國立交通大學

多媒體工程研究所

碩 士 論 文

基於小波混沌分析法之癲癇預測及電路實現

Wavelet-Based Chaos Analysis for Epileptic Seizure Prediction

and Circuit Implementation

研 究 生：王舒愷

指導教授：林進燈　教授
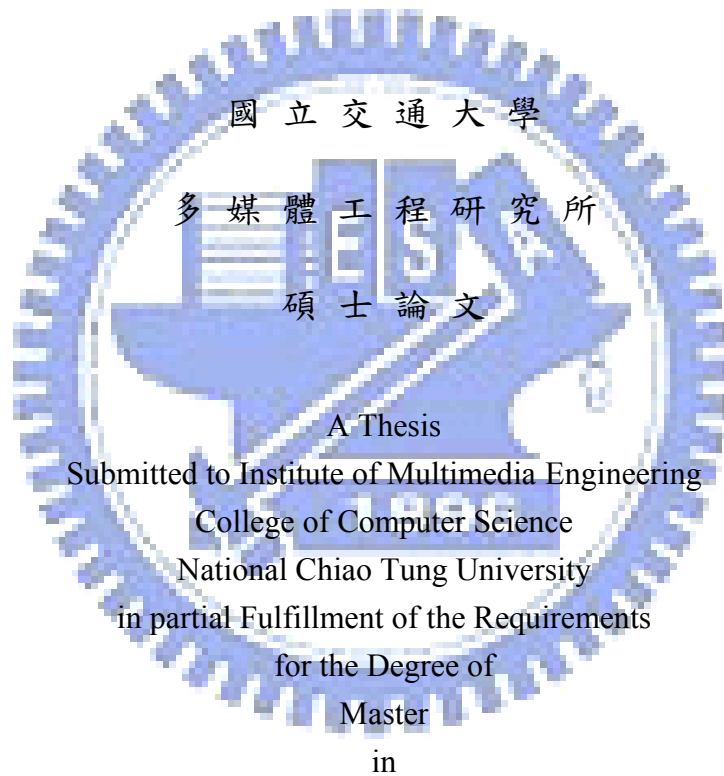
中 華 民 國 九 十 八 年 一 月

基於小波混沌分析法之癲癇預測及電路實現

Wavelet-Based Chaos Analysis for Epileptic Seizure Prediction and Circuit Implementation

研 究 生：王舒愷　　　　　Student：Shu-Kai Wang

指導教授：林進燈　　　　　Advisor：Chin-Teng Lin

國 立 交 通 大 學

多 媒 體 工 程 研 究 所

碩 士 論 文

A Thesis
Submitted to Institute of Multimedia Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer Science

January 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年一月

基於小波混沌分析法之癲癇預測及電路實現

學生：王舒愷　　　　　　　指導教授：林進燈 博士

國立交通大學多媒體工程研究所

中文摘要

在大腦生理訊號分析的研究中，如何從長時間的腦電波訊號找出可靠的特徵來進行癲癇疾病的預測是目前熱門的研究議題，有鑑於傳統的統計分析方法對於非穩定、非線性動態系統的訊號，容易因錯誤的結論而影響預測準確性。

本論文提出一個基於小波及混沌理論分析的架構包含離散小波轉換、相關維度、及相關係數。因小波具有多解析度、時頻分析的特性，相對於傅立葉轉換更適合用在非穩定訊號。而混沌理論對於非穩定、非線性動態系統的基本推論，比傳統的統計學方法能更有效地重建腦波的特徵，有助於腦波訊號的分析。因此，結合小波與混沌理論將可有效提高癲癇預測的準確性及降低誤判率。

本論文提出的作法是先將訊號分解成若干子頻帶，並針對子頻帶，利用腦電波在癲癇發作前後，其相關維度收歛速度的不同來做為預測的依據。經實驗結果，本論文所提出的演算法在 11 位具癲癇病患的測試中，可達到 87%的預測準確率，誤判率為 0.24 次/小時，其平均預測時間約在發病前 27 分鐘。

為了能讓本論文提出的方法能應用於可攜式生理監控設備，我們將其設計成癲癇分析的處理電路，並針對演算法特性，使用提升式小波轉換、改良記憶體定址、及其他算數化簡來降低電路面積及功耗，也可於將來進一步整合到生醫的訊號處理器中。

Wavelet-Based Chaos Analysis for Epileptic Seizure Prediction and Circuit Implementation

Student：Shu-Kai Wang　　　　Advisor：Dr. Chin-Teng Lin

Institute of Multimedia Engineering, College of Computer Science

National Chiao-Tung University

Abstract

The Epilepsy and epileptic seizure prediction algorithm by extracting useful features from Electroencephalography (EEG) is a hot topic in the current research of physiological signals. In view of the erroneous conclusions from the traditional statistical analysis methods for non-stationary and non-linear dynamics system of signals may affect the accuracy of forecasts.

This thesis presents a novel architecture based on wavelet and chaos theory, including Discrete Wavelet Transform (DWT), correlation dimension, and correlation coefficient. The wavelet transform is more suitable for non-stationary signals than Fast Fourier Transform (FFT) due to the ability of multi-resolution and time-frequency analysis. The fundamentals of Chaos theory for non-stationary and non-linear dynamics systems are more in line with the characteristics of brain waves than statistics. Therefore, combining DWT and Chaos analysis can achieve a high prediction rate.

In this thesis, first EEG signals are decomposed into several subbands. We predict seizures by the difference of convergent radius between the correlation dimension of EEG before a seizure and the one during a seizure for each subband.

The proposed algorithm is evaluated with intracranial EEG recordings from a set of eleven patients with refractory temporal lobe epilepsy. In the experimental results, the algorithm with global settings for all patients predicted 87% of seizures with a false prediction rate of 0.24/h. Seizure warnings occur about 27 min ahead the ictal on average.

To apply the algorithm proposed to a portable physiological monitoring device, a seizure analysis circuit is also designed. Some techniques, such as lifting wavelet transform, enhanced memory addressing, and arithmetic reduction etc., are used to reduce circuit area and power consumption of circuit. In the future, the seizure analysis circuit can be further integrated into a digital signal processor for biomedical applications.
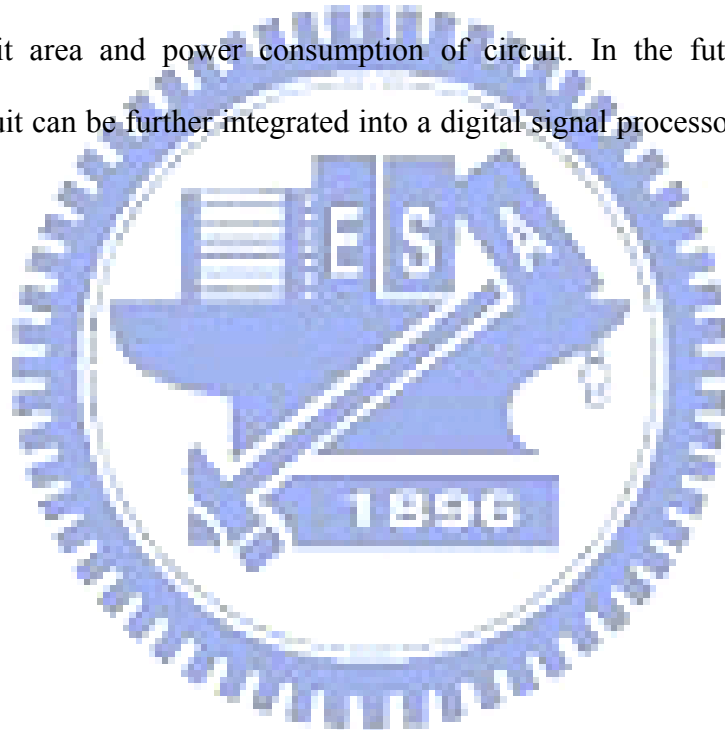
誌 謝

本篇論文研究能夠順利完成，要感謝許多人的不管是課業上或精神上的鼓勵及幫助，讓我在這兩年中獲得專業知識及研究方法。

首先要感謝的就是我的指導教授-林進燈老師。老師在許多領域都有優異的研究成果，在國內也是頗負盛名，給予我們豐富的學術資源，以及完善的研究環境，更指導我們面對研究時的態度應該小心謹慎。

另外，也非常感謝帶領指導我們的范倫達教授以及鍾仁峰學長，在畢業方向的討論上，提供許多重要的意見及建議，在研究過程中，對於學生的疑問或遇到的問題，總能很有耐心的提供不同面向的觀察，甚至是原本研究忽略掉的細節，使我面對困難時能有更全面的見解。

也很感謝這段時間在實驗室一起努力的夥伴，智文學長、德瑋學長、紹航學長、煒忠、儀晟、孟修、寓鈞、孟哲、毓廷、建昇、俊彥以及一群可愛的學弟們，因為有你們在研究上的扶持與鼓勵，使得我們在實驗室可以一起成長進步。

最感謝的當然還是在背後默默支持我的爸爸、媽媽、爺爺、奶奶、及妹妹，讓我即使在外求學，因為有你們的關心而感到溫暖，讓我更有信心及勇氣來面對一切的挑戰。

CONTENTS

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Epilepsy and Epileptic Seizure

Epilepsy is a common chronic neurological disorder that is characterized by recurrent unprovoked seizures, called Epileptic Seizure (ES). It may be related to a brain injury or a family tendency, but most of the time the cause is unknown. About 50 million people worldwide have epilepsy at any one time, and that would have a profound impact on the quality of life of epilepsy suffers. That's why people want to know about it and even control it.
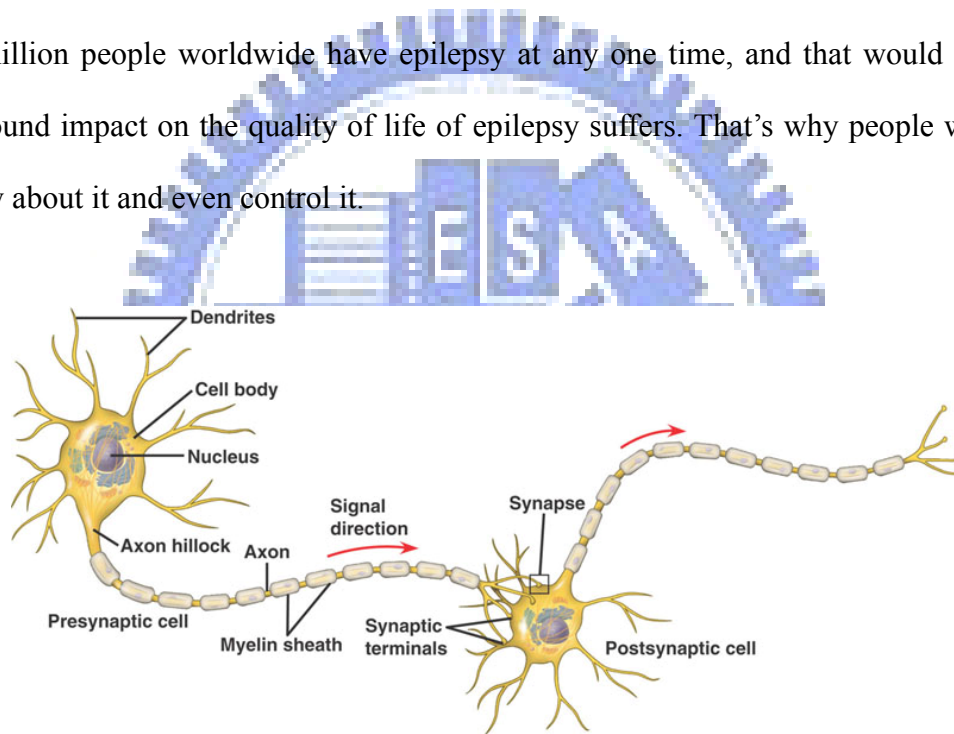
Fig. 1-1 Neurons diagram

(From: http://kvhs.nbed.nb.ca/gallant/biology/neuron_structure.html)

These seizures are transient signs or symptoms due to abnormal, excessive or synchronous neuronal activity in the brain. Generally, the brain continuously generates tiny electrical impulses in an orderly pattern. These impulses travel along the network of nerve cells, called neurons, in the brain and throughout the whole body via chemical messengers called neurotransmitters. A seizure occurs when the brain's

nerve cells misfire and generate a sudden, uncontrolled surge of electrical activity in the brain. Epilepsy should not be understood as a single disorder, but rather as a group of syndromes with vastly divergent symptoms but all involving episodic abnormal electrical activity in the brain.

Treatments are available that can successfully prevent seizures for most people with epilepsy. The first treatment is almost always one of the many seizure medicines (also called an antiepileptic drug or AED) that are now available. These medicines do not actually "fix" the problems that cause seizures. Instead, they work by stopping the seizures from occurring. The second, surgical treatment can be an option for epilepsy when an underlying brain abnormality, such as a benign tumor or an area of scar tissue (e.g. hippocampal sclerosis) can be identified. The abnormality must be removable by a neurosurgeon. Surgery is usually only offered to patients when their epilepsy has not been controlled by adequate attempts with multiple medications.

Fig. 1-2 Vagus nerve stimulation (VNS)

(From: http://www.medgear.org/page/4/)

Vagus Nerve Stimulation (VNS), shown in Fig. 1-2, is a recently developed form of seizure control which uses an implanted electrical device. It is similar in size, shape and implant location to a heart pacemaker, which connects to the vagus nerve in

the neck. The vagus nerve is part of the autonomic nervous system, which controls functions of the body that are not under voluntary control. Once in place the device can be set to emit electronic pulses, stimulating the vagus nerve at pre-set intervals and milliamp levels. A new research on VNS was a major focus at annual meeting of the American Epilepsy Society in 2000. At a symposium on neurostimulation, it was reported that long-term efficacy studies lasting up to 5 years show that VNS can help a wide array of epilepsy patients who do not respond to seizure medicines and cannot be treated with epilepsy surgery. Overall, the studies indicated that 34% to 48% of these adult patients experienced at least a 50% reduction in seizure frequency after 2 to 5 years of follow-up.

1.2 Motivation

Neurobehavioral disorders can profoundly affect the lives of epilepsy suffers. Thus, identification and treatment of cognitive and behavioral disorders are essential. To enhance the conventional treatments, we think incorporating the VNS or seizure medicines with a seizure prediction algorithm or more accurate to say a seizure precursory analysis algorithm is a feasible solution. Thus, the ability to predict seizures would play an important role to improve the quality of life of the people with epilepsy. In recent years, more and more researches focus on seizure analysis, and a lot of valuable papers are presented, such as Similarity Index [1], Sync decrease [2], Approximate entropy (ApEn) [3] etc. But we can say there is no robust enough algorithm has been published to date.

In view of the erroneous conclusions from the traditional statistical analysis methods for non-stationary and non-linear dynamics system of signals may affect the accuracy of forecasts. To determine the main precursory anomalies from brain waves

more precisely, this thesis presents an architecture based on wavelet and chaos theory, including Discrete Wavelet Transform (DWT), correlation dimension, and correlation coefficient. The wavelet transform is more suitable for non-stationary signals than Fast Fourier Transform (FFT) due to its ability of multi-resolution, and time-frequency analysis. The fundamentals of Chaos Theory for non-stationary and non-linear dynamics systems are more in line with the characteristics of brain waves than statistics. Therefore we can achieve a high prediction rate by combining the DWT and Chaos analysis.

For applying the algorithm proposed into a portable physiological monitoring device, we develop a seizure analysis circuit. Some techniques, such as lifting wavelet transform, an enhanced memory addressing, and arithmetic reduction etc., are used in the design to reduce the area and the power consumption of circuit. In the future, we even can integrate it into a digital signal processor of biomedical applications.

1.3 Organization of the Thesis

This thesis is organized as follows. Chapter 2 introduces the theory of chaos analysis and related works. The proposed algorithm for seizure prediction is described in Chapter 3. Chapter 4 describes the implementation techniques of seizure circuit design. Finally, the experimental results and discussions are presented in Chapter 5.

Chapter 2
Fundamentals of Seizure Analysis Algorithm

This chapter will introduce the seizure analysis algorithm based on the Chaos Theory. First, we explain what the Electroencephalography (EEG) is, then use the Chaos Theory to model EEG for analysis. After this, two well-known algorithms as short-term Lyapunov exponential and correlation dimension are introduced.

2.1 Electroencephalography

Electroencephalography (EEG) is the measurement of electrical activity produced by the brain as recorded from electrodes. So-called scalp EEG is collected from tens to hundreds of electrodes positioned on different locations at the surface of the head. EEG signals shown in Fig. 2-2 are amplified and digitalized for post processing.

In some situations, such as epileptic studies, when deeper brain activity needs to be recorded with more accuracy than provided by scalp EEG, clinicians use an invasive form of EEG known as intracranial EEG (icEEG) where electrodes are placed directly inside the skull (see Fig. 2-1). In some cases, a grid of electrodes is laid on the external surface of the brain, on dura mater yielding epidural EEG but in other cases, a depth electrode known as subdural EEG (sdEEG) and electrocorticography (ECoG) is placed into brain structures, such as the amygdala or hippocampus. Because of the filtering characteristics of the skull and scalp, icEEG activity has a much higher spatial resolution than surface EEG.
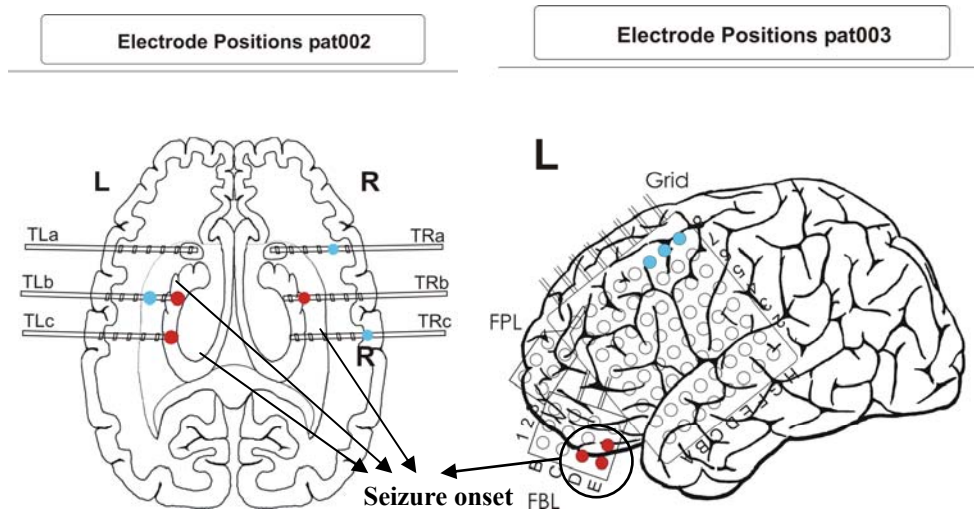
Fig. 2-1 Electrodes positions: contacts in red are chosen from the seizure onset zone and contacts in blue are selected as not involved or involved latest during seizure spread.
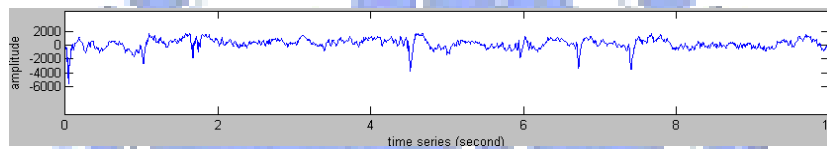
Fig. 2-2 EEG recordings

Then, we have to classify the EEG by symptoms for the observation of the relations between the normal states and abnormal states of the brain, and evaluating our algorithm later. The time intervals of the EEG in different states are defined as follows:

(a) Pre-ictal: the period prior to the start of the seizure.

(b) Ictal: the seizure onset.

(c) Inter-ictal: the period between seizures.

(d) Post-ictal: the period after a seizure.
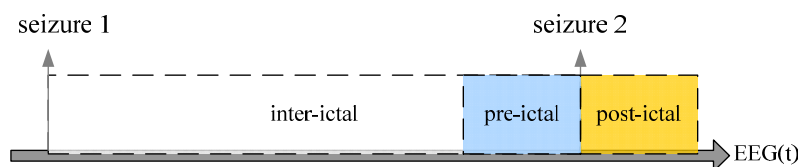
Fig. 2-3 Time intervals of EEG recordings

2.2   Chaotic Modeling for EEG

Many studies have shown that non-linear analysis could characterize the dynamics of neural network underlying EEG which cannot be obtained with conventional linear approach. In the section, we will explain the properties of non-linear dynamics system and how to build an EEG model for chaotic analysis will be explained in detail.

2.2.1   Introduction to Chaos Theorem

Recall that a Newtonian deterministic system is a system whose present state is fully determined by its initial conditions [4] (at least, in principle), in contrast to a stochastic (or random) system, for which the initial conditions determine the present state only partially, due to noise or other external circumstances beyond our control. For a stochastic system, the present state reflects the past initial conditions plus the particular realization of the noise encountered along the way. So, in view of classical science, we have either deterministic or stochastic systems.

For a long time, scientists avoided the irregular side of nature, such as disorder in a turbulent sea, in the atmosphere, and in the fluctuation of wild–life populations. Later, the study of this unusual results revealed that irregularity, nonlinearity, or chaos was the organizing principle of nature.

A modern scientific term deterministic chaos depicts an irregular and unpredictable time evolution of many (simple) deterministic dynamical systems, characterized by nonlinear coupling of its variables. Given an initial condition, the dynamic equation determines the dynamic process, i.e., every step in the evolution. However, the initial condition, when magnified, reveals a cluster of values within a certain error bound. For a regular dynamic system, processes issuing from the cluster

are bundled together, and the bundle constitutes a predictable process with an error bound similar to that of the initial condition. In a chaotic dynamic system, processes issuing from the cluster diverge from each other exponentially, and after a while the error becomes so large that the dynamic equation losses its predictive power.

For example, in 1960s, Ed Lorenz from MIT created a simple weather model in which small changes in starting conditions led to a marked changes in outcome, called sensitive dependence on initial conditions, or popularly, the butterfly effect (i.e., "the notion that a butterfly stirring the air today in Peking can transform storm systems next month in New York, or, even worse, can cause a hurricane in Texas"). Thus long–range prediction of imprecisely measured systems becomes impossibility.

The character of chaotic dynamics can be illustrated with the logistic map as follows [5] :

$$x_{n+1} = r x_n (1 - x_n),$$ (2.1)

a discrete-time analog of the logistic equation for population growth. Here, $x_n \geq 0$ is a dimensionless measure of the population in the n_{th} generation, and $r \geq 0$ is the intrinsic growth rate. We restrict the control parameter r to the range $0 \leq r \leq 4$ so that (2.1) maps the interval $0 \leq x \leq 1$ into itself.

A. Period-Doubling

Suppose we fix *r*, choose some initial population x_0, and then use (2.1) to generate the subsequent x_n. For the growth rate, we can consider cases as follows.

(1) When $r < 1$, the population always goes extinct: $x_n \to 0$ as $n \to \infty$.

(2) When $1 < r < 3$ the population grows and eventually reaches a non-zero steady state, called a period-1 cycle.

(3) Table 2-1 shows the results of logistic map of initial condition $x_0 = 0.4$ and $x_0 = 0.8$, and we can find that even though there is a huge difference between the initial conditions, the two series as shown in Fig. 2-4 converge

to the same value in a moment.

