輔仁學誌—理工類

中華民國九十三年十二月　　第三十八期

目錄

FU JEN STUDIES

SCIENCE AND ENGINEERING

NO. 38, DEC. 2004

CONTENTS

X-ray Absorption Techniques in Examination the Structural Properties of Nanocrystalline Materials

Hong-Ming Lin*
Department. of Materials Engineering
Tatung University
Taipei 104, Taiwan, R.O.C.

Chung-Kwei Lin
Department of Materials Science
Feng-Chia University
Taichung 407, Taiwan.

Yeukuang Hwu
Institute of Physics
Academia Sinica Nankang
Taipei 115, Taiwan, R. O. C.

Abstract

The local atomic environment and electronic structure of nanocrystalline materials can be investigated by the X-ray absorption spectroscopy (XAS). In an XAS spectrum, the spectral feature near the absorption edge, or the X-ray Absorption Near-Edge Structure (XANES), is dominated by the details of the final density of state, the transition probability, and resonance etc. At higher photon energy side further above the absorption edge, specific oscillatory spectral features, known as Extended X-ray Absorption Fine Structures (EXAFS), can be used to extract the information about local coordination number, interatomic distance and other structural parameters. Both types of information can be obtained without requiring sufficient long-range order of the sample atomic structure and therefore

* Corresponding author. Tel.:+886-2-25866030; Fax:+886-2-25936897
E-mail address : hmlin@ttu.edu.tw

particularly suitable of examine nanostructured materials. The goal of this article is to review the XAS techniques and to illustrate through selected examples their practical use to examine the structural properties of nanocrystalline materials.

Keywords: X-ray Absorption Spectroscopy, Near Edge X-ray Absorption Fine Structure, Extended X-ray Absorption Fine Structure, Nanocrystalline Material.

1. INTRODUCTION

At 1962, Kubo[1] pointed out that the physical properties of ultrafine metallic particles changes with the particle size. This characteristic of ultrafine metallic particles has drawn much attention and eventually motivated the development of many new functional materials. Nanocrystalline materials are now commonly defined as materials made of fine particles or crystalline grains of the size between 1 and 100 nanometers. The characteristics of nanocrystalline materials are often emphasized: 1) It has a different atomic structure, 2) different properties, such as optical, magnetic, heat transfer, diffusion, and mechanical, from those of the traditional crystal or non-crystal materials. These characteristics can often be attributed to that nanocrystalline materials exhibit a higher ratio of atoms on the grain or interphase boundaries to the bulk.[2] For example, nanocrystalline materials can be easily mixed with otherwise immiscible materials or compound and form new types of alloys or solid solutions. Due to such deviation from bulk materials, the properties of nanocrystalline materials such as the strength/hardness, modulus, thermal expansion, electrical resistivity, surface catalysis, and corrosion, etc., have become important areas of research.[3-7]

One of the most attractive aspects of nanocrystalline materials, besides its fundamental role in bridging atoms and bulk solid, and the exhibition of quantum confinement behavior, is that the size of particles can be used as a materials engineering parameter. For example, the advantage of the high surface area on catalysis is obvious, as its application as gas sensors--more than half of the atoms are at grain boundaries or surfaces when the grain size is smaller than 10 nm. In other words, without taking advantage of other properties special to nanoscale materials, the smaller particle size alone can be advantageous in terms of materials processing.

Given the technological and fundamental advances of nanostructured materials fabrication, however, the creation and the characterization of nanostructured materials are not always trivial. Even the most important parameter in the nanocrystalline materials, the distribution of the sizes, is already a difficult quantity to measure. This has a profound impact not only to the characterization, but also to the fabrication of nanocrystalline materials and to control their properties. The necessity of adapting a multi-technique approach signifies the challenges one faces in the characterization of nanoscale materials. Specifically, the reduction of the dimension of particles may reach a scale smaller than what conventional characterization techniques consider the limit. For example, X-ray diffraction technique using conventional laboratory sources is not sufficient to distinguish crystalline structure with a grain size smaller than 10 nm from amorphous. The fact that it is still widely used as the only source providing structural information more or less indicates the shortage of proper structural tools for nanocrystalline materials.

In the case of multi-component materials such as bimetallic alloy, the diminishing XRD peaks due to the small sizes of each component make the analysis of the structure even more ambiguous. Meanwhile, in the early development stage of the materials synthesis, it is generally difficult to prepare high purity composite materials suitable for precise characterization. This may be a general problem in synthesizing new materials, but in the case of nanostructured materials, it is compounded with the lacking of sufficient tools for characterization.

We review here a general approach of multi-technique characterization using X-ray absorption spectroscopy as the core component. This approach has generated reasonable success in overcome the problems in characterizing nanostructured, particularly the nanocrystalline materials. A brief introduction of the spectroscopy technique is presented and many examples are used afterward to illustrate the practical usage of this approach.

2. SYNCHROTRON X-RAY ABSORPTION SPECTROSCOPY

Synchrotron radiation is the electromagnetic radiation emitted by charged particles, such as electrons or positrons, moving at relativistic velocities along a curved trajectory. A typical electron accelerator or storage ring emits synchrotron radiation in a very broad range of photon energies, from microwaves to hard-X-rays and γ-rays. It provides electromagnetic radiation in certain spectral regions, such as the ultraviolet/soft-X-ray range, for which no other

usable sources exist. It is also by far the best source of hard-X-rays, even though other sources exist for this range.[8] The characteristic of broadband radiation also enables one to select photon energy with a monochromator in a continuous fashion and make possible many powerful spectroscopy techniques.

X-ray absorption spectroscopy (XAS) is one of the most used and in our opinion, most relevant to the characterization of nanostructured materials. The absorption of X-rays by atoms is smoothly varying with photon energy except at specific discrete energies where abrupt increases (i.e., absorption edges) corresponding to the X-ray photons with sufficient energy to excite additional electrons from the core levels of the atoms into unoccupied states. An XAS spectrum, basically the measured X-ray absorption coefficient as a function of the incoming photon energy, is generally divided, by the role of multiple scattering of photoelectron, into two different regions: X-ray Absorption Near-Edge Structure (XANES) and Extended X-ray Absorption Fine Structure (EXAFS). XANES, also referred as NEXAS (Near Edge X-ray Absorption Fine Structure), can be used to investigate the electronic structure of specific elements, while EXAFS is used to investigate mainly the local atomic structure concerning the type and number of element, and nearest neighbor bond length.[9,10,11] The results of X-ray absorption spectroscopy (electronic structure and local atomic environment) can complement those obtained by X-ray diffraction (long range order) and transmission electron microscopy (local, although non-statistical, atomic structure) to provide a more precise and complete picture of the structure of the nanocrystalline materials.

The energy region which an XANES spectrum concerns usually extends to energy of the order of 50 eV above the edge (Fig. 1) that includes the unoccupied part of band structure just above the Fermi level. Thus, certain aspects of the electronic structure of detected element can be revealed. While, EXAFS concerns the oscillation in the absorption coefficient spectrum on the higher energy side of absorption edges. By inverting these data, the location of the first few atoms surrounding the absorbing atom can be obtained.

In the EXAFS spectral region, the interaction between the photoelectrons and the back-scattered photoelectrons is relatively weak due to their high kinetic energy, so it can be considered as a single scattering process. In the XANES region, however, the interaction is much stronger, and multiple scattering events must be considered. As the photoelectron wavefunction is strongly dependent on the scattering potential around the absorber, which in turns in-

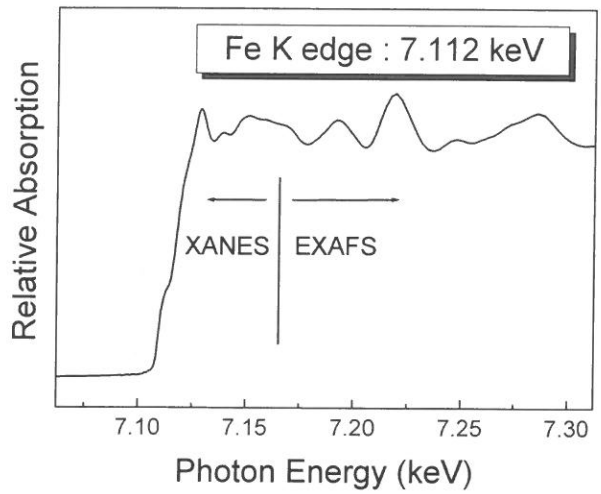


Fig. 1. Fe K edge X-ray absorption spectrum on standard Fe foil of 12 μm thick.

fluenced by the neighboring atoms, in principle XANES can be used to examine the coordination geometry of the absorber as well. Nevertheless, the most powerful use of XANES is still to explore the electronic structure, for example, the oxidation states of a constituent atom. The position of absorption edge changes as the electronic structure of the absorber changes which can be correlated to the chemical shifts as detected in X-ray photoelectron spectroscopy (XPS).[12]

Figure 2 illustrates the interference process of EXAFS. Interference between the outgoing wave and the backscattered wave of photoelectron produces the extended oscillations observed above the absorption edge. To the first approximation, the period of such interference, related closely with the separation of the absorber atom and its neighbors. Some information can be obtained about the second nearest neighbor coordination shell, or sometimes more distant shells by more sophisticate mathematical treatment of the data. Because only atoms close to the emitting atom produce EXAFS oscillations, EXAFS studies are not limited to system having long-range order, like single crystals. Systems that have a well-defined coordination around a central atom can already be studied. These include polycrystalline and amorphous, glasses, gel and solutions; examples including amorphous semiconductor, nanocrystalline materials, supported catalysts and biological systems.[13,14]

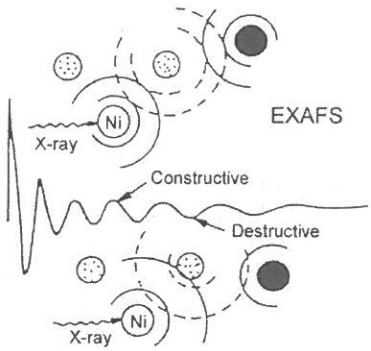


Fig. 2. Interference process of EXAFS.

Although these spectroscopies can be performed with laboratory sources, with the use of synchrotron radiation, the acquisition and quality of the XANES and EXAFS data is greatly improved and therefore corresponding to the majority applications. Currently these techniques have been successfully applied to many fields including inorganic chemistry, biochemistry, catalysis, and materials science. Detailed information is available in many excellent reviews and textbooks.[15-20] In this article, we concentrate to the applications of XAS to the studies of nanocrystalline materials.

3. APPLICATIONS

3.1 Nanocrystalline Metal and Alloys

In this section, we will use the work concerning nanocrystalline Ag-Ni or Ag-Fe powders and explain the data process in greater detail as an example to illustrate how the structural information can be obtained in a XAS based multi-technique approach. Nanocrystalline powder samples were prepared by a gas condensation method. By rapidly evaporating the metals in the tungsten boat, small particles are condensed in helium gas and then transported via the convection of helium gas and deposited on the liquid nitrogen cooled cold trap. The distance between the cold trap and the evaporation source as well as the inert gas pressure in the chamber can be optimized to ensure the smallest and most evenly distributed nanocrystalline particles.

XAS measurements were performed at the Taiwan Synchrotron Radiation Research Center (SRRC), with photons from the 6m-HSGM (high-energy spherical grating mono-

chromator), DCM (double crystal monochromator) and Wiggler beamlines.[21] XAS measurements at high energy photon range were performed on nanocrystalline powders using fluorescence and transmission modes that were detected using a Lytle detector.

Steps in the EXAFS data analysis included deglitching, background removal and normalization, the determination of the experimental threshold, conversion of E (energy, eV) to k (photoelectron wave vector, A^{-1}), weighting, and Fourier transformation. For instance, the EXAFS spectrum (Fe K edge) from a standard iron foil is shown in Fig. 3a where the spectrum has already been deglitched, background removed and normalized. After determining the experimental threshold (i.e., E_o), conversion into k space, and weighting, the corresponding k^2X (k) vs. k spectrum is shown in Fig. 3b. Finally the Fourier transformation of the k^2((k) into real space is performed and Fig. 3c shows the radial distribution function (RDF).

Measurements that were performed with different detection schemes were compared to ensure that the values of, at least, the derived first nearest neighbor bond lengths were consistent. The background removal and the extraction of the oscillation were performed following established EXAFS analysis practice[19] and a commercial software (WINXAS) were used to Fourier transform the spectra into radial distribution functions (RDF).

Both Ag-Fe and Ag-Ni binary systems are immiscible, i.e., they have limited solubility with each other at low temperatures. Ag_xFe_{100-x} and Ag_xNi_{100-x} (x=20, 35, 50, 65, 80) powders were prepared by a gas condensation method. Nanocrystalline Ag-Ni (or Ag-Fe) solid solutions, however, may be formed during the evaporation-condensation process. The first evidence of the formation of an AgNi solid solution were suggested by the absence of the Ni(200) peak in the X-ray diffraction (XRD) measurement. The $Ag_{1-x}Ni_x$ samples were prepared with different evaporation source contents. As shown in Fig. 4, Ni(200) is only visible in the sample of high Ni concentration.[22] Although there is no known Ni-lattice diffraction peaks can be seen, the measured Ni K edge absorption intensity remains nearly proportional to the Ni content. This excludes the possibility that Ni was evaporated during the condensation process. Figure 4 shows the EXAFS spectra of the corresponding Ag_xNi_{100-x} powders, where slight variation of the amplitude oscillation may be noticed. As shown in the inserted graph in Fig. 5, the variation of the nearest-neighbor bond length (the major peak position after Fourier transformation of the EXAFS spectra), shows a nearly linear relation with respect to the Ni composition. Similar to the Vegard's rule used to deduce the alloy composition from the lattice

constant,[23] such a relation indicates the formation of the AgNi solid solution.

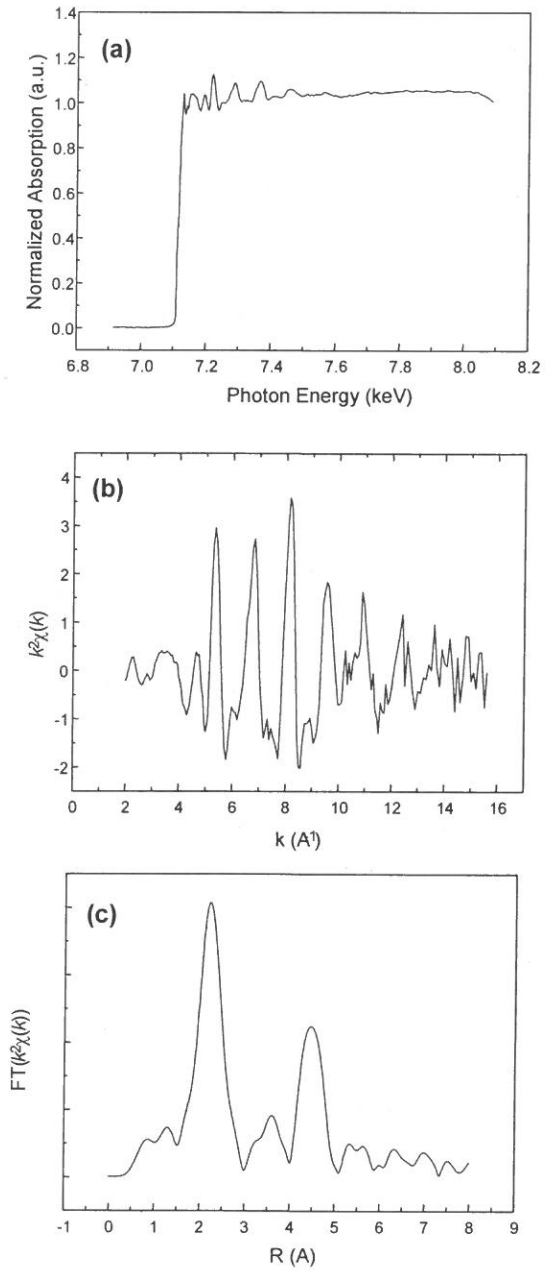


Fig. 3. Typical EXAFS analysis of a Fe foil: (a) EXAFS spectra at Fe K edge, (b) variations of wave number (k) after weighing, and (c) radial distribution functions (RDF) of Fe.

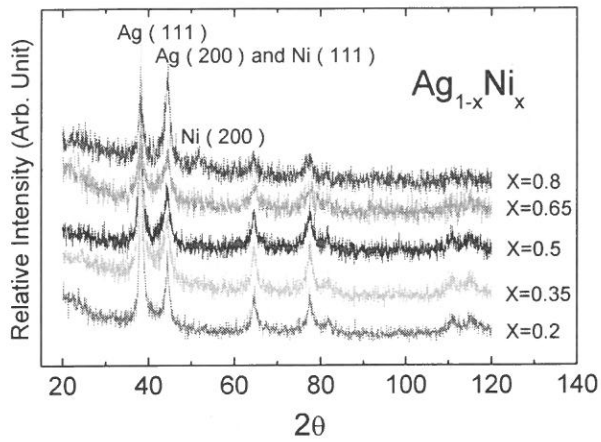


Fig. 4. X-ray diffraction of nanocrystalline AgNi powders.[22]

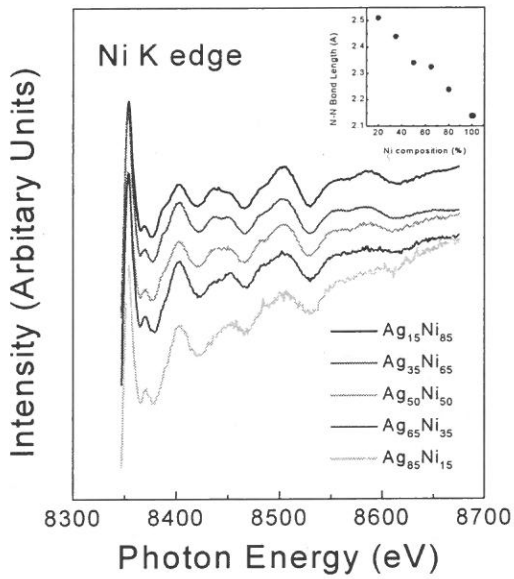


Fig. 5. EXAFS spectra of nanocrystalline AgNi powders. The inserted graph shows the nearest bond length (obtained from RDF) versus Ni composition for the corresponding nanocrystalline AgNi powders.[24]

The sharp feature in the edge region - see the top spectrum in Fig. 6 - at a photon energy of approximately 8365 eV is another strong indication that the Ni atoms of our samples are in a very different electronic structure than those of commercial Ni powders (ULVAC UFP)[24] which are also in a nanocrystalline form. Samples prepared with different noble gas pressure,

which usually leads to different particle sizes, does not exhibit strong differences in the X-ray absorption spectra, while the subsequent heat treatment caused drastic changes in the near-edge region and the EXAFS region.

The drastic decrease of the Ni $L_{2,3}$ absorption spectra intensity of the nanocrystal AgNi samples upon heating (Fig. 7),[24] indicates that a strong segregation of Ag and Ni occurred at the temperature above 320°C. Since the free energies of Ag and Ni increase with temperature and become eventually larger than the free energy gained by alloying in the nanocrystals, the segregation becomes energetically favorable above a certain temperature. We also observed different behavior on the samples of different particle sizes upon heating. For particle size of 50 nm and larger, the gain in the surface energy is not enough to form substantial amount of

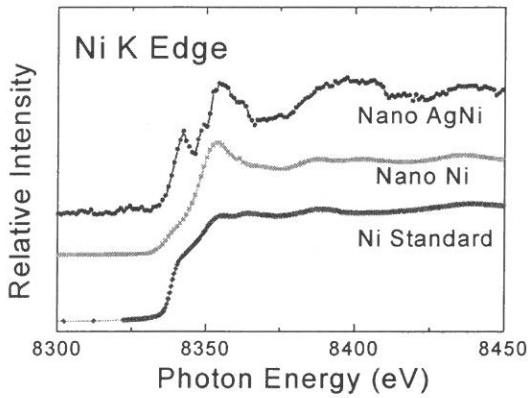


Fig. 6. XANES spectra of Ni standard foil, nanocrystalline Ni and AgNi powders.[24]

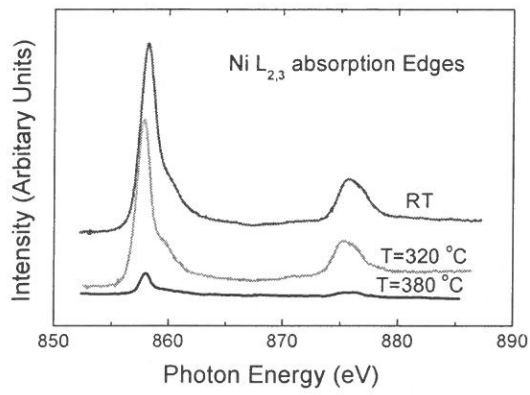


Fig. 7. Ni $L_{2,3}$ spectra for as-prepared and heat-treated nanocrystalline AgNi powders.[24]

AgNi solid solution and, therefore, no segregation can be detected at 300°C.

A simple empirical rule can be used to examine the surface segregation of transition metal alloys:[25] the constituent with lower elemental surface tension enriches the alloy surface. The same argument leads to a surface enrichment of Ag since the surface tension of the Ag in the liquid phase is lower than that of Ni.[26] The segregation temperature measured in our present experiment is consistent, within the accuracy of our temperature measurement, with that derived from the temperature dependent X-ray powder diffraction measurements. The result of the XRD of the sample after heated to above 420°C followed by a rapid quenching is shown in Fig. 8,[22] where the Ni(200) can be clearly seen. Scanning electron microscopy micrographs (not shown here) showed that the segregation made the sample thin-film-like, i.e., small particles no longer existed and there was a sharp division between the segregated Ag and Ni, with Ni buried under Ag. The segregation explains the decrease in the Ni L absorption signal.

Similar results were noticed for AgFe and Ag-Co (not reported here) binary systems. Figure 9 shows the X-ray diffraction patterns for as-prepared nanocrystalline AgFe powders.[27] We also noted that there are no Fe diffraction peaks can be distinguished. The EXAFS spectra are shown in Fig. 10a, where variations in the EXAFS regions can be noticed. It is more interesting to note that the XANES region of the corresponding nanocrystalline AgFe powders as shown in Fig. 10b. It can be noted that the Fe K absorption edges of these nanocrystals show a systematic change at photon energy around 7115 eV which indicate that the formation of nanocrystalline solid solution for immiscible AgFe system.

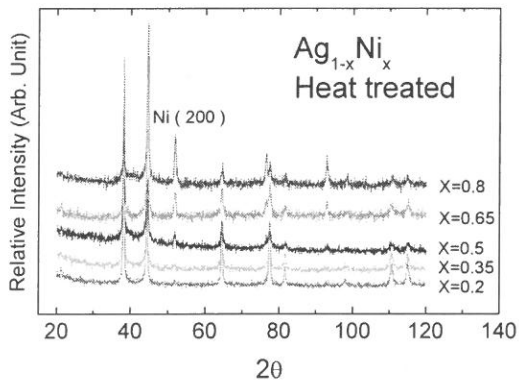


Fig. 8. X-ray diffraction of nanocrystalline AgNi powders after heat treatment above 420°C.[22]

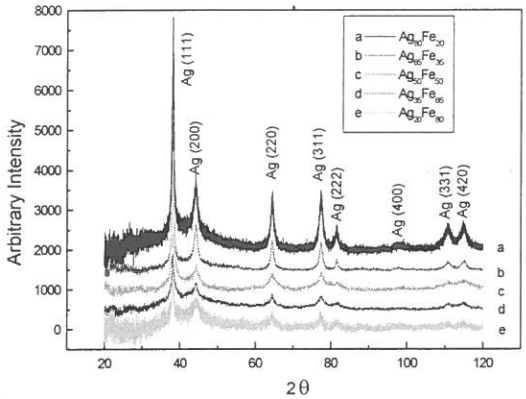


Fig. 9. X-ray diffraction of nanocrystalline AgFe powders.[27]

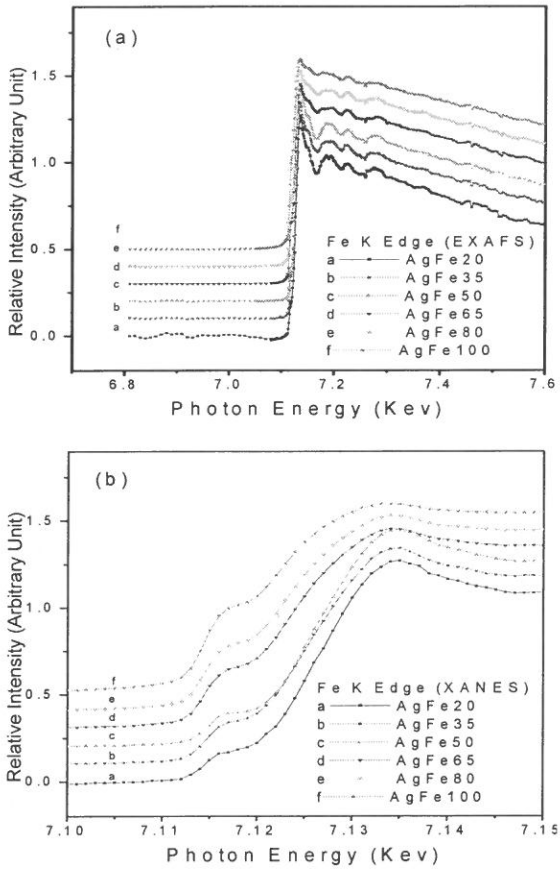


Fig. 10. (a) EXAFS, and (b) XANES spectra of nanocrystalline AgFe powders. [27]

3.2 Nano-Ceramics and Carbon Nanotubes

Liu *et al.*[28] used the X-ray absorption technique to examine the valences of V and Mo of $LiVMoO_6$. The XANES spectra of $LiVMoO_6$ and the standard samples at the V and Mo K-edges are shown in Figs. 11 and 12, respectively. The differences between the energy values corresponding to half height of normalized absorption ($\Delta\mu$ x) can usually be used to compare the oxidation states of the metal cations. According to the chemical shift of $LiVMoO_6$ spectra, the oxidation number of the vanadium ion is about +4.5 (as shown in Fig. 11) while that of molybdenum ion is +6 (as shown in Fig. 12).

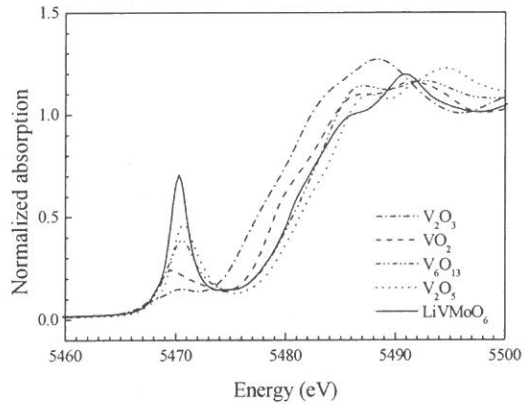


Fig. 11. Normalized V K-edge XANES spectra of $LiVMoO_6$ and those of the standard samples.[28]

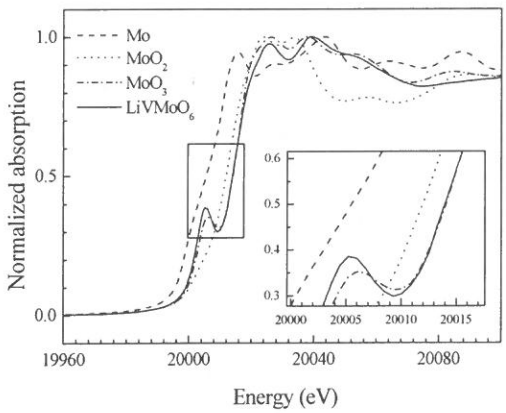


Fig. 12. Normalized Mo K-edge XANES spectra of $LiVMoO_6$ and those of the standard samples. [28]

The pre-edge feature can be assigned to the forbidden transition 1*s* to 3*d*, the lower-energy shoulder to the 1*s* to 4*p* shakedown transition and the strong peak to the dipole-allowed transition 1*s* to 4*p*. The initial 1*s* state being a ground state, the 1*s* to 3*d* transition is dipole forbidden in the regular octahedral VO_6 units with a center of inversion. When the symmetry of the VO_6 is lowered to distort octahedral (as in V_2O_3) or distorted square pyramidal (as in V_2O_5), the inversion center is broken. The pre-edge absorption is then no longer prohibited by the dipole transition rule due to the combination of stronger 3*d*-4*p* mixing and overlap of the vanadium 3*d* orbital with the 2*p* orbital of oxygen. This intense absorption occurring in the pre-edge region of the vanadium K-edge in $LiVMoO_6$ is shown in Fig. 11.

Kim, Choi and Lee[29] prepared TiO_2/SiO_2 by impregnated silica (Aldrich 23 683-7) with a solution of titanium tetrabutoxide (Aldrich 24 411-2) and toluene. Titania supported on silica samples were prepared with loadings of 1-20 wt % by Ti metal. X-ray absorption near edge structure (XANES) of Ti K-edge of Ti-Si mixed oxides and titania supported on silica with various Ti contents was studied to investigate the fractions of Ti-O-Si and Ti-O-Ti bonds quantitatively by fitting the pre-edge of Ti K-edge with the linear combination of two reference XANES spectra. Figure 13 is the Ti K-edge XANES spectra of TiO_2/SiO_2. In mixed oxides, the fraction of Ti-O-Ti was increased up to 0.55 when Ti/Si was varied from 0.04 to 0.5. The greatest change of each fraction occurred around 0.15-0.2 of Ti/Si, which was coincident with the formation of anatase titania as observed by XRD. This work demonstrated the possibility of the quantification of Ti-O-Si and Ti-O-Ti bonds in Ti-Si binary oxides by using the linear combination of reference XANES spectra.

Sanchez *et al.*[30] examined the nano-TiO_2 and nano-clay by EXAFS technique. Fourier filtering of the spectra between 0.8 and 2.3 Å extracted the first shell contribution. An average distance of 1.96 Å for the first oxygen shell of anatase was considered and two fitting procedures have been used. The second and third shells corresponding to four Ti-Ti at 3.04 Å, four Ti-Ti at 3.78 Å and eight Ti-O at 3.86 Å. This result can be explained considering that Ti in fine size TiO_2 and clays is present in two different phases: small particles of anatase and a mix of amorphous Ti(IV) oxo-ions specie with a broad distribution of Ti-O interatomic distances that smears out the EXAFS signal. Capel *et al.*[31] has conducted the similar research, to investigate Ti structure in TiO_2 doped stabilized tetragonal zirconia by the XANES and EXAFS techniques. From the EXAFS results, i.e. Ti-O and Ti-Ti distances, the decrease in

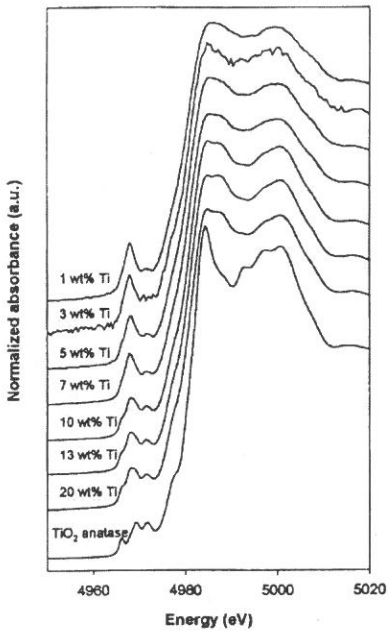


Fig. 13. Ti K-edge XANES spectra of TiO_2/SiO_2 prepared from impregnation method with different Ti loadings.[29]

conductivity was attributed to the formation of two kinds of cation-oxygen vacancy associations with different diffusion dynamics, leading to a decrease in the global concentration of moving oxygen vacancies.

Nano-CeO_2 had been heat treated at various temperatures to yield different particle sizes, the local structure around Ce atom was investigated by EXAFS at Ce K-edge(~40 keV).[32] It has been shown that EXAFS analysis at Ce K-edge is simpler and more straightforward than the procedures for Ce L_3(~5.7 keV)-EXAFS due to the absence of undesired spectral features at Ce K-edge. The first shell Ce-O distance was found to be slightly shorter with small CeO_2 particles, indicating a contraction of the lattice. There exists a significant increase in coordination number of the second cell Ce-Ce when the particle size became lager.

The local structure around both Ce and Zr of nano CeO_2-ZrO_2 samples was examined by using the X-ray absorption fine structure (XAFS) method in order to clarify the cation-cation (cation=Ce, Zr) network and the oxygen environment around the cation.[33] The different oxygen storage/release capacities (OSCs) of CeO_2-ZrO_2 were characterized by means of the Ce K-edge and Zr K-edge.

EXAFS is used to investigate local structure of nanosized ZrO_2 and Y_2O_3.[34] Fourier transform of EXAFS compares nanosized grain with coarse grain. The large strongly specific surface area ($75m^2/g$) is a first approach one may interpret that a reduction of coordination number in nano powder, amplitude is correlated with thermal and structural disorder. In nanosized ZrO_2, after compacting and sintering n-ZrO_2 (5-30nm), it will growth to ZrO_2 (80nm), the EXAFS oscillation is nearly identical to mono grain powder. Fourier filtered and back transformed oxygen shell can use to distinguish between the small distance (2.16Å) in the seven fold coordinated monoclinic part and large distance of the eight fold coordinated tetragonal part of the n-ZrO_2 powder. Results of quantitative data analysis prove the distance in n-Y_2O_3 is good agreement with the distance in coarse grain Y_2O_3 powder. A single shell fit gives reasonable result even though the metastable polymorph present several interatomic distances.

The oxidation state and coordination environment of copper ion-exchanged in ZSM-5 Zeolite can also be examined by EXAFS spectra.[35] When CO is adsorbed on the Cu^+ species, the coordination number of the carbon atom in the adsorbed CO is estimated to be 1.4, and a value of 1.89Å is obtained as a distance of Cu-C. The coordination number of oxygen is estimated to be 2.5, this value means that two- and three- coordinated oxygen atoms are involved. When CO gas (13.3 kPa) is introduced, three bands appear at distances of 1.89, 2.05, and 2.91Å. By strong interaction between the Cu^+ species and CO molecules, the former species is attracted to the latter molecule situated in a position distant from the lattice oxygen, resulting in a stable configuration. The variation of EXAFS spectra, the bond length, and the coordination number can be reasonably explained by taking account of the CO adsorption.

The structure of nano-NiO coated activated carbon fibers (ACF) had been examined by XANES and EXAFS spectra.[36] A dispersion of fine NiO particles can enhance methane adsorptivity of activated carbon fiber. The k^3 weighted Fourier transforms of $\chi(k)$ extended X-ray adsorption fine structures (EXAFS) spectra of NiO and dispersed NiO on ACF has five peak. The first peak of NiO-ACE has a little larger peak width than pure NiO. It means the specific structure that nearest-neighbor atomic distance of NiO is larger than that of powdered NiO. The results indicate that the electrons of the ligand valence band in the dispersed NiO are tightly bound than those of the powdered NiO.

To understand the synergistic effect of Co and Mo in synthesis of single wall carbon

nanotubes (SWCNT), X-ray absorption spectroscopy (EXAFS and XANES) has been used to characterize the state of Co and Mo on the catalysts before and after the production of SWCNT.[37] An important observation that may be linked to the deactivation process is the parallel trend observed in the oxidation state of Co with the rate of SWCNT growth. XANES/EXAFS showed that oxidized Co was present while the growth rate of SWCNT was high. After 30 minutes in reaction, when the SWCNT growth slower down, all Co became metallic. The EXAFS/XANES results on the Co:Mo (1:2) may shed some light in the combined action of Co and Mo. These results show that Co suffers a drastic transformation under reaction conditions. Before reaction, but after pretreatment in H_2 at 500°C and then in He at 700°C, a large fraction of Co is in the oxide state. After 10 min under reaction conditions, a significant growth in metallic Co was observed, although some oxidized Co was still present. After 30 min, the particles were even larger and essentially all the Co became metallic.

3.3 *In situ* XAS

The high penetration of X-ray allows measurement of the sample in their most natural state. This has great advantage of other techniques that require the measurement to be performed on specific environments, such as high vacuum or dried condition. Therefore, this characteristic is often taken to study reaction in an in situ fashion. To increase the time resolution to allow real time measurement of fast chemical reaction, a energy dispersive scheme of detection has been developed. We will show few examples that specifically targeted nanocrystalline materials.

Chemical oscillations during the oxidation of CO at atmospheric pressure on supported Pd catalysts were studied *in situ* by employing X-ray absorption spectroscopy (XAS) in an energy-dispersive mode.[38] From the Pd absorption spectra, a constant phase correlation of edge position (threshold energy) with respect to the observed temperature oscillations could be revealed. Furthermore, an oscillatory behavior in height and position of the first Pd peak was found in the radial distribution function FT($x(k)$).

A surface activation/deactivation mechanism for chemical oscillations during CO oxidation on supported Pd catalyst can be confirmed by XAS spectra. The activity of the catalyst appears to be decreased by a surface oxidation leading to an increasing concentration of linearly bonded CO molecules. The measured temperature oscillations within the range of a few

degrees indicate a kinetically controlled regime (i.e. thermokinetic oscillations). Figure 14 displays X-ray absorption spectra measured during CO oxidation and reduction on Pd/C with an integration time of 15 s each.

Figure 15 indicates two residual FT for the catalyst in the high- and low- activity branches of reaction. The Pd foil is also shown for reference. Obviously, three additional shells can be distinguished. The first shell corresponds to either C or oxygen from the reactant or the support. The second and third shell correspond to bridged and linearly bonded CO on the Pt cluster's surface, respectively.

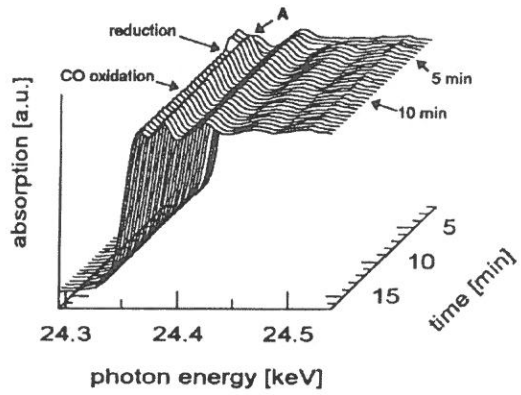


Fig. 14. Evolution of the Pd K-edge spectra during the reduction of PdO/C with CO (@ 5 min) and the starting oxidation of CO and Pd/C at 180°C (@ 10 min) (integration time per spectrum: 15s). [38]

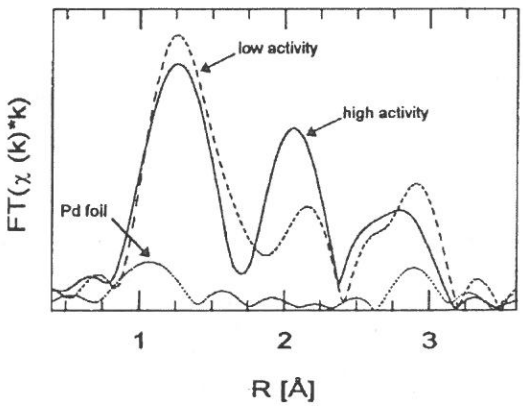


Fig. 15. Fourier transferred residual χ(k) of Pd/C catalyst in the high (solid) and low (dashed) activity stage of the oscillatory reaction. [38]

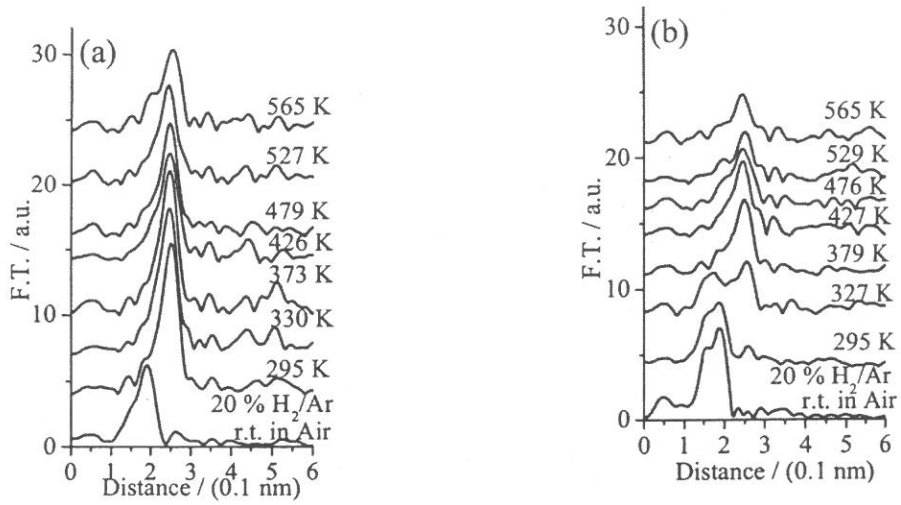


Fig. 16. Fourier transform of *in-situ* EXAFS spectra ($k^3\chi(k)$) observed for (a) Pd-Pt/SiO_2, and (b) Pd-Pt/Al_2O_3 during reduction in 20% H_2/Ar with increasing temperature at a rate of 5 K/min. The flow rate was 120 ml/min.[39]

Also, *in situ* EXAFS observation of nano-size bimetallic Pd-Pt particles supported on silica and alumina (Pd-Pt/SiO_2, Pd-Pt/Al_2O_3) was carried out.[39] Pd K-edge and Pt L_{III}-edge EXAFS spectra were measured during reduction process under a flow of 20% H_2/Ar from ambient temperature to 573 K. Metallic species were immediately formed by contact with H_2 at ambient temperature for Pd-Pt/SiO_2. On the other hand, for Pd-Pt/Al_2O_3, Pd and Pt atoms remained in an oxidized state under H_2 at ambient temperature and metal particles were formed by subsequent treatment at 427 K. Smaller metal particles were formed after reduction on Al_2O_3, compared with those on SiO_2. The kinds of the support have significant influence on both the process of reduction and the structure of reduced metal particles. Above difference in both structure and reactivity is supposed to be related to higher activity for hydrodearomatization (HDA) of diesel fuel than a silica-supported one in the presence of sulfur. Figure 16 shows changes Fourier transform of Pd K-edge XANES (X-ray absorption near edge structure) spectra observed during reduction process for Pd-Pt/SiO_2 and Pd-Pt/Al_2O_3. The bottom spectrum in both series of spectra was observed before starting reduction. In the Pd K-edge spectra, there was a peak at 24357 eV (room temperature, in air), and its shape is similar to that of $PdCl_2$. The Cl atoms were initially included in precursors as ligands and were supposed to remain on the surface after calcination. Owing to continuous monitoring of

spectra, different behaviors of supported nano-species in the two catalysts under reductive conditions at ambient temperature could be obtained. Metal atoms on SiO_2 were reducible and formed metal particles even at room temperature and formed relatively large particles after reduction. On the other hand, those on Al_2O_3 gradually reduced under H_2 and formed smaller particles. Above difference in reducibility is ascribed to difference in interaction between metal species and the supports. And in both catalysts, Pd metal particles were mainly composed of Pd, because relatively good fit was obtained only by Pd-Pd scattering. Both Pt and Pd surrounded Pt atoms. It is speculated that there are two kinds of domains: one is Pd rich and the other is mixture of Pd and Pt. And the difference in structures of reduced catalysts is supposed to be closely related to of the difference in catalytic activity for HDA.

3.4 Nano-Catalytic Materials

The reaction between colloidal palladium and cupric acetate in 2-ethoxyethanol leads, sequentially to the reduction of Cu(II) to Cu(0) at the surface of the palladium particle and dissolution of the deposited copper in the palladium matrix.[40] The Cu K near-edge X-ray absorption spectrum shows reduction of Cu(II) to Cu(0). After deposition on the palladium surface the copper is zero valent, and subsequently it migrates into the palladium particle but only to within a few atomic layers of the surface. Results of EXAFS analysis were able to determine the structural environment of copper in the "plated" colloid $Pd_{90}Cu_{10}$-P and to compare this result with the analogous copper environments in the "homogeneous" alloy particles $Pd_{90}Cu_{10}$-H and $Pd_{80}Cu_{20}$-H. Between the number of copper atoms in the average palladium coordination sphere, N^{PdCu}, and the number of palladium atoms in the coordination sphere of the average copper atom, N^{PdCu}, in a bimetallic PdCu matrix with a stoichiometry defined by the atomic fractions X_{Pd} and X_{cu}.

$$N^{PdCu} = (X_{cu} / X_{Pd}) N^{PdCu}$$

The combined EXAFS and XANES results also confirm that the copper in these dilute palladium-copper alloy colloids is zero valent and internal to the particle, as previously concluded on the basis of infrared spectroscopy of adsorbed CO. The reactivity of Cu(II) toward Pd(0) surfaces is confirmed, and the pathway from colloidal palladium to the final PdCu alloy particle has been delineated. The deposition of one metal at the surface of a second colloidal

metal is a potentially fruitful approach to the controlled synthesis of catalyst surfaces with controlled stoichiometries. The careful selection of precursors and deposition conditions is necessary, but by using this approach, a number of potentially interesting colloidal bimetallics should be accessible.

