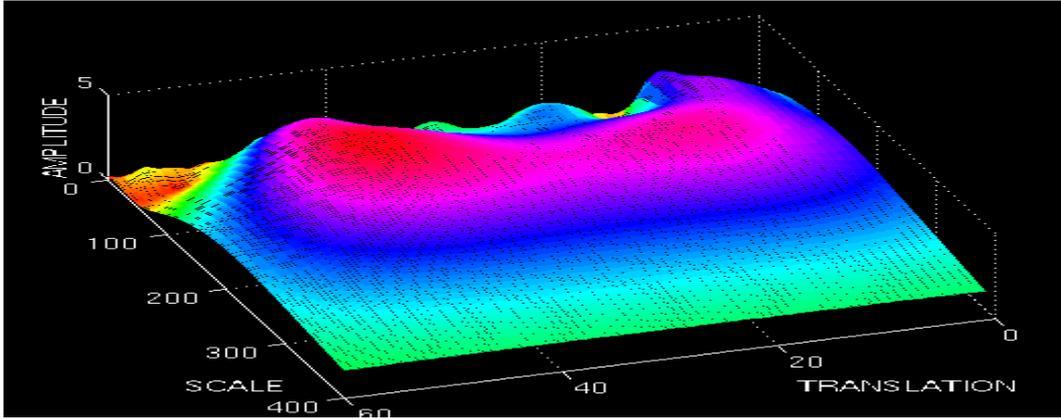


Figure 65. CWT Of An Event Related Potential Of A Patient Diagnosed With Alzheimer's Disease from another angle

6.2.3.7 The Wavelet Synthesis

The continuous wavelet transform is a reversible transform, provided that eq. (28) is satisfied. Fortunately, this is a very non-restrictive requirement. The continuous wavelet transform is reversible if eq. (28) is satisfied, even though the basis functions are in general may not be orthonormal. The reconstruction is possible by using the following reconstruction formula:

$$x(t) = \frac{1}{c_\psi^2} \int \int \Psi_x^\psi(\tau, s) \frac{1}{s^2} \psi\left(\frac{t-\tau}{s}\right) dt ds \quad \text{Equation 27}$$

Inverse Wavelet Transform

where c_ψ is a constant that depends on the wavelet used. The success of the reconstruction depends on this constant called, the admissibility constant, to satisfy the following admissibility condition:

$$c_\psi = \left\{ 2\pi \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi \right\}^{\frac{1}{2}} < \infty \quad \text{Equation 28}$$

Admissibility Condition

where $\hat{\psi}(\xi)$ is the FT of $\psi(t)$. Eq. (28) implies that $\hat{\psi}(0) = 0$, which is

$$\int \psi(t) dt = 0 \quad \text{Equation 29}$$

As stated above, eq. (29) is not a very restrictive requirement since many wavelet functions can be found whose integral is zero. For eq. (29) to be satisfied, the wavelet must be oscillatory.

6.2.3.8 Discretization of the Continuous Wavelet Transform: The Wavelet Series

In today's world, computers are used to do most computations (well,...ok... almost all computations). It is apparent that neither the FT, nor the STFT, nor the CWT can be practically computed by using analytical equations, integrals, etc. It is therefore necessary to discretize the transforms. As in the FT and STFT, the most intuitive way of doing this is simply sampling the time-frequency (scale) plane. Again intuitively, sampling the plane with a uniform sampling rate sounds like the most natural choice. However, in the case of WT, the scale change can be used to reduce the sampling rate.

At higher scales (lower frequencies), the sampling rate can be decreased, according to Nyquist's rule. In other words, if the time-scale plane needs to be sampled with a sampling rate of N_1 at scale s_1 , the same plane can be sampled with a sampling rate of N_2 , at scale s_2 , where, $s_1 < s_2$ (corresponding to frequencies $f_1 > f_2$) and $N_2 < N_1$. The actual relationship between N_1 and N_2 is

$$N_2 = \frac{s_1}{s_2} N_1 \quad \text{Equation 30}$$

or

$$N_2 = \frac{f_1}{f_2} N_1 \quad \text{Equation 31}$$

In other words, at lower frequencies the sampling rate can be decreased which will save a considerable amount of computation time.

It should be noted at this time, however, that the discretization can be done in any way without any restriction as far as the analysis of the signal is concerned. If synthesis is not required, even the Nyquist criteria does not need to be satisfied. The restrictions on the discretization and the sampling rate become important if, and only if, the signal reconstruction is desired. Nyquist's sampling rate is the minimum sampling rate that allows the original continuous time signal to be reconstructed from its discrete samples. The basis vectors that are mentioned earlier are of particular importance for this reason.

As mentioned earlier, the wavelet $\psi(\tau,s)$ satisfying eq. (28), allows reconstruction of the signal by eq. (27). However, this is true for the continuous transform. The question is: can we still reconstruct the signal if we discretize the time and scale parameters? The answer is "yes", under certain conditions.

The scale parameter s is discretized first on a logarithmic grid. The time parameter is then discretized with respect to the scale parameter, i.e., a different sampling rate is used for every scale. In other words, the sampling is done on the dyadic sampling grid shown in fig. (66).

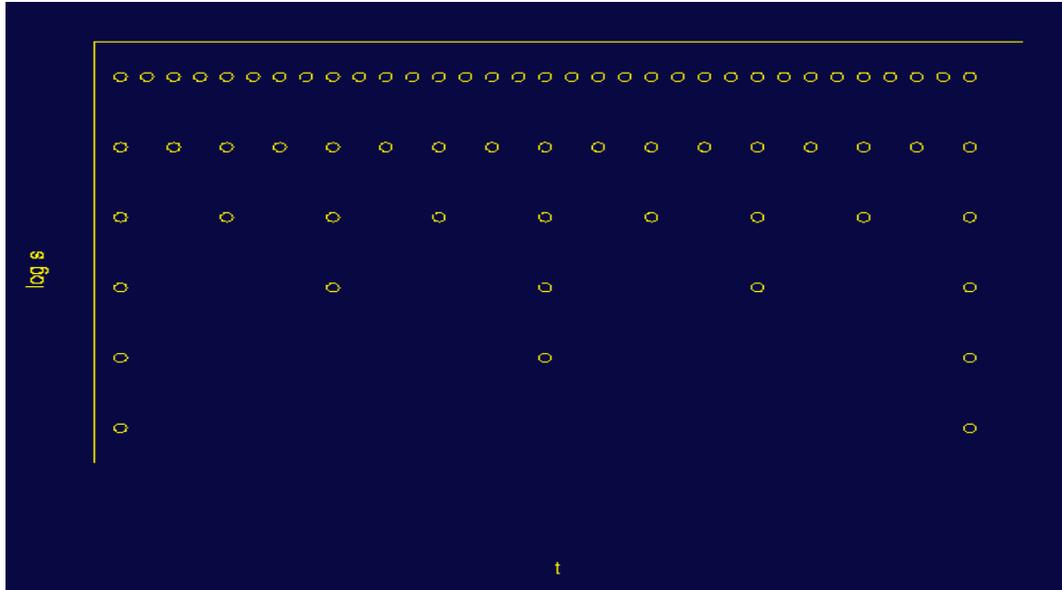


Figure 66. Dyadic Sampling Grid

Think of the area covered by the axes as the entire time-scale plane. The CWT assigns a value to the continuum of points on this plane. Therefore, there are an infinite number of CWT coefficients. First consider the discretization of the scale axis. Among that infinite number of points, only a finite number are taken, using a logarithmic rule. The base of the logarithm depends on the user. The most common value is 2 because of its convenience. If 2 is chosen, only the scales 2, 4, 8, 16, 32, 64...etc. are computed. If the value was 3, the scales 3, 9, 27, 81, 243...etc. would have been computed. The time axis is then discretized according to the discretization of the scale axis. Since the discrete scale changes by factors of 2, the sampling rate is reduced for the time axis by a factor of 2 at every scale.

Note that at the lowest scale ($s=2$), only 32 points of the time axis are sampled (for the particular case given in fig. (66)). At the next scale value, $s=4$, the sampling rate of time axis is reduced by a factor of 2 since the scale is increased by a factor of 2, and therefore, only 16 samples are taken. At the next step, $s=8$ and 8 samples are taken in time, and so on.

Although it is called the time-scale plane, it is more accurate to call it the translation-scale plane, because “time” in the transform domain actually corresponds

to the shifting of the wavelet in time. For the wavelet series, the actual time is still continuous.

Similar to the relationship between continuous Fourier transform, Fourier series and the discrete Fourier transform, there is a continuous wavelet transform, a semi-discrete wavelet transform (also known as wavelet series) and a discrete wavelet transform.

Expressing the above discretization procedure in mathematical terms, the scale discretization is $s = s_0^j$, and translation discretization is $\tau = k \cdot s_0^j \cdot \tau_0$ where $s_0 > 1$ and $\tau_0 > 0$. Note, how the translation discretization is dependent on scale discretization with s_0 .

The continuous wavelet function

$$\Psi_{\tau,s} = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right) \quad \text{Equation 32}$$

$$\Psi_{j,k} = s_0^{-\frac{j}{2}} \Psi(s_0^{-j}t - k\tau_0) \quad \text{Equation 33}$$

by inserting $s = s_0^j$, and $\tau = k \cdot s_0^j \tau_0$.

If $\{\Psi_{j,k}\}$ constitutes an orthonormal basis, the wavelet series transform becomes

$$\Psi_x^{\Psi_{j,k}} = \int x(t) \Psi_{j,k}^*(t) dt \quad \text{Equation 34}$$

or

$$x(t) = c_\Psi \sum_J \sum_k \Psi_x^{\Psi_{j,k}} \Psi_{j,k}(t) \quad \text{Equation 35}$$

A wavelet series requires that $\{\Psi_{(j,k)}\}$ are either orthonormal, biorthogonal, or frame. If $\{\Psi_{(j,k)}\}$ are not orthonormal, eq. (34) becomes

$$\Psi_x^{\Psi_{j,k}} = \int x(t) \hat{\Psi}_{j,k}^*(t) dt \quad \text{Equation 36}$$

where $\{\hat{\Psi}_{j,k}^*(t)\}$, is either the dual biorthogonal basis or dual frame (Note that * denotes the conjugate).

If $\{\Psi_{(j,k)}\}$ are orthonormal or biorthogonal, the transform will be non-redundant, where as if they form a frame, the transform will be redundant. On the other hand, it is much easier to find frames than it is to find orthonormal or biorthogonal bases.

The following analogy may clear this concept. Consider the whole process as looking at a particular object. The human eyes first determine the coarse view which depends on the distance of the eyes to the object. This corresponds to adjusting the scale parameter s_0^{-j} . When looking at a very close object, with great detail, j is negative and large (low scale, high frequency, analyses the detail in the signal). Moving the head (or eyes) very slowly and with very small increments (of angle, of distance, depending on the object that is being viewed), corresponds to small values of $\tau = k \cdot s_0^j \tau_0$. Note that when j is negative and large, it corresponds to small changes in time, τ , (high sampling rate) and large changes in s_0^{-j} (low scale, high frequencies, where the sampling rate is high). The scale parameter can be thought of as magnification too.

How low can the sampling rate be and still allow reconstruction of the signal? This is the main question to be answered to optimize the procedure. The most convenient value (in terms of programming) is found to be “2” for s_0 and “1” for τ . Obviously, when the sampling rate is forced to be as low as possible, the number of available orthonormal wavelets is also reduced.

The continuous wavelet transform examples that were given in this chapter were actually the wavelet series of the given signals. The parameters were chosen depending on the signal. Since the reconstruction was not needed, the sampling rates were sometimes far below the critical value where s_0 varied from 2 to 10, and τ_0 varied from 2 to 8, for different examples.

This concludes the basic theory on the Wavelet Transform, however there is one thing left to be discussed. Even though the discretized wavelet transform can be computed on a computer, this computation may take anywhere from a couple seconds to couple hours depending on your signal size and the resolution you want. An amazingly fast algorithm is actually available to compute the wavelet transform of a signal. The discrete wavelet transform (DWT) is introduced in the next and final clause of this section.

6.2.3.9 Discrete Wavelet Transform

Although the discretized continuous wavelet transform enables the computation of the continuous wavelet transform by computers, it is not a true discrete transform. As a matter of fact, the wavelet series is simply a sampled version of the CWT,

and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy, on the other hand, requires a significant amount of computation time and resources. The discrete wavelet transform (DWT), on the other hand, provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time.

The DWT is considerably easier to implement when compared to the CWT. The basic concepts of the DWT will be introduced in this section along with its properties and the algorithms used to compute it. As in the previous chapters, examples are provided to aid in the interpretation of the DWT.

The foundations of the DWT go back to 1976 when Croiser, Esteban, and Galand devised a technique to decompose discrete time signals. Crochiere, Weber, and Flanagan did a similar work on coding of speech signals in the same year. They named their analysis scheme as subband coding. In 1983, Burt defined a technique very similar to subband coding and named it pyramidal coding which is also known as multiresolution analysis. Later in 1989, Vetterli and Le Gall made some improvements to the subband coding scheme, removing the existing redundancy in the pyramidal coding scheme. Subband coding is explained below. A detailed coverage of the discrete wavelet transform and theory of multiresolution analysis can be found in a number of articles and books that are available on this topic, and it is beyond the scope of this paper.

6.2.3.10 The Subband Coding and The Multiresolution Analysis

The main idea is the same as it is in the CWT. A time-scale representation of a digital signal is obtained using digital filtering techniques. Recall that the CWT is a correlation between a wavelet at different scales and the signal with the scale (or the frequency) being used as a measure of similarity. The continuous wavelet transform was computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating over all times. In the discrete case, filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies.

The resolution of the signal, which is a measure of the amount of detail information in the signal, is changed by the filtering operations, and the scale is changed

by upsampling and downsampling (subsampling) operations. Subsampling a signal corresponds to reducing the sampling rate, or removing some of the samples of the signal. For example, subsampling by two refers to dropping every other sample of the signal. Subsampling by a factor n reduces the number of samples in the signal n times.

Upsampling a signal corresponds to increasing the sampling rate of a signal by adding new samples to the signal. For example, upsampling by two refers to adding a new sample, usually a zero or an interpolated value, between every two samples of the signal. Upsampling a signal by a factor of n increases the number of samples in the signal by a factor of n .

Although it is not the only possible choice, DWT coefficients are usually sampled from the CWT on a dyadic grid, i.e., $s_0 = 2$ and $\tau_0 = 1$, yielding $s=2^j$ and $\tau =k \cdot 2^j$, as described in the previous clause. Since the signal is a discrete time function, the terms function and sequence will be used interchangeably in the following discussion. This sequence will be denoted by $x[n]$, where n is an integer.

The procedure starts with passing this signal (sequence) through a half band digital lowpass filter with impulse response $h(n)$. Filtering a signal corresponds to the mathematical operation of convolution of the signal with the impulse response of the filter. The convolution operation in discrete time is defined as follows:

$$x(n) * h(n) = \sum_{k=-\infty}^{\infty} x(k) \cdot h(n - k) \quad \text{Equation 37}$$

A half band lowpass filter removes all frequencies that are above half of the highest frequency in the signal. For example, if a signal has a maximum of 1000 Hz component, then half band lowpass filtering removes all the frequencies above 500 Hz.

The unit of frequency is of particular importance at this time. In discrete signals, frequency is expressed in terms of radians. Accordingly, the sampling frequency of the signal is equal to 2π radians in terms of radial frequency. Therefore, the highest frequency component that exists in a signal will be π radians, if the signal is sampled at Nyquist's rate (which is twice the maximum frequency that exists in the signal); that is, the Nyquist's rate corresponds to π rad/s in the discrete frequency domain. Therefore using Hz is not appropriate for discrete signals. However, Hz is used whenever it is needed to clarify a discussion, since it is very common to think of frequency

in terms of Hz. It should always be remembered that the unit of frequency for discrete time signals is radians.

After passing the signal through a half band lowpass filter, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has a highest frequency of $\pi/2$ radians instead of π radians. Simply discarding every other sample will subsample the signal by two, and the signal will then have half the number of points. The scale of the signal is now doubled. Note that the lowpass filtering removes the high frequency information, but leaves the scale unchanged. Only the subsampling process changes the scale. Resolution, on the other hand, is related to the amount of information in the signal, and therefore, it is affected by the filtering operations. Half band lowpass filtering removes half of the frequencies, which can be interpreted as losing half of the information. Therefore, the resolution is halved after the filtering operation. Note, however, the subsampling operation after filtering does not affect the resolution, since removing half of the spectral components from the signal makes half the number of samples redundant anyway. Half the samples can be discarded without any loss of information. In summary, the lowpass filtering halves the resolution, but leaves the scale unchanged. The signal is then subsampled by 2 since half of the number of samples are redundant. This doubles the scale.

This procedure can mathematically be expressed as

$$y(n) = \sum_{k=-\infty}^{\infty} h(k) \cdot x(2n - k) \quad \text{Equation 38}$$

Having said that, we now look how the DWT is actually computed: The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and highpass filters, respectively. The decomposition of the signal into different frequency bands is simply obtained by successive highpass and lowpass filtering of the time domain signal. The original signal $x(n)$ is first passed through a halfband highpass filter $g(n)$ and a lowpass filter $h(n)$. After the filtering, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has a highest frequency of $\pi/2$ radians instead of π . The signal can therefore be subsampled by 2, simply by discarding every other sample. This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{\text{high}}(k) = \sum_n x(n) \cdot g(2k - n) \quad \text{Equation 39}$$

$$y_{\text{low}}(k) = \sum_n x(n) \cdot h(2k - n) \quad \text{Equation 40}$$

where $y_{\text{high}}(k)$ and $y_{\text{low}}(k)$ are the outputs of the highpass and lowpass filters, respectively, after subsampling by 2.

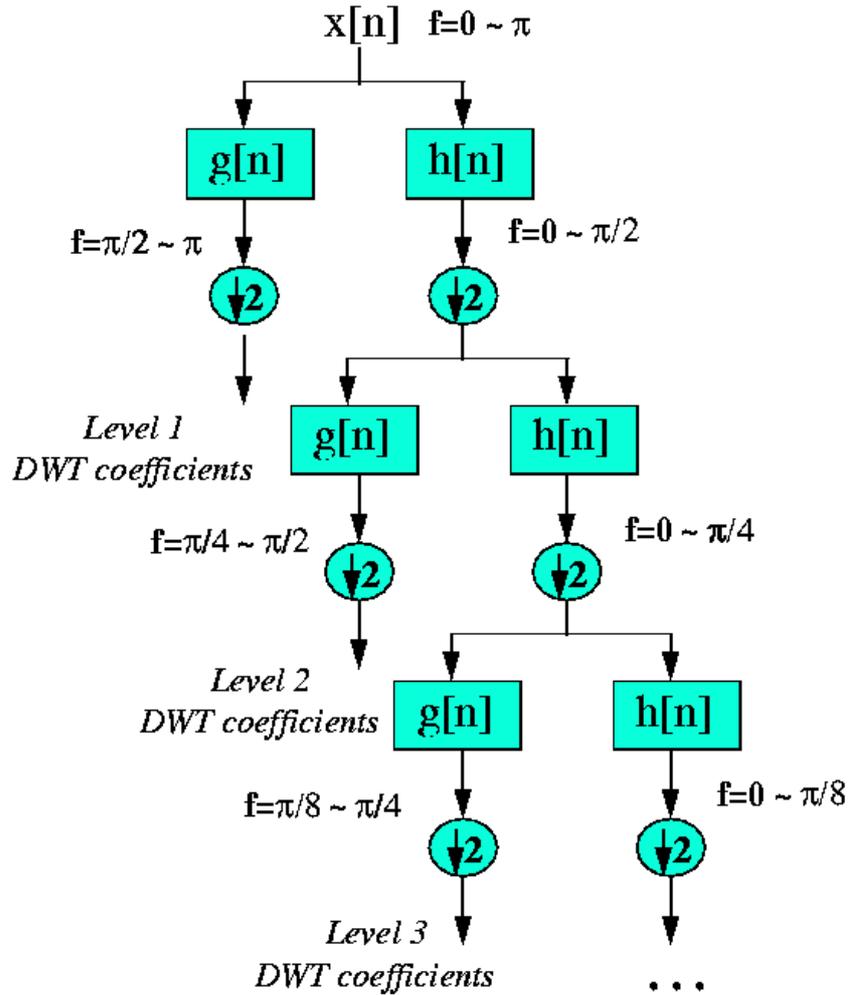


Figure 67. The Subband Coding Algorithm

This decomposition halves the time resolution since only half the number of samples now characterizes the entire signal. However, this operation doubles the frequency resolution, since the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency by half. The above procedure, which is also known as the subband coding, can be repeated for further decomposition. At every level, the filtering and subsampling will result in half

the number of samples (and hence half the time resolution) and half the frequency band spanned (and hence double the frequency resolution). Fig. (67) illustrates this procedure, where $x(n)$ is the original signal to be decomposed, and $h(n)$ and $g(n)$ are lowpass and highpass filters, respectively. The bandwidth of the signal at every level is marked on the figure as “F”.

As an example, suppose that the original signal $x(n)$ has 512 sample points, spanning a frequency band of zero to π rad/s. At the first decomposition level, the signal is passed through the highpass and lowpass filters, followed by subsampling by 2. The output of the highpass filter has 256 points (hence half the time resolution), but it only spans the frequencies $\pi/2$ to π rad/s (hence double the frequency resolution). These 256 samples constitute the first level of DWT coefficients. The output of the lowpass filter also has 256 samples, but it spans the other half of the frequency band, frequencies from 0 to $\pi/2$ rad/s. This signal is then passed through the same lowpass and highpass filters for further decomposition. The output of the second lowpass filter followed by subsampling has 128 samples spanning a frequency band of 0 to $\pi/4$ rad/s, and the output of the second highpass filter followed by subsampling has 128 samples spanning a frequency band of $\pi/4$ to $\pi/2$ rad/s. The second highpass filtered signal constitutes the second level of DWT coefficients. This signal has half the time resolution, but twice the frequency resolution of the first level signal. In other words, time resolution has decreased by a factor of 4, and frequency resolution has increased by a factor of 4 compared to the original signal. The lowpass filter output is then filtered once again for further decomposition. This process continues until two samples are left. For this specific example there would be 8 levels of decomposition, each having half the number of samples of the previous level. The DWT of the original signal is then obtained by concatenating all coefficients starting from the last level of decomposition (remaining two samples, in this case). The DWT will then have the same number of coefficients as the original signal.

The frequencies that are most prominent in the original signal will appear as high amplitudes in that region of the DWT signal that includes those particular frequencies. The difference of this transform from the Fourier transform is that the time localization of these frequencies will not be lost. However, the time localization will have a resolution that depends on which level they appear. If the main information of the signal lies in the high frequencies, as happens most often, the time localization of

these frequencies will be more precise, since they are characterized by more number of samples. If the main information lies only at very low frequencies, the time localization will not be very precise, since few samples are used to express signal at these frequencies. This procedure in effect offers a good time resolution at high frequencies, and good frequency resolution at low frequencies. Most practical signals encountered are of this type.

The frequency bands that are not very prominent in the original signal will have very low amplitudes, and that part of the DWT signal can be discarded without any major loss of information, allowing data reduction. Fig. (68) illustrates an example of how DWT signals look like and how data reduction is provided. Fig. (68.a) shows a typical 512-sample signal that is normalized to unit amplitude. The horizontal axis is the number of samples, whereas the vertical axis is the normalized amplitude. Fig. (68.b) shows the 8 level DWT of the signal in fig. (68.a.) The last 256 samples in this signal correspond to the highest frequency band in the signal, the previous 128 samples correspond to the second highest frequency band and so on. It should be noted that only the first 64 samples, which correspond to lower frequencies of the analysis, carry relevant information and the rest of this signal has virtually no information. Therefore, all but the first 64 samples can be discarded without any loss of information. This is how DWT provides a very effective data reduction scheme.

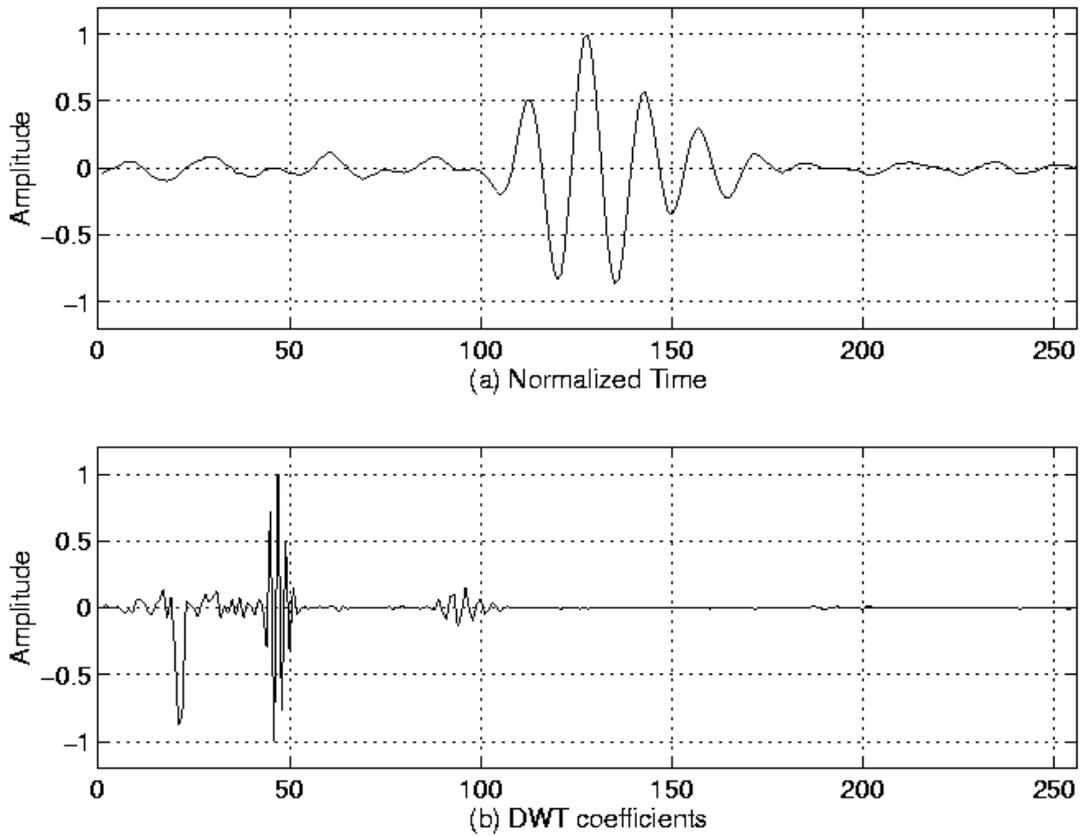


Figure 68.a,b Example of a DWT

We will revisit this example, since it provides important insight to how DWT should be interpreted. Before that, however, we need to conclude our mathematical analysis of the DWT.

One important property of the discrete wavelet transform is the relationship between the impulse responses of the highpass and lowpass filters. The highpass and lowpass filters are not independent of each other, and they are related by

$$\mathbf{g(L-1-n)} = \mathbf{(-1)^n \cdot h(n)} \quad \text{Equation 41}$$

where $g(n)$ is the highpass, $h(n)$ is the lowpass filter, and L is the filter length (in number of points). Note that the two filters are odd index alternated reversed versions of each other. Lowpass to highpass conversion is provided by the $(-1)^n$ term. Filters satisfying this condition are commonly used in signal processing, and they are known as the Quadrature Mirror Filters (QMF). The two filtering and subsampling operations can be expressed by

$$y_{\text{high}}(k) = \sum_n x(n) \cdot g(-n + 2k) \quad \text{Equation 42}$$

$$y_{\text{low}}(k) = \sum_n x(n) \cdot h(-n + 2k) \quad \text{Equation 43}$$

The reconstruction in this case is very easy since halfband filters form orthonormal bases. The above procedure is followed in reverse order for the reconstruction. The signals at every level are upsampled by two, passed through the synthesis filters $g'(n)$, and $h'(n)$ (highpass and lowpass, respectively), and then added. The interesting point here is that the analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each layer)

$$x(n) = \sum_{k=-\infty}^{\infty} (y_{\text{high}}(k) \cdot g(-n + 2k) + y_{\text{low}}(k) \cdot h(-n + 2k)) \quad \text{Equation 44}$$

However, if the filters are not ideal halfband, then perfect reconstruction cannot be achieved. Although it is not possible to realize ideal filters, under certain conditions it is possible to find filters that provide perfect reconstruction. The most famous ones are the ones developed by Ingrid Daubechies, and they are known as Daubechies' wavelets, which by the way are used on the mathematical examples of this paper.

Note that due to successive subsampling by 2, the signal length must be a power of 2, or at least a multiple of power of 2, in order this scheme to be efficient. The length of the signal determines the number of levels that the signal can be decomposed to. For example, if the signal length is 1024, ten levels of decomposition are possible.

Interpreting the DWT coefficients can sometimes be rather difficult because the way DWT coefficients are presented is rather peculiar. To make a real long story real short, DWT coefficients of each level are concatenated, starting with the last level. An example is in order to make this concept clear:

Suppose we have a 256-sample long signal sampled at 10 MHz and we wish to obtain its DWT coefficients. Since the signal is sampled at 10 MHz, the highest frequency component that exists in the signal is 5 MHz. At the first level, the signal is passed through the lowpass filter $h(n)$, and the highpass filter $g(n)$, the outputs of which are subsampled by two. The highpass filter output is the first level DWT coefficients. There are 128 of them, and they represent the signal in the [2.5 5] MHz range. These 128 samples are the last 128 samples plotted. The lowpass filter output, which also has 128 samples, but spanning the frequency band of [0 2.5] MHz, are further decomposed by passing them through the same $h(n)$ and $g(n)$. The output of the second highpass filter is the level 2 DWT coefficients and these 64 samples precede

the 128 level 1 coefficients in the plot. The output of the second lowpass filter is further decomposed, once again by passing it through the filters $h(n)$ and $g(n)$. The output of the third highpass filter is the level 3 DWT coefficients. These 32 samples precede the level 2 DWT coefficients in the plot.

The procedure continues until only 1 DWT coefficient can be computed at level 9. This one coefficient is the first to be plotted in the DWT plot. This is followed by 2 level 8 coefficients, 4 level 7 coefficients, 8 level 6 coefficients, 16 level 5 coefficients, 32 level 4 coefficients, 64 level 3 coefficients, 128 level 2 coefficients and finally 256 level 1 coefficients. Note that less and less number of samples is used at lower frequencies, therefore, the time resolution decreases as frequency decreases, but since the frequency interval also decreases at low frequencies, the frequency resolution increases. Obviously, the first few coefficients would not carry a whole lot of information, simply due to greatly reduced time resolution. To illustrate this richly bizarre DWT representation let us take a look at a real world signal. Our original signal is a 256-sample long ultrasonic signal, which was sampled at 25 MHz. This signal was originally generated by using a 2.25 MHz transducer, therefore the main spectral component of the signal is at 2.25 MHz. The last 128 samples correspond to [6.25 12.5] MHz range. As seen from the plot, no information is available here, hence these samples can be discarded without any loss of information. The preceding 64 samples represent the signal in the [3.12 6.25] MHz range, which also does not carry any significant information. The little glitches probably correspond to the high frequency noise in the signal. The preceding 32 samples represent the signal in the [1.5 3.1] MHz range. As you can see, the majority of the signal's energy is focused in these 32 samples, as we expected to see. The previous 16 samples correspond to [0.75 1.5] MHz and the peaks that are seen at this level probably represent the lower frequency envelope of the signal. The previous samples probably do not carry any other significant information. It is safe to say that we can get by with the 3rd and 4th level coefficients, that is we can represent this 256 sample long signal with $16+32=48$ samples, a significant data reduction which would make your computer quite happy.