Table 2-1 Results of logistic map (r=2.8)

Logistic growth equation (r=2.8)

X(0)	X(1)	X(2)	X(3)	...	X(22)	X(23)	X(24)	X(25)
0.4	0.672	0.617	0.661	⋯	0.642	0.643	0.642	0.642
0.8	0.448	0.692	0.596	⋯	0.643	0.642	0.643	0.642
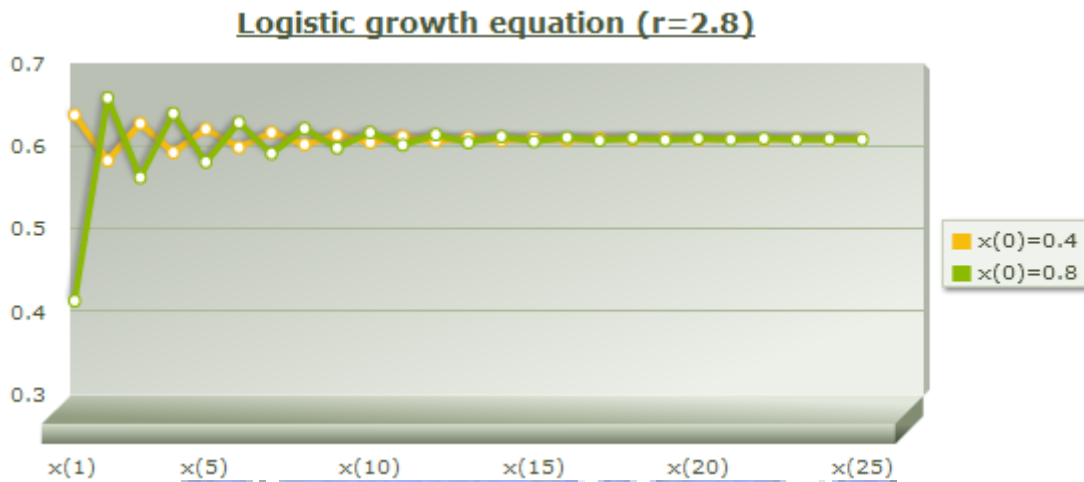
Fig. 2-4 Logistic map of period-1 cycle

(4) When $r = 3.14$, the population builds up again but now oscillates about the former steady state, alternating between a large population in one generation and a smaller population in the next. This type of oscillation, in which x_n repeats every two iterations, is called aperiod-2 cycle. Table 2-2 shows the results of logistic map of initial condition $x_0 = 0.1$ and $x_0 = 0.3$, and the series as shown in Fig. 2-5 reach the same two states after a while.

Table 2-2 Results of logistic map (*r*=3.14)

Logistic growth equation (*r*=3.14)								
X(0)	**X(1)**	**X(2)**	**X(3)**	**...**	**X(22)**	**X(23)**	**X(24)**	**X(25)**
0.1	0.2826	0.6365	0.7264	⋯	0.5385	0.7803	0.5382	0.7804
0.3	0.6594	0.7050	0.6527	⋯	0.7792	0.5402	0.7799	0.5389
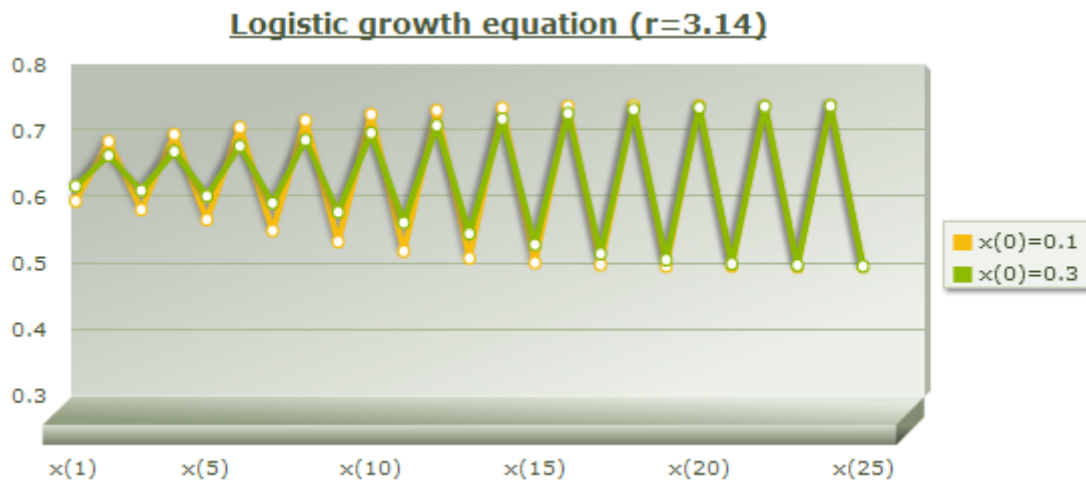
Fig. 2-5 Logistic map of period-2 cycle

Further period-doublings to cycles of period 8, 16, 32, . . . , occur as *r* increases. Specifically, let r_n denote the value of *r* where a 2^n-cycle first appears. Then computer experiments reveal that $r_1 = 3$, $r_2 = 3.449$, $r_3 = 3.54409$, …, $r_\infty = 3.5699$. The convergence is essentially geometric: in the limit of large *n*, the distance between successive transitions shrinks by a constant factor (2.2).

$$\delta = \lim_{n \to \infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = 4.669 \tag{2.2}$$

In fact, the same convergence rate appears no matter what unimodal map is iterated. In this sense, the number δ is universal. It is a new mathematical constant, as basic to period-doubling as n is to circles.

When $r > r_\infty$, the answer turns out to be complicated: For many values of *r*, the sequence $\{x_n\}$ never settles down to a fined point or a periodic orbit instead the

long-term behavior is aperiodic. Table 2-3 shows the results of logistic map. It is interesting to note that even though the difference of the initial conditions is only 0.0001, the two series as shown in Fig. 2-6 are totally divergent in a minute.

Table 2-3 Results of logistic map (r=3.9)

Logistic growth equation (r=3.9)

X(0)	X(1)	X(2)	X(3)	...	X(22)	X(23)	X(24)	X(25)
0.4	0.936	0.23362	0.6982	⋯	0.9184	0.2919	0.8062	0.609
0.4001	0.93607	0.23336	0.6977	⋯	0.6759	0.8542	0.4856	0.9741
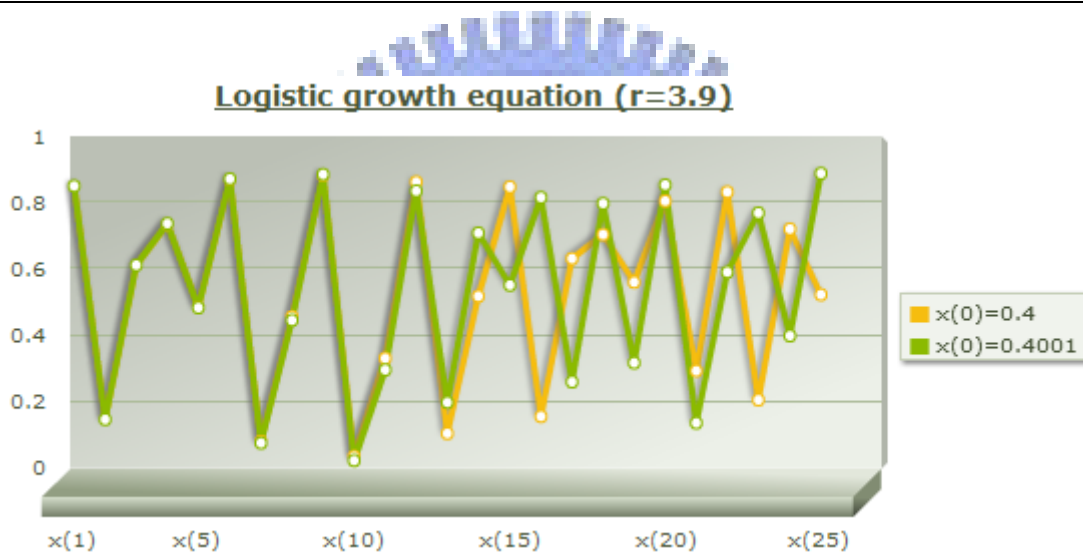
Fig. 2-6 Logistic map of r=3.9

A bifurcation diagram summarizes the above phenomenon (Fig. 2-7). The horizontal axis shows the values of the parameter r while the vertical axis shows the possible values of x.

At $r = 3.4$, the attractor is a period-2 cycle, as indicated by the two branches. As r increases, both branches split simultaneously, yielding a period-4 cycle. This splitting is the period-doubling bifurcation mentioned earlier. A cascade of further period-doublings occurs as r increases, yielding period-8, period-16, and so on, until at $r = r_\infty \approx 3.57$, the map becomes chaotic and the attractor changes from a

11

finite to an infinite set of points.
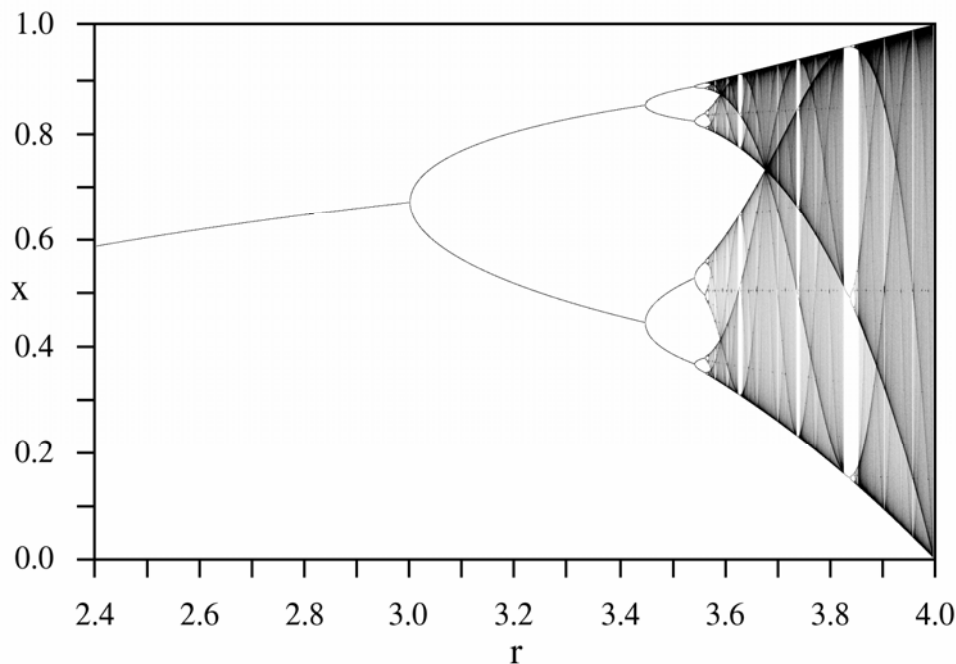
Fig. 2-7 Bifurcation diagram for the Logistic map

B. Basic Terms of Nonlinear Dynamics

Recall that nonlinear dynamics is a language to talk about dynamical systems. Here, brief definitions are given for the basic terms of this language [4].

- *Dynamical system*: A part of the world which can be seen as a self–contained entity with some temporal behavior. Mathematically, a dynamical system is defined by its state and by its dynamics.

- *Phase space*: In mathematics and physics, a phase space, introduced by Willard Gibbs in 1901, is a space in which all possible states of a system are represented, with each possible state of the system corresponding to one unique point in the phase space.

- *Attractor*: An attractor is a 'magnetic set' in the system's phase space to which all neighboring trajectories converge. More precisely, we define an attractor to be a subset of the phase space with the following properties:

12

(1)   It is an invariant set;

(2)   It attracts all trajectories that start sufficiently close to it;

(3)   It is minimal (it cannot contain one or more smaller attractors).

A *strange attractor* shown in Fig. 2-8 is defined to be an attractor that exhibits sensitive dependence on initial conditions. Geometrically, an attractor can be a point, a curve, a manifold, or even a complicated set with a fractal structure known as a strange attractor.
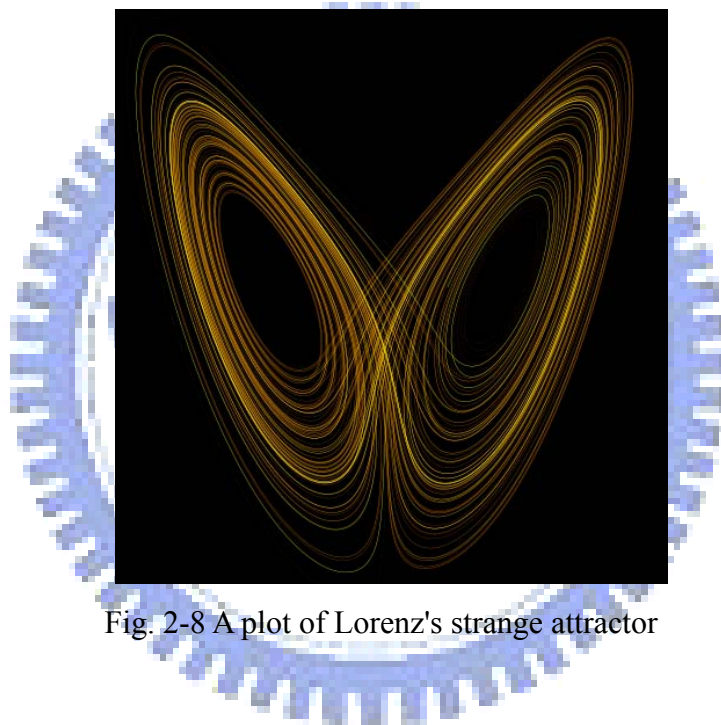
Fig. 2-8 A plot of Lorenz's strange attractor

- *Fractal*: Roughly speaking, fractals are complex geometric shapes with fine structure at arbitrarily small scales. Usually they have some degree of self-similarity. In other words, if we magnify a tiny part of a fractal, we will see features reminiscent of the whole. Sometimes the similarity is exact; more often it is only approximate or statistical.

Fractals are of great interest because of their exquisite combination of beauty, complexity, and endless structure. They are reminiscent of natural objects like mountains, clouds, coastlines, blood vessel networks, and even broccoli, in a

way that classical shapes like cones and squares can't match.

- *Embedding dimension*: The number of variables needed to characterize the state of the system. Equivalently, this number is the dimension of the phase space.

- *Fractal dimension*: The strange attractors typically have fractal microstructure. The attractor dimension counts the effective number of degrees of freedom in the dynamical system, described by a non integer dimension.

C. The link between EEG and chaos

Within the context of brain dynamics [4], there are suggestions that "the controlled chaos of the brain is more than an accidental by–product of the brain complexity" and that "it may be the chief property that makes the brain different from an artificial intelligence machine". Namely, Chaos drives the human brain away from the stable equilibrium, thereby preventing the periodic behavior of neuronal population bursting.

The EEG, being the output of a multidimensional system [6], has statistical properties that depend on both time and space. Components of the brain (neurons) are densely interconnected and the EEG recorded from one site is inherently related to the activity at other sites. This makes the EEG a multivariable time series. The analysis of such nonlinear dynamical systems from time series involves state space reconstruction, and we will introduce in the next section.

If prediction becomes impossible, it is evident that a chaotic system can resemble a stochastic system, say a Brownian motion. However, the source of the irregularity is quite different. For chaos, the irregularity is part of the intrinsic dynamics of the system, not random external influences. Usually, though, chaotic

systems are predictable in the short–term. This short–term predictability is useful.

Chaos theory has developed special mathematical procedures to understand irregularity and unpredictability of low–dimensional nonlinear systems. Lyapunov exponents and attractor dimension are some examples. Lyapunov exponents evaluate the sensitive dependence to initial conditions estimating the exponential divergence of nearby orbits, and we will discuss the method later. Correlation dimensions estimate the fractal dimension and will be described in Chapter 3.

2.2.2 Reconstruction of Attractors from Time Series

Roux et al. (in 1983) exploited a surprising data-analysis technique, now known as attractor reconstruction (Packard et al. 1980, Takens 1981). The claim is that for systems governed by an attractor, the dynamics in the full phase space can be reconstructed from measurements of just a single time series.

Construction of the embedding phase space from a data segment $x(t)$ of duration T is made with the method of delays [6]. The vectors X_i in the phase space are constructed as

$$X_i = (x(t_i), x(t_i + \tau) \ldots x(t_i + (p-1)\tau))^T, \tag{2.3}$$

where τ is the selected time delay between the components of each vector in the phase space, p is the selected dimension of the embedding phase space, and $t_i \in [1, T - (p-1)\tau]$. Obviously, the accuracy of computation depends on the sampling step Δt which decides the number of vectors N_a within a duration T data segment:

$$t_i = t_0 + (i-1) * \Delta t, \text{ where } i \in [1, N_a], \tag{2.4}$$

where t_0 is the initial time point of the fiducial trajectory and coincides with the time point of the first data in the data segment of analysis.

15

The embedding dimension p can be determined from (2.5) if the attractor dimension d is known.

$$p \geq 2d + 1 \qquad\qquad (2.5)$$

The choice of delay τ may also significantly affect the metric characteristics of an attractor. If τ is too small, the ith and the $(i+1)$th coordinates of a phase point are practically equal to each other. In this case, the reconstructed attractor is situated near the main diagonal of the embedding space, the latter complicating its diagnostics. When a value for τ is chosen that is too large, the coordinates become uncorrelated, and the structure of reconstructed attractor is lost.

2.3 Related Works

A chaotic attractor is an attractor where, on the average, orbits originating from similar initial conditions (nearby points in the phase space) diverge exponentially fast (expansion process); they stay close together only for a short time. If these orbits belong to an attractor of finite size, they will fold back into it as time evolves (folding process). The Lyapunov exponents measure the average rate of expansion and folding that occurs along the local eigen-directions within an attractor in phase space. For an attractor to be chaotic, the largest Lyapunov exponent (LLE) must be positive.

As we mentioned before, a relevant time scale should always be used in order to quantify the physiological changes occurring in the brain. Furthermore, the brain being a nonstationary system, algorithms used to estimate measures of the brain dynamics should be capable of automatically identifying and appropriately weighing existing transients in the data.

Iasemidis et al. developed a method [7] for estimation of short-term Lyapunov exponents (STL), an estimate of LLE for nonstationary data. It is well-known and

widely used in many researches. Here we will take an epileptic seizure prediction system, proposed by L. D. Iasemidis et al. in the recent years, for example to explain the STL in detail.

The short-term Lyapunov exponent (STL) is defined as:

$$L = \frac{1}{N_a \Delta t} \sum_{i=1}^{N_a} \log_2 \frac{\left| \delta X_{i,j}(\Delta t) \right|}{\left| \delta X_{i,j}(0) \right|} \qquad (2.6)$$

with
$$\delta X_{i,j}(0) = X(t_i) - X(t_j)$$
$$\delta X_{i,j}(\Delta t) = X(t_i + \Delta t) - X(t_j + \Delta t)$$

based on the reconstruction of attractors from time series, discussed in the last section, where:

- $X(t_i)$ is the point of the fiducial trajectory $\phi_t\left(X(t_0)\right)$ with $t = t_i$, and $X(t_j)$ is a properly chosen vector adjacent to $X(t_i)$ (see below).

- $\delta X_{i,j}(0) = X(t_i) - X(t_j)$ is the displacement vector at t_i, that is, a perturbation of the orbit at t_i, and $\delta X_{i,j}(\Delta t) = X(t_i + \Delta t) - X(t_j + \Delta t)$ is the evolution of this perturbation after time Δt.

- Δt is the evolution time for δX_{ij}, that is, the time one allows δX_{ij} to evolve in the phase space.
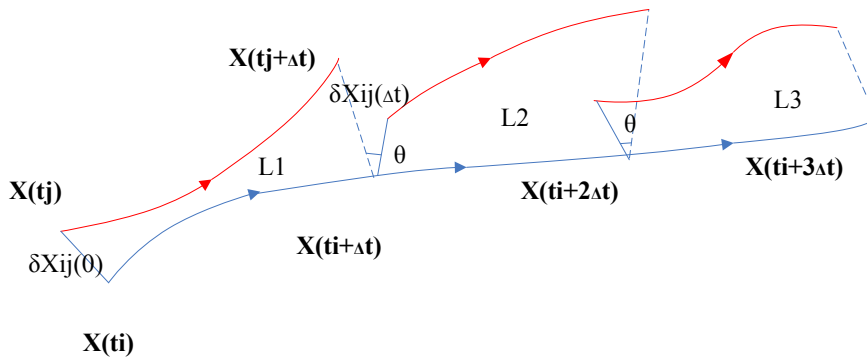
Fig. 2-9 Displacement vectors in the fiducial trajectory

The crucial parameter is the adaptive estimation in time and phase space of the magnitude bounds of the candidate displacement vector to avoid catastrophic replacements. The improvement in the estimates of L can be achieved by using the proposed modifications.

- For L to be a reliable estimate of STL, the candidate vector $X(t_j)$ should be chosen such that the previously evolved displacement vector $\delta X_{(i-1),j}(\Delta t)$ is almost parallel to the candidate displacement vector $\delta X_{i,j}(0)$, that is,

$$\left|V_{ij}\right| = \left|\left\langle \delta X_{i,j}(0), \delta X_{i-1,j}(\Delta t)\right\rangle\right| \leq V_{max} \tag{2.7}$$

where V_{max} should be small and $\left|\left\langle \varepsilon, \phi \right\rangle\right|$ denotes the absolute value of the angular separation between two vectors.

- For L to be a reliable estimate of STL, $\delta X_{i,j}(0)$ should also be small in magnitude in order to avoid computer overflow in the future evolution within very chaotic regions and to reduce the probability of starting up with points on separatrices. This means,

$$\left|\delta X_{i,j}(0)\right| = \left|X(t_i) - X(t_j)\right| < \Delta_{max} \tag{2.8}$$

with Δ_{max} assuming small values.

A typical long-term plot of STL versus time, obtained by analysis of continuous EEG, is shown in Fig. 2-10. This figure shows the evolution of STL at a focal electrode site, as the brain progresses from interictal to ictal to postictal states. There is a gradual drop in STL over approximately 2 hours preceding this seizure. The seizure, 2 minutes in duration, is characterized by a sudden drop in STL values with a

consequent steep rise. Postictal STL values exceed preictal values and slowly approach interictal values. This behavior of STL indicates a gradual preictal reduction in chaoticity, reaching a minimum shortly after seizure onset, and a postictal rise in chaoticity that corresponds to the reversal of the preictal pathological state. There will be more discussions about this character in Chapter 3.
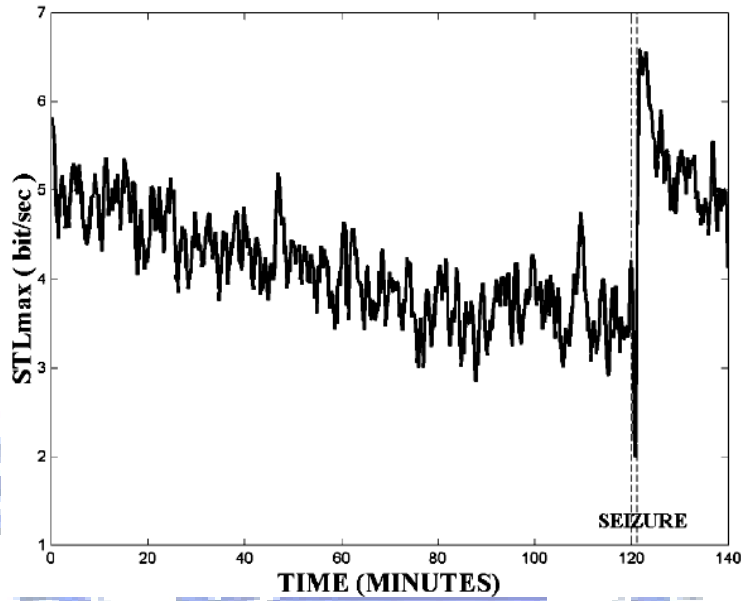
Fig. 2-10 Unsmoothed STL over time (140 min), including a 2-min seizure. [7]

Having estimated the STL temporal profiles at each electrode site, and as the brain proceeds toward the ictal state, the temporal evolution of the stability of each cortical site is quantified. However, since the brain is a system of spatial extent, information about the interactions of its spatial components should also be taken in consideration by the relations of the STL between different cortical sites.