The use of temperature-dependent EXAFS measurements is used to investigate the nature of bond relaxation phenomenon occurring within the small metallic nanoparticles.[41] To evaluate these complex structural behaviors, the disorder parameters are calculated from temperature-dependent EXAFS data and then subsequently compared to simple molecular graphics simulations of mechanisms involving either full cluster or surface relaxations. The average bond length and static disorder obtained by experiment appear to best fit a model involving dominant contributions made by surface atom bond relaxation.

A new catalyst for the copolymerization of ethylene and α-olefins, based on (η-C_5Me_5)$TiCl_3$, methanol, 2,6-di-*t*-butylphenol, and methylaluminoxane, had been examined by Wasserman *et al.*[42] To employ Ti K-edge EXAFS gain insights into the atomic environment around the transition metal atom in a half-sandwich titanium-based polyolefin catalyst. Data collection in the 50 eV window about the K-edge for our samples gave the XANES spectra. The edge energy E_0, which is part of the XANES peak-fitting output, found to vary significantly not only from sample to sample, but from fit to fit on a single sample. XPS and XANES analyses confirm that the average titanium center is more electron-poor when activated in the presence of methanol and/or DTBP; line shape fitting leads to a conclusion that DTBP limits titanium interestingly, all samples made from Cp*$TiCl_3$ and MAO, with or without MeOH or DTBP, appear to be fit best to the same structure as the four-component catalyst.

Yang *et al.*[43] studies the effect of photoreduction on the properties of Pt/TiO_2 catalysts can be investigated by XANES spectra. From the EXAFS fitting results, the growth of Pt particles with the photoreduction time could be monitored. The XPS spectra of Pt/TiO_2 catalysts at the Pt $4f_{5/2}$ and Pt $4f_{7/2}$ bands showed the presence of Pt^0 and Pt^{2+} states after photoreduction and the fraction of metallic Pt^0 was increased with increasing photoreduction time.

TiO_2 catalysts cannot absorb visible light and only make use of 3-5% of the solar beam that can reach the earth, necessitating the utilization of an ultraviolet light source. It is, therefore, necessary to develop a photocatalytic system that can be applied under visible and/or solar light irradiation. The metal ion-implantation using the accelerated metal ion beam has

been developed to modify the electronic structure of semiconductor electronic materials. By the metal ion-implantation method, metal ions are accelerated enough to have the high kinetic energy and can be implanted into the bulk of TiO_2 photocatalyst. TiO_2 photocatalyst, which can absorb visible light and work as a photocatalyst efficiently under visible light irradiation, were successfully prepared using this advanced technique.[44] XAFS studies have revealed that the implanted metal ions are located at the lattice positions of Ti^{4+} in TiO_2 after the calcinations. These spectroscopic studies show that the substitution of Ti ions in TiO_2 lattice with implanted metal ions is important to modify TiO_2 to be able to adsorb visible light and operate as photocatalyst under visible light irradiation.

The effect of milling is investigated on structural changes in a physical mixture of CuO and Cr_2O_3 (denoted as CuO)/ Cr_2O_3), which is used as an active catalyst component for the low-temperature methanol synthesis via methyl formate.[45] X-ray absorption fine structure (XAFS) reveals that long-term milling causes intimate contact between fine grains of CuO and Cr_2O_3 and lattice disorder in CuO crystallites. XAFS analysis of CuO/ Cr_2O_3 milled for 50h was slightly different from the spectra for non-milled CuO and 1h-milled CuO/ Cr_2O_3, that suggesting that some structural changes would occur during the course of prolonged milling. FT-XAFS shows that the bond distances of Cu became longer and that the peak heights for the bonds became lower with increased milling time. The structural parameters calculated by curve fitting clearly revealed that the bond distances were lengthened by prolonged milling from 1.94 to 2.07Å for Cu-O and 2.78 to 2.93Å for Cu-Cu. These findings indicate that the lattice disorder (amorphization) took place in the surroundings of Cu atoms in the CuO crystallites. The XAFS study clearly displayed that the lattice disorder occurred in CuO crystallites in CuO/Cr_2O_3 samples milled for prolonged times. This agreed with the disappearance of the CuO peaks in the XRD profiles and the abrupt increase in the CuO lattice strain for long-term-milled CuO/Cr_2O_3.

4. CONCLUDING REMARKS

The analysis of X-ray absorption spectra reveals element-specific information on both, the electronic structure of the absorbing species and its local environment, even in disordered systems. Using state-of-the-art methods of measurement and computational modeling, it is

possible to develop a detailed understanding of the shape and morphology of the nanoscale materials. The XAS technique is of intrinsic and practice value for structural and electronic studies in nanomaterials for catalytic and surfaces. Except the static XAS measurement, the very exciting future developments will come from dynamic studies on the chemical reaction of nanoscale materials. We reviewed using powerful tools of extended X-ray absorption fine structure (EXAFS) and X-ray absorption near-edge structure (XANES) to study the structural properties of metals, alloys and ceramics, *in situ* studies of surface reaction, and examining the catalytic activity of these nanoscale materials. We believe that XAS analysis for dynamics and surfaces can provide a valuable contribution toward present and future fundamental and practical research for development of nano science and technology.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support of this work by the National Science Council of Republic of China. The authors also thank the entire staff at SRRC (Hsinchu, Taiwan, R.O.C.) for the expert assistance.

REFERENCES

(1) Kubo, *J. Phys. Soc. Japan*, 17 (1962) 975.
(2) X. Zhu, R. Birringer, U. Herr and H. Gleiter, *Physics Review* B, 35 (1989) 9085.
(3) H. Froes and C. Suryanarayana, *JOM*, (1989)[6] 12.
(4) R. Birringen, *Materials Science and Engineering*, A117 (1989) 33.
(5) R. P. Andres, R. S. Arerback, et al., *J. Mater. Res.*, 4(3) (1989) 704.
(6) R. Dagani, "Nanostructured Materials Promise to Advance Range of Technologies", *C&EN*, Nov. 23, 1992.
(7) B. H. Kear, *Navel Research Review*, Four (1994), 4.
(8) G. Margaritondo, *Introduction to Synchrotron Radiation*, Oxford University Press, New York Oxford, 1988.
(9) E.A. Stern, *Phys. Rev.* B, 10[8] (1974) 3027.
(10) F.W. Lytle, D.E. Sayers, and E.A. Stern, *Phys. Rev. B*, 11[12] (1975) 4825.
(11) E.A. Stern, D.E. Sayers, and F.W. Lytle, *Phys. Rev. B*, 11[12] (1975) 4836.

(12) N. Kosugi and T. Fujikawa in *X-ray Absorption Fine Structure for Catalysts and Surfaces*, Y. Iwasawa (Ed.), World Scientific, Singapore, 1996. Chapter 4: Theory and Analysis of XANES.

(13) T. Yokoyama in *X-ray Absorption Fine Structure for Catalysts and Surfaces*, Y. Iwasawa (Ed.), World Scientific, Singapore, 1996. Chapter 2: Theory and Parameters of EXAFS.

(14) K. Asakura in *X-ray Absorption Fine Structure for Catalysts and Surfaces*, Y. Iwasawa (Ed.), World Scientific, Singapore, 1996. Chapter 3: Analysis of EXAFS.

(15) E.A. Stern and S.M. Heald, "Basic Principles and Applications of EXAFS", pp955-1014 of *Handbook of Synchrotron Radiation*, vol. 1B, E.E. Koch (Ed.), North-Holland Pub. Co., NY, 1983.

(16) B.T. Teo, *EXAFS: Basic Principles and Data Analysis*, Springer-Verlag, NY, 1986.

(17) J. Wong, "Extended X-ray Absorption Fine Structure", pp407-419 of *Metals Handbook* 9th Ed., Vol. 10, ASM, Metals Park, OH., 1986.

(18) Y. Iwasawa, *X-ray Absorption Fine Structure for Catalysts and Surfaces*, World Scientific, River Edge, NJ., 1996.

(19) D.C. Koningsberger and R. Prins, *X-ray Absorption Principle, Applications, Techniques of EXAFS, SEXAFS, and XANES*, John Wiley & Sons, NY., 1988.

(20) J.E. Penner-Hahn, *Coordination Chemistry Reviews*, 190-192 (1999) 1101.

(21) Please visit http://www.srrc.gov.tw

(22) C.Y. Tung, H.M. Lin, J.M. Gu, P.Y. Lee, *Nanostructured Materials*, 9 (1997) 117.

(23) B. D. Cullity, *Element of X-ray diffraction*, 2nd. Ed. (Addison Wesley, Menlo Park, 1988) p. 84.

(24) C.Y. Tung, J.M. Gu, H. M. Lin, Y. Hwu, and N.F. Cheng, *Nanostructured Materials*, 9 (1997) 351.

(25) A. Zangwill, *Physics at Surfaces*, Cambridge University Press, Cambridge, 1988, p.87.

(26) J. N. Schmit, Ph. D. Thesis, University of Liege (unpublished).

(27) H.M. Lin, C. Y. Tung, S. J. Tzeng, W. L. Tsai, Y. Hwu, C. K. Lin and P. Y. Lee, *Nanostructured Materials*, 10 (1998) 457.

(28) R.S. Liu, C.Y. Wang, S.F. Hu, L.Y. Jang, and J.F. Lee, *The Chinese Journal of Process Engineering*, 2 Suppl. (2002) 485.

(29) W.B. Kim, S.H. Choi, and J.S. Lee, *J. Phys. Chem.* B, 104 (2000) 8670.

(30) M.C. Sanchez, J. Garcia, J.A. Mayoral, J. Blasco, M.G. Proietti, *Physica* B, 208-209 (1995) 702.

(31) F. Capel, C. Moure, P. Duran, A. R. González Eilpe, and A. Caballero, *J. European Ceram. Soc.*, 19 (1999) 765.

(32) J.F.Lee, M.T.Tang, W.C. Shih, and R.S. Liu, *Mater. Res. Bulletin*, 37 (2002) 555.

(33) Y. Nagai, T. Yamamoto, T. Tanaka, S. Yoshida, T. Nonaka, T. Okamoto, A. Suda, M. Sugiura, *Catalysis Today*, 74 (2002) 226.

(34) R. Nitsche, M. Winterer and H. Hahn, *Nanostructured Materials*, 6 (1995) 679.

(35) R. Kumashiro, Y. Kuroda, and M. Nagao, *J. Phys. Chem. B*, 103 (1999) 89.

(36) Y. Hattori, T. Konishi, and K. Kaneko, *Chemical Physics Letters*, 355 (2002) 37.

(37) W.E. Alvarez, B. Kitiyanan, A. Borgna, and D.E. Resasco, *Carbon*, 39 (2001) 547.

(38) T. Ressler, M. Hagelstein, Ulrich Hatje, and W. Metz, *J. Phys. Chem. B*, 101 (1997) 6680.

(39) K. K. Bando, L. Le Bihan, H. Yasuda, K. Sato, T. Tanaka, F. Dumeignil, M. Imamura, N. Matsubayashi, and Y. Yoshimura, *Analytical Sciences*, 17 Suppl. (2001) i127.

(40) J.S. Bradley, G.H. Via, L. Bonneviot, and E.W. Hill, *Chem. Mater.*, 8 (1996) 1895.

(41) A.I. Frenkel, C.W. Hills, and R.G. Nuzzo, *J. Phys. Chem. B*, 105[51] (2001) 12689.

(42) E.P. Wasserman, A.D. Westwood, Z. Yu, J.H. Oskam, and S.L. Duenas, *J. Molecular Catalysis A: Chemical*, 172 (2001) 67.

(43) J.C. Yang, Y.C. Kim, Y.G. Shul, C,H, Shin, and T.K. Lee, *Applied Surface Sciences*, 121/122 (1997) 525.

(44) H. Yamashita, M. Harada, J. Misaka, M. Takeuchi, K. Ikeue, and M. Anpo, *J. Photochenistry and Photobiology A: Chemistry*, 148 (2002) 257.

(45) S. Ohyama and H. Kishida, *Applied Catalysis A: General*, 184 (1999) 239.

received October 29, 2004
accepted November20, 2004

X 光吸收光譜技術檢測奈米結晶材料之結構性質

林鴻明
大同大學材料工程學系

林中魁
逢甲大學材料科學系

胡宇光
中央研究院物理所

摘　要

X 光吸收光譜學可以用來探測奈米結晶材料之局部區域之原子及電子結構性質。在 X 光吸收光譜中，X 光近緣結構或延伸 X 光吸收精細結構光譜主要受到電子密度分佈、電子轉移之機率、共振等因素影響。高於 X 光吸收邊緣之光子能量，顯示有特殊明顯之震盪光譜，即延伸 X 光吸收精細結構光譜，此光譜可以分析局部區域原子之配位數、原子間距以及其他原子結構參數。X 光近緣結構或延伸 X 光吸收精細結構光譜分析不需要長程規則之原子結構，因此，特別適合應用於奈米結構材料之分析。本文主要回顧 X 光吸收光譜技術同時舉實例說明如何檢測奈米結晶材料之結構性質。

關鍵詞：X 光吸收光譜學，X 光近緣結構，延伸 X 光吸收精細結構，奈米結晶材料。

Making scientific knowledge accessible to all: a role for a Catholic University in Taiwan

Joseph Arul (Dept. of CSIE) *
John P. Selvamani (Holistic Education)
Fu Jen Catholic University,
Taipei, Taiwan.

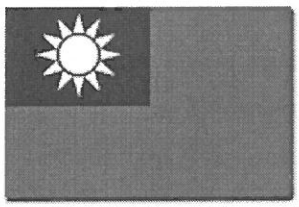
Abstract

Taiwan, an island nation, has become an important player in science and technology development. Realizing the potential in biotechnology and information technology advances, researches in these areas have been a priority among universities and research institutions in the country. At the same time, there is an awareness of the need to educate people on the impact of scientific advances on Taiwanese society and enable them to evaluate and judge the role of science and technology in human society. In this regard, biotechnology and genomic medicine have received much attention. Science, Technology and Society (STS) network and Ethical, Legal, Social Issues (ELSI) research centers have been established to educate the public and promote researches in the ethical implications of biotechnology and genomic medicine in Taiwan.

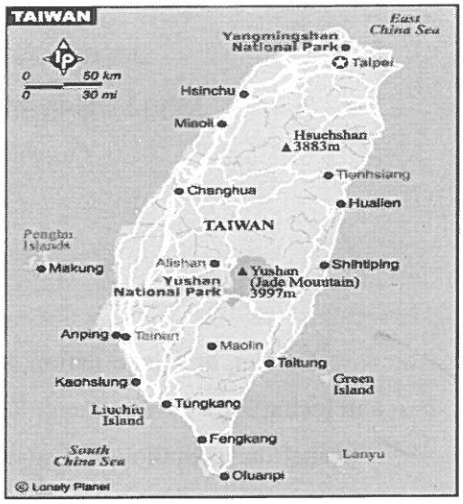
Many universities in the country have included courses on science, technology and society for non-science students to promote better understanding of science in the future generation. In this paper we assess the courses offered in science to non-science students in selected universities in Taiwan. We evaluate, in detail, the science courses offered through the **Holistic education** curriculum in Fu Jen Catholic University and propose ways by which a Catholic university could play an important role in pro-

* Corresponding author. Tel: 2-29053896; Fax:2-29023550
E-mail:arul@csie.fju.edu.tw

moting a scholarly and healthy discussion on the impact of science on human society and values in Taiwan.



National Flag



1. INTRODUCTION

The development of science and technology is a blessing and a burden for human society. Not an hour passes by in the life of a modern human being, which has not been enhanced by science, but at the same time there is a fear of being overridden by unregulated technology. The influence of science on human society is so deep that scientists alone cannot decide the future direction of science. To enable wider participation in deliberation on the role of science in human society and values, it is important to disseminate scientific knowledge to the public. There is a need for the public to understand and judge what those people in white coats do behind the closed doors of laboratories. Communication between scientists and the public is necessary to minimize the burdens of science. The universities must be a place, which can give a lead, and be a guiding force to build a structure for the government organization, church and other non-governmental organizations. Taiwan, as one of the leading players in the development of science, has felt the need to educate the public on science that they are well equipped to make an informed decision. To achieve this goal, a "General education cur-

riculum" has been included for university students, by which they are expected to possess basic understandings of science and technology. Since this is the first workshop of this kind to form a sectorial group to understand various realms from different continents; we give you a brief description of where Taiwan stands in this area of general education, which is meant for all people. Before we outline the impact of the "general education curriculum" and the role of Fu Jen Catholic University in promoting such a curriculum, we would like to introduce you to science education in Taiwan and to Fu Jen Catholic University.

2. BACKGROUND

a. Information on Taiwan

Taiwan, which is an island 394 km long and 144 km at its widest point, is shaped like a tobacco leaf. Taiwan, formerly known as ***Ilha Formosa*** or "beautiful island," one of the islands of this chain, is the largest body of land between Japan and the Philippines. Taiwan has a total area of nearly 36,000 sq. km, separated from China by the Taiwan Strait. It is about 220 km at its widest point and 130 km at its narrowest. As of November 2002, Taiwan had a population of 22.51 million people. Its population density is about 622 persons per square kilometer. Its economy had experienced a steady growth due to the spread of the universal education throughout the island. The Year Book of Taiwan 2003 states that, "From 1951 to 2001, the number of university students, including those at private colleges and universities increased by more than 100 times to 677,171." It further states that, "there were only five M.A candidates in 1950 and Taiwan did not have its first Ph.D student until 1956, but by 2001 there were 87,251 students in 3,250 master's programs and 15,962 students studying in 960 Ph.D programs." In 1951, 34.60 percent of the population six years and older was illiterate. By 1969 it dropped to 15.30 percent. Currently about 5.34 percent of the population is illiterate, mostly the elderly. Taiwan's constitution is based on the three principles of the people known as principles of nationalism, democracy and social well being formulated by Dr. Sun Yat-Sen the founding father of the Republic of China (ROC).

b. Education system in Taiwan

The Ministry of Education (MOE), a governing body chosen by the Central Government, centrally manages the education in Taiwan. The MOE sets the national educational policy and directly oversees departments and bureaus of education at the local and municipal levels. In 1950, only seven institutions offered higher education programs to 6,665 students. One University had three graduate-level departments. By 1974, 13 public and 19 private higher education institutions had been opened. The table 1 below presents the number of students studying in different levels in the 2001 academic year (SY2001). The number of higher education institutions reached 154, consisting of 57 Universities, 78 independent colleges and 9 junior colleges. A total of 1,187,225 undergraduates were enrolled in these institutions. There were 103,213 graduate students in 1,668 graduate programs. Many problems facing Taiwan's educational system at that time centered around the inadequacy of resources, especially high student-to-teacher ratios and high student-to-classroom ratio.

Table 1. Different levels of institutions and the number enrolled in SY2001.

Institution	Total	Private	Enrolled
Junior Colleges	19	16	79,371
Universities and Graduate Schools	57	30	580,898
Independent Colleges	78	-	526,956

Among the leading schools founded by the Catholic Church in Taiwan are 2 universities; Fu Jen Catholic University 天主教輔仁大學, and Providence University 靜宜大學. Among other Catholic institutions and high schools are Blessed Imelda's School 靜修女子高級中學, Kuang Jen Middle School 光仁中學, Fu Jen Middle School 天主教輔仁中學, the Cardinal Tien School of Nursing & Midwifery 耕莘高級護理助產職業學校, and the WenZao Ursuline College of Modern Language 文藻外語學院.

Fu Jen Catholic University is a large, comprehensive university located in the suburb of Hsin Chuang ten kilometers from Taipei, Taiwan. The total enrollment is nearly 23,000 students. It has a day and evening division. The day division consists of 8 colleges: Liberal Arts, Arts, Medicine, Science and Engineering, Foreign Languages, Human Ecology, Law and Management. The Evening division has 14 departments. Currently, there are 7 Ph.D programs, 35 master's programs and 44 undergraduate departments. More than 97,000 students have

graduated from Fu Jen, since it's re-opening in Taiwan some 40 years ago. The faculty consists of about 780 full time faculty members, 140 professors, 260 Associate professors and about 135 lecturers and more than 40 specialists in various fields. The students in Taiwan enter colleges through an island-wide joint entrance examination. Almost 37% of Fu Jen students say they have no religious faith. Besides, we can also clearly note the low number of Christians, both Catholic and Protestant. So, Fu Jen is surely not a Catholic or Christian University in that a large percentage of its students are Christian. In 1994, a University committee re-wrote Fu Jen's mission statement and educational goals. The mission statement of Fu Jen states that, "Fu Jen is committed to a dialogue leading to the integration of Chinese culture and Christian faith; to academic research and the promotion of genuine knowledge; to the development of society and the advancement of humankind." How does the University try to reach the average student? One of the ways in which the University tries to reach these students is through the so-called "Holistic educational program." The unique slogan of "holistic education" has been taken up by several other Christian universities in Taiwan as well as Hong Kong as defining their Christian character.

c. Religious Background

The figures released by the Ministry of Interior (MOI) in June 2001 indicated that about 12.96 million people in Taiwan were religious believers. During the Japanese occupation in Taiwan, Roman Catholicism experienced a rather slow development. It could have been due to the suppression by the Japanese colonial government; however there is no concrete evidence to support this theory. By 1945, there were just 10,000 Roman Catholics, 52 churches or missions, and 20 missionaries in Taiwan. When the island was returned to the Republic of China in 1948, the number of believers stood at 13,000. In 1949, when the Central government was moved to Taiwan, multitudes of Roman Catholic clergy and believers followed, with a new strength and vigor. The number of converts grew rapidly during the years 1953-63, from 27,000 to 300,000. In the year 1969, the number of Roman Catholics reached a maximum of 306,000, with seven dioceses. As of June 2001, based on the yearbook of 2003, there were 1,135 Catholic churches, 677 clergymen, and 664 foreign missionaries in Taiwan serving about 298,000 believers. One important factor that influences the religious nature in Taiwan is the eclectic nature of the people's view. People here have seldom felt it necessary to

exclude aspects of other faiths from their personal or collective religious beliefs. As of June 2001, there were 11 primary schools, 48 high schools, six colleges, 13 universities and 90 monasteries and seminaries operated by various religious groups.

d. Science and Technology Development in Taiwan

The government has initiated a 2008 National Development plan, which stresses innovation, research and development, including private investment in research and development. It plans to spend 3% of its gross domestic product in research and development and thus make Taiwan a best location for innovation, research and development in Asia. In the year 2002, the National Science Council's (NSC) budget was US$699.25 million. The NSC budget was spent as shown in fig.1. Taiwan ranks sixth in the world in machinery and automation. Taiwan also cooperates with leading institutions abroad such as Stanford and MIT. It co-operates with Stanford in producing a lower-power Integrated Circuit (IC) and biomaterials and Drug delivery system project at MIT. In Oct. 2002, the science and Technology Advisory Group of the Executive Yuan allocated US$857million to establish 75 biotech companies in Taiwan over the next eight years.

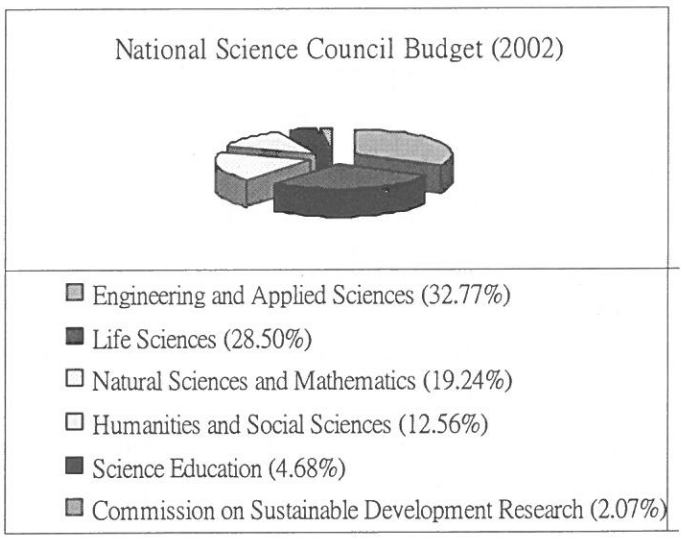


Fig. 1. Taiwan's NSC budget for the year 2002

Taiwan has recognized the importance of the biotechnology industry in development for

the 21st century. Taiwan's biotechnology industry has continuously increased investment in R&D, and attracted strong private sectors. As of 1997, there were around 70 centers with 1,500 employees. Taiwan's achievements in science and technology and in the number of scientific dissertations and patents are similar to those in developed countries. The success in science and technology has been mainly due to the result of strong encouragement by the NSC and government measures to promote the specialized areas. It is appropriate to say that science and technology development in the R.O.C has accomplished considerable achievements and that Taiwan can be promoted as a scientifically advanced nation.

3. GENERAL EDUCATION CURRICULUM – MAKING SCIENCE ACCESSIBLE TO ALL

a. The Beginning of the General Education Curriculum

In 1982, during a symposium on theory and practice in university liberal education, it was proposed to give importance to 'personality education' along with professional education in the universities. In the following year, National Taiwan University (NTU) officially began a "general education" program by offering introductory courses on natural sciences, social sciences and appreciation of art. The Ministry of education realized that the university education at that time had become so specialized that students lack understanding of other disciplines and thus were unable to possess a comprehensive view on modern development. In 1984, the Ministry of Education in Taiwan promulgated amendments to the education policy of the country introducing the "general education curriculum" [3]. The general education curriculum was designed after the liberal education model of US, to provide the students with basic knowledge in arts and literature, physical and natural sciences and technology, logic and philosophy, history and culture. Thus every university student was required to choose 8 credits in courses under the general education curriculum.

The aims of the general education curriculum were as follows:

1. To foster interdisciplinary learning and experience among students and a habit of lifelong learning.
2. Foster students' sense of self-respect and help them to understand and reflect upon the

meaning of human dignity and genuine care for the society, people and the environment.

3.To help students to develop a holistic and happy life in the modern society, to broaden their horizon, develop cultural awareness, and the ability to think and a healthy personality.

b. The Organization

At this time, the general education curriculum in a university in Taiwan is organized by a center for general education. The center has its own full-time and part-time faculty who are involved in preparation, teaching and research on the general education curriculum. The center also invites faculties from different disciplines to offer courses under the general education curriculum. For example, a professor from the computer science department would be asked to teach introduction to computer science under this general curriculum course. A professor from the chemistry department would be asked to teach a course in ethical issues involving chemical plants in the city, and so on. To strengthen the general education curriculum nationwide, in 1994, Chinese Academy of General Education was created to plan and promote general education among universities in Taiwan. This society publishes the journal of General Education to promote research and exchange of ideas on general education in Taiwan [5]. The general education curriculum consists of courses from three broader sections of humanities and arts, Natural sciences and technology and social sciences. The courses on humanities and art are attempt to inspire and increase self-understanding of the students, foster appreciation of art and literature and instill understanding of the meaning and value of life. The natural sciences and technology courses aspire to help students to recognize the contribution of modern science and understand scientific approaches and promote harmony between human and nature. The social science courses promote good inter-personal relationships, inculcate concern for society and involvement in social issues and foster positive ethical and moral values [4].

Thus the general education curriculum in Taiwan plays an essential role in making the knowledge and developments in sciences and humanities available to all. But in reality the curriculum does not seem to have received equal status, and is often looked upon as secondary to regular curriculum by the staff and students, and hence there is a lack of quality in planning and teaching the courses [6]. On the other hand, the government, as a part of its action plan for building a technologically advanced nation, encourages the harmonious devel-

opment of sciences and the humanities by promoting interdisciplinary collaboration and research [7]. The developmental plan of the National Science Council (NSC) in Taiwan outlines strategies to improve and foster scientific thinking and awareness among the public[8].

c. General Education Curriculum on Science in Taiwan Universities

To understand the quality of general education curriculum on science in Taiwan universities, we assessed the science courses offered in selected universities. The courses offered during this semester only were considered. Information about the courses were obtained mainly from various university websites [12-17]. Since enrollment for the courses is often done on the Internet in most universities in Taiwan, the information we obtained is expected to be current. The syllabi of some of the courses were downloaded to understand their content. A sample of courses offered in Science under general education curricula in these universities is presented in Table 2.

Table 2. Sample of courses offered in Science under general education curriculum

1.Space and ocean world	1.Public Health
2.Consumer chemistry	2.Bioethical questions
3.Search for extraterrestrial life	3.Sex and life
4.Understanding natural disasters	4.Discoveries in geology
5.Sustainable development and ecology	5.Achievements of Science in 20th Century
6.Bioethics	6.Biodiversity and conservation
7.History of western science	7.Human resources
8.Introduction to Science fiction	8.Environmental pollution management
9.Economics of Nature	9.Modern technology and society
10.Discussion on Biotechnology	10.Applied biotechnology
11.Development of Engineering and social change	11.Biology and life
12.Dialogue between Science and Humanities	12.Ecological pollution and environmental science
13.Engineering ethics	13.Agricultural technology and life
14.Animal ethics and public policy	14.Engineering and life
15.Advances in modern biology	15.medicine and health
16.Statistics and public policy	16.Artificial intelligence and modern science

Since the general curriculum is aimed to provide students with fundamental knowledge on sciences, all the universities offer introductory science courses. These courses offer the students basic understanding of different branches of science. The students who might have not studied science since their high school years get an opportunity to understand modern sci-

ence through these courses.

The importance of biological sciences in modern times is evident from the fact that there are more courses on biological sciences than any other branch of science in the general education in these universities. Ecology and biotechnology related courses are offered in all the universities. Surprisingly, not many courses on computers and information technology have been offered. Only a course on modern technology and society from Chung Shan University had a small influence on information technology in society [9]. Though there are not many courses on bio or medical ethics as such, most of the courses on biotechnology do include ethical questions raised by advances in bio-medical sciences.

There is a lack of courses on physics, chemistry and engineering in the general education curriculum on sciences in Taiwan's universities. One of the reasons could be the difficulty in popularizing science and in making it intelligible and interesting to non-science students. The success of the general education curriculum in science depends on the generosity and willingness of regular faculty from respective departments. But the demand on the science faculty to do research and publish is high, and faculty members may not have enough time and energy to teach in general education.

4. FU JEN CATHOLIC UNIVERSITY –HOLISTIC EDUCATION CENTER

a. General Education in Fu Jen – Holistic Education Center

Fu Jen's Holistic Education Center originated from the common courses program of the Division of Academic affairs in 1996. Other than making adjustments to the required courses and credits of the particular departments, Fu Jen has also enhanced the integration of the different curricula. These courses are motivated to create awareness and educate the students on the impact of scientific advances on Taiwanese Society and enable them to evaluate and judge the role of science and technology in human society. Fu Jen's Holistic education curriculum has three major parts: 1).Holistic Education Fundamental Courses, 2).Language and Cultural Development courses. 3).General Education courses. The general education courses comprise three different areas such as humanities and art, natural science and technology and social science.

b. Special Contribution of Fu Jen to Education in Taiwan

As noted earlier, the general education curriculum was aimed to provide a holistic education to the university students and thus prepare them with skills needed to be a complete human person. The special contribution of Fu Jen to the general education curriculum is the holistic fundamental courses, which are offered at different periods of students' lives.

Introduction to university studies is offered to the first year students at the beginning of the academic year. This course helps students to understand the meaning and purpose of university education, the mission and goal of Fu Jen. This helps the students to get acquainted with respective departments, extra-curricular activities, library system on and off campus and other resources, and foster good study habits.

Philosophy of Life is offered in the second year, for two semesters and it aims to guide students towards self-understanding, independent thinking and a formation of character. It helps them to reflect upon religion, art, ethics, rationality, ecology, society, politics, physical and mental health and understand the value and dignity of human life.

The course on professional ethics, offered in the third year, fosters a sense of responsibility and morality. It helps students to recognize ethical issues in their profession and cultivate skills to analyze and make appropriate decisions when faced with ethical dilemma.

These courses are well organized in Fu Jen, with a committee to plan and supervise each of them. Workshops are conducted periodically to assess and update the curricula and the skills of the teachers. The response of the students to these courses has been positive [10]. Initially the students may not be comfortable in discussions or may actually be uninterested or bored. But nevertheless, these special courses are ways Fu Jen uses to promote its mission, to share genuine knowledge, and nurture the holistic growth of the whole human person.

c. General Education Courses in Science at Fu Jen

The general education courses in science at Fu Jen are not very different from those in other universities (see Table 3). Here too, the introductory courses on Biology and Computer Science dominate the scene. Courses on Psychology are included in this section, since the Department of Psychology is part of the College of Science and Engineering. Fundamental and applied courses on computer science are offered as well, which was not the case in other

universities examined. At Fu Jen as well, there is no course which directly deals with ethical questions raised by science and technology except for the course on current topics on Bio-medical sciences, which has a chapter on ethical questions. But on the other hand, some of the required courses like Philosophy of life and Professional ethics do concentrate briefly on some of the ethical questions raised by science and technology.

Table 3. Science and Technology Courses in Holistic Education at Fu Jen

Natural Science and Technology courses	
1. Introduction to Computer Science.	18. Electricity and living.
2. Applied computer technology.	19. Fascinating mathematics
3. Computer and office automation	20. The enlightenment of mathematics.
4. Use of internet resources.	21. Chemistry and our world.
5. Introduction to computing and Networking application.	22. Introduction to ethology.
	23. Introduction to medicine and genetics.
6. Introduction to computer networks.	24. Current topics in biomedical science.
7. Information competency.	25. Medicine and living.
8. Electronic commerce security concepts.	26. Immunity and life.
9. Homepage and web server design.	27. Nutrition and health.
10. Use of library software.	28. Health promotion and protection.
11. Digital photography and image processing.	29. Diet and health.
12. Introduction to natural science.	30. Body weight control and health.
13. Introduction to biotechnology.	31. Environmental pollution and health.
14. Introduction to bioinformatics.	32. Environmental protection.
15. History of modern science.	33. Current and future global environment.
16. Introduction to life sciences.	34. Psychology.
17. Introduction to optoelectronic industry.	35. Textiles and fashion.

d. A Role for Fu Jen University in Taiwan

The education environment in Taiwan has changed since the re-foundation of Fu Jen in Taiwan [10].

Some of the important changes are:

1. The number of universities both private and public universities and technical colleges has increased dramatically in the past 10 years, and the number of placements available will be more than the number of applicants. That means that Fu Jen as an individual university is not offering a valuable service in education but finds itself in competition with other institutions.

2.Fu Jen has also become a research university with a well qualified faculty with research funding and publications. Similar developments have occurred in other universities as well and the public universities are better funded and equipped. It is nearly impossible to match up to them. Thus, the relevance of Fu Jen as a Catholic university in Taiwan is under review.

3.The Catholic character and identity of Fu Jen has changed as the number of Catholic students and staff has declined along with a drop in the number of members of religious orders working in the departments. The maintenance and revival of the Catholic character of the university depends upon the commitment of Catholic administrators and faculty and entities like holistic education centre, campus ministry and offices for mission and identity.

Under the above-mentioned environment, Fu Jen has moved from being a unique university, which offered high quality education and personal development to one being one among many good universities. Fu Jen is at a decisive point of history to be relevant and find a place for itself in education in Taiwan. We propose that, on a leading Catholic university with a strong science faculty, Fu Jen has a special role in making science accessible to all people. This can be achieved through the following:

1.The holistic education centre has an important role in achieving this goal. The natural science and technology section of the general education curriculum has many basic courses, which introduce the fundamentals of science and technology. But courses on relationship between science, technology and society must be developed to educate the students to be aware of the technology influence on our daily lives. For example the courses on computer science often focus on computer applications, software and Internet, but no course which deals with the impact of information technology on society. The authors of this paper are putting together a course on the impact of information technology and biotechnology on human values.

2.Introducing topics on technology and human society in the philosophy of life course would help the university have a wider impact. Being a required course in the university, the philosophy of life course has wide impact. This course was designed to introduce moral and religious values to the students and until now often taught by Catholics, priests, or members of religious orders. This course often includes introduction to major religions, topics on human relationships and moral questions related to the meaning of life. There is a need

to modify the syllabus of philosophy of life courses to include sections on science and technology, and with the help of the college of Science and Engineering, a course in which teachers could prepare materials more relevant for modern times.

3. Applied ethics is an area in which Taiwan's secular society will look to institutions like Fu Jen for inspiration and guidance [10]. The professional ethics course is a beginning, but there is a need to develop dialogue with wider society on ethical issues. There is an increased awareness of the need for understanding of the social and ethical impact of science and technology. The Science, Technology and Society (STS) network in Taiwan and the Ethical, legal and social issues (ELSI) branch in Taiwan in recent years have began to educate the public and promote researches on ethical implications of biotechnology and genomic medicine. Fu Jen University has the potential to be a leader in such research and promotion. Fu Jen has a strong philosophy and religion department, which could collaborate with other science and engineering departments to develop and promote relevant ethical thinking for Taiwanese society.
4. The Fu Jen Centre for Science and Religion can play an important role in promoting research on religion and science questions in the context of Chinese culture. This centre was established in 2001 and has been involved in research on science and religion. It has organized an international conference on science and religion and has invited international scholars in the field for public talks. This centre has a potential to promote the relationship between major religions in Taiwan and science. There is a dearth of information on Buddhist and Taoist responses and understanding of developments in science. This center could bring scholars in religion and science together to develop a spirit of dialogue and sharing of responses. Fu Jen, in its role as a leading private university run by a religious group, can play an important role in promoting well-founded public opinion and government policies.
5. The Fu Jen Catholic University does have a place of honor in Taiwan society as one of the older and famous universities. A leadership role in the promotion of ethical aspects of science and technology would be expected, and this could be a new niche for Fu Jen in Taiwan society.

5. CONCLUSION

As Taiwan becomes one of the developed countries in science and technology, the need for education of the public in science becomes apparent. In order to enjoy the benefits of science and avoid misuse and abuse of science, and to promote and regulate scientific advances, there is a need to increase scientific literacy of the people. The general education curriculum for universities in Taiwan has made holistic education of people as its mission. The ideals of this curriculum are yet to be realized in the classrooms and in the lives of students. In this setting, as a leading Catholic university, Fu Jen has a leadership role to play in Taiwan. With the strength of its committed science and humanities faculty, through its courses in philosophy of life, professional ethics and a well-planned general education curriculum in science, the university can make more of an impact in bringing science to the public and building a well-informed technically advanced nation.

REFERENCES

(1) Shun-Fen CHEN, "General Education in Taiwan's Universities: Retrospect and Prospect" (Article written in chinese), *Bulletin of Educational Research*, Vol. 47 , pp 283 - 300 , 2001

(2) Yi-Ying Huang, Re-knowing Integrated Curriculum and the System-fractal Model for the Development of Integrated Curriculum: Searching Another Way Out (Article written in chinese) *Chinese Journal of Science Education*, Vol. 11 No. 2 , Pages 211 - 234 , 2003

(3) Ding-Mao ZHONG, "Ideals and reality of General education" (Article in Chinese) Providence University, 2003

(4) http://www.hec.fju.edu.tw/school_1/englishmain.html (Fu Jen's Holistic Center Websites)

(5) http://www.ncu.edu.tw/~cage/ (Chinese academy of general education website)

(6) Jin-Jei, HUANG and Kun-Gin HUANG, "The report of evaluation of General Education at universities in Taiwan", (Article in Chinese) Ministry of Education, 1999. (website: http://www.ncu.edu.tw/~cage/learning/plan03.html

(7) ROC Executive Yuan, 1998 "Action plan for building a technologically advanced

nation" (Website: http://www.nsc.gov.tw/pub/techNation/tcchNationE.htm)

(8) National Science Council, R.O.C., National Science and Technology Development Plan (Fy2001 – Fy2004), abstract (in English), Website: http://www.nsc.gov.tw/tc/abstract.html

(9) http://tselcrs.nsysu.edu.tw/menu5/oldoutlinedown.asp?yrsm=922&CrsNo=A4107

(10) Budenholzer, Frank, 2003 "Reflections on Fu Jen as a Catholic University" in Proc. of 2003 workshop on mission and identity, Fu Jen University, Taiwan,

(11) Budenholzer, Frank, 2002 "Teaching professional ethics to science and engineering undergraduates" ACUCA sponsored lecture.

(12) Providence Universisty : http://www.pu.edu.tw/~pu0410/

(13) National Sun Yat-Sen university: http://www.general.nsysu.edu.tw/

(14) National Central University: http://www.ncu.edu.tw/~cge/index1.htm

(15) Tong Hai University: http://www2.thu.edu.tw/~ge/ge.html

(16) National Taiwan university: http://ccms.ntu.edu.tw/~cge/english/

(17) National Taiwan Normal university: http://www.ntnu.edu.tw/aa/aa5/

received September 24, 2004
revised October 13, 2004
accepted October 15, 2004

天主教大學在台灣應扮演讓科學知識普及化的角色

周 賜 福
輔仁大學資訊工程系

司 馬 忠
輔仁大學全人教育中心

摘 要

科技發展在台灣扮演著一個非常重要的角色。因體認生物科技與資訊工程的潛能，台灣的大學與研究中心重視這方面的研究；同時，人們也需要瞭解辨識科技發展對社會的影響。因此，政府建立科技與社會網路 (STS)、倫理法律與社會等人文議題 (ELSI)的相關網站提供公衆討論。

台灣教育部推薦各大學開設一些科技方面的通識課程，使學生瞭解科技發展對社會的影響。在這論文裡，我們評估部份大學，尤其是輔仁大學的自然科技通識課程，建議天主教大學在科技與社會的教育上如何扮演重要的角色。

Synthesis of a New Series of Mesogenic Compounds of Cyclohexyl 2-(4'-Alkylphenyl) Nicotinates

Win-Long Chia, * Chung-Sheng Lin, Yung-An Wang

Department of Chemistry

Fu Jen Catholic University

Taipei, Taiwan, Republic of China 242

Abstract

This study describes an efficient approach to synthesize a new series of mesogenic compounds by using of 4-alkylphenyl magnesium bromide to react with cyclohexyl *N*-phenyloxycarbonylnicotinium chloride and subsequent oxidation by *o*-chloranil to give regioselective cyclohexyl 2-(4'-alkylphenyl)nicotinates （CnPNCH, n=4-8）in a fair to good yield (49-68%).

Keywords: Synthesis; pyridine-containing compounds; mesogenic compounds.

INTRODUCTION

The synthesis of substituted pyridines by the reaction of Grignard reagents and pyridine is not a practical method due to the strenuous conditions required for addition, the low yields obtained, and frequent lack of regioselectivity[1]. Recently, we have been studying the synthesis of substituted pyridines via the addition of 1-acylpyridinium salts[2]. The 1-acylpyridinium salts are so reactive toward Grignard reagents that addition to the pyridine ring will occur in the presence of other reactive functional groups such as ketones, halides and esters[3]. Application of this methodology toward synthesis for liquid crystalline compounds have been proven to be quite successfully in our group[4].

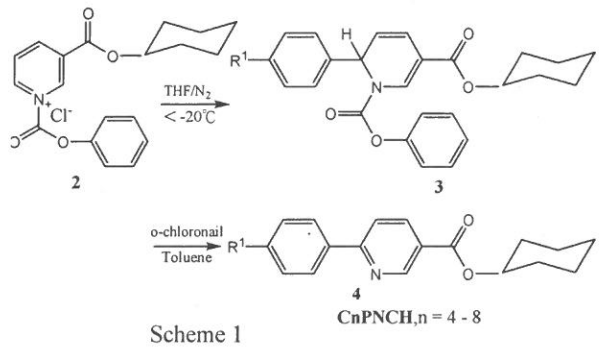
* Corresponding author. Tel: 2-29053567; Fax:2-29023209
E-mail:chem1008@mails.fju.edu.tw

Liquid crystalline phases can be exhibited by molecules having several different general molecular shapes. Only calamatic liquid crystals have found any technological importance[5]. The anisotropic shape and resulting anisotropic forces of a rod-shaped molecule give rise to the formation of liquid crystal phases. Nearly, all liquid crystal molecules contain at least one terminal alkyl chain. The most commercially re-known liquid crystal molecule is 4-n-pentyl-4'-cyanobiphenyls (5-CB) which was introduced in 1972[6]. Since then, a large amount of liquid crystalline molecules were synthesized. Pyridine-containing liquid crystals were extensively noticed recently[7]. In this paper, we would like to report the synthesis of a series of mesogenic compounds of cyclohexyl 2-(4'-alkylphenyl)nicotinates (CnPNCH, n=4-8) , which, however, were found do not show any mesogenic phase thermotropically.