One area that has benefited the most from this particular property of the wavelet transforms is image processing. As you may well know, images, particularly high-resolution images, claim a lot of disk space. As a matter of fact, if a web page is taking a long time to download, that is mostly because of the images. DWT can be used to reduce the image size without losing much of the resolution. Here is how:

For a given image, you can compute the DWT of, say each row, and discard all values in the DWT that are less than a certain threshold. We then save only those DWT coefficients that are above the threshold for each row, and when we need to reconstruct the original image, we simply pad each row with as many zeros as the number of discarded coefficients, and use the inverse DWT to reconstruct each row of the original image. We can also analyze the image at different frequency bands, and reconstruct the original image by using only the coefficients that are of a particular band. I will try to put sample images hopefully soon, to illustrate this point.

7 Application and Results

The purpose of this thesis, as written in the beginning, was to find, present and test three power quality monitoring algorithms on real signals. By real signals it is meant signals that are taken from a real power system, to be precise from the Power Supply Station in Katsampas at Heraklion, Crete, via the 5500 Series DualNode, and not artificially implemented in the lab by a simulation program. As it will be distinct on the graphs taken from the capacitors' 13.8kV busbar of the facility that will be presented later on, there is a great difference between a simulated signal and a real one. The reason is that the simulated signal is pure and any disturbance can be easily be detected by most of the classification methods, even the less sophisticated ones. Moreover, if there is any noise inserted in the signal in order to lose its purity, it can easily be detected and isolated as it is created by an algorithm with a logical, mathematical, way that can be localized by the classification algorithm. However, if the signal has random noise by its own, like notching for example, that has been inserted to it by the Power System itself then it is very difficult, and in some case even more impossible, to "purify" it in order to detect the disturbance and only the disturbance. Thus more effective and robust algorithms need to be mobilized in order to be able to separate, in the monitored signal, the disturbance from the noise that even though it exists in the signal, it fluctuates in such low levels that does not allow it to be characterised as a disturbance.

The three algorithms that were finally selected and implemented are:

- "Adaline", which, as presented earlier is a neural network structure
- "Wavelet-Based PNN, Probabilistic Neural Network", which is also a neural network structure, and finally the
- "Wavelet-Based ANFIS, Adaptive Neuro-Fuzzy Inference System" which is a fuzzy network structure.

The criteria based on which those algorithms were selected, were firstly that they should be relatively recent. It is obvious that the use of algorithms that are out-of-date, is, if nothing else, useless. Another criterion was that the algorithms should be as fast as possible, at least during the laboratory experiments. That was because it is essential to detect and point out the disturbance during its occurrence, so if any actions were to be made, they would be made instantly. Finally, the algorithms should

be flexible enough in order to be able to distinguish the noise in the signal, which is inserted by the Power System itself and is not a disturbance, from a small disturbance when it happens.

Let's see now how those techniques are being utilized in the service of power quality monitoring.

7.1.1 Power Quality Event Detection Using Adaline

The purpose of this chapter is to introduce a simple, yet rigorous Adaline technique as a tool for power quality event detection. The simplicity introduced by this technique is due to the ease in calculations that facilitates its hardware implementation. The simplicity of the Adaline makes it a very competitive choice for the algorithms currently used in power quality instrumentation.

7.1.1.1 Adaline architecture

Artificial neural networks have been extensively used in power system applications including classification of different power quality phenomena (76) and monitoring of power system operation (77). The Adaline can be thought of as the smallest, linear building block of the artificial neural networks. Adaline had been applied successfully in many power applications, including harmonic estimation (78), frequency deviation estimation (79) and feature extraction (80, 81). In this section, Adaline is introduced as an efficient tool for disturbance detection. An Adaline is a p-input, single-output, signal-processing element, which can be thought of as a simple model of a non-branching biological neuron. Graphically, an artificial neuron is represented by the construction shown in fig. (69), where y is the Adaline output, W is the weight matrix and X is the Adaline input matrix.

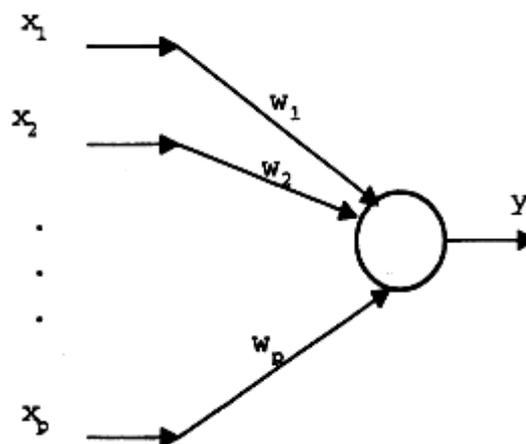


Figure 69. Adaline Construction

7.1.1.2 Adaline training

Training is the process of tuning the weights of the Adaline so that its output matches a desired outcome. The main advantage of the Adaline architecture is the ability of the Adaline to be trained on-line, eliminating the need for repetitive off-line training. Although most neural networks have the ability of on-line training, the Adaline is superior because of its simple structure and speed. The training process is carried out by minimizing a performance criterion $J(w)$, which represents the sum of the squared error over the whole training patterns N . Mathematically $J(w)$ is expressed as:

$$J(w) = \frac{1}{N} \sum_N E(k)^2 = \frac{1}{2N} e \cdot e^T \quad \text{Equation 45}$$

where the output error is given as:

$$e(k) = \hat{y}(k) - y(k) \quad \text{Equation 46}$$

After a brief mathematical operation, eq. (45) could be simplified in the following form:

$$J(w) = \frac{1}{2} \left(\frac{\|D\|^2}{N} - 2qw^T + wRw^T \right) \quad \text{Equation 47}$$

where $q = (D \cdot X^T)/N$, the $1 \times p$ cross-correlation vector and $R = (X \cdot X^T)/N$, the $p \times p$ input correlation matrix. In order to find the optimal weight vector which minimizes the mean-squared error, $J(w)$, the gradient of J with respect to w should be calculated as:

$$\nabla J(w) = \frac{\partial}{\partial w} J = \left[\frac{\partial J}{\partial w_1} \dots \frac{\partial J}{\partial w_p} \right] = -q + wR \quad \text{Equation 48}$$

Setting the gradient of mean square error equal to zero, then:

$$0 = -q + w_{opt}R \quad \text{Equation 49}$$

where w_{opt} is the optimum weight matrix. Calculation of the optimal weights, which minimize the performance index, $J(w)$, is mathematically demanding because of the calculation of the correlation matrices, q and R in addition to the inversion of the input correlation matrix. In order to avoid matrix inversion, the optimal weight vector for which performance index, $J(w)$, attains a minimum value should be found. This task is achieved through an iterative modification of the weight vector for each training example in the direction opposite to the gradient of the performance index, $J(w)$. This procedure is known as the steepest decent method (82,83). Once the weight

vector attains the optimal value for which the gradient is zero, the iteration process is terminated. More precisely, the iterations are specified as:

$$w(k+1) = w(k) + \Delta w(k) \quad \text{Equation 50}$$

where the weight adjustment, $\Delta w(n)$ is proportional to the gradient of the mean-squared error:

$$w(k) = -n\nabla J(w(k)) \quad \text{Equation 51}$$

where n is the learning rate of the Adaline.

In order to further simplify the calculations, the least mean-square learning law (LMS) replaces the gradient of the mean-squared error in eq. (51) with the gradient update is driven in the following form:

$$w(k+1) = w(k) + ne(k)x^T(k) \quad \text{Equation 52}$$

7.1.1.3 Power Quality Event Detection Using Adaline

The idea behind using the Adaline in detection of power quality disturbances is to represent the Adaline as an adaptive signal predictor. The input to this predictor is time-delayed samples of the signal and the output of the Adaline is the predicted value of the signal. The Adaline algorithm possesses a highly tracking capability. Yet when a power quality disturbance occurs, the abrupt change in the signal gives rise to the error signal generated by the Adaline and the weight values experience variation until it settles down to the new values. Both the alterations in the error signal and the sum of the variation of the weight values can aid in the detection of power quality events. This algorithm is described in fig. 70.

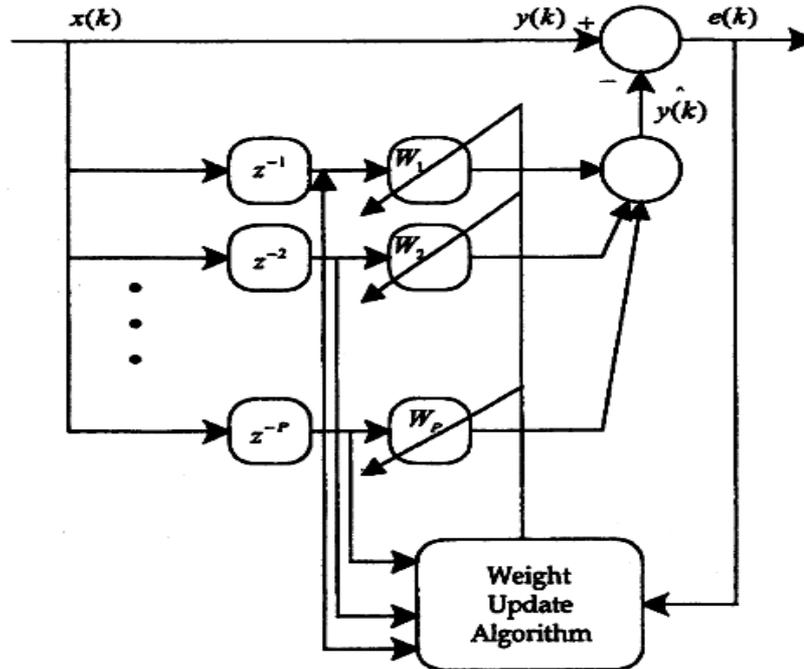


Figure 70. Adaline Based Detection Scheme

The following section demonstrates the capability of the Adaline in detecting power quality disturbances. In order to accomplish this task, different power quality events has been simulated using Matlab 6.0 software. Initially, the number of inputs to the Adaline P was chosen to be four. These inputs represent the most recent samples of the voltage. The Adaline learning rate η was chosen to be equal to 0.4. It was found that the performance of the Adaline is highly dependent on the chosen number of inputs and learning rate. Therefore, the rule of choosing an appropriate value for those two variables is investigated in a later section. The sampling frequency was chosen to be equal to $f = 60$ sample per cycle during all numerical examples.

7.1.1.4 Numerical examples

In order to obtain representative signals for the most common power quality events to serve the purpose of testing the Adaline detector, power quality event simulations were created mathematically. Five categories of events have been simulated, namely, sudden sag, sudden swell, harmonics, oscillatory transient and interruptions. These waveforms have been generated at a sampling rate of 7680 Hz. In order to create different waveforms for each disturbance category, some unique parameters for each disturbance type have been introduced. Parameters, such as magnitude, duration, frequency and damping, were implemented.

7.1.1.4.1 Voltage sag

According to IEEE 1159-1995 standards, voltage sag is a sudden reduction (between 10 and 90%) of the voltage magnitude at a point in the electric system and lasting from 0.5 cycles to a few seconds. Either switching operations or any type of fault as well as fault clearing process can cause a voltage dip.

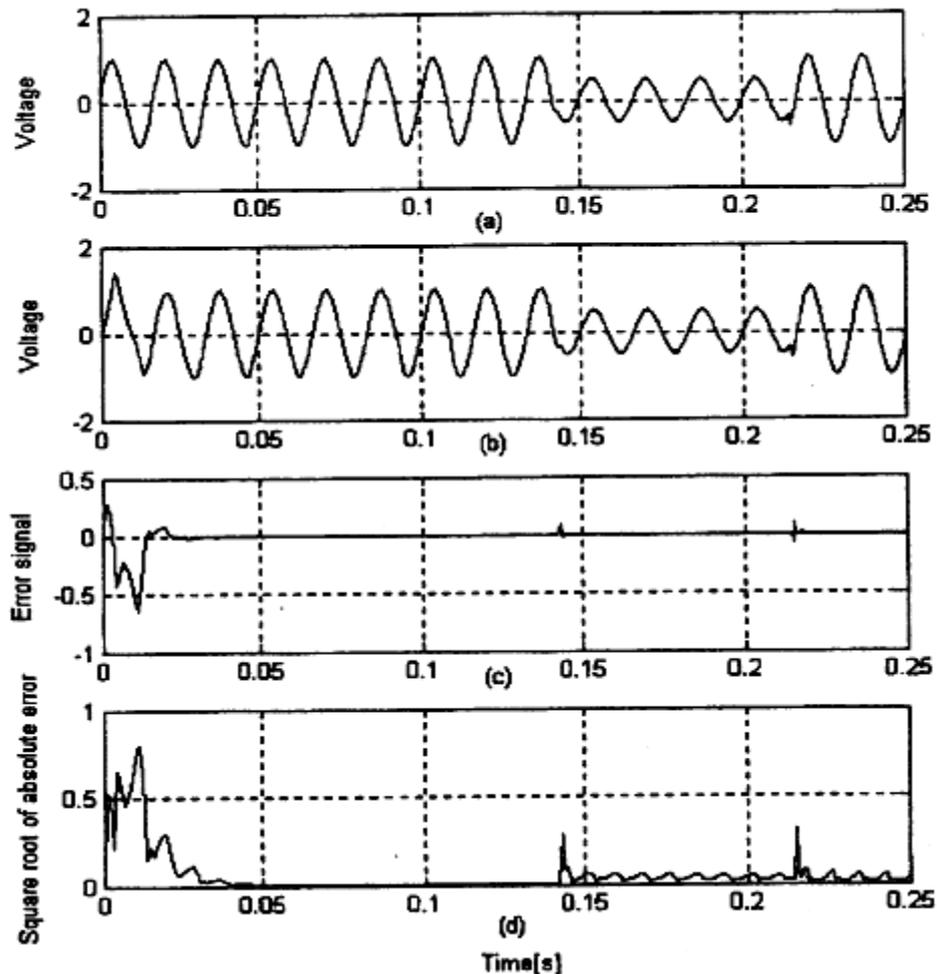


Figure 71. Voltage Sag Event (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

Fig. (71a, b) shows the input voltage and the predicted signal waveforms for typical sag. The corresponding error signal (difference between the actual and predicted signals) and the square root of it are shown in fig. (71.c, d), respectively. The square root of the error signal is added just to magnify the effect and make it visual. As shown in fig. (71.d), there are two large spikes associated with the sag start and end. In actual power system, this separation should be clustered most of the time around the time of the operation of protective relays.

7.1.1.4.2 Voltage swell

According to IEEE 1159-1995 standards, voltage swells are defined as an increase of the supply voltage between 1.1 and 1.8 from its normal rated rms value. The duration of any voltage swell is typically between 0.5 cycles up to 1 min. Voltage swell may appear due to switching off large loads or large capacitor energizing. Moreover, during single line to ground faults, the unfaulted phases may experience a voltage swell. Voltage swell may cause damage to electronic equipment and may cause redundant protective relays operation.

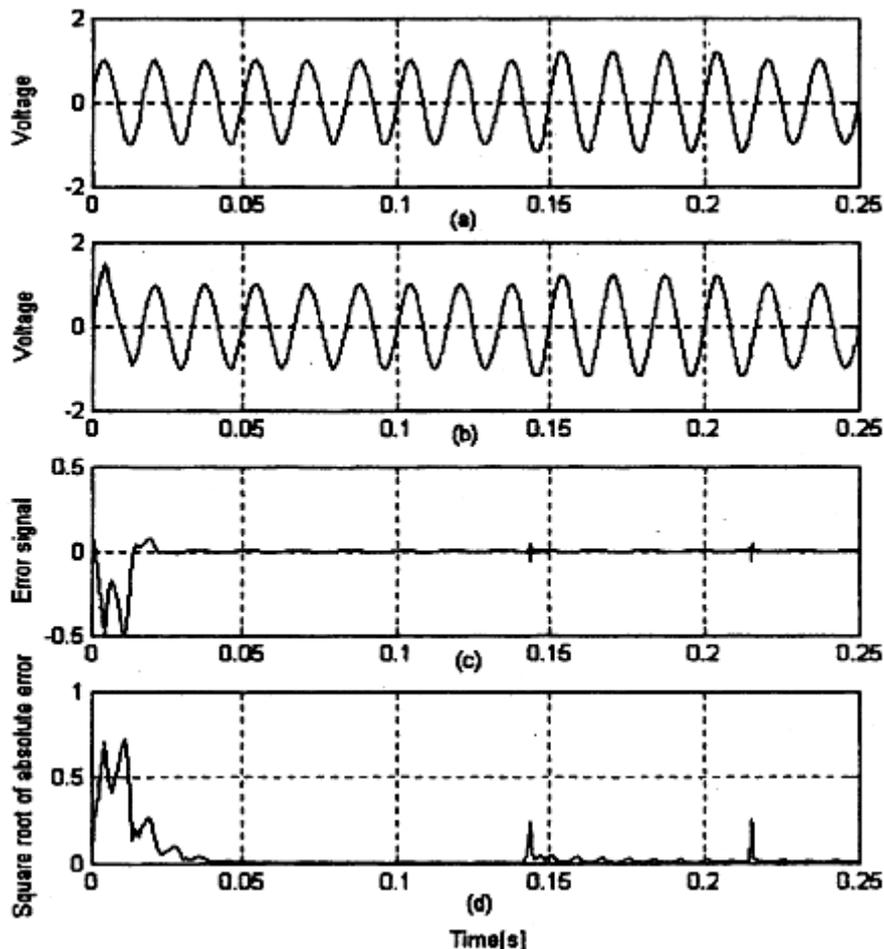


Figure 72. Voltage Swell Event (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

Fig. (72.a, b) shows the input voltage and the predicted signal waveforms for a typical swell. The corresponding error signal and the square root of the error signal are shown in fig. (72.c, d), respectively. Fig. (72) offers a precise determination of the start and the end of the voltage swell using the square root of the absolute value of the prediction error generated from Adaline.

7.1.1.4.3 Momentary interruptions

According to IEEE 1159-1195 standards, a momentary interruption belongs to short duration voltage variation. During momentary interruptions, the supply voltage reaches a value $<10\%$ of its normal rated rms value. Momentary interruptions may occur due to power system faults or equipment breakdown. As in the case of swell and sag, interruption is associated with two spikes denoting the start and end instant, as shown in fig. (73).

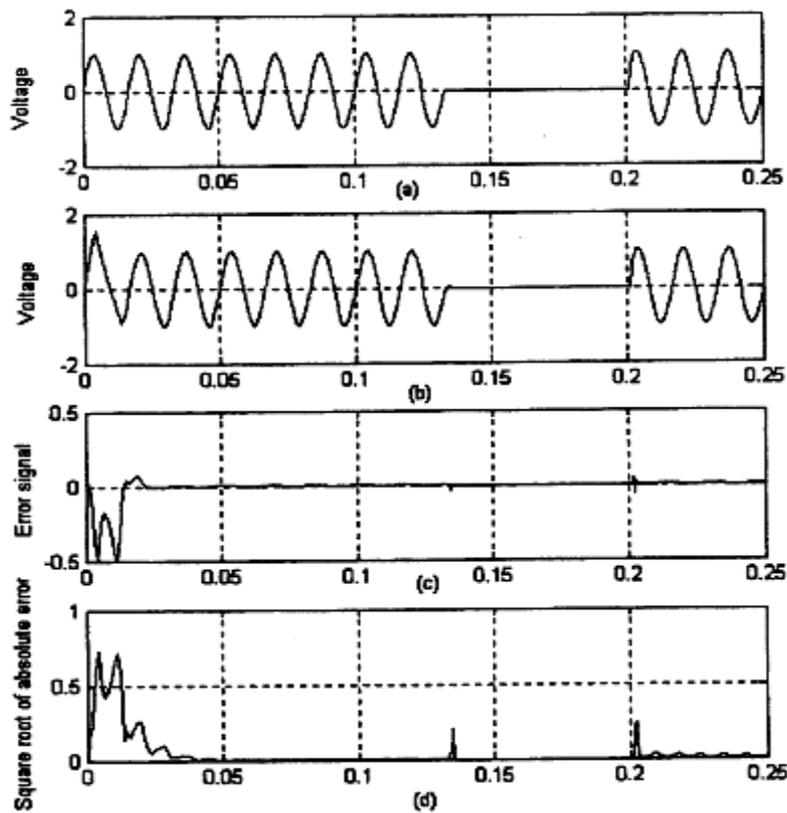


Figure 73. Voltage Momentary Interruption Event (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.1.4.4 Harmonics

Harmonics are sinusoidal voltages or currents having frequencies that are multiple of the fundamental system frequency. Distorted waveforms could be decomposed into a sum of fundamental frequency and harmonics. Harmonic distortion originates from the nonlinear characteristics of devices and loads on the power system. Harmonic distortion levels are described by the complete harmonic spectrum with magnitudes and phase angle of each harmonic component. Fig. (74) shows the input and the predicted voltage waveforms for a typical signal contaminated with harmonic distortion. In addition, fig. (74) offers the corresponding error signal and the

square root of the error signal. The presence of the harmonics in the waveform resulted in sustained oscillations, as shown in fig. (74.d).

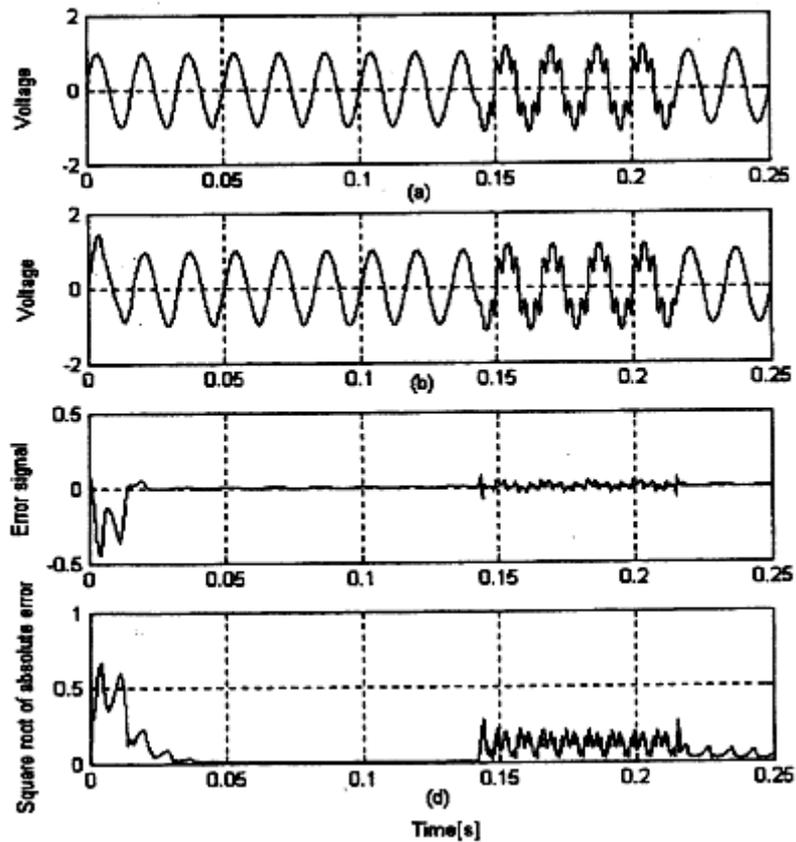


Figure 74. Voltage Harmonic Event (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.1.4.5 Transient

Transients are one of the most important and severe power quality events that lead to malfunction of electronic equipment. Transients can be originated from surges, huge motors starting or utility capacitor switching. These transients can be detected by the presence of large spikes on the square root of the error signal produced by the Adaline, as shown in fig. (75).

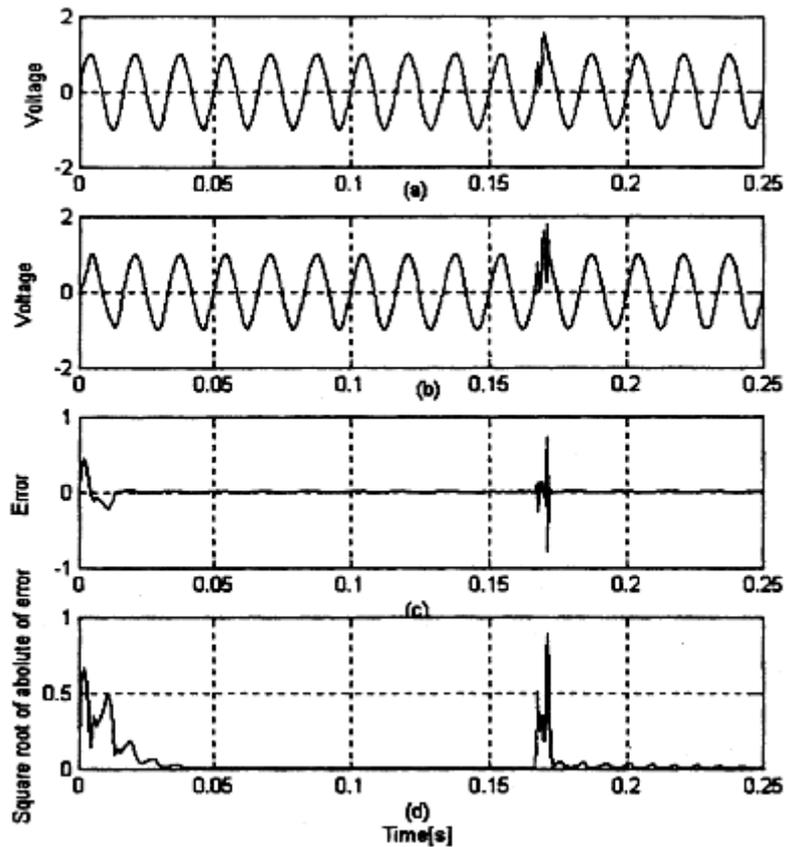


Figure 75. Voltage Transient Event (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.1.4.6 Combined Events

In this clause, a combined scenario is assumed to validate the continuous operation of the Adaline. Figure 76 shows a voltage waveform, which is contaminated with harmonics and is subjected to transient and interruption events. The corresponding prediction error signal is given in figure 76(c, d). The variation of the square root of the error signal represents a good indication to detect different power quality phenomena, as shown in fig. (76).

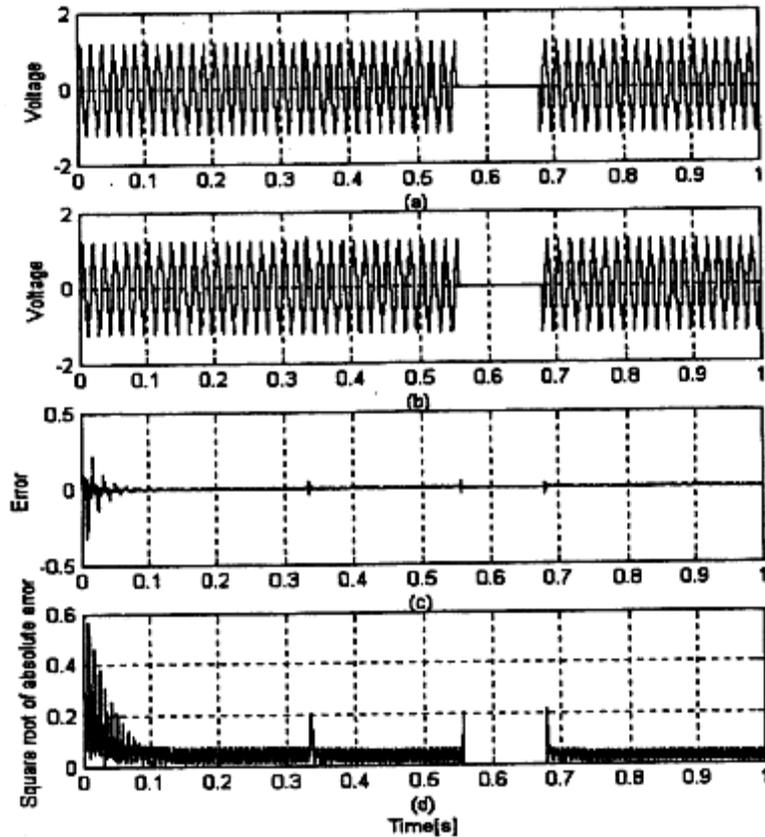


Figure 76. Combined Disturbance Events (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.1.5 Sensitivity analysis

There are two major variables which controls the Adaline performance in detecting power quality events. These factors are the learning rate h and the number of inputs P introduced to the Adaline. The following section is dictated to explore the effect of changing both h and P on the Adaline accuracy in detecting the power quality disturbances.

7.1.1.5.1 Effect of learning rate (h)

Changing of the learning rate has a direct effect on the detection process. Small learning rates will lead to small error signal and a small variation of the sum of the weight when an abrupt change occurs to the system. This might lead to an error signal appearing in normal operation of power distribution system. On the other hand, large learning rates produce a large error signal and large variation of sum of the weight. Yet, caution should be taken because this may lead the Adaline to lose its ability to track the signal if the learning rate is chosen too high. It was found that a

learning rate h falling between 0.3 and 0.6 yields the most accurate results for detection of power quality events. In order to study the effect of the learning rate on the Adaline performance, a voltage signal contaminated with harmonic distortion was utilized. The learning rate is varied while keeping the number of inputs to the Adaline constant.

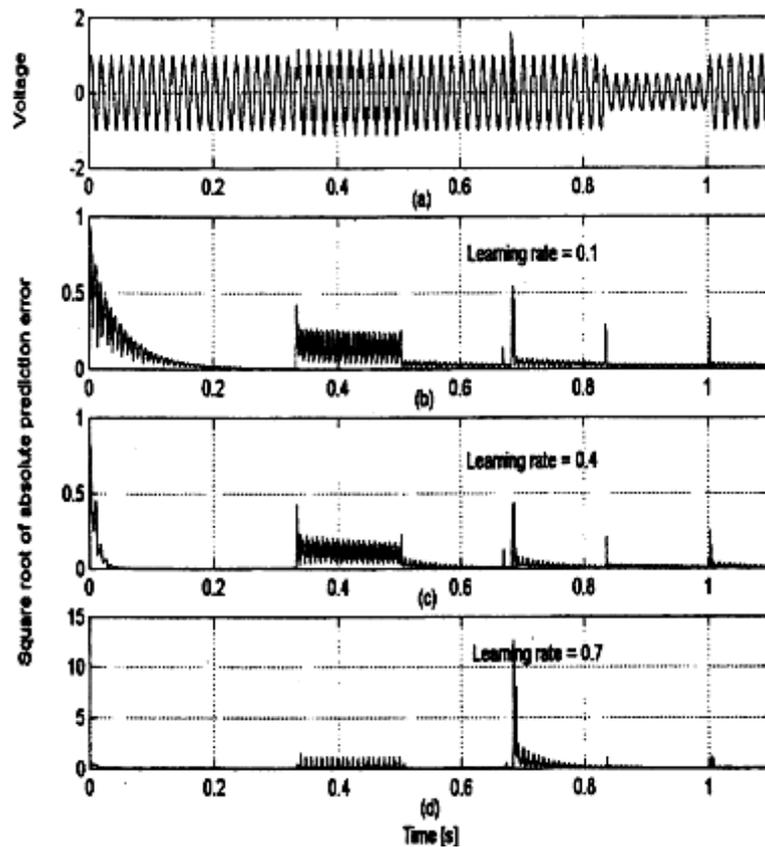


Figure 77. Effect Of Learning Rate On Adaline Operation

Fig. (77) shows the effect of changing the learning rate on the error signal. As can be observed from the graphs, learning rate around 0.4 is the most suitable to achieve the desired detection task with the highest accuracy. It should be noted that at $n=0.7$, a large spike occurs in the error signal indicating an unstable system operation. Consequently, the Adaline will fail to track the dynamics of the system.

7.1.1.5.2 Effect of choosing P

The effect of changing the number of inputs to the Adaline P on the Adaline performance is carried out while keeping the learning rate constant at 0.4. The value of P is allowed to vary between 2 and 10. It was found that a small value of P , such as 2, would result in a negligible and unnoticeable error signal due to the lack of interaction with the change in the power signal. This will lead to poor identification process.

On the other hand, utilizing large number of inputs might lead to model over fitting resulting in poor results. A reasonable range of inputs was found to range between 3 and 6. Fig. (78) presents the results for Adaline inputs of 2, 4, 6, 8 and 10, respectively.