The *T*- index at time t between electrode sites *i* and *j* is then defined as:

$$T_{i,j} = E\left\{\left|STL_i(t) - STL_j(t)\right|\right\} \div \frac{\sigma_{i,j}(t)}{\sqrt{N}}, \tag{2.9}$$

where $E\{\ \}$ denotes the average of all absolute differences $\left|STL_i(t) - STL_j(t)\right|$

within a moving window $w_t(\lambda)$ defined as:

$$w_t(\lambda) = 1 \ \ for \ \ \lambda \in [t\text{-}N\text{-}1,t] \ \ and \ \ w(\lambda) = 0 \ \ for \ \ \lambda \notin [t\text{-}N\text{-}1,t]$$

where N is the length of the moving window. Then, $\sigma_{i,j}(t)$ is the sample standard

deviation of the STL differences between electrode sites i and j within the moving

window.

A dynamical transition toward a seizure is announced at time t^* when the

T-indexes of sites over time transits from a value above threshold T_1 at times $t < t'$,

to a value below threshold T_2 at time t^*, as shown in Fig. 2-11.
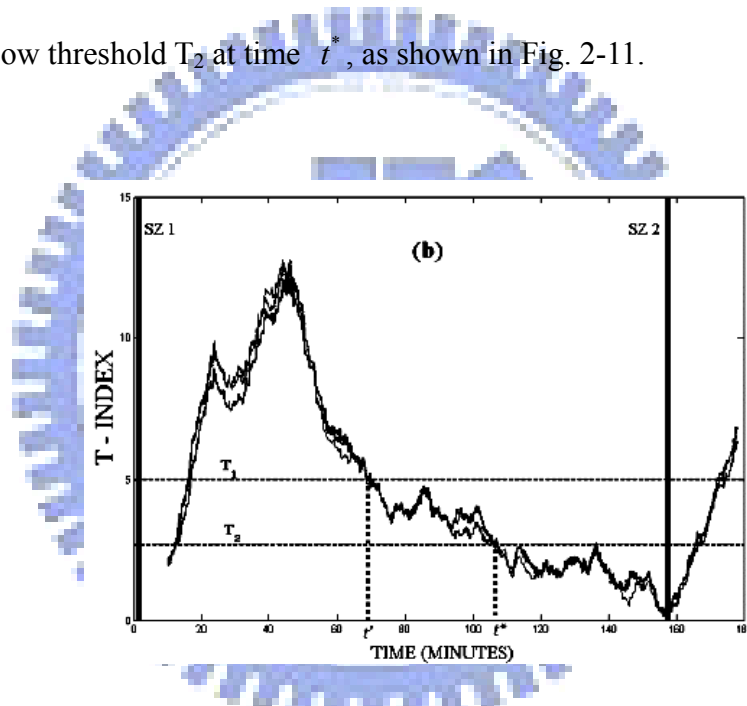
Fig. 2-11 The T-index curves denoting entrainment 55 min before seizure SZ2 [7].

The method presented achieved amazing results with high prediction sensitivity,

and pretty low false prediction rate. More details and comparison results with other

algorithms and ours will be listed in Chapter 5.

Chapter 3
Wavelet-Correlation Dimension Seizure Prediction

In this chapter a real-time seizure prediction method based on correlation dimension analysis is presented, including the system architecture, data flow, and algorithms.

3.1   Architecture of Seizure Prediction

Before starting the prediction processing, first we observe the EEG signal whether exists any clue around the seizures. For example, as show in Fig. 3-1, for different clinical states including pre-ictal, ictal, and post-ictal states, the corresponding properties of intracranial EEG recordings are different.
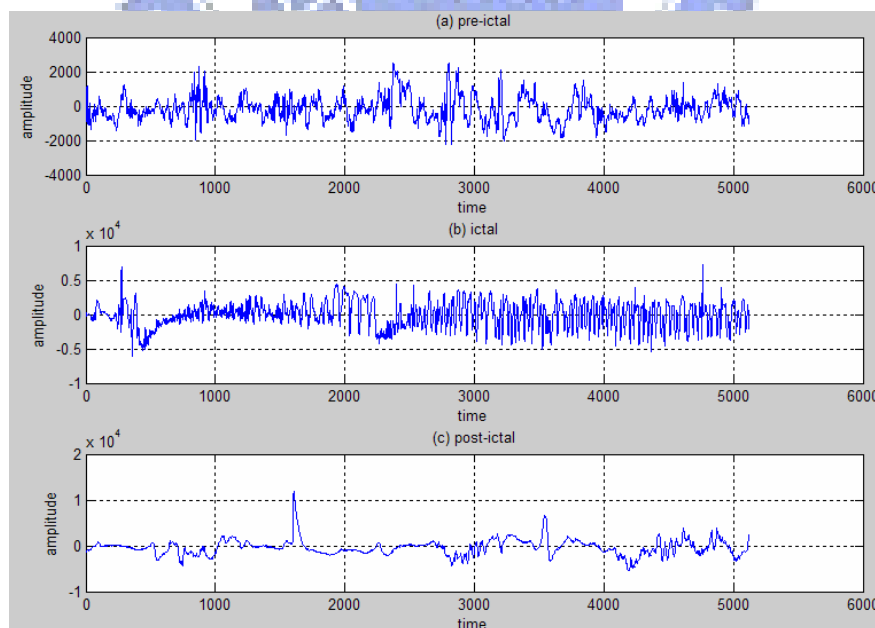
Fig. 3-1 Typical EEG waveforms corresponding to epilepsy:

(a) pre-ictal, (b) ictal, and (c) post-ictal

In the pre-ictal state the EEG signal is of the chaotic nature. As we approach the epileptic seizure the signals are less and less chaotic and take the regular shape. These findings imply that seizures may represent spatiotemporal transitions of the epileptic brain from chaos-to-order-to-chaos. Therefore the chaoticity measure of the signal is a good prognostic of the incoming seizure. In fact, this phenomenon is confirmed by STL in the last chapter, but we try to use another method to prove it.

In this chapter, we would propose a Wavelet-Correlation Dimension based Seizure Prediction system, called WCDSP, as shown in Fig. 3-2. The system is consisted of three primary parts:

- Discrete Wavelet Transform analysis: Wavelet is used to decompose the EEG into several sub-bands.

- Chaos analysis: Correlation dimension is used to measure the EEG complexity.

- Feature extraction: The correlation coefficient is used to be the main feature for prediction rules and to decide the seizure states.
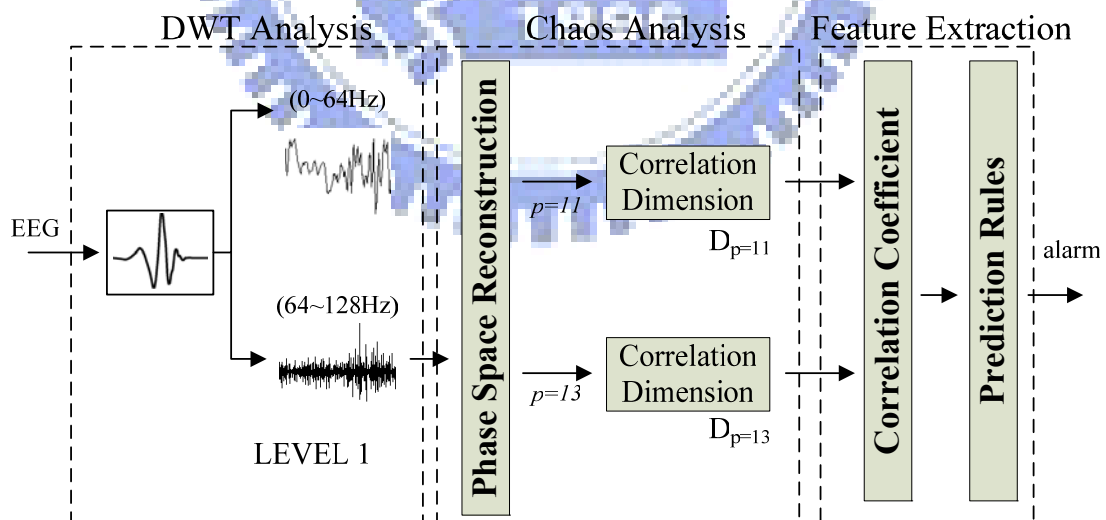
Fig. 3-2 System architecture of seizure prediction

3.2 Discrete Wavelet Transform

We would introduce the discrete wavelet transform (DWT) in this section, and briefly discuss the properties of the discrete Fourier transform (DFT), the short-time Fourier transform (STFT), and the wavelet transform.

3.2.1 Discrete Fourier Transform

In mathematics, the discrete Fourier transform (DFT) is one of the specific forms of Fourier analysis, and is widely employed in signal processing and related fields to analyze the frequencies contained in a sampled signal. As such, it transforms one function into another, which is called the frequency domain representation of the original function (which is often a function in the time domain).

The sequence of N complex numbers $x_0, ..., x_{N-1}$ is transformed into the sequence of N complex numbers $X_0, ..., X_{N-1}$ by the DFT according to the formula

$$\hat{f} = X(k) = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} , \quad k = 0, ..., N-1 \tag{3.1}$$

The importance of the Fourier transform stems not only from the significance of their physical interpretations, such as time-frequency analysis of signals, but also from the fact that Fourier analytic techniques are extremely powerful.

While Fourier analysis forces us to choose between time on one side of the transform and frequency on the other, "…our everyday experiences insist on a description in terms of both time and frequency," Gabor wrote. To analyze a signal in both time and frequency, he used the windowed Fourier transform. The idea is to study the frequencies of a signal segment by segment; the way, one can at least limit the span of time during witch something is happening. The "window" that defines the size of the segment to analyzed — and which remains fixed in size — is a little piece of curve.

One of the downfalls of the STFT is that it has a fixed resolution. The width of the windowing function relates to the how the signal is represented — it determines whether there is good frequency resolution (frequency components close together can be separated) or good time resolution (the time at which frequencies change).

3.2.2 Heisenberg uncertainty principle

We want to construct a function f whose energy is well localized in time and whose Fourier transform $\hat{f}$ has an energy concentrated in a small frequency neighborhood.

The Heisenberg principle [8] says the following. For every function $f(t)$, such that

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = 1 \tag{3.2}$$

The product of the variance of t and the variance of τ (the variable of $\hat{f}$) is at least $\dfrac{h}{16\pi^2}$, where h is the Planck's constant:

$$\underbrace{\left(\int_{-\infty}^{\infty} (t - t_m)^2 |f(t)|^2 \, dt \right)}_{\text{variance of } t} \underbrace{\left(\int_{-\infty}^{\infty} (\tau - \tau_m)^2 |\hat{f}(t)|^2 \, d\tau \right)}_{\text{variance of } \tau} \geq \frac{h}{16\pi^2} \tag{3.3}$$

These variances measure to what extent t and τ take values far from their average values, t_m and τ_m. Thus the shorter-lived a function, the wider the band of frequencies given by its Fourier transform; the narrower the band of frequencies of its Fourier transform, the more the function is spread out in time. Time and frequency energy concentrations are restricted by the Heisenberg uncertainty principle.

3.2.3 Wavelet and Multiresolution Analysis

Unlike the Fourier transform, whose basis functions are sinusoids, wavelet transforms are based on small waves, called wavelets [9], of varying frequency and limited duration. This allows them to provide the equivalent of a musical score for a signal, revealing not only what notes (or frequencies) to play but also when to play them. Conventional Fourier transforms, on the other hand, provide only the notes or frequency information; temporal information is lost in the transformation process.

In 1987, wavelets were first shown to be the foundation of a powerful new approach by Mallat to signal processing and analysis called multiresolution theory. Multiresolution theory incorporates and unifies techniques from a variety of disciplines, including subband coding from signal processing, quadrature mirror filtering from digital speech recognition, and pyramidal image processing. As its name implies, multiresolution theory is concerned with the representation and analysis of signals at more than one resolution. The appeal of such an approach is obvious—features that might go undetected at one resolution may be easy to spot at another.

A. Background

When we look at images, generally we see connected regions of similar texture and gray level that combine to form objects. If the objects are small in size or low in contrast, we normally examine them at high resolutions; if they are large in size or high in contrast, a coarse view is all that is required. If both small and large objects—or low and high contrast objects—are present simultaneously, it can be advantageous to study them at several resolutions. This, of course, is the fundamental motivation for multiresolution processing.

(1) Image Pyramids

An image pyramid is a collection of decreasing resolution images arranged in the shape of a pyramid. As can be seen in Fig. 3-3, the base of the pyramid contains a high-resolution representation of the image being processed; the apex contains a low-resolution approximation. As you move up the pyramid, both size and resolution decrease.
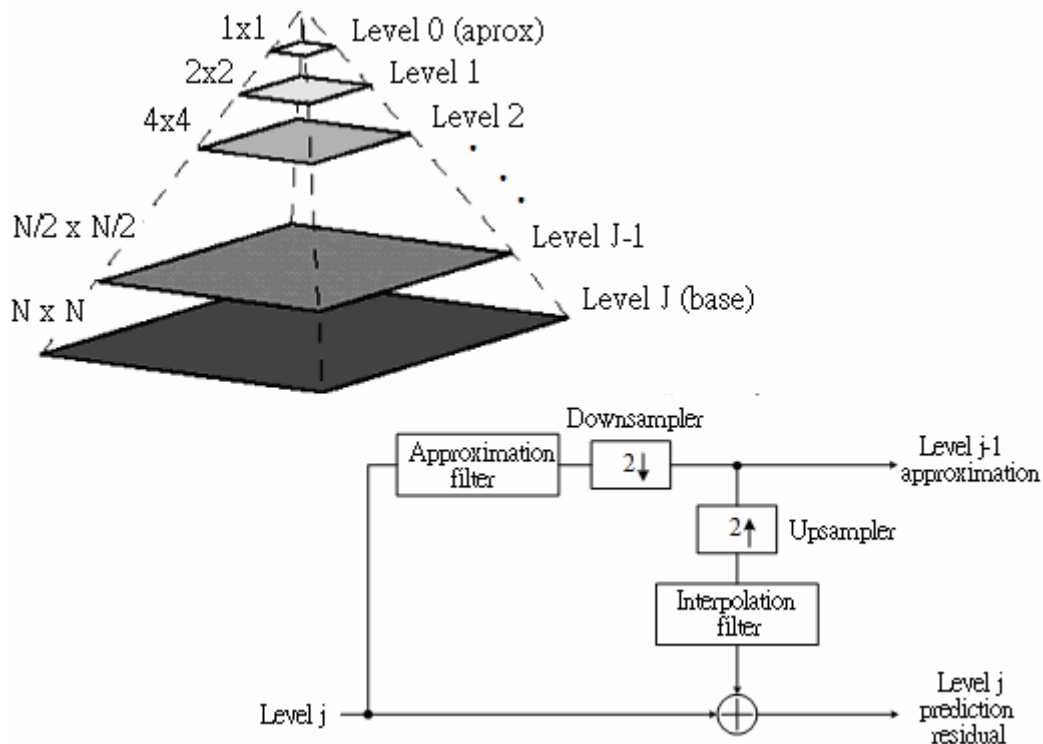
Fig. 3-3 A pyramidal image structure and system block diagram for creating it

The level $j-1$ approximation output is used to create approximation pyramids, which contain one or more approximations of the original image. The level j prediction residual output is used to build prediction residual pyramids.

For example, Fig. 3-4 shows one possible approximation (Gaussian) and prediction residual (Laplacian) pyramid for the vase. The Laplacian pyramid contains the prediction residuals needed to compute its Gaussian counterpart. To build the Gaussian pyramid, we begin with the Laplacian pyramid's level j 64 by 64

approximation image, predict the Gaussian pyramid's level $j+1$ 128 by 128 resolution approximation (by upsampling and filtering), and add the Laplacian's level $j+1$ prediction residual. This process is repeated using successively computed approximation images until the original 512 by 512 image is generated.
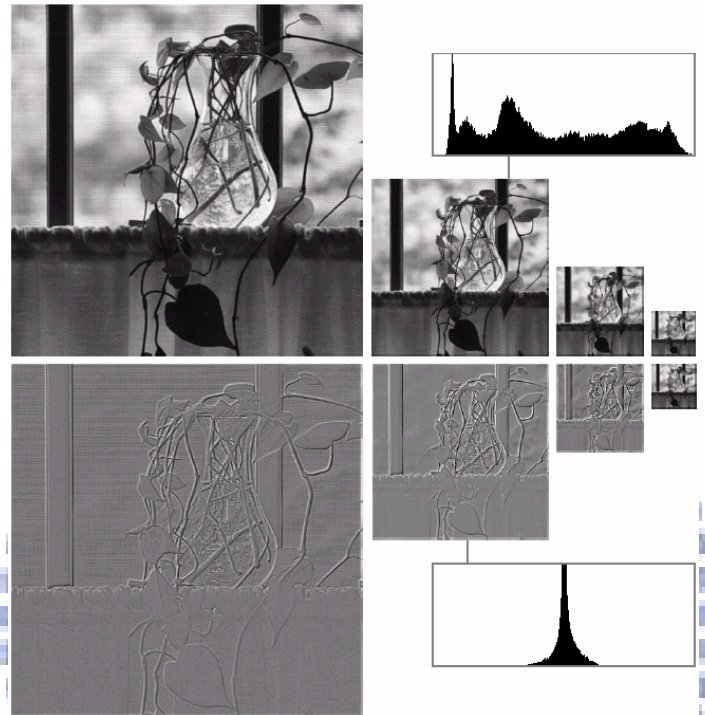
Fig. 3-4 Two image pyramids and their statistics: a approximation (Gaussian) pyramid and a prediction residual (Laplacian) pyramid

(2) Subband Coding

Another important imaging technique with ties to multiresolution analysis (MRA) is subband coding. In subband coding, an image is decomposed into a set of band-limited components, called subbands, which can be reassembled to reconstruct the original image without error. Since the bandwidth of the resulting subbands is smaller than that of the original signal, the subbands can be downsampled without loss of information. Reconstruction of the original signal is accomplished by upsampling, filtering, and summing the individual subbands.

Fig. 3-5 shows the principal components of a two-band subband coding and

decoding system. The input of the system is a one-dimensional, band-limited discrete-time signal $x(n)$ for $n = 0,1,2,...$; the output sequence, $\hat{x}(n)$, is formed through the decomposition of $x(n)$ into $y_0(n)$ and $y_1(n)$ via analysis filters $h_0(n)$ and $h_1(n)$, and subsequent recombination via synthesis filters $g_0(n)$ and $g_1(n)$. Note that filters $h_0(n)$ and $h_1(n)$ are half-band digital filters whose idealized transfer characteristics, H_0 and H_1, are shown in Fig. 3-5(b).
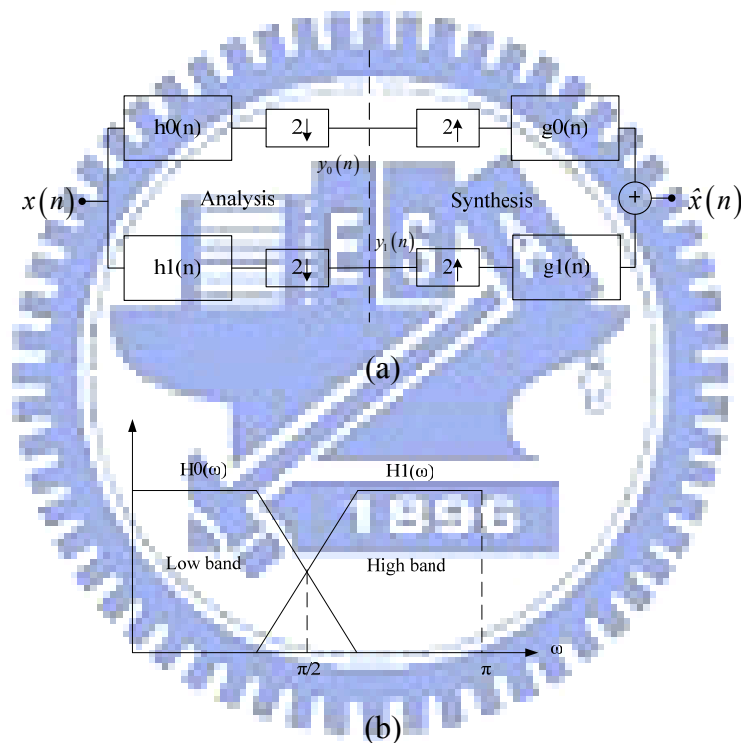
Fig. 3-5 (a) A two-band filter bank for one-dimensional subband coding and decoding and (b) its spectrum splitting properties

Filter H_0 is a low-pass filter whose output is an approximation of $x(n)$; filter H_1 is a high-pass filter whose output is the high frequency or detail part of $x(n)$. We wish to select $h_0(n)$, $h_1(n)$, $g_0(n)$, and $g_1(n)$ so that the input can be reconstructed perfectly. That is, so that $\hat{x}(n) = x(n)$.

We can express the system's output as:

$$\hat{X}(z) = \frac{1}{2}\left[H_0(z)G_0(z) + H_1(z)G_1(z)\right]X(z) \tag{3.4}$$

$$+ \frac{1}{2}\left[H_0(-z)G_0(z) + H_1(-z)G_1(z)\right]X(-z)$$

For error reconstruction of the input, $\hat{x}(n) = x(n)$ and $\hat{X}(z) = X(z)$. Thus, we

impose the following conditions:

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 2 \tag{3.5}$$
$$H_0(-z)G_0(z) + H_1(-z)G_1(z) = 0$$

Both can be incorporated into the single matrix expression

$$\left[G_0(z) \quad G_1(z)\right]H_m(z) = \left[2 \quad 0\right] \tag{3.6}$$

where analysis modulation matrix $H_m(z)$ is

$$H_m(z) = \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \tag{3.7}$$

Assume $H_m(z)$ is nonsingular, we can transpose (3.6) and left multiply by inverse

$\left(H_m^T(z)\right)^{-1}$ to get

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \frac{2}{det(H_m(z))}\begin{bmatrix} H_1(-z) \\ -H_0(-z) \end{bmatrix} \tag{3.8}$$

For FIR filters, the determinate of the modulation matrix is a pure delay.

$$det(H_m(z)) = \alpha \cdot z^{-(2k+1)}$$

Ignoring the delay and let $\alpha = 2$

$$g_0[n] = (-1)^n h_1[n]$$
$$g_1[n] = (-1)^{n+1} h_0[n]$$

Define $P(z) = G_0(z)H_0(z) = \dfrac{2}{det(H_m(z))}H_0(z)H_1(-z)$

Then, $G_1(z)H_1(z) = \dfrac{-2}{det(H_m(z))}H_0(-z)H_1(z) = P(-z)$

$\Rightarrow H_0(z)G_0(z) + H_0(-z)G_0(-z) = 2$

$\Rightarrow \sum_k g_0[k]h_0[n-k] + (-1)^n \sum_k g_0[k]h_0[n-k] = 2\delta[n]$

$\Rightarrow \sum_k g_0[k]h_0[2n-k] = <g_0[k], h_0[2n-k]> = \delta[n]$

Similarly, we can show that

$$< g_1[k], h_1[2n-k] > = \delta[n]$$
$$< g_0[k], h_1[2n-k] > = 0$$
$$< g_1[k], h_0[2n-k] > = 0$$

That is,

$$< h_i[2n-k], g_j[k] > = \delta[i-j]\delta[n] \qquad i,j = \{0,1\} \qquad (3.9)$$

Filter banks satisfying this condition are called *biorthogonal*. Moreover, the analysis and synthesis filter impulse responses of all two-band, real-coefficient, perfect reconstruction filter banks are subject to the biorthogonality constraint.