RESULTS AND DISCUSSION

Synthesis of cyclohexyl 2-(4'-alkylphenyl)nicotinates :

Efficient syntheses of cyclohexyl 2-(4'-alkylphenyl)nicotinates (CnPNCH, n=4-8) 4 were carried out first by using Grignard reagent 4-alkylphenylmagnesium bromide 1 to react with cyclohexyl *N*-phenyloxycarbonylnicotinium chloride 2 to give crude 1,2- dihyhdropyridines 3, which were then subsequently aromatized by reaction with *o*-chloranil to acquire regioselective cyclohexyl 2-(4'-alkylphenyl)nicotinates 4 (Scheme 1). Yields of 4 were good (49-68%, Table 1) indicates the high reactivity of α-position of pyridinium ring that prohibit the attack of Grignard reagent toward ester functional group.



Scheme 1

Table 1. Synthesis of cyclohexyl 2-(4'-alkylphenyl)nicotinates 4a-4e (CnPNCH, n = 4-8).

Entry	n	Mp (℃)	Yield[a](%)	Color
4[a]	4	74	68	pale yellow
4[b]	5	63.1	58	white
4[c]	6	62.7	62	white
4[d]	7	60.1	49	white
4[e]	8	54	66	white

[a]Isolated yields by column chromatography (dichloromethane/hexane=2:1) on silica gel.

A representative synthetic procedure of compounds CnPNCH (n=4-8) is described[8]. All compounds obtained are further purified through column chromatography (dichloromethane : hexane = 2 : 1) and re-crystallization (dichloromethane : methanol = 1 : 5) . Chemical structures of the pure products were identified from ^{1}H-NMR, ^{13}C-NMR and infrared spectra. One ^{13}C-NMR spectrum of cyclohexyl 2-(4'-pentylphenyl)nicotinates (CnPNCH, n=5) shown in Figure 1, other spectra are listed in the reference[9]. All peaks in Figure 1 could be assigned clearly from an authentic spectrum of cyclohexanol and from the knowledge of ^{13}C-NMR. The substituted aromatic carbon atoms can be distinguished from the unsubstituted aromatic carbon atom by its decreased peak height; that is, it lacks a proton and thus suffers from a longer T_1 and a diminished NOE. Shifts of the aromatic carbon atom directly attached to the substituent have been correlated with substituent electronegativity after correcting for magnetic anisotropy effects.[10]

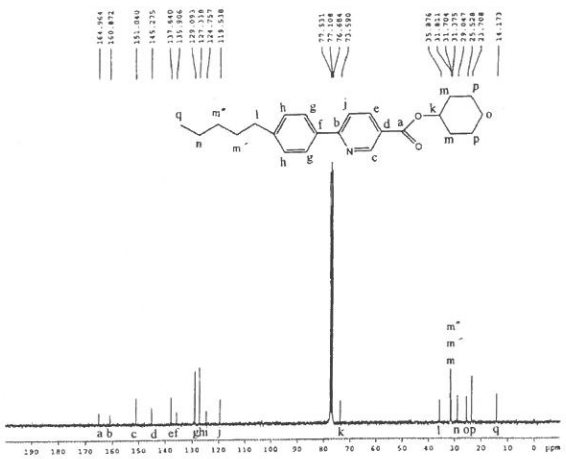


Fig. 1. ^{13}C-NMR spectrum of of cyclohexy 2-(4'-pentylphenyl)nicotinates. (CnPNCH, n-5)

This series of compounds (CnPNCH, n=4-8) , were not synthesized and studied thermotropically before and subjected to thermal analysis by differential scanning calorimetry and polarized optical microscopy. It was found surprisingly that no mesophase under thermo-scan from their solid forms to corresponding isotropic liquid forms, although these compounds are calamatic and mesogenic in nature. Compared to the analogous phenyl 2-(4'-alkylphenyl)nicotinates (CnPNP, n=4-8)[4b], which show a strong tendency of a semectic A phase in a wide temperature range (100-70°C) when cooling.

In conclusion, this new homologous series of pyridine-containing mesogenic compounds, cyclohexyl 2-(4'-alkylphenyl)nicotinates (CnPNCH, n=4-8), indicated no mesophases, which can be observed under polarized optical microscope. By contrast, its benzene analogs, phenyl 2-(4'-alkylphenyl)nicotinates, has shown mainly in semectic A phase. Further research is undertaken based upon these studies, to synthesize a new series of promising liquid crystalline compounds, 5-cyano-2-(4'-alkylphenyl)pyridines, which is an analogous series of n-CB.

Acknowledgement: Financial support of our work by the cross cultural center and the SVD section of Fu Jen Catholic University is gratefully acknowledged.

REFERENCE

(1) (a)Goetz-Luthy, N.; *J. Am. Chem. Soc.* 1949, *71*, 2254. (b) Comins, D. L.; Abdullah, A. H.; *J. Org. Chem.* 1982, *47*, 4312.

(2) (a) Chia, W-L., Shiao, M-J., 1991, *Tetrahedron Lett.*, 32, 2033-2034. (b) Shing, T-L., Chia, W-L., Shiao, M-J., Chau, T-Y., 1991, *Synthesis,* 849-850. (c) Shiao, M-J., Chia, W-L., Peng, C-L.: Shen, C-C., 1993, *J. Org. Chem.,* 58, 3162-3164. (d) Shiao, M-J., Shih, L-H., Chia, W-L., Chau, T-Y., 1991, *Heterocycles*, 32, 2111-2118.

(3) (a)Comins, D. L.; Stroud, E. D.; Herric, J. J.; *Heterocycles*, 1984, *22*, 151. (b)Comins, D. L.; Stroud, E. D.; *Heterocycles*, 1986, *24*, 3199.

(4) (a)Chia, W-L.; Shen, S-W.; Lin, H-C.; *Tetrahedron Lett.* 2001, *42*, 2177. (B)Chia, W-L., Hsia, C.-P., Lin, H.-C., was sent to Heteroatom Chemistry.

(5) (a)Bahadur, B.; *Liquid Crystals: Applications and Uses*, vol.1, 1990, Singapore: World Scientific. (b)Collings, P. J.; Hird, M.; *Introduction to Liquid Crystals: Chemistry and*

Physics, 1997, London: Taylor & Francis Ltd.

(6) Gray, G. W.; Harrison, K. J.; Nash, J. A.; *Electronics Letters*, 1973, *9*, 130.

(7) (a)Burrow, M. P.; Gray, G. W.; Lacey, D.; Toyne, K. J. *Liquid Crystals* 1988, *3*, 1643. (b)Reiffenrath, V.; Bremer, M.; *Angew. Chem. Int. Ed. Engl.* 1994, *33*, 1386.

(8) For 3a: To a solution of 1-bromo-4-butylbenzene (10 mmol) in 20 ml THF was added freshly dried magnesium granules (11 mmol) under an inert atmosphere. The Grignard solution 1 was then slowly added by syringe into a preformed solution of nicotinium chloride 2 (10 mmol phenyl chloroformate, 10 mmol cyclohexyl nicotinate, 20 ml dry THF at -20℃, 0.5 h) at -20℃. The resulting solution was warmed slowly to room temperature and stirred for another 8 h. After evaporation of the THF, the residue was extracted with ether. The crude 3a was obtained by washing twice with 10% HCl solution and brine and by drying with magnesium sulfate. For 4a (CnPNCH, n=4): To a solution of 20 ml dry toluene and crude 3a was added about 1.5eq. *o*-chloranil. The reaction mixture was heated to reflux for a number of hours under inert atmosphere and then quenched by adding 25 ml 1N NaOH solution and 25 ml ethyl ether and filtered through celite. Normal aqueous work up and isolation with column chromatography (SiO_2, methylene chloride: hexane=2:1) affords the product 4a (68%). The analytically pure pale yellowish crystals 4a were obtained after several times of re-crystallizations from the combination solvent methylene chloride: methanol (1:5).

(9) ^{1}H NMR (300MHz, $CDCl_3$). 4a: 9.10 (d, 1H, J=1.8 Hz), 8.15 (dd, 1H, J_1=8.4 Hz, J_2=2.1 Hz), 7.81 (d, 1H, J=8.1 Hz), 7.61 (d, 1H, J=8.4 Hz), 7.14 (d, 1H, J=8.1 Hz), 4.91 (quin, 1H, J=3.9 Hz), 2.51 (t, 2H, J=7.5 Hz), 1.70-1.85 (m, 2H), 1.55-1.70 (m, 2H), 1.18-1.51 (m, 10H), 0.80 (t, 3H, J=7.2 Hz); 4b: 9.27 (d, 1H, J=2.1Hz), 8.32 (dd, 1H, J_1=8.4 Hz, J_2=2.1 Hz), 7.97 (d, 1H, J=8.1 Hz), 7.78 (d, 1H, J=8.1 Hz), 7.31(d, 1H, J=8.1 Hz), 5.08 (quin, 1H, J=3.9 Hz), 2.67(t, 2H, J=7.2 Hz), 1.90-2.05 (m, 2H), 1.75-1.90 (m, 2H), 1.20-1.75 (m, 12H), 0.91 (t, 3H, J =4.8 Hz); 4c: 9.27 (d, 1H, J =2.1 Hz), 8.32 (dd, 1H, J_1=8.4 Hz, J_2=2.1 Hz), 7.97 (d, 2H, J =8.1 Hz), 7.78 (d, 1H, J =8.1 Hz), 7.31 (d, 2H, J =8.1 Hz), 5.08 (quin, 1H J=3.9 Hz), 2.67 (t, 2H, J=7.5 Hz), 1.90-2.05 (m, 2H), 1.70-1.90 (m, 2H), 1.20-1.70 (m, 14H), 0.89 (t, 3H, J=6.6 Hz); 4d: 9.27 (s, 1H), 8.32 (dd, 1H, J_1=8.4 Hz, J_2=2.1 Hz), 7.97 (d, 2H, J =8.1 Hz), 7.78 (d, 1H, J =8.4 Hz), 7.31 (d, 2H, J =8.1 Hz), 5.08 (quin, 1H, J=3.9 Hz), 2.67 (t, 2H, J=7.2 Hz), 1.90-2.05 (m, 2H), 1.75-1.90 (m, 2H),

1.20-1.70 (m, 16H), 0.88 (t, 3H, *J*=6.6 Hz); 4e: 9.27 (d, 1H, *J*=2.1Hz), 8.32 (dd, 1H, J_1=8.4 Hz, J_2=2.1 Hz), 7.97 (d, 2H, *J* =8.1 Hz), 7.78 (d, 1H, *J* =8.4 Hz), 7.31 (d, 2H, *J* =8.1 Hz), 5.08 (quin, 1H, *J*=3.9 Hz), 2.67 (t, 2H, J=7.2 Hz), 1.90-2.05 (m, 2H), 1.75-1.90 (m, 2H), 1.20-1.70 (m, 18H), 0.88 (t, 3H, *J*=6.9 Hz). ^{13}C-NMR (75MHz, $CDCl_3$). 4a: 164.85, 160.76, 150.91, 145.15, 137.77, 135.77, 128.99, 127.24, 124.67, 119.47, 73.5, 35.46, 31.60, 25.43, 23.61, 13.92; 4b: 164.96, 160.87, 151.04, 145.27, 137.84, 135.91, 129.09, 127.34, 124.76, 119.54, 73.59, 35.88, 31.81, 31.70, 31.38, 29.05, 25.53, 23.71, 14.17; 4c: 164.97, 160.88, 151.05, 145.28, 137.85, 135.91, 129.10, 127.35, 124.76, 119.55, 73.60, 35.88, 31.82, 31.71, 31.38, 29.06, 25.54, 23.72, 22.70, 14.18; 4d: 164.96, 160.87, 151.04, 145.27, 137.83, 135.90, 129.09, 127.33, 124.75, 119.52, 73.58, 35.88, 31.97, 31.70, 31.41, 29.35, 29.26, 25.52, 23.70, 22.75, 14.17; 4e: 164.98, 160.89, 151.06, 145.30, 137.85, 135.92, 129.11, 127.35, 124.77, 119.55, 73.60, 35.90, 31.99, 31.72, 31.43, 29.58, 29.41, 29.36, 25.54, 23.72, 22.77, 14.20. FTIR (KBr). 4a: 3034, 2931, 1711, 1596, 778. 4b: 3029, 2927, 1715, 1596, 778. 4c: 3034, 2926, 1711, 1596, 777. 4d: 3034, 2923, 1711, 1595, 777. 4e: 3030, 2923, 1711, 1595, 774.

(10) Silverstein, R. M.; Bassler, G. C.; Morrill, T. C.; *Spectrometric Identification of Organic Compounds*, 4th ed., 1981, New York: John Wiley & Sons Inc.

received October 08, 2004
revised November 16, 2004
accepted November27, 2004

一系列環己烷基 6-(4-烷苯基)尼古丁酯新液晶化合物的合成

賈文隆 林忠生 王永安
中華民國台灣台北輔仁大學化學系

摘 要

這個研究是在描述一有效的含比啶液晶化合物的合成，其方法是利用烷基苯溴鎂與環己烷基-氮-苯氧醯基氯化尼古丁反應，所得之中間物再用氯化鄰苯二酮氧化，便可得一很好的產率且具區域位置選擇性的一系列新化合物環己烷基 6-(4-烷苯基)尼古丁酯。此系列化合物雖具有液晶化合物的本質，但卻沒有呈現任何液晶相。

關鍵詞：合成，含比啶化合物，液晶化合物

Cumulant-based Blind Signal Separation with Subspace Projection Techniques in Multipath Propagation

Jung-Lang Yu*, Ding-Bing Lin**

Abstract

This paper presents the eigenspace-based CBOB algorithm which utilizes the eigenspace projection beamforming techniques to enhance the performances of the cumulant-based blind beamforming (CBOB) algorithm. In the beamforming stage of the CBOB algorithm, the minimum variance distortionless response (MVDR) beamformer is used to extract each signal while suppressing interferences. However, the MVDR beamformer is very sensitive to the steering vector errors, which cause the signal cancellation effects. We use the eigenspace-based beamformer instead of the MVDR beamformer to reduce signal cancellation and increase the output signal-to-interference-plus-noise ratio (SINR). Computer simulations are given to demonstrate that the eigenspace-based CBOB outperforms the CBOB.

Keywords: blind signal separation, CBOB, MVDR beamformer, ESPRIT algorithm

I. INTRODUCTION

Wireless communication systems are rapidly growing in volume and range of services. A major challenge for these systems today is the limited radio frequency spectrum available. Approaches that increase spectrum efficiency are therefore of great interest. One promising

* Department of Electronic Engineering, Fu-Jen Catholic University Hsinchuang, Taipei Hsien 24205, Taiwan, R.O.C.
Phone:+886-2-29052102, Fax:+886-2-29042638
Email: yujl@ee.fju.edu.tw
** Department of Electronic Engineering, National Taipei University of Technology Chuang Hsiao East Road, Taipei 106, Taiwan, R.O.C.
Phone: +866-2-27712171-2274 Fax:+866-2-227317120
Email: dblin@en.ntut.edu.tw

approach is to use antenna arrays at cell sites. Array processing techniques are used to receive multiple signals that are separated in space. Hence, multiple co-channel users can be supported per cell to increase capacity. Besides, the blind signal separation techniques are used to save the cost of training or pilot sequences. For example, clustering algorithms [1][2], finite alphabet methods [3]-[5], constant modulus algorithms [6][7] and high-order statistics algorithms [8]-[14] have been proposed recently.

This paper combines the eigenspace projection beamforming techniques [15] with the CBOB algorithm [13] to blindly separate the co-channel digital signals in multipath propagation. In CBOB, the generalized steering vectors are first estimated. Then using the estimated generalized steering vectors to compute the MVDR beamforming weights and to extract the signals of interest. However, the MVDR beamformer is very sensitive to the steering vector errors, which cause the signal cancellation effect on the signal of interest. A lot of approaches are proposed to reduce the effects of steering vector errors [15][16]. Among them the eigenspace-based beamformer [15] projects the MVDR beamforming weight vector onto the signal subspace to reduce the increase of the output noise power due to imperfection. It has been shown that the eigenspace-based beamformer is less sensitive to the steering vector errors than the MVDR beamformer. Here, we present the eigenspace-based CBOB by applying the eigenspace projection beamforming techniques to the CBOB. It is shown in simulation results that the proposed method has better performance in blind source separation than the CBOB.

II. PROBLEM FOUMULATION

Consider a signal scenario consisting D statistically independent narrowband source signals in the presence of multipath propagation. Assume that total of $L = \sum_{i=1}^{D} L_i$ multipath signals for L_i paths for each source $s_i(n)$ impinge on an M-element antenna array. It is assumed that the number of sources is less than the number of array sensors (i.e., $D < M$). We also assume that the channel is constant over the available N symbol periods. With these assumptions, the signal received by the array at time n is

$$\mathbf{r}(n) = \mathbf{A}\tilde{\mathbf{s}}(n) + \mathbf{v}(n) \tag{1}$$

where $\mathbf{r}(n) = [r_1(n), \ldots, r_M(n)]^T$, $\tilde{\mathbf{s}}(n)$, is a $L \times 1$ signal vector, $\mathbf{v}(n)$ is additive white noise with covariance $\sigma^2\mathbf{I}$, $\mathbf{a}(\theta_{d,k})$ is $M \times 1$ steering vector of the k-th path of the d-th source and $\mathbf{A}$ is an $M \times L$ unknown steering matrix

$$\mathbf{A} = [\mathbf{a}(\theta_{1,1}), \mathbf{a}(\theta_{1,2}), \ldots, \mathbf{a}(\theta_{1,L_1}), \mathbf{a}(\theta_{2,1}), \ldots, \mathbf{a}(\theta_{2,L_2}), \ldots, \mathbf{a}(\theta_{D,L_D})] \tag{2}$$

The signal vector $\tilde{\mathbf{s}}(n)$ can be expressed by

$$\tilde{\mathbf{s}}(n) = \mathbf{G}\tilde{\tilde{\mathbf{s}}}(n) \tag{3}$$

Where $\mathbf{G}$ is a $L \times L$ diagonal matrix, with diagonal term $\{c_{d,k}\}$ indicating fading factor

$$\mathbf{G} = \text{diag}[c_{1,1}, c_{1,2}, \ldots, c_{1,L_1}, c_{2,1}, \ldots, c_{2,L_{21}}, \ldots, c_{D,1}, \ldots c_{D,L_D}] \tag{4}$$

and $\tilde{\tilde{\mathbf{s}}}(n)$ is given by

$$\tilde{\tilde{\mathbf{s}}}(n) = [s_{1,1}(n), s_{1,2}(n), \ldots, s_{1,L_1}(n), s_{2,1}(n), \ldots, s_{2,L_2}(n), \ldots, s_{D,1}(n), \ldots, s_{D,L_D}(n)]^T \tag{5}$$

Where $s_{d,k}(n) = s_d(n - \tau_{d,k})$ represents the k-th path of source d signal and $\tau_{d,k}$ indicates the time delay. On the other hand, if propagation delay of multipath signal of each source is smaller than the inverse bandwidth of the signals, the frequency-selective multipath propagation is degenerated to the frequency-flat multipath propagation. In the frequency-flat multipath environment, $s_{d,1}(n) \approx \ldots \approx s_{d,L_d}(n) \equiv s_d(n)$ for all $d = 1 \ldots D$. Hence the signal vector in (3) is rewritten as

$$\tilde{\mathbf{s}}(n) = \begin{bmatrix} \mathbf{c}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{c}_D \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_D(n) \end{bmatrix} \equiv \mathbf{Q}\mathbf{s}(n) \tag{6}$$

where $\mathbf{c}_i = [c_{i,1}\, c_{i,2} \cdots c_{i,L_i}]^T$, $\mathbf{s}(n) = [s_1(n) \cdots s_D(n)]^T$ and $\mathbf{Q}$ is $L \times D$ fading factor matrix. In this paper, we focus our discussions on the case of frequency-flat multipath propagation. Hence the received signal vector in (1) becomes

$$\mathbf{r}(n) = \mathbf{A}\mathbf{Q}\mathbf{s}(n) + \mathbf{v}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{v}(n) \tag{7}$$

where $\mathbf{H} = \mathbf{A}\mathbf{Q} = [\mathbf{h}_1 \cdots \mathbf{h}_D]$. Columns of $\mathbf{H}$, $\mathbf{h}_i$, i=1$\cdots$D, are called generalized steering vectors. It is observed from (7) that the signals for frequency-flat multipath fading channel are equivalent to statistically independent narrowband sources with a modified steering matrix. Our objective is to blindly recover the source signals $\{s_i(n)\}_{i=1}^{D}$ by designing suitable beamformers. There have many methods been presented to achieve the goal [13][17][18]. Among them, the CBOB [13] uses fourth-order cumulants to estimate the generalized steer-

them, the CBOB [13] uses fourth-order cumulants to estimate the generalized steering vectors followed by the MVDR beamformers which are designed to extract the signals of interest. Here we propose the eigenspace-based CBOB algorithm, which combines the eigenspace projection beamforming technique with the CBOB, to enhance the performance of signal separation.

III. EIGENSPACE-BASED CBOB

In CBOB, the generalized steering vectors are first estimated by using fourth-order cumulant techniques and then the MVDR beamformers are designed to recover each source. However, the MVDR beamformer is very sensitive to the steering vector errors, which cause the signal cancellation effects. We utilize the eigenspace-based beamformer instead of the MVDR beamformer to alleviate the signal cancellation effect in what follows. First, we need to estimate the generalized steering vectors. The CBOB applies the ESPRIT algorithm [19] onto suitably defined cumulant matrices to derive the generalized steering vectors. The procedure is briefly summarized as follows:

Step 1: From the $M \times 1$ received data, estimate the $M \times M$ cumulant matrices

$$\begin{aligned} \mathbf{C}_1 &\equiv cum(r_1(n), r_1^*(n), \mathbf{r}(n), \mathbf{r}^H(n)) \\ \mathbf{C}_2 &\equiv cum(r_2(n), r_1^*(n), \mathbf{r}(n), \mathbf{r}^H(n)) \end{aligned} \tag{8}$$

and define the $2M \times M$ matrix $\mathbf{C}$ as $\mathbf{C} \equiv [\mathbf{C}_1^T \ \mathbf{C}_2^T]^T$.

Step 2: Perform SVD on $\mathbf{C}$. The number of nonzero singular values gives the number of independent sources D. Keep the first $2M \times D$ submatrix of the left singular vectors of $\mathbf{C}$, and denote this submatrix as $\mathbf{U}_1$.

Step 3: Partition $\mathbf{U}_1$ into two $M \times D$ submatrices $\mathbf{U}_{11}$ and $\mathbf{U}_{12}$.

Step 4: Perform SVD on $[\mathbf{U}_{11} \ \mathbf{U}_{12}]$. Stack the last D right singular vectors of $[\mathbf{U}_{11} \ \mathbf{U}_{12}]$ into the $2D \times D$ matrix denoted $\mathbf{F}$.

Step 5: Partition $\mathbf{F}$ as $\mathbf{F} \equiv \begin{bmatrix} \mathbf{F}_x \\ \mathbf{F}_y \end{bmatrix}$, where $\mathbf{F}_x$ and $\mathbf{F}_y$ are $D \times D$ matrices.

Step 6: Perform eigendecomposition on $-\mathbf{F}_x\mathbf{F}_y^{-1}$. Let the eigenvector and eigenvalue matrices of $-\mathbf{F}_x\mathbf{F}_y^{-1}$ be $\mathbf{E}$ and $\mathbf{D}$, respectively.

Step 7: An estimate of $\mathbf{H}$ is obtained to within a diagonal as follows:

$$\mathbf{HZ}^{-1} = \frac{1}{2}(\mathbf{U}_{11}\mathbf{E} + \mathbf{U}_{12}\mathbf{E}\mathbf{D}^{-1}) \tag{8}$$

where $\mathbf{Z}$ is an arbitrary diagonal matrix with nonzero entries.

After performing these steps above, we obtain the estimated generalized steering vectors $\{\hat{\mathbf{h}}_1,\ldots\hat{\mathbf{h}}_D\}$. With the generalized steering vector $\hat{\mathbf{h}}_i$, the weight vector of the *i*-th MVDR beamformer is computed by

$$\mathbf{w}_{i,MVDR} = \alpha_i \mathbf{R}^{-1}\hat{\mathbf{h}}_i \tag{9}$$

where $\alpha_i = (\hat{\mathbf{h}}_i^H \mathbf{R}^{-1}\hat{\mathbf{h}}_i)^{-1}$ and $\mathbf{R}$ is correlation matrix

$$\mathbf{R} = E[\mathbf{r}(n)\mathbf{r}^H(n)] \tag{10}$$

However, the performance of MVDR beamformer will be seriously degraded due to the imperfection of steering vector error. To enhance the performance of MVDR beamformer, the eigenspace-based beamformer generates its weight vector by projecting the MVDR beamforming weight vector onto the signal subspace [15],

$$\mathbf{w}_{i,sub} = \mathbf{U}_s\mathbf{U}_s^H\mathbf{w}_{i,MVDR} \tag{11}$$

where $\mathbf{U}_s$ is the signal subspace computed from the eigendecomposition of the correlation matrix. The matrix in (10) can be eigendecomposed as

$$\mathbf{R} = \sum_{i=1}^{M} \lambda_i \mathbf{e}_i \mathbf{e}_i^H \tag{12}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{D+1} = \cdots = \lambda_M = \sigma^2$ are eigenvalues in the descending order, and $\mathbf{e}_i, i = 1,\cdots,M$, are the corresponding orthonormal eigenvectors. The $\mathbf{e}_1,\ldots,\mathbf{e}_D$ eigenvectors span the subspace same as that spanned by $\mathbf{h}_1,\ldots,\mathbf{h}_D$, which is called the signal subspace. Therefore we have

$$\mathbf{U}_s = [\mathbf{e}_1\ \mathbf{e}_2\ \cdots \mathbf{e}_D] \tag{13}$$

The eigenspace-based beamformer can be also interpreted as the constraint projection beamformer [16], which projects the steering vector onto the signal subspace. The projected steering vector is given by

$$\hat{\mathbf{h}}_{i,sub} = \mathbf{U}_s \mathbf{U}_s^H \hat{\mathbf{h}}_i \tag{14}$$

Using (14), the constraint projection beamformer computes the weight vector by

$$\mathbf{w}_{i,sub} = \alpha_i \mathbf{R}^{-1} \hat{\mathbf{h}}_{i,sub} \tag{15}$$

With $\mathbf{w}_{i,sub}$ in (11) or (15), the i-th signal is estimated by

$$\hat{\mathbf{s}}_i(n) = \mathbf{w}_{i,sub}^H \mathbf{r}(n) \tag{16}$$

(16) can be performed in parallel for all D sources. $\hat{\mathbf{s}}_i(n)$ actually differs $\mathbf{s}_i(n)$ by an unknown scalar. This problem can be solved by using a finite training sequence or by the property of BPSK modulation scheme for $\mathbf{s}_i(n)$ [13].

IV. SIMULATION RESULTS

The performances of the eigenspace-based CBOB are examined in this section. A ten-element linear uniform array with half wavelength spacing is used for simulations. There are four BPSK source signals arriving at the array from two, three, three, and three different directions, respectively. Note that the total number of signals impinging on the array is 11, which is more than the number of sensors. The signal arrival angles and propagation constants were chosen as [55°, 35°] and [1, 0.7+j*0.6];[70°, 80°, 120°] and [1, 0.6-j*0.7, 0.5+j*0.8]; [40°, 90°, 60°] and [1, 0.8+j*0.5, 7+j*0.6]; [25°, 65°, 130°] and [1, 0.6+j*0.6, 7-j*0.7].

Fig. 1 and Fig. 2 show the signal constellations of CBOB and the eigenspace-based CBOB with sample number N=500 and SNR=-2 dB. It is shown that eigenspace-based CBOB has better signal constellation than the CBOB. Especially, the proposed method can extract the 4-th source successfully but the CBOB is failed. This is because the projection operation in (11) reduces the norm of weight vector $\mathbf{w}_{i,MVDR}$ with the array responses to the signals unchanged. Thus, $\mathbf{w}_{i,sub}$ produces lower output noise power than $\mathbf{w}_{i,MVDR}$ while both have the same output signal power and output interference power. That means the eigenspace-based CBOB has a higher output signal-to-noise-plus-interference ratio (SINR) than the CBOB method. Fig. 3 shows the beam patterns according to the weight vectors in (9) and (11) for all signal sources. It is observed that the eigenspace-based CBOB has higher gains in the

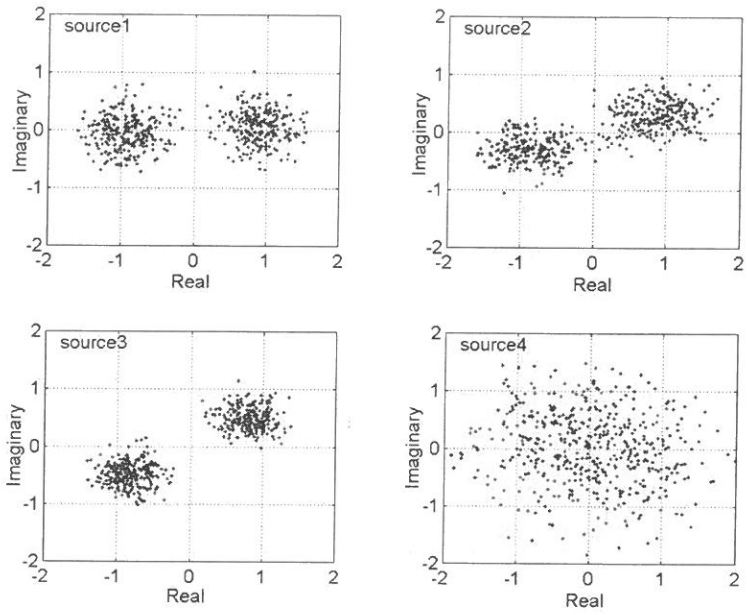


Fig. 1. Signal constellations of the CBOB method.

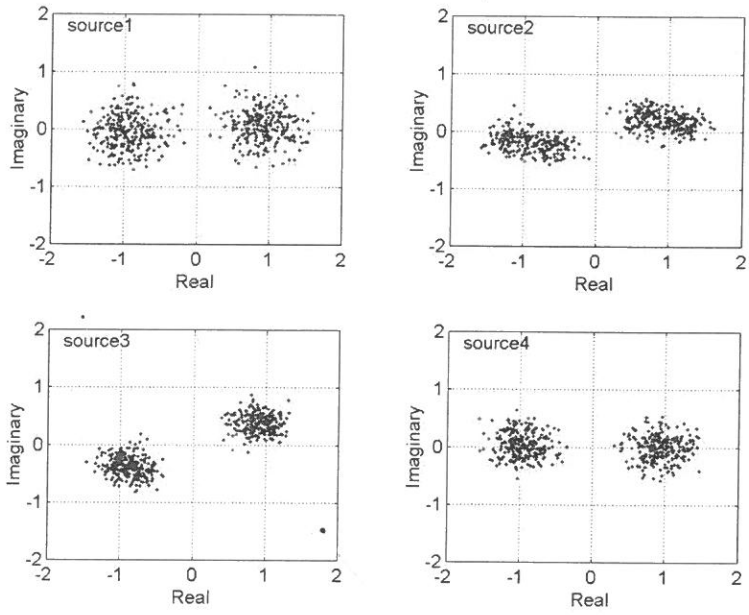


Fig. 2. Signal constellations of the eigenspace-Based CBOB method.

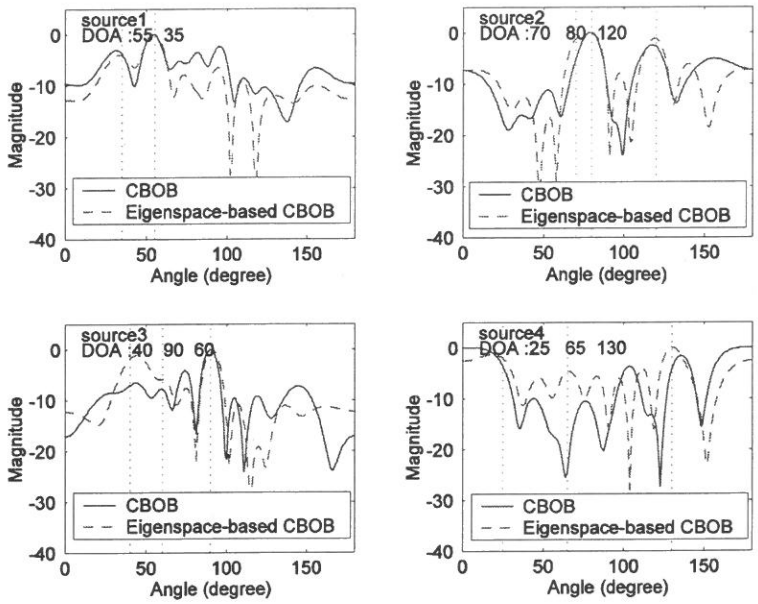


Fig. 3. Beam pattern for the weight vectors of the CBOB and the eigenspace-Based CBOB.

directions of desired signal than the CBOB. Further, the proposed method has a lower sidelobe level on average than the CBOB. Fig. 4 shows the output SINR versus the sample number when the input SNR=0dB. The dotdash lines indicate the optimal SINR calculated from the ideal correlation matrix and ideal generalized steering vector. The dotted and solid lines indicate the output SINRs computed by the proposed method and the CBOB method, respectively. It is found that the proposed method has better convergence rate than the CBOB. We also examine the BER versus the input SNR in Fig. 5. the input SNR varies from -10dB to 0dB. The proposed method still outperforms the CBOB.

V. CONCLUSIONS

We have presented the eigenspace-based CBOB algorithm that uses eigenspace projection beamforming technique to alleviate the signal cancellation due to imperfect steering vector. Since the projection operation in the eigenspace-based beamformer will reduce the output noise power while keep the output signal powers unchanged, the eigenspace-based CBOB generates higher output SINR than the CBOB. Simulation results were provided to demonstrate that the proposed approach has faster convergence speed and lower BER than the CBOB.

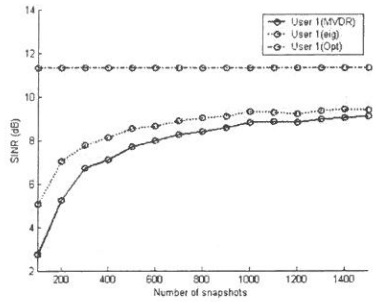


Fig. 4a. The output SINR of source 1.

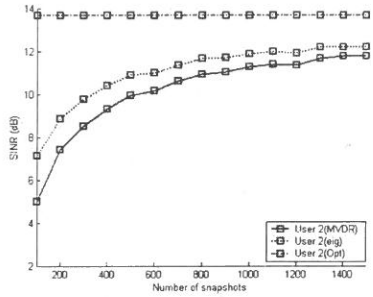


Fig. 4b. The output SINR of source 2.

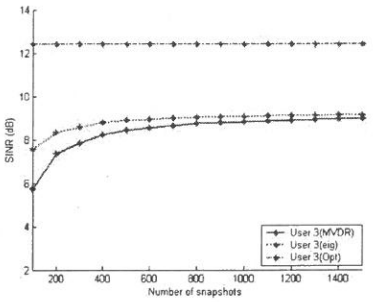


Fig. 4c. The output SINR of source 3.

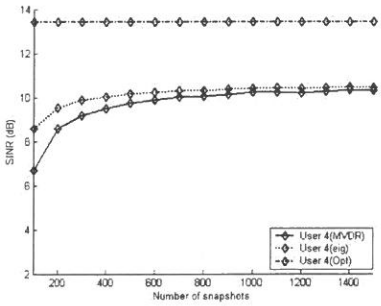


Fig. 4d. The output SINR of source 4.

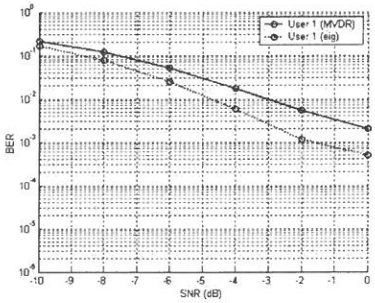


Fig. 5a. The BER of source 1

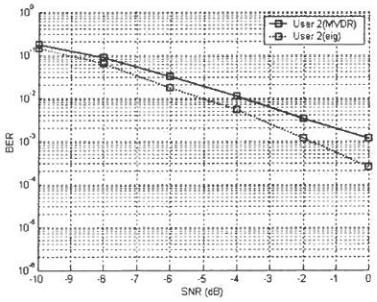


Fig. 5b. The BER of source 2.

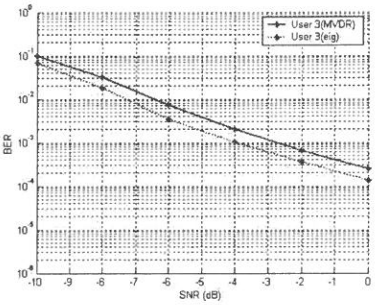


Fig. 5c. The BER of source 3.

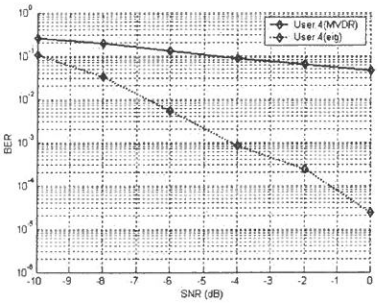


Fig. 5d. The BER of source 4.

REFERENCE

(1) A. Kannan, G..Mathew, and V. U. Readdy, "Blind separation of multiple co-channel bpsk signals arriving at an antenna array," *IEEE Transaction on Signal Processing Letters*, vol. SPL-2, no. 9, pp. 176-178, Sept. 1995.

(2) C. C. Lee and J. H. Lee, "An efficient method for blind digital signal separation of array data," *Signal Processing*, vol. 77, pp. 229-234, 1999.

(3) S. Talwar, M. Viberg and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Transaction on Signal Processing Letter*, vol. SPL-1, no. 2, pp. 29-31, Feb. 1994.

(4) S. Talwar, M. Viberg and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part I: Algorithms," *IEEE Transaction on Signal Processing*, vol. 44, no. 5, pp. 1184-1197, May 1996.

(5) S. Talwar, M. Viberg and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array — Part II: Performance Analysis," *IEEE Transaction on Signal Processing*, vol. 45, no. 3, pp. 706-718, Mar. 1997.

(6) A. J. van der Veen, "An analytical constant modulus algorithm," IEEE Transaction on Signal Processing, vol.44, no. 5, pp. 1136-1155, May 1996.

(7) L. Castedo, C. J. Escudero and A. Dapena, "A blind signal separation method for multiuser communications," *IEEE Transaction on Signal Processing*, vol. 45, no. 5, pp. 1343-1348, May 1997.

(8) J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceeding-F*, vol. 140, no. 6, pp. 362-370, Dec. 1993.

(9) J. F. Cardoso and B.H. Laheld, "Equivariant adaptive source separation," *IEEE Transaction on Signal Processing*, vol. 44, no. 12, pp. 3017-3029, Dec. 1996.

(10) C. B. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Transaction on Signal Processing*, vol. 48, no. 12, pp. 3508-3519, Dec. 1996.

(11) O. Shalviand E. Weinstein, "New criteria for blind deconvoluation of nonminimum phase systems (channels)," *IEEE Transaction on Information Theory*, vol. 36, no. 2, pp. 312-321, Mar. 1990.

(12) Z. Ding and T. Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming," *IEEE Transaction on Signal Processing*, vol. 48, no. 6, pp. 1587-1596, June 2000.

(13) E. Gonen and J. M. Mendel, "Applications of cumulants to array processing—Part III: Blind beamformer for coherent signals," *IEEE Transaction on Signal Processing*, vol. 45, no. 9, pp. 2252-2264, Sept. 1997.

(14) C. Y. Chi and C. Y. Chen, "Blind beamforming and maximum ratio combining by kurtosis maximization for source separation in multipath," *Proc. IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, pp. 20-23, Mar. 20-23, 2001.

(15) L. Chang and C. C. Yeh, "Performance of DMI and eigenspace-based beamformers," *IEEE Transaction on Antennas and Propagation*, vol. 40, no. 11, pp. 1336-1347, Nov. 1992.

(16) D. D. Feldman and L. J. Griffiths, "A projection approach for robust adaptive beamforming," *IEEE Transaction on Signal Processing*, vol. 42, no. 4, pp. 867-876, Apr. 1994.

(17) A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *IEEE Transaction on Signal Processing Magazine*, vol. 14, pp.49-83, Nov. 1997.

(18) Tsuhan Chen, editor, "Highlights of statistical signal and array processing," *IEEE Transaction on Signal Processing Magazine*, vol. 15, No. 5, pp.21-64, Sept. 1998.

(19) R. Roy and T. Kailth, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol.37, pp. 984-995, July 1989.

received, 2004
revised, 2004
accepted, 2004

在多路徑傳輸下使用子空間投射的統計盲目訊號分離技術

余金郎
輔仁大學電子工程系

林丁丙
台北科技大學電子工程系

摘　要

此篇論文提出特徵空間爲基礎的 CBOB 演算法則，它使用特徵空間投射波束形成器技術來加強 CBOB 演算法則的效能。在 CBOB 演算法則中，使用直接反矩陣波束形成器來抑除干擾並接收單一信號。但是直接反矩陣波束形成器對指向向量誤差非常靈敏，常造成訊號抵消效應。爲了改善此現象，並提高接收訊號的雜訊比，我們使用特徵空間波束形成器取代直接反矩陣波束形成器來達成盲目訊號分離的目的。計算機模擬證明我們提出的方法優於傳統的 CBOB 演算法則。

關鍵詞：盲目訊號分離，統計盲目波束形成器，最小變量無失眞響應波束形成器，方位估測演算法。

Nonlinear Acm Filters for Narrowband Interference Suppression in Spread Spectrum Systems

Jenq-Tay Yuan, Shih-Hsuan Lo, and Sheng-Lung Tsai*

Department of Electronic Engineering,
Fu Jen Catholic University,
24205 Taipei, Taiwan, R.O.C

Abstract

Our results show that the nonlinear approximate conditional mean (ACM) interpolator always outperforms its linear counterpart for narrowband interference (NBI) suppression for unknown interference statistics, whereas the nonlinear ACM predictor performs even worse than the linear predictor when an interferer is not extremely sharply peaked. Thus, interpolation is more appropriate than prediction when employing the nonlinear adaptive ACM filters for NBI suppression given the fact that the nonlinear ACM filter is much more computationally expensive than its linear counterpart.

Keywords: Spread spectrum communications, Narrowband interference suppression, Nonlinear ACM filters.

I. INTRODUCTION

Vijayan and Poor [1] proposed a nonlinear ACM filter (predictor or interpolator) for NBI suppression. When analyzing the performance of the nonlinear ACM filter, previous references [2]-[4] assume that it completely removes the spread-spectrum (SS) signal from the measurements via the soft decision feedback shown in Fig. 1. Consequently, they actually obtain an *upper bound* on SNR improvements of the nonlinear ACM filters, which reveals that

* Corresponding author. Tel.:+886-2-29053794; Fax: +886-2-29042638
E-mail address : yuan@ee.fju.edu.tw

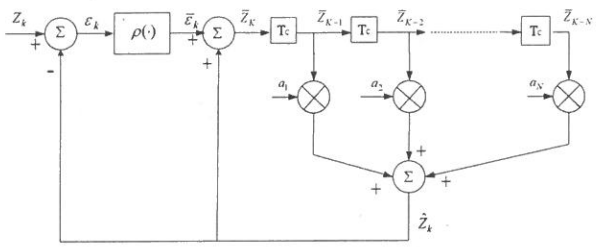


Fig. 1 Nonlinear ACM predictor

the nonlinear ACM filters almost always outperform linear ones. However, when the statistics of the interference are unknown to the receiver which is a practical scenario, our simulation results displayed that the nonlinear interpolator always outperforms its linear counterpart despite the bandwidth of an interferer, whereas the nonlinear predictor does not outperform the linear predictor except when an interferer is extremely sharply peaked. Hence the upper bound on SNR improvement of the nonlinear ACM filters may not reveal their actual performances. In this work, a *lower bound* on SNR improvement for the nonlinear ACM filters when the interference is modeled as an AR(1) process is presented. It is indicated that although the nonlinear ACM predictors are never able to attain the upper bound derived in [3], [4], the lower bound derived herein explains why the linear predictor never significantly outperforms its nonlinear counterpart either even with less peaked interferers. Furthermore, owing to the excellent performances that the nonlinear interpolator achieves consistently, interpolator may be more suitable than predictor when employing the nonlinear adaptive ACM filters for NBI suppression.

II. SYSTEM MODEL

The spread-spectrum and NBI model employed herein is the same as that used in [1], [2], i.e., the received signal at sample k is given by $z_k = s_k + n_k + i_k$, where the processes $\{s_k\}$, $\{n_k\}$, and $\{i_k\}$ are assumed to be mutually independent. The SS signal $\{s_k\}$ is a sequence of independent and identically distributed (i.i.d.) random variables taking on values +1 or –1 with equal probability. The noise $\{n_k\}$ is modeled as white Gaussian with variance σ_n^2 and the interference $\{i_k\}$ is modeled as an AR(1) process given by $i_k = \alpha \cdot i_{k-1} + e_k$, where $e_k \sim N(0, \sigma_e^2)$. Figure 1 presents the nonlinear function which is given by $\rho(\varepsilon_k) = \varepsilon_k - \tanh(\varepsilon_k / \sigma_k^2)$, where

$\varepsilon_k = z_k - \hat{z}_k$ in which $\hat{z}_k = \hat{i}(k \mid k-1) = \sum_{m=1}^{N} a_m \bar{z}_{k-m}$ is the interference estimated from the nonlinear ACM predictor, and σ_k^2 is the variance of ε_k.