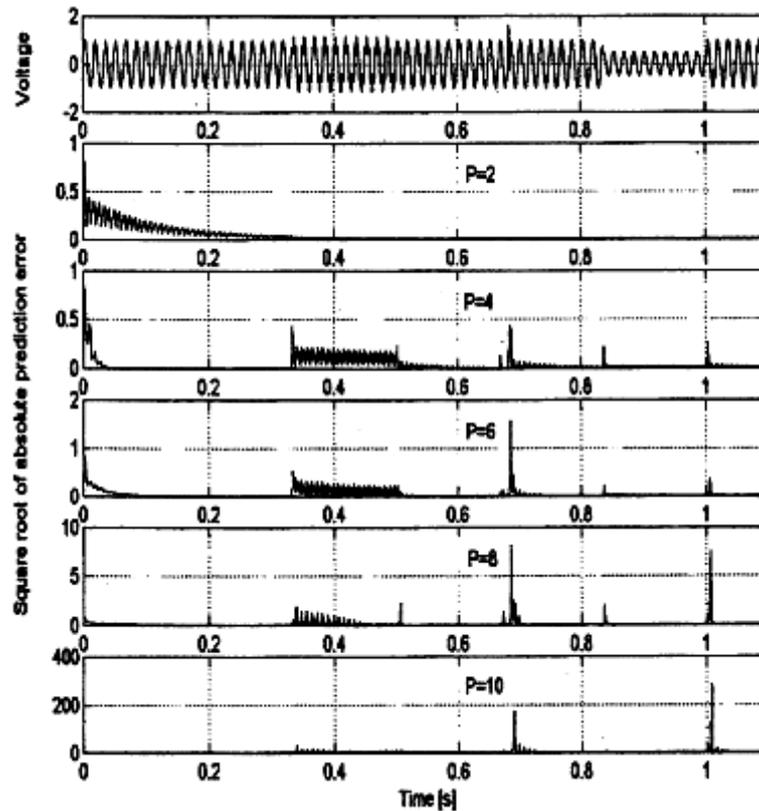


Figure 78. Effect Of Choosing Number Of Inputs On Adaline Operation

7.1.1.6 Testing

The algorithm was rigorously tested utilizing the Series 5500 DualNode on the Public Power Constitution at Katsampas at Heraklion, Crete. The DataNode was connected at the capacitors' 13.8kV busbar of the facility, via the current and voltage pods, the set up of the InfoNode was made according to the IEEE standards, described earlier regarding all the disturbance events, the data were obtained by the use of the program Dranview and were finally processed on Matlab 6.5v. Unfortunately due to technical difficulties, regarding the time of its use, the only disturbances that were recorded and afterwards processed involved swells, sags and transients. That is the reason why the experimental results, presented later on relate only to these three categories of power disturbance events.

7.1.1.6.1 Sag detection in harmonic contaminated signal

The first study carried out on this system was to detect the occurrence of voltage sag. The sag is created on the distribution system by initiating several short circuits with different durations and types on the busbar. The corresponding voltage signals, which are contaminated with harmonics, are measured and recorded at the bus. The Adaline algorithm was then utilized to detect the start and end of the sag. A sample result is shown in fig. (79). It is clear from fig. (79) that the Adaline succeeded in detecting the start and end of the sag. Moreover, the presence of the harmonics is also noticeable by the sustained oscillations in the error signal.

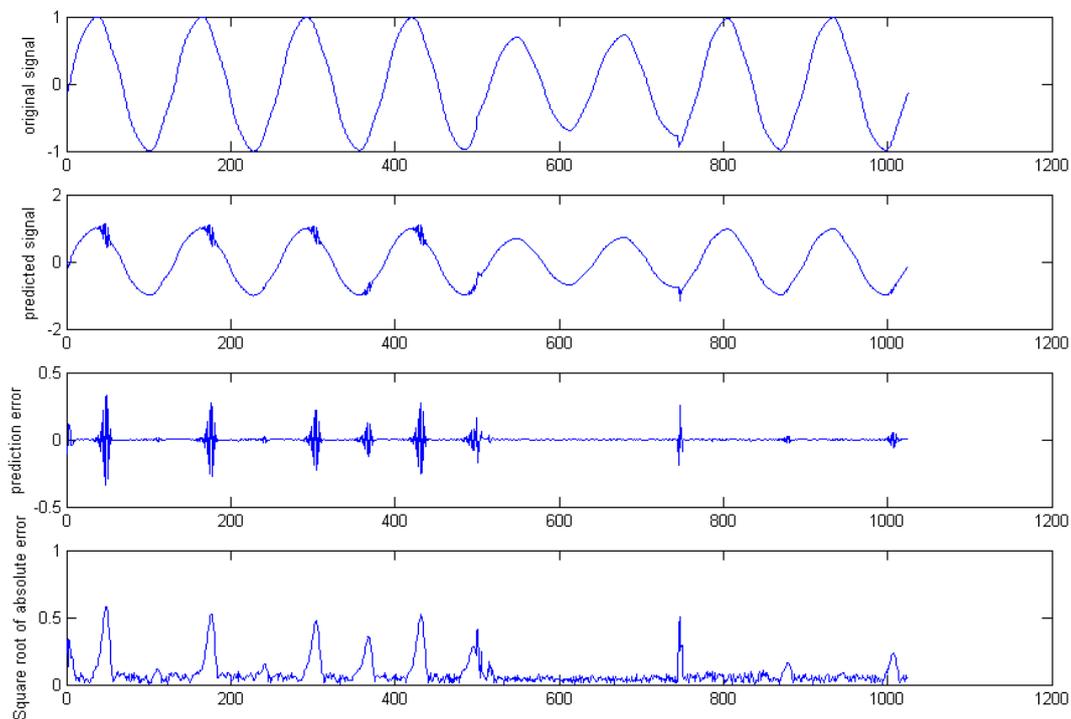


Figure 79. Voltage Sag Event in a signal contaminated with Harmonics (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Of Error

7.1.1.6.2 Swell detection in harmonic contaminated signal

The second study carried out on this system was to detect the occurrence of voltage swell. The swell is created on the distribution system as the result of load switching (e.g., switching off a large load), or variations in the reactive compensation on the system (e.g., switching on a capacitor bank). Poor system voltage regulation capabilities or controls result in swells. Incorrect tap settings on transformers can also result in system swells. The Adaline algorithm was then, again, utilized to detect the start and end of the swell. A sample result is shown in fig. (80). It is clear from fig.

(80) that the Adaline succeeded in detecting the start and end of the swell. Moreover, the presence of the harmonics is also noticeable by the sustained oscillations in the error signal.

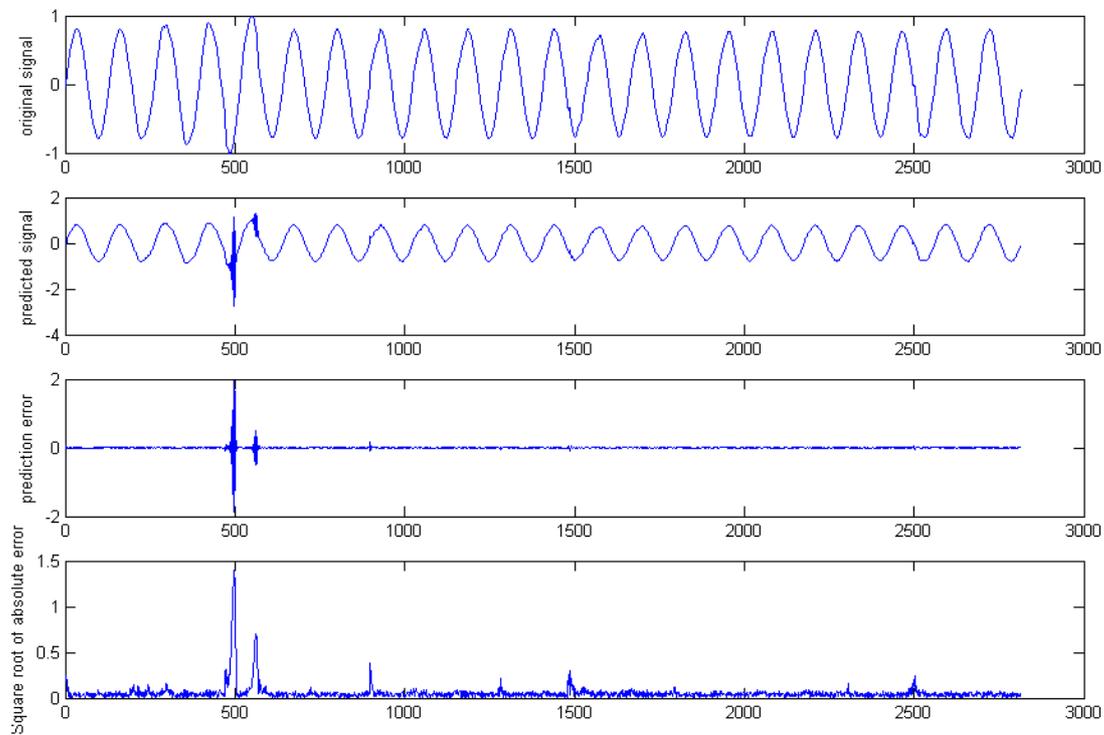


Figure 80. Voltage Swell Event in a signal contaminated with Harmonics (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.1.6.3 Transient detection in harmonic contaminated signal

The third and last study on the industrial distribution system was carried out to investigate the presence of transients in a harmonic contaminated voltage signal. The transients were created by switching several capacitors at different locations and recording the voltage at the busbar. The Adaline was used to detect the presence of these oscillatory transient due to switching the capacitors on and off. A test from the results is shown in fig. (81). The Adaline, once again, succeed in detecting the transient and it also identified the presence the harmonic contamination of the signal resulting from the non-linear loads of the system.

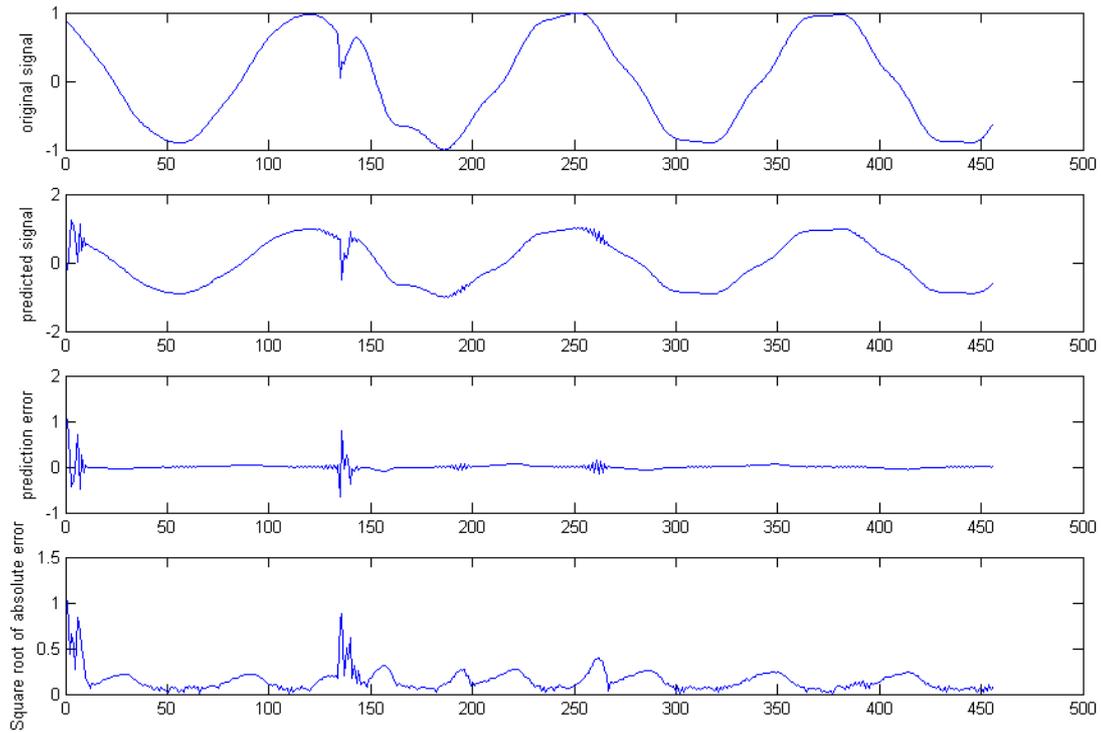


Figure 81. Voltage Transient Event in a signal contaminated with Harmonics (a) Input Voltage Signal, (b) Predicted Signal, (c) Error Signal, (d) Square Root Of Absolute Error

7.1.2 Power Quality Event Recognition and Classification Using a Wavelet-Based Neural Network

In this section, a wavelet-based neural network classifier for recognizing power-quality disturbances is implemented and tested under various power quality disturbance events. This novel classifier is constructed with the combined use of the discrete wavelet transform (DWT) technique and the probabilistic neural-network (PNN) model. First, the multiresolution-analysis technique of DWT and the Parseval's theorem are employed to extract the energy distribution features of the distorted signal at different resolution levels. Then, the PNN is employed to classify disturbance types according to the detailed energy distribution. Since the proposed methodology can reduce a great quantity of the distorted signal features without losing its original property, by means of the Parseval's theorem, less memory space is required. Moreover, because of the very fast learning efficiency of the PNN, less computing time is required, thus making the proposed algorithm suitable for real-time operation for fault diagnosis and signal classification problems.

7.1.2.1 Wavelet Transform

The wavelet analysis block transforms the distorted signal into different time-frequency scales. The wavelet transform (WT) uses the wavelet function φ and scaling function ϕ to perform simultaneously the multiresolution analysis (MRA) decomposition and reconstruction of the measured signal. The wavelet function φ will generate the detailed version (high-frequency components) of the decomposed signal and the scaling function ϕ will generate the approximated version (low-frequency components) of the decomposed signal. The wavelet transform is a well-suited tool for analyzing high-frequency transients in the presence of low-frequency components such as nonstationary and nonperiodic wideband signals [85].

7.1.2.1.1 Multiresolution Analysis (MRA) and Decomposition

The first main characteristic in WT is the MRA technique that can decompose the original signal into several other signals with different levels (scales) of resolution. From these decomposed signals, the original time-domain signal can be recovered without losing any information.

The recursive mathematical representation of the MRA is as follows:

$$V_j = W_{j+1} \oplus V_{j+1} = W_{j+1} \oplus W_{j+2} \oplus \dots \oplus W_{j+n} \oplus V_n \quad \text{Equation 53}$$

where

V_{j+1} is the approximated version of the given signal at scale $j+1$

W_{j+1} is the detailed version that displays all transient phenomena of the given signal at scale $j+1$

\oplus denotes a summation of two decomposed signals

N is the decomposition level.

7.1.2.1.2 Mathematical Model of DWT

Before the WT is performed, the wavelet function $\varphi(t)$ and scaling function $\phi(t)$ must be defined. The wavelet function serving as a highpass filter can generate the detailed version of the distorted signal, while the scaling function can generate the approximated version of the distorted signal. In general, the discrete $\varphi(t)$ and $\phi(t)$ can be defined as follows:

$$\phi_{j,n}[t] = 2^{\frac{j}{2}} \sum_n c_{j,n} \phi[2^j t - n] \quad \text{Equation 54}$$

$$\varphi_{j,n}[t] = 2^{\frac{j}{2}} \sum_n d_{j,n} \varphi[2^j t - n] \quad \text{Equation 55}$$

Where c_j is the scaling coefficient at scale j , and d_j is the wavelet coefficient at scale j .

Simultaneously, the two functions must be orthonormal and satisfy the properties as follows:

$$\begin{cases} \langle \phi \cdot \phi \rangle = \frac{1}{2^j} \\ \langle \varphi \cdot \varphi \rangle = \frac{1}{2^j} \\ \langle \phi \cdot \varphi \rangle = 0 \end{cases} \quad \text{Equation 56}$$

Assuming the original signal $x_j[t]$ at scale j is sampled at constant time intervals, thus $x_j[t] = (u_0, u_1, \dots, u_{N-1})$, the sampling number is $N=2^j$. J is an integer number. For $x_j[t]$, its DWT mathematical recursive equation (as) is presented as follows:

$$V_J = W_{J+1} \oplus V_{J+1} \quad \text{Equation 57}$$

DWT($x_J[t]$)=

$$\begin{aligned} \text{DWT}(x_J[t]) &= \sum_k x_J[t] \phi_{j,k}[t] = \\ &= 2^{\frac{(j+1)}{2}} \left(\sum_n u_{j+1,n} \phi[2^{j+1} t - n] + \sum_n w_{j+1,n} \varphi[2^{j+1} t - n] \right) \quad \text{Equation 58} \\ 0 &\leq n \leq \frac{N}{2^j} - 1 \end{aligned}$$

where

$$u_{j+1,n} = \sum_k c_{j,k} u_{j,k+2n}, \quad 0 \leq k \leq \frac{N}{2^j} - 1 \quad \text{Equation 59}$$

$$w_{j+1,n} = \sum_k d_{j,k} u_{j,k+2n}, \quad 0 \leq k \leq \frac{N}{2^j} - 1 \quad \text{Equation 60}$$

$$d_k = (-1)^k c_{2^j p - 1 - k}, \quad p = \frac{N}{2^j} \quad \text{Equation 61}$$

where

$u_{j+1,n}$ is the approximated version at scale $j+1$,

$w_{j+1,n}$ is the detailed version at scale $j+1$ and

j is the translation coefficient.

According to the orthonormal wavelet functions and eq.(58), the signal $x_j(t)$ can be reconstructed from both u_{j+1} and w_{j+1} coefficients using the inverse discrete wavelet transform IDWT, as $V_{j+1} \oplus W_{j+1} = V_j$.

Fig. (82) illustrates the three decomposed/reconstructed levels of the DWT algorithm. At each decomposition level, the length of the decomposed signals (e.g., u_1 and w_1) is half that of the signals (x_0) in the previous stage.

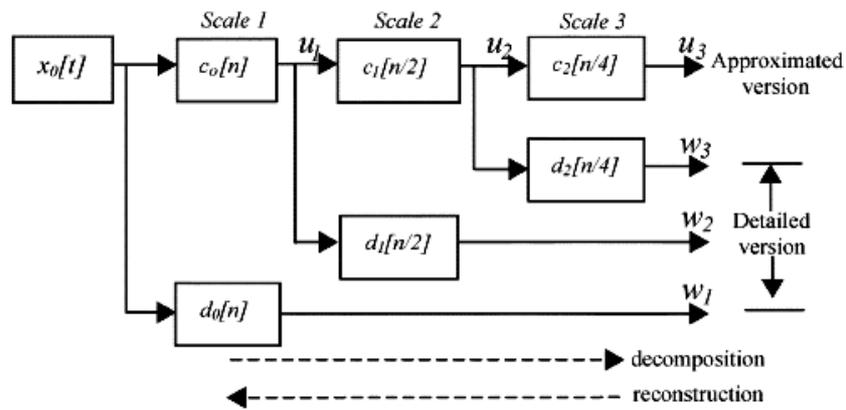


Figure 82. Three decomposed/reconstructed levels of DWT

7.1.2.2 Parseval's Theorem In DWT Application

In Parseval's theorem, assuming a discrete signal $x[n]$ is the current that flows through the $1-\Omega$ resistance, then the consumptive energy of the resistance is equal to the square sum of the spectrum coefficients of the Fourier transform in the frequency domain

$$\frac{1}{N} \sum_{n=\langle N \rangle} |x[n]|^2 = \sum_{k=\langle N \rangle} |a_k|^2 \quad \text{Equation 62}$$

Where N is the sampling period, and a_k is the spectrum coefficients of the Fourier transform.

To apply the theorem to the DWT, we use eq.(58) and eq.(61) to obtain eq.(62) which is the Parseval's theorem in the DWT application

$$\frac{1}{N} \sum_t |x[t]|^2 = \frac{1}{N_j} \sum_k |u_{j,k}|^2 + \sum_{j=1}^J \left(\frac{1}{N_j} \sum_k |w_{j,k}|^2 \right) \quad \text{Equation 63}$$

Hence, through the DWT decomposition, the energy of the distorted signal is shown by eq.(63). The first term on the right of eq.(63) denotes the average power of the approximated version of the decomposed signal, while the second term denotes that of the detailed version of the decomposed signal. The second term giving the energy distribution features of the detailed version of distorted signal will be employed to extract the features of power disturbance.

7.1.2.3 Feature Extraction, Recognition And Classification

7.1.2.3.1 Feature Extraction

Detailed Energy Distribution: As seen in eq.(63), the energy of the distorted signal can be partitioned at different resolution levels in different ways depending on the power-quality problem. Therefore, we will examine the coefficient w of the detailed version at each resolution level to extract the features of the distorted signal for classifying different power-quality problems. The process can be represented mathematically as followed

$$P_j = \frac{1}{N_j} \sum_k |w_{j,k}|^2 = \frac{\|w_j\|^2}{N_j} \quad \text{Equation 64}$$

where $\|w_j\|$ is the norm of the expansion coefficient w_j .

Four special properties in eq.(64) need further explanation.

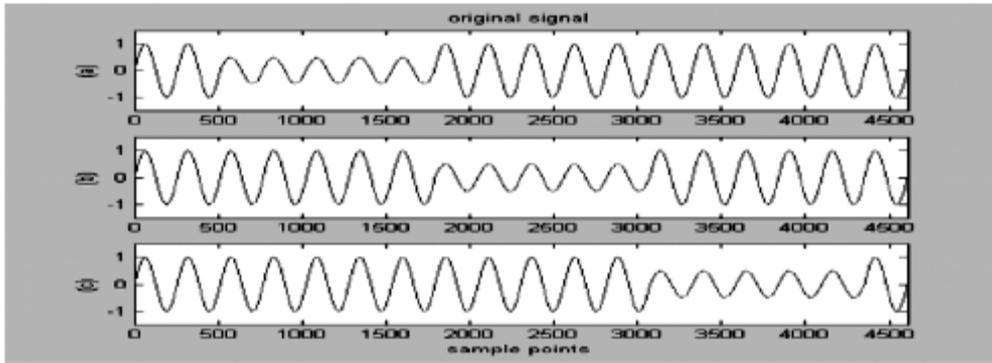
- The Daubanchie “db4” wavelet function was adopted to perform the DWT, thus resulting in the larger energy distributions of the decomposition levels 6, 7, and 8. However, using different wavelet functions will generate different results.
- The energy distribution remains unaffected by the time of disturbance occurrence.

- The outline of energy distribution remains the same despite variations in the vibration amplitude of the same disturbance type.
- The low-level energy distribution will show obvious variations when the distorted signal contains high-frequency elements. On the contrary, the high-level energy distribution will show obvious variations when the distorted signal contains low-frequency elements.

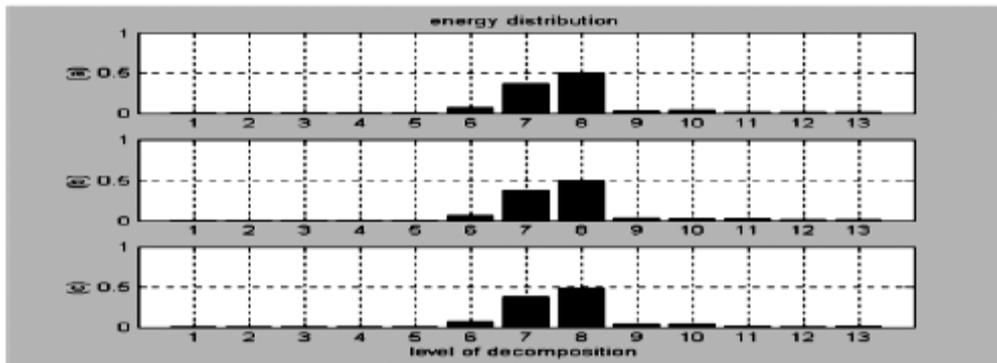
To display clearly the characteristics of the above properties, we normalize eq.(64) by eq.(65)

$$P_j^D = (P_j)^{\frac{1}{2}} \quad \text{Equation 65}$$

For example, Fig. 83(a) shows a voltage sag disturbance in a simulation power system (frequency = 50 Hz, amplitude = 1 p.u.) with three different times of occurrence. Employing eq.(64) and eq.(65) to analyze the three distorted signals (sampling rate is 256 points/per cycle) shows that when the disturbance intervals are the same (five cycles), the detailed energy distributions ($P_1^D \sim P_{13}^D$) of the given signals are also the same as shown in Fig. 83(b). Similarly, when the amplitude of vibration (0.8 p.u., 0.5 p.u., and 0.2 p.u.) of the sag disturbances are changed as seen in Fig. 84(a), the detailed energy distributions ($P_1^D \sim P_{13}^D$) are the same as shown in Fig. 84(b). The energy distribution of the distorted signal with low-frequency elements is shown in Fig. 85, while that with high-frequency elements is shown in Fig. 86. These experimental results depict clearly the properties of energy distribution of Parseval's theorem in DWT applications.

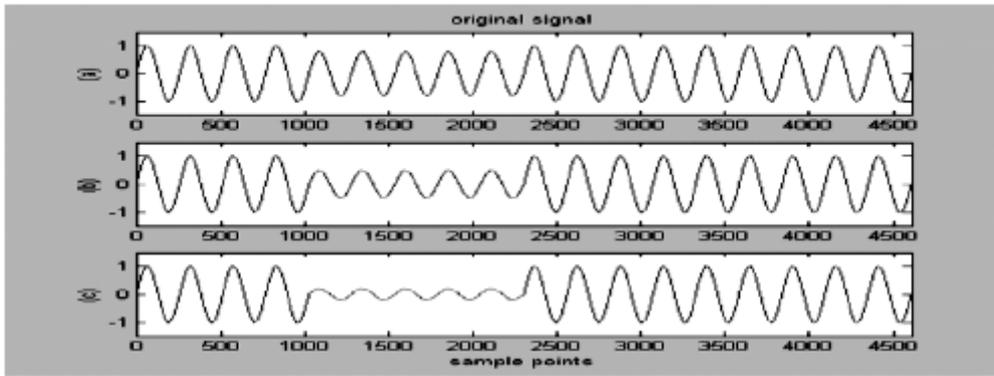


a.

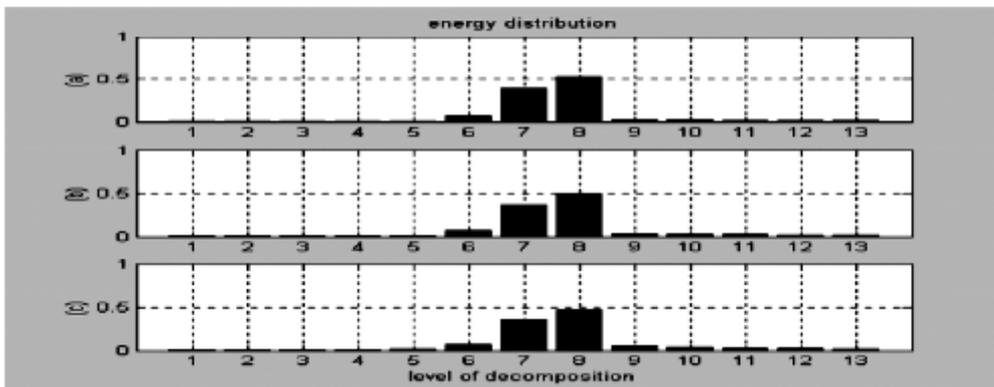


b.

Figure 83. Voltage sag with different times of disturbance occurrence. (a) Different times of disturbance occurrence. (b) Energy distribution diagram.



a.



b.

Figure 84. Voltage sag with different times amplitude magnitudes. (a) Different amplitude magnitude. (b) Energy distribution diagram.

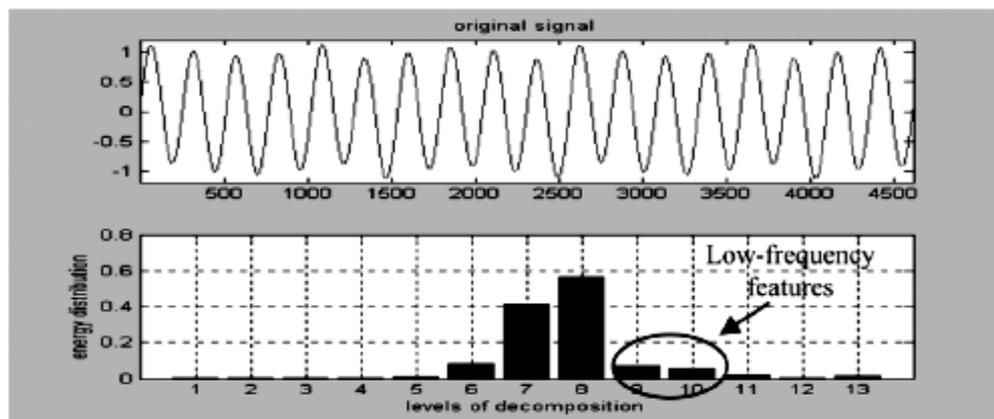


Figure 85. Voltage fluctuation with low-frequency elements.

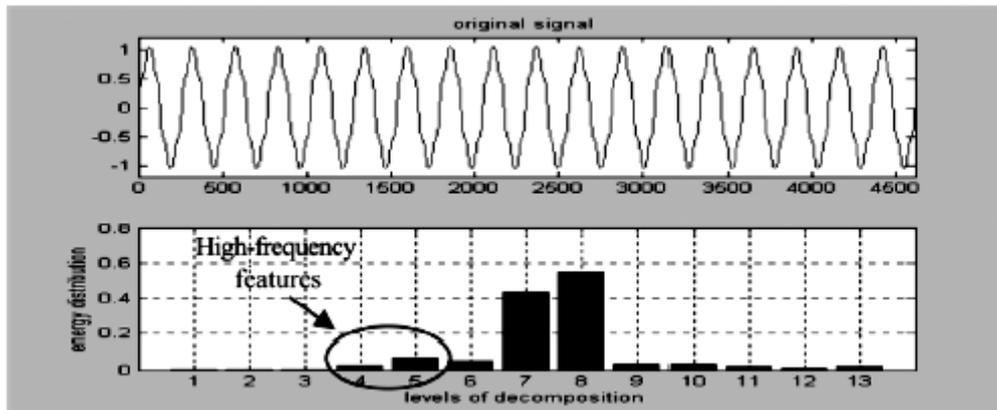


Figure 86. Distorted voltage with high-frequency elements.

7.1.2.3.2 Probabilistic Neural Network (PNN)

The probabilistic neural-network (PNN) model is one of the supervised learning networks, and has the following features distinct from those of other networks in the learning processes [86].

- It is implemented using the probabilistic model, such as Bayesian classifiers.
- A PNN is guaranteed to converge to a Bayesian classifier provided that it is given enough training data.
- No learning processes are required.
- No need to set the initial weights of the network.
- No relationship between learning processes and recalling processes.
- The differences between the inference vector and the target vector are not used to modify the weights of the network.

The learning speed of the PNN model is very fast, making it suitable for fault diagnosis and signal classification problems in real time. Fig. 87 shows the architecture of a PNN model that is composed of the radial basis layer and the competitive layer.