Orthonormal filter banks:

$$< h_i[2n-k], g_j[k] > = \delta[i-j]\delta[n] \qquad i,j = \{0,1\}$$
$$< g_i[k], g_i[n+2m] > = \delta[i-j]\delta[m] \qquad i,j = \{0,1\} \qquad (3.10)$$

$$g_1[n] = (-1)^n g_0[2k-1-n]$$
$$h_i[n] = g_i[2k-1-n], \quad i = \{0,1\} \qquad (3.11)$$

B. Multiresolution Expansions

In MRA, a scaling function is used to create a series of approximations of a function, each differing by a factor of 2 from its nearest neighboring approximations. Additional functions, called wavelets, are then used to encode the difference in information between adjacent approximations.

A function $f(x)$ can be decomposed into a linear combination of expansion functions as follows.

$$f(x) = \sum_k \alpha_k \varphi_k(x)$$

If the expansion is unique, the $\varphi_k(x)$ are called basis functions. The expressible function forms a *function space* $V = \overline{span}_k \{\varphi_k(x)\}$.

For any function space V and corresponding expansion set $\{\varphi_k(x)\}$, there exist a set of dual functions, denoted $\{\tilde{\varphi}_k(x)\}$, which can be used to compute the α_k coefficients for any $f(x) \in V$.

$$\alpha_k = <\tilde{\varphi}_k(x), f(x)> = \int \tilde{\varphi}_k^*(x) f(x) dx$$

Case 1: **Orthonormal basis**

$$<\varphi_j(x), \varphi_k(x)> = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

The basis and its dual are equivalent and $\alpha_k = <\varphi_k(x), f(x)>$.

Case 2: **Biorthogonal basis**

$$<\varphi_j(x), \tilde{\varphi}_k(x)> = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

- For the case of wavelet expansion, we restrict ourselves to forming the basis functions by binary scaling (shrinking by factors of two) and dyadic translation (shifting by the amount k/2^j)

- Consider the set of expansion functions composed of integer translations and binary scalings of the real, square-integrable function $\varphi(x)$:

$$\varphi_{j,k}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k) \quad \text{for all } j, k \in Z \text{ and } \varphi(x) \in L^2(R)$$

-- $\varphi(x)$ is called a scaling function

-- $L^2(R)$: the set of all measurable, square-integrable functions

-- By choosing $\varphi(x)$ wisely, $\{\varphi_{j,k}(x)\}$ can be made to span $L^2(R)$

-- If we restrict j to a specify value, $j = j_0$ the resulting expansion set

$\{\varphi_{j_0,k}(x)\}$ is a subset of $\{\varphi_{j,k}(x)\}$.

$$V_{j_0} = \overline{\underset{k}{span}\{\varphi_{j_0,k}(x)\}}$$

More generally, we denote

$$V_j = \overline{\underset{k}{span}\{\varphi_{j,k}(x)\}}$$

Four fundamental requirements of multiresolution analysis:

MRA requirement 1:

The scaling function is orthogonal to its integer translates.

MRA requirement 2:

The subspaces spanned by the scaling function at low scales are nested within those spanned at higher scales.

$$V_{-\infty} \subset \ldots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots \subset V_\infty$$

MRA requirement 3:

The only function that is common to all V_j is $f(x) = 0$.

$$V_{-\infty} = \{0\}$$

MRA requirement 4:

Any function can be represented with arbitrary precision.

$$V_\infty = \{L^2(R)\}$$

- $$\varphi_{j,k}(x) = \sum_n \alpha_n \varphi_{j+1,n}(x) = \sum_n h_\varphi[n] \cdot 2^{(j+1)/2} \varphi\left(2^{j+1}x - n\right)$$

$$\varphi(x) = \varphi_{0,0}(x) = \sum_n h_\varphi[n] \cdot \sqrt{2} \cdot \varphi(2x - n),$$

which is called as refinement equation (MRA equation, dilation equation).

$h_\varphi[n]$: scaling function coefficients

- Given a scaling function that meets the MRA requirements, we can define a wavelet function $\psi(x)$.

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi\left(2^j x - k\right) \qquad \text{for all} \quad j,k \in Z$$

and $W_j = \overline{span_k\{\psi_{j,k}(x)\}}$

$$V_{j+1} = V_j \oplus W_j$$

$$<\varphi_{j,k}(x), \psi_{j,l}(x)> = 0 \quad \text{for all} \quad j,k,l \in Z$$

$$L^2(R) = V_0 \oplus W_0 \oplus \ldots$$
$$= V_1 \oplus W_1 \oplus W_2 \oplus \ldots$$
$$= \ldots \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \ldots$$

$$\psi(x) = \sum_n h_\psi[n] \cdot \sqrt{2} \cdot \varphi(2x - n)$$

$h_\psi[n]$: wavelet function coefficients

C. Wavelet Transform

We can now formally define several closely related wavelet transformations: the generalized wavelet series expansion, the discrete wavelet transform, and a computationally efficient implementation of the discrete wavelet transform called the fast wavelet transform.

(1) The Wavelet Series Expansion

We begin by defining the wavelet series expansion of function $f(x) \in L^2(R)$

relative to wavelet $\psi(x)$ and scaling function $\varphi(x)$.

$$f(x) = \sum_k c_{j_0}[k] \cdot \varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_j[k] \cdot \psi_{j,k}(x), \qquad (3.12)$$

$c_{j_0}[k]$: approximation (or scaling) coefficients

$d_j[k]$: detail (or wavelet) coefficients

For orthonormal bases and tight frames,

$$c_{j_0}[k] = <f(x),\varphi_{j_0,k}(x)> = \int f(x)\varphi_{j_0,k}(x)dx$$
$$d_j[k] = <f(x),\psi_{j,k}(x)> = \int f(x)\psi_{j,k}(x)dx$$

For biorthogonal bases,

$$c_{j_0}[k] = <f(x),\tilde{\varphi}_{j_0,k}(x)> = \int f(x)\tilde{\varphi}_{j_0,k}(x)dx$$
$$d_j[k] = <f(x),\tilde{\psi}_{j,k}(x)> = \int f(x)\tilde{\psi}_{j,k}(x)dx$$

(2) Discrete Wavelet Transform

Like the Fourier series expansion, the wavelet series expansion maps a function of a continuous variable into a sequence of coefficients. If the function being expanded is a sequence of numbers, like samples of a continuous function $f(x)$, the resulting coefficients are called the *discrete wavelet transform* (DWT) of $f(x)$.

$$f(x) = \frac{1}{\sqrt{M}} \sum_k W_\varphi[j_0,k]\varphi_{j_0,k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_\psi[j,k]\psi_{j,k}(x), \qquad (3.13)$$

$W_\varphi[j_0,k]$: approximation (or scaling) coefficients

$W_\psi[j,k]$: detail (or wavelet) coefficients

For orthonormal bases and tight frames,

$$W_\varphi[j_0,k] = \frac{1}{\sqrt{M}} \sum_x f[x]\varphi_{j_0,k}(x)$$

$$W_\psi[j,k] = \frac{1}{\sqrt{M}} \sum_x f[x]\psi_{j,k}(x)$$

34

For biorthogonal bases,

$$W_\varphi[j_0,k] = \frac{1}{\sqrt{M}}\sum_x f[x]\tilde{\varphi}_{j_0,k}(x)$$

$$W_\psi[j,k] = \frac{1}{\sqrt{M}}\sum_x f[x]\tilde{\psi}_{j,k}(x)$$

(3) The Fast Wavelet Transform (FWT)

The fast wavelet transform (FWT) is a computationally efficient implementation of the discrete wavelet transform (DWT) that exploits a surprising but fortunate relationship between the coefficients of the DWT at adjacent scales, also called Mallat's herringbone algorithm (Mallat [1989]).

Consider again the multiresolution refinement equation:

$$\varphi(x) = \sum_n h_\varphi[n]\cdot\sqrt{2}\cdot\varphi(2x-n)$$

$$\varphi(2^j x - k) = \sum_n h_\varphi[n]\cdot\sqrt{2}\cdot\varphi(2(2^j x - k)-n)$$

$$= \sum_m h_\varphi[m-2k]\cdot\sqrt{2}\cdot\varphi(2^{j+1}x-m)$$

Similarly, $\psi(2^j x - k) = \sum_m h_\psi[m-2k]\cdot\sqrt{2}\cdot\psi(2^{j+1}x-m)$

$$W_\psi[j,k] = \frac{1}{\sqrt{M}}\sum_x f[x]\psi_{j,k}(x) = \frac{1}{\sqrt{M}}\sum_x f[x]2^{j/2}\psi(2^j x - k)$$

$$= \frac{1}{\sqrt{M}}\sum_x f[x]2^{j/2}\left[\sum_m h_\psi[m-2k]\sqrt{2}\varphi(2^{j+1}x-m)\right]$$

$$= \sum_m h_\psi[m-2k]\left[\frac{1}{\sqrt{M}}\sum_x f[x]2^{(j+1)/2}\psi(2^{j+1}x-k)\right]$$

$$\Rightarrow W_\psi[j,k] = \sum_m h_\psi[m-2k]\cdot W_\varphi[j+1,m]$$

$$W_\psi[j,k] = h_\psi[-n]*W_\varphi[j+1,m]$$

Similarly,

$$W_\varphi[j,k] = h_\varphi[-n]*W_\varphi[j+1,n]$$

35

Fig. 3-6 reduces these operations to block diagram form. We note that the filter bank can be "iterated" to create multistage structures for computing DWT coefficients.
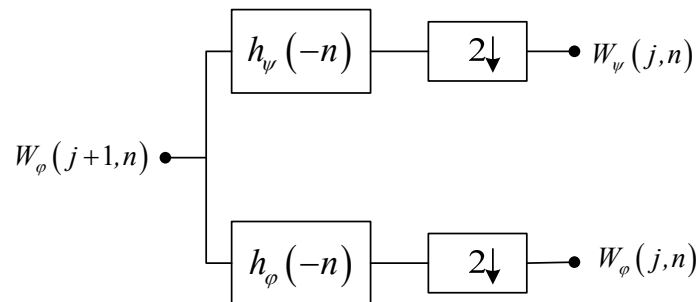

Fig. 3-6 FWT analysis bank

D. Time-Frequency Analysis

Fig. 3-7 shows the time-frequency tiles for (a) a delta function (i.e., conventional time domain) basis, (b) a sinusoidal (FFT) basis, and (c) an FWT basis.

(a)   (b)   (c)

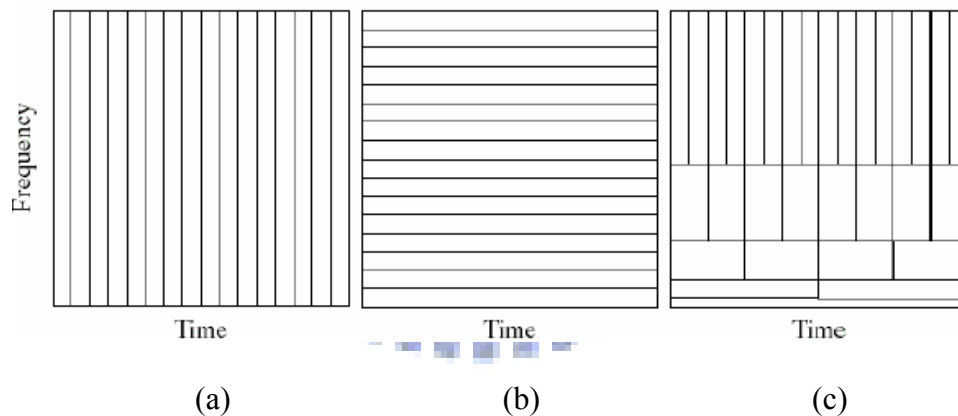Fig. 3-7 Time-frequency tilings for (a) sampled data, (b) FFT, and (c) FWT basis functions.

Note that the standard time domain basis pinpoints the instants when events occur but provides no frequency information. A sinusoidal basis, on the other hand, pinpoints the frequencies that are present in events that occur over long periods but provides no time resolution. The time and frequency resolution of the FWT tiles vary.

At low frequencies, the tiles are shorter (i.e., have better frequency resolution) but are wider (which corresponds to poorer time resolution). At high frequencies, tile width is smaller (so the time resolution is improved) and tile height is greater (which means the frequency resolution is poorer). This fundamental difference between the FFT and FWT was noted in the introduction to the section and is important in the analysis of nonstationary functions whose frequencies vary in time.

In this research, a two-level Daubechies 4 (db4) wavelet is used for the EEG recordings. Fig. 3-8 shows (a) the corresponding wavelet structure, and (b) the time-frequency tilings of EEG signal.
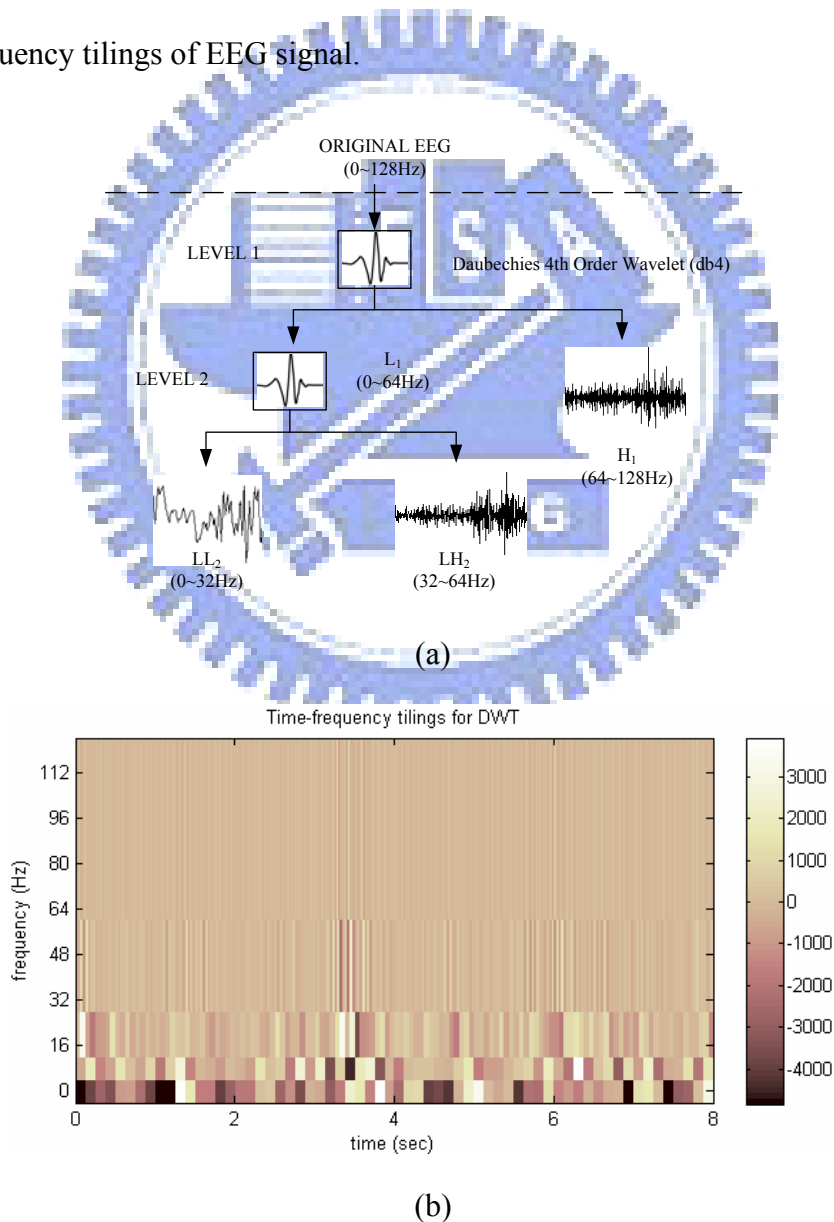

(a)

(b)

Fig. 3-8 (a) Two-level Daubechies 4 wavelet and (b) Time-frequency tilings of EEG

The corresponding analysis and synthesis filters are shown in Fig. 3-9. We list the analysis part, including a low-pass and a high-pass filter, as follows:

$$h_0(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + h_3 z^{-3}$$
$$h_1(z) = h_3 z^2 - h_2 z^1 + h_1 - h_0 z^{-1}, \qquad (3.14)$$

where $h_0 = \dfrac{1+\sqrt{3}}{4\sqrt{2}}, h_1 = \dfrac{3+\sqrt{3}}{4\sqrt{2}}, h_2 = \dfrac{3-\sqrt{3}}{4\sqrt{2}}, h_3 = \dfrac{1-\sqrt{3}}{4\sqrt{2}}$
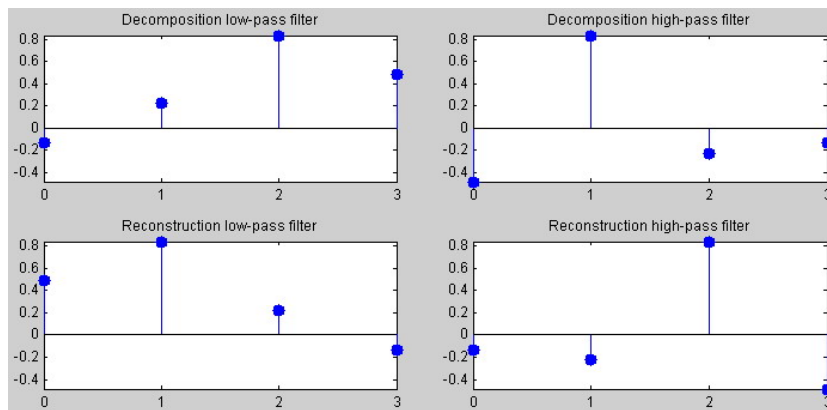

Fig. 3-9 Analysis and synthesis filters

We decompose the original EEG into several subbands, including L_1, H_1, LL_2, and LH_2 (Fig. 3-10). Each subband may contain some specific characteristic of the brain dynamics. In the next section, we will use these subbands for advanced analysis.
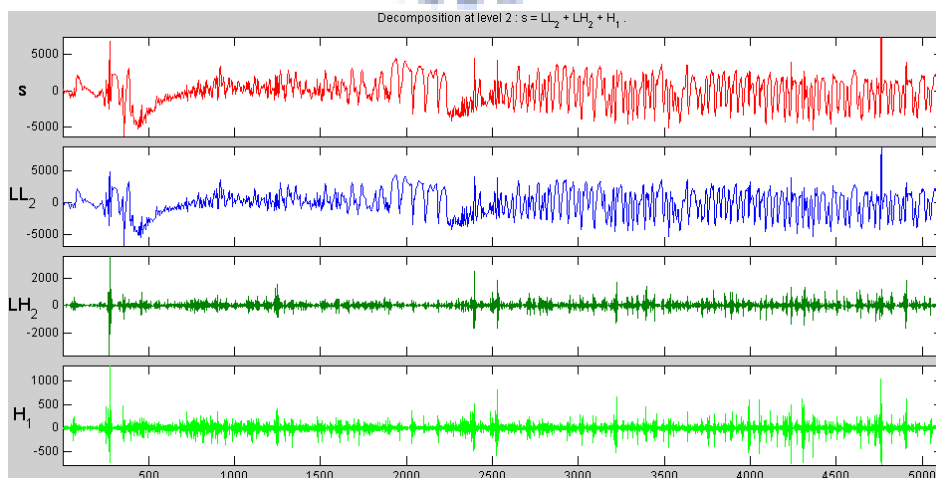

Fig. 3-10 Decompose the EEG into many subbands

3.3   Correlation Dimension

Estimating the fractal dimension of a strange attractor from a corresponding time series has attracted considerable attention in the past few years and has become one of the main tools in the analysis of the underlying dynamics. Of all types of dimensions, most attention has been given to the *correlation dimension* (D_c). This is mainly because this type of dimension is easier to estimate than others and also because it provides a good measure of the complexity of the dynamics, i.e. of the number of active degrees of freedom.

First, we have to reconstruct the attractor on the phase space, introduced in Chapter 2. We divide the EEG into non-overlapping time blocks which is long enough for a good estimation of D_c, shown in Fig. 3-11(a). Each block contains 256 states of the attractor ($N_a = 256$), and the distance between each state is 9 ($\Delta t = 9$), shown as Fig. 3-11(b). The dimension of the phase space is p, and the delay $\tau = 1$, shown in Fig. 3-11(c).
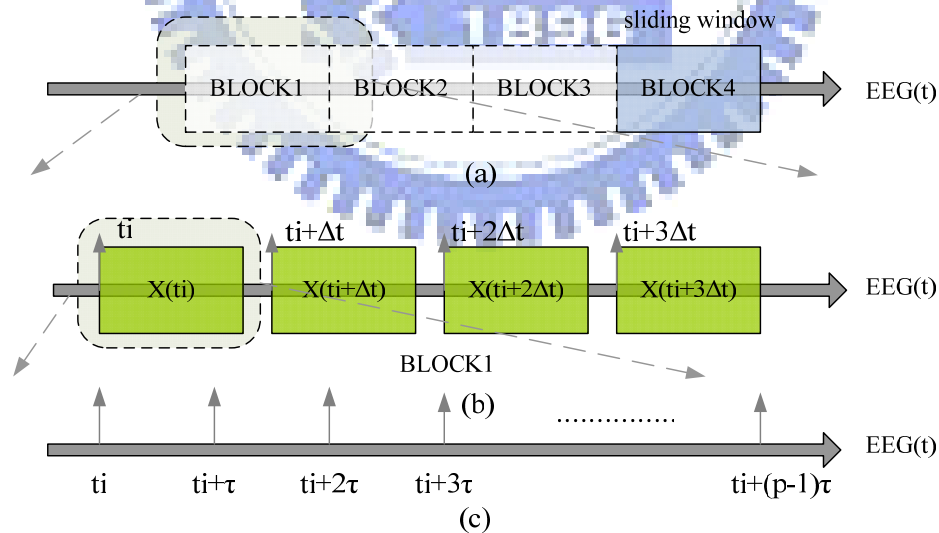
$$X(ti) = (x(ti), x(ti+\tau), x(ti+2\tau), \cdots, x(ti+(p-1)\tau)$$

Fig. 3-11 (a) Time blocks, (b) States, and (c) Embedding dimension

Grassberger and Proccacia suggested a procedure of estimating which became widely used by mathematicians and applied scientists immediately [5]. Fix a point x on the attractor A. Let $N_x(\varepsilon)$ denote the number of points on A inside a ball of radius ε about x (Fig. 3-12).
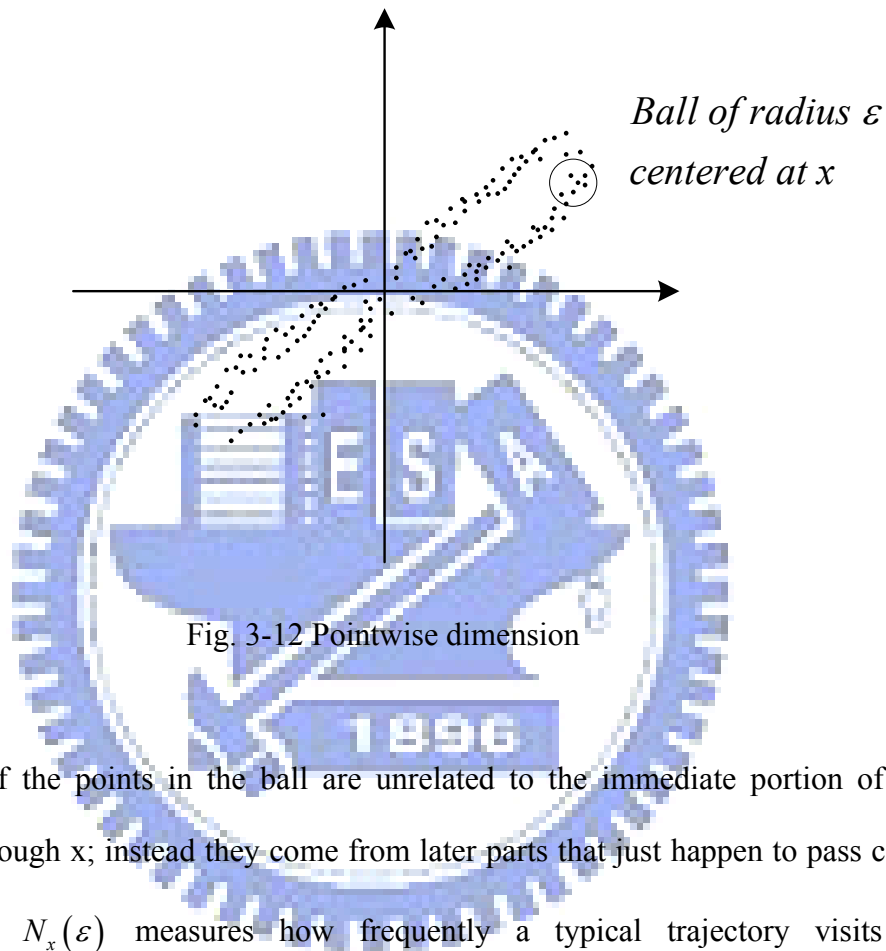

Ball of radius ε centered at x

Fig. 3-12 Pointwise dimension

Most of the points in the ball are unrelated to the immediate portion of the trajectory through x; instead they come from later parts that just happen to pass close to x. Thus $N_x(\varepsilon)$ measures how frequently a typical trajectory visits an ε-neighborhood of x.