III. LOWER BOUND ON SNR IMPROVEMENT

As stated previously, in order to simplify the analysis of the nonlinear ACM predictor, it was generally assumed that the estimate of s_k, $\hat{s}_k = \tanh(\varepsilon_k / \sigma_k^2) \cong s_k$, and the upper bound on SNR improvement is thus obtained. However, this assumption is based on the condition that σ_k^2 is very small (i.e., sharply peaked interference with $\alpha > 0.99$) [3], [4]. Herein, the primary focus is the case when σ_k^2 is not very small and thus the nonlinear predictor fails to produce good estimates [i.e., tanh(.) is in the linear region of operation]. Consequently, $\hat{s}_k \neq s_k$ and $\hat{s}_k$ is no longer i.i.d. From Fig. 1, we thus have

$$\varepsilon_k = z_k - \hat{z}_k = z_k - [\hat{i}(k \mid k-1) - \hat{s}(k \mid k-1)] \tag{1}$$

$$\bar{z}_k = (s_k - \hat{s}_k) + n_k + i_k \tag{2}$$

where $\hat{s}(k \mid k-1)$ is the predictable part of $\hat{s}_k$ and is related to $\hat{s}_k$ by

$$\hat{s}_k = \hat{s}(k \mid k-1) + w_k \tag{3}$$

In (3), w_k is the unpredictable part of $\hat{s}_k$ that is orthogonal to $\hat{s}(k \mid k-1)$. Since tanh(.) is in the linear region of operation, the following occurs

$$\hat{s}_k \cong \frac{\varepsilon_k}{\sigma_k^2} \tag{4}$$

where the variance of the innovation ε_k, σ_k^2, can be expressed as

$$\sigma_k^2 = E\{\varepsilon_k^2\} - 1 = S + P + 2C + \sigma_n^2 \tag{5}$$

in which $S = E\left\{\left[\hat{s}(k \mid k-1)\right]^2\right\}$, $C = E\left\{\hat{s}(k \mid k-1)\left[i_k - \hat{i}(k \mid k-1)\right]\right\}$

$$P = E\{[i_k - \hat{i}(k \mid k-1)]^2\} = \sigma_e^2 \frac{\left[(1-\alpha\beta) + \beta^{2N+1}(\alpha - \beta)\right]}{\left[(1-\alpha\beta)^2 - \beta^{2N}(\alpha - \beta)^2\right]}, \text{ with}$$

$\beta = \frac{1}{2\alpha}\cdot\left[\left(1+\alpha^2\right)+\frac{\sigma_e^2}{\sigma_n^2}\right]-\sqrt{\frac{1}{4\alpha^2}\left[\left(1+\alpha^2\right)+\frac{\sigma_e^2}{\sigma_n^2}\right]^2-1}$ and N is the filter size [3]. Through (3)-(5) it is evident that $S = E\left\{\left[\hat{s}(k\mid k-1)\right]^2\right\} \le E\left\{\hat{s}_k^2\right\} \cong \left[S+P+2C+\sigma_n^2+1\right]/\sigma_k^4 \cong 1/\sigma_k^2$. In addition, by using the Schwarz inequality, we obtain $C^2 \le E\left\{\left[\hat{s}(k\mid k-1)\right]^2\right\}E\left\{\left[i_k-\hat{i}(k\mid k-1)\right]^2\right\} \le P/\sigma_k^2$. When P is not very small (e.g., $\alpha \le 0.99$), it dominates the terms in σ_k^2 in (5). Consequently, $E\left\{\hat{s}_k^2\right\} \cong 1/\sigma_k^2 \cong 1/P$ and $|C| \le 1$ are produced. Equation (5) is then upper-bounded by

$$E\left\{\left(\varepsilon_k - s_k\right)^2\right\} \le \frac{1}{P} + P + \sigma_n^2 + 2 \cong P + \sigma_n^2 + 2 \tag{6}$$

When $0.99 < \alpha < 1$, $\hat{s}_k = \tanh(\varepsilon_k/\sigma_k^2) \cong s_k$ which is i.i.d. Accordingly, $\hat{s}(k\mid k-1) \cong 0$ and (5) becomes

$$E\left\{\left(\varepsilon_k - s_k\right)^2\right\} \cong P + \sigma_n^2 \tag{7}$$

By using (6) and (7), the SNR improvement using the nonlinear predictor is therefore bounded by

$$10\log\frac{\sigma_i^2+\sigma_n^2}{P+\sigma_n^2+2} \le 10\log\frac{E\left(|z_k-s_k|^2\right)}{E\left(|\varepsilon_k-s_k|^2\right)} \le 10\log\frac{\sigma_i^2+\sigma_n^2}{P+\sigma_n^2} \tag{8}$$

where $\sigma_i^2 = \frac{\sigma_e^2}{1-\alpha^2}$ is the power of interference. Similarly, the SNR improvement employing the nonlinear interpolator is bounded by

$$10\log\frac{\sigma_i^2+\sigma_n^2}{I+\sigma_n^2+2} \le 10\log\frac{E\left(|z_k-s_k|^2\right)}{E\left(|\varepsilon_k-s_k|^2\right)} \le 10\log\frac{\sigma_i^2+\sigma_n^2}{I+\sigma_n^2} \tag{9}$$

where $I = E\{[i_k - \hat{i}_{p,f}(k)]^2\}$ in which $\hat{i}_{p,f}(k)$ representing the interference estimated by employing the nonlinear interpolator using the p past and f future neighboring samples of $\bar{z}_k$ ($N = p+f$ is assumed implicitly) [4]-[5]. Notably, I is usually much smaller than P as interpolation more effectively utilizes the correlation between the nearest neighboring samples than prediction does, nonlinear interpolator yields much more accurate estimates than its pre-

diction counterpart through $\hat{S}_k = \tanh(\varepsilon_k / \sigma_k^2)$. Consequently, the SNR improvement through the nonlinear interpolator closely approaches its upper bound.

IV. COMPUTER SIMULATIONS

To verify (8) and (9), computer simulations were performed by using the adaptive linear and nonlinear QRD-LSL predictors (interpolators) of order $N = 10$ $[(p, f) = (5,5)]$ (with the forgetting factor $\lambda = 0.998$) that was developed in [5], [6]. The jammer was an AR(1) process with $0.8 \le \alpha \le 0.999$ and σ_e^2 was varied with a chosen α to ensure that the interference power was kept at 100. The noise power was held constant at $\sigma_n^2 = 0.01$. The power of the SS signal was also held constant with amplitudes ± 1. The linear and nonlinear predictors (interpolators) were run for 5000 samples. Figure 2 illustrates the simulation results of SNR improvements for predictors and interpolators. Notably, these were computed using the last 500 samples over 200 independent trials. This figure demonstrates that the nonlinear predictor outperforms the linear predictor only when $\alpha \ge 0.995$ corresponds to an extremely sharply peaked interference. It appears that this result is caused by the increase in the Gaussianity of the received signal as α is decreased and also by the fact that the ACM filter is an approximate filter. In contrast, the nonlinear interpolator always outperforms the linear one even when α is very small. However, the nonlinear predictor's performance cannot be much worse than that of the linear one even when $\alpha << 0.995$ occurs due to the bounds that (8) produces

V. CONCLUSIONS

In this work, a *lower bound* on SNR improvement for the nonlinear ACM filters when the interference is modeled as an AR(1) process is presented. It is indicated that although the nonlinear ACM predictors are never able to attain the upper bound, the lower bound derived herein explains why the linear predictor never significantly outperforms its nonlinear counterpart either even with less peaked interferers. Furthermore, owing to the excellent performances that the nonlinear interpolator achieves consistently, interpolator may be more suitable than predictor when employing the nonlinear adaptive ACM filters for NBI suppression.

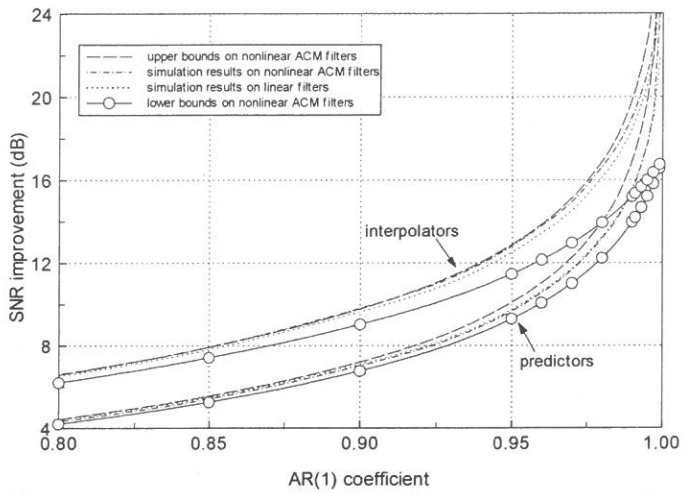


Fig. 2 **SNR improvements (dB) for predictors and interpolators ($\sigma_n^2 = 0.01$)**

REFERENCES

(1) R. Vijayan and H. V. Poor, "Nonlinear techniques for interference suppression in spread-spectrum systems," *IEEE Trans. Commun.*, vol. COM-38, pp. 1060-1065, July 1990.

(2) L. A. Rusch and H. V. Poor, "Narrowband interference suppression in CDMA spread spectrum communications," *IEEE Trans. Commun.*, vol. COM-42, pp. 1969-1979, February/March/April 1994.

(3) K. J. Wang, Z. C. Zhou, and Y. Yao, "Closed-form analytical results for interference rejection of nonlinear prediction filters in DS-SS systems," *Electron Lett.*, July 1997, 33, (16), pp. 1354-1355.

(4) K. J. Wang, Z. C. Zhou, and Y. Yao, "Performance analysis of nonlinear interpolation filters in DS spread spectrum systems under narrowband interference condition," *Electron Lett.*, July 1998, 34, (15), pp. 1464-1465.

(5) J. –T. Yuan, "QR-Decomposition-Based Least-Squares Lattice Interpolators," *IEEE Trans. on Signal Processing*, Vol 48, No.1, January 2000, pp. 70 - 79.

(6) J. –T. Yuan and J. –N. Lee, "Narrowband Interference Rejection in DS/CDMA Systems Using Adaptive (QRD-LSL)-Based Nonlinear ACM Interpolators," *IEEE Transactions on Vehicular Technology,* Vol. 52, No. 2, March 2003, pp. 374-379.

展頻通訊系統中非線性 ACM 適應性濾波器之窄頻干擾抑制

袁正泰　羅仕軒　蔡昇龍
輔仁大學電子工程系

摘　要

這篇論文結果顯示當使用運算量相對較大的非線性 ACM 適應性濾波器在展頻通訊系統窄頻干擾抑制時，使用雙向預估(interpolation)比單向預估(prediction)更適合。這是因爲不論干擾源頻寬如何變化，非線性 ACM 雙向預估濾波器的表現永遠比線性雙向預估濾波器佳。而相對的，當干擾源頻寬不是非常窄時，非線性 ACM 單向預估濾波器的表現甚至有可能比運算量相對較低的線性單向預估濾波器還差。

關鍵詞：展頻通訊，窄頻干擾抑制，非線性 ACM 濾波器

On Cumulative Hazard Function Estimation for Series System from Weibull Family

Sy-Mien Chen* and Hou-Yan Chu

Department of Mathematics

Fu-Jen Catholic University, Taipei, Taiwan, R.O.C.

Abstract

In this article, we consider the cumulative hazard function of a series system product which is composed by two independent components . We consider the case when the two independent components of the series are from Weibull distributions. We propose a direct estimator and an indirect estimator of the cumulative hazard function of the system. It is shown that the indirect estimator is better than the direct estimator in the sense of mean square error.

Keywords: Cumulative Hazard Function; Limiting Variance; Maximum Likelihood Estimator.

1. INTRODUCTION

Hazard function is an important measure of quality in reliability studies. Fail to identify high hazard function products or components could lead to a great loss. Diverse existing shapes of hazard function well describe the reliabilities of all kinds of products and provide solid grounds for decision making. See Lawless (1982), Lee and Max (1991), Meeker and Escobar (1998) and references therein for more details about hazard function.

It is interesting to note that the hazard function of an assembly product can be estimated by observations from the product directly or from the components indirectly. A practical question is which method is better, and that will mainly our goal for this research. We con-

* Corresponding author. Tel.:+886-2-29052451; Fax: +886-2-29044509
E-mail: math1013@mails.fju.edu.tw

sider a series system product which is composed by two independent components from Weibull families. In section two, we did a brief review of such family. Then we propose a direct and an indirect estimator of the cumulative hazard function and study their asymptotic performances. In section 3, some numerical study is demonstrated to give idea about both direct and indirect estimators. In section 4 we give some conclusions and suggestions.

2.WEIBULL FAMILY

The Weibull distribution is commonly used for analyzing lifetime data. It accommodates increasing and decreasing failure rates. The probability density function $f(y)$ and the hazard function $h(t)$ of Weibull family $W(\sigma,\alpha)$ are given by

$$\mathrm{f(y)} = \frac{\alpha}{\sigma}\left(\frac{y}{\sigma}\right)^{\alpha-1} \exp(-(y/\sigma)^{\alpha}) \quad \text{and} \quad h(t) = \frac{\alpha}{\sigma}\left(\frac{t}{\sigma}\right)^{\alpha-1},$$

respectively, for $0 < y < \infty$, $\alpha > 0$, $\sigma > 0$, $t > 0$. See Fig I-(a) (b) (c).The cumulative hazard function is defined by $H(t) = \int_0^t h(u)du$.

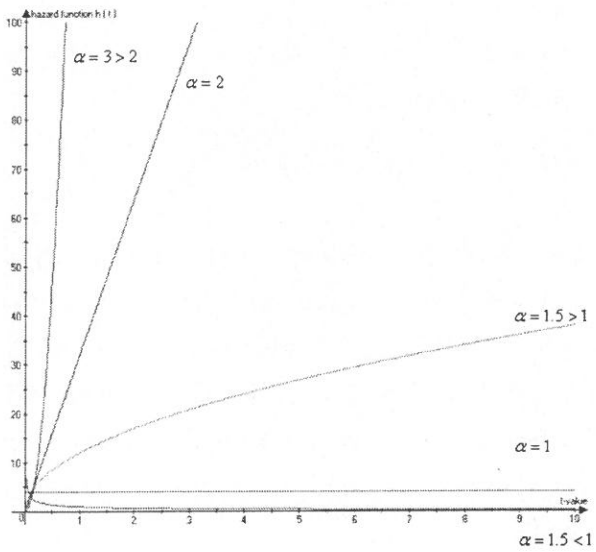


Fig I-(a). hazard function of Weibull distribution
$\sigma = 0.25$; $t \in [0,10]$; $h_T(t) \in [0,100]$

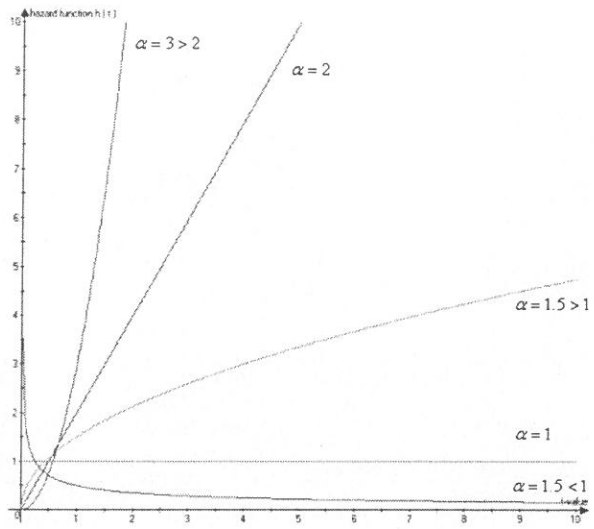


Fig I-(b). hazard function of Weibull distribution

$\sigma = 1$, $t \in [0,10]$; $h_T(t) \in [0,10]$

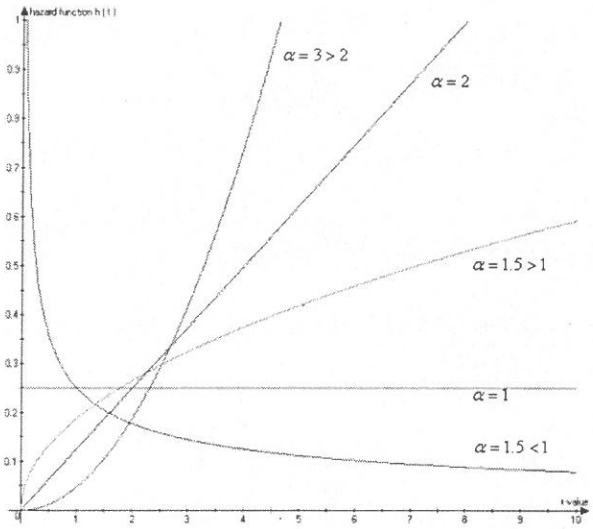


Fig I-(c). hazard function of Weibull distribution

$\sigma = 4$; $t \in [0,10]$; $h_T(t) \in [0,1]$

Consider a series assembly product which is composed by two independent components. Let *Z, X, Y* denote the life times of the series assembly product and the two components re-

spectively, then $Z=\min\{X,Y\}$. Assume that both X and Y are from the Weibull family, namely $X\sim$ W (α_1,σ_1), $Y\sim$ W (α_2,σ_2). Then the distribution functions and the cumulative hazard functions of X, Y, Z are defined as

$$F_X(t)=1-\exp[-(t/\sigma_1)^{\alpha_1}],\ F_Y(t)=1-\exp[-(t/\sigma_2)^{\alpha_2}],$$

$$F_Z(t)=1-\exp\{-[(t/\sigma_1)^{\alpha_1}+(t/\sigma_2)^{\alpha_2}]\},$$

$$H_X(t)=-\ln(1-F_X(t)),\ H_Y(t)=-\ln(1-F_Y(t)),\quad H_Z(t)=-\ln(1-F_Z(t)),$$

$H_Z(t)=H_X(t)+H_Y(t)$, for $0<t<\infty$, $\sigma_1>0,\sigma_2>0,\alpha_1>0,\alpha_2>0$. For the case when $\alpha=1$, i.e. the negative exponential distribution, see Chen and Hsu (2002) .

Since the maximum likelihood estimator is typically asymptotically efficient, another estimator cannot hope to beat its asymptotic variance. Hence in this research, we will concentrate on maximum likelihood estimators. Let $\Theta_1=(\sigma_1,\alpha_1)$, $\Theta_2=(\sigma_2,\alpha_2)$ denote the parametric spaces of X and Y, and $\hat{\Theta}_1=(\hat{\sigma}_1,\hat{\alpha}_1)$, $\hat{\Theta}_2=(\hat{\sigma}_2,\hat{\alpha}_2)$ be the maximum likelihood estimators of Θ_1 and Θ_2, respectively. Then

$$\sqrt{n}(\hat{\Theta}_1-\Theta_1)\xrightarrow[n\to\infty]{} N((0,0),I^{-1}(\Theta_1)),$$

$$\sqrt{n}(\hat{\Theta}_2-\Theta_2)\xrightarrow[n\to\infty]{} N((0,0),I^{-1}(\Theta_2)),$$

where $I^{-1}(\Theta_1)$, $I^{-1}(\Theta_2)$, are the inverse of the Fisher information matrices $I(\Theta_1)$, $I(\Theta_2)$ of X and Y, respectively. Let $I(\Theta_1)=\begin{pmatrix} I_{11}^X & I_{12}^X \\ I_{21}^X & I_{22}^X \end{pmatrix}$, $I(\Theta_2)=\begin{pmatrix} I_{11}^Y & I_{12}^Y \\ I_{21}^Y & I_{22}^Y \end{pmatrix}$ then

$$I_{11}^X=\left(\frac{\alpha_1}{\sigma_1}\right)^2,\qquad I_{12}^X=\frac{\alpha_1 Ln(\sigma_1)}{\sigma_1}-\frac{\alpha_1 E\left(t^{\alpha_1}Ln(t)\right)}{\sigma_1^{\alpha_1+1}}=I_{21}^X,$$

$$I_{22}^X=\frac{1}{\alpha_1^2}+\left(Ln(\sigma_1)\right)^2+\frac{E\left(t^{\alpha_1}\left(Ln(t)\right)^2\right)-2Ln(\sigma_1)E\left(t^{\alpha_1}Ln(t)\right)}{\sigma_1^{\alpha_1}},$$

and

$$I_{11}^{Y}=\left(\frac{\alpha_2}{\sigma_2}\right)^2,\ I_{12}^{Y}=\frac{\alpha_2 Ln(\sigma_2)}{\sigma_2}-\frac{\alpha_2 E\left(t^{\alpha_2}Ln(t)\right)}{\sigma_2^{\alpha_2+1}}=I_{21}^{Y},$$

$$I_{22}^{Y}=\frac{1}{\alpha_2^{\ 2}}+\left(Ln(\sigma_2)\right)^2+\frac{E\left(t^{\alpha_2}\left(Ln(t)\right)^2\right)-2Ln(\sigma_2)E\left(t^{\alpha_2}Ln(t)\right)}{\sigma_2^{\alpha_2}}.$$

Let $g_1(\Theta_1)=H_X(t)$, $g_2(\Theta_2)=H_Y(t)$, then by delta method, one has

$$\sqrt{n}\left(\hat{H}_X(t)-H_X(t)\right)=\sqrt{n}(g_1(\hat{\Theta}_1)-g_1(\Theta_1))\xrightarrow[n\to\infty]{}N((0,0),\Phi_1^{'}I^{-1}(\Theta_1)\Phi_1)$$

$$\sqrt{n}\left(\hat{H}_Y(t)-H_Y(t)\right)=\sqrt{n}(g_2(\hat{\Theta}_2)-g_2(\Theta_2))\xrightarrow[n\to\infty]{}N((0,0),\Phi_2^{'}I^{-1}(\Theta_2)\Phi_2)$$

where Φ_i is the differential of g_i, i=1, 2, i.e.

$$\Phi_1=\left(\frac{\partial H_X(t)}{\partial\sigma_1}\ \frac{\partial H_X(t)}{\partial\alpha_1}\right)=\left(-\frac{\alpha_1}{\sigma_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\ln\left(\frac{t}{\sigma_1}\right)\right)$$

$$\Phi_2=\left(\frac{\partial H_Y(t)}{\partial\sigma_2}\ \frac{\partial H_Y(t)}{\partial\alpha_2}\right)=\left(-\frac{\alpha_2}{\sigma_2}\left(\frac{t}{\sigma_2}\right)^{\alpha_2}\left(\frac{t}{\sigma_2}\right)^{\alpha_2}\ln\left(\frac{t}{\sigma_2}\right)\right).$$

Since $g(\Theta)=H_Z(t)=H_X(t)+H_Y(t)=g_1(\Theta_1)+g_2(\Theta_2)$, it implies that the indirect estimator $\hat{H}_Z^I$ of $H_Z(t)$ is $\hat{H}_Z^I(t)=\hat{H}_X(t)+\hat{H}_Y(t)$, and

$$\sqrt{n}(\hat{H}_Z^I(t))-H(t))\xrightarrow[n\to\infty]{}N((0,0),\Phi_1^{'}I^{-1}(\Theta_1)\Phi_1+\Phi_2^{'}I^{-1}(\Theta_2)\Phi_2).\qquad(1)$$

Let $\Theta=(\sigma_1,\alpha_1,\sigma_2,\alpha_2)$ and $\hat{\Theta}$ be the maximum likelihood estimator of Θ, then $\sqrt{n}(\hat{\Theta}-\Theta)\xrightarrow[n\to\infty]{}N((0,0),I^{-1}(\Theta))$. Let $\hat{H}_Z^D(t)$ be the direct estimator of $H(t)$, then by delta method, we have

$$\sqrt{n}(\hat{H}_Z^D(t)-H(t))\xrightarrow[n\to\infty]{}N((0,0),\Phi^{'}I^{-1}(\Theta)\Phi),\qquad(2)$$

where Φ is the differential of $g(\Theta)=H_Z(t)$, i.e.

$$\Phi=\left(\frac{\partial H_Z(t)}{\partial\sigma_1}\ \frac{\partial H_Z(t)}{\partial\alpha_1}\ \frac{\partial H_Z(t)}{\partial\sigma_2}\ \frac{\partial H_Z(t)}{\partial\alpha_2}\right)$$

$$= \left(-\frac{\alpha_1}{\sigma_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\left(\frac{t}{\sigma_1}\right)^{\alpha_1}\ln\left(\frac{t}{\sigma_1}\right) - \frac{\alpha_2}{\sigma_2}\left(\frac{t}{\sigma_2}\right)^{\alpha_2}\left(\frac{t}{\sigma_2}\right)^{\alpha_2}\ln\left(\frac{t}{\sigma_2}\right)\right),$$

and the Fisher information $I(\Theta)$ is given by $I(\Theta) = \begin{pmatrix} I_{11}^Z & I_{12}^Z & I_{13}^Z & I_{14}^Z \\ I_{21}^Z & I_{22}^Z & I_{23}^Z & I_{24}^Z \\ I_{31}^Z & I_{32}^Z & I_{33}^Z & I_{34}^Z \\ I_{41}^Z & I_{42}^Z & I_{43}^Z & I_{44}^Z \end{pmatrix}$, where

$$I_{11}^Z = -\alpha_1(\alpha_1 - 1)\alpha_2\sigma_1^{\alpha_1-2}\varepsilon + \alpha_1^2\alpha_2^2\sigma_1^{2\alpha_1-2}\eta - \frac{\alpha_1}{\sigma_1^2} + \frac{\alpha_1(\alpha_1+1)}{\sigma_1^{\alpha_1+2}}\beta$$

$$I_{12}^Z = -\alpha_2\sigma_1^{\alpha_1-1}\left(1 + \alpha_1 Ln(\sigma_1)\right)\varepsilon + \alpha_1\alpha_2^2\sigma_1^{2\alpha_1-1}Ln(\sigma_1)\eta + \frac{1}{\sigma_1} - \frac{\left(1-\alpha_1 Ln(\sigma_1)\right)}{\sigma_1^{\alpha_1+1}}\beta$$

$$-\frac{\alpha_1}{\sigma_1^{\alpha_1+1}}\delta + \alpha_1\alpha_2\sigma_1^{\alpha_1-1}\sigma_2^{\alpha_2}\phi + \alpha_1^2\alpha_2\sigma_1^{\alpha_1-1}\sigma_2^{\alpha_1}\lambda = I_{21}^Z$$

$$I_{13}^Z = \alpha_1^2\alpha_2^2\sigma_1^{\alpha_1-1}\sigma_2^{\alpha_2-1}\phi = I_{31}^Z$$

$$I_{14}^Z = -\alpha_1\sigma_1^{\alpha_1-1}\varepsilon - \alpha_1\alpha_2\sigma_1^{\alpha_1-1}\iota + \alpha_1^2\alpha_2\sigma_1^{\alpha_1-1}\sigma_2^{\alpha_2}Ln(\sigma_2)\phi + \alpha_1\alpha_2\sigma_1^{2\alpha_1-1}\eta + \alpha_1\alpha_2^2\sigma_1^{2\alpha_1-1}\psi = I_{41}^Z$$

$$I_{22}^Z = -2\sigma_2^{\alpha_2}\varsigma - \alpha_1\sigma_2^{\alpha_2}\xi - \alpha_2\sigma_1^{\alpha_1}\left(Ln(\sigma_1)\right)^2\varepsilon + \sigma_2^{2\alpha_2}\kappa + 2\alpha_1\sigma_2^{2\alpha_2}\vartheta + \alpha_1^2\sigma_2^{2\alpha_2}\tau$$
$$+ \alpha_2^2\sigma_1^{2\alpha_1}\left(Ln(\sigma_1)\right)^2\eta + \frac{\left(Ln(\sigma_1)\right)^2}{\sigma_1^{\alpha_1}}\beta - \frac{2Ln(\sigma_1)}{\sigma_1^{\alpha_1}}\delta + \frac{1}{\sigma_1^{\alpha_1}}\gamma$$
$$+ 2\alpha_2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}Ln(\sigma_1)\phi + 2\alpha_1\alpha_2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}Ln(\sigma_1)\lambda$$

$$I_{23}^Z = -\alpha_2\sigma_2^{\alpha_2-1}\rho - \alpha_1\alpha_2\sigma_2^{\alpha_2-1}\varsigma + \alpha_1\alpha_2\sigma_2^{2\alpha_2-1}\kappa$$
$$+ \alpha_1^2\alpha_2\sigma_2^{2\alpha_2-1}\vartheta + \alpha_1\alpha_2^2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2-1}Ln(\sigma_1)\phi = I_{32}^Z$$

$$I_{24}^Z = -\sigma_2^{\alpha_2}Ln(\sigma_2)\rho - \alpha_1\sigma_2^{\alpha_2}Ln(\sigma_2)\varsigma - \sigma_1^{\alpha_1}Ln(\sigma_1)\varepsilon$$
$$- \alpha_2\sigma_1^{\alpha_1}Ln(\sigma_1)\iota + \alpha_1\sigma_2^{2\alpha_2}Ln(\sigma_2)\kappa + \alpha_1^2\sigma_2^{2\alpha_2}Ln(\sigma_2)\vartheta$$
$$+ \sigma_1^{\alpha_1}\sigma_2^{\alpha_2}\left(1 + \alpha_1\alpha_2 Ln(\sigma_1)Ln(\sigma_2)\right)\phi + \sigma_1^{\alpha_1}\sigma_2^{\alpha_2}\left(\alpha_1 + \alpha_2\right)\lambda$$
$$+ \alpha_1\alpha_2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}\omega + \alpha_2\sigma_1^{2\alpha_1}Ln(\sigma_1)\eta + \alpha_2^2\sigma_1^{2\alpha_1}Ln(\sigma_1)\psi = I_{42}^Z$$

$$I_{33}^{Z} = -\alpha_1\alpha_2(\alpha_2-1)\sigma_2^{\alpha_2-2}\rho + \alpha_1^2\alpha_2^2\sigma_2^{2\alpha_2-2}\kappa - \frac{\alpha_2}{\sigma_2^2} + \frac{\alpha_2(\alpha_2+1)}{\sigma_2^{\alpha_2+2}}\zeta$$

$$I_{34}^{Z} = -\alpha_1\sigma_2^{\alpha_2-1}\left(1+\alpha_2 Ln(\sigma_2)\right)\rho + \alpha_1^2\alpha_2\sigma_2^{2\alpha_2-1}Ln(\sigma_2)\kappa + \frac{\left(\alpha_2 Ln(\sigma_2)-1\right)}{\sigma_2^{\alpha_2+1}}\zeta$$
$$+ \alpha_1\alpha_2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2-1}\phi + \alpha_1\alpha_2^2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2-1}\lambda + \frac{1}{\sigma_2} - \frac{\alpha_2}{\sigma_2^{\alpha_2+1}}\theta = I_{43}^{Z}$$

$$I_{44}^{Z} = -\alpha_1\sigma_2^{\alpha_2}\left(Ln(\sigma_2)\right)^2\rho - 2\sigma_1^{\alpha_1}\iota - \alpha_2\sigma_1^{\alpha_1}\upsilon$$
$$+ 2\alpha_1\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}Ln(\sigma_2)\phi + 2\alpha_1\alpha_2\sigma_1^{\alpha_1}\sigma_2^{\alpha_2}Ln(\sigma_2)\lambda + \sigma_1^{2\alpha_1}\eta + 2\alpha_2\sigma_1^{2\alpha_1}\psi + \alpha_2^2\sigma_1^{2\alpha_1}\mu$$
$$+ \alpha_1^2\sigma_2^{2\alpha_2}\left(Ln(\sigma_2)\right)^2\kappa + \frac{\left(Ln(\sigma_2)\right)^2}{\sigma_2^{\alpha_2}}\zeta - \frac{2Ln(\sigma_2)}{\sigma_2^{\alpha_2}}\theta + \frac{1}{\sigma_2^{\alpha_2}}\chi,$$

and

$$\beta = \mathrm{E}\left(t^{\alpha_1}\right);\ \delta = \mathrm{E}\left(t^{\alpha_1}Ln(t)\right);\ \gamma = \mathrm{E}\left(t^{\alpha_1}(Ln(t))^2\right);\ \zeta = \mathrm{E}\left(t^{\alpha_2}\right);\ \theta = \mathrm{E}\left(t^{\alpha_2}Ln(t)\right);$$

$$\chi = \mathrm{E}\left(t^{\alpha_2}(Ln(t))^2\right);\ \rho = E\left(\frac{t^{\alpha_1-1}}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_{X1}^{\alpha_1}t^{\alpha_2-1}}\right);\ \varsigma = E\left(\frac{t^{\alpha_1-1}Ln(t)}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}}\right);$$

$$\xi = E\left(\frac{t^{\alpha_1-1}\left(Ln(t)\right)^2}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}}\right); \qquad \varepsilon = E\left(\frac{t^{\alpha_2-1}}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}}\right);$$

$$\iota = E\left(\frac{t^{\alpha_2-1}Ln(t)}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}}\right); \qquad \upsilon = E\left(\frac{t^{\alpha_2-1}\left(Ln(t)\right)^2}{\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}}\right);$$

$$\kappa = \mathrm{E}\left(\frac{t^{2\alpha_1-2}}{(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1})^2}\right); \qquad \vartheta = \mathrm{E}\left(\frac{t^{2\alpha_1-2}Ln(t)}{(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1})^2}\right);$$

$$\tau = \mathrm{E}\left(\frac{t^{2\alpha_1-2}(Ln(t))^2}{(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1})^2}\right); \qquad \eta = \mathrm{E}\left(\frac{t^{2\alpha_2-2}}{(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1} + \alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1})^2}\right);$$

$$\psi = E\left(\frac{t^{2\alpha_2-2}Ln(t)}{\left(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1}+\alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}\right)^2}\right); \qquad \mu = \mathrm{E}\left(\frac{t^{2\alpha_2-2}(Ln(t))^2}{(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1}+\alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1})^2}\right);$$

$$\phi = E\left(\frac{t^{\alpha_1+\alpha_2-2}}{\left(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1}+\alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}\right)^2}\right); \qquad \lambda = E\left(\frac{t^{\alpha_1+\alpha_2-2}Ln(t)}{\left(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1}+\alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}\right)^2}\right);$$

$$\omega = E\left(\frac{t^{\alpha_1+\alpha_2-2}\left(Ln(t)\right)^2}{\left(\alpha_1\sigma_2^{\alpha_2}t^{\alpha_1-1}+\alpha_2\sigma_1^{\alpha_1}t^{\alpha_2-1}\right)^2}\right).$$

3. NUMERICAL ANALYSIS

Theoretically speaking, to compare the direct estimator with the indirect estimator of the cumulative hazard function of the series system, we can compare their asymptotic mean square error. The one with the smaller asymptotic mean square error performs better than the one with the larger asymptotic mean square error. Notice that $\Phi_1' I^{-1}(\Theta_1)\Phi_1 + \Phi_2' I^{-1}(\Theta_2)\Phi_2$ $= \Phi' J^{-1}(\Theta)\Phi$, where $J^{-1}(\Theta) = \begin{pmatrix} I^{-1}(\Theta_1) & 0 \\ 0 & I^{-1}(\Theta_2) \end{pmatrix}$. One way to compare is to study the difference between $\Phi_1' I^{-1}(\Theta_1)\Phi_1 + \Phi_2' I^{-1}(\Theta_2)\Phi_2$ and $\Phi' I^{-1}(\Theta)\Phi$, which is $\Phi'(J^{-1}(\Theta) - I^{-1}(\Theta))\Phi$. If we can show that $J^{-1}(\Theta) - I^{-1}(\Theta)$ is non-negative or non-positive definite, then we are done. On the other hand, we can study the asymptotic relative efficiency (ARE) of $\hat{H}_Z^D(t)$ with respect to $\hat{H}_Z^I(t)$, where

$$\mathrm{ARE}\left(\hat{H}_Z^D(t), \hat{H}_Z^I(t)\right) = \frac{\Phi_1' I^{-1}(\Theta_1)\Phi_1 + \Phi_2' I^{-1}(\Theta_2)\Phi_2}{\Phi' I^{-1}(\Theta)\Phi} = \frac{\Phi'(J^{-1}(\Theta) - I^{-1}(\Theta))\Phi}{\Phi' I^{-1}\Phi}.$$

In this paper we concentrate on the asymptotic relative efficiency (ARE) of $\hat{H}_Z^D(t)$ with respect to $\hat{H}_Z^I(t)$. By checking the formulas, we found that it is almost impossible to compare the ARE analytically because of their complexity. Fortunately, the integral can be carried out numerically. And that is exactly what we are going to do in the rest of this research. We use Visual Fortran , Mathematica and Excel for the computations, and use Golden Sofeware Grapher for graphing.

For simplicity, we concern only the case when $\alpha_1 \geq \alpha_2 > 0$. First, we consider the case when the two shape parameters α_1 and α_2 are the same. As we can see from Fig II (a)-(e), for

all curves, ARE are all greater than .67 at the beginning, increasing first but then drop soon especially when the difference between σ_1 and σ_2 are small, then increase again, finally become stable after some t. ARE increases when $\delta = \sigma_1 - \sigma_2 > 0$ increases, which means that for two components with the same shape parameters, the ARE is increases as long as the difference between their scale parameters is getting larger. When the value of the common shape parameter increases, the ARE increases. When the components are from exactly the same distributions, the corresponding ARE are quite low.

When $\alpha_1 = \alpha_2 = 0.5$, the hazard functions $h_X(t)$, $h_Y(t)$, $h_Z(t)$ for X , Y and Z are all decreasing functions. As we can see from Fig II (a), all the stable values of the ARE tend to be small. When the difference between the two scale parameters δ is increasing, the stable values are increasing too. This means the more these two scale parameters separate, the larger the relative efficiency of the direct estimator with respect to the indirect estimator is.

When $\alpha_1 = \alpha_2 = 1$, the hazard functions $h_X(t)$, $h_Y(t)$, $h_Z(t)$ for X , Y and Z are all constants. As we can see from Fig II (b), all the stable values of the ARE in this case are relatively greater than those when $\alpha_1 = \alpha_2 = 0.5$.

When $\alpha_1 = \alpha_2 = 1.5$, the hazard functions $h_X(t)$, $h_Y(t)$, $h_Z(t)$ for X , Y and Z are all increasing functions. As we can see from Fig II (c), when the value of increases to a certain value, e.g. 15, the value of ARE is almost 1 from the beginning, which is very interesting. And again, all the stable values of the ARE in this case are relatively greater than those in the previous two cases.

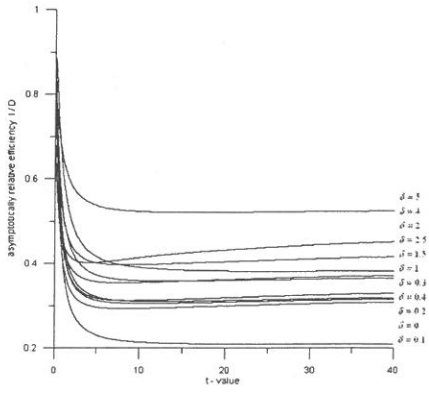


Fig II-(a). ARE when $\alpha_1 = \alpha_2 = 0.5$, $\sigma_2 = 0.5$, $\delta = \sigma_1 - \sigma_2$.

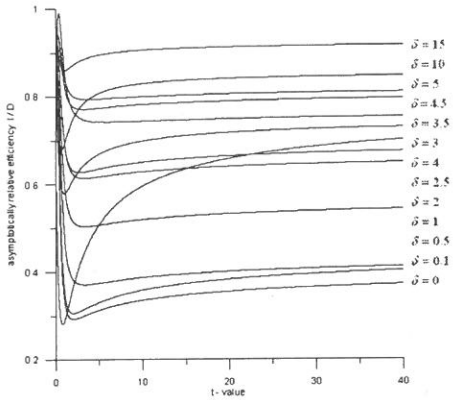


Fig II-(b) : ARE when $\alpha_1 = \alpha_2 = 1$, σ_2 =0.5 , $\delta = \sigma_1 - \sigma_2$.

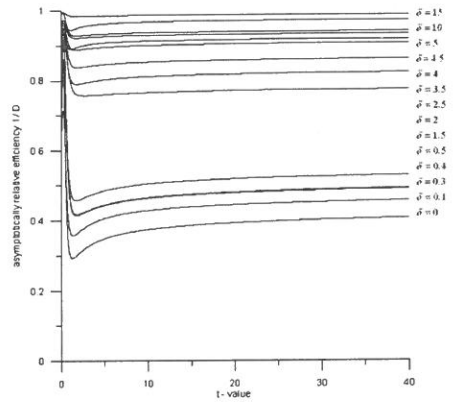


Fig II-(c) : ARE when $\alpha_1 = \alpha_2 = 1.5$, σ_2 =0.5 , $\delta = \sigma_1 - \sigma_2$.

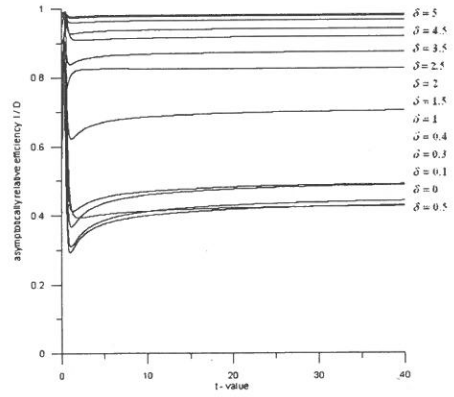


Fig II-(d) : ARE when $\alpha_1 = \alpha_2 = 2$, σ_2 =0.5 , $\delta = \sigma_1 - \sigma_2$.

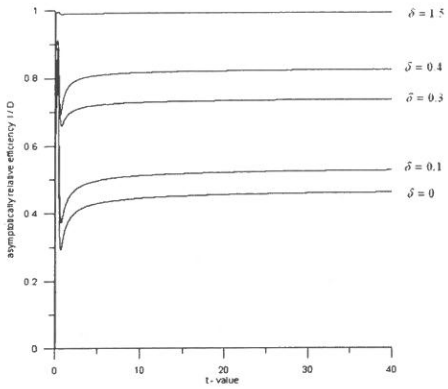


Fig II-(e) : ARE when $\alpha_1 = \alpha_2 = 4$, σ_2=0.5, $\delta = \sigma_1 - \sigma_2$.

When $\alpha_1 = \alpha_2$=4, the hazard functions $h_X(t)$, $h_Y(t)$, $h_Z(t)$ for X , Y and Z are all increasing functions. As we can see from Fig II (e), even when the value of is as small as 1.5, the value of ARE is almost 1 from the beginning, which means that the direct estimator is about the same good as the indirect estimator, i.e. when the shape parameters of the two components are large enough, the difference between the two parameters are no longer important for evaluating the efficiency of the direct and the indirect estimator.

In the case when $\sigma_1 = \sigma_2$, ARE increases as $\rho = \alpha_1 - \alpha_2$ is increasing, which is similar to the pattern in previous cases when $\alpha_1 = \alpha_2$. After the ARE reach stability, we can see that for $\alpha_1 > \alpha_2$,, the values of ARE increase as $\rho = \alpha_1 - \alpha_2$ increases, but slowly. This means that when the two scale parameters are the same, unless the differences between the two shape parameters are extremely large, the indirect estimator is superior than the direct estimator. For small value of t, the value of the common scale parameter affect the trend of ARE. As we can see, when $\sigma_1 = \sigma_2 =$ 0.5, the curves of ARE are quite smooth. But as the value increases, curves are not as smooth as before. Fortunately, they all reach stability latter.

It is interesting to see from Fig II and Fig III that when and , the values of ARE are all between 0.3 and 0.4. Which indicates that the indirect estimator is superior than the direct estimator.

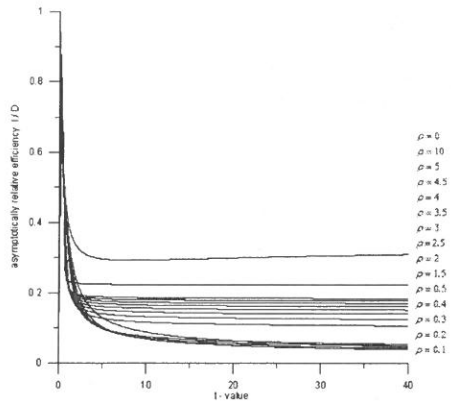


Fig III-(a) : ARE when $\sigma_1 = \sigma_2 = 0.5$, $\alpha_2 = 0.5$, $\rho = \alpha_1 - \alpha_2$.