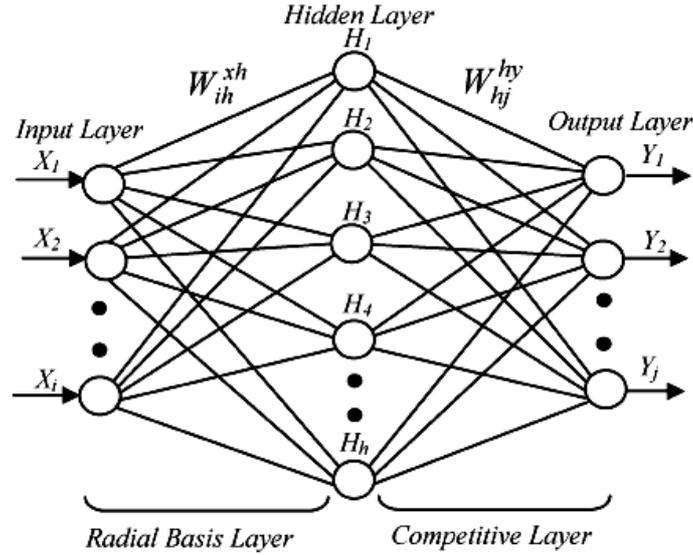


Figure 87. Architecture of a PNN

In the signal classification application, the training examples are classified according to their distribution values of probabilistic density function (PDF), which is the basic principle of the PNN. A simple PDF is defined as follows:

$$f_k = \frac{1}{N_k} \sum_{j=1}^{N_k} \exp\left(-\frac{\|X - X_{kj}\|}{2\sigma^2}\right) \quad \text{Equation 66}$$

Modifying and applying eq. (66) to the output vector H of the hidden layer in the PNN is as below

$$H_h = \exp\left(\frac{-\sum_i (X_i - W_{ih}^{xh})^2}{2\sigma^2}\right) \quad \text{Equation 67}$$

The algorithm of the inference output vector in the PNN is as follows

$$\text{net}_j = \frac{1}{N_j} \sum_h W_{hj}^{hy} \cdot H_h \quad \text{and} \quad N_j = \sum_h W_{hj}^{hy}, \quad \text{Equation 68}$$

if $\text{net}_j = \max_k(\text{net}_k)$ then $Y_j=1$, else $Y_j=0$

where

- i number of input layers
- h number of hidden layers
- j number of output layers
- k number of training examples
- N_k number of classifications (clusters)

σ	smoothing parameter (standard deviation), $0,1 < \sigma < 1$. In general, σ is set to be 0.5
X	input vector
$\ X - X_{kj}\ $	Euclidean distance between the vectors X and X_{jk} , i.e. $\ X - X_{kj}\ = \sqrt{\sum_i (X_i - X_{jk})^2}$
W_{ih}^{xh}	connection weight between the input layer and the hidden layer
W_{hj}^{hy}	is the connection weight between the hidden layer and the output layer .

The learning and recalling processes of the PNN for classification problems can be found in [86].

7.1.2.3.3 Classification of Transient Signals Using PNN Model

Though the PNN has some disadvantages, such as a large memory requirement and the recalling time being proportional to the quantity of training samples, we can overcome these drawbacks by employing Parseval's theorem to reduce the training inputs.

In this paper, we will perform a 13-level decomposition of each discrete distorted signal to obtain the detailed version coefficients ($w_1 \sim w_3$). Using eq. (64) and eq. (65), we can obtain each detailed energy distribution ($P_1^D \sim P_{13}^D$). These features would be applied to the PNN for recognizing and classifying the distorted signals. The calculation procedures of the proposed classifier are shown in fig. (88).

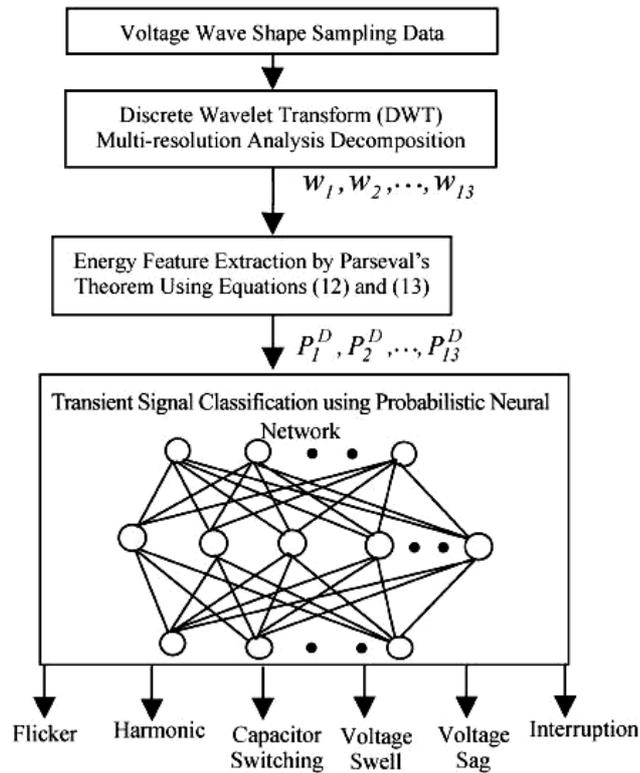


Figure 88. Diagram Of The Proposed Classifier

7.1.2.4 Applications And Results

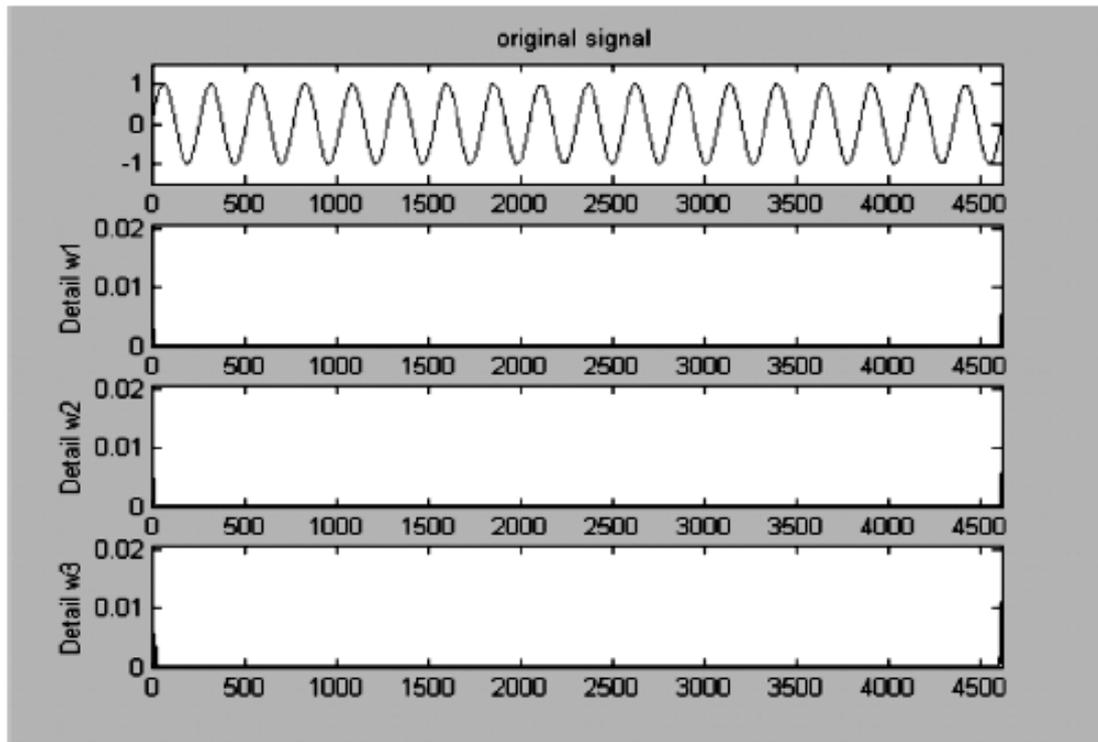
7.1.2.4.1 Testing of The Classifier By Laboratory Simulation

To verify the feasibility of the proposed method, we used the Power System Blockset Toolbox in Matlab to generate one pure sine-wave signal (frequency = 50 Hz, amplitude = 1 p.u.) and six sample transient distorted signals. These distorted signals included momentary interruption, transient due to capacitor switching, voltage sag/swell, harmonic distortion, and flicker. The sampling rate of the digital recorder was 256 points/per cycle. The Daubanchie “db4” wavelet was adopted to perform the DWT.

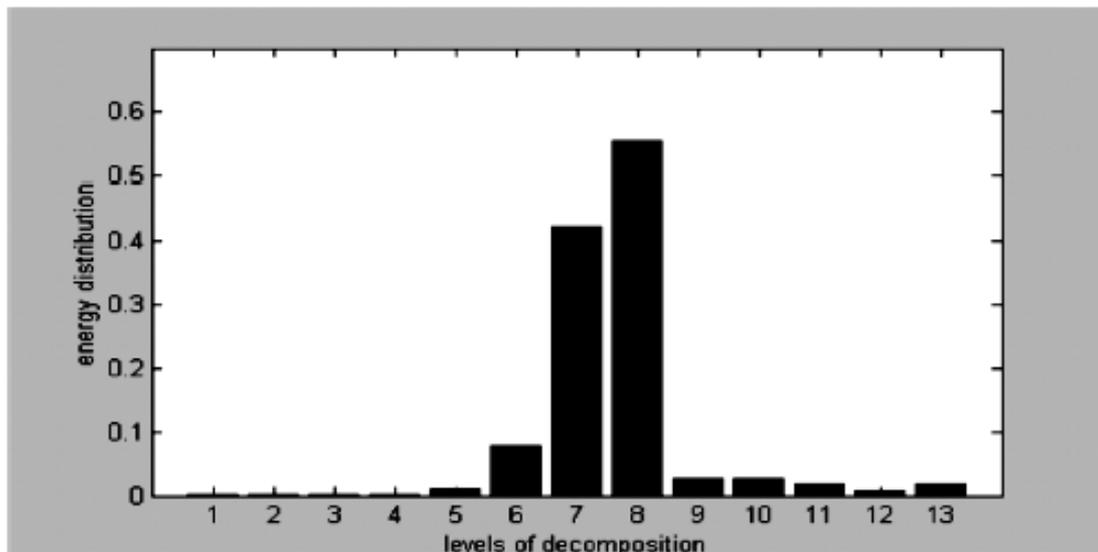
The PNN model was provided by the Neural Network Toolbox in Matlab. The proposed method was written in Matlab language and executed on a Pentium III 550 personal computer with 256-MB RAM.

Fig. 89 shows the detailed version of a three-level decomposition ($w_1 \sim w_3$) and the detailed energy distribution ($P_1^D \sim P_{13}^D$) of a pure sine wave. The X-axis is the sampled signal points and the Y-axis is the magnitude in Fig. 89(a). The X-axis is the de-

composition level and the Y-axis is the energy in Fig. 89(b). Figs. 90–95 show the detailed version of a three-level decomposition ($w_1 \sim w_3$) and the detailed energy distribution ($P_1^D \sim P_{13}^D$) of each given distorted signal. The simulation results are summarized in Table 4.

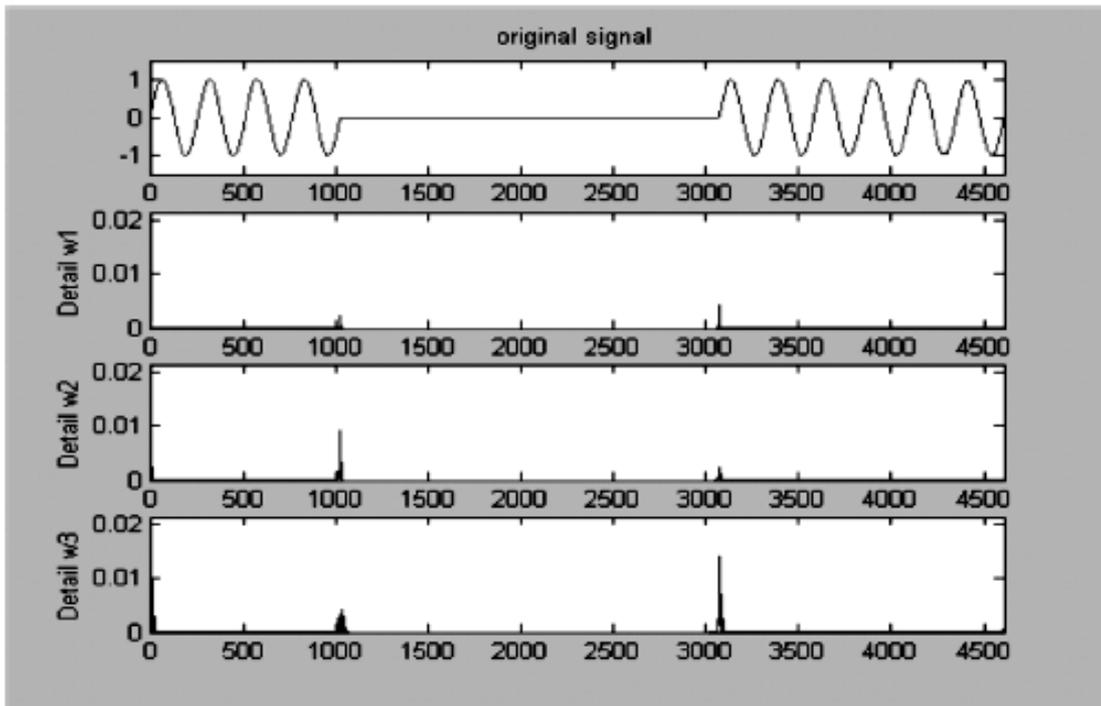


(a)

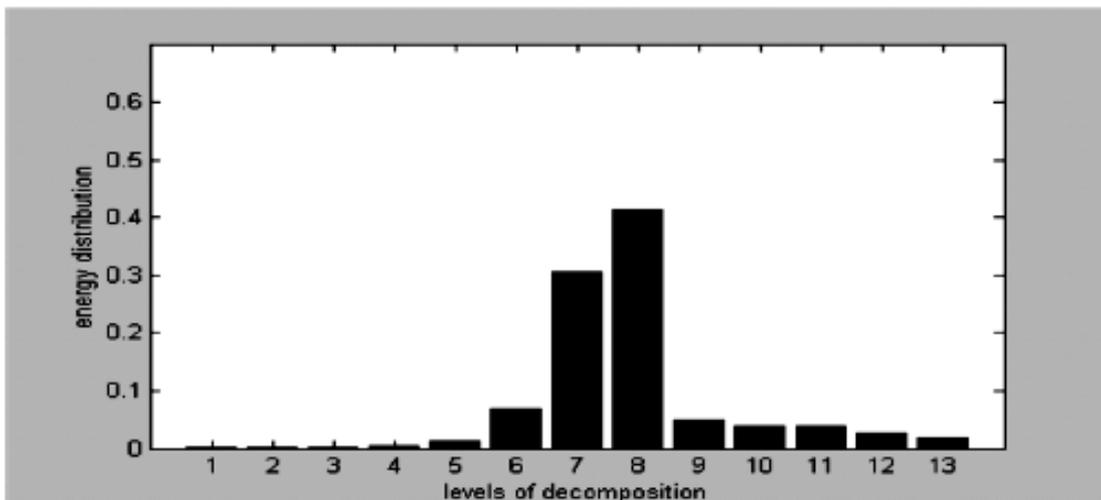


(b)

Figure 89. MRA Decomposition And The Detailed Energy Distribution Of A Pure Sine Wave

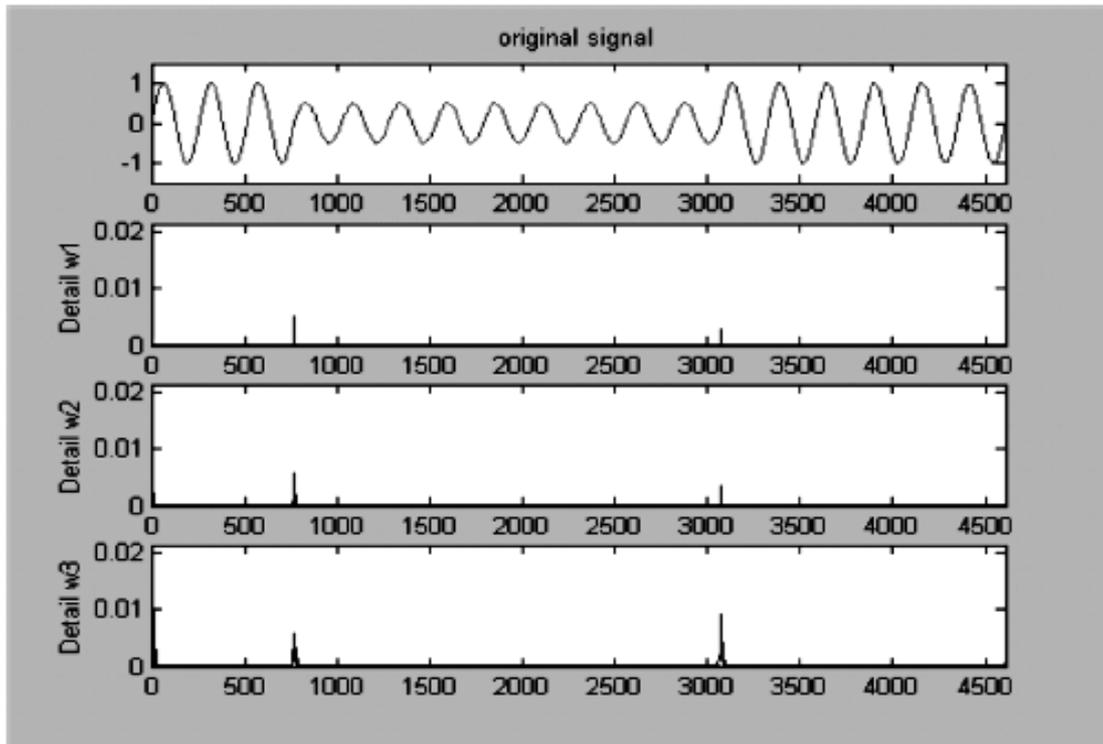


(a)

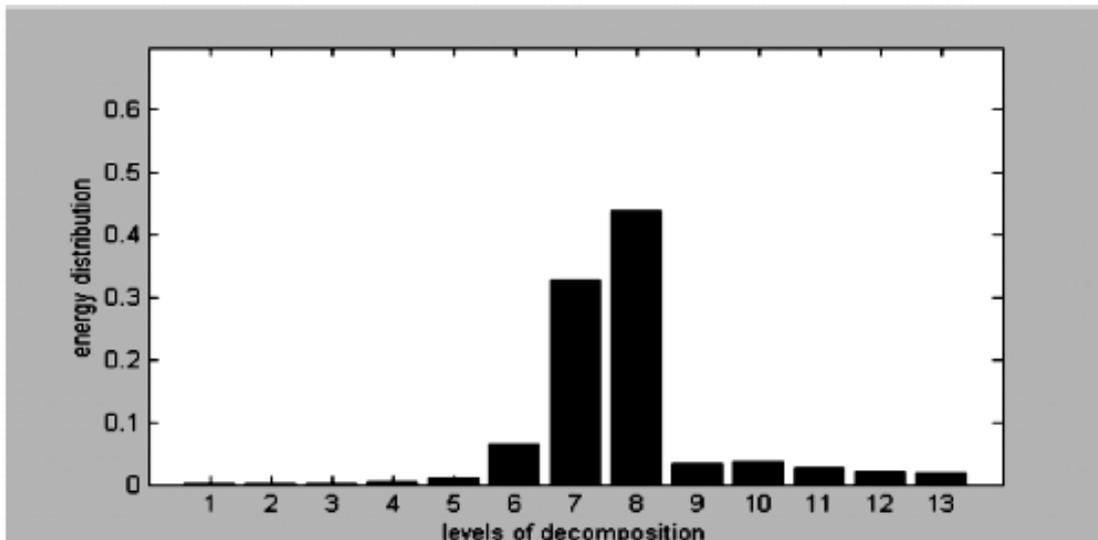


(b)

Figure 90. Detailed Energy Distribution Of A Momentary Interruption

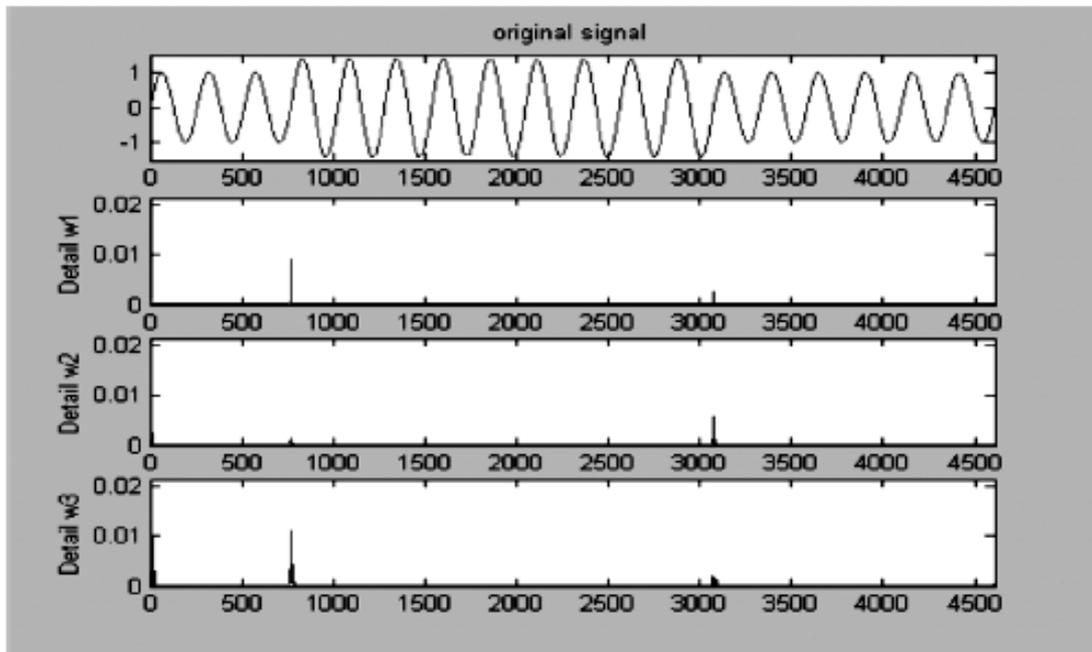


(a)

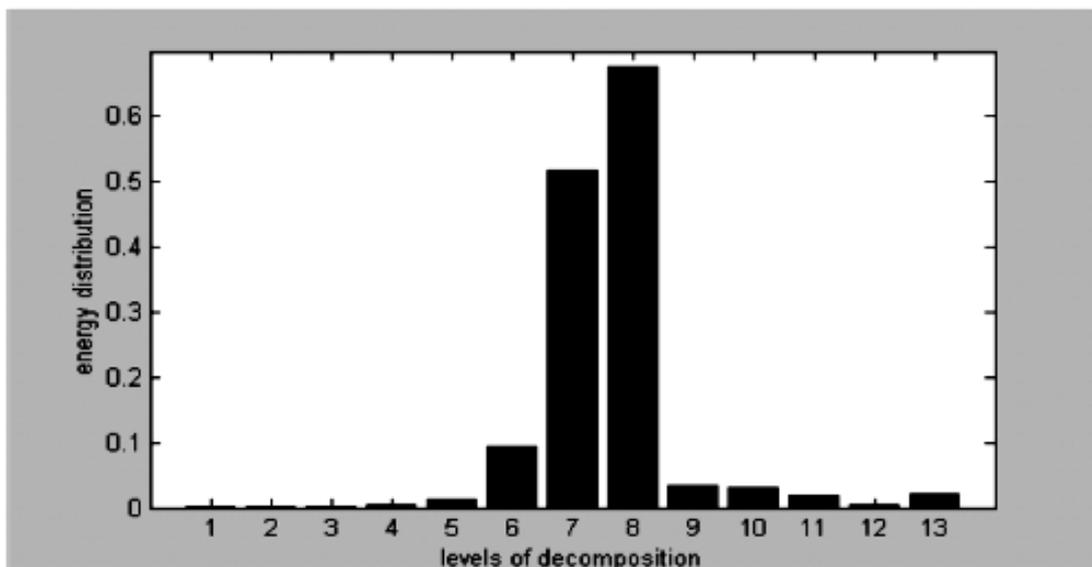


(b)

Figure 91. Detailed Energy Distribution Of A Voltage Sag

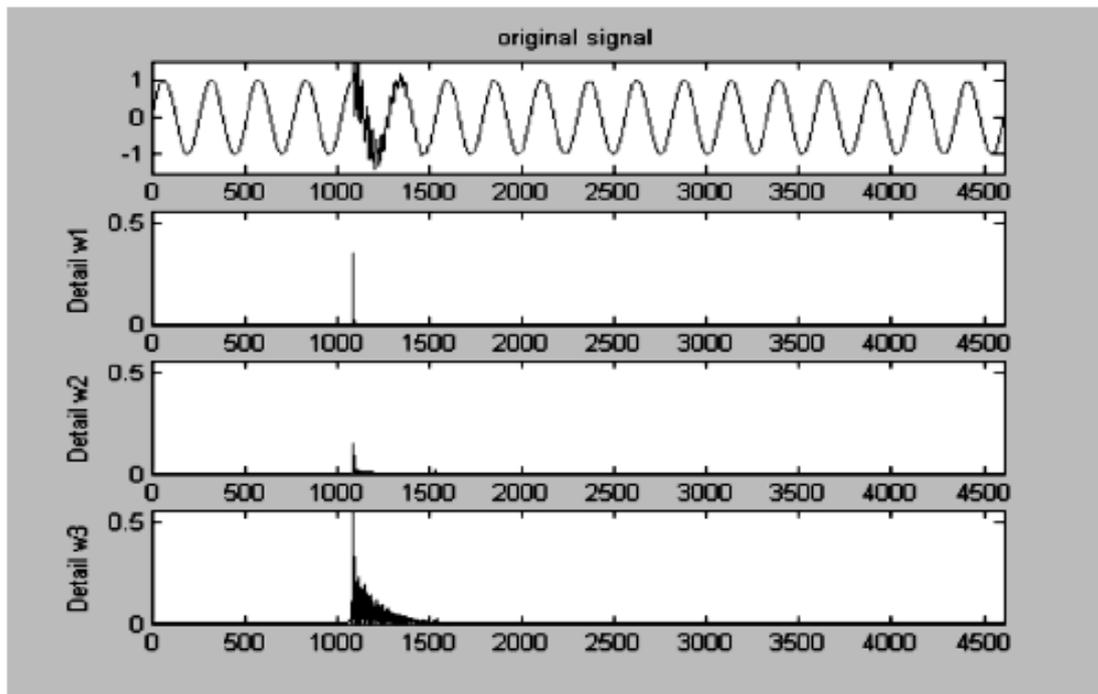


(a)

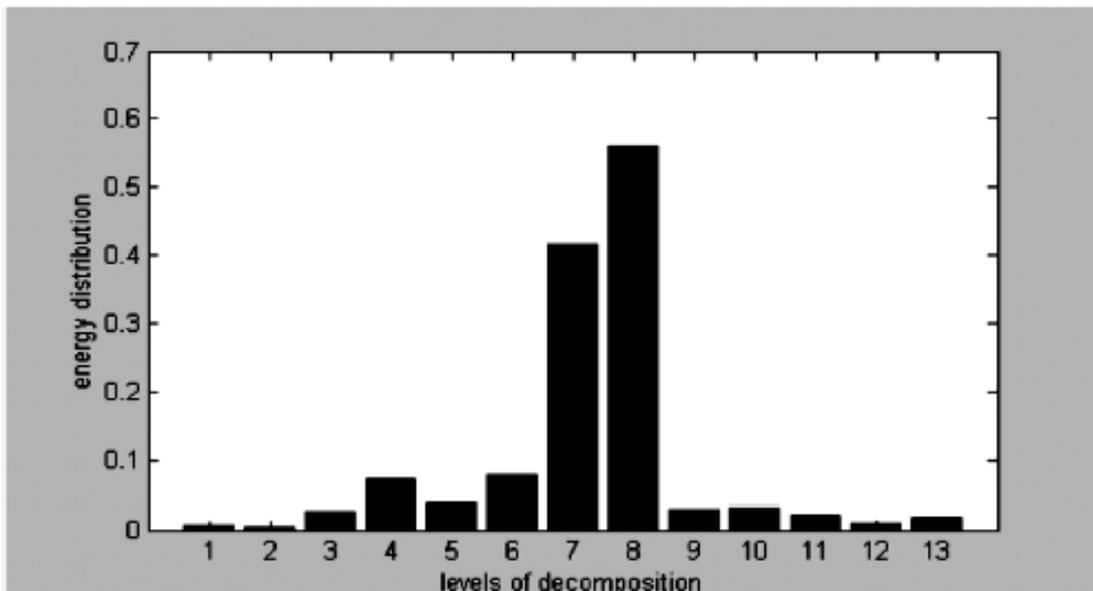


(b)

Figure 92. Detailed Energy Distribution Of A Voltage Swell

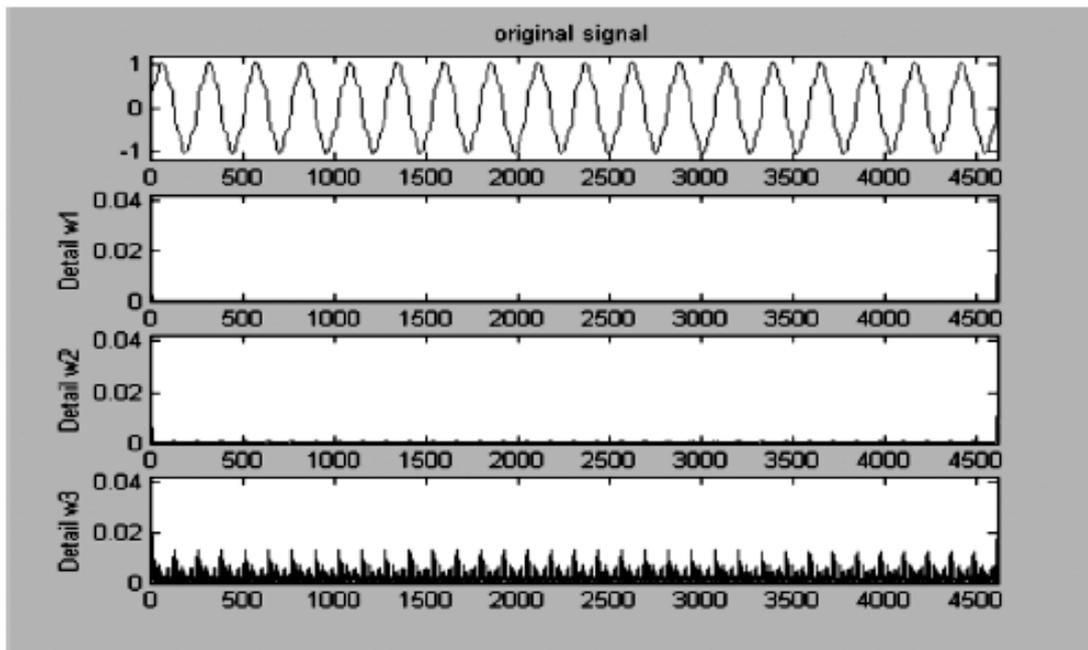


(a)

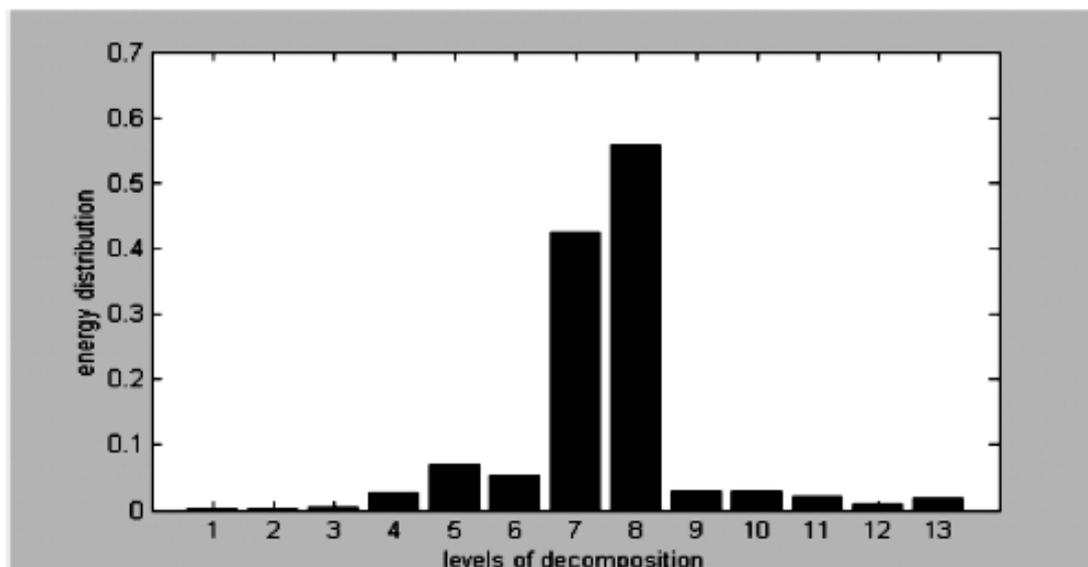


(b)