Now vary ε. As ε increases, the number of points in the ball typically grows as a power law:

$$N_x(\varepsilon) \propto \varepsilon^d, \tag{3.15}$$

where d is called the pointwise dimension at x. The pointwise dimension can depend significantly on x; it will be smaller in rarefied regions of the attractor. To get an overall dimension of A, one averages $N_x(\varepsilon)$ over many x. The resulting quantity

$C(\varepsilon)$ is found empirically to scale as

$$C(\varepsilon) \propto \varepsilon^{D_c} \tag{3.16}$$

where D_c is called the correlation dimension.

If the relation $C(\varepsilon) \propto \varepsilon^{D_c}$ were valid for all ε, we'd find a straight line of slope D_c.

$$D_c = \lim_{\varepsilon \to 0} \lim_{N_a \to \infty} \frac{\ln C(\varepsilon, N_a)}{\ln \varepsilon} \tag{3.17}$$

where $C(\varepsilon)$ is the correlation integral and defined as follows [10]:

$$C(\varepsilon) = \frac{1}{N_a^2} \sum_{1 < i < j < N_a} l\left(\|X_i - X_j\| \le \varepsilon \right) \tag{3.18}$$

For distance $\| \cdot \|$ one usually takes the maximum norm, i.e. for a k-dimensional vector x, $\|x\| = max|x_i|$, $1 \le i \le k$, and l is the Heaviside function.
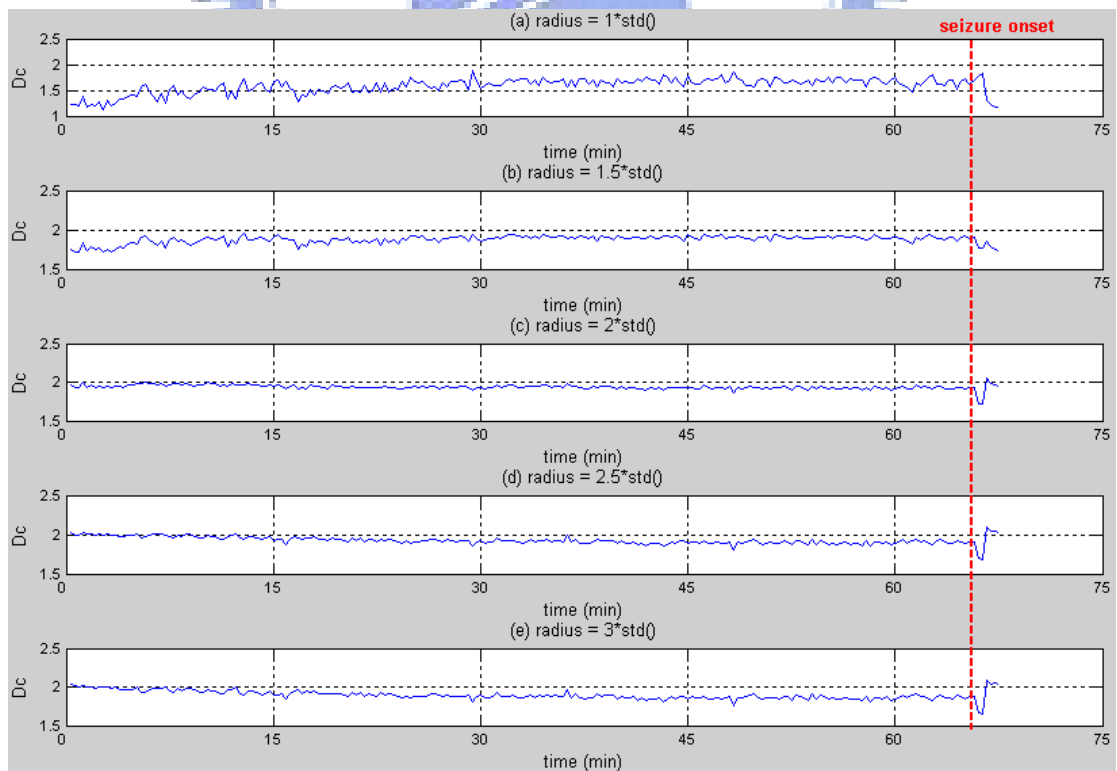

Fig. 3-13 Correlation dimension with different radius

It is known that when calculating the correlation dimension some restrictions are imposed on the value of ε. We use the standard deviation (std.) of each time block to estimate it as follows:

$$\varepsilon = k * std \ of \ the \ time \ block, \quad k=\{1, \ 1.5, \ 2, \ 2.5, \ 3\} \tag{3.19}$$

The results are shown in Fig. 3-13. When k is less than two, there is no clear trend of the wave before the seizure onset. With the increasing of k, we can find that there is a long-term decreasing before the seizure and a sudden drop during the ictal, and then a strong rise after the seizure. Recall the feature mentioned in the previous section, the EEG would decrease in the degrees of freedom during the pre-ictal. And we confirm this character again by correlation dimension.

Now, we have decided the basic parameters of the algorithm of correlation dimension. Next, we will calculate the D_c with different subbands of the EEG, including the unfiltered signal, L_1 band (0~63 Hz), H_1 band (64~128 Hz), LL_2 band (0~31 Hz), and LH_2 band (32~64 Hz), to find out the most appropriate one.
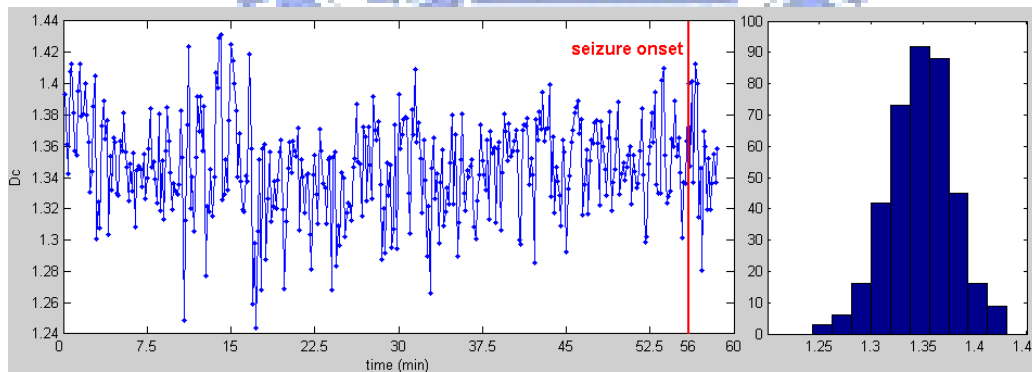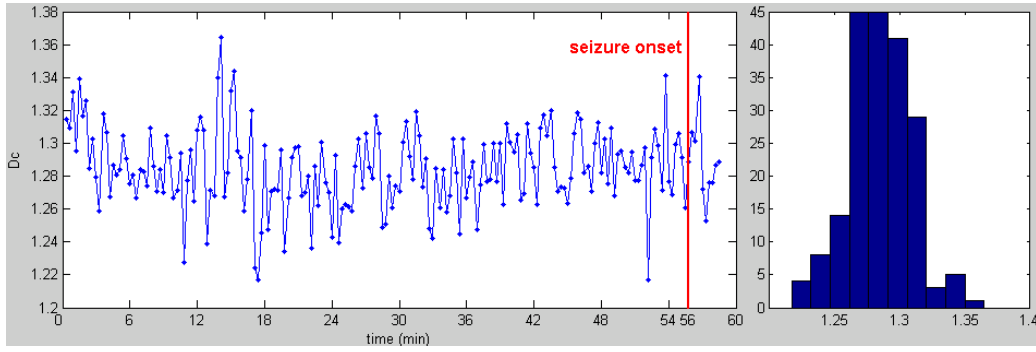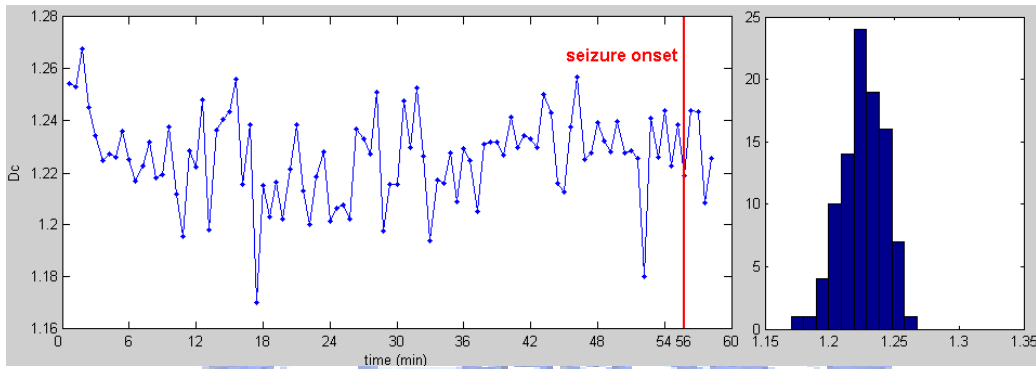

Fig. 3-14 Correlation dimension of unfiltered signal

(a)

(b)

Fig. 3-15 Correlation dimension of (a) L_1 and (b) LL_2 band

The results of unfiltered signal, L_1, and LL_2 are shown in Fig. 3-15 and Fig. 3-15. And the correlation dimensions of each band seem a normal distribution with a very narrow range (about 0.09~0.18). We can hardly find any features to identify the seizures.
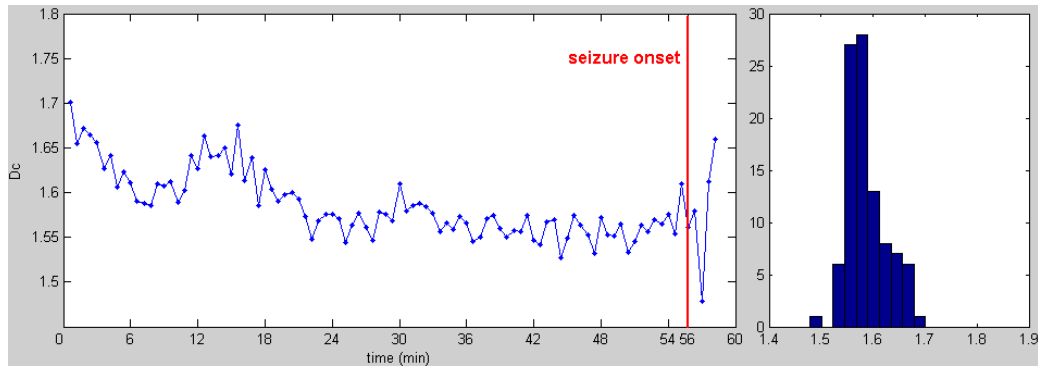
(a)
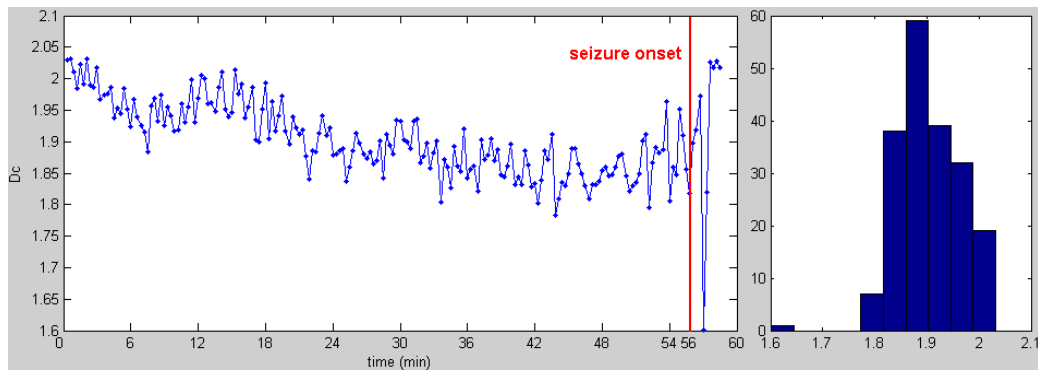
(b)

Fig. 3-16 Correlation dimension of (a) LH_2, and (b) H_1 band

The results of LH_2 and H_1 subbands are shown in Fig. 3-16. Obviously, correlation dimensions of each band decrease during the pre-ictal. The histograms mainly focus on the larger values reflecting sudden fallings at the onset of seizures. That is, the characteristics of the seizures likely embedded in the high frequency subbands not in low frequency ones.

In the following experiments, we will see the comparison of the correlation dimension between pre-ictal and inter-ictal.
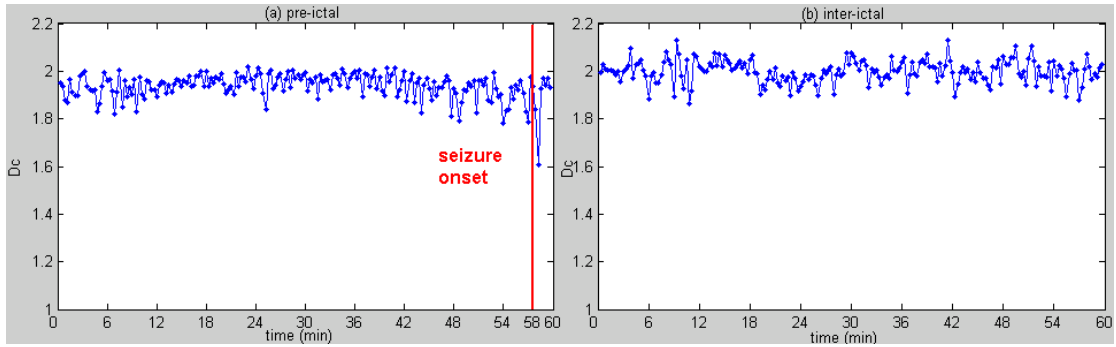
Fig. 3-17 Comparison result of patient1 between (a) pre-ictal and (b) inter-ictal
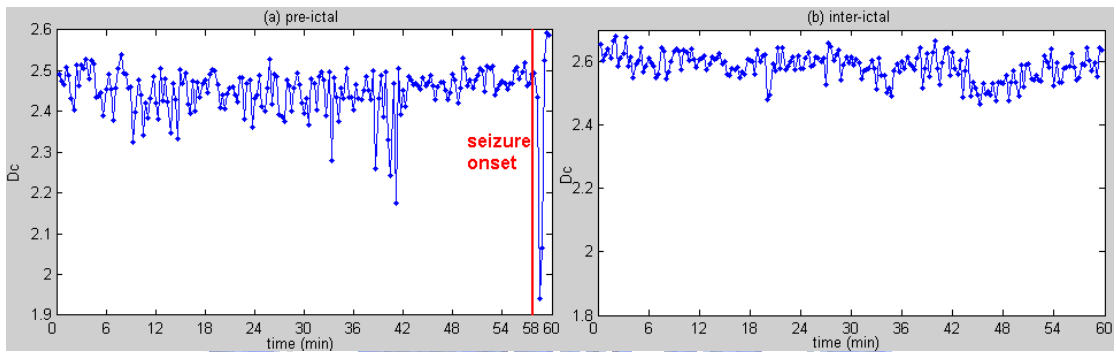
Fig. 3-18 Comparison Result of patient 10 between (a) pre-ictal and (b) inter-ictal

The results perfectly match the conclusions of STL. Although correlation dimension can reflect the loss in the degree of freedom, we need find out some clue before a seizure so that we can "predict" it. In the next section, we will introduce the concept of correlation coefficient to extract useful features.

3.4 Feature Extraction and Prediction Rules

In fact, there exists a relationship between the embedding dimension p and correlation dimension D_c. If, as p increases, D_c continues to rise then this is symptomatic of a stochastic system. If, however, the data generated by a deterministic process, then D_c will reach a finite limit at some relatively small p.

We suppose that when there is a transition from chaotic to periodic, the radius of convergence varies. Thus, we have tried to reconstruct the attractor with different

values of p from 7 to 25. We consider the correlation coefficient between two D_c using distinct embedding dimensions, p_1 and p_2, within a sliding window, denoted as CC_{p_1,p_2}.

$$CC_{p_1,p_2}(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} \quad (3.20)$$

where

$$X = (D_c(i), D_c(i+1),...,D_c(i+19)), D_c \; lay \; on \; p_1 - dimension$$
$$Y = (D_c(i), D_c(i+1),...,D_c(i+19)), D_c \; lay \; on \; p_2 - dimension \quad (3.21)$$
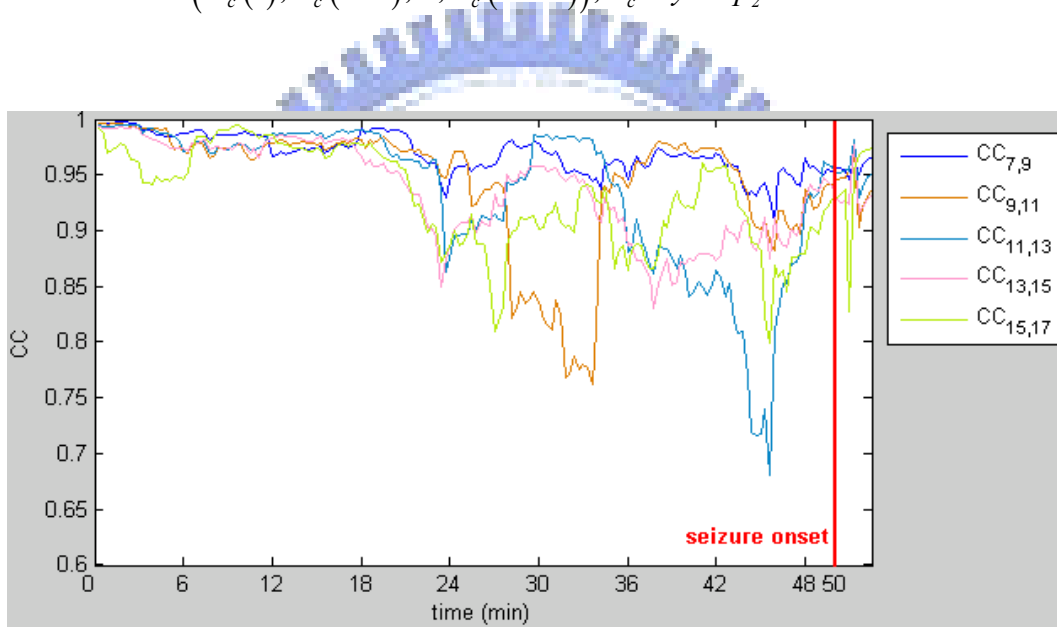
Fig. 3-19 Correlation coefficients between correlation dimensions in different embedding dimensions

Fig. 3-19 shows the results of $CC_{7,9}$, $CC_{9,11}$, …, and $CC_{15,17}$, and we can easily find that the correlation does decreases tremendously.

We evaluate the five different CCs with the following definition of prediction rule. We find the $CC_{11,13}$ is the best of all. Thus, the $CC_{11,13}$ is our final choice.
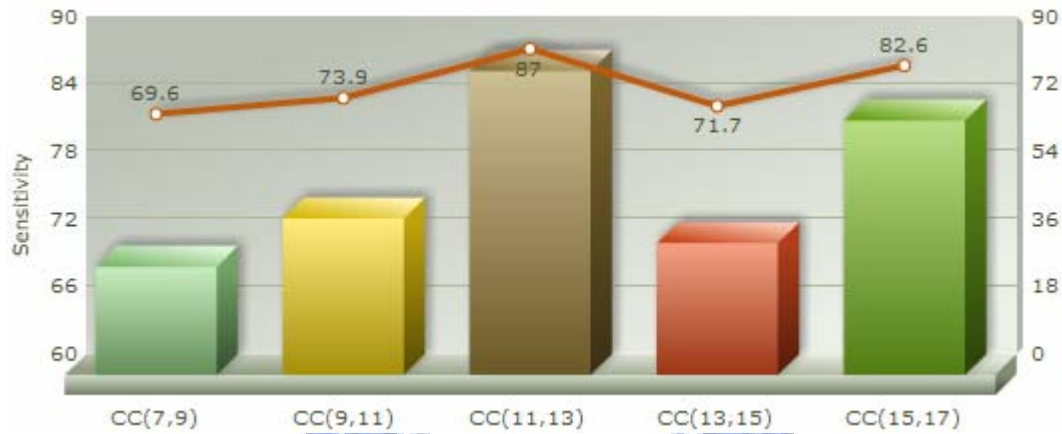
Fig. 3-20 The sensitivity with five different correlation coefficients

We define the timing to give an alarm if:

(1) The CC must transit from a value above threshold T_1 to a value below threshold T_2.

(2) The period of the CC below T_1 must longer than L.