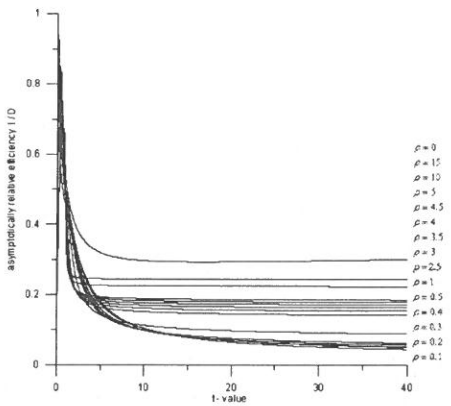


Fig III-(b) : ARE when $\sigma_1 = \sigma_2 = 1$, $\alpha_2 = 0.5$, $\rho = \alpha_1 - \alpha_2$.

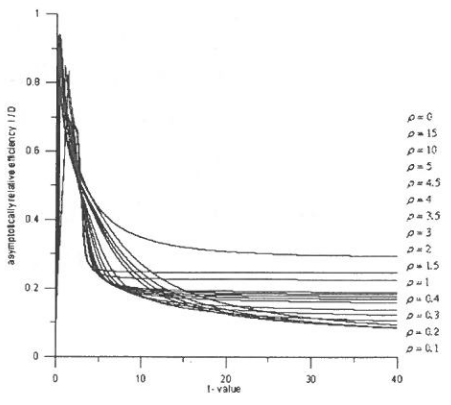


Fig III-(c) : ARE when $\sigma_1 = \sigma_2 = 3$, $\alpha_2 = 0.5$, $\rho = \alpha_1 - \alpha_2$.

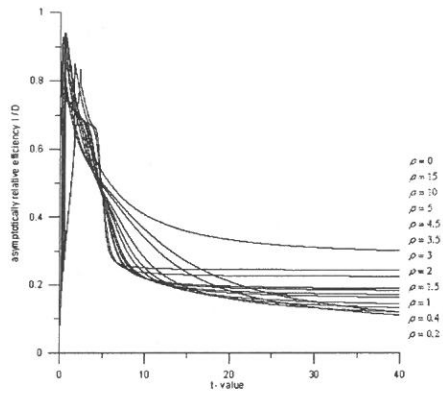


Fig III-(d) : ARE when $\sigma_1 = \sigma_2 = 5$, $\alpha_2 = 0.5$, $\rho = \alpha_1 - \alpha_2$.

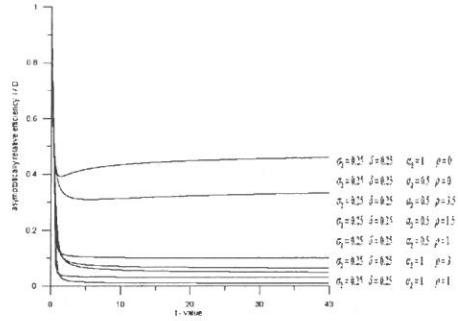


Fig IV : ARE when $\delta = \sigma_1 - \sigma_2$, $\rho = \alpha_1 - \alpha_2$.

Fig IV demonstrate the case when both δ and ρ are nonzero. From graph we see that the pattern in this case still the same. i.e. The indirect estimator is recommended.

4. CONCLUSION AND SUGGESTIONS

In this article, we consider the cumulative hazard function of a series system product which is composed by two independent components. We consider the case when the two independent components of the series are from Weibull distributions. We propose a direct estimator and an indirect estimator of the cumulative hazard function of the system.

Intuitively, indirect estimator is more informative. In this research, first we study the asymptotic performance of both estimators. Then conduct a numerical analysis. From the numerical study, we see that when the two shape parameters are the same but the two scale parameters are quite different, either indirect or direct estimators perform similarly. But in all

other cases, the indirect estimator is superior than the direct estimator. Hence, indirect estimator is still recommended.

ACKNOWLEDGE

This research is partially supported by the Division of academic research office of research and development at Catholic Fu-Jen University under research project 3923104484.

REFERENCES

(1) Chen S. M and Hsu Y. S. (2002), "Quality Estimations for Assembly Products." Technical report (339-1-2001-1-227) of the division of academic research office of research and development at Fu-Jen Catholic University.

(2) Efron, B.F., and Hinkley, D.V. (1978), "Assessing the Accuracy of the Maximum Likelihood Estimator Observed Versus Expected Fisher Information." Biometrika, 65, 457-487.

(3) Lawless, J. F. (1982), Statistical Models and Methods for Lifetime Data. Wiley.

(4) Lee, J.B. and Max, E (1991), "Statistical Analysis of Reliability and Life Testing Models." 2nd Ed. Marcel Dekker, New York.

(5) Lehmann , E. L. and Casella, G. (1998), "Theory of Point Estimation." 2nd Edition

(6) Meeker, W. Q.and Escobar, L. A.(1998), "Statistical Methods for Reliability Data." Wiley.

received October 2, 2003
revised November 6, 2003
accepted November20, 2003

韋伯分佈之累積失效函數之估計

陳思勉 朱厚恩
輔仁大學數學系

摘 要

本研究主要探討對於由兩個來自偉伯分佈之獨立零件所構成之串聯系統, 其累積故障函數之直接與間接估計式之比較.

關鍵詞：累積失效函數；近似變異；最大概似估計式

Realization of Current-Mode Highpass Lowpass and Bandpass Biquad Filters using Single CFCCII$_p$

Yung-Chang Yin Yueh-Chergn Liou*
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.

Abstract

A new circuit configuration for active current filter using single four-terminal p-type active current conveyors (CFCCII$_p$) is presented. The proposed current-mode single-CFCCII$_p$ filter would not require critical matching conditions. The circuit configuration can be used to synthesize lowpass, bandpass and highpass functions. The active and passive sensitivities of the proposed filters have been derived. Moreover, the quality factor ϑ and the central frequency ω_o of the proposed filters are insensitive to the current tracking error of a CFCCII$_p$. Finally, three experimental results are included to verify the theoretical prediction.

Keywords: four-terminal p-type active current conveyors, sensitivity

I. INTRODUCTION

Some recent literature has mentioned that the circuits based on current-mode amplifier will operate at high signal bandwidth, with greater linearity and with a larger dynamic range than their voltage-mode counterpart[1][2]. In addition to the above advantages, the current-mode approach can easily cascade the current-mode filters without additional matching circuits. Hence, the current-mode circuits have been receiving significant attention in the field of analogue signal processing. The four-terminal p-type active current conveyors (CFCCII$_p$) can circumvent the finite gain-bandwidth limitation of the conventional operational amplifier

* Corresponding author. Tel.:+886-2-29053801; Fax: +886-2-29042638
E-mail address : ee1004@mails.fju.edu.tw

[3]. A CFCCII$_p$ can offer both a constant bandwidth and a high slew rate (i.e. 2000 V/us)[4]. Recently, applications in the realization of various filter functions using some CFCCII$_p$s have been published [5]-[9]. However, the current-mode filter using only one CFCCII$_p$ has never been synthesized. In this paper, a circuit configuration for current-mode filter using single CFCCII$_p$ is presented. This proposed circuit can be configured to realize either a lowpass (LP), or a highpass (HP), or a bandpass (BP) filter function. These current-mode filters have high output impedance, so they can be cascaded without additional buffers. Furthermore, the quality factor ϑ and the central frequency ω_o of the proposed filters are insensitive to the current tracking errors of a CFCCII$_p$. Finally, three experimental results are given to confirm the afore mentioned theoretical analysis.

II.CIRCUIT DESCRIPTION

The circuit symbol for a CFCCII$_p$ is shown in Fig.1. The port relations of a CFCCII$_p$can be characterized as $i_z = i_x$, $v_x = v_y$, $i_y=0$ and $i_0 = i_x$. The proposed circuit using a single CFCCII$_p$ is shown in Fig.2 and the transfer function can be expressed as:

$$\frac{I_o}{I_{in}} = \frac{Y_2Y_4}{Y_1Y_2 + Y_1Y_3 + Y_2Y_3 + Y_2Y_4} \tag{1}$$

where $Y_1 - Y_4$ are the admittances.

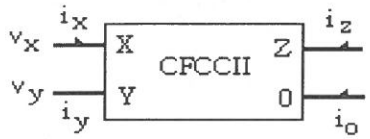


Fig.1. A CFCCII$_p$ symbol

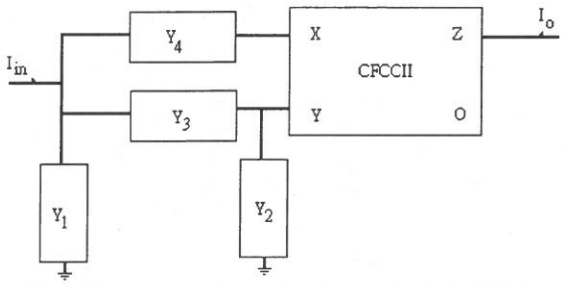


Fig.2. The proposed circuit using a single CFCCII$_p$

For (1), if $Y_1 = G_1$, $Y_2 = G_2$, $Y_3 = SC_3$ and $Y_4 = G_4$, as shown in Fig-3, then the transfer function has a biquadratic LP characteristic with

$$\frac{I_o}{I_{in}} = \frac{G_2G_4}{S^2C_1C_3 + S(C_1G_2 + C_3G_2) + G_2G_4} \tag{2}$$

the central frequency: $\omega_o = (\frac{G_2G_4}{C_1C_3})^{1/2}$

the quality factor: $\vartheta = \frac{1}{C_1 + C_3}(\frac{C_1C_3G_4}{G_2})^{1/2}$

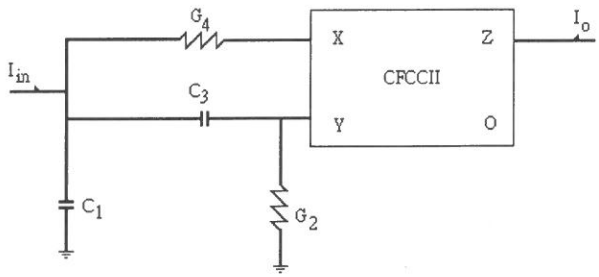


Fig.3. Second order lowpass filter

For (1), if $Y_1 = G_1$, $Y_2 = SC_2$, $Y_3 = G_3$ and $Y_4 = SC_4$, this circuit, as shown in Fig-4, is a biquadratic HP filter, with

$$\frac{I_o}{I_{in}} = \frac{S^2C_2C_4}{S^2C_2C_4 + S(C_2G_1 + C_2G_3) + G_1G_3} \tag{3}$$

the central frequency: $\omega_o = (\frac{G_1G_3}{C_2C_4})^{1/2}$

the quality factor: $\vartheta = \frac{1}{G_1 + G_3}(\frac{G_1G_3C_4}{C_2})^{1/2}$

Similarly, for (1), if $Y_1 = SC_1 + G_1$, $Y_2 = SC_2$, $Y_3 = SC_3 + G_3$ and $Y_4 = G_4$, this circuit, as shown in Fig-5, performs a second order BP function, with

$$\frac{I_o}{I_{in}} = \frac{SC_2G_4}{S^2(C_1C_2 + C_2C_3 + C_3C_1) + S(C_2G_1 + C_1G_3 + C_3G_1 + C_2G_3 + C_2G_4) + G_1G_3} \tag{4}$$

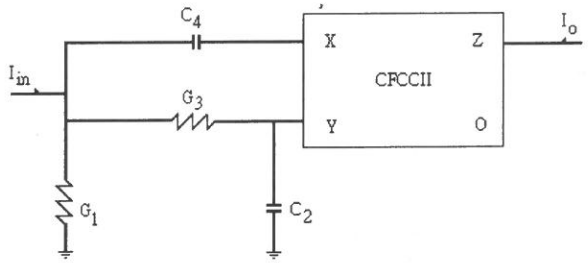


Fig.4. Second order highpass filter

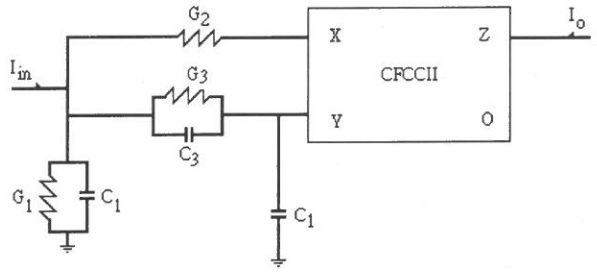


Fig.5. Second order bandpass filter

the central frequency: $\omega_o = \left(\frac{G_1 G_3}{C_1 C_2 + C_2 C_3 + C_3 C_1}\right)^{1/2}$

the quality factor: $\vartheta = \frac{G_1 G_3 (C_1 C_2 + C_2 C_3 + C_3 C_1)}{C_2 G_1 + C_1 G_3 + C_3 G_1 + C_2 G_3 + C_2 G_4}$

The quality factor ϑ and the central frequency ω_o of the proposed filters are derived.

The comparison between this paper and the recent ones〔12~14〕〔Gunes and Anday (1996), Chang and Tu (1998), Yin(2003)〕is shown in the table 1.

Number of CFCCII$_p$ components

Filter / Paper	Lowpass	highpass	bandpass
Gunes and Anday(1996)	3	3	3
Chang and Tu (1998)	4	4	4
Yin(2003)	2	2	2
Yin(present)	1	1	1

Table1. Comparison of the active components

From the above table, we can see that the required number of active components is indeed reduced.

III. SENSITIVITY ANALYSIS

Taking into consideration the non-idealities of $CFCCII_p$, namely $V_x = \alpha V_y$, $\alpha = \varepsilon_1$, $|\varepsilon_1| << 1$, denotes the voltage tracking error and $I_o = \beta I_x$, $\beta = 1 - \varepsilon_2$, $|\varepsilon_2| << 1$, denotes the output current tracking error, the transfer function (1) becomes

$$\frac{I_o}{I_{in}} = \frac{Y_2Y_4 + Y_3Y_4 - \alpha Y_3Y_4}{\beta(Y_1Y_2 + Y_1Y_3 + Y_2Y_3 + Y_2Y_4 + Y_3Y_4 - \alpha Y_3Y_4)} \tag{5}$$

By relating a sensitivity parameter F to the element of variation X_i by

$$S_{X_i}^{F} = \frac{X_i}{F}\frac{dF}{dX_i}$$

it is easy to show that the active and passive sensitivities of the parameters ω_0 and ϑ can be expressed as:

(1) *LP filter (for Fig.3):*

$$S_{G_2,G_4}^{\omega_0} = S_{G_4}^{\vartheta} = \frac{1}{2}$$

$$S_{C_1,C_3}^{\omega_0} = S_{G_2}^{\vartheta} = -\frac{1}{2}$$

$$S_{C_1}^{\vartheta} = \frac{1}{2} - \left(\frac{C_1}{C_1 + C_3}\right)$$

$$S_{C_3}^{\vartheta} = \frac{1}{2} - \left(\frac{C_3}{C_1 + C_3}\right)$$

$$S_{\alpha}^{\omega_0} = S_{\beta}^{\omega_0} = S_{\beta}^{\vartheta} = 0$$

$$S_{\alpha}^{\vartheta} = \frac{\alpha C_3G_4}{C_1G_2 + C_3G_2 + C_3G_4 - \alpha C_3G_4}$$

(2)*HP filter (for Fig.4):*

$$S^{\omega_0}_{G_1,G_3} = S^{\vartheta}_{C_4} = \frac{1}{2}$$

$$S^{\omega_0}_{C_2,C_4} = S^{\vartheta}_{C_2} = -\frac{1}{2}$$

$$S^{\vartheta}_{G_1} = \frac{1}{2} - \frac{G_1}{G_1 + G_3}$$

$$S^{\vartheta}_{G_3} = \frac{1}{2} - \frac{G_3}{G_1 + G_3}$$

$$S^{\omega_0}_{\alpha} = S^{\omega_0}_{\beta} = S^{\vartheta}_{\beta} = 0$$

$$S^{\vartheta}_{\alpha} = \frac{\alpha C_4 G_3}{C_2 G_3 + C_2 G_1 + C_4 G_3 - \alpha C_4 G_3}$$

(3)*BP filter (for Fig.5):*

$$S^{\omega_0}_{G_1,G_3} = \frac{1}{2},$$

$$S^{\omega_0}_{C_1} = -\frac{1}{2}\frac{C_1(C_2 + C_3)}{\Delta},$$

$$S^{\omega_0}_{C_2} = -\frac{1}{2}\frac{C_2(C_1 + C_3)}{\Delta}$$

$$S^{\vartheta}_{G_4} = -1,$$

$$S^{\omega_0}_{C_3} = -\frac{1}{2}\frac{C_3(C_1 + C_2)}{\Delta},$$

$$S^{\vartheta}_{G_1} = 1 - \frac{G_1(C_2 + C_3)}{\Delta_1}$$

$$S^{\vartheta}_{G_3} = 1 - \frac{G_3(C_1 + C_2)}{\Delta_1},$$

$$S^{\vartheta}_{C_1} = \frac{C_1(C_2 + C_3)}{\Delta} - \frac{\Delta G_3}{\Delta_1}$$

$$S^{\vartheta}_{C_3} = \frac{C_3(C_1 + C_2)}{\Delta} - \frac{\Delta G_1}{\Delta_1}$$

$$S^{\omega_0}_{\beta} = S^{\vartheta}_{\beta} = 0$$

$$S^{\omega_0}_{\alpha} = \frac{1}{2}$$

$$S_{C_2}^{\vartheta} = \frac{C_2(C_1+C_3)}{\Delta} - \frac{\Delta(G_1+G_3+G_4)}{\Delta_1}$$

$$S_{\alpha}^{\vartheta} = \frac{\alpha C_3 G_4}{C_2G_1 + C_1G_3 + C_3G_1 + C_2G_3 + C_2G_4 - \alpha C_3G_4 + C_3G_4}$$

where $\Delta_1 = C_1G_3 + C_2G_1 + C_3G_1 + C_2G_3 + C_2G_4$ and $\Delta = C_1C_2 + C_1C_3 + C_2C_3$.

Clearly, the passive sensitivities of the proposed filter are very low. Moreover, the natural frequency ω_o and the quality factor ϑ of the proposed current-mode filters are insensitive to the current tracking error of a $CFCCII_p$.

IV. EXPERIMENTAL RESULTS

To verify the theoretical prediction of the proposed circuit, three filter prototypes have been constructed with discrete components. The AD844 can be constructed as a $CFCCII_p$. All the experiments of the three mentioned filters are described below:

(a) A biquadratic LP filter, shown in Fig.3, was constructed with $G_2=G_4=10^{-3}\ \Omega^1$, $C_1=C_3=1uF$. The experimental results for the gain and phase responses are shown in Fig.6(a)(b).

(b) A biquadratic HP filter, shown in Fig.4, was constructed with $G_1=G_3=10^{-3}\ \Omega^1$, $C_2=C_4=1uF$. The experimental results for the gain and phase responses are shown in Fig.6(c)(d).

(c) A biquadratic BP filter, shown in Fig.5, was constructed with $G_1=G_3=G_4=10^{-3}\ \Omega^1$, $C_1=C_2=C_3=1uF$. The experimental results for the gain and phase responses are shown in Fig.6(e)(f). The theoretical analysis correlated with the measured results with few errors which due to the errors of the use of passive elements. Experimental results confirm the results of the theoretical analysis.

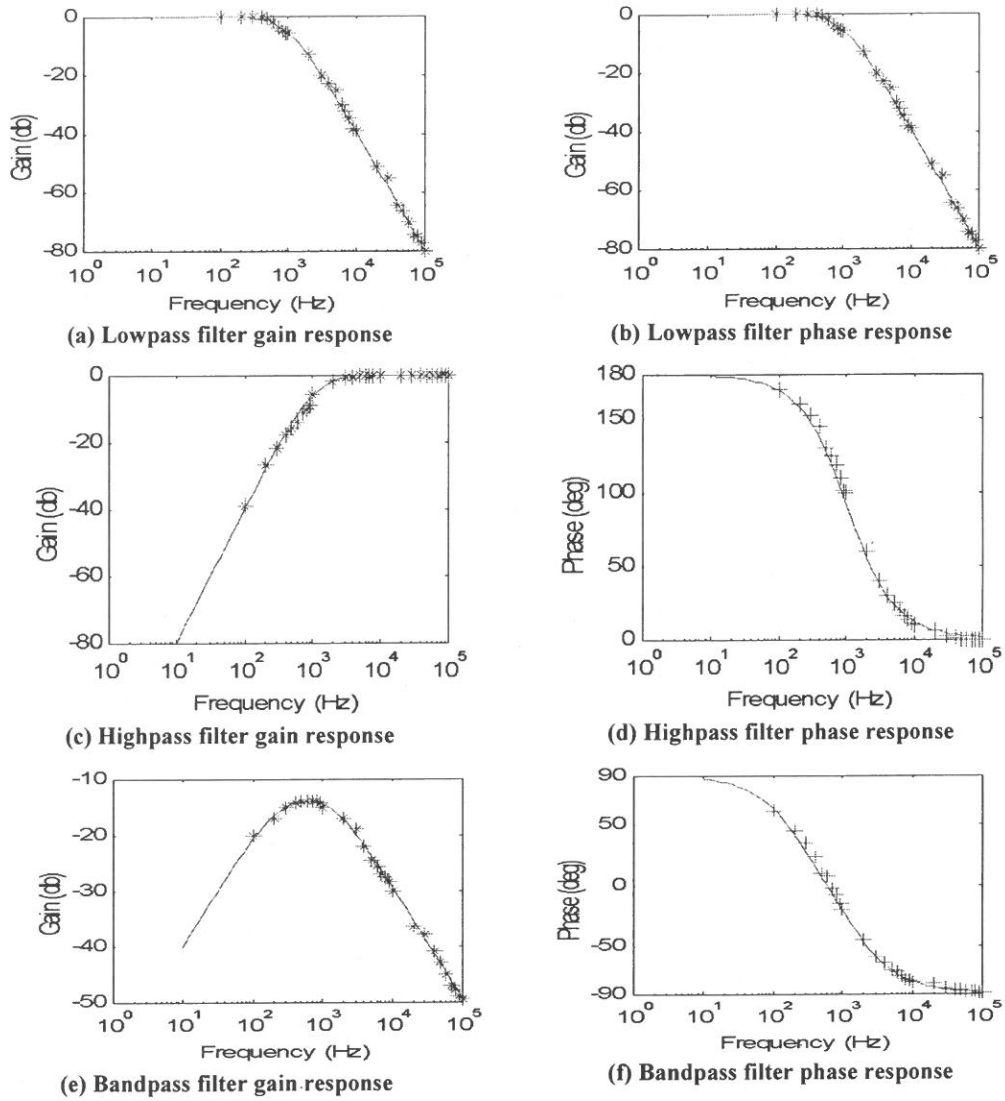


Fig.6. (a): Lowpass filter gain response (b): Lowpass filter phase response (c): Highpass filter gain response (d): Highpass filter phase response (e): Bandpass filter gain response (f): Bandpass filter phase response 0: experimental result for gain +: experimental result for phase

V.CONCLUSION

A new configuration circuit of bandpass, lowpass and highpass filters using single a CFCCII$_p$ has been presented. These filters can be achieved from the same configuration. The proposed current-mode single- CFCCII$_p$ filter can be easily cascaded with matching circuits.

The central frequency ω_o and the quality factor ϑ of the proposed are insensitive to the current tracking error of a $CFCCII_p$. Finally, three experimental results confirmed the theoretical analysis.

VI.ACKNOWLEDGMENT

The authors wish to express their thanks for the financial support of the Office of Research and Development Fu Jen Catholic University.

REFERENCES

(1) B.Wilson, "Recent Developments in Current Conveyors and Current-Mode Circuits", *IEE Proc-G*, vol. 137, no. 2, pp.63-77, 1990.

(2) G.W.Rober and A.S.Sedra, "All Current-Mode Frequency Selective Circuits", *Electron. Lett.*, Vol.25, pp.759-761, 1989.

(3) C.Toumazou and E.J.Lidgey, "Universal active filter using current conveyous", *Electron. Lett.*,vol. 22, pp.662-664, 1986.

(4) Y.Sun and J.K.Fidler, "Versatile active biquad based on second-generation current conveyors", *Int. J Electron.*, vol. 76, pp91-98 , 1994.

(5) V.K.Singh and R.Senani, "New multifunction active configuration employing current conveyors", *Electron. Lett.*, vol. 26, pp.1814-1816, 1990.

(6) G.W.Robert and A.S.Sedra, "A general class of current amplifier-based biquadratic filter circuits", *IEEE Transactions on Circuits and Systems.,* vol. 39, pp.257-263 , 1992.

(7) C.M.Chang, "Universal active current filters using single second-generation current conveyors", *Electron. Lett.*, vol. 27, no.18, pp.1614-1617 , 1991

(8) C.M.Chang and P.C.Chen, "Universal active current filter with three inputs and one output using current conveyors", *Int. J Electron.*, vol. 71, no.5, pp.817-819 , 1991.

(9) R. Senani, "New current-mode biquad filter", *Int. J Electron.*, vol. 73, no. 4, pp.735-742 , 1992.

(10) Y.C. Yin, "A Current-mode Biquad filter with multi-output using CCIIs", *Journal of the Chinese Institute of Electrical Engineering.*, vol. 3, no. 4, pp.367-370 ,1996.

(11) C.L. Hou and J.S. Wu, "Universal cascadable current-mode biquad using only four four

CCIIs", *Int. J Electron.*, vol. 82, no.2, pp.125-129 , 1997.

(12) E.O.Gues and F. Anday, "Realization of current-mode universal filter using CFCCII$_p$s ", *Electron. Lett.*, vol. 32, pp.1081-1082 , 1996.

(13) C.M. Chang and S.H. Tu, "Universal current-mode filters employing CFCCII$_p$s", *Int. J Electron.*, vol. 85, no.6, pp.749-754 , 1998.

(14) Y.C. Yin, "Current-Mode Biquad Using Two CFCCII$_p$s ", Fu Jen Studies., vol. 3, no. 4, pp.367-370 , 2003.

received October 13, 2004
revised November 04, 2004
accepted December 03, 2004

使用單一四端 P 型主動電流傳輸器合成電流式高通低通帶通濾波器

鄞永昌　劉岳乘
輔仁大學大學電子工程學系

摘　要

本文提出使用單一四端 P 型主動電流傳輸器合成新結構二階式電流形態濾波器，此單一電流迴授放大器合成之濾波器無需元件匹配條件。本文提出之濾波器可合成二階式低通， 高通和帶通三種功能。其主動和被動靈敏度在文中已被推導出， 而且品質因素與中心頻率沒有電流之軌跡誤差問題。最後以三個實驗驗證本文之理論預測。

關鍵詞：四端 P 型主動電流傳輸器，靈敏度

Generation of Deep-Ultraviolet Coherent Light Below 200 nm in $KBe_2BO_3F_2$ (KBBF) Crystal

Lien-Bee Chang*
Department of Physics,
Fu Jen Catholic University,
Taipei 242, Taiwan, ROC

S. C. Wang
Institute of Electro-Optical Engineering,
National Chiao Tung University,
Hsinchu 300, Taiwan, ROC

A. H. Kung
Institute of Atomic and Molecular Sciences,
Academia Sinica
Taipei 106, Taiwan, ROC

Abstract

By use of a $KBe_2BO_3F_2$ (KBBF) crystal with a thickness of 0.4 mm and a special prism-coupled technique, fourth-harmonic generation was obtained by direct frequency doubling of the second harmonics of a tunable pulse-amplified cw Ti:sapphire laser system that generated 8-ns pulses at a repetition rate of 30 Hz. The tuning range was from 186 to 200 nm, and the maximum conversion efficiency was reached at the fourth-harmonic generation wavelength of 193 nm.

Keywords: Frequency conversion, KBBF, Deep-ultraviolet, Second-harmonic generation, Fourth-harmonic generation

* Corresponding author, Tel.,+886-2-29052018; fax,+886-2-29021038
E-mail address: lbchang@mails.fju.edu.tw

INTRODUCTION

With the development of 193-nm and even 157-nm photolithography as well as various applications in photochemistry, photobiology, spectroscopy, and other applications that require high photon energy, coherent light sources in the deep-UV (DUV) range, in particular below 200 nm, have become more significant. Although excimer lasers can emit coherent light with high average power output, efficient tunable solid-state lasers are still needed because of their narrower spectral bandwidth, and better beam quality for the inspection of optics and tunability for high-resolution spectroscopy. Fourth-harmonic generation of the widely tunable (800-700 nm) output of Ti:sapphire lasers with good beam quality is an attractive approach to generate DUV coherent radiation below 200 nm.

There are several UV nonlinear optical (NLO) crystals that can produce DUV coherent light below 200 nm by sum-frequency generation (SFG) such as (-BaB_2O_4 (BBO) [1-4], LiB_3O_5 (LBO) [5-6], $Li_2B_4O_7$ (LB4) [7], CsB_3O_5 (CBO) [8-9], $CsLiB_6O_{10}$ (CLBO) [10-13], $KB_5O_8 \cdot 4H_2O$ (KB5) [14-15], and $K_2Al_2B_2O_7$ (KABO or KAB) [16]. However, these SFG techniques always need two laser beams, one's wavelength is short and the other's long. The requirement for two beams can be inconvenient when one employs SFG for practical applications. In addition, more than two nonlinear optical crystals are needed to achieve DUV generation.

For example, three widely used approaches for the generation of 193-nm radiation are described as follows, One approach is based on Nd:YAG laser system, and the short wavelength is the fifth harmonic of the fundamental radiation, while the long one is obtained by the idler wave of an optical parametric oscillator [17]; another is based on the eighth harmonics of Er^{3+}-doped fiber amplifier [18]; the final one is based on Ti:sapphire laser [19]. For the first two methods, it would require more than five nonlinear optical crystals to achieve the 193-nm generation, whereas more than two NLO crystals are needed to complete the generation for the third approach. Furthermore, better spatial and temporal overlapping is also required to increase the conversion efficiency for pulsed laser systems.

The main reason SFG techniques are still used to generate DUV coherent radiation below 200 nm is that no NLO crystals, as mentioned above, are able to perform direct frequency doubling below 200 nm. Fortunately, this situation has been changed since the

$KBe_2BO_3F_2$ (KBBF) crystal was developed. The KBBF crystal, which has a large birefringence and a transmittance down to 155 nm, is expected to generate radiation below 160 nm. A comparison of these NLO crystals is given in Table 1, and the related references are shown in this table. From Table 1 KBBF is the only NLO crystal we can employ to obtain the coherent radiation below 200 nm by direct frequency doubling. The crystal belongs to the negative uniaxial class and its transparency is down to 155 nm. In addition, it is the only NLO crystal to date which can realize direct frequency doubling to DUV below 200 nm by type-I second-harmonic generation (SHG). SHG wavelength as short as 165 nm may be obtained by use of this crystal [20-21].

Table 1. Nonlinear optical crystals applicable for SFG below 200 nm. The transparency is given for the zero level transmission.

NLO crystals	Transparency (nm)	Nonlinear coefficient [pm/V]	Shortest Type-I SHG (nm)
β-BaB2O4 (BBO) (ref. 29)	189-3500	d11 = 2.3	205
LiB3O5 (LBO) (ref. 29)	160-2600	d31 =1.05	277
Li2B4O7 (LB4) (refs. 33-34)	160-3500	d31 =0.15	243.8
CsB3O5 (CBO) (refs. 35-37)	170-3000	d14 =1.08	272.8
CsLiB6O10 (CLBO) (ref. 38)	180-2750	d36 = 0.95	235
KB5O8 ·4H2O (KB5) (refs. 39-41)	162-1500	d31 =0.045	217
K2Al2B2O7 (KABO or KAB) (ref. 42)	180-3600	d11 =0.45	232.5
KB2BO3F2 (KBBF) (ref. 20)	155-3500	d11 = 0.76	164

Consequently, the coherent light below 200 nm can be generated by two sequential fre-

quency-doubling stages using KBBF as the frequency-quadrupling NLO crystal if the wavelength of fundamental light is shorter than 800 nm.

A number of important properties of this crystal, such as phase-matching angles, effective nonlinear coefficients and group-velocity match for both type-I and type-II SHG, have theoretically been investigated [22-24]. Those studies show that KBBF is attractive for ultrashort pulse generation, particularly in DUV region because of its small group-velocity dispersion and large spectral bandwidth. However, the thickness of KBBF for efficient frequency conversion in nanosecond or sub-nanosecond regime is too thin so far. The main problem of the crystal lies in its layered structure that makes it very difficult to grow [25-26].

KBBF crystal is currently too thin to be cut at any phase-matching angle. Up to now the thickest crystal available is only 1.8 mm (being the length along the z-axis). If the phase-matching condition is realized by tuning the external angle of incidence, the angle will increase to 90° at SHG wavelength of 235 nm as shown in Fig. 1(a) where n_2 and n_1 are the refractive index of KBBF crystal and the medium where the wave is incident to the interface, and α and β are the external angle of incidence and the related phase-matching angle, respectively. In other words, phase-matching condition cannot be realized for the wavelength of SHG below 235 nm by directly tuning the angle of incidence. Another problem is the high Fresnel reflection loss associated with large angle of incidence. To overcome this problem, Lu et al. adopted a special prism-coupled technique, described in detail in reference 27 and shown in Fig. 1(b), to avoid the difficulty in cutting the crystal. The present shortest SHG wavelength of 172.5 nm and SFG wavelength of 163 nm have been achieved by this technique [27].

KBBF possesses high damage threshold [28]. As a result, more tightly focusing can be used in experiment to achieve higher conversion efficiency. A comparison between LBO, BBO and KBBF is shown in Figs. 2 and 3 (Type-I SHG phase-matching angle vs. wavelength curve of KBBF is illustrated in Fig. 4). From the comparison it is found that KBBF crystals possess moderate birefringence, and the value of $n_o(\lambda_f)$-$n_e(\lambda_{SH})$ does not drop as much as that of BBO. As a result, the shortest Type-I SHG for KBBF can reach down to 164 nm. In addition, the internal angular bandwidth for KBBF is also medium, compared to those for LBO and BBO. Thus the requirement for laser beam quality is not as high as that for BBO. These results are calculated according to Sellmeier's equations in reference 29 for LBO and BBO,

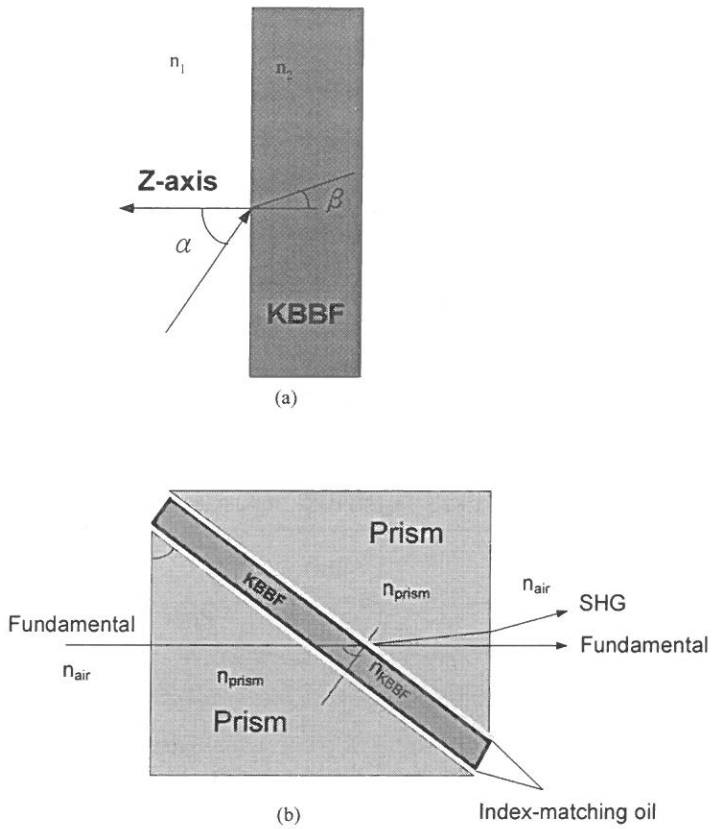


Fig. 1 (a) General SHG technique in KBBF crystal, and (b) the prism- coupled technique (PCT)

and in references 20 and 28 for KBBF.

In this paper we will describe the generation of deep UV coherent light. By use of a $KBe_2BO_3F_2$ (KBBF) crystal with a thickness of 0.4 mm and a special prism-coupled technique, fourth harmonics was obtained by direct frequency doubling of the second harmonics of a tunable pulse-amplified CW Ti:sapphire laser system. The tuning range was from 200 to 186 nm.

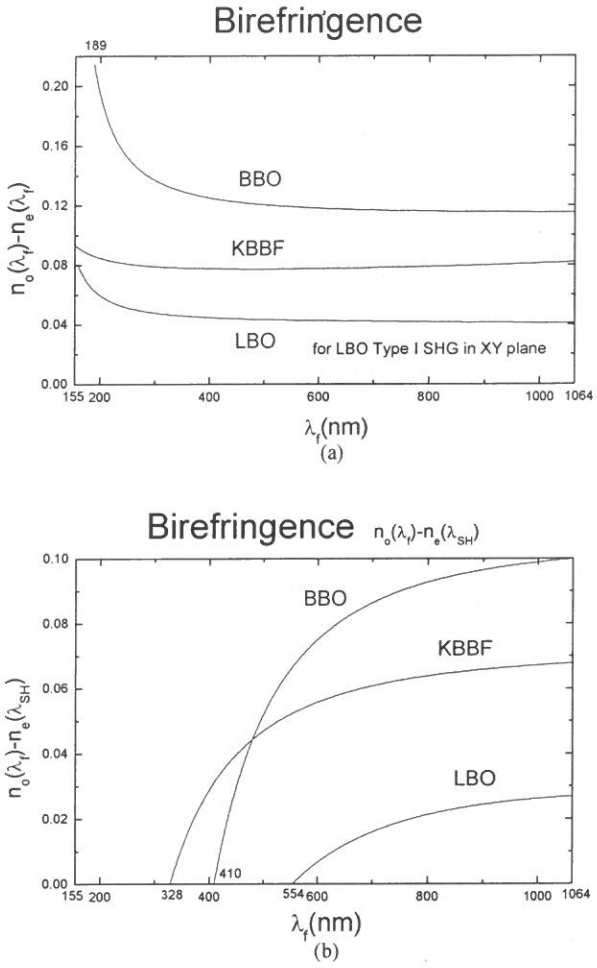


Fig. 2 A comparison of birefringence between BBO, LBO and KBBF.

EXPERIMENTAL ARRANGEMENT

The schematic of the experimental arrangement for the generation of 186-196 nm radiation is shown in Fig. 5. A tunable pulse-amplified single-longitudinal-mode cw Ti:sapphirre laser system with a repetition rate of 30 Hz was used as the pump laser [30]. The pulses of the laser had a pulse width ($1/e^2$) of approximately 8 ns. Figure 6 shows the block diagram of the pump laser. Two harmonic generation crystals were arranged in tandem to generate the fourth-harmonic radiation of the fundamental light of the laser system by two successive fre-

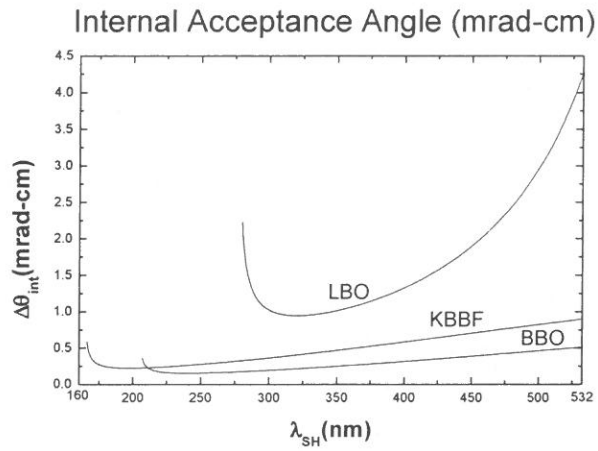


Fig. 3. A comparison of internal angular bandwidth between BBO, LBO and KBBF.

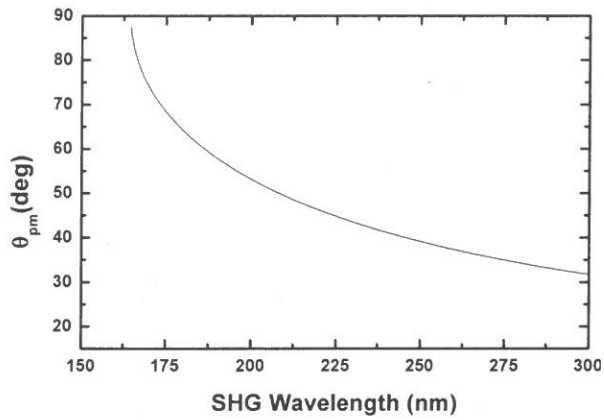


Fig. 4. The type-I SHG phase-matching curve of KBBF crystal.

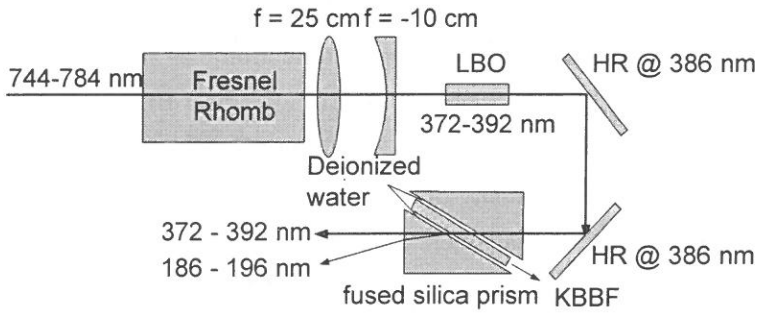


Fig. 5. Schematic of experimental setup for the generation of tunable DUV coherent light.

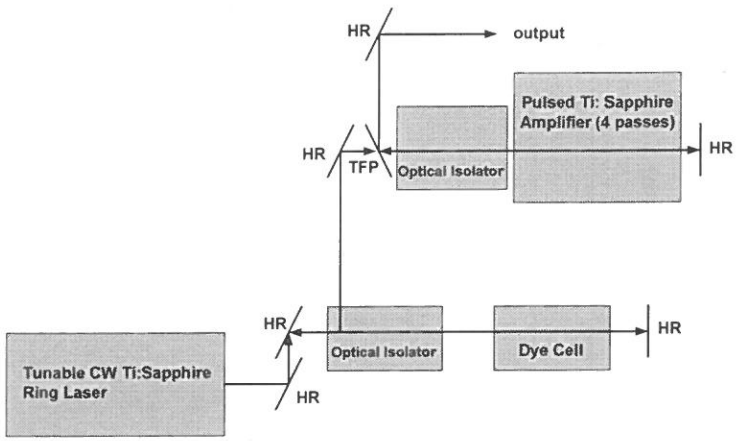


Fig. 6. The block diagram of tunable pulse-amplified CW Ti:sapphire laser system.

quency-doubling stages. The SHG crystal was a 5x5x15 mm^3 type-I LBO whose end faces were dual antireflection coated for 772 and 386 nm. It was oriented so that the generated light was vertically polarized. The crystal for four-harmonic generation was a 0.4 mm thick KBBF crystal.

Two 63-degree prisms made of Corning 7980 ArF laser grade fused silica were used in the experiment. The incident angle to the prism was about 2 to 10 degree because the phase-matching angles of KBBF are from 53 to 60 degree. Both the front and back interfaces of KBBF were filled with deionized water for convenience. The Fresnel reflection loss was about 9% for second-harmonic radiation incident on the interfaces between the prism, deionized water and KBBF because the polarization of the radiation was perpendicular to the plane of incidence. However, the polarization of fourth-harmonic radiation was parallel to the plane of incidence, which would minimize the Fresnel reflection loss when the angles of incidence on these interfaces between KBBF, water and prism were near Brewster's angle. The loss was estimated less than 3 %. The energy of each fundamental pulse was approximately 8 mJ with the resulting input peak power density of ~ 280 MW/cm^2. The powers and the full widths at half maximum (FWHM) of the fundamentals, the second harmonics and the fourth harmonics were measured by a calibrated thermopile with a time constant of ~ 0.7 s and by a 500-MHz-bandwidth oscilloscope, respectively.

RESULTS AND DISCUSSION

The experimental results without further focusing are shown in Table 2. It is interesting to note that the maximum conversion efficiency of ~ 0.46% was reached at the fourth-harmonic generation wavelength of 193 nm. This was due to the minimum Fresnel loss for the process of the fourth-harmonic generation at 193 nm. The input second-harmonic peak power densities were estimated approximately 180 MW/cm^2, which was in good agreement with the numerical calculation [31]. When the fourth harmonics move to the shorter end, the loss due to the absorption of deionized water, besides the Fresnel one, increases so seriously that the output energy of each fourth-harmonic pulse dropped from 12.3 μJ at 188 nm to 7.7 μJ at 186 nm. Such low conversion efficiencies were attributed to the short effective interaction length. In addition, due to high conversion efficiency from fundamentals to second harmonics (40-45%), the pulse widths of the second harmonics increased to ~ 6.5 ns from $\frac{8}{\sqrt{2}}$ ns, corresponding to the small-signal case [32]. In contrast, the pulse widths of the fourth harmonics reduced to ~ $\frac{6.5}{\sqrt{2}}$ ns because of extremely low conversion efficiency for the second doubling.