Figure 93. Detailed Energy Distribution Of A Transient Due To Capacitor Switching

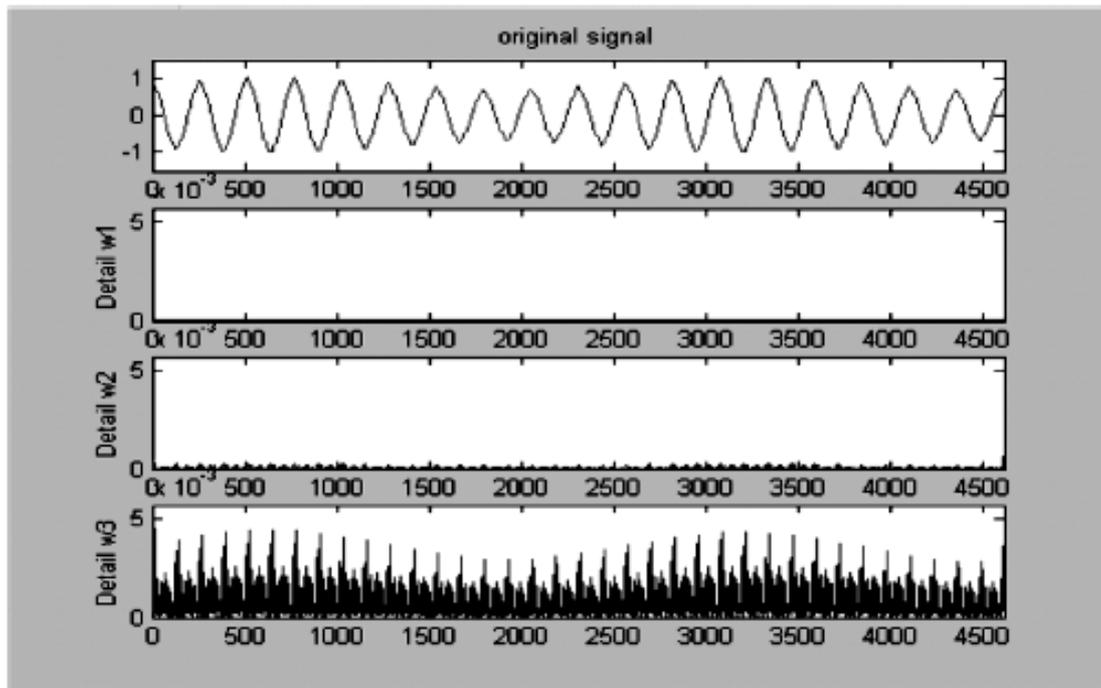


(a)

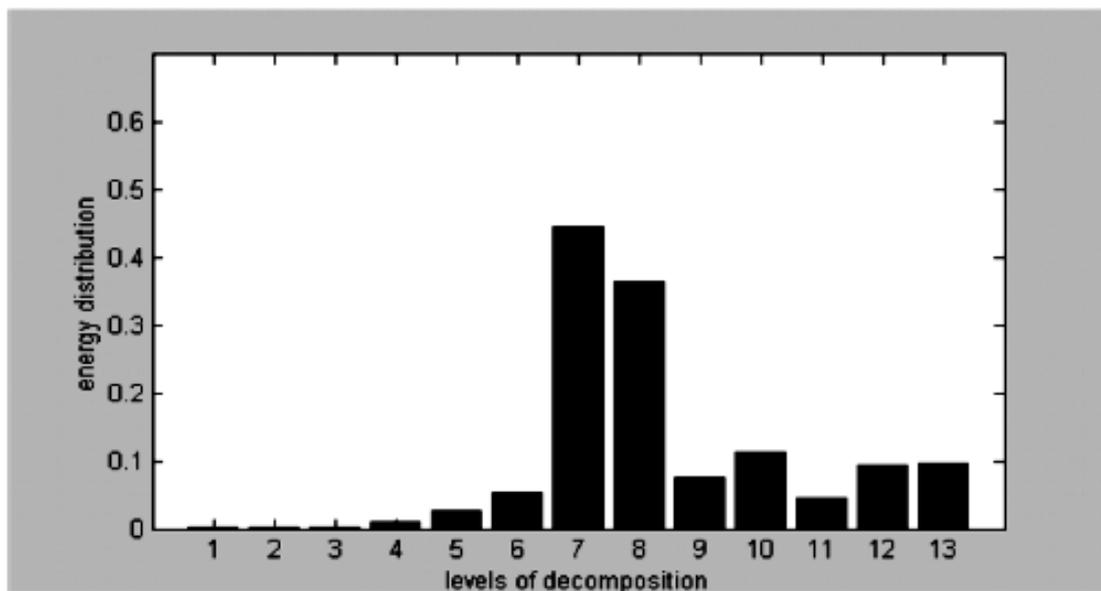


(b)

Figure 94. Detailed Energy Distribution Of A Harmonic Distortion



(a)



(b)

Figure 95. Energy Distribution Of A Flicker

Signal type	Features of distorted signal													Expected output Classification
	P_1^D	P_2^D	P_3^D	P_4^D	P_5^D	P_6^D	P_7^D	P_8^D	P_9^D	P_{10}^D	P_{11}^D	P_{12}^D	P_{13}^D	
Sine Wave	0.0001	0.0004	0.0008	0.0022	0.0085	0.0773	0.4203	0.5576	0.0261	0.0258	0.0182	0.0072	0.0161	0
Interruption	0.0001	0.0005	0.0009	0.0038	0.0126	0.0666	0.3042	0.4133	0.0482	0.0388	0.0384	0.0261	0.0179	1
Sag	0.0001	0.0005	0.0008	0.0029	0.0098	0.0645	0.3282	0.4390	0.0335	0.0369	0.0254	0.0195	0.0187	2
Swell	0.0002	0.0004	0.0008	0.0028	0.0109	0.0941	0.5156	0.6767	0.0332	0.0300	0.0169	0.0039	0.0206	3
Capacitor Switching	0.0077	0.0052	0.0264	0.0713	0.0378	0.0785	0.4171	0.5591	0.0281	0.0301	0.0203	0.0099	0.0190	4
Harmonic	0.0002	0.0010	0.0040	0.0247	0.0697	0.0523	0.4242	0.5580	0.0278	0.0287	0.0210	0.0091	0.0180	5
Flicker	0.0000	0.0001	0.0012	0.0099	0.0246	0.0532	0.4452	0.3639	0.0747	0.1113	0.0445	0.0929	0.0954	6

Table 4. Energy Features Of Distorted Signals

From Figs. 89–95 and Table 4, we can categorize three properties of energy distribution of the given distorted signals. These properties become the foundations for classifying the disturbance type.

- When a sag or swell or interrupt occurs, P_6^D , P_7^D and P_8^D will show great variations.
- When the voltage suffers a transient disturbance of the high-frequency elements such as capacitor switching and harmonic distortion, P_3^D , P_4^D and P_5^D will show obvious variations.
- When the voltage suffers a transient disturbance of the low-frequency elements such as flicker, P_9^D , P_{10}^D and P_{11}^D will show obvious variations.

Fig. 96 shows orderly the energy distributions of seven signals on the same three-dimensional (3-D) coordinate axis (using piecewise plot style). Thus, we can clearly observe the differences in energy distribution between different signals.

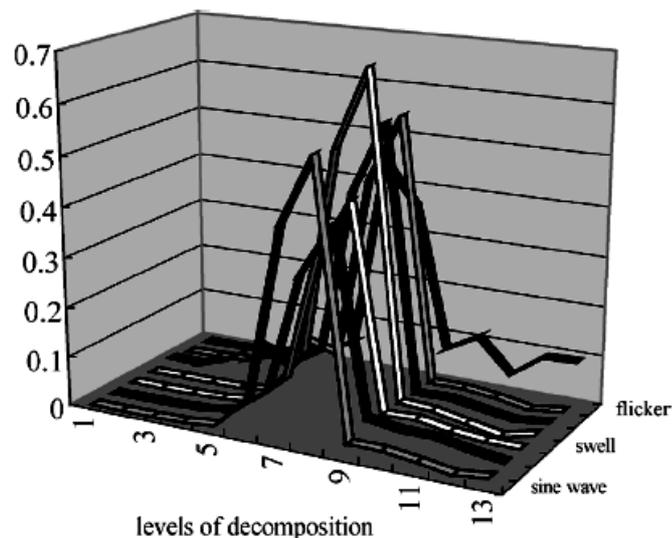


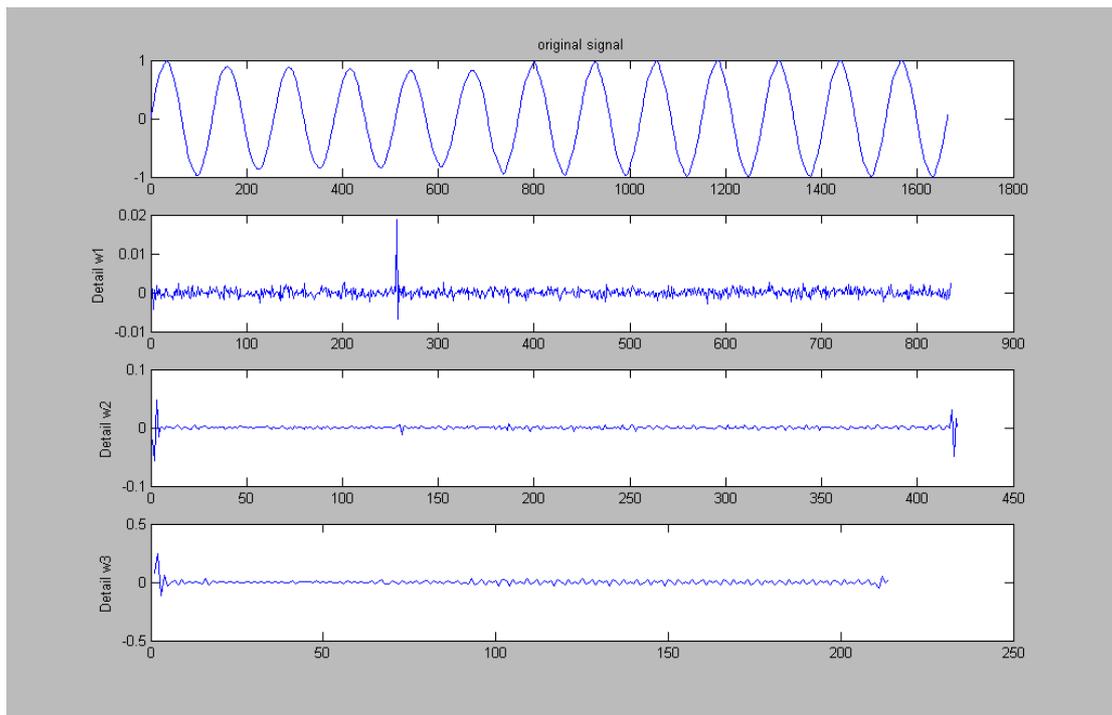
Figure 96. Differences In Energy Distribution Of All Signals

7.1.2.4.2 Application Of The Classifier On A Power System

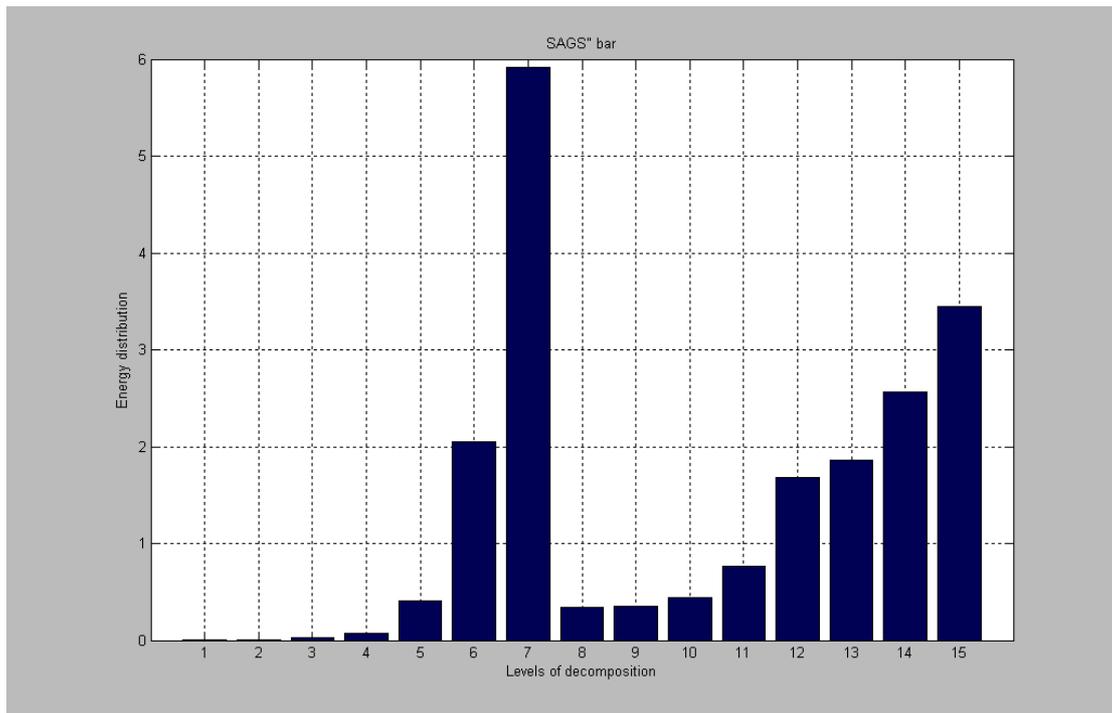
All the above results were obtained by making use of the Matlab in order to create both the pure sine wave and the distorted signals as well. Let’s have a look at the results that were obtained by utilizing the Series 5500 DualNode on the Public Power Constitution at Katsampas at Heraklion, Crete. The DataNode was connected at the capacitors’ 13.8kV busbar of the facility, via the current and voltage pods, the set up of the InfoNode was made according to the IEEE standards, described earlier

regarding all the disturbance events, the data were obtained by the use of the program Dranview and were finally processed on Matlab 6.5v. Unfortunately due to technical difficulties, regarding the time and duration of its use, the only disturbances that were recorded and afterwards processed involved swells, sags and transients. That is the reason why the experimental results, presented later on relate only to these three categories of power disturbance events.

As it will be shown at the figures presented later on, fig. (97-99), the waves that were received from the facility's busbar weren't so "pure" as those that were created at the laboratory. That was mainly because of the intense harmonic distortion of the signal, which may not exceeded the IEEE limits but however altered the signal, and moreover from the electromagnetic insertions of the facility's power instruments. That is the reason that during the training and the testing of the PNN network the following algorithm was chosen: for each group of distorted signals we created a loop. From the beginning till the end of that loop in each circle that was made one signal was set as the testing one while all the others were set as the training examples. That means that in each circle the signal that was previously set as testing example was now set as training example while the following signal that was previously set as training example was now set as testing and so on till the end of the signals. That algorithm led to the results presented on fig. (100-103) and Table 5.

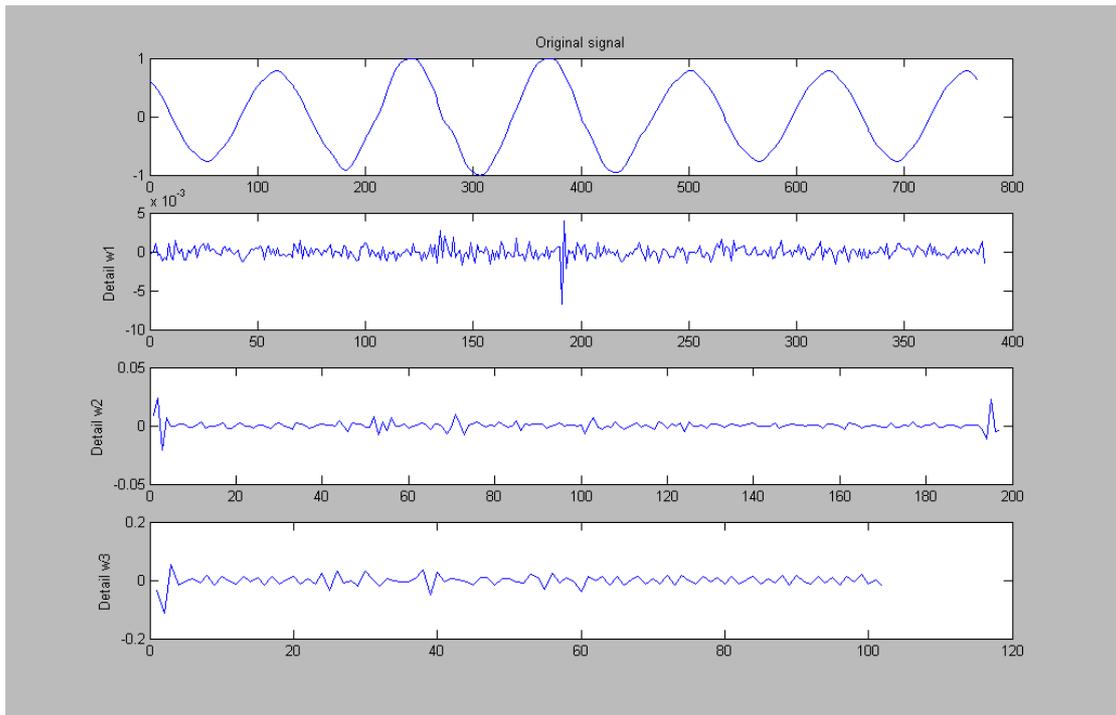


a)

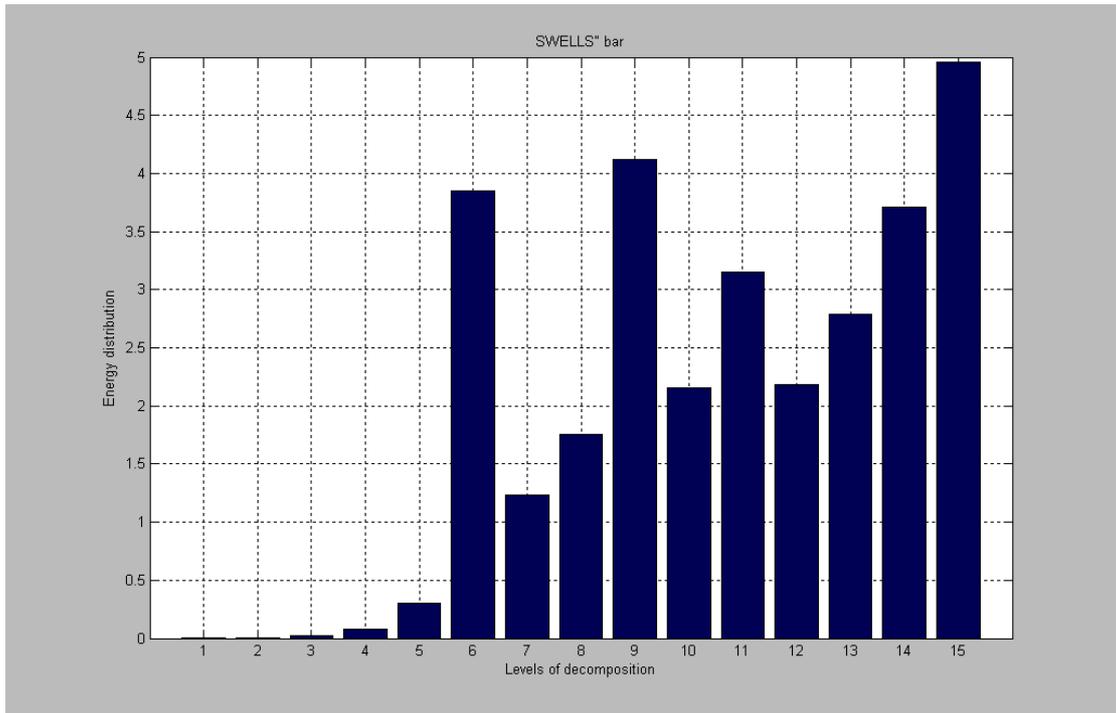


b)

Figure 97. Detailed Energy Distribution Of A Voltage Sag

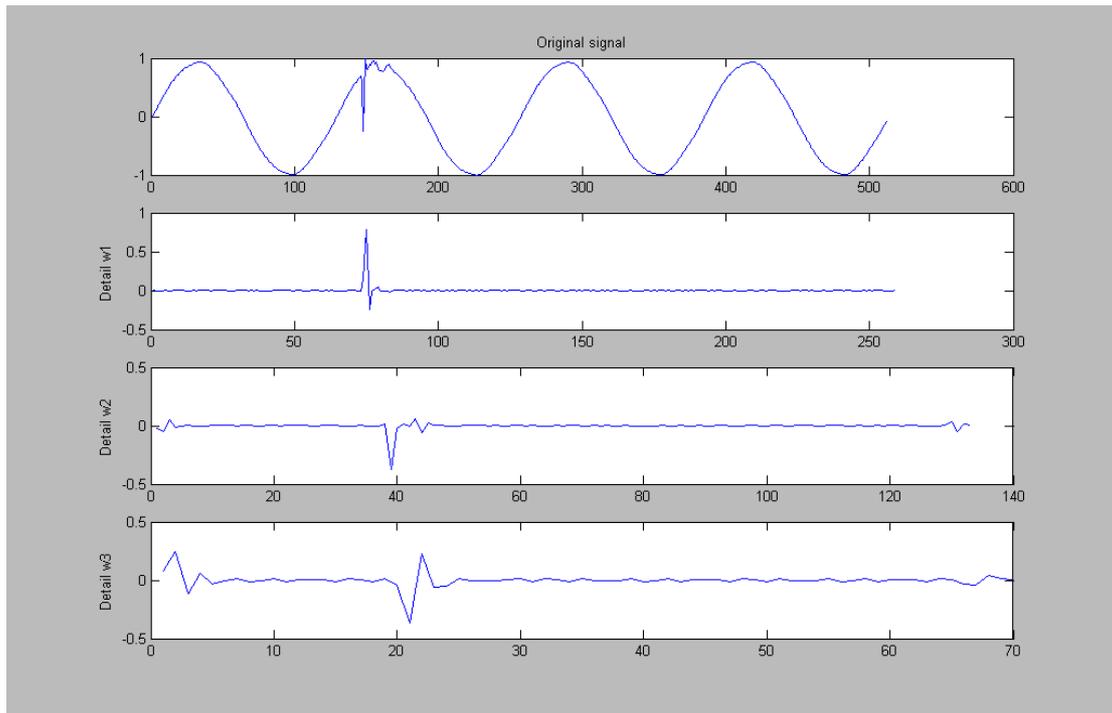


a)

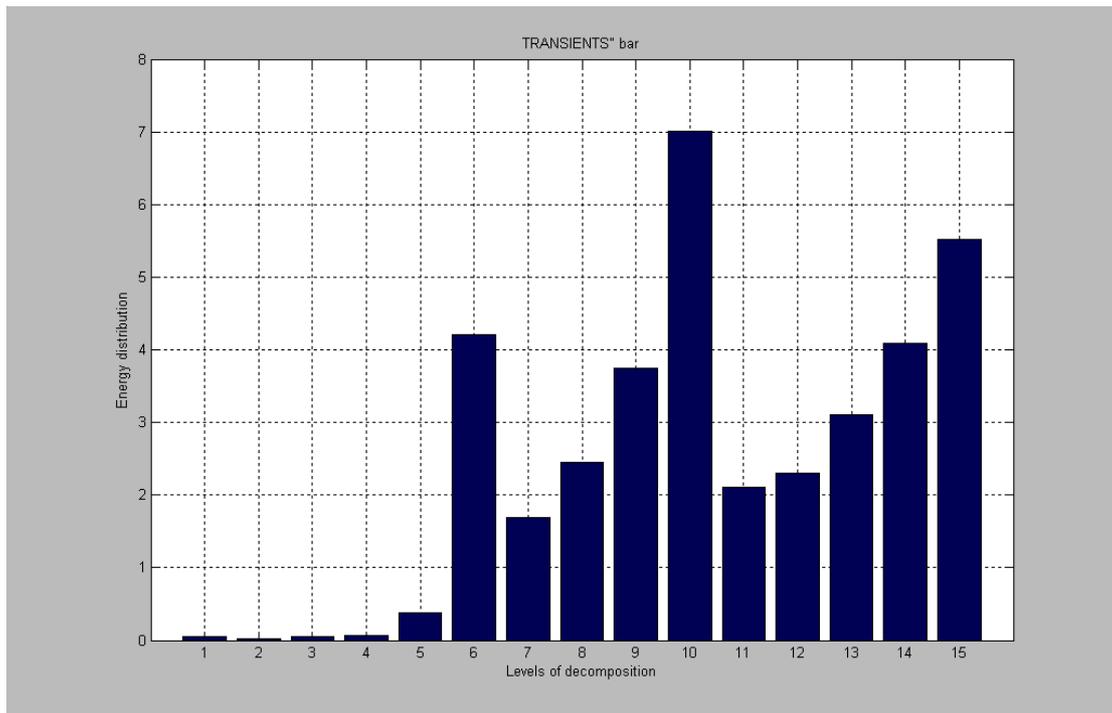


b)

Figure 98. . Detailed Energy Distribution Of A Voltage Swell



a)



b)

Figure 99. Detailed Energy Distribution Of A Transient

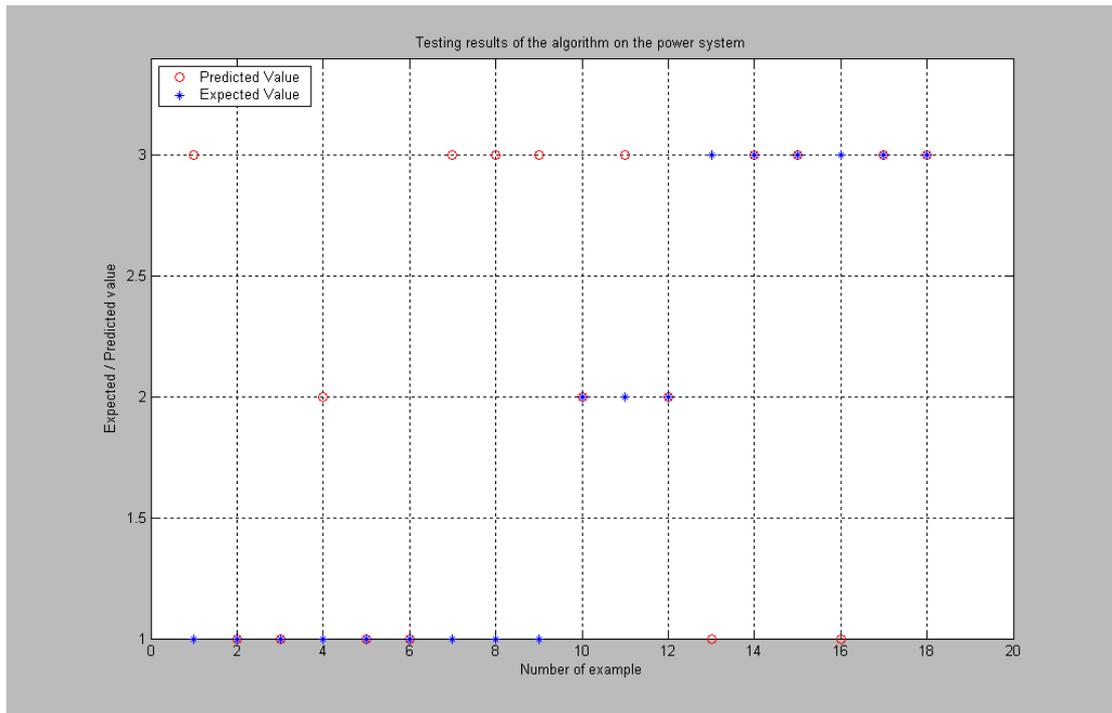


Figure 100. Testing Results Of The PNN Algorithm On The Power System Using Only One Third Of The Training Examples

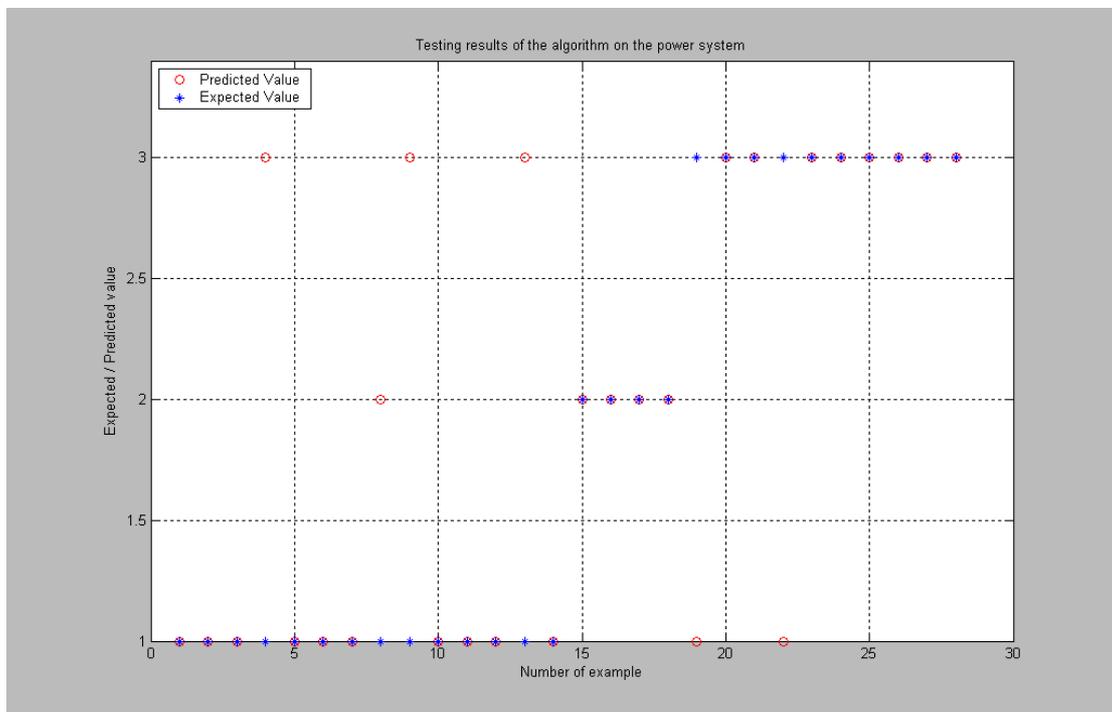


Figure 101. Testing Results Of The PNN Algorithm On The Power System Using Only Half Of The Training Examples

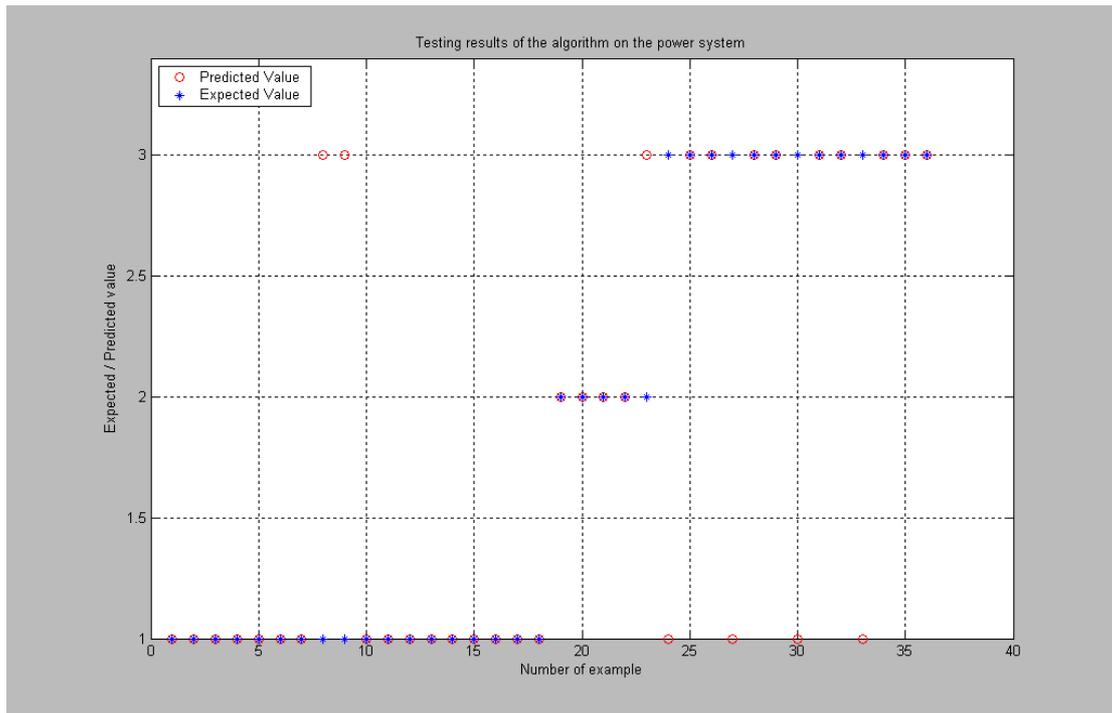


Figure 102. Testing Results Of The PNN Algorithm On The Power System Using Only Two Thirds Of The Training Examples

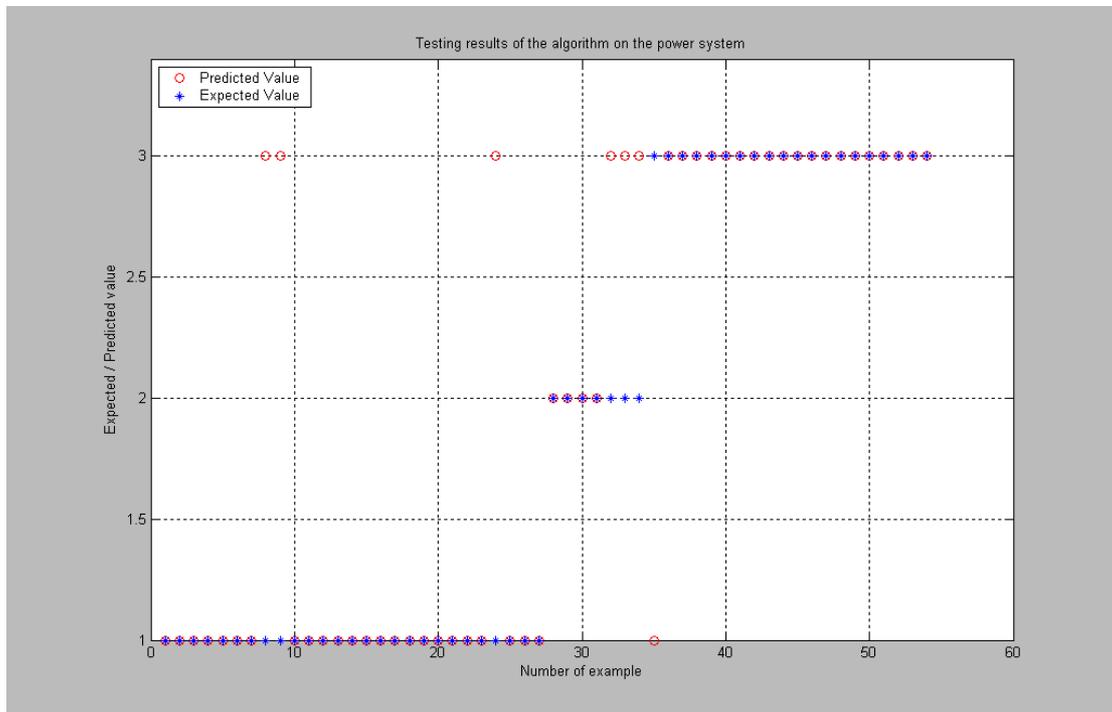


Figure 103. Testing Results Of The PNN Algorithm On The Power System Using All The Training Examples

<i>Number of training examples</i>	Transients	9	14	18	27
	<i>Sags</i>	3	4	5	7
	<i>Swells</i>	6	10	13	20
	Total	18	28	36	54
Number of testing examples		1	1	1	1
Learning time (sec.)		0.00599	0.0608	0.0611	0.0616
Classification time of a new signal (sec.)		0.0061	0.0056	0.0057	0.0061
Accuracy rate		55.556%	78.571%	80.556%	87.037%

Table 5. Test Results Of The PNN Algorithm On The Power System

As it's obvious from the above table the more training examples we have, the better results we obtain. That was expected, because the classification skills of the PNN are directly connected to the number of its training samples, thus the increase of the training samples help the network to adjust its parameters better so that its accuracy of classification gets even higher. However a large number of training samples could complicate the structure of the network in such a way that it could take longer to classify the sample. Therefore it is important for the user to specify if it's of greater importance for him to have a fast algorithm with a satisfying accuracy or a little slower algorithm with an even more satisfying accuracy!