Fig. 3-21 and Fig. 3-22 are the prediction results for patient 1 and patient 7, and the system gives an alarm about 7 and 37 minutes earlier than the seizure onset respectively. On the other hand, the two D_c are always highly correlated in inter-ictal cases. That is, the architecture based on wavelet decomposition, and correlation dimension could achieve the goal we expect.
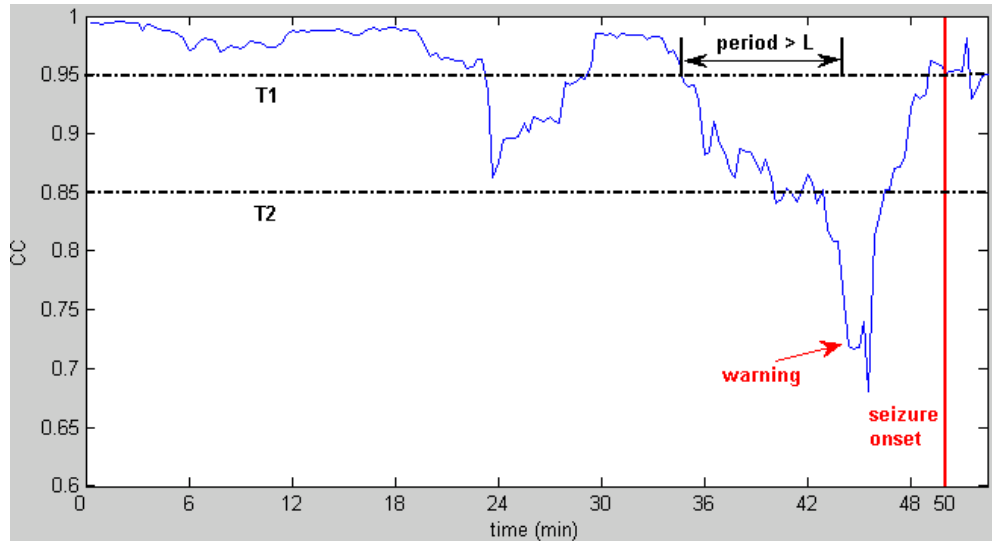
(a)
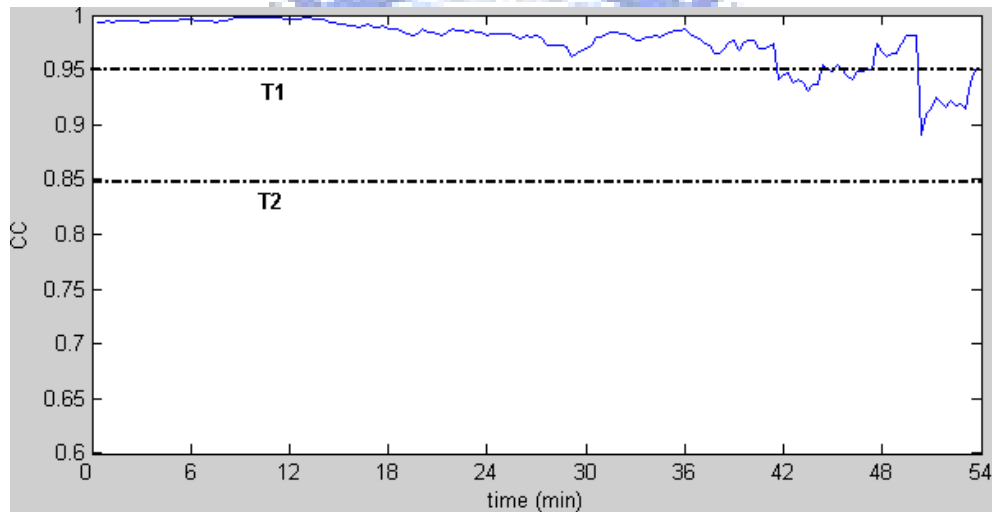
(b)

Fig. 3-21 Prediction results for patient 1 (a) pre-ictal, and (b) inter-ictal
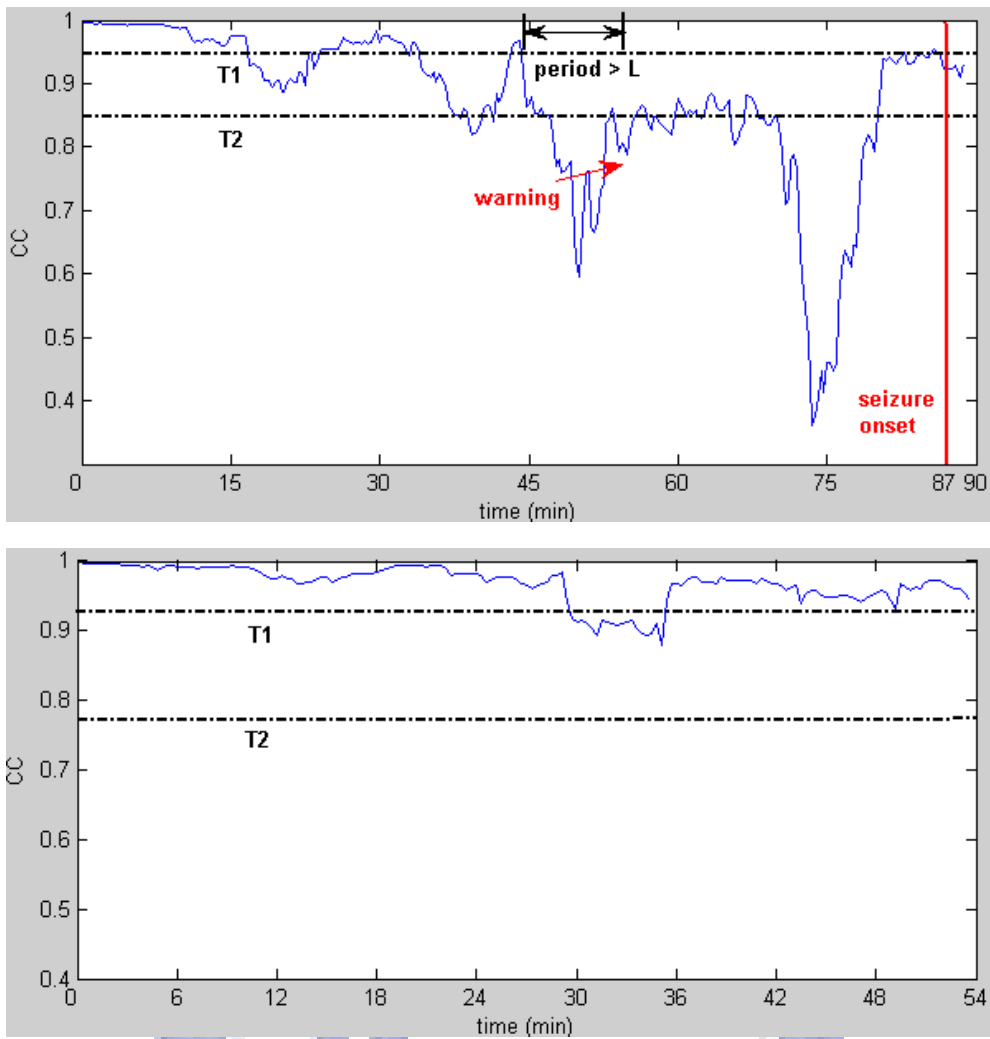
Fig. 3-22 Prediction results for patient 7 (a) pre-ictal, and (b) inter-ictal

More statistical information, such as prediction sensitivity and false prediction rate, and comparisons with other algorithms will be discussed in Chapter 5.

Chapter 4
VLSI Implementation and Verification

In the chapter, VLSI implementation for the components in the seizure chip is introduced. The system is designed for biomedical portable device, so we consider about not only the speed, but also the area and power consumption. Top level hardware architecture describes how the WCDSP play its role and corporate with VNS, shown in Fig. 4-1.
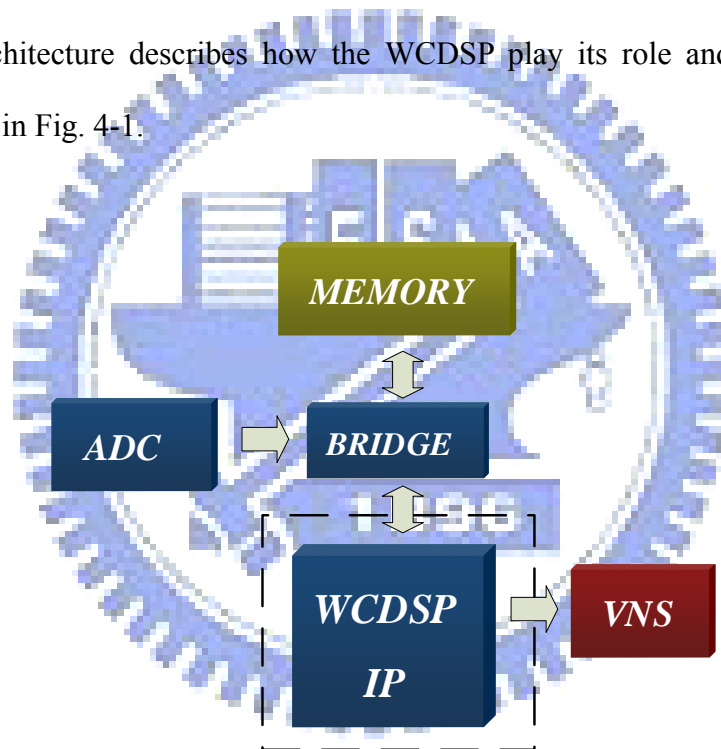
Fig. 4-1 Top level hardware architecture

4.1   Architecture of the Real-Time Seizure Circuit

The overall architecture shown in Fig. 4-2 is made up of two parts: the arithmetic functional unit, and system control unit.

Arithmetic function units mainly consist of wavelet, correlation dimension, correlation coefficient circuits. To satisfy the requirement of accessing the memory at

a high frequency when computing the correlation dimension, we need a more efficient memory access method. Therefore, a memory controller is represented as well to optimize the addressing of memory, reduce the hardware costs and the power consumption certainly.

The system control units set the system by control words from users. In order to saving the most power, we also need a power management to control the activation of each component.
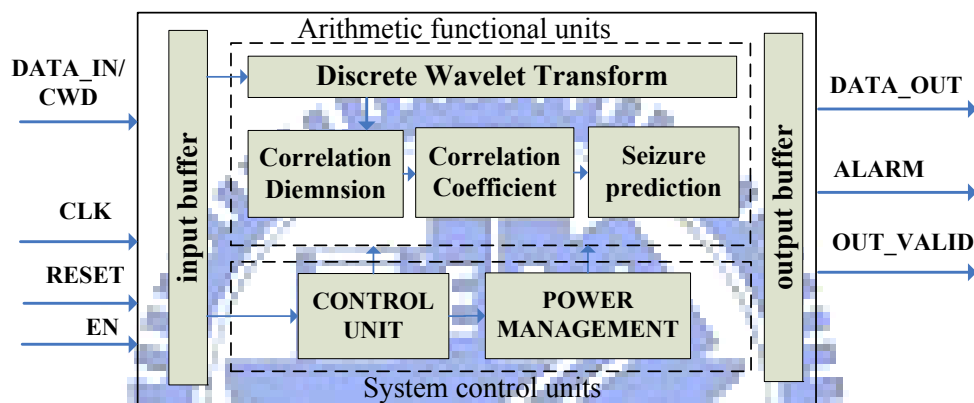
Fig. 4-2 WCDSP system architecture

4.2 Arithmetic Functional Units

In this section, implementations of primary function units will be discussed, such as using lifting wavelet to reduce the computation complexity, correlation dimension with an optimized memory controller, etc.

4.2.1 Lifting Wavelet

Various techniques to construct wavelet bases, or to factor existing wavelet filters into basic building blocks are known. One of these is lifting. The original motivation for developing lifting was to build second generation wavelets. First generation wavelets are all translates and dilates of one or a few basic shapes; the

Fourier transform is then the crucial tool for wavelet construction. A construction using lifting, on the contrary, is entirely spatial and therefore ideally suited for building second generation wavelets when Fourier techniques are no longer available. When restricted to the translation and dilation invariant case, or the "first generation," lifting comes down to well-known ladder type structures and certain factoring algorithms.

The basic idea of wavelet transforms is to exploit the correlation structure present in most real life signals to build a sparse approximation. The correlation structure is typically local in space (time) and frequency; neighboring samples and frequencies are more correlated than ones that are far apart. Traditional wavelet constructions use the Fourier transform to build the space-frequency localization. However, as the following simple example shows, this can also be done in the spatial domain.

Consider a signal $X = (x_k)$. Let us split it in two disjoint sets which are called the polyphase components: the even indexed samples $X_e = (x_{2k})$, or "evens" for short, and the odd indexed samples $X_o = (x_{2k+1})$, or "odds." Typically these two sets are closely correlated. Thus it is only natural that given one set, e.g., the odd, one can build a good predictor P for the other set, e.g., the even. Of course, the predictor does not need to be exact, so we need to record the difference or detail d :

$$d = X_o - P(X_e)$$

Given the detail d and the odd, we can immediately recover the odd as

$$X_o = P(X_e) + d$$

The operation of computing a prediction and recording the detail is called a *lifting step*. This idea connects naturally with wavelets as follows. The prediction steps can

take care of some of the spatial correlation, but for wavelets we also want to get some separation in the frequency domain. The frequency separation is poor since X_e is obtained by simply subsampling so that serious aliasing occurs. In particular, the running average of the X_e is not the same as that of the original samples X. To correct this, we propose a second lifting step, which replaces the evens with smoothed values s with the use of an update operator U applied to the details as

$$S = X_e + U(d)$$

Again this step is trivially invertible as

$$X_e = s - U(d)$$

and then X_o can be recovered as explained earlier. This illustrates one of the built-in features of lifting: no matter how P and U are chosen, the scheme is always invertible and thus leads to critically sampled perfect reconstruction filter banks. The block diagram of the two lifting steps is given in Fig. 4-3.
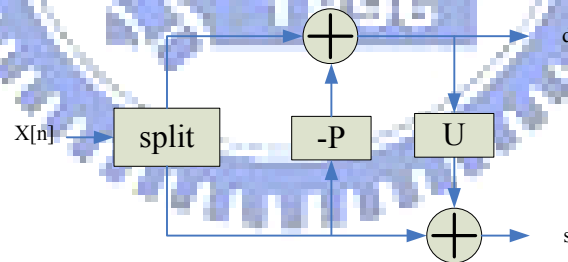
Fig. 4-3 Block diagram of predict and update lifting steps.

A natural question now is how much of the first generation wavelet families can be built with the lifting framework. It turns out that every FIR wavelet or filter bank can be decomposed into lifting steps. This can be seen by writing the transform in the polyphase form. Statements concerning perfect reconstruction or lifting can then be made using matrices with polynomial. A lifting step then becomes a so-called

elementary matrix. It is a well known result in matrix algebra that any matrix with polynomial entries and determinant one can be factored into such elementary matrices.
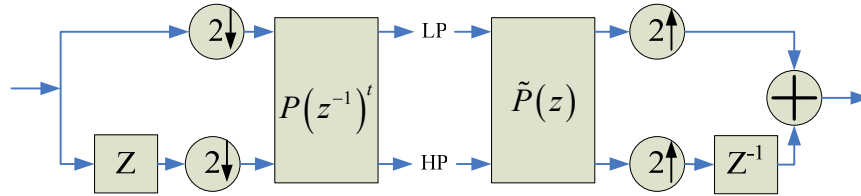
Fig. 4-4 Polyphase representation of wavelet transform

Supposing the analysis filters are $h(z)$, $g(z)$, and synthesis filters are $\tilde{h}(z)$, $\tilde{g}(z)$. The polyphase representation of a filter $\tilde{h}$ is given by

$$\tilde{h}(z) = \tilde{h}_e(z^2) + z^{-1}\tilde{h}_o(z^2)$$

where h_e contains the even coefficients, and h_o contains the odd coefficients:

$$\tilde{h}_e(z^2) = \frac{\tilde{h}(z) + \tilde{h}(-z)}{2} \quad and \quad \tilde{h}_o(z^2) = \frac{\tilde{h}(z) - \tilde{h}(-z)}{2\,z^{-1}}$$

We assemble the polyphase matrix as

$$\tilde{P}(z) = \begin{bmatrix} \tilde{h}_e(z) & \tilde{g}_e(z) \\ \tilde{h}_o(z) & \tilde{g}_o(z) \end{bmatrix}$$

We define $P(z)$ similarly. For the Daubechies 4 wavelet, the polyphase matrix is

$$P(z) = \tilde{P}(z) = \begin{bmatrix} h_0 + h_2 z^{-1} & -h_3 z^1 - h_1 \\ h_1 + h_3 z^{-1} & h_2 z^1 + h_0 \end{bmatrix}$$

The analysis polyphase matrix is factored as:

$$P(1/z)^t = \begin{bmatrix} \dfrac{\sqrt{3}+1}{\sqrt{2}} & 0 \\ 0 & \dfrac{\sqrt{3}-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ z^{-1} & 1 \end{bmatrix} \begin{bmatrix} 1 & \dfrac{\sqrt{3}}{4} + \dfrac{\sqrt{3}-2}{4}z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\sqrt{3} & 1 \end{bmatrix}$$

The corresponding circuit of db4 wavelet is shown in Fig. 4-5.
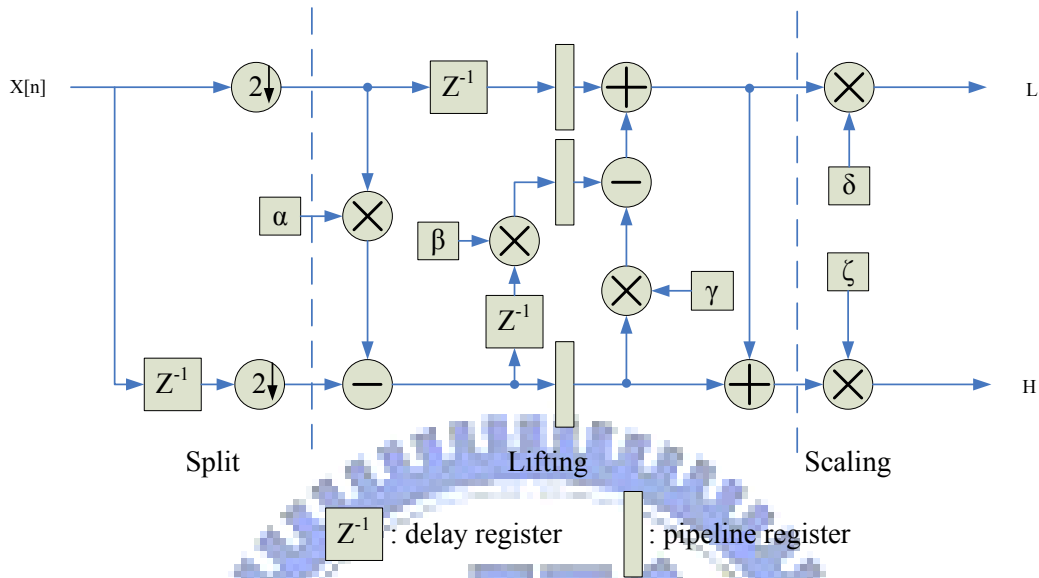
$$\alpha = -1.73201, \ \beta = 0.43301, \ \gamma = -0.06698, \ \delta = 1.93185, \text{ and } \zeta = 0.51763$$

Fig. 4-5 Pipelined lifting wavelet architecture of db4

Lifting wavelet transform is very suitable for VLSI implementation. As a comparison base we use the standard algorithm, which corresponds to applying the polyphase matrix. This already takes advantage of the fact that the filters will be subsampled and thus avoids computing samples that will be subsampled immediately. The unit we use is the cost, measured in number of multiplications and additions. Take db4 wavelet for example. The cost of applying a filter h is $|h|$ multiplications and $|h-1|$ additions. The cost of the standard algorithm thus is

$$2(|h|+|g|)-2 = 2(4+4)-2 = 14$$

However, by using the lifting wavelet we need only 9 units. It reduces the complexity of about 56%.

In the next, we consider the accuracy of the computation. Floating-point adders

and multipliers definitely perform the best, but they require a lot of hardware resource and they are usually slower. We use 16-bit fixed-point numeric, which consist of two bits for the integer part, and fourteen bits for the fractional part. The rough precision of the fixed point number representation can reach 0.000030518.

Since the coefficients of the lifting are constants and represented by fixed-point numeric, multiplication can be realized by shifts generating partial products that subsequently are added together. The number of nonzero partial products is determined by the number of nonzero bits in the multiplier coefficient. Consequently, if the number of nonzero bits in a coefficient is reduced, the number of partial products is reduced and therefore a smaller, faster, and less power consuming summation can be achieved. If a signed digit representation with the digits −1, 0, and 1 is used for the coefficient the number of partial products may be reduced. The canonical signed digit code (CSD) is a signed digit representation with minimal Hamming weight, i.e., it has a minimum number of ones, and contains no adjacent nonzero digits.

The conversion of a two's complement number into CSD code is done according to Table 4-1.

Table 4-1 Two's complement to CSD conversion

b_{i+1}	b_i	c_i	a_i	c_{i+1}
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	0	0
1	0	1	-1	1
1	1	0	-1	1
1	1	1	0	1

The b_i is the bit of the two's complement number to be converted and the a_i is the CSD code after the conversion. The c_i is the carry generated in step $i-1$ and c_{i+1} is the carry out at step i.

The coefficients converted to CSD and the circuits are listed below:

(1)　$\alpha = -1.73201_{10} = 10.01000100100111_{2's} = \bar{1}0.010001001010 0\bar{1}_{CSD}$

$$y = \bar{1}0.010001001010 0\bar{1} * x$$
$$= (-2^2 + 2^{-2} + 2^{-6} + 2^{-9} + 2^{-11} - 2^{-14})x$$
$$= -(x << 2) + (x >> 2) + (x >> 6) + (x >> 9) + (x >> 11) - (x >> 14)$$
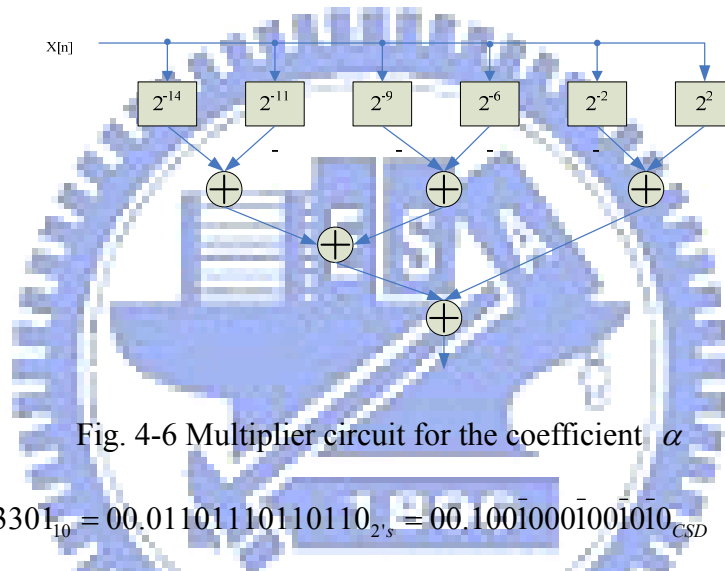

Fig. 4-6 Multiplier circuit for the coefficient α

(2)　$\beta = 0.43301_{10} = 00.01101110110110_{2's} = 00.100\bar{1}000\bar{1}00\bar{1}0\bar{1}0_{CSD}$

$$y = 00.100\bar{1}000\bar{1}00\bar{1}0\bar{1}0 * x$$
$$= (2^{-1} - 2^{-4} - 2^{-8} - 2^{-11} + 2^{-13})x$$
$$= (x >> 1) - (x >> 4) - (x >> 8) - (x >> 11) + (x >> 13)$$
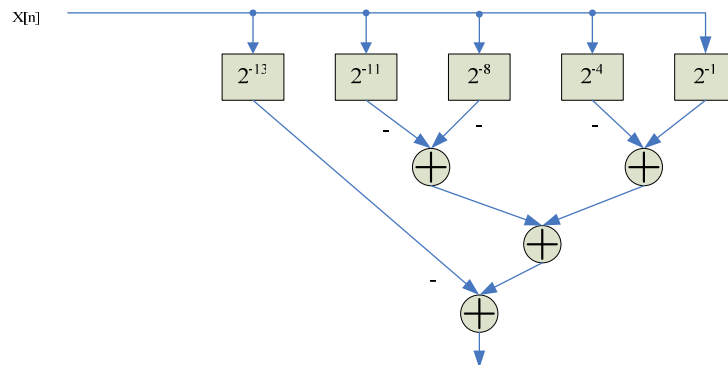
Fig. 4-7 Multiplier circuit for the coefficient β

(3) $\gamma = -0.06698_{10} = 11.11101110110111_{2's} = 00.000\bar{1}000\bar{1}00\bar{1}00\bar{1}_{CSD}$