Table 2. The input energy of fundamental light and output energies of second harmonic and 4H generated in the processes of SHG and 4HG for each pulse. Where f, SH and 4H stand for fundamental, second harmonic and fourth harmonic respectively.

λ_{4H} (nm)	186	188	190	193	196
E_f (mJ)	8	8	8	8	8
E_{SH} (mJ)	3.2	3.4	3.5	3.6	3.5
E_{4H} (μJ)	7.7	12.3	12.7	16.7	14.8

For small signal and plane-wave approximations, the second-harmonic generation conversion efficiency is proportional to the fundamental intensity; in contrast, the conversion efficiency is proportional to the square of the effective interaction length [32]. Therefore, extending the effective interaction length is more efficient for SHG than increasing the fundamental intensity. Furthermore, the latter would raise the possibility of damaging the crystal. In addition, for the wavelengths of fourth-harmonics shorter than 186 nm, the experiment should be performed in vacuum to avoid serious absorption by deionized water and air.

In this experiment we found that after a few minutes, the output power of the fourth harmonics dropped significantly. This is because the deionized water film between the two

prisms would vaporize gradually, which resulted from the absorption of the fourth-harmonic radiation by deionized water. The problem could be solved by put the bottom of the prisms in a tank filled with deionized water. By means of the molecular force, the molecules of deionized water in the tank can be pulled upwards and fill the vacancies caused by vaporization to make the output power of the DUV radiation more stable without serious Fresnel reflection loss.

CONCLUSON

This paper has described a tunable coherent light source from 196 to 186 nm generated by two successive frequency-doubling stages. More fourth-harmonic energies would be extracted if the second harmonics could be focused further. Thicker crystals can provide longer interaction lengths, thus increasing the conversion efficiency. However, at present the thickest crystal available is only 1.8 mm. A more efficient approach is to extend the effective interaction length using a double-pass configuration, as described in reference 34. This method may increase the conversion efficiency significantly without further focusing the second harmonics to avoid damaging the crystals.

REFERENCES

(1) M. Watanabe, K. Hayasaka, H. Imajo, J. Umezu and S.Urabe, Appl. Phys. B 53, p. 11, 1991.

(2) J. Ringling, O. Kittelmann, F. Noack, G. Korn and J. Squier, Opt. Lett. 18, p. 2035, 1993.

(3) J. Lublinski, M. Muller, F. Laeri and K. Vogler, Appl. Phys. B 61, p. 529, 1995.

(4) H. Kouta and Y. Kuwano, Opt. Lett. 17, p. 1230, 1999.

(5) F. Seifert, J. Ringling, F. Noack, V. Petrov and O. Kittelmann, Opt. Lett. 19, p. 1538, 1994.

(6) B. Wu, F. Xie, C. Chen, D. Deng and Z. Xu, Opt. Commun. 88, p. 451, 1992.

(7) V. Petrov, F. Rotermund, F. Noack, R. Komatsu, T. Sugawara and S. Uda, J. Appl. Phys. 84, p. 5887, 1998.

(8) K. Kato, IEEE J. Quantum Electron. 31, p. 169, 1995.

(9) Y. Kagebayashi, K. Deki, Y. Morimoto, S. Miyazawa and T. Sasaki, Jpn. J. Appl. Phys. 39, p. L1224, 2002.

(10) N. Umemura and K. Kato, APPl. Opt. 36, p. 6794, 1997.

(11) V. Petrov, F. Noack, F. Rotermund, M. Tanaka and Y. Okada, Appl. Opt. 39, p. 5076, 2000.

(12) J. Sakuma, K. Deki, A. Finch, Y. Ohsako and T. Yokota, Appl. Opt. 39, p. 5505, 2000.

(13) S. B. Mirov, V. V. Fedorov, B. Boczar, R. Frost and B. Pryor, Opt. Commun. 198, p. 403, 2000.

(14) N. Umemura and K. Kato, Appl. Opt. 35, p. 5332, 1996.

(15) V. Petrov, F. Rotermund and F. Noack, Electron. Lett. 34, p. 1748, 1998.

(16) K. Kato and E. Takaoka, Conference on Lasers and Electro-Optics, 2002 OSA Technical Digest (Optical Society of America, Long Beach, 2002) p. 331.

(17) C. E. Hamilton, C. B. Doughty, P. M. Roper, R. D. Mead and S. C. Tidwell, Lasers and Electro-Optics Society Annual Meeting, 1998, LEOS'98 IEEE, Vol. 1, p. 322.

(18) H. Kitano, H. Kawai, T. Ohtsuki and S. Owa, The 4th Pacific Rim Conference on Lasers and Electro-Optics 2001, Chiba, Japan. Technical Digest, Vol. 2, p. 394.

(19) W. B. Yan, T. F. Steckroat, R. A. Frost, J. C. Walling and D. F. Heller, Conference on Lasers and Electro-Optics, 1997 OSA Technical Digest (Optical Society of America, Baltimore, 1997) p. 485.

(20) C Chen, Z. Xu, D. Deng, J. Zhang, G. K. L. Wong, B. Wu, N. Ye and D. Tang, Appl. Phys. Lett. 68, p. 2930, 1996.

(21) J. LU, G. Wang, Z. Xu, C. T. Chen, J. Wang, C. Zhang and Y. Liu, Opt. Commun. 200, p. 415, 2001.

(22) F. Zhang and M. Yonemura, Jpn. J. Appl. Phys. 36, p. 6353, 1997.

(23) V. Petrov, F. Rotermund, F. Noack, J. Ringling, O. Kittelmann and R. Komatsu, IEEE J. Selec. Topics Quantum Electron. 5, p. 1532, 1999.

(24) K. Nagashima and L. Liu, Opt. Laser Technol. 33, p. 611, 2001.

(25) C. Chen, Y, Wang, Y. Xia, B. Wu, D. Tang, K. Wu, W. Zeng, L. Yu and L. Mei, J. Appl. Phys. 77, p. 2268, 1995.

(26) D. Tang, Y. Xia, B. Wu and C. Chen, J. Crystal Growth 222, p. 125, 2001.

(27) T. Togashi, T. Kanai, T. Sekikawa, C. Chen, C. Zhang, Z. Xu and J. Wang, Opt. Lett. 28

(2003) 254.

(28) B. Wu, D, Tang, N. Ye and C. Chen, Opt. Mater. 5, p. 105, 1996.

(29) CASTECH. Crystal catalog (1999/2000).

(30) C. K. Ni and A. H. Kung, Appl. Opt. 37, p. 530, 1998.

(31) L. B. Chang and S. C. Wang, Jpn. J. Appl. Phys. 43, p. 550, 2004.

(32) J. A. Armstrong, N. Bloembergen, J. Ducuing and P. S. Pershan, Phys. Rev. 127, p. 1918, 1962.

(33) T. Sugawara, R. Komatsu, and S. Uda: Solid State Commun. 107, p. 233, 1998.

(34) V. Petrov, F. Rotermund, F. Noack, R. Komatsu, T. Sugawara, and S. Uda: J. Appl. Phys. 84, p. 5887, 1998.

(35) Y. Wu, T. Sasaki, S. Nakai, A. Yokotani, H. Tang, and C. Chen: Appl. Phys. Lett. 62, p. 2614, 1993.

(36) K. Kato: IEEE J. Quantum Electron. 31, p. 169, 1995.

(37) H. Kitano, T. Matsui, K. Sato, N. Ushiyama, M. Yoshimura, Y. Mori, and T. Sasaki: Opt. Lett. 28, p. 263, 2003.

(38) Y. Mori, I. Kuroda, S. Nakajima, T. Sasaki, and S. Nakai: Appl. Phys. Lett. 67, p. 1818, 1995.

(39) J. A. Paisner, M. L. Spaeth, and D. C. Gerstenberger, and I. W. Ruderman: Appl. Phys. Lett. 32, p. 476, 1978.

(40) H. J. Dewey: IEEE J. Quantum Electron. QE-12, p. 303, 1976.

(41) N. Umemura and K. Kato: Appl. Opt. 35, p. 5332, 1996.

(42) N. Ye, W. Zeng, J. Jiang, B. Wu, C. Chen, B. Feng, and X. Zhang: J. Opt. Soc. Am. B. 17, p. 764, 2000.

藉由 KBBF 晶體產生短於 200 nm 之同調光

張連璧
輔仁大學物理系

王興宗
交通大學光電工程研究所

孔慶昌
中研院原分所

摘　要

我們利用一片厚度只有 0.4 mm 的 KBBF 非線性光學晶體並應用稜鏡耦合的技術 (prism-coupled technique)，直接將二倍頻光倍頻，而基頻雷射光爲一個將可調式之連續藍寶石雷射經由脈衝放大(pulse-amplified)所構成的雷射系統，放大後其脈衝寬度約爲 8 奈秒，重複率爲 30 Hz。經由連續兩次二倍頻的非線性轉換，成功的產生奈秒級的脈衝可調之深紫外同調光，可調的範圍從 196 nm 到 186 nm。在波長爲 193 nm 的四倍頻達到最大的轉換率。

關鍵詞：頻率轉換、KBBF、深紫外光、二倍頻、四倍頻。

200 萬畫素 4 片式影像鏡頭設計

洪國書 王昌偉 曾詠傑
大葉大學電機工程研究所

韓 斌*
國立中興大學精密工程研究所

摘 要

本文是以 4 片式光學結構方式設計，提出適用於兩百萬畫素 CMOS Sensor(Complementary Metal Oxide Semiconductor)之手機相機鏡頭設計。由於使用一片玻璃球面鏡片再加上三片塑膠材質的鏡片來設計，與目前市售使用玻璃非球面鏡頭比較，此方式可大幅降低鏡頭成本又可滿足系統要求。

關鍵詞 ：鏡頭設計、光學設計、MTF 值、TV 歪曲像差。

一、前 言

手機在目前人類的生活中已經成爲必備的通訊裝置，從早期只是單純的通話連絡，到近年來手機搭配相機攝影的興起，使得手機的應用越來越廣泛。本文使用一片玻璃鏡片及三片塑膠鏡片搭配非球面方式，適用於兩百萬畫素手機相機鏡頭使用。

二、理論分析

本研究採用 CMOS Sensor，其長、寬分別爲 5.14mm x 3.85mm，對角線長度爲 6.422mm。光學傳遞含數 MTF 中心值（Field=0）必須在 155Lp/mm 時辨識率大於 30% 以上才能符合系統要求。MTF 的計算式如下：

* Corresponding author. Tel: +886-2-22850405; Fax:+886-2-22858362
E-mail: pin@dragon.nchu.edu.tw

MTF = 1/[2x(5.14/1600)] = 155.6 Lp/mm ≒ 155 Lp/mm

表一 本研究之系統要求

No	項目	規格
1	視角 (Field Angle)	60° ±1°
2	MTF (Field = 0) Center	155lp/mm > 30%
3	MTF (Field = 1.0)	155lp/mm > 25%
4	主光線角度 (Chief Ray Angle)	< 16°
5	TV 歪曲像差 (TV Distortion)	< 0.5 %
6	相對照度 (Relative Illumination)	> 60 %

從幾何光學理論得知，y-nu 方法是屬於近軸光線的追跡方法，在高斯光學的範圍內，討論光線在系統中的行進路線。光線在系統之中的行進路線可分爲二種方式：(1)折射過程 (2)轉移過程 [1]

(1) 折射過程 (Refracted Procedure)

如圖 1 爲一球面，兩邊的界質爲 n 及 n'，一條與光軸夾 u 角度的光線入射至球面上，光線在面上的高度爲 y，光線經過球面折射後，以與光軸夾角 u'的角度從球面射出。M 與 M'分別爲入射、出射光線和光軸的焦點，因此滿足

$$\frac{n}{s}+\frac{n'}{s'}=\frac{n'-n}{r} \tag{A-1}$$

將上式改寫成

$$\frac{n'}{s'}=\frac{n'-n}{r}=-\frac{n}{s} \tag{A-2}$$

將(A-2)式中的每一項乘 y

$$n'\frac{y}{s'}=y\left(\frac{n'-n}{r}\right)-n\frac{y}{s} \tag{A-3}$$

利用近軸條件(A-3)可改寫成

$$n'u'=nu+yc\,(n'-n) \tag{A-4}$$

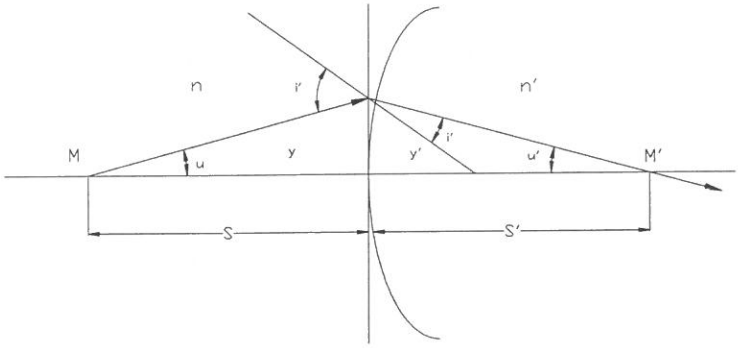


圖 1　折射過程

(2) 轉移過程 (Transfer Procedure)

轉移過程是討論光線經過球面折射後，行進到下一個球面之間的關係。如圖 2 假設兩球面間距爲 t，光線在第一個球面上的高度爲 y1，第二個球面上的高度爲 y2，兩者之間的關係

$$y2 = y1 - u't \tag{B-1}$$

將(B-1)式改寫成

$$y2 = y1 + n'1u'1\left(\frac{-t}{n'1}\right) \tag{B-2}$$

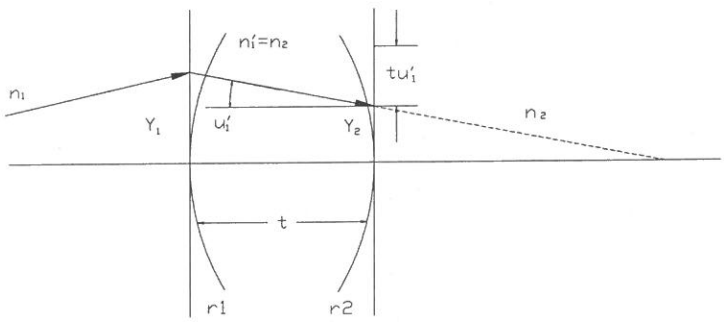
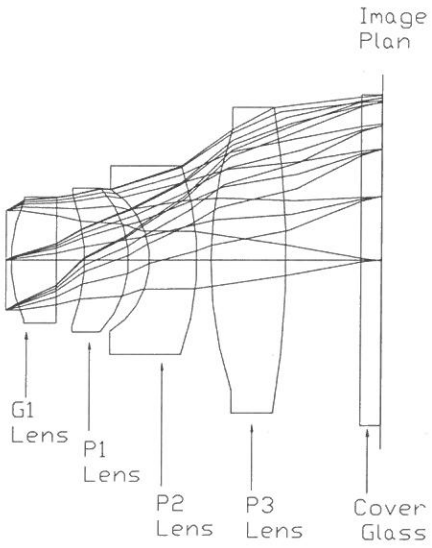


圖 2　轉移過程

利用(A-4)式及(B-2)式爲基礎，帶入 Excel 計算表中執行計算與分析，可以將光線在系統中所走的軌跡追蹤出來。使用公式計算所得到之初始資料，輸入光學設計軟體 Zemax 中，進行光學最佳化(Optimization)求得最終數據。

下圖爲使用 Zemax 軟體設計之二百萬畫素鏡頭光路圖。



圖一　光學系統光路圖(1G3P)

三、優化過程

本研究設計採用光學設計軟體 Zemax 來進行整體的優化過程，優化的過程中使用的變數有曲率、厚度、材料、二次曲面等等。Zemax 軟體在最佳化(Optimization)的應用上使用阻尼最小二乘法，求得最終數據。這種演算法能夠優化加權目標值組成的評價函數 [2]：

設定優化過程必須考慮以下三項要素 [3]：

步驟一：選擇一個適當的光學系統做為出發點

所謂適當的光學系統很難界定，一般會以軟體所附的標準範例、設計者以前的設計、專利公報等資料中，選擇適當的出發點，做爲優化的開始。起始點的設定影響整體光學系統設計非常的大，必須要多加注意。

步驟二：定義系統上所需要的變數

選定鏡片的曲率半徑、厚度、材質、K 值、非球面係數等做爲設計變數。如果在環境上有特別的要求，例如：使用環境溫度變化大，此時必須將溫度補償因素考慮進去才行。

步驟三：定義績效函數(Merit Function)

績效函數(Merit Function)是使光學系統接近我們指定目標的一種數值表示，理想的績效函數值爲零。定義預設的績效函數和權重值(Weight)是優化最重要的關鍵。

首先採用內建的優化類型均方根值(RMS)及預設的優化資料波前(Wavefront)再加上以中心(Centroid)爲參考點來進行優化，搭配 EFFL 函數將有效焦距目標值訂在系統要求，使用像差函數 SPHA、COMA、ASTI、DIST、AXCL、LACL、FCUR...等來控制軸上及軸外像差，TOTR、MNET、MNCT...等函數做爲鏡頭資料的約束條件，使鏡片形狀易於加工。

表二　光學鏡頭設計之系統資料。

項目	數值	項目	數值
Surfaces	12	Paraxial Image Height	3.3
Stop	1	Paraxial Magnification	-0.009
System Aperture	F/# = 2.8	Entrance Pupil Diameter	2.007
Effective Focal Length	5.62 (in air)	Entrance Pupil Position	0
Back Focal Length	1.5	Exit Pupil Diameter	2.063
Total Track	7.603	Exit Pupil Position	-5.83

表三爲使用 ZEMAX 優化後所得到之鏡片參數值，在物距的設定上以三個位置爲設定基準：(1)無窮遠位置[Infinity] (2) 600 mm (3) 100 mm [Macro 機構近拍使用]，表三所顯示爲當物距設定在 600mm 所得之結果。在正常情形下，手機相機鏡頭的物距會以無窮遠及 600 mm 位置爲設計基準；但是考慮市場需求，光學模組廠普遍會要求鏡頭廠商提供鏡頭時同時提供近拍機構(Macro)，因此在設計時必須加入物距 100mm 位置做爲近拍機構設計的參考依據。

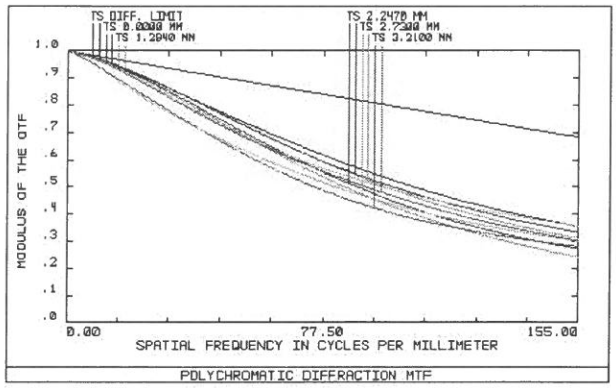
表三　使用 ZEMAX 優化後所得到之鏡片參數值

Surf	Type	Radius	Thickness	Glass	Conic
OBJ	STANDARD	Infinity	600		0
STO	STANDARD	Infinity	0.1		0
2	STANDARD	2.95	0.9	LAL52	0
3	STANDARD	70	0.57		0
4	EVENASPH	-3.60	0.9	ZEONEX	-3.91
5	EVENASPH	-2.02	0.44		-0.71
6	EVENASPH	-1.1	0.98	PC	-0.82
7	EVENASPH	-2.42	0.25		-0.69
8	EVENASPH	2.87	1.54	ZEONEX	-8.56
9	EVENASPH	3.48	1.5		-6.51
10	STANDARD	Infinity	0.4	Cover Glass	0
11	STANDARD	Infinity	0.0098		0
IMA	STANDARD	Infinity			0

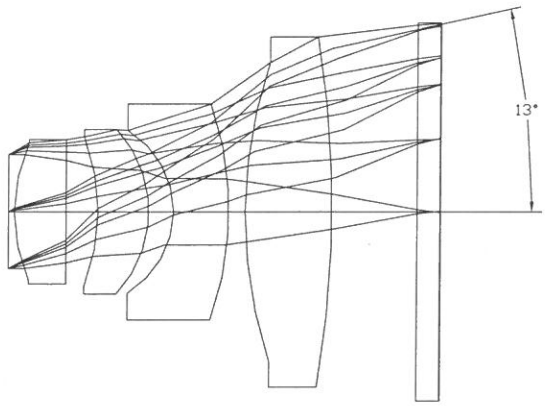
表四　非球面係數參數值

Surface 4	EVENASPH	Surface 5	EVENASPH	Surface 6	EVENASPH
Coeff on r 4	- 0.017	Coeff on r 4	- 0.021	Coeff on r 4	0.056
Coeff on r 6	0.004	Coeff on r 6	0.012	Coeff on r 6	- 0.0002
Coeff on r 8	0.0004	Coeff on r 8	- 0.004	Coeff on r 8	0.002
Coeff on r 10	- 0.0001	Coeff on r 10	0.0007	Coeff on r 10	- 0.002
Coeff on r 12	0	Coeff on r 12	0	Coeff on r 12	0.0002
Surface 7	EVENASPH	Surface 8	EVENASPH	Surface 9	EVENASPH
Coeff on r 4	0.016	Coeff on r 4	- 0.005	Coeff on r 4	- 0.013
Coeff on r 6	0.009	Coeff on r 6	- 0.0003	Coeff on r 6	0.001
Coeff on r 8	- 0.001	Coeff on r 8	0.0002	Coeff on r 8	- 0.0001
Coeff on r 10	0.0001	Coeff on r 10	- 1.96E-05	Coeff on r 10	2.14E-05
Coeff on r 12	- 1.92E-05	Coeff on r 12	- 8.90E-07	Coeff on r 12	- 1.44E-06

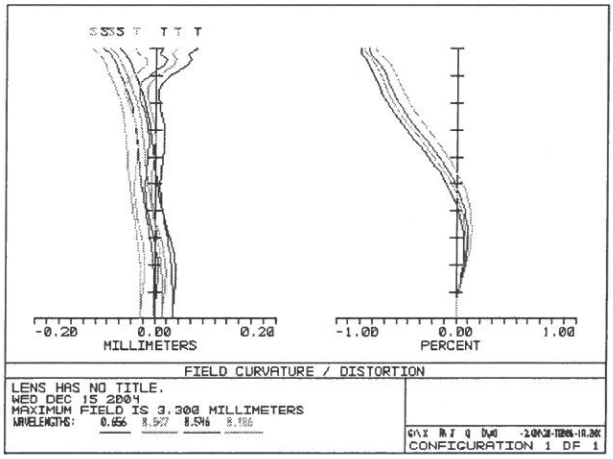
圖二爲 MTF 圖，由圖中可以瞭解中心位置(Field=0)設計值 35%大於系統要求 30%，符合規格。在周邊位置(Field=1.0)設計值 25%大於系統要求 20%，也滿足系統之要求。圖三表示主光線角度，系統要求小於 16° ，使用 Zemax 軟體設計後，所得到的值爲 13° ，符合系統要求。圖四爲場曲及歪曲像差資料。



圖二　MTF 資料



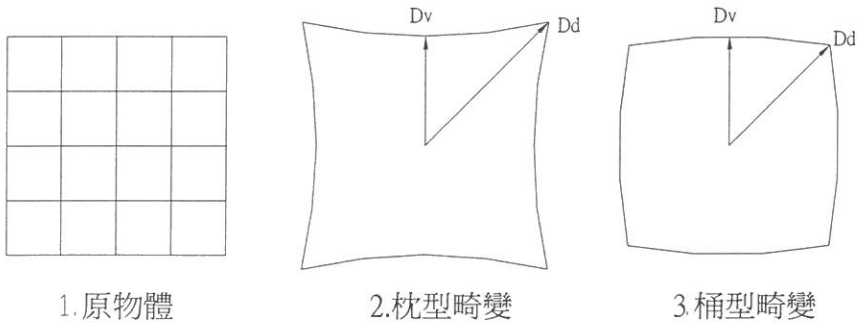
圖三　主光線角度 (Chief Ray Angle)



圖四　爲場曲及歪曲像差

在手機相機鏡頭評價上以 TV Distortion (TV 歪曲像差)取代以往 Distortion (歪曲像差) 評價，計算方式如下：

公式 : $TV = \frac{D_d - D_v}{2}$



圖五　TV 歪曲像差(TV Distortion)示意圖

在設計時考慮時效性，因此採用簡易計算方式其計算如下：

TV-Distortion 計算方式 : (參考表四)

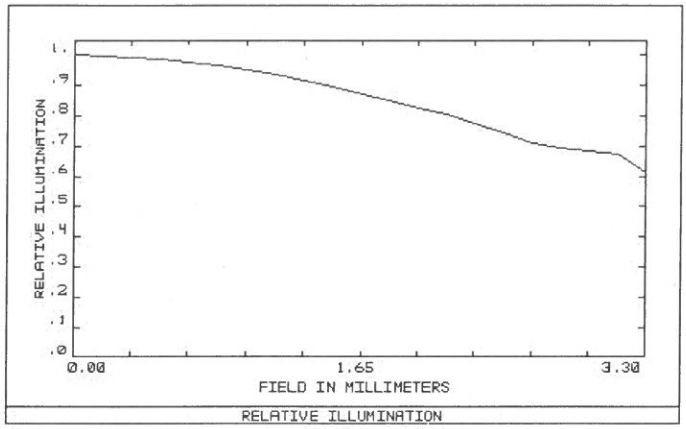
$$\text{TV 歪曲像差} = \frac{(1.0Field - 0.6Field)}{2} = \frac{-0.74-(-0.15)}{2} = -0.445\% \cong -0.45\%$$

由上式資料可知 TV 歪曲像差爲 - 0.48%，滿足系統規格小於 0.5%要求。

表四　歪曲(Distortion)像差值

影像高度 (Image Height)	場 (Field)	歪曲值 (Distortion)
1.98	0.6	- 0.15%
3.3	1.0	- 0.74%

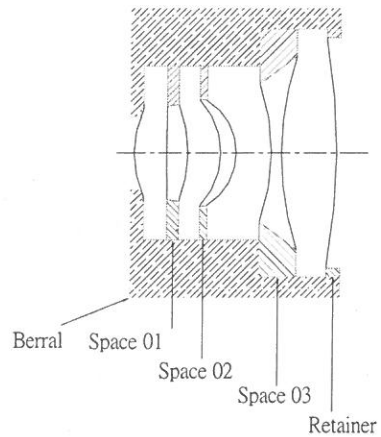
下圖爲相對照度圖型，系統要求大於 55%，使用軟體優化後得到設計值 61%，符合系統要求規格。



圖七　相對照度圖

組立對於光學鏡頭廠是一項很重要的工作，當所有的鏡片、相關部品(EX:Berral、Space、Retainer...等等)完成後，經由組立治工具將相關鏡片組合完成，下圖八爲二百萬畫素鏡頭組立完成品。

由圖八可以瞭解此次組立方式，首先將前三片鏡片外徑設定相同寸法，再搭配墊片(Space01、Space02)一同組立至鏡筒(Berral)，第二部分再將墊片(Space03)及最後一片鏡片組立，最後以壓環(Retainer)將上述所有光學元件固定於鏡筒中。墊片(Space)及壓環(Retainer)可以視其厚度以不同材料來製作(EX:軟性塑膠薄片、塑膠成型、銅片...等等)。鏡筒(Berral)部分，試做時以金屬方式驗證設計是否合乎性能要求，當試做完成後再決定是否以塑膠部品方式大量生產。



圖八　二百萬畫素鏡頭組立完成品。

四、討論與結論

針對塑膠鏡片的缺點如：溫度特性差、抗刮力差，於本設計對策方式如下：

- 溫度特性差——使用較佳之塑料 ex：ZEONEX 材料來改善。
- 抗刮力差 ——將玻璃鏡片放置於第一片，避免塑膠鏡片與外界直接磨擦造成表面傷痕，影響整體影像精度。

在手機相機鏡頭光學設計的領域中，使用四片式結構來設計兩百萬畫素手機相機鏡頭是眾多設計中其中的一種方式。若單純以球面鏡片來製作，4 片球面鏡片型式將無法達到系統要求。若以塑膠非球面鏡片來製作，四片式塑膠鏡片所設計的鏡頭雖可達到系統要求，但是實際生產時，因爲鏡頭的設計敏感度較大，在大量生產時，不良率會偏高。此時就必須針對鏡片精度、偏芯、肉厚、外徑做更嚴謹的控制，而在相對部品上鏡筒(Berral)及間隔環(Space)也必須採取更高品質的設計及製作，此時所投入的人力、物力、財力等等，將會額外的增加數倍費用，在設計時必須多加考慮。

爲了使光學設計的敏感度降低、鏡頭整體低成本下降、達成系統要求的規格以及在生產上能將不良率降低等要求，使用 1 片玻璃球面鏡片及 3 片塑膠非球面鏡片的排列組合來設計。使用此方式設計，成像品質如：MTF、TV Distortion、Chief Ray Angle、Relative Illumination 都能符合系統要求並能兼顧量產，此種設計方式提供手機相機鏡頭在設計上另一項好的選擇。

五、參考文獻

(1) 張弘 (1987)，"幾何光學"，第七章，東華書局。
(2) ZEMAX 光學設計程式使用手冊，第 15 章，2004。
(3) 黃俊明、葉嘉宏，單眼相機鏡頭模組逆向工程設計，Chp. 2-6 2004 年模具技術與論文發表會論文集。
(4) 孫慶成 (2001)，"光電概論"，第一章，全華科技圖書。
(5) W. J. Smith，"Modern optical engineering"Second Edition McGraw-Hill，Inc.，Chp. 13，1990。
(6) G.H.Smith，"Practical Computer-Aided"Willmann-Bell，Inc.，Chp. 15，1998
(7) 黃旭華，"光學優化設計的數值方法探討"，國立中央大學光電所，碩士論文，中華民國 85 年。

(8) 陳昭先，“對給定初階相差目標值之光學薄透鏡組設計方法與應用”，國立交通大學光電所，博士論文，中華民國 86 年。
(9) 鄭伊凱，”非球面透鏡的最佳化與分析 ”， 國立成功大學物理所，碩士論文，中華民國 90 年。
(10) 葉茂勳，“變焦鏡頭高斯光學之設計與分析”，國立交通大學光電所，博士論文，中華民國 86 年。

received October 20, 2004
revised December 01, 2004
accepted December17, 2004

2M Pixel Size Four Pieces Equation Image Lens Design

Rex.Hung Wang Chang Wei Smart Tseng
Da-Yeh University Electrical Engineering delegacy
112 Shan-jeau Rd.,Da-Tsusn,Chang-Hwa

Pin Han
National ChungHsing University Institute of Precision Engineeting
250,Kou Kuang Rd.,Taichung

Abstract

This thesis discusses the design formed by 4 pieces of optical structures, witch offers cell-phone lens design applied to 2 millions pixels CMOS sensor. As a result of the design which is with 1 piece of glass-sphere lens and 3 pieces of plastic lens, compared with the glass- aspheric lens in present market, this mode can not only lower the lens cost but content system request.

Keywords: Lens Design、Optical Design、MTF、TV Distortion

高功率場發射限流二極體之準穩態分析

林銘杰*
輔仁大學物理學系

摘　要

本研究使用自洽方法（self-consistent approach）分析相對論場發射式二極體的準穩態，此二極體採用高穿透率網格式陽極（mesh anode）。場發射過程以滿足量子力學的 Fowler-Nordheim 方程式來描述。冷陰極（cold cathode）的表面特性與游離效應以等效功函數近似來考慮。空間電荷效應以 Poisson 方程來描述並包含相對論修正。高穿透率網格式陽極的游離現象予以忽略。二極體的準穩態呈現一截止電壓，在高壓區，表面電場趨於飽和且只與等效功函數有關。

關鍵詞：場發射、等效功函數、相對論、二極體、Fowler-Nordheim 方程式

一、前　言

在高功率眞空電子元件如虛陰極振盪器（virtual cathode oscillators）[1-3]、Super-Reltrons[4,5]和電子注入器（electron injectors）中[6]，通常使用相對論二極體（relativistic diodes）來加速電子以提供高能電子束。非相對論二極體的基本性質可用一理想模型來描述，即 Child-Langmuir 平面二極體[7]。此模型假設陰極可以提供足夠的電子以達到空間電荷限制（space-charge-limited），在這條件下，陰極表面電場爲零而電流達到最大值。Jory 和 Trivelpiece[8]兩位曾研究一維相對論平面二極體中的電流限制發射現象（current-limited emission）。他們得到一解析解，可描述電流限制及電荷限制現象。然而，電子發射可由很多種機制產生如熱電子發射（thermionic emis-

* Corresponding author. Tel: +886-2-29052585; Fax:+886-2-29021038
E-mail:phys2023@mails.fju.edu.tw

sion）[9]、光電子發射（photoemission）、二次電子發射（secondary emission）、場發射（field emission）及爆炸式發射（explosive emission）等[10]，其中我們有興趣的是場發射。雖然場發射在微觀下可藉由控制場增強因子（field enhancement factor）來提升電子發射量，如增加表面突出物分佈、塗佈氧化層及表面污染等；在巨觀下，電流密度通常無法達到空間電荷限制條件。在此情況下，我們稱之爲場發射限流（field-emission-limited）。近年來，由於高品質電子束的需求與日俱增，網狀陽極（mesh anodes）大量使用於上述之高功率器件[1-6]。研究發現：使用高穿透率網狀陽極可降低電子截獲（low interception）、減少陽極電漿（reducing of anode plasma）及避免二極體短路（ avoiding of diode short）等優點。

在本研究中，我們探討相對論場發射限流二極體（relativistic field-emission-limited diodes ，RFELDs），此二極體使用高穿透率網狀陽極。場發射過程以滿足量子力學的Fowler-Nordheim 方程式來描述[11-18]。冷陰極（cold cathode）的表面特性與游離效應以等效功函數近似（effective work function approximation）來考慮。空間電荷效應以Poisson 方程來描述並包含相對論修正。高穿透率網格式陽極的游離現象予以忽略[19,20]。我們可以由以下的自洽方法（self-consistent approach）得到二極體之準穩態（quasi-stationary state）[21]。

二、理論分析

讓我們來考慮相對論二極體中場發射電子行爲。冷陰極表面的場發射現象可用導帶電子（conduction electrons）之量子穿隧（quantum mechanical tunneling）現象來描述，即引入 Fowler-Nordheim 方程如下[11-18, 22]：

$$J = \frac{AE_s^2}{\phi t^2(y)} \exp\left(\frac{-Bv(y)\phi^{3/2}}{E_s}\right), \tag{1}$$

此處 A 和 B 是 Fowler-Nordheim 常數，而 ϕ 是等效功函數。我們假設它是一個與陰極材料、表面粗糙度及游離效應（ionization effects）有關的常數。陰極表面正向電場 E_s 將由此方程解得並作爲求解 Poission 方程的邊界條件。函數 t（y）和 v（y） 由Spindt 等人引入[22]， 假設如下：

$$t^2(y) = 1.1, \tag{2}$$

$$v(y) = 0.95 - y^2, \tag{3}$$

$$y = 3.79\times10^{-5} E_s^{1/2}/\phi, \tag{4}$$

此處 y 是 Schottky 效應降低功函數能障因子。

爲了分析高能電子在平面二極體中的運動，我們求解 Poisson 方程及相對論能量守恆方程，如下所示：

$$\nabla^2\Phi = -\frac{\rho}{\varepsilon_0}, \tag{5}$$

$$e\Phi = (\gamma - 1)mc^2, \tag{6}$$

此處ρ 是電荷密度，而 γ 是相對論因子（relativistic factor）。Z 方向的電子流密度如下：

$$J = -\rho v. \tag{7}$$

結合（5）-（7）式可得

$$\frac{d^2\gamma}{dz^2} = \frac{eJ}{\varepsilon_0 mc^3\sqrt{1-\gamma^{-2}}}. \tag{8}$$

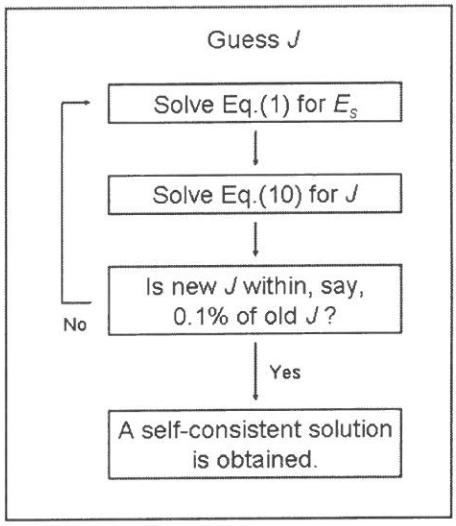
此方程可積分如下：

$$\frac{d\gamma}{dz} = [\frac{2eJ}{\varepsilon_0 mc^3}(\gamma^2-1)^{1/2} + \frac{e^2}{m^2c^4}E_s^2]^{1/2}. \tag{9}$$

積分常數 E_s 是陰極表面電場（$z = 0$ and $\Phi = 0$）。再由陰極表面（$z = 0$）積分到陽極表面（$z = d$），方程式（9）成爲

$$J = \frac{\varepsilon_0 mc^3}{2ed^2}\left\{\int_1^{\gamma_0}\left[(\gamma^2-1)^{1/2} + \frac{\varepsilon_0 e}{2mcJ}E_S^2\right]^{1/2} d\gamma\right\}^2, \tag{10}$$

此處 $\gamma_0 (= 1 + e\Phi_0/mc^2)$ 是電子在陽極時的相對論因子對應於外加電壓Φ_0。實際上，方程式（10）應考慮陽極電漿擴張產生之間隙關閉（gap closure）效應，以（d-$v_c t$）取代 d，此處 v_c 爲關閉（closure）速度[10]。然而，我們考慮系統在時間尺度甚小於 gap closure 發生的時間，此即 $v_c t \ll d$，所以方程式(10)可近似地用來研究二極體的準穩態現象。當 $E_s = 0$ 時，方程式(10)可藉由標準積分表來積分；當 $E_s \neq 0$ 時，需透過變數變換仍可將方程式(10)積分求出。給定一 E_s，電流密度即可由求解方程式(10)得到。



圖一 自洽法求解流程圖

自洽法求解流程如下，如圖一所示，先猜一電流密度 J，表面電場 E_s 由 Fowler-Nordheim 方程求出，作爲 Poission 方程之邊界條件進而解出一新電流密度，再代回 Fowler-Nordheim 方程。反複迭代求解方程式(1)和(10)直到我們達到一穩定的自洽解。

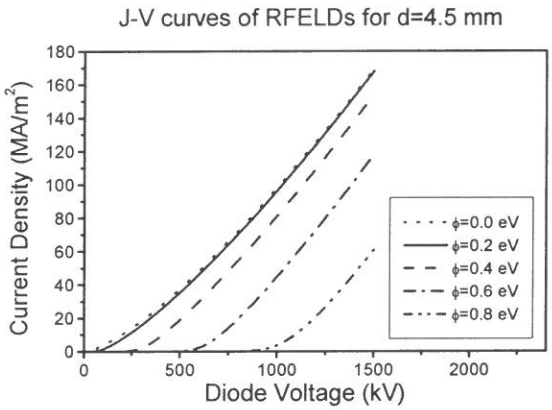
三、結果與討論

我們由以上的自洽法得到了相對論場發射限流二極體之準穩態。圖二是我們計算得到 RFELDs 的 J-V 曲線，對應二極體間隙 d = 4.5 mm，等效功函數ϕ=（0, 0.2, 0.4, 0.6, 0.8） eV。對應ϕ = 0 eV（爆炸式發射）的 J-V 曲線趨近於正確極限，爲空間電荷限制，其他爲場發射限制。我們可以由圖二看出，對應ϕ = 0.2 eV 的曲線非常接近空間電荷限制。

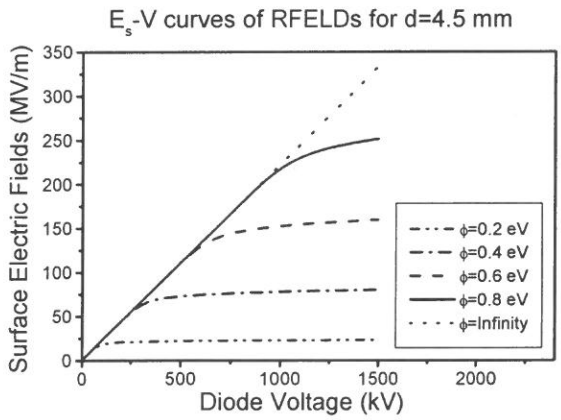
然而，對應於場發射限流的 J-V 曲線皆存在一截止電壓（cutoff voltage）。在截止電壓下，幾乎沒有電流，此與空間電荷限流不同。

圖三爲我們計算得到的 E_s-V 曲線，對應二極體間隙 d = 4.5 mm，等效功函數ϕ=（0.2, 0.4, 0.6, 0.8, Infinity） eV。對應ϕ= Infinity 的曲線趨近於正確極限，爲無發射態（nonemission），表面電場 E_s 與二極體電壓成正比。其他四條曲線在二極體電壓高於截止電壓後，表面電場隨電壓增加的速率開始減緩；在高壓區，表面電場迅速達到一

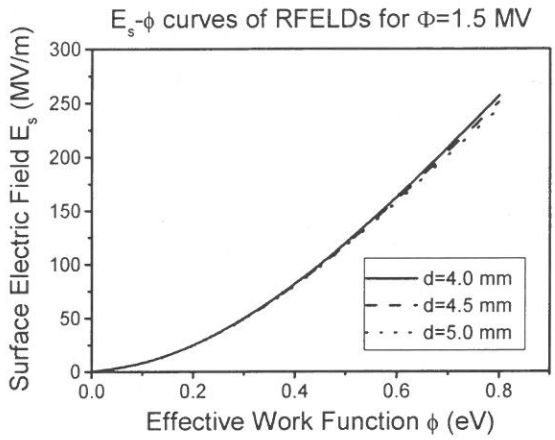
飽和值。在此飽和區，表面電場與外加二極體電壓無關。圖四爲我們計算得到的 E_s-ϕ 曲線，對應二極體電壓Φ = 1.5 MV 且二極體間隙 d =（4.0, 4.5, 5.0） mm。表面電場幾乎由等效功函數決定，與二極體間隙長度無關。所以，我們可將表面電場視爲只依賴等效功函數的函數。



圖二　RFELDs 的 *J-V* 曲線，對應二極體間隙 d = 4.5 mm，等效功函數ϕ=（0, 0.2, 0.4, 0.6, 0.8） eV。



圖三　RFELDs 的 E_s-*V* 曲線，對應二極體間隙 d = 4.5 mm，等效功函數ϕ=（0.2, 0.4, 0.6, 0.8, Infinity） eV。



圖四　RFELDs 的 E_s-ϕ 曲線，對應二極體電壓 Φ = 1.5 MV 且二極體間隙 d =（4.0, 4.5, 5.0）mm。

四、結　語

我們以自洽法研究相對論場發射式二極體。場發射過程以滿足量子力學 Fowler-Nordheim 方程式來描述。冷陰極（cold cathode）的表面特性與游離效應以等效功函數近似來考慮。空間電荷效應以 Poisson 方程來描述並包含相對論修正。高穿透率網格式陽極的游離現象予以忽略。相對論場發射式二極體與相對論空間電荷限流二極體的差別在於前者的準穩態呈現一截止電壓；另外，在高壓區，表面電場趨於飽和且只與等效功函數有關。

作者想感謝 SLAC 研究人員 George Caryotakis， Glenn Scheitrum， Daryl W. Sprehn， 和 Alex Burke 等人的支持與鼓勵；另外，Cecilio Vazquez 和 Rafael J. Gomez 在電腦上的協助在此一併致謝。This work was partially supported by the National Science Council, Taiwan, under Grant No. NSC 93-2112-M-030-007.

參考文獻

(1) C. S. Hwang, M. S. Yang, F. C. Lin, M. W. Wu, and W. S. Hou, Pulse Power Conf., 7th, 951 (1989).

(2) M. W. Wu, T. C. Guung, C. Y. Chen, and C. S. Hwang, Pulse Power Conf., 8th, 815 (1991).