7.1.3 Power Quality Event Recognition and Classification Using a Wavelet-Based Adaptive Neuro-Fuzzy Inference System

In this section, a wavelet-based adaptive neuro-fuzzy inference system is being used instead of the neural based classifier that was used in the above section. The main idea of the algorithm remains the same, only the classifying mean changes. That means that once again the multiresolution-analysis technique of DWT and the Parseval's theorem are employed to extract the energy distribution features of the distorted signal at different resolution levels, however this time ANFIS is employed to classify disturbance types according to the detailed energy distribution.

7.1.3.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The purpose of this clause is to present a novel architecture called Adaptive-Network-Based Fuzzy Inference Systems, or simply ANFIS, which can serve as a basis for constructing a set of fuzzy if-then rules with appropriate membership functions to generate the stipulated input-output pairs.

7.1.3.1.1 Fuzzy If-Then Rules And Fuzzy Inference Systems

7.1.3.1.1.1 Fuzzy If-Then Rules

Fuzzy *if-then* rules or *fuzzy conditional statements* are expressions of the form IF A THEN B, where A and B are labels or fuzzy sets characterized by appropriate membership functions. Due to their concise form, fuzzy if-then rules are often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. An example that describes a simple fact is

If pressure is high, then volume is small

where *volume* and *pressure* are linguistic variables [98], *high* and *small* are linguistic values or labels that are characterized by membership functions.

Another form of fuzzy if-then rule, proposed from Takagi-Sugeno [99], has sets involved only in the premise part. By using Takagi and Sugeno's fuzzy if-then rule, the resistant force on a moving object can be described as follows:

If velocity is high, then force = $k \cdot \text{velocity}^2$

where, again, *high* in the premise part is a linguistic label characterized by an appropriate membership function. However, the consequent part is described by a nonfuzzy equation of the input variable, *velocity*.

Both types of fuzzy if-then rules have been used extensively in both modelling and control. Through the use of linguistic labels and membership functions, a fuzzy if-then rule can easily capture the spirit of a "rule of thumb" used by humans. From another angle, due to the qualifiers on the premise parts, each fuzzy if-then rule can be viewed as a local description of the system under consideration. Fuzzy if-then rules form a core part of the of the fuzzy inference system to be introduced below.

7.1.3.1.1.2 Fuzzy Inference systems

Fuzzy inference systems are also known as *fuzzy-rule-based systems*, *fuzzy models*, *fuzzy associative memories (FAM)*, or *fuzzy controllers* when used as controllers. Basically a fuzzy inference system is composed of five functional blocks, as presented at the following figure.

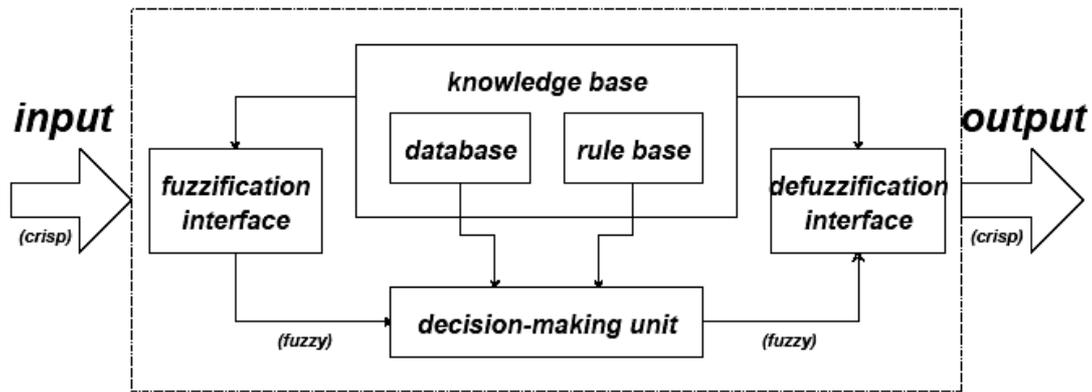


Figure 104. Fuzzy Inference System

- i) a **rule base** containing a number of fuzzy if-then rules,
- ii) a **database** which defines the membership functions of the fuzzy sets used in the fuzzy rules,
- iii) a **decision-making unit** which performs the inference operations on the rules,
- iv) a **fuzzification interface** which transforms the crisp inputs into degrees of match with linguistic values,
- v) a **defuzzification interface** which transform the fuzzy results of the inference into a crisp output.

Usually, the rule base and the database are jointly referred to as the *knowledge base*.

The steps of *fuzzy reasoning* (inference operations upon fuzzy if then rules) performed by fuzzy inference systems are:

1. Compare the input variables with the membership functions on the premise part to obtain the membership values (or compatibility measures) of each linguistic label. This step is often called fuzzification.
2. Combine (through a specific T-norm operator, usually multiplication or min) the membership values on the premise part to get *firing strength (weight)* of each rule.
3. Generate the qualified consequent (either fuzzy or crisp) of each rule depending on the firing strength.
4. Aggregate the qualified consequents to produce a crisp output. This step is called defuzzification.

Several types of fuzzy reasoning [100,101] have been proposed in the literature. Depending on the types of fuzzy reasoning and fuzzy if-then rules employed, most fuzzy inference systems can be classified in three types, as presented.

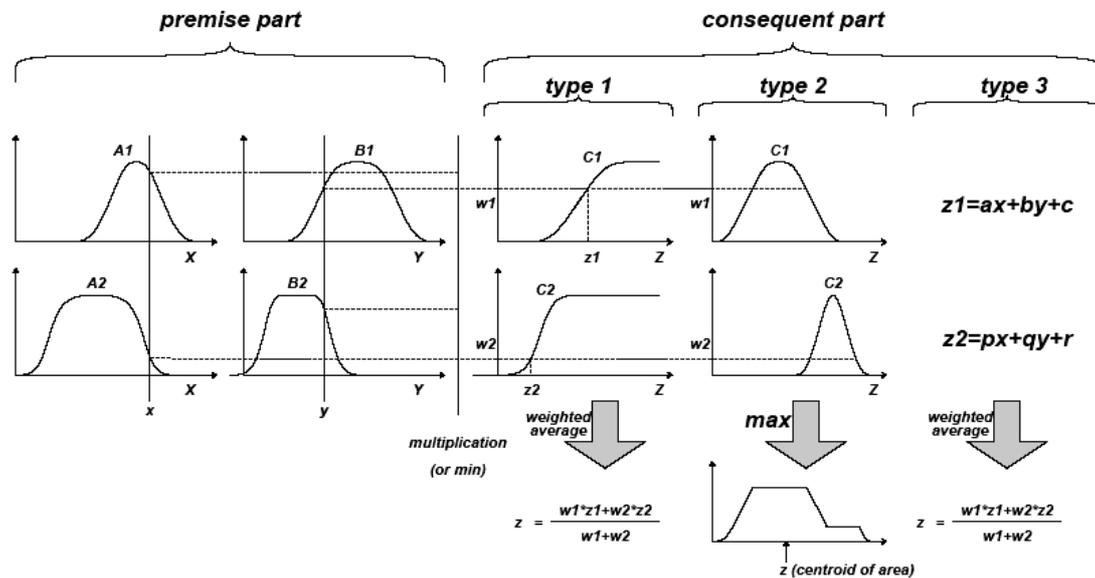


Figure 105. Commonly used fuzzy if-then rules and fuzzy reasoning mechanisms

Type 1: The overall output is the weighted average of each rule’s crisp output induced by the rule’s firing strength and output membership functions. The output membership functions used in this scheme must be monotonically non-decreasing [102].

Type 2: The overall fuzzy output is derived by applying “max” operation to the qualified fuzzy outputs, each of which is equal to the minimum of firing strength and the output membership function of each rule. Various schemes have been proposed to choose the final crisp output based on the overall fuzzy output; some of them are center of area, bisector of area, mean of maxima, maximum criterion etc [100,101].

Type 3: Takagi and Sugeno’s fuzzy if-then rules are used [103]. The output of each rule is a linear combination of input variables plus a constant term, and the final output is the weighted average of each rule’s output.

Fig. (105) utilizes a two-rule two-input fuzzy inference system to show different types of fuzzy rules and fuzzy reasoning mentioned above. Be aware that most of the differences come from the specification of the consequent part (monotonically non-decreasing or bell-shaped membership functions or crisp function) and thus the defuzzification schemes (weighted average, centroid of area, etc) are also different.

7.1.3.1.2 Adaptive Networks: Architectures And Learning Algorithms

This clause introduces the architecture and learning procedure of the adaptive network is in fact a superset of all kinds of feedforward neural networks with supervised learning capability. An adaptive network, as its name implies, is a network structure consisting of nodes and directional links through which the nodes are connected. Moreover, part or all of the nodes are adaptive, which means each output of these nodes depends on the parameter(s) pertaining to this node, and the learning rule specifies how these parameters should be changed to minimize a prescribed error measure.

The basic learning rule of adaptive networks is based on the gradient descent and the chain rule, which was proposed by Werbos [104] in the 1970's. However, due to the state of artificial neural network research at that time, Werbos' early work failed to receive the attention it deserved. In the following presentation, the derivation is based on the author's work [105, 106] which generalises the formulas in [107].

Since the basic learning rule is based on the gradient method which is notorious for its slowness and tendency to become trapped in local minima, here we propose a hybrid learning rule which can speed up the learning process substantially. Both the batch learning and the pattern learning of the proposed hybrid learning rule is discussed below.

7.1.3.1.2.1 Architecture And Basic Learning Rule

An adaptive network, fig. 106, is a multi-layer feedforward network in which each node performs a particular function (*node function*) on incoming signals as well as a set of parameters pertaining to this node. The nature of the node functions may vary from node to node, and the choice of each node function depends on the overall input-output function which the adaptive network is required to carry out. Note that the links in an adaptive network only indicate the flow direction of signals between nodes and no weights are associated with the links.

To reflect different adaptive capabilities, both square and circle nodes are used in an adaptive network. A square node, adaptive node, has parameters while a circle node, fixed node, has none. The parameter set of an adaptive network is the union of the parameter set of each adaptive node. In order to achieve a desired input-output mapping, these parameters are updated according to given training data and a gradient-based learning procedure described below.

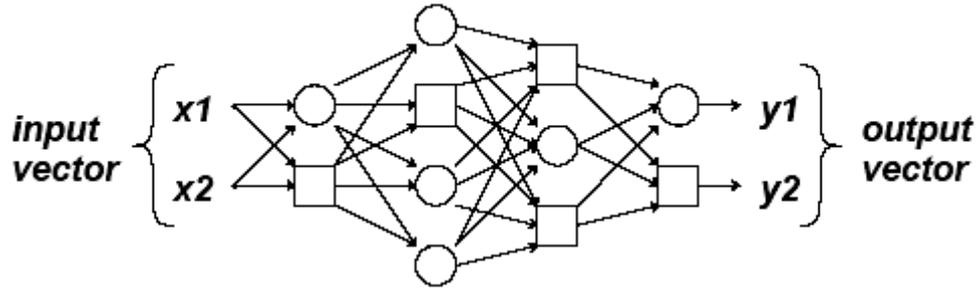


Figure 106. An Adaptive Network

Suppose that a given adaptive network has L layers and the k -th layer has $\#(k)$ nodes. We can denote the node in the i -th position of the k -th layer by (k,i) and its node function, or node output, by O_i^k . Since a node output depends on its incoming signals and its parameter set, there is

$$O_i^k = O_i^k(O_i^{k-1}, \dots, O_{\#(k-1)}^k, a, b, c, \dots) \quad \text{Equation 69}$$

where a, b, c, \dots are the parameters pertaining to this node. Note that O_i^k is used both as node function and output.

Assuming the given training data set has P entries, the *error measure* (or *energy function*) for the p -th ($1 \leq p \leq P$) entry of training data entry can be defined as the sum of squared errors:

$$E_p = \sum_{m=1}^{\#(L)} (T_{m,p} - O_{m,p}^L)^2 \quad \text{Equation 70}$$

where $T_{m,p}$ is the m -th component of p -th target output vector, and $O_{m,p}^L$ is the m -th component of actual output vector produced by the presentation of the p -th input vector. Hence the overall error measure is

$$E = \sum_{p=1}^P E_p \quad \text{Equation 71}$$

In order to develop a learning procedure that implements gradient descent in E over the parameter space, first the error rate $\frac{\partial E_p}{\partial O}$ for p -th training data and for each node output O must be calculated. The error rate for the output node at (L,i) can be calculated readily from eq. (70):

$$\frac{\partial E_p}{\partial O_{i,p}^L} = -2(T_{i,p} - O_{i,p}^L) \quad \text{Equation 72}$$

For the internal node at (k,i) the error rate can be derived by the chain rule:

$$\frac{\partial E_p}{\partial O_{i,p}^L} = \sum_{m=1}^{\#(k+1)} \frac{\partial E_p}{\partial O_{m,p}^{k+1}} \frac{\partial O_{m,p}^{k+1}}{\partial O_{i,p}^k} \quad \text{Equation 73}$$

where $1 \leq k \leq L-1$. That is, the error rate of an internal node can be expressed as a linear combination of the error rates of the nodes in the next layer. Therefore for all $1 \leq k \leq L$ and $1 \leq i \leq \#(k)$, $\frac{\partial E_p}{\partial O_{i,p}^k}$ can be found by eq. (72) and eq. (73).

Now, if α is a parameter of the given adaptive network, there is

$$\frac{\partial E_p}{\partial \alpha} = \sum_{O^* \in S} \frac{\partial E_p}{\partial O^*} \frac{\partial O^*}{\partial \alpha} \quad \text{Equation 74}$$

where S is the set of nodes whose outputs depend on α . Then the derivative of the overall error measure E with respect to α is

$$\frac{\partial E}{\partial \alpha} = \sum_{p=1}^P \frac{\partial E_p}{\partial \alpha} \quad \text{Equation 75}$$

Accordingly, the update formula for the generic parameter α is

$$\Delta \alpha = -\eta \frac{\partial E}{\partial \alpha} \quad \text{Equation 76}$$

in which η is a learning rate which can be further expressed as

$$\eta = \frac{k}{\sqrt{\sum_a \left(\frac{\partial E}{\partial a} \right)^2}} \quad \text{Equation 77}$$

where k is the *step size*, the length of each gradient transition in the parameter space. Usually, the value of k can be changed to vary the speed of convergence.

Actually there are two learning paradigms for adaptive networks. With the *batch learning* (or *off-line training*), the update formula for parameter is based on eq. (75) and the update action takes place only after the whole training data set has been presented., i.e., only after each *epoch* or *sweep*. On the other hand, if there is needed for the parameters to be updated immediately after each input-output has been presented, then the update formula is based on eq. (74) and it is referred to as the *pattern*

learning (or on-line learning). In the following paragraphs there will be presented a faster hybrid learning rule and both of its learning paradigms.

7.1.3.1.3 Hybrid Learning Rule: Batch (Off-Line) Learning

Here, a hybrid learning rule [88] is proposed which combines the gradient method and the least squares estimate (LSE) to identify parameters.

For simplicity, assume that the adaptive network under consideration has only one output

$$\text{output} = F(\bar{I}, S) \quad \text{Equation 78}$$

where \bar{I} is the set of input variables and S is the set of parameters. If there exists a function $H \circ F$ is linear in some of the elements of S , then these elements can be identified by the least squares method. More formally, if the parameter set S can be decomposed into two sets

$$S = S_1 \oplus S_2 \quad \text{Equation 79}$$

(where \oplus represents direct sum) such that $H \circ F$ is linear in the elements in S_2 , then upon applying H to eq. (78) we have

$$H(\text{output}) = H \circ F(\bar{I}, S) \quad \text{Equation 80}$$

which is linear in the elements in S_2 . Now given values of elements of S_1 , P training data can be plugged into eq. (80) and a matrix equation can be obtained:

$$AX = B \quad \text{Equation 81}$$

where X is an unknown vector whose elements are parameters in S_2 . Let $|S_2|=M$, then the dimensions of A , X and B are $P \times M$, $M \times 1$ and $P \times 1$, respectively. Since P (number of training data pairs) is usually greater than M (number of linear parameters), this is an overdetermined problem and generally there is no exact solution to eq. (81). Instead a *least squares estimate* (LSE) of X , X^* , is required to minimize the squared error $\|AX-B\|^2$. This is a standard problem that forms the grounds for linear regression, adaptive filtering and signal processing. The most well-known formula for X^* uses the pseudo-inverse of X

$$X^* = (A^T A)^{-1} A^T B \quad \text{Equation 82}$$

where A^T is the transpose of A , and $(A^T A)^{-1} A^T$ is the pseudo-inverse of A if $A^T A$ is non singular. While eq. (82) is concise in notation, it is expensive in computation when dealing with the matrix inverse and, moreover, it becomes ill-defined if $A^T A$ is singular. As a result, sequential formulas are employed to compute the LSE of X . this

sequential method of LSE is more efficient (especially when M is small) and can be easily modified to an –online version, as it will be presented below, for systems with changing characteristics. Specifically, let the i th row vector of matrix A defined in eq. (82) be a_i^T and the i th element of B be b_i^T , then X can be calculated iteratively using the sequential formulas widely adopted in the literature [89, 90, 91, 92]

$$\left. \begin{aligned} X_{i+1} &= X_i + S_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T X_i) \\ S_{i+1} &= S_i - \frac{S_i a_{i+1} a_{i+1}^T S_i}{1 + a_{i+1}^T S_i a_{i+1}}, i = 0, 1, \dots, P-1 \end{aligned} \right\} \quad \text{Equation 83}$$

where S_i is often called the *covariance matrix* and the least squares estimate X^* is equal to X_p . The initial conditions to bootstrap eq. (83) are $X_0=0$ and $S_0=\gamma I$, where γ is a positive large number and I is the identity matrix of dimension $M \times M$. When dealing with multi-output adaptive networks (output in eq. (78) is a column vector), eq. (83) still applies except that b_i^T is the i -th row of matrix B .

Now the gradient method can be combined and the least squares estimate to update the parameters in an adaptive network. Each epoch of this hybrid learning procedure is composed of a forward pass and a backward pass. In the forward pass, input data are supplied and functional signals go forward to calculate each node output until the matrices A and B in eq. (81) are obtained, and the parameters in S_2 are identified by the sequential least squares formulas in eq. (83). After identifying parameters in S_2 , the functional signals keep going forward till the error measure is calculated. In the backward pass, the error rates (the derivative of the error measure w.r.t. each node output) propagate from the output end toward the input end, and the parameters in S_1 are updated by the gradient method in equation

$$\frac{\partial E}{\partial a} = \sum_{p=1}^P \frac{\partial E_p}{\partial a} \quad \text{Equation 84}$$

where,

$$E \text{ is the overall error measure } E = \sum_{p=1}^P E_p, \quad \text{Equation 85}$$

P is the number of the entries of the training data sets and

a is the parameter pertaining to the nodes of the network

For given fixed values of parameters in S_1 , the parameters in S_2 found are guaranteed to be the global optimum point in the S_2 parameter space due to the choice of the squared error measure. Not only can this hybrid learning rule decrease the di-

mension of the search space in the gradient method, but, in general, it will also cut down substantially the convergence time.

Take for example an one-hidden-layer back-propagation neural network with sigmoid activation functions. If this neural network has p output units, then the output in eq. (78) is a column vector. Let $H(\cdot)$ be the inverse sigmoid function

$$H(x) = \ln\left(\frac{x}{1-x}\right) \quad \text{Equation 86}$$

Then eq. (80) becomes a linear (vector) function such that each element of $H(\text{output})$ is a linear combination of the parameters (weights and thresholds) pertaining to layer 2. In other words,

S_1 = weights and threshold of hidden layer,

S_2 = weights and threshold of output layer

Therefore the back-propagation learning rule can be applied to tune the parameters in the hidden layer, and the parameters in the output layer can be identified by the least squares method. However, it should be kept in mind that by using the least squares method on the data transformed by $H(\cdot)$, the obtained parameters are optimal in terms of the transformed squared error measure instead of the original one. Usually this will not cause practical problem as long as $H(\cdot)$ is monotonically increasing.

7.1.3.1.4 Hybrid Learning Rule: Pattern (On-Line) Learning

If the parameters are updated after each data presentation, we have the *pattern learning* or *on-line learning* paradigm. This learning paradigm is vital to the on-line parameter identification for systems with changing characteristics. To modify the batch learning rule to its on-line version, it is obvious that that the gradient descent should be based on Ep instead of E (see eq. (85)). Strictly speaking, this is not a truly gradient search procedure to minimize E , yet it will approximate to one if the learning rate is small.

For the sequential least squares formulas to account for the time varying characteristics of the incoming data, the effects of the old data pairs need to be decayed as new data pairs become available. Again, this problem is well studied in the adaptive control and system identification literature and a number of solutions are available [90]. One simple method is to formulate the squared error measure as weighted ver-

sion that gives higher weighting factors to more recent data pairs. This amounts to the addition of a *forgetting factor* λ to the original sequential formula.

$$\left. \begin{aligned} X_{i+1} &= X_i + S_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T X_i) \\ S_{i+1} &= \frac{1}{\lambda} \left[S_i - \frac{S_i a_{i+1} a_{i+1}^T S_i}{1 + a_{i+1}^T S_i a_{i+1}} \right] \end{aligned} \right\} \text{Equation 87}$$

where the value of λ is between of 0 and 1. The smaller λ is, the faster the effects of old data decay. But a small λ sometimes causes numerical instability and should be avoided.

7.1.3.1.5 ANFIS: Adaptive-Network-Based Fuzzy Inference System

The architecture and learning rules of adaptive networks have been described in the previous section. Functionally, there are almost no constraints on the node functions of an adaptive network except piecewise differentiability. Structurally, the only limitation of network configuration is that it should be of feedforward type. Due to these minimal restrictions, the adaptive network's applications are immediate and immense in various areas. In this clause, a class of adaptive networks is proposed which are functionally equivalent to fuzzy inference systems. The proposed architecture is referred to as *ANFIS*, standing for Adaptive-Network-based Fuzzy Inference System. The decomposition of the parameter set is described in order to apply the hybrid learning rule.

7.1.3.1.5.1 Adaptive Neuro-Fuzzy Inference System (ANFIS) Architecture

For simplicity it is assumed that the fuzzy inference system under consideration has two inputs x and y one output z . Suppose that that the rule base contains two fuzzy if-then rules of Takagi and Sugeno's (87) type:

Rule 1: If x is A_1 and y is B_1 then $f_1 = p_1 x + q_1 y + r_1$

Rule 2: If x is A_2 and y is B_2 then $f_2 = p_2 x + q_2 y + r_2$

then the type-3 fuzzy reasoning is illustrated in fig. (107.a) and the corresponding equivalent ANFIS architecture is shown in fig. (107.b).

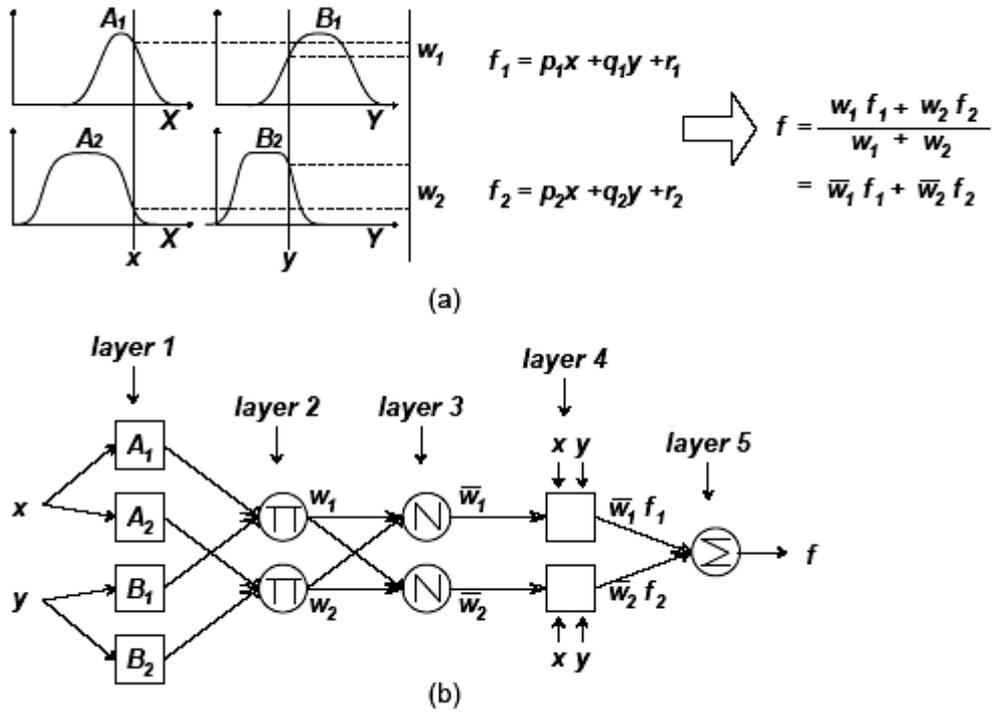


Figure 107. (a) Type-3 Fuzzy Reasoning, (b) Equivalent ANFIS (type-3 ANFIS)

The node functions in the same layer are of the same function family as described below:

Layer 1: Every node i in this layer is a square node with a node function

$$O_i^1 = \mu_{A_i}(x) \quad \text{Equation 88}$$

where x is the input to node i and A_i is the linguistic label associated with this node function. In other words O_i^1 is the membership function of A_i and it specifies the degree to which the given x satisfies the quantifier A_i . Usually $\mu_{A_i}(x)$ is chosen to be bell-shaped with maximum equal to 1 and minimum equal to 0 such as the generalized bell-function

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\frac{(x - c_i)^2}{a_i} \right]^{b_i}} \quad \text{Equation 89}$$

or the Gaussian function

$$\mu_{A_i}(x) = \exp \left\{ - \left[\frac{(x - c_i)^2}{a_i} \right]^{b_i} \right\} \quad \text{Equation 90}$$

where $\{a_i, b_i, c_i\}$ is the parameter set. As the values of these parameter set change, the bell functions vary accordingly, thus exhibiting various forms of member-

ship functions on linguistic label A_1 . In fact, any continuous and piecewise differentiable functions, such as commonly used trapezoidal or triangular-shaped functions, are also qualified candidates for node functions in this layer. Parameters in this layer are referred to as *premise parameters*.

Layer 2: every node in this layer is a circle node labeled Π which multiplies the incoming signals and sends the product out. For instance,

$$w_i = \mu_{A_1}(x) \times \mu_{B_1}(y), \quad i = 1, 2. \quad \text{Equation 91}$$

Each node output represents the firing strength of a rule. (in fact, other *T-norm* operators that perform generalized AND can be used as the node function in this layer)

Layer 3: every node in this layer is a circle node labeled N . The i -th node calculates the ratio of the i -th rule's firing strength to the sum of all rule's strengths:

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \quad \text{Equation 92}$$

For convenience, outputs of this layer will be called *normalized firing strengths*.

Layer 4: Every node i in this layer is a square node with a node function

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \quad \text{Equation 93}$$

where \overline{w}_i is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer will be referred to as *consequent parameters*.