$y = 00.000\bar{1}000\bar{1}00\bar{1}00\bar{1} * x$

$\quad = (-2^{-4} - 2^{-8} - 2^{-11} - 2^{-14})x$

$\quad = -(x >> 4) - (x >> 8) - (x >> 11) - (x >> 14)$
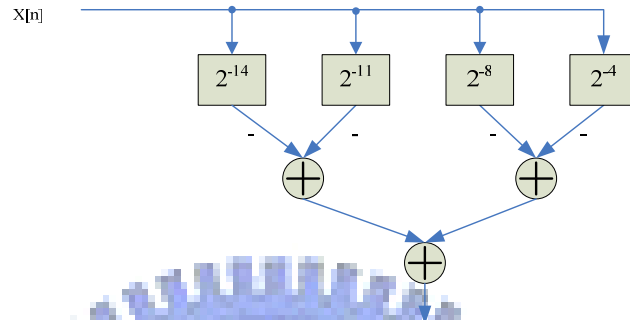
Fig. 4-8 Multiplier circuit for the coefficient γ

(4) $\delta = 1.93185_{10} = 01.11101110100011_{2's} = 10.000\bar{1}00\bar{1}0100010\bar{1}_{CSD}$

$y = 10.000\bar{1}00\bar{1}0100010\bar{1} * x$

$\quad = (2^2 - 2^{-4} - 2^{-7} + 2^{-9} + 2^{-12} - 2^{-14})x$

$\quad = (2^2 x - 2^{-4} x) - (2^{-7} x - 2^{-9} x) + (2^{-12} x - 2^{-14} x)$

$\quad = (2^2 x - 2^{-4} x) - (2^{-7}(x - 2^{-2} x)) + (2^{-12}(x - 2^{-2} x))$

$\quad = (x << 2) - (x >> 4) - (x_2 >> 7) + (x_2 >> 12)$

$x_2 = x - (x >> 2)$
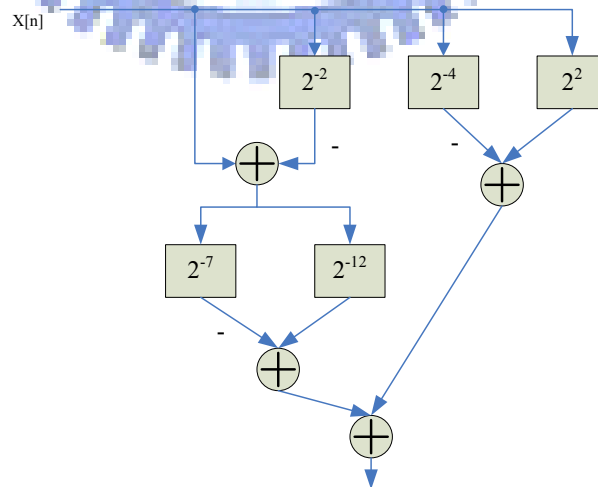
Fig. 4-9 Multiplier circuit for the coefficient δ

(5) $\zeta = 0.51763_{10} = 11.01111011100000_{2's} = 00.\bar{1}0000\bar{1}00\bar{1}00000_{CSD}$

$$y = 00.\bar{1}0000\bar{1}00\bar{1}00000 * x$$
$$= -(2^{-1} + 2^{-6} + 2^{-9})x$$
$$= -(x >> 1) - (x >> 6) - (x >> 9)$$
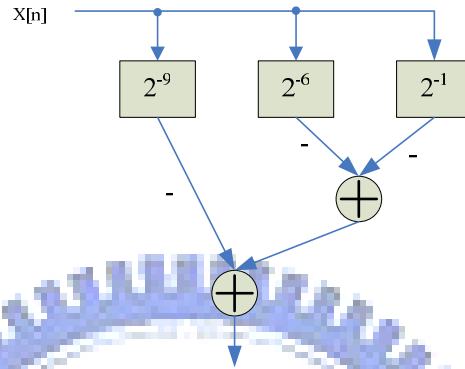
Fig. 4-10 Multiplier circuit for the coefficient ζ

As a result, we need only twenty four instead of forty two adders and shifters.

The hardware costs are about 42% less than using 2's complement representation.

4.2.2 Correlation Dimension
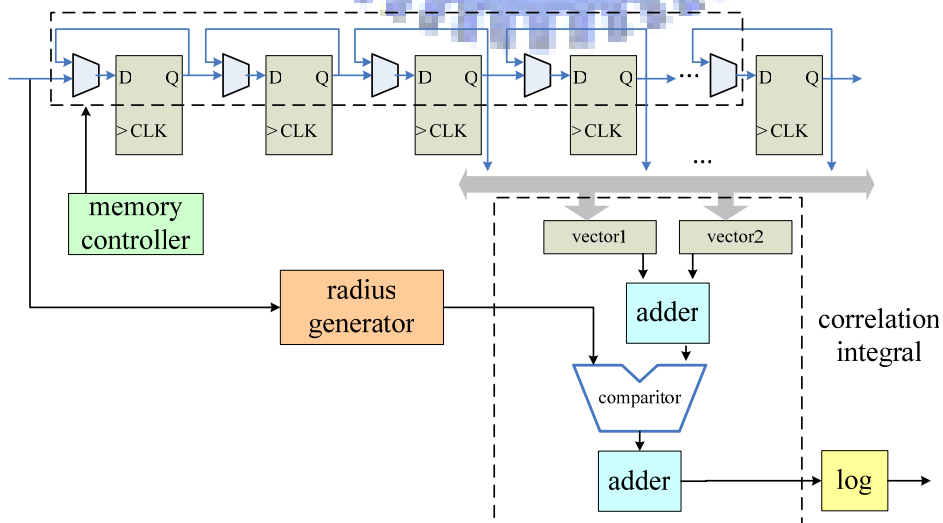
Fig. 4-11 Correlation dimension circuit

The correlation dimension circuit is composed of three components, including memory controller, radius generator, and correlation integral (Fig. 4-11) and operations in detail is shown in Fig. 4-12.
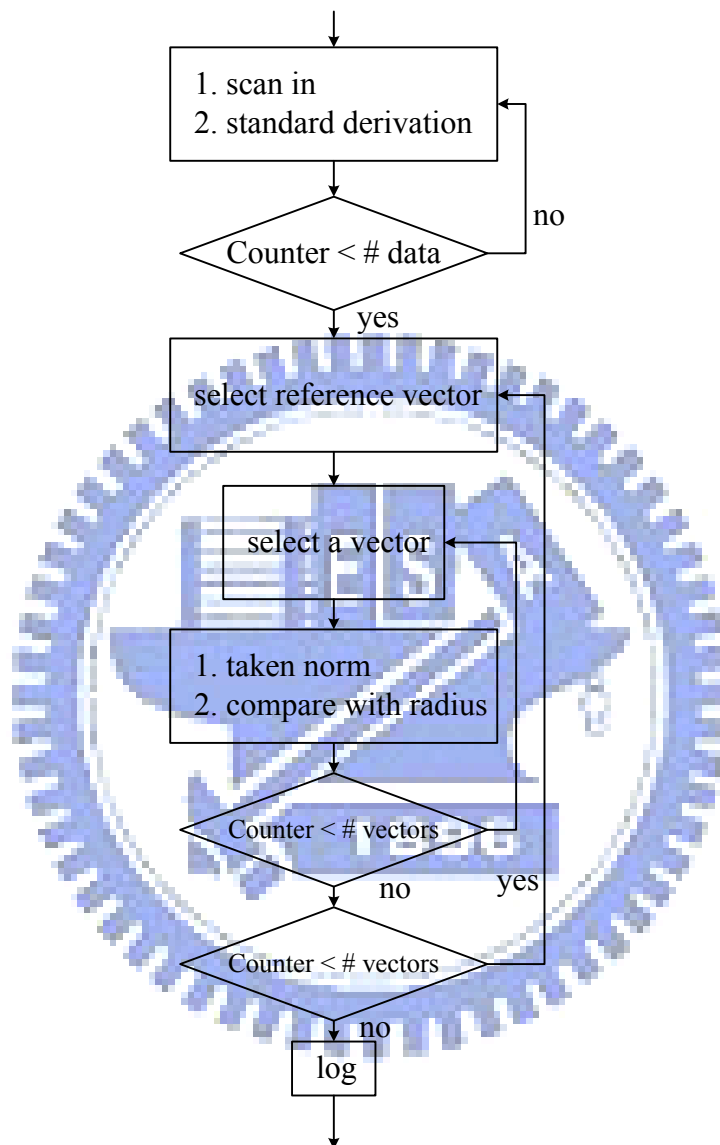
Fig. 4-12 Block diagram of correlation dimension

First, we scan in all data we need in a register chain, and calculate the standard deviation used for radius generation at the same time. Then calculate the taken norm pair by pair and compare with the radius. At last, we obtain the correlation dimension by log operation.

(1) Memory controller

There are two tasks of the controller. The one is to reduce the memory access costs. The memory used in the circuit needs not random access. If we use conventional memory addressing, it may:

➢ Infer a very large decoder when scanning in the data, and

➢ Infer a lot of (at least 13) large multiplexers when generating vectors.

Because the access is well-regulated, we can achieve it more easily by the following way:

➢ When scanning in the data:

Step1: put the data at the tail of the memory

Step2: shift left the whole memory by one unit until the data is all scanned in

➢ When generating vectors (suppose embedding dimension=13, $\Delta t = 9$):

Step1: select the first consecutive thirteen data as the vector

Step2: shift left the memory by nine units

Step3: return to step1 until the last loop

Step4: fix the positions of the data in the memory to the beginning state
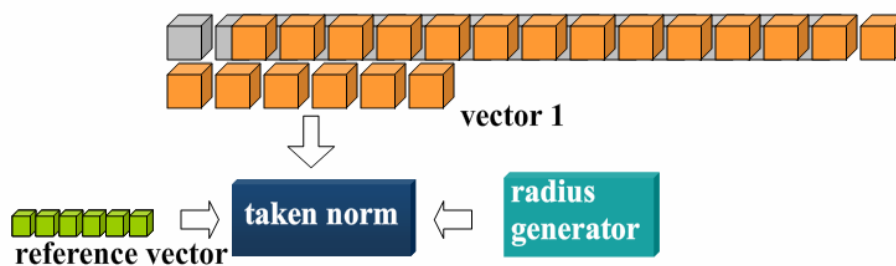
Fig. 4-13 Enhanced memory addressing

Therefore we can avoid a great amount of redundant hardware costs. The other task of the controller is the reconfiguration when applying different embedding dimensions, and then we can share the same circuit.

(2) Radius generator

When we generate the radius ε, standard deviation (std.) is used as mentioned before. But the original formula does not suitable for hardware implementation.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (4.1)$$

It can be divided into three stages:

Stage1: calculate sum for the mean

Stage2: calculate the difference $(x_i - \bar{x})$ and summation

Stage3: obtain the standard deviation

By this way, we find the stage1 overlaps the time of data scan_in, but the stage2 needs another cycles. And the system waits for it to go on.
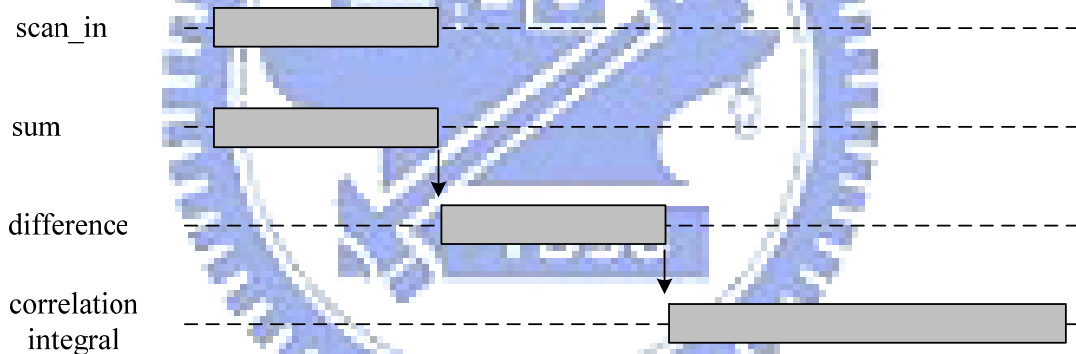
Fig. 4-14 Timing diagram for the original std. formula

To avoid this situation, we adjust the formula as follows:

$$\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \left(\sum_{i=1}^{N}x_i^2\right) - \left(2\bar{x}\sum_{i=1}^{N}x_i\right) + N\bar{x}^2$$

$$= \left(\sum_{i=1}^{N}x_i^2\right) - \left(2\bar{x}N\bar{x}\right) + N\bar{x}^2 \qquad (4.2)$$

$$= \left(\sum_{i=1}^{N}x_i^2\right) - N\bar{x}^2$$

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right) - \overline{x}^2} = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N}x_i\right)^2} \qquad (4.3)$$

It can be divided into two stages:

Stage1: calculate sum of x_i^2 and square of sum

Stage2: obtain the standard deviation

By this way, we can find the std. totally overlaps the time of data scan_in. Therefore we don't need any redundant cycles, and speed up the calculation of correlation dimension.
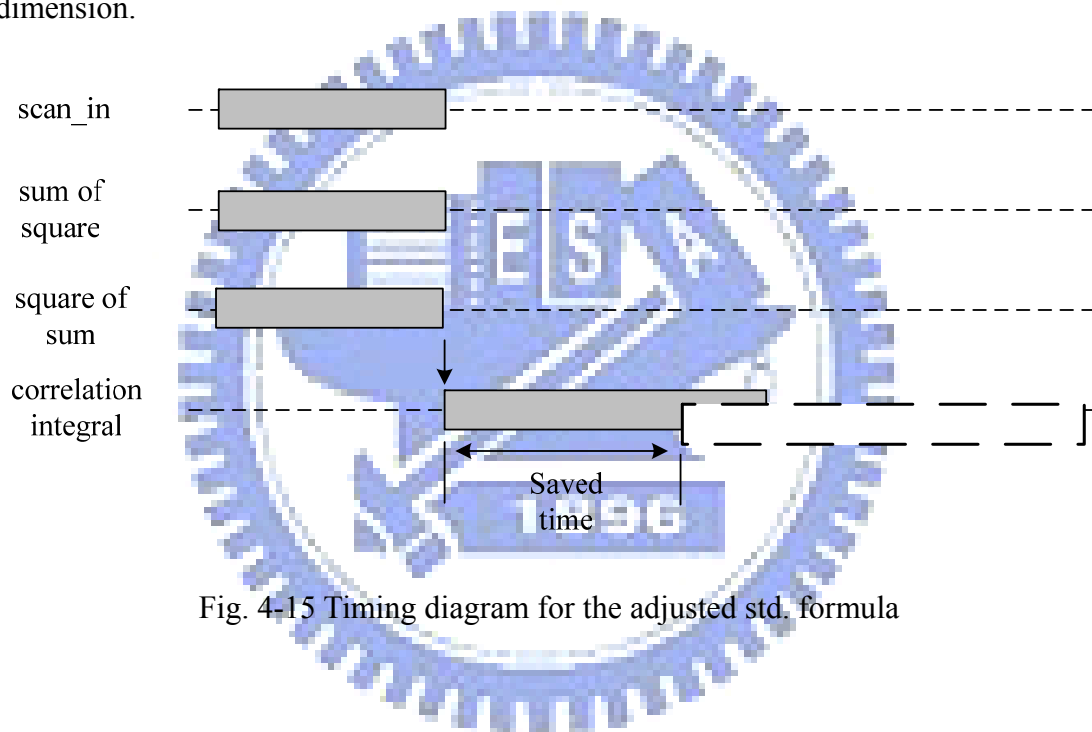
Fig. 4-15 Timing diagram for the adjusted std. formula

4.3 System Controller

The WCDSP system controller takes charge of two major control signals. One is the mode switching signal, and the other is control word signal.

In the architecture, we need two sets of wavelet and correlation dimension circuit for two embedding dimensions respectively. However, most part of the two sets of circuit is similar. Thus, we can slightly modify the two kinds of circuit to be applied to two different embedding dimensions.

The signal 'mode' is used to configure the circuit which embedding dimension

should be. When the mode='0', the embedding dimension is 11, and the one is 13 otherwise.

For example, we use two sets of register banks in one wavelet blocks to keep the values of pipeline registers and delay registers instead of two blocks. And we switch the register banks through mode signal (Fig. 4-16).
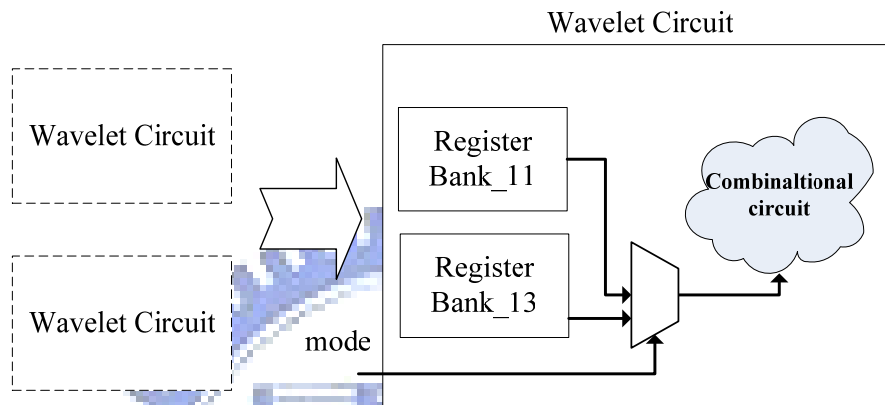
Fig. 4-16 Modified wavelet circuit with two register banks

For correlation dimension circuit, we need to increase the number of registers, and set the memory controller to adjust the shift operation, radius generator and comparator to configure the number of loops, as shown in Fig. 4-17.
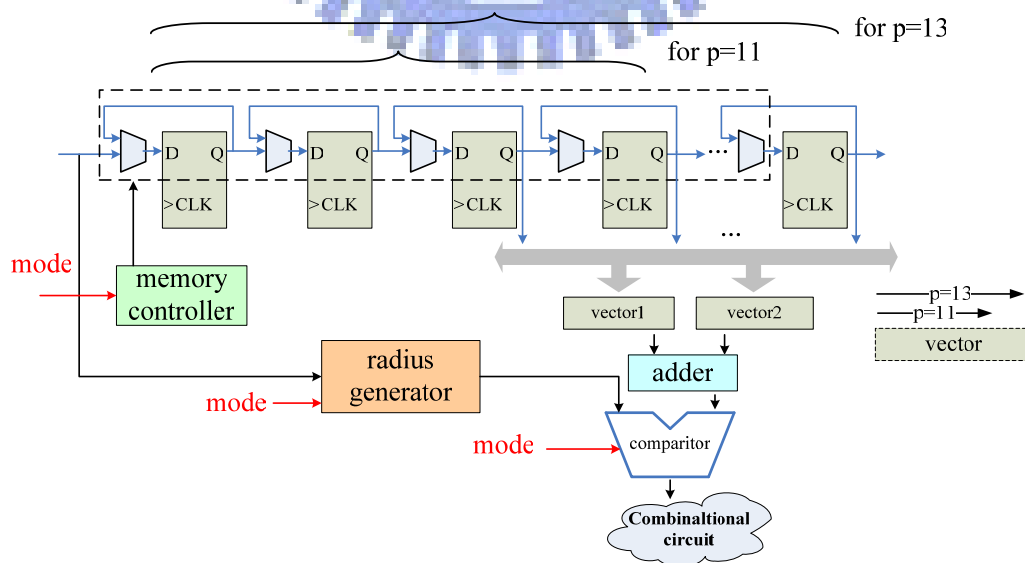
Fig. 4-17 Modified correlation dimension circuit

64

The other kind of signal decoded by system controller is control word signal. For the convenience of analysis and reconfiguration of the algorithm, we can probe the results of the internal components and applied to the new algorithm by CPU. Thus we set the control word to tell the controller which results should be scented. The detail description is in Table 4-2.
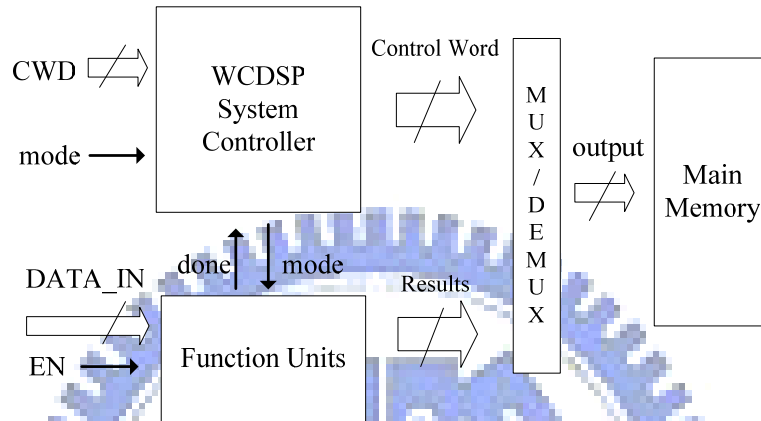
Fig. 4-18 Block diagram of system controller

Table 4-2 Description of control word

EN	CWD	OUTPUT
0	X	invalid
1	00	Result of seizure prediction
1	01	Result of correlation dimension
1	10	Result of wavelet (high frequency)
1	11	Result of wavelet (low frequency)

4.4 Simulation and Verification

We develop the algorithm by Verilog HDL which is commonly adopted. In the design, we use Synopsys Design Compiler with 90nm UMC/Faraday Design Kit for synthesis, and pre-simulation with ModelSim. In this section, we will show the simulation results in each component. It contains area report, timing report, and simulation report.