(3) Eun-Ha Choi et al., IEEE Trans. Plasma Sci., vol. 28, no. 6, 2128 (2000).
(4) R. B. Miller, W. F. McCullough, K. T. Lancaster, and C. A. Muehlenweg, IEEE Trans. Plasma Sci., vol. 20, no. 3, 332 (1992).
(5) R. B. Miller, IEEE Trans. Plasma Sci., vol. 26, no. 3, 340 (1998).
(6) G. Mamaev et al., Proc.1997 Part. Accel. Conf., vol. 1, 1263 (1998).
(7) C. D. Child, Phys. Rev. 32, 492 (1911); I. Langmuir, Phys. Rev. 2, 450 (1913).
(8) H. R. Jory and A. W. Trivelpiece, J. Appl. Phys. 40, 3924 (1969).
(9) S. Dushman, Phys. Rev. 21, 623 (1923).
(10) R. B. Miller, An Introduction to the Physics of Intense Charged Particle Beams (Plenum Press, New York, 1982).
(11) R. H. Fowler and L. W. Nordheim, Proc. R. Soc. London, Ser. A 119, 173 (1929).
(12) L. W Nordheim, Proc. R. Soc. London, Ser. A 121, 626 (1928).
(13) L. W. Nordheim, Z. Phys. 30, 177 (1929).
(14) W. Schottky, Z. Phys. 14, 63 (1923).
(15) N. H. Frank and L. A. Young, Phys. Rev. 38, 80 (1931).
(16) E. Guth and C. J. Mullin, Phys. Rev. 61, 339 (1942).
(17) D. V. Gogate and D. S. Kothari, Phys. Rev. 61, 349 (1942).
(18) S. Gasiorowicz, Quantum Physics, 2nd edition (Wiley, New York, 1996).
(19) S. A. Goldstein and R. Lee, Phys. Rev. Lett. 35, 1079 (1975).
(20) E. H. Choi, H. M. Shin, and D. I. Choi, J. Appl. Phys. 61, 2160 (1986).
(21) M. C. Lin and D. S. Chuu, Appl. Phys. Lett. 80, 4262 (2002).
(22) C. A. Spindt, I. Brodie, L. Humphrey, and E. R. Westerberg, J. Appl. Phys. 47, 5248 (1976).

received October 20, 2004
revised December 01, 2004
accepted December17, 2004

Study on Quasi-stationary States of a Relativistic Field-emission-limited Diode

Ming-Chieh Lin

Department of Physics, Fu Jen Catholic University
Hsinchuang, Taipei Hsien 24205, Taiwan, ROC.

Abstract

A relativistic field-emission-limited diode employing a high-transparency mesh anode is investigated via a self-consistent approach. The field emission process is described quantum mechanically by the Fowler–Nordheim equation. The cathode plasma and surface properties are considered within the framework of the effective work function approximation. Space-charge effects are described by Poisson's equation including relativistic effects. Ionization effects at the high-transparency mesh anode are ignored. The numerical calculations are carried out on a time scale much shorter than the emergence of the gap closure. The quasi-stationary state of the diode exhibits a cutoff voltage. The electric field on the cathode surface is found to be saturated in the high-voltage regime and determined by the effective work function only.

Keywords : Field emission, Effective work function, Relativistic diode, Fowler-Nordheim equation.

Extended C Preprocessor Directives for Device Driver Programming

Kuan Jen Lin* and Jian Lung Chen
Department of Electronic Engineering
Fu Jen Catholic University
Taiwan

Abstract

Writing device driver has always been tedious and error-prone. Traditionally, C programmers exploit preprocessor directives to facilitate the driver development. In this paper, we follow this programming style and extend preprocessor directives to improve the development. The extension facilitates the accesses and manipulations of individual data fields in IO registers. It makes the C driver code more readable and concise. Furthermore, our complier provides type-checking capability for low level device operations, which is not supported by the original C compiler. Current assessment for embedded Linux environment shows favorable results.

Keywords: Device driver, Programming language, Preprocessor, Linux OS, Software development.

1. INTRODUCTION

Device drivers are a set of software components inside the operating system which form a software layer to talk to peripheral hardware devices. To the user they provide an abstraction and adaptation of the interaction with devices and isolate the I/O interactions from the rest of the system. In Unix-like OS, the devices are treated as *device files* and application programs can communicate with them via file operations, which can also be viewed as Appli-

* Corresponding author. Tel.:+886-2-29052159
E-mail address : kjlin@mails.fju.edu.tw

cation Programming Interface (API). Fig. 1 shows such a file-IO scheme for Linux OS. As a file-IO operation is called, OS kernel will identify the device and pass the command to its corresponding driver to activate the service routine like *read*( ). Linux runs the driver in kernel mode, which means that the process calling the driver cannot be arbitrarily suspend and substituted with another process, unless it voluntarily relinquishes the CPU. For slower devices, the driver generally suspends itself and forces the process to relinquish the CPU. When the data is available, an interrupt signal from the device will cause the OS to activate the *interrupt*( ) routine in the driver, which then resumes the process.

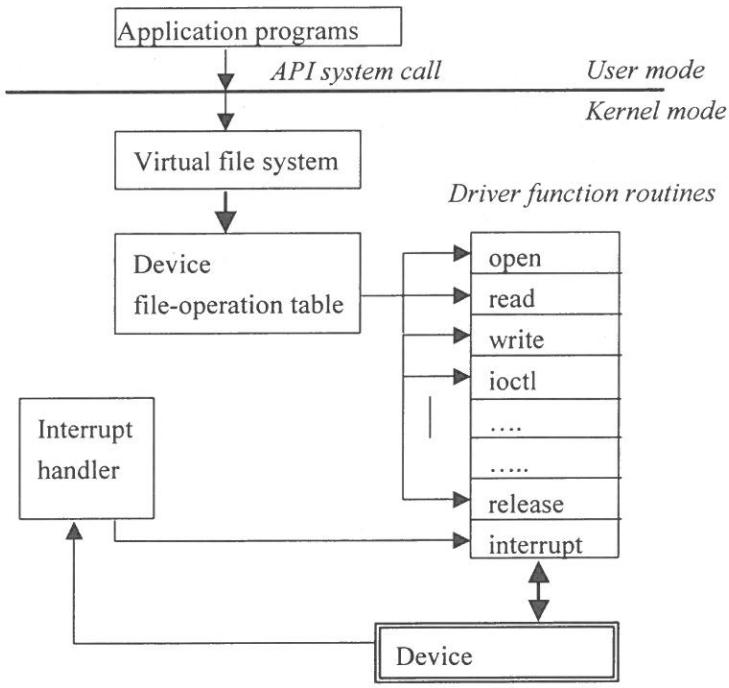


Fig. 1. Linux device driver model.

Writing device driver has always been considered as a tedious and error-prone task. The followings summarize various reasons why developing device drivers is difficult.

(1) **Kernel-mode programming**: Linux runs device drivers in the kernel mode. Only the functions that are exported by the kernel can be called. For example, an application program can call *printf*() function, while a driver must use *printk*(). Furthermore, a driver must

avoid *namespace pollution* within the whole kernel space. Note that a segmentation fault in a driver will hang up the whole system. All these make the driver hard to debug. As reported in [2], device drivers have been noted as a major source of faults in operating system codes.

(2) **Hardware knowledge**: Device drivers lie between the applications and the physical peripheral devices. A driver programmer must understand the communication mechanism, physical allocation of data (register layout) and even electronics, based on device documents. This requires an expertise in two different areas. On the one hand, programming skill is required. On the other hand, the capability of understanding low-level interface of the hardware and its internal behavior is also needed. As a matter of fact, few driver programmers can be considered expert in both domains.

(3) **Low level operations in languages**: Writing device drivers involves many low-level instructions such as direct memory access and bit-operation. It has been found that bit operations can represent up to 30% driver codes [9]. Traditionally, the C language is the most commonly used to write device drivers due to its support of such kinds of instructions. However, such operations in C are not checked for type-correctness. Moreover, they are fairly unreadable.

(4) **Portability**: The interface between device driver and operating system varies widely with the operating system. Nevertheless, different device vendors, even for the same kind of device, may have different programming interface. Hence, device drivers usually have to be largely rewritten when ported to another OS or vendor-dependent hardware.

To facilitate the development process and improve the reliability, a variety of approaches have been suggested from both industrial and academic communities. Current commercial assistant tools like Jungo's WinDriver [16] and Bsquare's WinDk [17] provide a graphical user interface for specifying the main features of a driver. They can automatically generate a code skeleton which is comprised of coarse-grained functions and libraries which wrap kernel functions. Most of research works in literature attempt to automatically generate fine-grained driver code from Domain-Specific Languages (DSLs). A DSL is a programming language tailored for a specific application and provides more expressive power over the application domain. The language GAL developed by Thibault et al. [12] is designed for specifying X Windows video driver. Though the driver code is reduced about 90%, GAL covers a very restricted domain. The language Devil developed by Merillon et al. [9] is designed for more

general classes of devices. Its specification is complied into a set of C procedures (or in-line function) for accessing registers and manipulating data. These procedures are called in a traditional C driver code and prevent programmers from directly dealing with low-level codes. The languages Dveil+ [15] and NDL [4] both follow the Devil and attempt to propose a complete language to replace General Programming Languages (GPL) like C for writing device drivers. Although their languages have higher level abstraction of device behaviors, they incur less flexible programming capability than GPLs. Interface HW/SW co-synthesis works [3, 10, 14] address the automatic generation of driver codes for dedicated device behaviors. P. Chou et al. [3] propose a system that can automatically generate device drivers from a high level specification and glue logic to connect hardware components. The generated software driver is not in the context of an OS, but rather to emulate the signaling of bus interface in software. O'Nil [10] proposes a Yacc-like language Program to model the communication protocol between the device and the software. FSMs are extracted from the specification to enable code synthesis. Shaojie Wang et al. [14] use the extended FSM to model device behavior and the codes responsible for the extended FSM are automatically synthesized.

We have presented a design framework to facilitate driver development in our earlier publication [7, 8]. The framework bases a structural two-stage approach and provides assistant tools and language for design. In the first stage, the code skeleton of a driver can be automatically generated from user-specified configurations. In the second stage, a user can continue to finish the driver using our proposed Extended C Preprocessor (ECP) directives and C language. This paper will elaborate on the design principle and the grammars of the ECP directives. The extension aims at simplifying the accesses and manipulations of individual data fields in IO registers. It makes the C driver code more readable and concise. Furthermore, the ECP complier performs type-checking for low level device operations, which is not supported by the original C compiler. Current assessment for embedded Linux environments will show favorable results.

The reset of the paper is organized in the following way. The framework will be overviewed in the next section. Section 3 will introduce the concept of device variables, which allows us to access individual data fields in IO registers conceptually as to access general variables in programming languages. The ECP directives are designed for the accesses and manipulations of device variables. Section 4 will describe the grammar of ECP directives. As-

sessment will be shown in Section 5. Final section concludes the paper.

2. DESIGN FRAMEWORK

Fig. 2 shows our design framework that defines a design flow and supports assistant tools for Linux driver programming. It consists of a two stage design process. The inputs are user-specified configurations for API calls as well as system resources (such as DMA channel and interrupt vector) associated with the device. API is an abstraction of functions provided by the peripheral device, through that operation system (on behalf of application programs) can communicate with devices. Linux OS has defined a standard API function prototype for each API routine [11]. We prepare a set of parameterized code templates for these routines. The templates are developed on a driver model which interprets device behavior in terms of control flow and data flow.

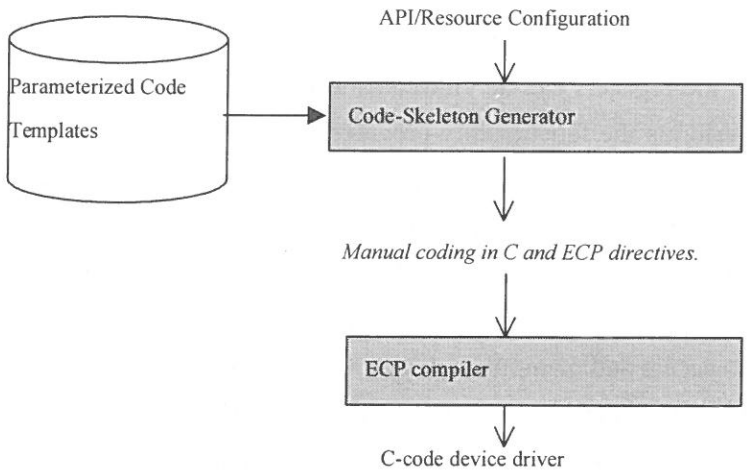


Fig. 2. The design framework for Linux driver

The parameters are used to define what a subset of API functions provided and what operation mode used in both control and data flows. The program skeleton generator parses the configuration file to extract the parameter values. Then it builds the driver code skeleton from the set of code templates. Basically, a skeleton is comprised of a set of the API routines, to-

gether with initialization and interrupt routines. All the routines contain not only the function interface but also the required statements to serve operations in both control and data flows. The details can be found in [1].

In the second stage, one can start from the generated code skeleton and continue to complete the driver with C language and our proposed ECP directives. It has always been a popular programming style to write device driver with the help of C preprocessor [6]. Our extension is expressive uniquely over the specific features of device drivers, which will be elaborated in later sections. The program ECP compiler translates preprocessor statements to C codes. The final output is a device driver in C language, which then can be fed into gcc to get executable codes.

3. DEVICE VARIABLES

The main tasks of a device driver are to configure the device operation mode and observe its status to manage the data transfer. The real physical part interacting with the driver is a dedicated controller (an IC or IP) that manages the peripheral device, as shown in Fig. 3. In the IO controller, the registers can be viewed as a programmable interface to the driver and the kernel part performs the functionality provided by the device. Through the registers, the driver configures the functions of the device and observes its statuses. For example, a typical UART controller contains a set of registers for setting data format and transfer rate (baud rate), as shown in Fig. 4. The data format is determined by the values of PMD, STB and WL, each of which occupies a bit range within the register ULCON0. They are referred to as device variables. To access and manipulate device variables involve low-level instructions such as direct memory access and bit-operation. Such operations in C are not checked for type-correctness and are fairly unreadable. ECP provides more readable and concise expressions to allow them to be read or written a5 any variable in C. For example, if we want to set the number of stop bits per frame to be 2 (i.e. STB=1) while keeping other variables unchanged, a typical sequence of C codes might be written as follows:

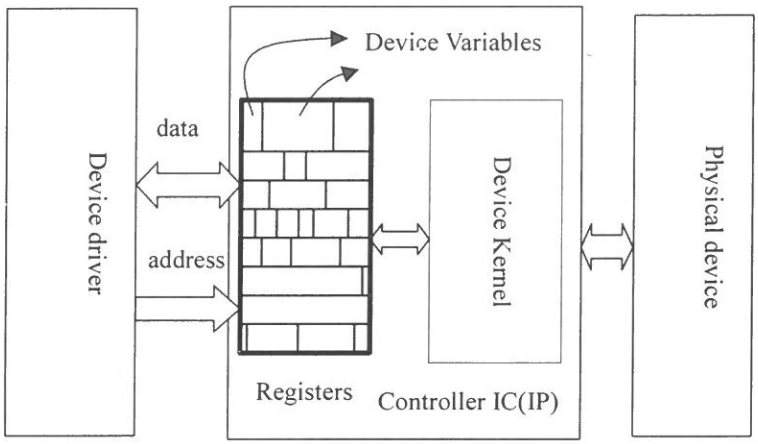


Fig. 3. A peripheral device model.

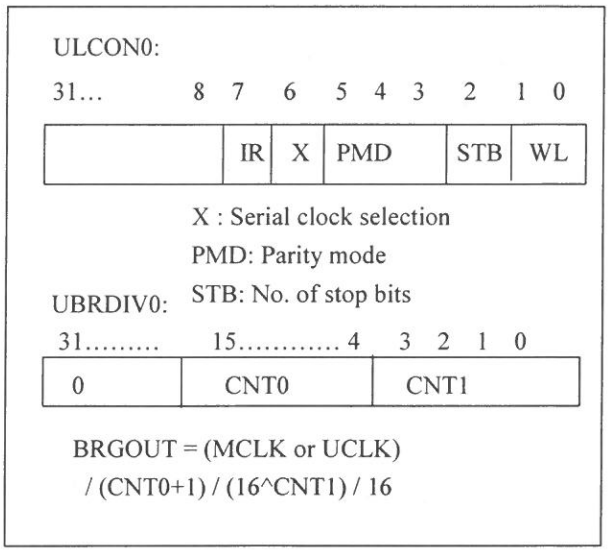


Fig. 4. UART registers (A real case from Samsung 3C4510B SOC).

```
temp = *(ULCON0);
temp = temp & 0xFFFFFFFB | (1 <<2);
*(ULCON0) = temp;
```

Using ECP directives, you can write a simple assignment in C to get equivalent result.

```
STB =1;
```

Group Variable

To set a function or get a status often involves several relevant device variables. We call the set of variables as *group variable*. For example, if we want to set a data format for UART communication, it is often to set PMD, STB and WL simultaneously. We can let a group variable "UARTFRAME" be composed of PMD, STB and WL. The ECP directives allow us to use the following statement to set the group.

```
#set(UARTFRAME [ODD, TwoS, Bit8])
```

And use the following statement to get the group.

```
#get(UARTFRAME[ x, y, z])
```

Concatenated Variable

A device variable could comprise several fragments residing in different registers. For example, the setting of baud rate is determined by three values: the X in ULCON0 register (it selects external clock MCLK or internal clock UCLK as a reference clock), the CNT1 and CNT0 in UBRDIV0 register. The calculation is shown in Fig. 4. If we concatenate the three variables to be a new variable BaudRate, a simple assignment can be used to set its vale as follows:

```
BaudRate  = U2400;
```

4. ECP DIRECTIVES

This section will describe the main portion of ECP grammar and give statement examples to show its application. Fig. 5 summarizes the grammar and Fig. 6 gives statements in ECP directives to specify a UART device shown in Fig. 4.

The syntax for device declaration begins with the nonterminal *device-statement*, as shown in Fig. 5. The directive **#dev** is used to define the name of a device, the bit-width of its registers and its base address (referring to line 1 in Fig. 6). The ending of a device declaration is with directive **#enddev**. The syntax for register declaration begins with the nonterminal *reg_statement*. The directive **#reg** (line 2) is used to define the name, RW attribute, the address offset and an optional bank-register ID. The RW attribute specifies the register to be read-only, write-only or read-write. The offset plus the device's based address equals the physical address of the register. In some device, several sets of registers are mapped to the same memory addresses. Each set usually is called a bank. A bank register is used to select

driver →*device_statement* (*device_statement* | *getfun* | *setfun* | *flush*)*

devic_statement → ***#dev*** *dev_line dev_content* ***#enddev***

dev_line → *dev_id* (*reg_bitwidth*) *iobase_addr*

dev_content → *(reg_statement)*+ *(convar_statement* | *gupvar_statement)**

reg_statement → ***#reg*** *reg_line reg_content* ***#endreg***

reg_line →*reg_id* (**R**|**W**|**RW**) *offset (bank_id*[*index*])?

reg_content → *(var_statement)*+

var_statement → ***#var*** *var_line ((var_content)*+ ***#endvar*** *)?*

var_line →*(-* | *var_id) reg_id* [*bit_position*]

var_content → *val_id* := *num*

convar_statement → ***#convar*** *convar_line ((convar_content)*+ ***#endconvar)***?

convar_line → *convar_id* (*covar_member (,convar_member)**)

convar_member → *var_id* | *convar_id* | *reg_id* [*bit_position*]

convar_content → *conval_id* := *(num* | *val_id* | *conval_id)(, (num* | *val_id* | *conval_id))**

gupvar_statement → ***#gupvar*** *gupvar_id* (*gupvar_membert (,gupvar_member)**)

gupvar_member → *var_id* | *convar_id*

getfun → ***#get*** (*gupvar_id* [*c_var_id (, c_var_id)**])

setfun → ***#set*** (*gupvar_id* [*gupvar_para (, gupvar_para)**])

flush → ***#flush***

Fig. 5. The main portion of ECP grammar.

which set to be used. The UART device shown here does not use such addressing. The lines between #**reg** and #**endreg** define all the device variables within the register.

The syntax for a device variable declaration begins with the nonterminal *var_statement*. The directive #**var** is used to define a variable. As shown in line 9 of Fig. 6, the statement defines "PMD" to represent the value in bit5~bit3 in register ULCON0. Following the declaration, an enumeration of identifiers that represent all the permissible values of the variable can be defined (line 10~15), ending with directive #**endvar**. Such a structure can be thought of as enumerated type in C. Only these values in the type can be assigned to the device variable. Our ECP compiler performs a type-checking to reject other values. This facility provides safety type-checking, that is not supported in C language. If the type definition is not used, the #**endvar** is not needed. The variable then can accept any positive integer less than 2^N,

```
1   #dev        UART(32) 0x03FF0000
2   #reg        ULCON0(RW)              0xD000
3   #var        -                       ULCON0[31:8]
4   #var                IR                      ULCON0[7]
5               Disable  := 0
6               InfraRed := 1
7   #endvar
8   #var                X                       ULCON0[6]
9   #var                PMD             ULCON0[5:3]
10              Disable:= '0**'
11              Odd         := '100'
12              Even := '101'
13              As1         := '110'
14              As0         := '111'
15  #endvar
16  #var        STB                     ULCON0[2]
17              OneS        := 0
18              TwoS        := 1
19  #endvar
20  #var        WL                      ULCON0[1:0]
21              Bit5        := 0
22              Bit6        := 1
23              Bit7        := 2
24              Bit8        := 3
25  #endvar
26  #endreg
27  #reg        UBRDIV0(RW)             0xD014
28
29  #var        CNT0                    UBRDIV0[15:4]
30  #var        CNT1                    UBRDIV0[3:0]
31  #endreg
32  #convar BaudRate(X,CNT0,CNT1)
33              U1200 := 1,1735,1
34              U2400 := 1,867,1
35              U57600 := 1,35,0
36  #endconvar
37  #gupvar UARTFRAME(PMD, STB, WL)
38  ..................
39  #enddev
```

Fig. 6. A portion of the device declaration for the UART in Fig. 4.

where N is its bit length. Unused bits in a register can be declared with symbol "-" as shown in line 3, that defines unused bits 31~8 in register ULCON0. After a variable declaration, we

can write a simple assignment to access the variable. The following statement assigns "TwoS" to variable STB to configure the UART to use two stop bits in a date frame:

STB = TwoS;

ECP compiler will convert the statement into the following:

*(0x03ffd000) = *(0x03ffd000) & 0xFFFFFFFB | (1 <<2);

The part between #**convar** and #**endconvar** (line 32~36) defines a concatenated variable *Baudrate* and its enumerated data type, only a small portion shown here. We can use the following simple expression to set value.

BaudRate = U2400;

ECP complier converts the statement into:

**(0x03ffd000) = *(0x03ffd000) & 0xFFFFFFFB | (1 <<6);*

**(0x03ffd014) =867 << 4 | 1 ;*

The directive #**gupvar** is used to define a group variable. As shown in line 37, the group *UARTFRAME* is composed of *PMD*, *STB* and *WL*. Then we can use the following statement to set their values:

#set(UARTFRAME [ODD, TwoS, Bit8])

This statement is converted into:

**(0x03ffd000)= *(0x03ffd0000) & 0xFFFFFFFC0 | ((4<<3) | (1 << 2) | 3);*

Reading the group at a time is specified as follows:

#get(UARTFRAME[x, y, z])

This statement is converted into:

*temp = *(0x03ffd000);*

x = (temp & 0x38) >> 3;

y = (temp & 0x04) >> 2;

z = (temp & 0x03);

The statements accessing device variables in the same register can be grouped into one instruction if their access orders do not affect the device behavior. If the order must be obeyed, the directive **#flush** can be used to tell the compiler that all the statements preceding it must be completed before doing succeeding statements.

5. ASSESSMENT

The ECP compiler has been written in C using Flex and Berkeley Yacc to generate a front-end parser. To assess the utility of ECP directives, we have implemented four drivers, as listed in Table 1. The first two are designed for a Samsung 3C4510B-based platform [18], which runs under uClinux [19]. The UART_DMA is a UART device with DMA transfer mode. The LCD is a 16 character by 2 line device. The last two are designed for an X-Hyper250B platform [20], which runs under Linux 2.4.18 kernel. The RTC (Real Time Clock) device is used to configure a clock source with a wide range of frequencies. The I2C device enables the processor to communicate with I2C peripherals. Table 1 shows the result in terms of lines of code. All original codes are designed in C. The number of lines of C preprocessor (CPP) is counted for only the part dealing with the target devices. The number of lines of configuration file used to specify API and system resource is given in "CF" column. The number of lines in ECP directives is given in "ECP" column. The "C" column in ours indicates how many lines of C code are required to finish the driver, not including the code produced by the skeleton generator and ECP compiler. The required numbers in average are significantly shorter than the original ones. Currently, we have not developed optimization technique for performance-sensitive drivers. However, examples in current assessment do not sacrifice performance as compared to original ones.

Table 1. Code comparison.

Device	Original		Ours		
	C	CPP	CF	ECP	C
UART_DMA	43	19	6	220	2
LCD	107	29	5	64	65
RTC	424	26	23	24	317
I2C	666	31	14	90	617

6. CONCLUSION

Writing device driver has always been tedious and error-prone. In this paper, we extend preprocessor directives to improve the driver development. The extension is based on the concept of device variables and facilitates the accesses and manipulations of them. It makes the C driver code more readable and concise. The ECP compiler has been written in C using

Flex and Berkeley Yacc to generate a front-end parser. It performs type-checking for accessing device variables, which is not supported by the original C compiler. Current assessment for embedded Linux environment shows favorable results.

REFERENCES

(1) J. L. Chen, "Assistant Tools and Language for Device Driver Development," Ms. Thesis, Fu Jen Catholic University, 2004.

(2) A. Chou et al., "An Empirical Study of Operation System Errors," Proceedings of the 18th ACM Symposium on Operation System Principles, pp. 73-88, 2001.

(3) P. Chou, R. B. Ortega and G. Borriello, "Interface Co-synthesis Techniques for Embedded Systems," Proceedings of IEEE/ACM International Conference on Computer-Aided Design, pp. 280-287, 1995.

(4) C. L. Conway and S. A. Edwards, "NDL: A Domain-Specific Language for Device Drivers," Proceedings of 2004 ACM SIGPLAN/SIGBED conference on Languages, compilers, and tools, pp. 30-36, 2004.

(5) A. Deursen et al., "Domain-Specific Languages: An Annotated Bibliography," ACM SIGSOFT Software Engineering Notes, pp. 26-36, June 2000.

(6) A. Mansky, "Using The C Preprocessor For Device Control," C/C++ Users Journal, Dec. 1990.

(7) K. J. Lin, S. W. Chen and J. L. Chen "A Design Framework for Embedded Linux Drivers," National Computer Symposium, Taiwan, pp. 1663-1668, 2003.

(8) K. J. Lin, J. L. Chen and C. H. Huang, "An Extension of C Preprocessor Directives for Device Programming," International Computer Symposium, pp. 1279-1284, Taiwan, 2004.

(9) F. Merillon, L. Reveillere, C. Consel, R. Marlet, and G. Muller, "A DSL Approach to Improve Productivity and Safety in Device Drivers Development," Proceedings of the 15th International Conference on Automated Software Engineering, pp. 101-109, 2000.

(10) M. O'Nils and A. Jantsch , "Operating system sensitive device driver synthesis from implementation independent protocol specification," Proceedings of Design, Automation and Test in Europe Conference and Exhibition, pp. 563-567, 1999.

(11) A. Rubini and J. Corbet, Linux Device Driver, 2nd Edition O'Reilly, 2001.

(12) S. A. Thibault, R. Marlet and C. Consel, "Domain-Specific Language: From Design to Implementation Application to Video Device Drivers Generation," IEEE Tran. On Software Engineering, pp. 363-377, May/June 1999.

(13) E. Tuggle, "Writing Device Drivers," Embedded Systems Programming, pp. 42-65, Jan 1993.

(14) Shaojie Wang, Malik S., Bergamaschi, R.A. "Modeling and Integration of Peripheral Devices in Embedded Systems," Deing, Automation and Test in Europe Conference and Exhibition, pp.136-141, 2003.

(15) Q. L. Zhang, M. Y. Zhu and S. Y. Chen, "Automatic Generation of Device Drivers," ACM SIGPLAN Notices, pp.60-69, June 2003.

(16) Jungo Ltd, WinDriver V5 Users Guide, URL: http://www.jungo.com

(17) Bsquare, WinDk Users Manual, URL: http://www.bsquare.com.

(18) http://www.wiscore.com.tw

(19) http://www.uclinux.org.

(20) http://www.hybus.net/en/

received October 20, 2004
revised December 08, 2004
accepted December15, 2004

擴充 C 語言前置處理命令以輔助撰寫驅動程式

林寬仁* 陳建隆
輔仁大學電子工程系

摘 要

驅動程式是介於應用軟體與實際裝置之間的軟體，開發驅動程式需要核心程式及硬體裝置相關知識，一直是件冗長乏味且易於發生錯誤的工作。傳統上，C 語言程式工程師使用前置處理命令來簡化其開發工作。此篇論文延續此程式風格，提出擴展 C 前置處理器(Extended C Preprocessor (ECP))語法來進一步改善其開發。其主要目的係簡化暫存器內個別資料欄位之存取與運算。前置處理語言的擴展使得程式較易閱讀與精簡，並且其編譯器針對 IO 裝置動作，提供原來 C 編譯器並未支援之型態檢查能力。此做法已針對嵌入式 Linux 環境進行使用評估試驗，獲致令人滿意的成果。

關鍵詞：驅動程式、前置處理器、程式語言、Linux 作業系統、軟體開發

從音樂資料中找出近似重複樣型的 curve-based 演算法[1]

劉益助 徐嘉連[2]
輔仁大學資訊工程學系

摘 要

從一首歌曲裡找尋「近似重複樣型（approximate repeating patterns）」，是近幾年音樂資訊查詢（music information retrieval）領域裡的重要研究目標。

在這篇論文裡，我們提出一個 curve-based method，能從一個音樂物件（music object）裡快速地找出符合音樂特性的「重要而近似的重複樣型（nontrivial and approximate repeating patterns）」。我們用 sliding window 的方式將音程序列（interval sequence）分割多段，再對每一片段作離散餘弦轉換（DCT）運算把它對映（mapping）到低維度的特徵向量空間（feature space）中的點。我們把 feature space 中距離相近的點逐步組成（self-joining）較長的軌跡（trail），相似的片段會對映成相近軌跡（similar trails），並視爲一近似重複樣型的答案候選者（candidate）。我們的方法能留下較少的 candidates，再做詳細的驗證以找出 nontrivial and approximate repeating patterns。最後，我們做實驗展示在不同的參數條件下 curve-based method 的效果。

關鍵詞：重要而近似的重複樣型（nontrivial and approximate repeating patterns），相近軌跡（similar trails），離散餘弦轉換（DCT）。

---

[1] 本論文研究爲國科會補助之研究成果，計畫編號爲 NSC-93-2213-E-030-005.
[2] Corresponding author. Tel: +886-2-29053817; fax: +886-2-29023550
E-mail: alien@csie.fju.edu.tw

1.序　論

擷取重要的音樂內涵資訊作爲特徵值，一直是音樂資訊查詢領域裡的研究重心，找歌曲裡的 repeating patterns 是近幾年來很重要的研究題目，並被視爲是一首歌裡最重要的特徵值之一。

1.1.主題與近似的重複樣型的意義

爲了加速音樂查詢與作有效的音樂分類，我們都需要擷取歌曲裡重要的音樂特徵值來儲存或做分析。在古典音樂裡，「主題（theme）」是最能呈現一首歌曲特色的一段有意義的旋律（melody），它不只是呆板地一成不變的重複，作曲家經常依據一、兩個主題旋律來發展整首樂曲，再融入一些變奏技巧，使主題有變化而不完全一致地「重現」在後段的歌曲中，稱之爲變奏（variation），如 Figure 1.1。變奏是主題的一個重現（occurrence），它雖然大致上保有主題旋律的主要結構，但是會在旋律、節奏上作變化，因此需要發展一個快速而有效的方法分辨主題與每一個變奏的相似關係。我們根據主題會近似地重現的特性，擷取音樂物件中所有的 approximate repeating patterns，我們也認爲除了能快速地找出 approximate repeating patterns 之外，從中挑選較有意義、較具代表性的 patterns 是另一個更重要的目標。

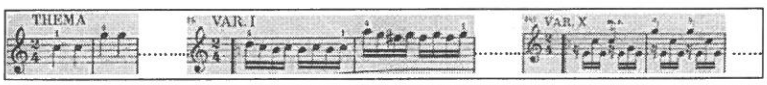


Fig. 1.1.節錄自〈莫札特：小星星主題變奏曲(KV 265)〉之主題與第 1、10 變奏 (excerpt from theme, var.1, and var.10 of ZWÖLF VARIATIONEN by Wolfgang Amadeus Mozart (A.D. 1756~1791), KV 265)

1.2.相關研究

樂曲片段會重複（repetition）的特性被認爲是最有意義的音樂特徵值之一，有一些研究據此意義從 midi 或樂譜裡擷取音高序列（pitch sequence），利用一致性的（exact）或相似性的（approximate）字串比對方式以找出音樂物件中的重複樣型（repeating pattern）或近似重複樣型（approximate repeating pattern）：

Hsu [6]用一 correlative-matrix 找出所有 nontrivial and exact repeating patterns。並且提出 string-join 的方法重複地 join 較短的 repeating patterns 以快速地找到最長的 exact repeating patterns。

Hsu [7]接續之前的研究，提出 pattern 與 occurrence 之間會有 longer_length、shorter_length、equal_length 三種近似的方式，並用 pattern-join 的方式找出符合 longer_length 這個近似方式的 approximate repeating patterns。

Medina [13]用 Self-similarity matrix 找尋所有出現多次的 repeating patterns。這個方法會找出重複的 patterns，所以最後需要在 patterns 之間作最後檢查，此外，也移除那些是長度較長之 patterns 的子字串的 patterns。Richard 認爲大部分的歌曲其主題的變奏也時常會一致地重現，所以這個方法只找出 exact repeating patterns，再用兩 patterns 之間連續的 overlapping 個數判斷是否互爲主題與變奏的關係，但是對於節奏變化較大的變奏，這個方法無法找出。

Pienimäki [15]也用 join 的方式以長度 k 的 repeating patterns 組成長度 k+1 的 repeating patterns，並根據 pattern 的長度、次數、對比程度做排序進而挑選較重要的 patterns。Pienimäki 的方法可以從 polyphonic data 裡找 repeating patterns，他把所有的音符依時間、音樂聲部排先後順序，允許 pattern 裡有 gap，這使跨音樂聲部的 patterns 有機會被找出。

Lartillot [10]認爲一段一段短的樂句彼此在音高與節奏上有鮮明的對比，他用 interval 與 duration 先對歌曲做分段，並用 interval 與 duration 訂出 formula 衡量任意兩段樂句之間的相似度，藉以找出每一個 pattern 的所有重現（occurrence），也把 overlap 的兩 patterns join 成一較長的 pattern。對於兩段較長的樂句，若前半部分很相似但後半部差異大，Lartillot 認爲人類的聽覺感知仍然會感覺它們相似，因此他允許一個 pattern 在每一次的 occurrence 裡可以有較大的音程差異。

Kadota [8]認爲以 note-based 方式擷取音高序列，往往用 dynamic programming 的方法來比對 note sequences 之間的相似度，但可能使兩段很不相似的 sequences 卻發生 match 的情形。他提出一 frame-based 的比對方法，在他的實驗裡把 3 首歌所有的主題與變奏預先擷取出來且混合在一起，測試每段變奏是否能正確的被分類到各自所屬的主題。Kadota 認爲 frame-based 方式有較準確的結果。

這篇論文的內容組織如下：第 2 章簡介主題的變奏方式與 note-based 找尋樣型的方法，並說明音符的拍長規則在變奏裡的重要意義。第 3 章描述問題、名詞定義，也說明 frame-based 取樣方式。在第 4 章說明 curve-based method 的主要架構，也介紹離散餘弦轉換的特性，以及運用它把音程序列對映（mapping）到 feature space 的概念。此外也提出改進方式以篩選出更少而較重要的 patterns。第 5 章展示我們的實驗結果。第 6 章是本篇論文的結論與我們未來研究方向。

2.變奏的方式與 Note-based 找尋樣型的方法

2.1.主題的旋律變奏方式

旋律變奏是歌曲裡很常見的變奏技巧，作曲家通常會維持和聲的結構，加入裝飾音或倚音以修飾主題且改變主題原來的節奏，但是仍保留主題旋律的音高結構。以莫札特的小星星主題變奏曲爲例，如 Figure 2.1，主題的第 1、2 小節的每一拍都是 4 分音符，在第 1 變奏的時候每一拍改由 4 個 16 分音符組成並且產生節奏上的變化，但是仍然保有主題的四個音與演進順序：$C_5 - C_5 - G_5 - G_5$ 。

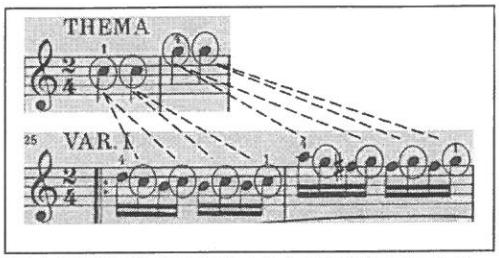


Fig. 2.1. 節錄自(莫札特：小星星主題變奏曲(KV 265))之主題與第 1 變奏（excerpt from theme and var.1 of ZWÖLF VARIATIONEN by Wolfgang Amadeus Mozart（A.D. 1756~1791）, KV 265）

2.2. Note-based 方法與瓶頸

爲了擷取歌曲中 repeating patterns，已經有一些 note-based 的方法被提出。這些方法以音符爲單位來擷取音高值，依時間順序排列得到一 note-based 音高序列，再發展近似於 text-based 的方法從音高序列裡找出 repeating patterns。而爲了擷取「approximate」repeating patterns，這些方法允許 pattern 的每一個 occurrence 裡可以插入少許個音符（Gap），允許有 Gap 恰巧能夠解決旋律變奏的問題，我們還是以小星星變奏曲的主題與第 1 變奏的第 1、2 小節爲例，如 Figure 2.1，主題的第一小節 $C_5 - C_5$ 在第 1 變奏裡被擴展成 $D_5 - C_5 - B_4 - C_5 - B_4 - C_5 - B_4 - C_5$，其中穿插了 6 個音符，若要辨識出 $D_5 - C_5 - B_4 - C_5 - B_4 - C_5 - B_4 - C_5$ 爲 $C_5 - C_5$ 的 occurrence，則 Gap 值至少要設爲 6，然而 pattern：$C_5 - C_5$ 的長度卻僅是 2 而已。顯然 note-based 方法的效果受制於 Gap 值的設定，Gap 值太大會有過多的假警報，太小則無法有效辨識旋律變奏，我們再以小星星變奏曲爲例，主題一共有 16 個音符，但是第 1 變奏已擴展成 114 音符，使用前

述的方法且設定 Gap 值爲 98，將不是個符合效益的作法，因爲過大的 Gap 值會增加運算的時間。

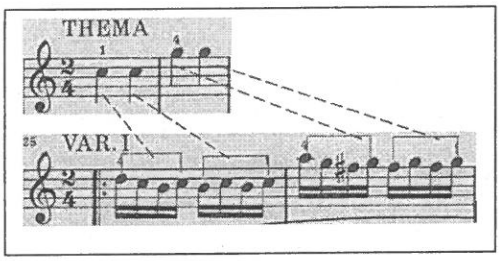


Fig. 2.2. 旋律變奏—每 4 分音符擴展爲四個 16 分音符

我們從樂理得知，旋律變奏不是單純地在主題裡穿插幾個音符使其重現而已，一段旋律變奏是以主題旋律的音高爲依據，用相異於主題的音符類型，組成與主題有相同時間長度的旋律。我們再以小星星變奏曲的主題與第 1 變奏的第一小節爲例，主題的第一個音符是一個 4 分音符 C_5，經過變奏成爲第 1 變奏的四個 16 分音符 $D_5 - C_5 - B_4 - C_5$，雖然插入了 3 個音符，但是合起來仍然是一個 4 分音符長，絕不會是二個或八個 16 分音符，如 Figure2.2。我們已經知道旋律變奏裡有此「時間長度規則」，但是 note-based 的方法只擷取音符的音高值，沒有包含時間的資訊，因此我們需要用包含音符的音高、時間長度的取樣方式，發展更符合樂理的方式以解決旋律變奏的情形。

3.問題描述、名詞定義與 Frame-based 取樣方式

3.1.問題描述

我們的研究目的是依據使用者給訂的三個參數：長度範圍 $\ell b \le L \le ub$ （$2 \le \ell b$）、dissimilarity ratio ε 、frequency threshold f ，從一 interval sequence S 裡快速地找出「重要而近似的重複樣型（nontrivial and approximate repeating pattern）」。一個 approximate repeating pattern 也稱爲 interval substring，它近似的出現在 S 裡至少 f 次，每一次出現都稱爲這個 pattern 的 occurrence。

我們在 3.2 節描述 interval sequence 的取樣方式，也在 3.3 節更詳細地定義

nontrivial and approximate repeating patterns 與相關的名詞。

3.2. Frame-based 取樣方式

我們在第 2.2 節提到，一般 note-based 方法的取樣方式是以音符爲單位擷取音高序列，只包含了音高的資訊，不具有節奏內涵。

爲了擷取兼具旋律與節奏的資訊，並符合旋律變奏的意義，我們採用 frame-based 的取樣方式從 midi 裡擷取音高序列，以時間爲單位，每隔一固定時間長度從 midi 裡擷取音高值，這種方式產生的音高序列具有拍長的意義，也表現出節奏快慢的變化。作曲家爲主題作變奏時，通常會保留主題旋律的音高，基於此特性我們認爲主題與其變奏會有相似的音樂輪廓（music contour）。以 Figure 2.1 的小星星的主題與第 1 變奏爲例，用 frame-based 的方式分別擷取音高序列，其音樂輪廓的呈現如 Figure 3.1 所示，我們不受音符個數不同所影響，可以正確的得到相同時間長度的兩音樂輪廓。爲了解決移調（transposition invariance）的問題，我們計算任意兩相鄰的 pitch 的差值，得到一音程序列（interval sequence），再從中找出「approximate repeating patterns」

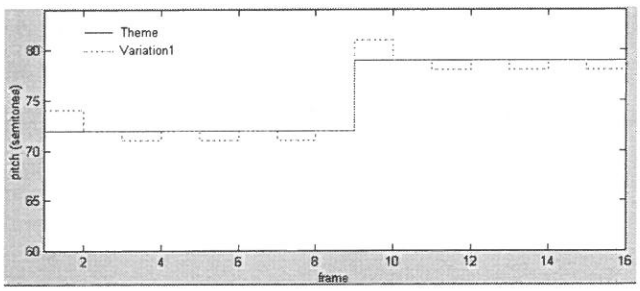


Fig. 3.1. The music contour of theme and variation 1

3.3.「重要而近似的重複樣型」的定義

在介紹我們提出的方法之前，先定義「重要而近似的重複樣型（nontrivial and approximate repeating pattern）」與相關名詞。

Definition 1：distance of interval substrings

- Let A and B are two substrings in S, and $|A| = |B| = L$. The distance is defined as follows.

  $D(A, B) = \sqrt{\sum_{i=1}^{L} (A(i) - B(i))^2}$,

  where $A(i)$ and $B(i)$ is the i_{th} symbol in A and B, respectively

- If $D(A, B) \le L \times \varepsilon$, We call the music contours of A and B are similar.

Example 1：

兩音程序列分別是 $A = \{-2, -2, 0, -3, 2\}$與 $B = \{-2, -2, 1, -3, 3\}$，則 A 與 B 之間的 distance $D(A, B) = \sqrt{2}$ 。

Definition 2：frequency of substring

- The frequency of a substring $P = S[a,b]$ in S is denoted as $freq(P)$. $freq(P) = \sum occurrence(P)$, where $occurrence(P)$

  $= \begin{cases} 1, \text{ if } \exists S[c, d] \in S, \text{ s.t. } D(P, S[c, d]) < k \times \varepsilon, \text{ where } |P| = k \text{ and } a \le c \\ 0, \text{ otherwise} \end{cases}$

Example 2：

對於一音程序列：$S = \{-2, -2, 0, -3, 2, 1, -2, -2, 1, -3, 3\}$，若 $\varepsilon = 0.4$ ，則在 S 中有一 substring $P = S[1,5] = \{-2, -2, 0, -3, 2\}$，$freq(P) = 2$，因爲 $D(P, S[1,5]) = 0 \le 2$ and $D(P, S[7,11]) = \sqrt{2} \le 2$ ，並且我們稱 $S[1,5]$與 $S[7,11]$是 P 的 occurrence。

Definition 3：approximate repeating patterns

- A substring P is approximate repeating pattern if $freq(P) \ge f$, where f is a given frequency threshold.