Layer 5: The single node in this layer is a circle node labeled Σ that computes the overall output as the summation of all incoming signals i.e.,

$$O_i^5 = \text{overall output} = \sum_i \overline{w}_i f_i = \frac{\sum_i \overline{w}_i f_i}{\sum_i \overline{w}_i} \quad \text{Equation 94}$$

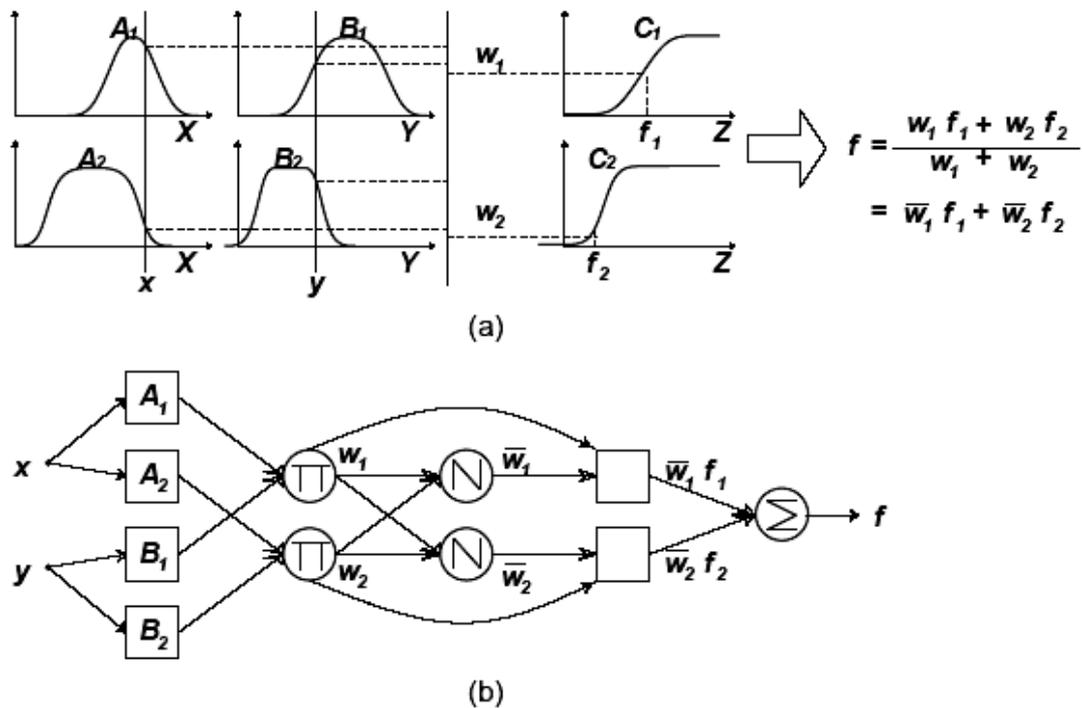


Figure 108. (a) Type-1 fuzzy reasoning, (b) equivalent ANFIS (type-1 ANFIS)

Thus an adaptive network has been constructed which is functionally equivalent to a type-3 fuzzy inference system. For type-1 fuzzy inference systems, the extension is quite straightforward and the type-1 ANFIS is shown in fig. 108 where the output of each rule is induced jointly by the output membership function and the firing strength.

Fig. 109 shows a 2-input, type-3 ANFIS with 9 rules. Three membership functions are associated with each input, so the input space is portioned into 9 fuzzy subspaces, each of which is governed by a fuzzy if-then rule. The premise part of a rule delineates a fuzzy subspace, while the consequent part specifies the output within this fuzzy subspace.

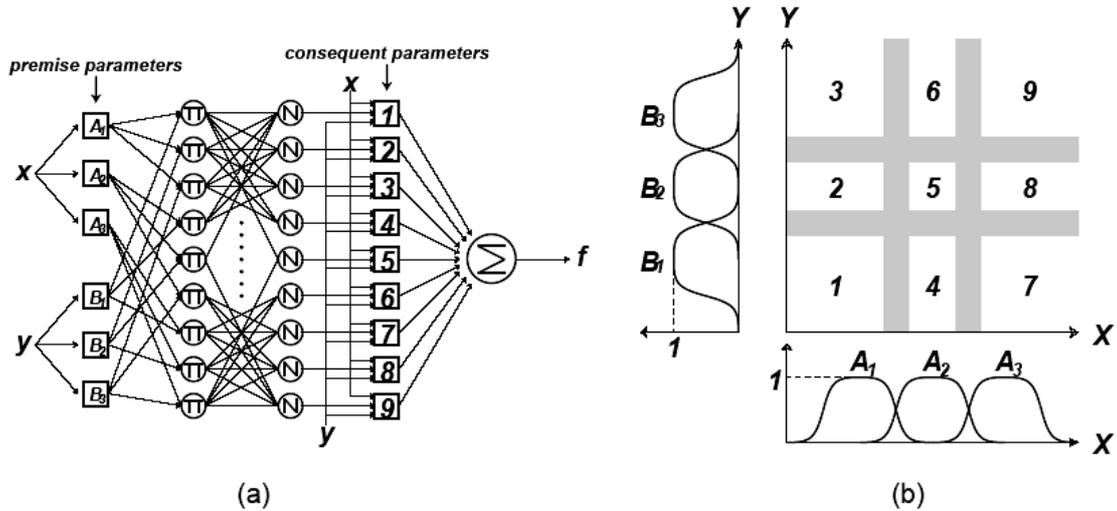


Figure 109. a)2-Input Type-3 ANFIS With 9 Rules, b) Corresponding Fuzzy Subspaces

7.1.3.1.5.2 Hybrid Learning Algorithm

From the proposed type-3 ANFIS architecture, Fig. 109, it is observed that given the values of premise parameters, the overall output can be expressed as linear combinations of the consequent parameters. More precisely, the output f in Fig. 109 can be rewritten as

$$\begin{aligned}
 f &= \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \\
 &= \overline{w_1} f_1 + \overline{w_2} f_2 \quad \text{Equation 95} \\
 &= (\overline{w_1} x) p_1 + (\overline{w_1} y) q_1 + (\overline{w_1}) r_1 + (\overline{w_2} x) p_2 + (\overline{w_2} y) q_2 + (\overline{w_2}) r_2,
 \end{aligned}$$

which is linear in the consequent parameters (p_1, q_1, r_1, p_2, q_2 and r_2). As a result we have in eq. (77).

S = set of total parameters,

S_1 = set of premise parameters,

S_2 = set of consequent parameters,

$H(\cdot)$ and $F(\cdot)$ in eq. (80) are the identity function and the function of the fuzzy inference system, respectively. Therefore the hybrid learning algorithm developed in the previous paragraphs can be applied directly. More specifically, in the forward pass, the error rates propagate backward and the premise parameters are updated by the gradient descent. Table 6 summarizes the activities in each pass.

As mentioned earlier, the consequent parameters thus identified are optimal (in the consequent parameter space) under the condition that the premise parameters are fixed. Accordingly the hybrid approach is much faster than the strict gradient descent

and it is worthwhile to look for the possibility of decomposing in the manner of eq. (79). For type-1 ANFIS, this can be achieved if the membership function on the consequent part of each rule is replaced by a piecewise linear approximation with two consequent parameters fig. (110). In this case, again, the consequent parameters constitute set S_2 and the hybrid learning rule can be employed directly.

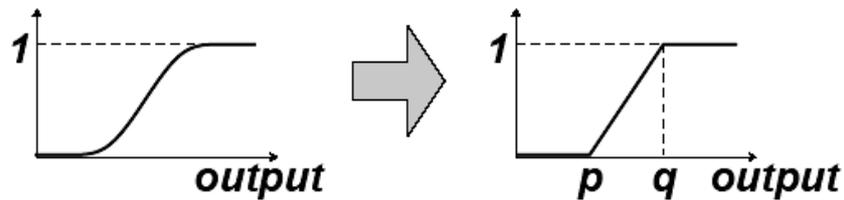


Figure 110. Piecewise Linear Approximation Of Membership Functions On The Consequent Part Of Type-1 ANFIS

However, it should be noted that the computation complexity of the least squares estimate is higher than that of the gradient descent. In fact, there are four methods to update the parameters, as listed below according to the computation complexities:

1. Gradient descent only: all parameters are updated by the gradient descent
2. Gradient descent and one pass of LSE: the LSE is applied only once at the very beginning to get the initial values of the consequent parameters and then the gradient descent takes over to update all parameters
3. Gradient descent and LSE: this is the proposed hybrid learning rule
4. Sequential LSE only: the ANFIS is linearized w.r.t. all parameters and the extended kalman filter algorithm is employed to update all parameters. This has been proposed in the neural network literature [93, 94].

The choice of above methods should be based on the trade-off between complexity and resulting performance.

-	Forward pass	Backward pass
premise parameters	fixed	gradient descent
consequent parameters	least squares estimate	fixed
signals	node outputs	error rates

Table 6. Two Passes In The Hybrid Learning Procedure Of ANFIS

Because the update formulas of the premise and consequent parameters are decoupled in the hybrid learning rule, as seen in Table 6, further speed up of learning is possible by using other versions of the gradient method on the premise parameters, such as conjugate gradient descent, second-order back-propagation [95], quick propagation [96], nonlinear optimization[97] and many others.

7.1.3.1.6 Application Of The Classifier On A Power System

The proposed classifier was tested on data that were obtained by utilizing the Series 5500 DualNode on the Public Power Constitution at Katsampas at Heraklion, Crete. The DataNode was connected at the capacitors' 13.8kV busbar of the facility, via the current and voltage pods, the set up of the InfoNode was made according to the IEEE standards, described earlier regarding all the disturbance events, the data were obtained by the use of the program Dranview and were finally processed on Matlab 6.5v. Unfortunately due to technical difficulties, regarding the time and duration of its use, the only disturbances that were recorded and afterwards processed involved swells, sags and transients. That is the reason why the experimental results, presented later on relate only to these three categories of power disturbance events.

As it was shown at the figures presented earlier, fig. (97-99), the waves that were received from the facility's busbar weren't so "pure" as those that were created at the laboratory. That was mainly because of the intense harmonic distortion of the signal, which may not exceeded the IEEE limits but however altered the signal, and moreover from the electromagnetic insertions of the facility's power instruments. That is the reason that during the training and the testing of the ANFIS network the following algorithm was chosen: for each group of distorted signals we created a loop. From the beginning till the end of that loop in each circle that was made one signal was set as the testing one while all the others were set as the training examples. That means that in each circle the signal that was previously set as testing example was now set as training example while the following signal that was previously set as training example was now set as testing and so on till the end of the signals. That algorithm led to the results presented on fig. (111-114) and Table 7.

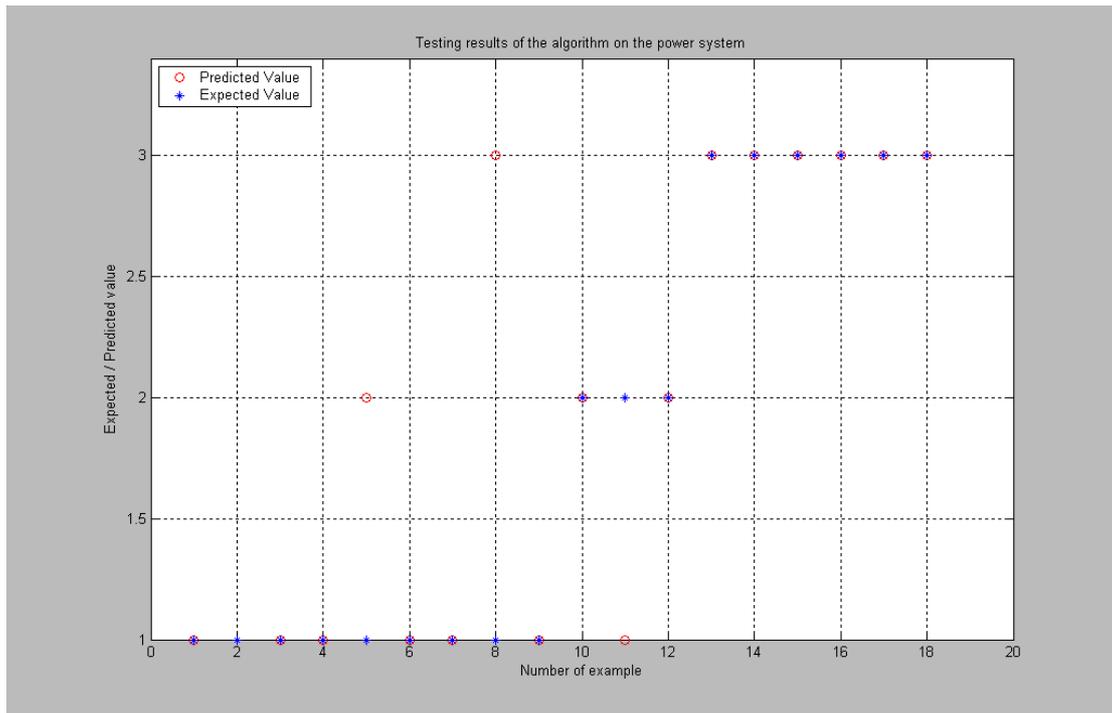


Figure 111. . Testing Results Of The ANFIS Algorithm On The Power System Using Only One Third Of The Training Examples

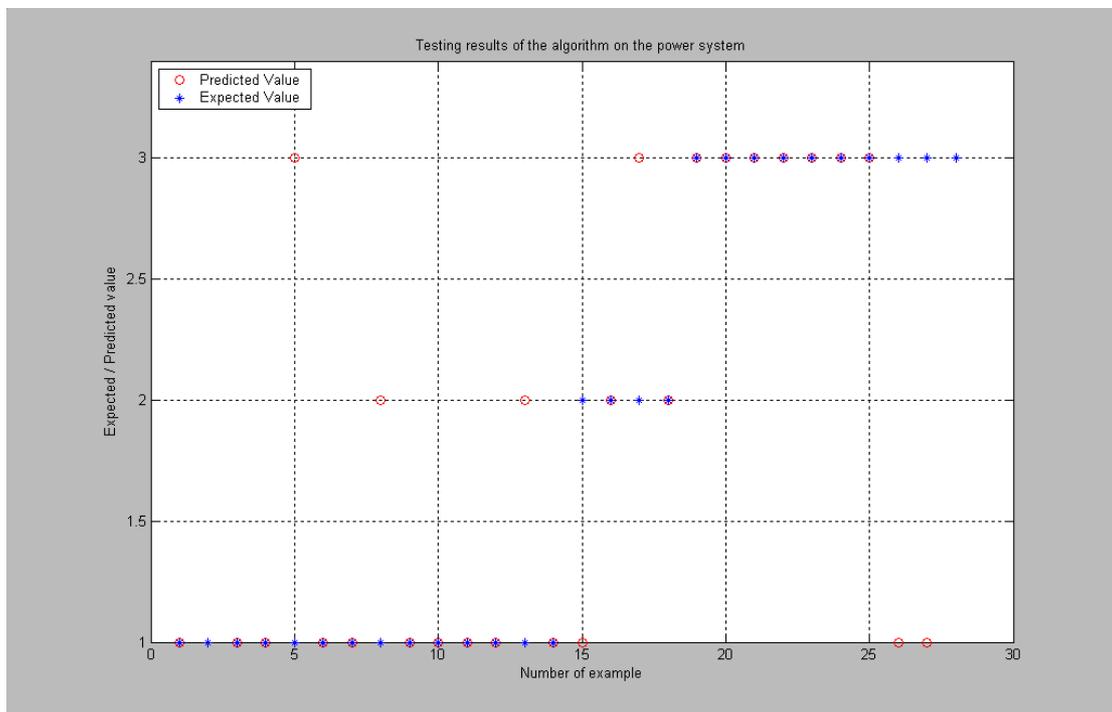


Figure 112. Testing Results Of The ANFIS Algorithm On The Power System Using Only Half Of The Training Examples

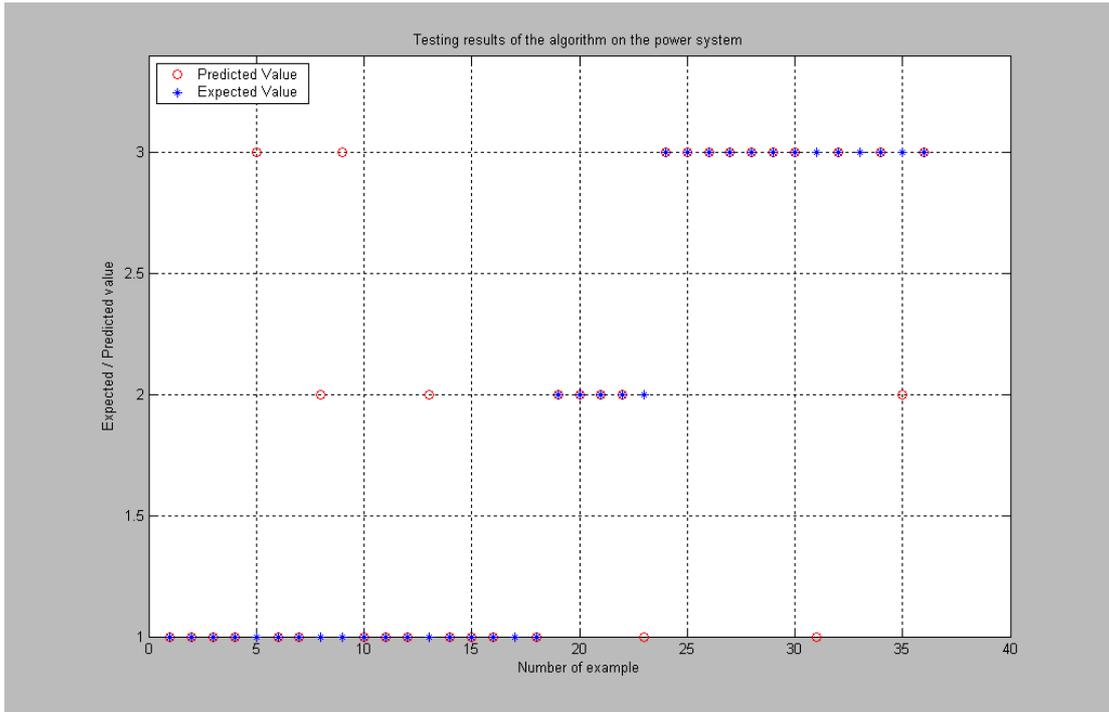


Figure 113. Testing Results Of The ANFIS Algorithm On The Power System Using Only Two Thirds Of The Training Examples

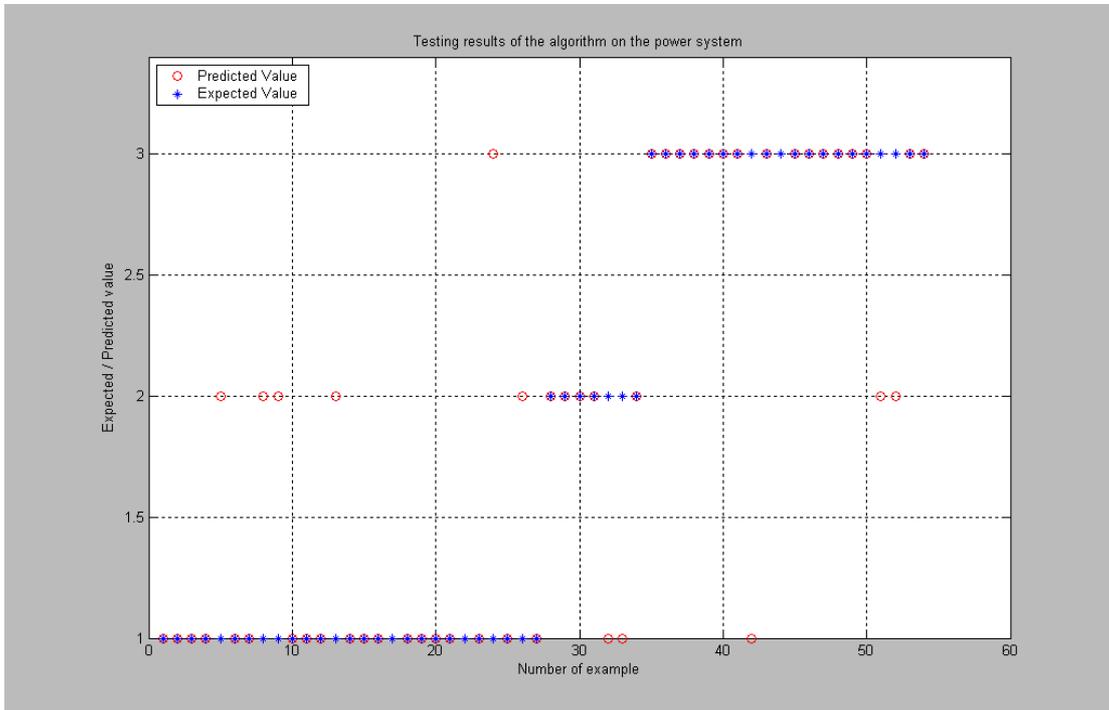


Figure 114. Testing Results Of The ANFIS Algorithm On The Power System Using All The Training Examples

<i>Number of training examples</i>	Transients	9	14	18	27
	Sags	3	4	5	7
	Swells	6	10	13	20
	Total	18	28	36	54
Number of testing examples		1	1	1	1
Learning time (sec.)		27.903	32.134	35.296	46.339
Classification time of a new signal (sec.)		0.0547	0.0413	0.0467	0.0489
Accuracy rate		77.7778%	67.8571%	75.0000%	74.0741%

Table 7. Test Results Of The ANFIS Algorithm On The Power System

As it's obvious from the above table the ANFIS classifier does not work as well as the PNN one. It can easily be seen that its accuracy rate does not rise as the number of the training samples grows. That happens, as it will be graphically presented later on in the next section, because the increase of the number of the training samples causes the increase of the complexity of the network's structure as well. That means that the network becomes mathematically "too heavy" and also easy to be dis-oriented. That is also why both learning and classification time rise as the number of the training samples grows.

8 Conclusions

After having read the theoretical background regarding the three above algorithms on power quality monitoring, applied them on the Public Power Constitution at Katsampas at Heraklion, Crete by utilizing the Series 5500 DualNode, in order to acquire the distorted signals, and implementing the program on a PC, by the use of Matlab the following conclusions came out.

8.1 Power Quality Event Detection Using Adaline

The simplicity introduced by this technique is due to the ease in calculations that facilitates its hardware implementation. The fact of being easy to execute makes the Adaline a very competitive choice for the algorithms currently used in power quality instrumentation.

The main advantage of the Adaline architecture is the ability of the Adaline to be trained on-line, eliminating the need for repetitive off-line training. Although most neural networks have the ability of on-line training, the Adaline is superior because of its simple structure and thus its high speed.

The Adaline succeed in fast and relatively accurate detection of the most common power quality disturbances. It was found that Adaline is sensitive for both number of delayed input and the value of learning rate. Small learning rate may lead to slow converge time, while large learning rate may cause Adaline to lose the ability of tracking power signal. In addition, large numbers of input may worsen the operation of Adaline. Typical values of learning rate and number of inputs were found to vary from 0.2 to 0.5 and from 4 to 6, respectively. The proposed algorithm was tested utilizing generated signals using Matlab as well as with actual bus industrial distribution system loaded with different types of linear and non-linear loads. In both cases, the Adaline yields a satisfactory detection result of different power quality events.

However, the Adaline may detect all kinds of power quality event disturbances but it can't classify them. In all the graphs that were presented, at the corresponding chapter, the Adaline, as it is easily noticed, indicates whenever there is a disturbance in the network but it doesn't clarify what kind of disturbance has occurred.

That yields the need of an algorithm that should both detect and classify a power quality disturbance at the network that is being monitored. That means that whenever there would be an anomaly in the signal, the algorithm should provide the

specific time of occurrence of the disturbance and moreover the category in which the disturbance lies within.

Two different algorithms that fulfill these conditions were presented later on at the following sections: “Power Quality Event Recognition and Classification Using a Wavelet-Based Neural Network” and “Power Quality Event Recognition and Classification Using a Wavelet-Based Adaptive Neuro-Fuzzy Inference System”.

8.2 Power Quality Event Recognition and Classification Using a Wavelet-Based Neural Network

This technique proposes a prototype of wavelet-based neural-network classifiers for power disturbance recognition and classification. That means that on the contrary to the Adaline algorithm this one fills in the lack of the classification problem. Though someone would believe that this advantage of the new algorithm would create a disadvantage, regarding its duration, experimental results showed the apposite! Thus not only did it not increase its duration but even more it can reduce as well. This happens because a completely different methodology is used which doesn't first predict the signal and then compares it with the measured one but instead uses the properties of the distorted ones as a comparison to decide whether the measured one is distorted as well or not.

The proposed method can reduce the quantity of extracted features of distorted signal without losing its property, thus requiring less memory space and computing time for proper classification of disturbance types. The experimental results showed that the proposed method has the ability of recognizing and classifying different power disturbance types efficiently, and it has the potential to enhance the performance of the power transient recorder with real-time processing capability.

Moreover, as it can easily be seen from Table 6 the proposed algorithm responds better as the number of the training examples increases, as it was expected. However, the learning time of the system as well as the classification time of a new signal also increases with the number of the training samples, as it was also expected. Thus, there is a fragile balance between the number of the samples that are selected to train the system and the time that is afterwards needed from the system to classify a new one. A suggestion of a way to partially resolve this problem would be to select a small number of representative samples of each category of distorted signals to be the training ones. In addition, whenever a new known distorted signal would come up

whose variation from the training group was significant then the algorithm by its own would list it to the training samples.

Comparing the algorithm's theoretical results, the results that came up when the classifier was given artificially distorted signals, created in the laboratory with a PC, with the real ones, the results that came up when the classifier was really tested on the Public Power Constitution, it can easily be said that the resulting deviation, regarding the efficiency of the algorithm, was not negligible. This leads us to the conclusion that the proposed algorithm has a drawback: the lack of robustness which is especially needed in combined events, which in our case is the most used scenario. This could be solved by the use of carefully designed filters, in order to isolate the distorted signal. Carefully though, because from one hand the signal should be "clean" from external interferences but on the other hand not too "clean" because there is the risk of influencing the distortion in such a way that it would be unnoticeable!

8.3 Power Quality Event Recognition and Classification Using a Wavelet-Based Adaptive Neuro-Fuzzy Inference System

This technique also proposes a prototype of wavelet-based neuro-fuzzy system for power disturbance recognition and classification. This algorithm works in a similar way to the previous one, thus also giving the advantage of classification, over against the Adaline algorithm.

The proposed method can also, as it was seen with the wavelet-based neural network algorithm, reduce the quantity of extracted features of distorted signal without losing its property, thus requiring less memory space and computing time for proper classification of disturbance types.

The motive for the creation of the specific algorithm was the lack of robustness of the wavelet-based neural method, which by definition of the fuzzy networks would be, if not extinguished, at least constrained. However, as the results on Table 7 showed, that wasn't confirmed, or at least not in all cases.

So, as it was proven the fuzzy algorithm, comparing to the neural algorithm, responded better, regarding the efficiency, but only where there weren't many training samples and that was unfortunately the only case where fuzzy networks showed their flexibility.

As the number of the samples increased though, the algorithm initially became unstable requiring more time to carry out its processes and moreover did not come out

more precise, as it was expected. The cause of this unbalance could be the complicated structure of the network that came with the increased number of inputs, due to the fact that the algorithm is dealing with combined events. In figs. 115-117 is presented the increase of the complication of the structure of an AFIS model as the number of its inputs grows. So, a proposed way to enhance the algorithm would be to “lighten” the network, by reducing its inputs or similarly to the previous case to filter the signals in such a way that the features of each category of the distorted signals would be more extensive. However, once again, attention should be paid in order to maintain the features of the distorted signals.

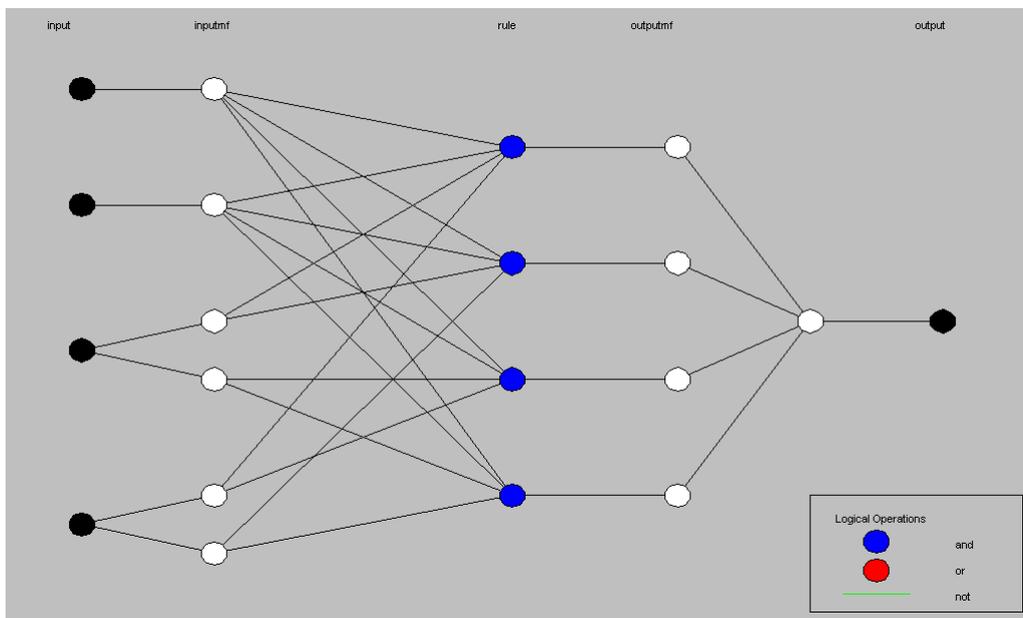


Figure 115. Example Of Anfis Model Structure With 4 Inputs

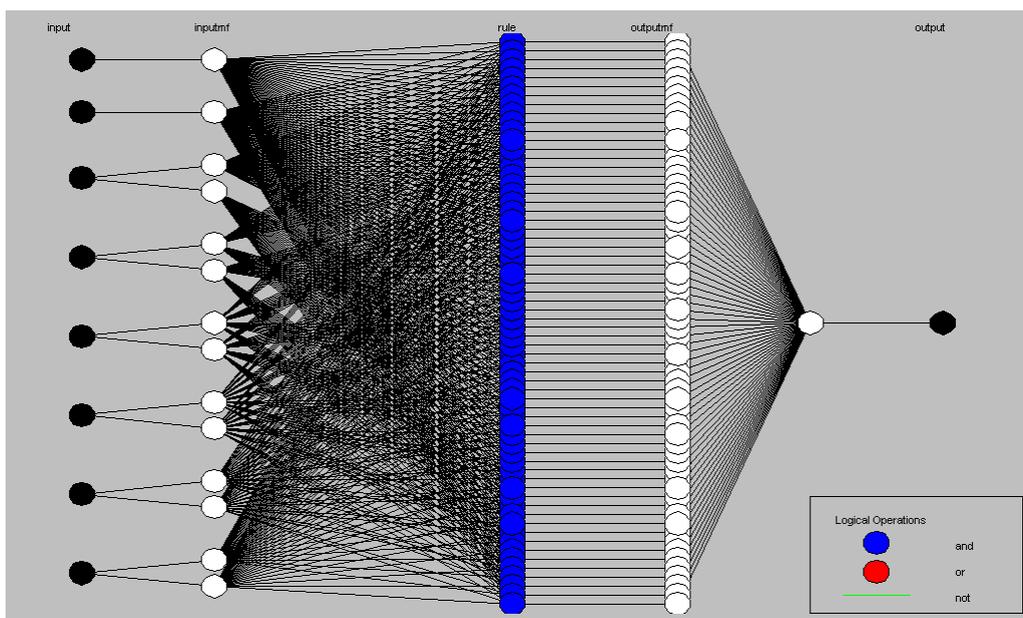


Figure 116. Example Of Anfis Model Structure With 8 Inputs

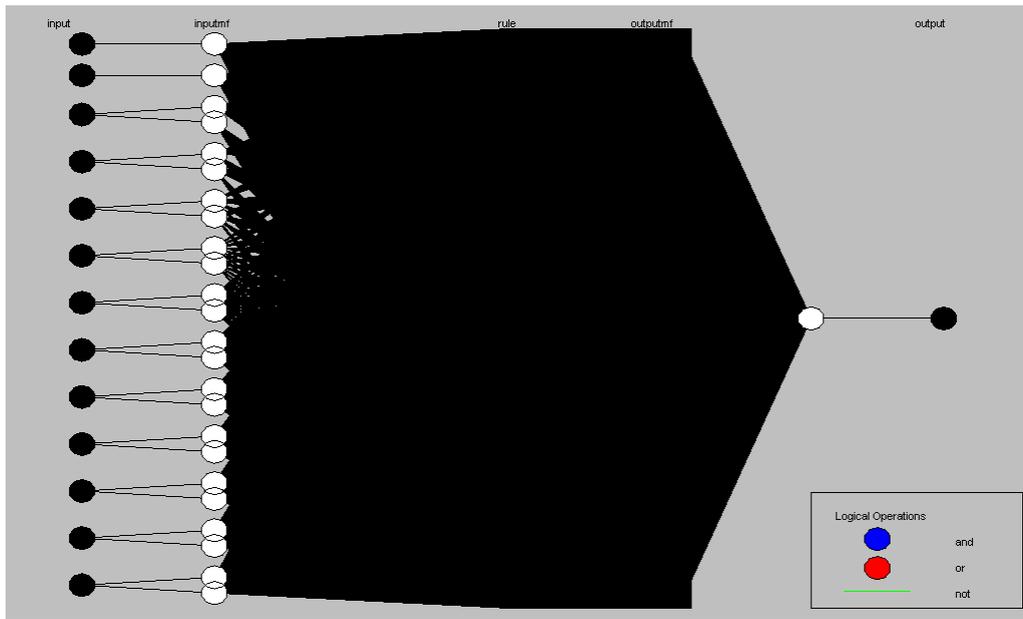


Figure 117. Structure Of The Anfis Model Used on This Paper, With 13 Inputs

To sum up, while one might expect that the ANFIS classification algorithm would behave better than the PNN classification algorithm, due to the increased robustness and flexibility that characterizes it, which by the way has been proven more than once in other applications, that didn't happen. Fig. 118 shows a graphical comparison between those two classifiers which denotes that for the specific application the ANFIS algorithm is not recommended.