4.5.1 Simulation in Wavelet Circuit

Fig. 4-19 shows the gate-level simulation of wavelet circuit. The signals ca_o and cd_o are the low-frequency coefficients and high-frequency coefficients respectively. And we use two sets of registers to keep the values in different modes.
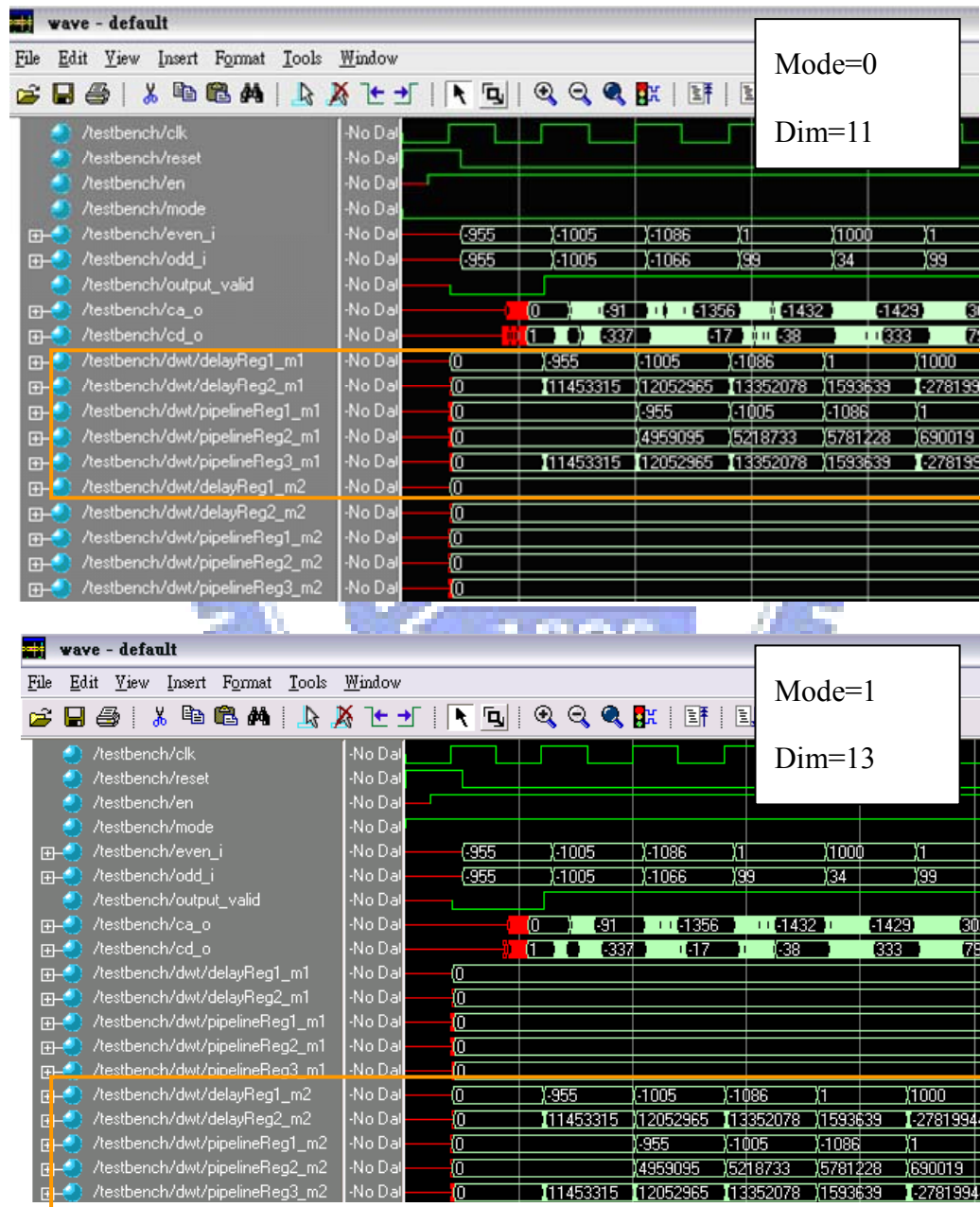
Fig. 4-19 Wavelet gate-level simulation with two modes

4.5.2 Simulation in Correlation Dimension Circuit

Fig. 4-20 shows the waves of correlation dimension circuit for the first state. The state READ_IN takes 2306 (for embedding dimension=11)/2308 (for embedding dimension=13) cycles, and we can see the calculation time of standard deviation overlaps the READ_IN state.
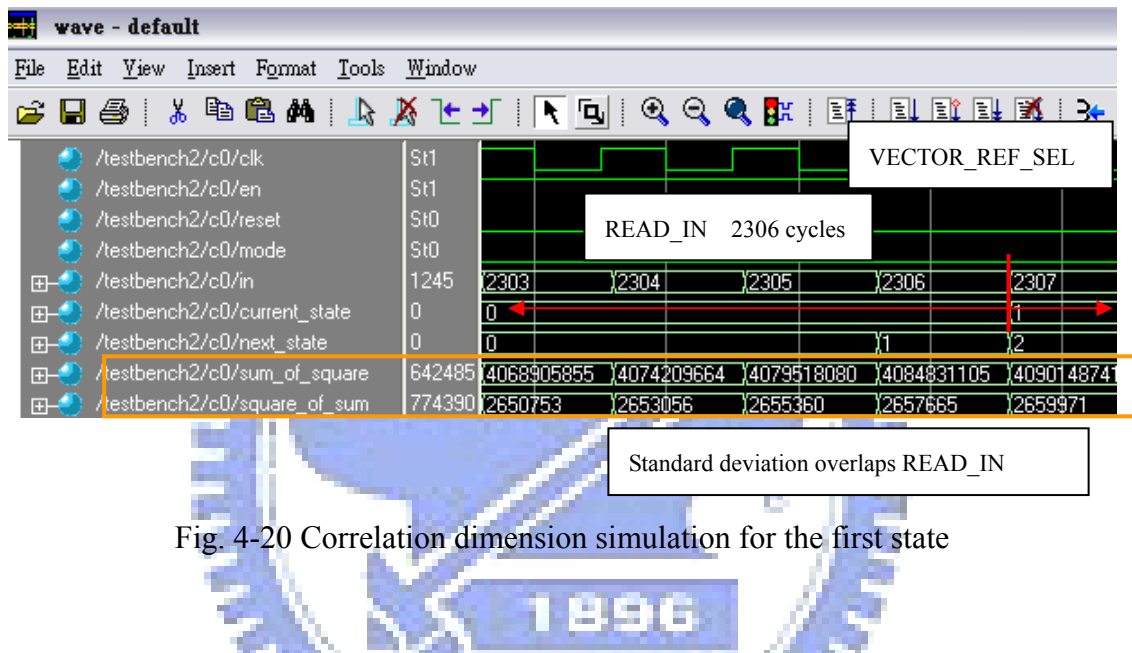
Fig. 4-20 Correlation dimension simulation for the first state
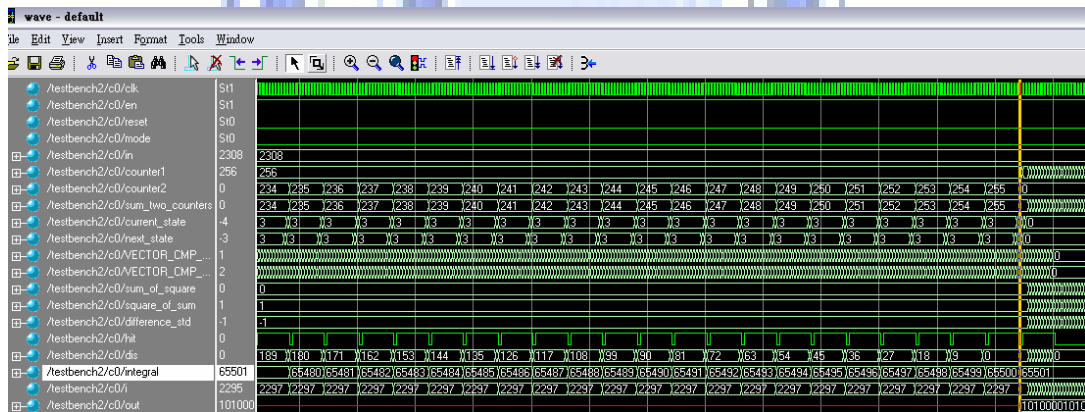
Fig. 4-21 Correlation dimension simulation results

Chapter 5
Experimental Results

In this chapter, we will show the experiment results of the algorithms, and the comparison with other important algorithms in recent years.

5.1 EEG Data and Patient Characteristics

The EEG data that we use are invasive EEG recordings of 11 patients suffering from medically intractable temporal lobe epilepsy. The data were recorded during an invasive pre-surgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany (http://www.fdm.uni-freiburg.de/EpilepsyData/).

In eight patients, the epileptic focus was located in neocortical brain structures, in two patients in the hippocampus, and in one patient in both. In order to obtain a high signal-to-noise ratio, fewer artifacts, and to record directly from focal areas, intracranial grid-, strip-, and depth-electrodes were utilized. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analogue-to-digital converter. Notch or band pass filters have not been applied.

For each of the patients, there are datasets called "ictal" and "interictal", the former containing files with epileptic seizures and at least 50 min pre-ictal data. the latter containing approximately 24 hours of EEG-recordings without seizure activity. At least 24 h of continuous interictal recordings are available for eight patients. For the remaining patients interictal invasive EEG data consisting of less than 24 h were joined together, to end up with at least 24 hours per patient (Table 5-1).

Table 5-1 Patient characteristics

Patient	Sex	Age	Seizure type	H/NC	Origin	Electrodes	#Seizures	Interictal (h)
1	f	15	SP, CP	NC	Frontal	g, s	4	24
2	m	14	SP, CP	NC	Frontal	g, s	5	24
3	f	42	SP, CP, GTC	H	Temporal	d	3	25
4	f	32	SP, CP	NC	Frontal	g, s	2	24
5	f	10	SP, CP, GTC	NC	Parietal	g, s	4	24
6	f	42	SP, CP, GTC	H	Temporal	d, g, s	4	25
7	f	41	CP, GTC	H and NC	Frontal/ Temporal	d, s	4	24
8	m	28	SP, CP, GTC	NC	Temporal	s	5	24
9	f	25	SP, CP	NC	Frontal	s	5	25
10	m	33	SP, CP, GTC	NC	Tempo/Parietal	d, s	5	26
11	m	13	SP, CP	NC	Temporal	s	5	24

Seizure types and location: simple partial (SP), complex partial (CP), generalized tonic-clonic (GTC), hippocampal (H), neocortical (NC).

Electrodes: grid (g), strip (s), depth (d).

5.2   Seizure Prediction Statistics

5.2.1   Terminology

A seizure prediction method has to forecast an impending epileptic seizure by raising an alarm in advance of the seizure onset. A perfect prediction method indicates the exact point in time when a seizure occurs. This ideal behavior is not expected for current prediction methods that analyze EEG data. The uncertainty can be considered by use of the seizure occurrence period, SOP, which is defined as a time period during which the seizure is to be expected (Fig. 5-1). In addition, to permit a therapeutic intervention, a minimum window of time between the alarm raised by the prediction method and the beginning of SOP is essential. This window of time is called the seizure prediction horizon, SPH.

Taking into account the two time periods SPH and SOP, a correct prediction is

defined as follows: after the alarm signal, during SPH, no seizure has occurred yet. During SOP, a seizure occurs.
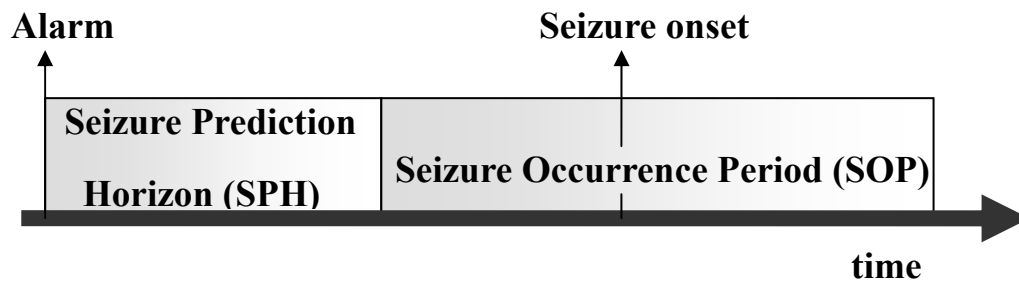
Fig. 5-1 Defining the SPH and SOP

Thus, the sensitivity was defined as the number of seizures predicted divided by the total number of seizures recorded.

$$Sensitivity = \frac{the\ number\ of\ seizures\ predicted}{total\ number\ of\ seizure\ recorded} \tag{5.1}$$

The False Positive Rate (FPR) was defined as the average number of warnings that no seizure occurs within SOP after SPH per hour.

$$FPR = \frac{the\ number\ of\ warnings\ that\ no\ seizure\ occurs}{EEG\ recording\ length\ (h)} \tag{5.2}$$

5.2.2 Seizure Prediction Results

To evaluate the WCDSP, a prediction was considered to be true if a seizure occurred within 1 h after a warning was observed and false otherwise. That is, a time horizon of 1-h period was chosen for the evaluation of the prediction results of the algorithm. Long time horizons obviously improve the sensitivity of the algorithm but also increase the uncertainty about the exact time of the next seizure. Short time horizons decrease the sensitivity and specificity.

Table 5-2 Performance of WCDSP for the optimal setting over all patients
(SPH = 1h, SOP = 18s, p1 = 11, p2 = 13, threshold1 = 0.95, threshold2 = 0.85,
interval = 33)

Patient	Sensitivity	False Positive Rate (False pre hr.)	Average Prediction Time (min)
1	4/4	0.3	14
2	5/5	0.6	34
3	2/3	0.3	20
4	2/2	0.2	28
5	4/4	0.6	29
6	3/4	0.2	28
7	3/4	0.1	20
8	4/5	0.1	33
9	4/5	0	22
10	5/5	0.4	30
11	4/5	0	40
Total	40/46 (86.96%)	0.2545	27

The algorithm was tested under two cases. In the first case, we evaluated a range of parameter settings (threshold1, threshold2, interval) to find the optimal result, when applied to all eleven patients (see Table 5-2). Under this condition, we obtain the parameter settings with p1 = 11, p2 = 13, threshold1 = 0.95, threshold2 = 0.85, and interval = 33. The sensitivity ranged from 66% (patient 3) to 100% (patient 2, etc.), with an average of 87% sensitivity overall. For example, the algorithm correctly predicted 75% (3/4) of seizures in patient 7 with FPR = 0.1/h, and 100% (4/4) of seizures in patient 1 with FPR = 0.3/h. On average, the algorithm gives alarms approximate 27 minutes before each seizure.

In the other case, we have the optimal parameter settings for each individual patient.

5.3　Comparison with other Prediction Methods

There have been few studies about seizure prediction algorithm. In the thesis, we will discuss with the difference between our proposed and others in Table 5-3.

In 2008, Bruno Direito1 [12] proposed an algorithm based on energy-wavelet with a sensitivity of 40% and false prediction rate of 0.4/h. V. Navarro's [13] and Le van Quyen's [1] algorithms using similar index achieve sensitivities of 83% and false prediction rates of 0.3/h in 2002 and 1999, respectively. In 2003, F. Mormann proposed an algorithm [2] based on synchronization decrease with a sensitivity of 81%. Maryann D'Alessandro et al. [14] presented a method of hybrid-feature with a sensitivity of 62.5% and false prediction rate of 0.27/h. Leon D. Iasemidis [7] proposed an algorithm by using short-term Lyapunov exponential to estimate the LLE, called ASPA, with a sensitivity of 84% and false prediction rate of 0.12/h.

Table 5-3 Comparison with other algorithms

	2008[12] B.D.'s	2002[13] V.N.'s	2003[2] F.M.'s	1999[1] Le V.Q.'s	2003[14] M.D.'s	2003[7] Iasemidis's	This Work
Algorithm	Energy Wavelet	Similarity Index	Sync. decrease	Similarity Index	Hybrid Feature	ASPA	WCDSP
Sensitivity	40%	83%	81%	83%	62.5%	84%	86.96%
False Positive Rate (False pre hr.)	0.4	0.3	N/A	N/A	0.2775	0.12	0.254
Prediction Time (min)	5.1	7.54	4 - 221	5.75	3.455	74.4	27
Number of Patients	19	11	18	13	4	5	11
Number of Seizures	17/42	34/41	26/32	19/23	N/A	42/50	40/46
Interictal (h)	N/A	12-60	49	N/A	N/A	70	110

Our proposed algorithm achieves a higher sensitivity of 86.69% with a slightly larger false prediction rate of 0.254 than others. Moreover we have the prediction time of 27 minutes witch is much longer than most of the others for a therapeutic intervention.
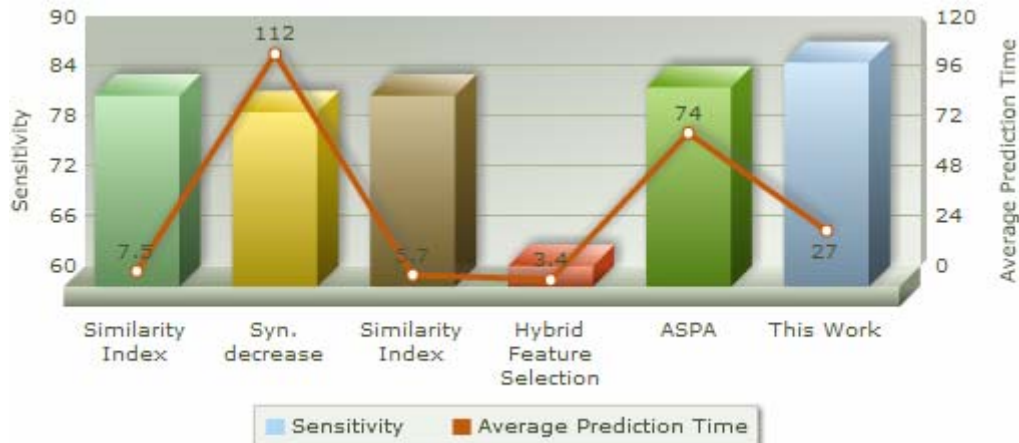

Fig. 5-2 Sensitivity and FPR comparisons with other algorithms

In Fig. 5-2, we show the comparison results of sensitivity and FPR. The bars of the chart represent the result of sensitivity and the line represents the one of FPR.

Chapter 6
Conclusions and Future Works

In this thesis, we have proposed the WCDSP algorithm based on wavelet analysis, and chaos theory. The experiment results for several patients were shown with a high sensitivity with respect to prediction of epileptic seizures. The time horizon for a seizure prediction was set at 1 h and the average prediction time over all patients was about 27 min/seizure. The interval is sufficient for a therapeutic intervention. Not only a more reliable algorithm is presented, a VLSI implementation of the seizure analysis IP is also made for applications in portable device, such as VNS.

In the future, we may improve the prediction sensitivity and lower the false positive rate (FPR) by the two approaches:

(1) Hilbert-Huang transform:

The Hilbert-Huang transform (HHT) is NASA's designated name for the combination of the empirical mode decomposition (EMD) and the Hilbert spectral analysis (HSA).

It is an adaptive data analysis method, which improves accuracy by using an adaptive basis to preserve intrinsic properties of data, designed specifically for analyzing data from nonlinear and non-stationary processes, e.g. EEG signals. It yields results with more physical meaning and a different perspective than existing transforms.

Table 6-1 Comparison between Fourier, wavelet, and HHT

	Fourier	**Wavelet**	**Hilbert**
Basis	a priori	a priori	adaptive
Frequency	convolution: global, uncertainty	convolution: regional, uncertainty	differentiation: local, certainty
Presentation	energy-frequency	energy-time-frequency	energy-time-frequency
Nonlinear	no	no	yes
Non-Stationary	no	yes	yes
Feature Extraction	no	discrete: no, continuous: yes	yes
Theoretical Base	theory complete	theory complete	empirical

By HHT, we may obtain more important information to enhance the prediction results.

(2) Independent Component Analysis (ICA) :

In recent years, Independent Component Analysis (ICA) has been proved as a powerful algorithm to solve blind source separation (BSS) problems in a variety of signal processing applications such as speech, image, or biomedical signal processing.

The EEG is composed of electrical potentials arising from several sources. Each source (including separate neural clusters, blink artifact, or pulse artifact) projects a unique topography onto the scalp, called "scalp maps." These maps are mixed according to the principle of linear superposition.

Independent component analysis (ICA) attempts to reverse the superposition by separating the EEG into mutually independent scalp maps, or components.
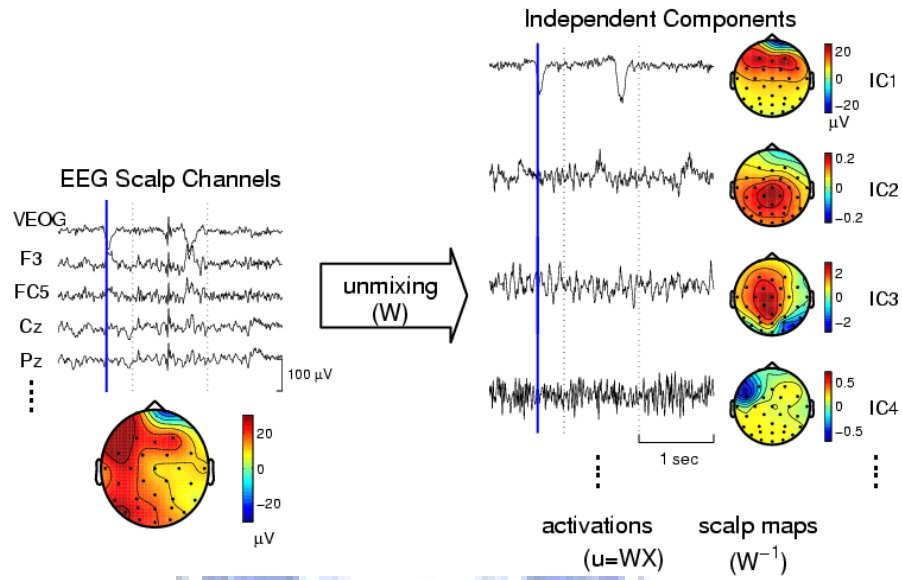
ICA decomposition

Fig. 6-1 ICA decomposition

We can apply ICA to multi-channel EEG recordings and remove a wide variety of artifacts from EEG records by eliminating the contributions of artifactual sources onto the scalp maps. It may lead to noiseless signals and we could more easily extract useful features to increase the sensitivity.

76

References

[1] M. Le van Quyen, J. Martinerie, M. Baulac, F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings", *Neuroreport 10,* pp2149–2155, 1999.

[2] F. Mormann, T. Kreuz, R.G. Andrzejak, P. David, K. Lehnertz, C.E. Elger, "Epileptic seizures are preceded by a decrease in synchronization", *Epilepsy Res. 53 173–185, 2003.*

[3] Vairavan Srinivasan, Chikkannan Eswaran, Natarajan Sriraam, "Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks", *IEEE Transactions on Information Technology in Biomedicine,* vol. 11, no. 3, MAY 2007.

[4] Vladimir G. Ivancevic, "High-Dimensional Chaotic and Attractor Systems - Intelligent Systems, Control and Automation: Science and Engineering", *Springer Publishers,* 2007.

[5] Steven H. Strogatz, "Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering", *Westview Press Publishers,* 1994.

[6] L. D. Iasemidis, J. C. Principe, J. C. Sackellares, "Measurement and Quantification of Spatio-Temporal Dynamics of Human Epileptic Seizures", *Nonlinear Signal Processing in Medicine, IEEE Press*, pp. 1-27, 1999.

[7] Leon D. Iasemidis*, Deng-Shan Shiau, Wanpracha Chaovalitwongse, J. Chris Sackellares, Panos M. Pardalos, Jose C. Principe, Paul R. Carney, Awadhesh Prasad, Balaji Veeramani, and Konstantinos Tsakalis, "Adaptive Epileptic Seizure Prediction System", *IEEE Transactions on Biomedical Engineering*, vol.50, no. 5, MAY 2003.

[8] Barbara Burke Hubbard, The World According to Wavelets: The Story of a Mathematical Technique in the Making, AK Peters, Ltd, 1996.

[9] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, *Prentice Hall Publishers,* 2002.

[10] SVETLANA BOROVKOVA, "Estimation and Prediction for Nonlinear Time Series", *PhD Thesis, University of Groningen,* 1998.

[11] Ingrid Daubechies, and Wim Sweldens, "Factoring wavelet transforms into lifting steps", *Journal of Fourier Analysis and Applications*, 2008.

[12] Bruno Direito1, António Dourado1, Marco Vieira1, Francisco Sale, "Combining energy and wavelet transform for epileptic seizure prediction in an advanced computational system", *International Conference on BioMedical Engineering and Informatics*, Volume 4, Number 3 247-269, 1998.

[13] V. Navarro, J. Martinerie, M. Le Van Quyen, S. Clemenceau, C. Adam, M. Baulac, F. Varela, "Seizure anticipation in human neocortical partial epilepsy", *Brain 125,* pp 640–655, 2002.

[14] Maryann D'Alessandro, Rosana Esteller, George Vachtsevanos, Arthur Hinson, Javier Echauz, and Brian Litt, "Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients", *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, MAY 2003.

[15] Hojjat Adeli, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr, "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy", *IEEE Transactions on Biomedical Engineering, vol. 54, no. 2, FEBRUARY 2007.*

[16] Samanwoy Ghosh-Dastidar, Hojjat Adeli, and Nahid Dadmehr, "Mixed-Band Wavelet-Chaos-Neural Network Methodology for Epilepsy and Epileptic Seizure Detection", *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, SEPTEMBER 2007.

[17] Thomas Maiwald, Matthias Winterhalder, Richard Aschenbrenner-Scheibe, Henning U. Voss, Andreas Schulze-Bonhage, and Jens Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic", Physica D 194 357–368, 2004.

[18] L.D. IASEMIDIS, P. PARDALOS, J.C. SACKELLARES, and D.-S. SHIAU, "Quadratic Binary Programming and Dynamical System Approach to Determine the Predictability of Epileptic Seizures", *Journal of Combinatorial Optimization,* 5, 9–26, 2001.