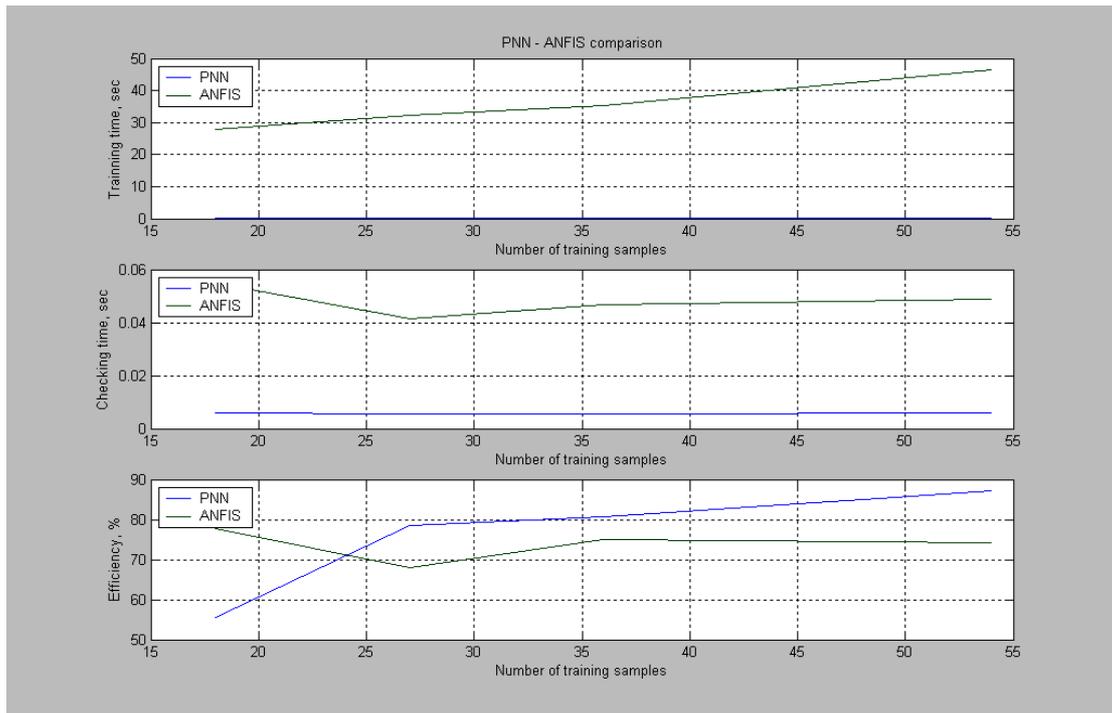


Figure 118. Comparison Between PNN – ANFIS Regarding The Training Time, Checking Time And The Efficiency

9 Future Work

Hereto, a state of the art techniques oriented to automatic classification of power systems events has been shown. More work has to be done to get a better understanding of voltages and currents behavior under different event conditions. Additionally, a better understanding of relations between voltage and currents in different voltage levels and its propagation in the network is still required.

Moreover, feature definition and extraction have to be enhanced in order to improve classification algorithms. As a result of this work, a universal set of accepted features for each type of event could be drawn. Although, the work done in the definition and proposition of classifiers based on power system techniques is still immature, statistical approach for classification algorithms showed interesting results.

Finally, advances in automatic classification in power quality events should influence or change the way that hardware for power quality is currently defined which is discussed in the following paragraphs.

The scheme of a power quality monitoring equipment hasn't changed during all these years. A basic power quality monitor could be defined as shown in fig. (119).

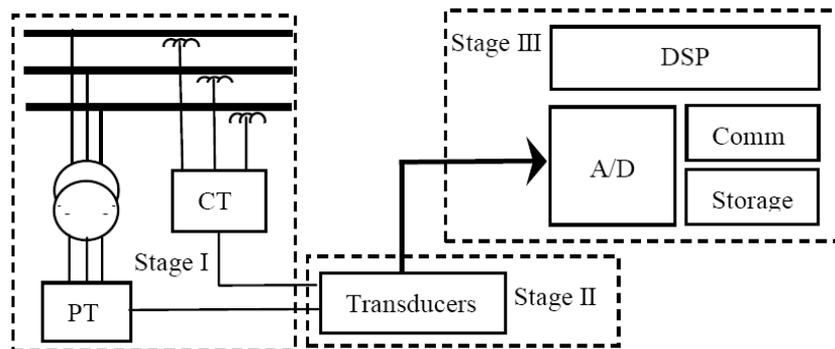


Figure 119. Power Quality Monitoring Equipment Scheme

In stage I, the electrical phenomenon occurs in the network (e.g. fault, capacitor switching), high voltage and current transformers (PT and CT) convert these three phase signals to an intermediate voltage or current level.

Then, in stage II, transducers or signal conditioning devices convert intermediate signals to low-level voltages.

Up to here, signals are still analog or in the continuous time domain. In stage III, signals are discretized by an analog to digital (A/D) converter. This device usually works with low voltage levels $\pm 2-10$ pp Vac and has a limited resolution in function of 2-b (LSB), where b is the number of bits to represent the signal. Nowadays, A/D

converters up to 24 bits, 8 channels and high sampling frequencies are available. At least 6 channels are required for power quality monitoring purposes in a 3-phase power system.

Once the signal is converted from the continuous time domain to the discrete time domain by a digital signal processor (DSP) or digital computer, the data obtained can be stored in a mass storage device or sent through any communication device to a central station.

Note that once data acquisition is made, signals are band limited, which means that the information contained in the signal is trustable until the Nyquist frequency (half the sampling frequency), e.g. if the sampling frequency is 3600 Hz, the higher frequency component of the signal is 1800 Hz. Higher frequency transients and traveling waves can hardly be studied from these signals.

On the other hand, hardware for power quality monitoring is not a problem anymore, unless, high frequency phenomenon studies were required. In this case, the scheme shown in fig. 119 will require some modifications. PT and CT transformers are working as low or high pass filters, mapping of its frequency response should be required.

Power quality equipments are still working as photographic cameras recording events sample per sample and producing enormous quantity of data. Nowadays, with the software and hardware technologies available, on-site processing of the information and analysis could be performed.

Instead of centralizing all the information events of a power network in a central station and store recordings sample by sample in huge databases, the analysis of the event could be distributed between all the monitors located in the nodes involved in the event. So that, this distributed system could give the diagnostics of the event or even better forecasting the next event. It's believed that the next generation of power quality monitoring equipment is becoming a reality.

10 REFERENCES

1. M.H.J. Bollen, What is power quality? , Electric Power Systems Research 66 (2003) 5_/14
2. IEEE Std 1159-1995, IEEE Recommended Practice for Monitoring Electric Power Quality
3. ANSI C84.1-1989, American National Standard for Electric Power Systems and Equipment – Voltage Ratings (60 Hz)
4. UIE-DWG-3-92-G, Guide to Quality of Electrical Supply for Industrial Installations, “Part 1: General Introduction to Electromagnetic Compatibility (EMC), Types of Disturbances and Relevant Standards” Advance UIE Edition, Disturbances Working Group GT 2, 1994.
5. IEC Technical Committee 77, Working Group 6 (Secretariat) 110-R5, Classification of Electromagnetic Environments. Draft, January 1991
6. ANSI C84.1-1989, American National Standard for Electric Power Systems and Equipment – Voltage Ratings (60 Hz)
7. IEEE Std 519-1992, IEEE Recommended Practices and Requirements for Harmonic Control in Electric Power Systems (ANSI)
8. IEC Technical Committee 77, Working Group 6 (Secretariat) 110-R5, Classification of Electromagnetic Environments. Draft, January 1991
9. UIE-DWG-2-92-D, UIE Guide to Measurements of Voltage Dips and Short Interruptions Occurring in Industrial Installations, October 1993
10. UIE-DWG-3-92-G, Guide to Quality of Electrical Supply for Industrial Installations, “Part 1: General Introduction to Electromagnetic Compatibility (EMC), Types of Disturbances and Relevant Standards”, Advance UIE Edition, Disturbances Working Group GT 2, 1994
11. IEEE Surge Protection Standards Collection (C62), 1992 Edition

12. IEC 1000-3-3 (1994) Electromagnetic compatibility (EMC)-Part 3: Limits. Section 3: Limitation of voltage fluctuations and flicker in low-voltage supply systems for equipment with rated current ≤ 16 A.
13. IEEE Std 141-1993, IEEE Recommended Practice for Electric Power Distribution for Industrial Plants (Red Book) (ANSI).
14. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=26>
15. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=23>
16. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=22>
17. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=14>
18. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=13>
19. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=41>
20. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=51>
21. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=50>
22. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=42>
23. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=43>
24. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=1>
25. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=48>
26. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=47>
27. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=3>
28. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=2>
29. <http://www.dranetz-bmi.com/products/prodspec.cfm?prod=31>
30. <http://www.admmicro.com/adm-3311.asp>
31. http://64.37.103.99/cgi-bin/pdc/viewprod.cgi?pid=1890&tid=new_e&type=elec
32. <http://www.amprobe.com/cgi-bin/pdc/viewprod.cgi?pid=1852&tid=2&type=hvac#features>
33. <http://www.instruments4hire.co.uk/amprobe-skylab-three-phase-power-analyser.htm>
34. <http://www.edmi.com.au/Products/PowerQuality/mk3.cfm?Menu=Products>
35. <http://www.edmi.com.au/Products/PowerQuality/genius.cfm?Menu=Products>
36. <http://www.edmi.com.au/Products/PowerQuality/mk6cd.cfm?Menu=Products>

37. <http://www.edmi.com.au/Products/PowerQuality/psqm.cfm?Menu=Products>
38. <http://www.elcomponent.co.uk/products/product.cfm?productid=43>
39. <http://www.elcomponent.co.uk/products/product.cfm?productid=32>
40. http://www.bis.fm/bis/products/Dent_Elite_Pro.asp
41. <http://www.enetics.com/prod-LM5400.html>
42. http://www.instruments4hire.co.uk/fluke_43b.htm#
43. <http://www.testequity.com/Specifications/flk41b.phtml>
44. <http://www.powermonitors.com/iVS-3-600E.htm>
45. <http://www.specs.com/products/stm/STM-scanner.htm>
46. <http://www.powermonitors.com/VP2.asp>
47. <http://www.powermonitors.com/VS-3S-Series.htm>
48. http://www.instruments4hire.co.uk/reliable_power_meter.htm
49. Afroz Khan, "Monitoring Power for the future". Power Engineering Review Journal April 2001
50. Mark McGranaghan, "Trends in Power Quality Monitoring". IEEE Power Engineering Review, October 2001
51. David Hart, David Uy, Damir Novosel, Steven Kunsman, Carl LaPlace and Marco Tellarini, "Improving Power Quality". ABB Review, 4/2000
52. Emmanouil Styvaktakis, Math Bollen and Irene Gu, "Automatic Classification of Power System Events Using RMS Voltage Measurements". Submitted for publication in Transactions on Power Delivery
53. Adly Girgis, Bi Chang and Elham Makram, "A Digital Recursive Measurement for On Line Tracking of Power Systems Harmonics". IEEE transactions on Power Delivery, vol 6, N° 3 July 1991
54. P. Dash, A. Pradham and G. Panda, "Frequency Estimation of Distorted Power System Signals Using Extended Complex Kalman Filter". IEEE Transactions on Power Delivery, vol 14, N° 3, July 1999
55. Haili Ma and Adly Girgis. "Identification and Tracking of Harmonics Sources in a Power System Using Kalman Filter". IEEE Transactions on Power Delivery, vol 11, N° 3, July 1996

56. P. Ribeiro, "Wavelet Transform: An Advanced Tool for Analyzing Non Stationary Harmonic Distortion in Power Systems". Proceeding IEEE International Conference On Harmonics in Power Systems, 1994
57. Tongxin Zheng, Elham Makran and Adly Girgis. "Power System Transient and Harmonic Studies Using Wavelet Transform". IEEE Transactions on Power Delivery, vol 14, N° 4, October 1999
58. Gerald Heydt and A. Galli, "Transient Power Quality Problems Analyzed using Wavelets". IEEE Transactions on Power Delivery, vol 12, N° 2, April 1997
59. Shyh Huang, Cheng Hsieh and Ching Huang. "Application of Morlet Wavelets to Supervise Power System Disturbances". IEEE Transactions on Power Delivery, vol 14, N° 1, January 1999
60. O. Poisson, P. Rioual and M. Meunier, "New signal Processing Tools Applied to Power Quality Analysis". IEEE Transactions on Power Delivery, vol 14, N° 2, April 1999
61. Irene Gu, Math Bollen, "Time Frequency and Timescale domain analysis of Voltage Disturbances". IEEE Transactions on Power Delivery, vol 15, N° 4 October 2000
62. Surya Santoso, Jeff Lamoree, Mack Grady. Edward Powers and Siddharth. Bhatt. "A Scalable PQ Event Identification System". IEEE Transactions on Power Delivery, vol 15, N° 2, April 2000
63. Emmanouil Styvaktakis, Math Bollen and Irene Gu, "Expert System for Voltage Dip Classification And Analysis". Power Engineering Society Summer Meeting, 2001, vol 1 , 2001
64. A. Gaouda M. Salama, M. Sultan and A.Chikhani, "Power Quality Detection and Classification using Wavelet Multi resolution Signal Decomposition", IEEE Transactions on Power Delivery, vol 14, N° 4, October 1999
65. Surya Santoso, Peter Hoffman, "Power Quality Assessment Via Wavelet Transform Analysis". IEEE Transactions on Power Delivery, vol 11, N° 2, April 1996
66. Chen Xiangxun, "Wavelet Based Detection, Localization, Quantification and Classification of Short Duration Power Quality Distur-

- bances". Power Engineering Society Winter Meeting, 2002, vol 2, 2002
67. Masoud Karami, Hossein Mokthari and Reza Iravani, "Wavelet Based On Line Disturbance Detection for Power Quality Applications". IEEE Transactions on Power Delivery, vol 15, N° 4 October 2000
 68. Jaehak Chung, Edward Powers, Mack Grady and Siddharth Bhatt. "Power Disturbance Classifier Using a Rule-Based Method and Wavelet Packet-Based Hidden Markov Model". Transactions on Power Delivery, vol 17, N° 1, January 2002
 69. S. Vasilic, M. Kezumovic. "An Improved Neural Network Algorithm for Classifying The Transmission Line Faults". Power Engineering Society Winter Meeting, 2002, vol 2, 2002
 70. Widrow B. and S. D. Sterns, Adaptive Signal Processing, New York: Prentice-Hall 1985
 71. Hagan, M. T., H. B. Demuth, and M. H. Beale, Neural Network Design, Boston, MA: PWS Publishing, 1996
 72. Widrow, B., and S. D. Sterns, Adaptive Signal Processing, New York: Prentice-Hall, 1985
 73. Chen, S., C.F.N. Cowan, and P. M. Grant, "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks," IEEE Transactions on Neural Networks, vol. 2, no. 2, March 1991, pp. 302-309
 74. P.D. Wasserman, Advanced Methods in Neural Computing, New York: Van Nostrand Reinhold, 1993, pp. 35-55
 75. Howard Demuth, Mark Beale, Neural Network Toolbox for use with Matlab, Version 4, 2002, pp. 247-249
 76. A.K. Ghosh, D.L. Lubkeman, The classification of power system disturbance waveforms using a neural network approach, IEEE Trans. PWRD 10 (1) (1995) 109_115
 77. L.C. Min, L. Qing, An enhanced adaptive neural network control scheme for power systems, IEEE Trans. Energy Convers. 12 (2) (1997) 166_174

78. P.K. Dash, D.P. Swain, A. Routray, A.C. Liew, Harmonic estimation in a power system using adaptive perceptrons, *IEE Proc. A.C. Gen. Trans. Dist.* 143 (6) (1996) 565_574
79. P.K. Dash, D.P. Swain, A. Routray, A.C. Liew, An adaptive neural network approach for the estimation of power system frequency, *J. Electr. Power Sys. Res.* 41 (3) (1997) 203_210
80. P.K. Dash, S. Mishra, M.M.A. Salama, A.C. Liew, Classification of power system disturbances using a fuzzy expert system and a Fourier linear combiner, *IEEE Trans. PWRD* 15 (2) (2000) 472_477
81. B. Widrow, M.E. Hoff, Adaptive switching circuits, WESTCON Convention, Record Part IV, 1960, pp. 96_104
82. B. Widrow, S. Stearns, Adaptive Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1985
83. B. Widrow, R. Winter, Neural networks for adaptive filtering and adaptive pattern recognition, *IEEE Comp.* 1988 (1988) 25_39
84. E.F. El-Saadany, T.K. Abdel-Galil, M.M.A. Salama, Application of wavelet transform for assessing power quality in medium voltage industrial distribution system, *IEEE TD Proc.* 1 (2001) 554_560
85. S. Santoso, E. J. Powers, W. M. Grady, and P. Hofmann, "Power quality assessment via wavelet transform analysis," *IEEE Trans. Power Delivery*, vol. 11, pp. 924–930, Apr. 1996.
86. D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol., no. 1, pp. 109–118, 1990.
87. T. Takagi and M. Sugeno. Derivation of fuzzy control rules from human operator's controls actions. Proc. Of the IFAC Symp. on Fuzzy Information, Knowledge representation and Decision analysis, Pages 55-60, July 1983
88. J.-S. Roger Jang. Fuzzy modelling using generalised neural networks and Kalman filter algorithm. In Proc. Of the Ninth National Conference on Artificial Intelligence (AAAI-91), pages 762-767, July 1991
89. K. J. Astrom and B. Wittenmark. Computer Controller Systems: Theory and Design. Prentice-Hall, Inc., N.J., 1984

90. G. C. Goodwin and K. S. Sin. Adaptive filtering prediction and control. Prentice-Hall, Englewood Cliffs, N.J., 1984
91. L. Ljung. System identification: theory for the user. Prentice-Hall, Englewood Cliffs, N.J., 1987
92. P. Strobach. Linear prediction theory: a mathematical basis for adaptive systems. Springer-Verlag 1990
93. S. Shar and F. Palmieri. MEKA-a fast, local algorithm for training feedforward neural networks. In Proc. Of International Joint Conference on Neural Networks, pages III 41-46, 1990
94. S. Singhal and L. Wu. Training multilayer perceptrons with the extended kalman algorithm. In David S. Touretzky, editor, Advances in neural information processing systems I, pages 133-140. Morgan Kaufman Publishers, 1989
95. D. B. Parker. Optimal algorithms for adaptive networks: Second order back-propagation, second order direct propagation, and second order Hebbian learning. In *Proc. Of IEEE International Conference on Neural Networks*, pages 593-600, 1987
96. S. E. Fahlman. Faster-learning variations on back-propagation: an empirical Study. In D. Touretzky, G. Hinton and T. Sejnowski, editors, *Proc. Of the 1988 Connectionist Models Summer School*, pages 38-51, Carnegie Mellon University, 1988
97. R. L. Watrous. Learning algorithms for connectionist network: applied gradient methods for nonlinear optimization. In Proc. Of IEEE International Conference on Neural Networks, pages 619-627, 1991
98. L.A. Zadenh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. On Systems, Man, and Cybernetics*, 3(1): 28-44, January 1973.
99. T.Takagi and M.Sugeno. Derivation of fuzzy control rules from human operator's control actions. Proc, of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis, pages 55-60, July 1983.
100. C.-C. Lee. Fuzzy Logic in Control Systems: Fuzzy Logic Controller-part-1. *IEEE Trans. On Systems, Man, and Cybernetics*, 20(2):404-418, 1990.

101. C.-C. Lee. Fuzzy Logic in Control Systems: Fuzzy Logic Controller-part-2. IEEE Trans. On Systems, Man, and Cybernetics, 20(2):419-435, 1990.
102. Y. Tsukamoto. An approach to fuzzy reasoning method. In Madan M. Gupta, Rammohan K. Ragade, and Ronald R. Yager, editors, Advances in Fuzzy Set Theory and Applications, pages 137-149. North-Holland, Amsterdam, 1979.
103. T. Takagi and M. Sugeno. Derivation of fuzzy control rules from human operator's control actions. Proc. Of the IFAC Symp. On Fuzzy Information, Knowledge Representation and Decision Analysis, pages 55-60, July 1983.
104. P. Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. PhD Thesis, Harvard University, 1974
105. J.-S. Roger Jang. Fuzzy modelling using generalized neural networks and Kalman filter algorithm. In Proc. Of the Ninth National Conference on Artificial Intelligence (AAAI-91), pages 762-767, July 1991
106. J.-S. Roger Jang. Rule extraction using generalized neural networks. In Proc. Of the 4th IFSA World Congress, July 1991.
107. D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. In D.E. Rumelhart and James L. McClelland, editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, volume 1, chapter 8, pages 318-362. The MIT Press, 1986.

11 INDEX

<u>1</u>	<u>INTRODUCTION.....</u>	<u>1</u>
<u>2</u>	<u>POWER SYSTEMS IN GENERAL.....</u>	<u>4</u>
<u>3</u>	<u>POWER QUALITY</u>	<u>7</u>
3.1	ELECTROMAGNETIC COMPATIBILITY	7
3.2	GENERAL CLASSIFICATION OF PHENOMENA.....	7
3.3	DETAILED DESCRIPTIONS OF PHENOMENA.....	11
3.3.1	TRANSIENTS	11
3.3.1.1	Impulsive Transient.....	12
3.3.1.2	Oscillatory Transient.....	13
3.3.2	SHORT-DURATION VARIATIONS	15
3.3.2.1	Interruption.....	16
3.3.2.2	Sags (Dips).....	17
3.3.2.3	Swells	20
3.3.3	LONG DURATION VARIATIONS	21
3.3.3.1	Overvoltage.....	21
3.3.3.2	Undervoltage.....	21
3.3.3.3	Sustained Interruptions.....	22
3.3.4	VOLTAGE IMBALANCE	22
3.3.5	WAVEFORM DISTORTION	23
3.3.5.1	DC Offset	24
3.3.5.2	Harmonics	24
3.3.5.3	Interharmonics.....	25
3.3.5.4	Notching.....	26
3.3.5.5	Noise	26
3.3.6	VOLTAGE FLUCTUATIONS.....	27
3.3.7	POWER FREQUENCY VARIATIONS.....	28
<u>4</u>	<u>MONITORING OBJECTIVES.....</u>	<u>30</u>
4.1	NEED FOR MONITORING POWER QUALITY.....	30
4.2	EQUIPMENT TOLERANCES AND EFFECTS OF DISTURBANCES ON EQUIPMENT	31

4.3	EQUIPMENT TYPES	31
4.4	EFFECT ON EQUIPMENT BY PHENOMENA TYPE	32
4.4.1	TRANSIENTS	32
4.4.2	SHORT DURATION VARIATIONS	32
4.4.2.1	Interruptions	32
4.4.2.2	Sags	33
4.4.2.3	Swells	33
4.4.3	LONG DURATION VARIATIONS	34
4.4.3.1	Sustained Interruptions.....	34
4.4.3.2	Undervoltages	34
4.4.3.3	Overvoltages	34
4.4.4	VOLTAGE IMBALANCE	35
4.4.5	WAVEFORM DISTORTION	35
4.4.6	VOLTAGE FLUCTUATIONS	36
4.4.7	POWER-FREQUENCY VARIATIONS	36
5	<u>MEASUREMENT INSTRUMENTS</u>	37
5.1	HANDHELD DEVICES	38
5.1.1	ANALYST 2060	38
5.1.2	ANALYST Q70	39
5.1.3	POWERXPLOER PX5-400	40
5.1.4	POWERGUIDE 4400	42
5.1.5	POWER PLATFORM PP-4300.....	44
5.2	PORTABLE DEVICES	46
5.2.1	658 POWER QUALITY ANALYZER	46
5.2.2	POWER PLATFORM PP1	47
5.3	FIXED SYSTEMS.....	48
5.3.1	3100 PQ PAGER.....	48
5.3.2	7100/7100S	49
5.4	SERIES 5500 DUALNODE	51
5.4.1	SIGNATURE SYSTEM.....	51
5.4.2	INFO NODE: THE CENTRAL COMPONENT.....	52
5.4.3	SIGNATURE SYSTEM INFO NODE GRAPHICAL USER INTERFACE.....	53
5.4.4	THE EPQ DATA NODE SERIES	55
5.4.5	SERIES 5500 DUALNODE: THE ABSOLUTE COMBINATION	55

6 CLASSIFICATION METHODS..... 61

6.1 STATE OF THE ART 62

6.1.1 STATE OF THE ART ON THE CLASSIFICATION OF POWER QUALITY EVENTS 63

6.1.2 BASIC TOOLS FOR SIGNAL COMPONENTS ESTIMATION..... 64

6.1.3 MORE ADVANCE TOOLS FOR SIGNAL COMPONENTS ESTIMATION 65

6.2 MOST USED TECHNIQUES 65

6.2.1 NEURAL NETWORKS 65

6.2.1.1 Adaptive Linear Neuron Networks (ADALINE)..... 66

6.2.1.1.1 Linear Neuron Model..... 66

6.2.1.1.2 Adaptive Linear Network Architecture..... 67

6.2.1.1.3 Single ADALINE..... 68

6.2.1.1.4 Mean Square Error 69

6.2.1.1.5 LMS Algorithm..... 70

6.2.1.1.6 Adaptive Filtering 70

6.2.1.1.7 Multiple Neuron Adaptive Filters 71

6.2.1.1.8 Summary 73

6.2.1.2 Probabilistic Neural Networks (PNN)..... 74

6.2.1.2.1 Radial Basis Networks 74

6.2.1.2.2 Probabilistic Neural Networks (PNN)..... 76

6.2.1.2.3 Summary 77

6.2.2 FUZZY NETWORKS 78

6.2.2.1 Adaptive Neuro – Fuzzy Inference System ANFIS 78

6.2.2.1.1 FIS Structure and Parameter Adjustment..... 78

6.2.2.1.2 Familiarity Breeds Validation; Knowledge of the Data..... 79

6.2.2.1.3 Constraints of ANFIS..... 80

6.2.3 WAVELET TRANSFORM 80

6.2.3.1 Computation Of The Cwt..... 87

6.2.3.2 Time And Frequency Resolutions..... 94

6.2.3.3 The Wavelet Theory: A Mathematical Approach 95

6.2.3.4 Basis Vectors..... 96

6.2.3.5 Inner Products, Orthogonality, and Orthonormality 97

6.2.3.6 Examples..... 99

6.2.3.7 The Wavelet Synthesis..... 101

6.2.3.8 Discretization of the Continuous Wavelet Transform: The Wavelet Series 102

6.2.3.9 Discrete Wavelet Transform 105

6.2.3.10 The Subband Coding and The Multiresolution Analysis 106

7 APPLICATION AND RESULTS..... 116

7.1.1 POWER QUALITY EVENT DETECTION USING ADALINE 117

7.1.1.1 Adaline architecture 117

7.1.1.2 Adaline training..... 118

7.1.1.3 Power Quality Event Detection Using Adaline..... 119

7.1.1.4 Numerical examples..... 120

7.1.1.4.1 Voltage sag..... 121

7.1.1.4.2 Voltage swell..... 121

7.1.1.4.3 Momentary interruptions..... 122

7.1.1.4.4 Harmonics 123

7.1.1.4.5 Transient..... 124

7.1.1.4.6 Combined Events 125

7.1.1.5 Sensitivity analysis..... 126

7.1.1.5.1 Effect of learning rate (h)..... 126

7.1.1.5.2 Effect of choosing P..... 127

7.1.1.6 Testing..... 128

7.1.1.6.1 Sag detection in harmonic contaminated signal..... 129

7.1.1.6.2 Swell detection in harmonic contaminated signal..... 129

7.1.1.6.3 Transient detection in harmonic contaminated signal..... 130

7.1.2 POWER QUALITY EVENT RECOGNITION AND CLASSIFICATION USING A WAVELET-BASED NEURAL NETWORK..... 131

7.1.2.1 Wavelet Transform..... 132

7.1.2.1.1 Multiresolution Analysis (MRA) and Decomposition 132

7.1.2.1.2 Mathematical Model of DWT 132

7.1.2.2 Parseval’s Theorem In DWT Application..... 134

7.1.2.3 Feature Extraction, Recognition And Classification..... 135

7.1.2.3.1 Feature Extraction 135

7.1.2.3.2 Probabilistic Neural Network (PNN)..... 139

7.1.2.3.3 Classification of Transient Signals Using PNN Model..... 141

7.1.2.4 Applications And Results..... 142

7.1.2.4.1 Testing of The Classifier By Laboratory Simulation..... 142

7.1.2.4.2 Application Of The Classifier On A Power System..... 150

7.1.3 POWER QUALITY EVENT RECOGNITION AND CLASSIFICATION USING A WAVELET-BASED ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM 157

7.1.3.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)..... 157

7.1.3.1.1 Fuzzy If-Then Rules And Fuzzy Inference Systems..... 158

7.1.3.1.1.1	Fuzzy If-Then Rules.....	158
7.1.3.1.1.2	Fuzzy Inference systems	158
7.1.3.1.2	Adaptive Networks: Architectures And Learning Algorithms.....	161
7.1.3.1.2.1	Architecture And Basic Learning Rule	161
7.1.3.1.3	Hybrid Learning Rule: Batch (Off-Line) Learning.....	164
7.1.3.1.4	Hybrid Learning Rule: Pattern (On-Line) Learning.....	166
7.1.3.1.5	ANFIS: Adaptive-Network-Based Fuzzy Inference System.....	167
7.1.3.1.5.1	Adaptive Neuro-Fuzzy Inference System (ANFIS) Architecture	167
7.1.3.1.5.2	Hybrid Learning Algorithm	171
7.1.3.1.6	Application Of The Classifier On A Power System.....	173
8	<u>CONCLUSIONS</u>	<u>177</u>
8.1	POWER QUALITY EVENT DETECTION USING ADALINE.....	177
8.2	POWER QUALITY EVENT RECOGNITION AND CLASSIFICATION USING A WAVELET-BASED NEURAL NETWORK.....	178
8.3	POWER QUALITY EVENT RECOGNITION AND CLASSIFICATION USING A WAVELET-BASED ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM	179
9	<u>FUTURE WORK</u>	<u>183</u>
10	<u>REFERENCES.....</u>	<u>185</u>
11	<u>INDEX.....</u>	<u>193</u>