Example 3：

有一長度是 11 的 interval sequence $S = \{-2, -2, 0, -3, 2, 1, -2, -2, 1, -3, 3\}$，若 $\varepsilon = 0.4$ 、$f = 2$ 、$3 \le L \le 5$ ，則所找到的 approximate repeating patterns 如下：

$PS_4 = \{S[1, 4]\} = \{\{-2, -2, 0, -3\}\}$，因爲 $D(S[1, 4], S[1, 4]) = 0 \le 4 \times 0.4$ 且 $D(S[1, 4], S[7, 10]) = 1 \le 4 \times 0.4$

$PS_5 = \{S[1, 5]\} = \{\{-2, -2, 0, -3, 2\}\}$，因爲 $D(S[1, 5], S[1, 5]) = 0 \le 5 \times 0.4$ 且 $D(S[1, 5],$

$S[7, 11) = \sqrt{2} \leq 5 \times 0.4$

PS_L 是所有長度爲 L 的 approximate repeating pattern 所成的集合。

Definition 4：nontrivial and approximate repeating pattern

- Let $PS = \{P_1, P_2, ..., P_m\}$ is a set of approximate repeating patterns found from S. P_i is trivial, if $P_i \subset P_j$ and $freq(P_i) \leq freq(P_j)$. In this paper, we want to find all nontrivial and approximate repeating patterns from S.

Example 4：

接續 **Example 3** 的結果，所找到的兩個 approximate repeating pattern $S[1, 4]$ = {–2, –2, 0, –3}與 $S[1, 5]$ = {–2, –2, 0, –3, 2}因爲：

$S[1, 4]$是 $S[1, 5]$的 substring 且 $freq(S[1, 4]) = freq(S[1, 5])$ ，所以 $S[1, 4]$是一個 trivial pattern。對於 interval sequence S 而言，$S[1, 5]$是唯一的 nontrivial and approximate repeating pattern。

4.Curve-based method

4.1.Naïve 的方法

這個部分我們說明從一 interval sequence 中找出 approximate repeating patterns 的 Naïve method。

若要找尋的 patterns 的長度範圍是 $\ell b \leq L \leq ub$ ，Naïve method 先找出長度爲 ℓb 的 approximate repeating patterns，然後逐步地找出直到長度爲 ub 的所有 approximate repeating patterns。更仔細的來說，在用 frame-based 的取樣方式得到一長度是 n 的 interval sequence 後，在每一回合裡（要找長度 k 的 patterns），分別以第 1, 2, …, $n-k$ 個 frame 爲起始位置，各自與後面的連續 $k-1$ 個 frames 組成長度 k 的 substrings $S[1, k]$, $S[2, k+1]$, …, $S[n-k, n]$，我們首先考慮 $S[1, k]$，計算 $S[1, k]$與其它 substrings 之間的 Euclidean distance，若計算結果不大於 distance threshold 的 substrings 個數不少於 frequency threshold，那麼我們就得到一 approximate repeating pattern $S[1, k]$，以此類推找到所有長度 k 的 patterns，接著依序地完成每一回合找到所有符合長度範圍的 approximate repeating patterns。

Example 5：

針對一個 interval sequence S = {–2, –2, 0, –3, 2, 1, –2, –2, 1, –3, 3}，若要找長度是 L 的 approximate repeating pattern（$\varepsilon = 0.4$, $f = 2$, $3 \le L \le 5$），Naïve method 逐步地找出 $L = 3$、$L = 4$、$L = 5$ 的 patterns，逐一地檢查每個 substring 的 frequency 是否大於等於 f，我們舉找尋 $L = 5$ 的 patterns 爲例：

對於 S[1, 5]，逐一計算 $D(S[1, 5], S[1, 5])$, $D(S[1, 5], S[2, 6])$, …, $D(S[1, 5], S[7, 11])$ 是否小於 $\varepsilon \times L$，以求出 $freq(S[1, 5])$。

對於 S[2, 6]，逐一計算 $D(S[2, 6], S[2, 6])$, $D(S[2, 6], S[3, 7])$, …, $D(S[2, 6], S[7,11])$是否小於 $\varepsilon \times L$，以求出 $freq(S[2,6])$。

對於 S[6, 10]，計算 $D(S[6, 10], S[6, 10])$與$(S[6, 10], S[7, 11])$是否小於 $\varepsilon \times L$ 以求出 $freq(S[6, 10])$。

這個例子裡，只有 S[1, 5]的 frequency 大於等於 2，所以 $S[1, 5]$ = {–2, –2, 0, –3, 2} 是一個 approximate repeating pattern，S[1, 5]與 S[7, 11]是其 occurrences。

Naïve method 巨細靡遺地比對將耗費大量時間，在 4.2 節我們介紹 curve-based 方法，用以改進 Naïve method 在時間上的瓶頸。

4.2.Curve-based method 的概念與架構說明

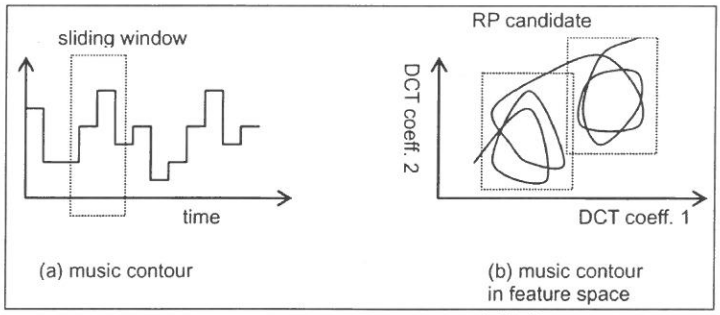


Fig. 4.1. Curve-based method 之概念

我們提出的 curve-based method 的概念是：首先以 frame-based 方式擷取 interval sequence，可表示成一 music contour（如 Figure 4.1（a））。用 sliding window 的方式從 interval sequence 取出一段一段 interval substrings，再做 DCT 運算取前幾個係數把它們對映（mapping）至低維度的特徵向量空間（feature space）中。這個 mapping 的特性使相似的 substrings 會在 feature space 中聚集在相近的位置。連續的數個點會形成一個

軌跡（trail），我們在 feature space 中找出每一組相似軌跡（similar trails），並視爲一 approximate repeating pattern 的 candidate（如 Figure 4.1（b）），再對這些 candidates 做後續的處理以找出「重要而近似的重複樣型（nontrivial and approximate repeating patterns）」。

Curve-based method 分成四個階段：1. Frame-based sampling、2. 利用離散餘弦轉換之對映（DCT mapping）、3. 找尋近似的頻繁軌跡（Similar frequent trails discovery）4. 近似重複樣型的驗證與刪除（Validation and pruning）。整個程式架構如 Figure 4.2 所示。在第 3 階段， DCT mapping 的結果使 interval sequence 轉換成 feature space 中的點，我們用一 point-based self-joining 方法找出 similar frequent trails，主要的概念是把 k 個點的 similar frequent trails 組成 k+1 個點的 similar frequent trails，Figure 4.3 是第 3 階段的 pseudo code。在 4.4 節我們分成三個部分更詳細地描述我們的方法。

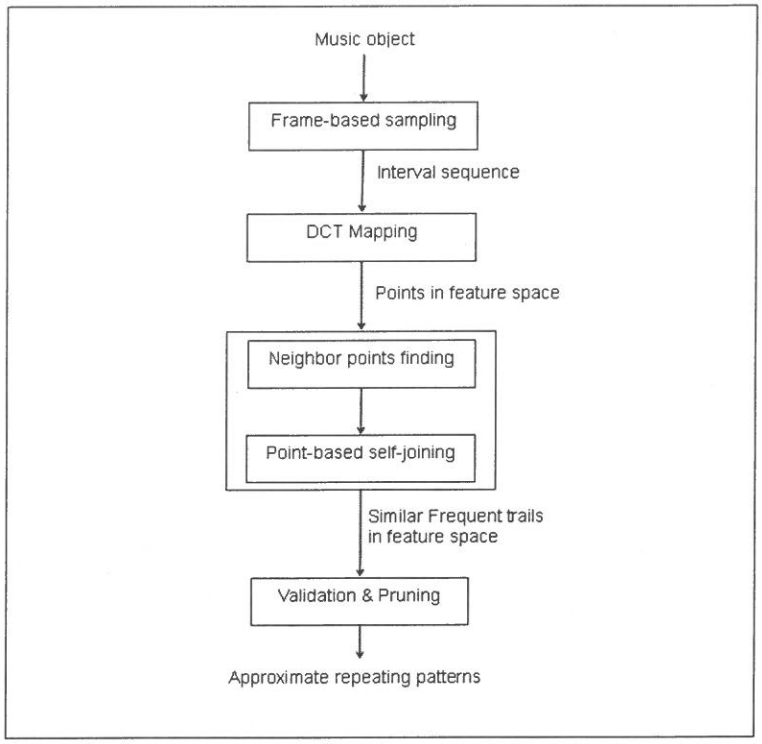


Fig. 4.2. Curve-based method 之流程圖

```
Similar_Frequent_Trails_Discovery (P, d)          // finding similar frequent trails
//   input : a set of points  P={p1, p2, ..., pn}  in the feature space
//   output : a set of all frequent  k -point trails
Begin
1.   d = threshold distance
2.   Find _ near_points ( P, d )                   // find all of the near-points set
3.   Find all of frequent 1-point pattern  FT1
4.   for ( k = 2; FTk-1 ≠ 0; k++ ) do
5.        Generate frequent k-point trails  FTk  from  FTk-1
6.   loop
End
```

Fig. 4.3. 找尋近似的頻繁軌跡之 procedure code

4.3.離散餘弦轉換的對映

我們的方法希望能擷取少數的且夠好的特徵值來代表一個 interval substring，基於我們的問題定義，所謂好的特徵值需具備兩個要素：

1.Preserve Euclidean distance（歐幾里德距離）

2.擁有序列裡絕大多數的重要資訊

對於一長度爲 n 的 interval substring，在對它作離散餘弦轉換（DCT）的運算後，我們取前幾個係數作爲特徵值。因爲絕大部分的重要資訊會保留在前面的係數並且依據 Parseval's theorem [14]，這個特徵值將保留（preserve）歐幾里德距離。離散傅立葉轉換（DCT）與離散餘弦轉換（DCT）都適用 Parseval's theorem [4]。

Parseval's theorem：Let $\vec{X}$ be the Discrete Fourier Transform of the sequence $\vec{x}$. Then we have

$$\sum_{i=0}^{n-1}|x_i|^2=\sum_{f=0}^{n-1}|X_f|^2$$

雖然保留越多的係數包含的資訊也越多，也能夠挑出更少的 candidates 與減少驗證階段的處理時間，但同時也加重「預先刪除」步驟的處理時間。我們的策略是挑選個數很少但是包含夠多資訊的係數當作特徵值，使我們能夠快速地篩選出 candidates 且有較少的假警報，讓我們的方法有更快的處理效率。

C. Faloutsos 曾經針對 subsequence matching in time-series databases 的問題，將

subsequences 作 DFT 的運算後取前兩項係數，把整個 sequence 對映（mapping）到 2-dimensional feature space，接著用 spatial access method 加速查詢的比對速度[3]。C. Faloutsos 的實驗結果發現，當係數個數是 2 或 3 時，可以在時間效率上得到最好的效果。我們延伸上述的觀念用一 curve-based method 找 approximate repeating patterns，也測試不同的係數個數對 curve-based method 在執行時間上的影響。

4.4.找尋近似的頻繁軌跡（Similar frequent trail）

我們的問題是要從 interval sequence S 裡找出長度爲 L（$\ell b \le L \le ub$）的 approximate repeating patterns。一個 pattern P 與 occurrence O（$|P|=|O| = L$）各可用一 sliding window（window size: w）取出 k（$k = L-w+1$）個 substrings $P[1, w]$, $P[2, w+1]$, ⋯, $P[k, L]$與 $O[1, w]$, $O[2, w+1]$, …, $O[k, L]$，這些 substrings 必然滿足此關係：

$$D(P[1, w], O[1, w]) \le \varepsilon \times ub,\ D(P[2, w+1], O[2, w+1]) \le \varepsilon \times ub,\ \ldots,\ D(P[k, L], \text{ and } O[k, L]) \le \varepsilon \times ub \quad (1)$$

我們分別對上述 P 與 O 的 substrings 作 DCT 運算並取前 t 個係數 mapping 到 t-dimensional feature space 中的連續 k 個點 $p_1, p_2, ..., p_k$ 與 $o_1, o_2, ..., o_k$，且滿足下述的關係[3]：

$$D(p_1, o_1) \le D(P[1, w], O[1, w]),\ D(p_2, o_2) \le D(P[2, w+1], O[2, w+1]),\ \ldots, \text{ and } D(p_k, o_k) \le D(P[k, L], O[k, L]) \quad (2)$$

由（1）與（2）可推得，P 與 O 必然滿足 Condition 1。

Condition 1： $D(p_i, o_i) \le \varepsilon \times ub$, where $i = 1, 2, ..., k$

若有任意兩 interval substrings M 與 N，mapping 成連續 k 個點 $m_1, m_2, ..., m_k$ 與 $n_1, n_2, ..., n_k$ 並且滿足 Condition 1，則我們稱 $m_1, m_2, ..., m_k$ 與 $n_1, n_2, ..., n_k$ 互爲彼此的相似軌跡（similar trail）。舉 Example 6 爲例：

Example 6：有一 interval sequence $S = \{-2, -2, 0, -3, 2, 1, -2, -2, 1, -3, 3\}$, 當 $\varepsilon = 0.4$，則 $S[1, 5]$與 $S[7, 11]$是相似的（$\because D(S[1, 5], S[7, 11]) \le \varepsilon \times 5$）。

若取 window size $w = 3$，對兩 substrings 做 DCT 運算並取前兩項係數，可 mapping 到 2-dimensional feature space 中 6 個點，如 Figure 4.4 (a)。因爲 $D(S[1, 5], S[7, 11]) \le \varepsilon \times 5$，所以 $D(S[1, 3], S[7, 9]) \le \varepsilon \times 5$、$D(S[2, 4], S[8, 10]) \le \varepsilon \times 5$、$D(S[3, 5], S[9, 11])$

$\leq \varepsilon\times5$ 必皆成立，可推得 $D(p_1, p_7) \leq \varepsilon\times5$、$D(p_2, p_8) \leq \varepsilon\times5$、$D(p_3, p_9) \leq \varepsilon\times5$，如 Figure 4.4 (b)。我們稱這兩個 substrings 在 feature space 中形成相似軌跡（similar trail）。

對於一 interval substring P，可以 mapping 成連續 k 個點 $p_1, p_2, ..., p_k$，若 feature space 中存在包括 $p_1, p_2, ..., p_k$ 的至少 f 個 similar trails（f is the frequency threshold），則我們稱 $p_1, p_2, ..., p_k$ 是一 similar frequent trail，也稱爲一 frequent k-point trail，表示成此形式：$\{p_1, freq(p_1, p_2, ..., p_k), (\alpha_1, \alpha_2, ..., \alpha_m)\}$，$p_1$ 是 trail $p_1, p_2, ..., p_k$ 起始點的編號，$freq(p_1, p_2, ..., p_k)$ 是 similar trail 的個數，$(\alpha_1, \alpha_2, ..., \alpha_m)$ 是每一個 similar trail 的起始點編號，且 $\alpha_1 < \alpha_2 < ... < \alpha_m$。

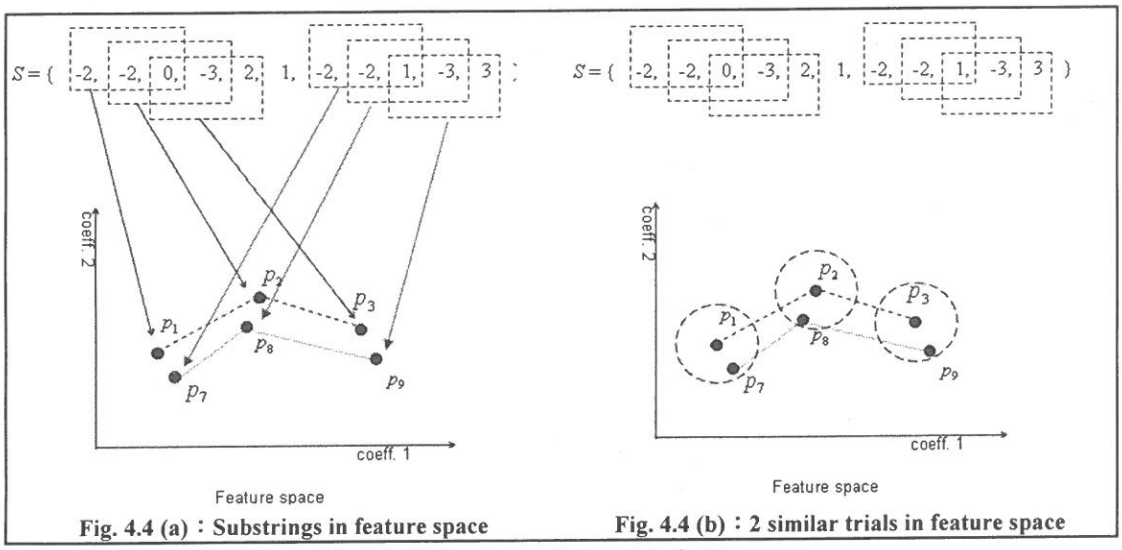


Fig. 4.4 (a)：Substrings in feature space **Fig. 4.4 (b)：2 similar trials in feature space**

Fig. 4.4. Similar trials discovery in feature space

4.4.1.Feature Space 中距離相近的點

爲了找出 similar frequent trails，我們先對 feature space 中每一點個別找出與其距離相近的點，如果某一點 p_i 及與它距離相近的點總個數不小於 f（f is the frequency threshold），則 p_i 是一個 frequent 1-point trail。

要找出每個點及個別與其距離相近的點，一一地計算任意兩點之間的 Euclidean distance 不是一個聰明的作法。對於在 2-dimensional feature space 中的點 (x_i, y_i)，若有另一點 (x_j, y_j) 與 (x_i, y_i) 距離相近，至少要滿足 Condition 2.1、Condition 2.2，所以我

們先挑出滿足二 conditions 的 candidates，再計算這些點與 (x_i, y_i) 之間眞正的 Euclidean distance，決定眞正鄰近 (x_i, y_i) 的點。

Condition 2.1：$|x_i - x_j| \le \varepsilon * ub$

Condition 2.2：$|y_i - y_j| \le \varepsilon * ub$

挑出鄰近點的 candidates 的詳細做法是：我們分別依 feature space 中的每個點的 x、y 座標值由小到大對點作排序，會得到兩個序列 sort_x、sort_y。當我們要找尋點 (x_i, y_i) 的所有鄰近點時，首先到 sort_x 裡找出點 (x_i, y_i)，然後循序地向前與向後找出所有滿足 Condition 2.1 的點；相似地做法再到 sort_y 裡找出滿足 Condition 2.2 的點，接著對找出的兩組點取 intersection 找到 candidates，最後再檢查點 (x_i, y_i) 與這些 candidates 之間眞正的 Euclidean distance，以找出眞正的鄰近點。舉 Example 7 爲例來說明：

Example 7：有一 interval sequence 利用 DCT 對映至 feature space 中的 13 個點分別爲：$p_1(9,10)$、$p_2(12,12)$、$p_3(20,16)$、$p_4(25,13)$、$p_5(20,9)$、$p_6(22,3)$、$p_7(10,3)$、$p_8(4,2)$、$p_9(2,4)$、$p_{10}(4,11)$、$p_{11}(8,14)$、$p_{12}(15,17)$、$p_{13}(19,23)$，我們對每個點找出與之距離相近的點（若 $\varepsilon \times ub = 6$、$f = 2$）。

Step 1：分別依 x、y 座標值由小到大對點作排序（Point_num = 1 代表 p_1，以此類推。）

Sorting by X-coordinate	
Point_num	Coordinate
9	(2,4)
8	(4,2)
10	(4,11)
11	(8,14)
1	(9,10)
7	(10,3)
2	(12,12)
12	(15,17)
13	(19,23)
3	(20,16)
6	(22,3)
4	(25,13)
5	(29,9)

Sorting by Y-coordinate	
Point_num	Coordinate
8	(4,2)
6	(22,3)
7	(10,3
9	(2,4)
5	(29,9)
1	(9,10)
10	(4,11)
2	(12,12)
4	(25,13)
11	(8,14)
3	(20,16)
12	(15,17)
13	(19,23)

Step 2：找尋與 $p_1(9,10)$ 距離相近的 candidates

Sorting by X-coordinate	
Point_num	Coordinate
9	(2,4)
8	(4,2)
10	(4,11)
11	(8,14)
1	(9,10)
7	(10,3)
2	(12,12)
12	(15,17)
13	(19,23)
3	(20,16)
6	(22,3)
4	(25,13)
5	(29,9)

Sorting by Y-coordinate	
Point_num	Coordinate
8	(4,2)
6	(22,3)
7	(10,3
9	(2,4)
5	(29,9)
1	(9,10)
10	(4,11)
2	(12,12)
4	(25,13)
11	(8,14)
3	(20,16)
12	(15,17)
13	(19,23)

對 p_1 而言，滿足 Condition 2.1 的點集合是：$\{1,2,7,8,10,11,12\}$。

對 p_1 而言，滿足 Condition 2.2 的點集合是：$\{1,2,3,4,5,9,10,11\}$。

因此與 p_1 距離相近的 candidates 有：$\{1,2,7,8,10,11,12\} \cap \{1,2,3,4,5,9,10,11\} = \{1,2,10,11\}$，再計算 Euclidean distance 以驗證後，p_1、p_2、p_{10}、p_{11} 確實是 p_1 的相鄰點。以此類推可找出每個點的距離相近點，藉此可產生所有 frequent 1-point trails：$\{\{1,4,(1,2,10,11)\},\{2,3,(2,11,12)\},\{3,3,(3,4,12)\},\{4,2,(4,5)\},\{8,2,(8,9)\},\{10,2,(10,11)\}\}$

4.4.2.Point-based self-joining

我們逐步地 join 較短的 similar frequent trails 來得到較長的 similar frequent trails。從 frequent k-point trails 裡，正確地組合兩個 trails 可得到一 frequent (k+1)-point trail，稱之爲 self-joining。詳細的組合方式是，若有二 frequent k-point trails $\{p_1, freq(p_1,p_2,...,p_k),(\alpha_1,\alpha_2,...,\alpha_m)\}$、$\{q_1, freq(q_1,q_2,...,q_k),(\beta_1,\beta_2,...,\beta_n)\}$，其中 $p_1+1=q_1$，並且有 F（$F \geq f$）個 similar trails 的起始點 $po_1, po_2, ..., po_F$ 在 $(\alpha_1,\alpha_2,...,\alpha_m)$、$qo_1, qo_2, ..., qo_F$ 在 $(\beta_1,\beta_2,...,\beta_n)$，使得 $po_i+1=qo_i, 1 \leq i \leq F$，那麼這二 frequent k-point trails 可以 self-joining 成一 frequent (k+1)-point trail：$\{p_1, freq(p_1,p_2,...,p_k,q_k),(po_1,po_2,...,po_F)\}$，我們舉 Example 8 爲例：

Example 8：繼續 Example 7 的工作，在找出 frequent 1-point trails 之後，以 self-joining 可以逐步組成 frequent 2-point trails 與 frequent 3-point trails 如下：

frequent 1-point trails：$\{\{1,4,(1,2,10,11)\},\{2,3,(2,11,12)\},\{3,3,(3,4,12)\},$
$\{4,2,(4,5)\},\{8,2,(8,9)\},\{10,2,(10,11)\}\}$

frequent 2-point trails：$\{\{1,3,(1,10,11)\},\{2,2,(2,11)\},\{3,2,(3,4)\}\}$

frequent 3-point trails：$\{\{1, 2, (1,10)\}\}$

我們以一 frequent 3-point trail：$\{1, 2, (1,10)\}$爲例，它由兩個 frequent 2-point trails：$\{1, 3, (1,10,11)\}$、$\{2, 2, (2,11)\}$組成，代表在 feature space 中有一 2-point trail：1, 2，它有另二 similar trail：10, 11 與 11, 12；另有一 2-point trail：2, 3，它有另一 similar trail：11, 12，它們可以 join 成一 3-point trail：1, 2, 3 而且有兩個 similar trails（1, 2, 3 與 10, 11, 12），$freq(1, 2, 3) = 2$。

4.4.3.近似重複樣型的驗證與刪除

對於任一 frequent k-point trail：$\{p_1, freq(p_1, p_2, ..., p_k), (po_1, po_2, ..., po_F)\}$，它是一個長度爲（$k+w-1$）的 approximate repeating pattern 的 candidate（window size is w），代表在一 interval sequence 中有一 substring $S[p_1, p_{k+w-1}]$，它的可能的 occurrences 有：$S[p_{po_1}, p_{po_1+k+w-2}]$、$S[p_{po_2}, p_{po_2+k+w-2}]$、…、$S[p_{po_F}, p_{po_F+k+w-2}]$。我們需要計算 $S[p_1, p_{k+w-1}]$ 與這 F 個 occurrences 之間的 Euclidean distance 以驗證 $S[p_1, p_{k+w-1}]$ 是否眞正爲一個 approximate repeating pattern，Figure 4.5 爲此部分的執行程序。

```
Validation (FT_k)
//  input : the set of all frequent k - point trails
//  output : the set of all repeating (k+w-1) - patterns
Begin
1.  for each frequent k-point trail {p_i, freq(p_i, p_{i+1}, ..., p_{i+k-1}), (po_1, po_2, ..., po_F)}
in FT_k  do
2.      for each occurrence_i of pattern_{p_1}  do
3.          if (D(pattern_{p_1}, occurrence_i) ≤ d) then    // D(pattern_{p_1}, occurrence_i) is
the Euclidean distance, d is the threshold distance
4.              frequency++
5.          end if
6.      loop
7.      if (frequency ≥ f) then     // f is the frequency threshold
8.          add {p_i, freq(p_i, p_{i+1}, ..., p_{i+k-1}), (po_1, po_2, ..., po_F)} in RP_{k+w-1}
9.      end if
10. loop
11. return RP_{k+w-1}
End
```

Fig. 4.5.『驗證 patterns』之 procedure code

一個長度爲 k 的 approximate repeating pattern $S[i, i+k-1]$ 也稱爲一 repeating k-pattern，可以表示成此形式：$\{i, freq(S[i, i+k-1]), (a_1, a_2, ..., a_f)\}$，代表 $S[i, i+k-1]$ 有 f

個 occurrences：$S[a_1, a_1+k-1]$ 、$S[a_2, a_2+k-1]$ 、…、$S[a_f, a_f+k-1]$ 。

在找出的 approximate repeating patterns 之中還留有許多 trivial patterns，在 Removal of trivial pattern 的階段，我們依據 4.3 節的 Definition 4，剔除所有 trivial patterns，只保留 nontrivial patterns。對於一 pattern P_j，在找尋 P_j 的 sub-pattern P_i（$P_i \subset P_j$）的過程裡我們只要檢查 P_i 與 P_j 的起始位置，就可以很快的確認是否滿足 $P_i \subset P_j$ 的關係，Figure 4.6 爲此部分的執行程序。舉 Example 9 爲例接續 Example 8 的工作，最後要找到所有 nontrivial and approximate repeating patterns。

```
Prune (RP_1, RP_2, ..., RP_k)  // pruning trivial patterns in RP_1 ∪ RP_2 ∪ ... ∪ RP_{k-1}
//   input : repeating patterns sets : RP_1, RP_2, ..., RP_k.
//   output : nontrivial repeating patterns set : RP_1, RP_2, ..., RP_k  at present.
Begin
1.   for each repeating k - pattern {p_i, freq(p_i, p_{i+1}, ..., p_{i+k-1}), (po_1, po_2, ..., po_F)} in
RP_k
2.     for (m = 1; m ≤ k-1; m++) do
3.       for each repeating m-pattern {q_j, freq(q_j, q_{j+1}, ..., q_{j+m-1}), (qo_1, qo_2, ..., qo_F)}
in  RP_m
4.         if(q_j ≤ (p_i + k - m)  &&  freq(p_i, p_{i+1}, ..., p_{i+k-1}) ≥ freq(q_j, q_{j+1}, ..., q_{j+i-1}))
5.             remove {q_j, freq(q_j, q_{j+1}, ..., q_{j+i-1}), (qo_1, qo_2, ..., qo_F)} from  RP_m
6.         end if
7.       loop
8.     loop
9.   loop
End
```

Fig. 4.6. 『刪除 trivial patterns』之 procedure code

Example 9：接續 Example 8，假設所找出的每一組 similar frequent trails 經過驗證後全部皆爲 approximate repeating patterns 如下，我們要進一步挑出 nontrivial patterns（假設 window size $w = 3$）：

repeating 3-patterns：$\{\{1, 4, (1, 2, 10, 11)\}, \{2, 3, (2, 11, 12)\}, \{3, 3, (3, 4, 12)\}, \{4, 2, (4, 5)\}, \{8, 2, (8, 9)\}, \{10, 2, (10, 11)\}\}$

repeating 4-patterns：$\{\{1, 3, (1, 10, 11)\}, \{2, 2, (2, 11)\}, \{3, 2, (3, 4)\}\}$

repeating 5-patterns：$\{\{1, 2, (1, 10)\}\}$

對於任一 repeating k-pattern P，逐步地檢查所有 repeating 1-patterns 至 repeating (k-1)-patterns 之中是否有 P 的 sub-patterns。

Step 1：找尋 repeating 5-patterns 的 sub-patterns，並刪除 trivial patterns

Step 1.1：檢查 repeating 4-patterns

對於 repeating 5-pattern：$\{1,2,(1,10)\}$（起始位置是第 1 個 frame），有兩 repeating 4-patterns：$\{1,3,(1,10,11)\}$、$\{2,2,(2,11)\}$的起始位置不大於 $1+(5-4)$，所以它們是 repeating 5-pattern：$\{1,2,(1,10)\}$的 sub-patterns，又因爲$\{2,2,(2,11)\}$的 frequency 不大於 repeating 5-pattern：$\{1,2,(1,10)\}$，因此 repeating 4-patterns：$\{2,2,(2,11)\}$、是 trivial patterns，我們只保留 repeating 4-patterns：$\{1,3,(1,10,11)\}$、$\{3,2,(3,4)\}$。

Step 1.2：檢查 repeating 3-patterns

Repeating 3-patterns：$\{1,4,(1,2,10,11)\}$、$\{2,3,(2,11,12)\}$是 repeating 5-pattern $\{1,2,(1,10)\}$的 sub-patterns（起始位置不大於 $1+(5-3)$），但它們的 frequency 都大於 repeating 5-pattern：$\{1,2,(1,10)\}$，因此 repeating 3-patterns：$\{1,4,(1,2,10,11)\}$、$\{2,3,(2,11,12)\}$目前不算是 trivial repeating patterns。

Step 2：找尋 repeating 4-patterns 的 sub-patterns，並刪除 trivial patterns

Step 1.1：檢查 repeating 3-patterns

對於 repeating 4-pattern：$\{3,2,(3,4)\}$（起始位置是第 3 個 frame）有兩 repeating 3-patterns：$\{3,3,(3,4,12)\}$、$\{4,2,(4,5)\}$的起始位置不大於，$3+(4-3)$，所以它們是 repeating 4-pattern：$\{3,2,(3,4)\}$的 sub-patterns，其中，repeating 3-pattern：$\{4,2,(4,5)\}$的 frequency 不大於 repeating 4-pattern：$\{3,2,(3,4)\}$的 frequency，所以是一個 trivial pattern。

最後所找出的 nontrivial and approximate repeating patterns 有：

repeating 3-patterns：$\{\{1,4,(1,2,10,11)\},\{2,3,(2,11,12)\},$
$\{3,3,(3,4,12)\},\{8,2,(8,9)\},\{10,2,(10,11)\}\}$

repeating 4-patterns：$\{\{1,3,(1,10,11)\},\{3,2,(3,4)\}\}$

repeating 5-patterns：$\{\{1,2,(1,10)\}\}$

4.5.改善效果的策略（Improving strategy）

依據 3.3 節的定義，我們從一 frame-based interval sequence 裡找出 approximate repeating pattern，並從中刪除了 trivial patterns，最後留下較重要的 nontrivial and approximate repeating pattern。然而在其中包含了許多不符合樂理的 fragmental patterns，而且我們的方法高度的受給定的參數值所影響，可能會找出太多的 patterns。我們的目

標是希望保留越少而越重要的 patterns，在這一章裡我們提出兩個方式改善上述兩因素造成的影響。

4.5.1.刪除不完整的樣型（Fragmental patterns）

我們要刪除所有不符合音樂特性的 fragmental patterns。在 3.2 節曾經提到我們用 frame-based 方式擷取音程序列，以小星星變奏曲的主題爲例，它的一拍時間有 120 ticks，若我們以每 30 ticks 爲一個 frame 時間長度，那麼第一小節的第一個 4 分音符會表示爲四個 frames $C_5 - C_5 - C_5 - C_5$，這四個 frames 共同代表一個四分音符 C_5，不可分割，此時若有一個以第二或第三或第四個 frame 爲起使位置的近似重複樣型，我們稱這是一個不完整的樣型（fragmental pattern），它不符合樂理意義。以 Figure 4.7 爲例，對於小星星主題變奏曲的第 1、2 小節的四個音符 $C_5 - C_5 - G_5 - G_5$，我們用 frame-based 的方式擷取出 16 個 frames，其中第 1 至第 4、第 5 至第 8、第 9 至第 12、第 13 至第 16 frame 分別代表 1 個四分音符，所以只有從第 1、第 5、第 9、第 13 個 frame 起始的 pattern 才是一個符合樂理的 approximate repeating pattern，其餘都屬 fragmental pattern，不可能是一個 Theme，也不適合用來建立索引結構。在程式實際撰寫時，我們在驗證一個 candidate 是否眞爲 pattern 之前檢查它起始的 frame 編號，若不爲一個音符的起始 frame，我們就刪除這個 candidate，因爲即使它是一個 approximate repeating pattern 也只是一個 fragmental pattern。一首歌裡可能包含很多的 fragmental patterns，這個方式可以省略不必要的驗證計算以加速程式的執行。

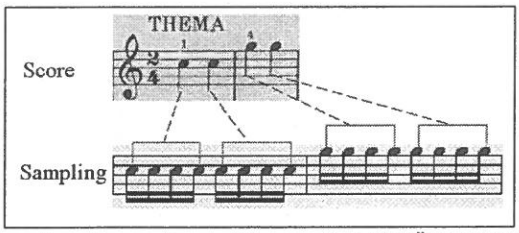


Fig. 4.7. Frame-based sampling to Measure 1-2, ZWÖLF VARIATIONEN

4.5.2.近似重複樣型的排序

由於我們所提出之方法的效果高度的受使用者給定的參數所影響，可能因此產生數量太多的 nontrivial and approximate repeating patterns，所以我們要對所找到的 patterns 作排序以挑出更爲重要的 patterns。此外，一些的論文研究裡會挑選長度最長的 patterns 當作古典歌曲主題，但是當 frequency threshold 過低時，將誤使我們太過看

重 pattern 長度而挑選不符眞實的 nontrivial and approximate repeating patterns 作爲查詢索引，因此，我們需要一個較適當而符合樂理的排序方式。

對於一個 nontrivial and approximate repeating pattern 而言，它的 occurrences 之間發生 overlapping 的程度可以衡量此 pattern 的「重要性程度（importance rate）」。若 occurrences 彼此互不發生 overlapping，此 pattern 越具重要性；反之，此 pattern 越不重要。

對於一長度爲 L 的 pattern P，我們以它的長度與「重現度（occurrence rate）」來計算它的重要性程度，其公式如下：

Importance rate（P）$= x \times L + y \times$ Occurrence rate(P)，其中 $0 \le x,y \le 1$。

Occurrence rate $= \dfrac{num(P) + \sum num(O_i)}{freq(P)}$，其中

$num(P)$ = 不與 pattern P 發生 overlapping 的重現（occurrence）個數

$num(O_i)$ = 對於每個重現 O_i，其餘重現不與之發生 overlapping 的個數。

我們舉 Example 10 爲例：

Example 10：有一長度爲 5 的 pattern $P = \{1, 4, (1, 5, 8, 14)\}$，$P$ 的 importance rate 計算如下：（x = 0.4, y = 0.6）

$num(P) = 2$，$\sum num(O_i) = 2 + 1 + 2 + 3 = 8$

$\Rightarrow$ Occurrence rate $= (2 + 8) \div 4 = 2.5$

$\Rightarrow$ Importance rate $= 0.4 \times 5 + 0.6 \times 2.5 = 3.5$

我們依 patterns 的 importance rate 對它們作排序，據此可以挑出更重要的 nontrivial and approximate repeating patterns。

5.實驗分析

我們提出的 curve-based method 目標是希望盡可能預先刪除一部分絕不可能的答案、留下較少的 candidates，藉此可以減少驗證 Euclidean distance（在任意兩相同長度的 substrings 之間）的計算量，使我們的方法可以耗費較少的時間以找出一首歌裡的 approximate repeating patterns。我們進行一些實驗，測試 curve-based method 在各種變

數下的預先刪除效果，也比較 curve-based method 與 Naïve method 在時間上的效率。

5.1.實驗設定

我們從莫札特的小星星主題變奏曲的 midi 檔案裡擷取第 3 個 track 的 melody 作爲測試資料，除了第 9 變奏的主旋律發生跨聲部的情形之外，其餘變奏的主旋律都落在第 3 個 track。第 3 個 track 全部有 2866 個音符，我們擷取出長度爲 5403 的 interval sequence，欲從中找出所有 nontrivial and approximate repeating patterns。

5.2.實驗結果

影響 curve-based method 的效果優劣的變因有四點：（1）dissimilarity ratio ε、（2）frequency threshold f、（3）the range of the length：$\ell b \le L \le ub$ 、（4）the number of DCT coefficients：m。我們在四個不同參數條件下統計從小星星主題變奏曲找出的 approximate repeating pattern 數量，也比較 naïve method 與 curve-based method 在時間上的效率，其中不包含作 DCT 運算所花的時間（在 Matlab 執行約 1.42 秒）。在這些實驗裡，window size $w = 31$。

5.2.1.時間上的效率

Figure 5.1 展示不同的 DCT 係數個數對 curve-based method 在時間效率上的影響。實驗結果顯示至少要取前 8 個係數，curve-based method 在時間效率上才會超前 naïve method，而且取的係數個數越多，耗費的時間越少（除了取前 30 個係數時，時間略爲攀升之外），這個實驗結果與 C. Faloutsos 的結論有極大差異，這是因爲我們對每個 interval substring（in a sliding window）作 DCT 運算後，前 3 個係數並未含有足夠多的 information，甚至用前 7 個係數都無法讓 curve-based method 有效地找出較少的 candidates，致使在時間效率上落後 naïve method。

Figure 5.2 展示不同的 dissimilarity ratio 對耗費時間的影響。隨著 dissimilarity ratio 越大，curve-based method 花費的時間也越多，因爲我們把 substrings（in a sliding window）對映到 feature space 後，找尋鄰近點需滿足的 distance threshold（$\varepsilon \times ub$）條件變的越寬鬆了，所以留下越多的 candidates 也花費越多的時間在 Validation 的階段。當 dissimilarity ratio 在 0.1 至 0.18 之間，curve-based method 在時間上有較快的執行效率。

Figure 5.3 展示不同的 frequency threshold 對耗費時間與找出的 pattern 個數的影響。雖然 frequency threshold f 越大所找出的 patterns 也越少，甚至當 $f = 80$ 時 patterns

個數只有 11 個，但是在找尋 candidates 時，絕大多數的 frequent k-point trails 的 frequency 都超過 80，以致 frequency threshold 在 10 至 80 之間對於 candidates 的個數多寡影響不大，因此 curve-based method 耗費的時間並沒有顯著地減少。

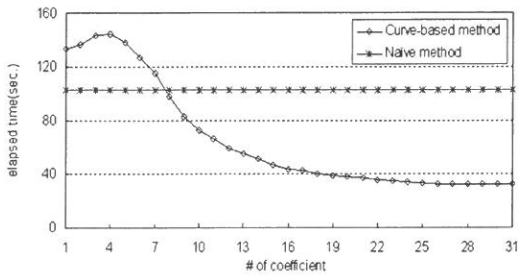


Figure 5.1：Elapsed time vs. # of coeff.（$\varepsilon = 0.14$、$f = 10$、$31 \le L \le 63$）

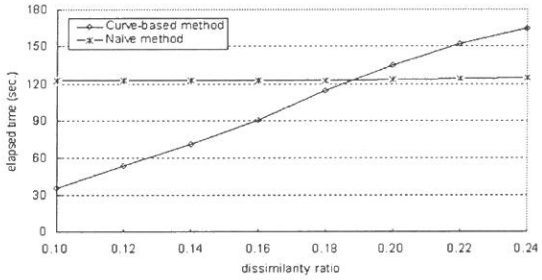


Figure 5.2：Elapsed time vs. dissimilarity ratio（$f = 20$, $31 \le L \le 63$, $m = 12$）

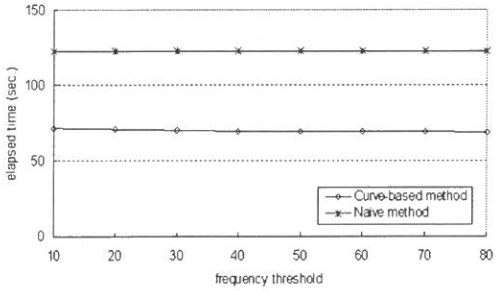


Fig. 5.3. Elapsed time vs. frequency pattern length（$\varepsilon = 0.14$、$f = 20$、$m = 12$）

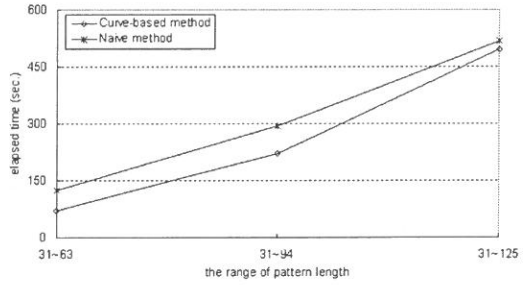


Fig. 5.4. Elapsed time vs. the range of threshold（$\varepsilon = 0.14$、$31 \le L \le 63$、$m = 12$）

Figure 5.4 展示要找尋的 pattern 長度範圍對耗費時間的影響。當長度範圍的 upper bound 越大，找尋鄰近點需滿足的 distance threshold（$\varepsilon \times ub$）條件變的越寬鬆，在長度範圍的 upper bound 大於 156 的時候，curve-based method 在找尋 similar frequent trail 的階段花費更多的時間卻留下了太多 candidates，使的總體的執行時間大過 Naïve method。但是我們對照小星星主題變奏曲的樂譜，其主題旋律有的 a、b 兩段，長度各是 63 個 frames，而 curve-based method 即使在尋找長度範圍是 $31 \le L \le 125$ 的 approximate repeating patterns 時仍然有較快的執行效率，當要找的 pattern 長度範圍是 $31 \le L \le 63$ 時，curve-based method 在時間上比 Naïve method 快 1.74 倍。

5.2.2.Patterns 的數量

Figure 5.5 統計在不同的 dissimilarity ratio ε，從小星星主題變奏曲裡找出的 nontrivial and approximate repeating pattern 數量。當 dissimilarity ratio 越大，所找出的 approximate repeating pattern 數量會急遽攀升，但是絕大多數是 trivial patterns，我們只需要保留 nontrivial and approximate repeating patterns，不會失去 trivial patterns 的 melody information。

Figure 5.6 統計在不同的 frequency threshold f 之下，從小星星主題變奏曲裡找出的 nontrivial and approximate repeating pattern 數量。當 frequency threshold 越大所找到的 patterns 也越少，當 $f = 20$，nontrivial and approximate repeating pattern 的數量從 338 個大幅減少爲 202 個，這代表小星星主題變奏曲裡會重現 10 次的 patterns 太普遍，較重要的 patterns 至少要重現 20 次以上。

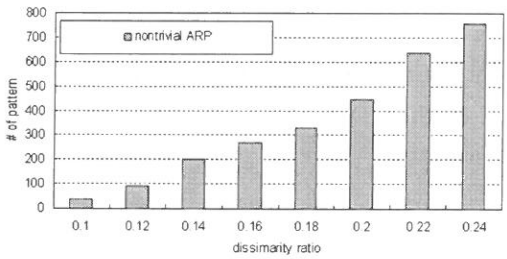


Fig. 5.5. # of Pattern vs. dissimilarity ratio（$f = 20$、$31 \le L \le 63$、$m = 12$）

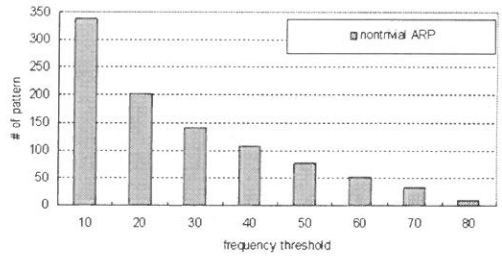


Fig. 5.6. # of pattern vs. frequency threshold（ε = 0.14、$31 \le L \le 63$、$m = 12$）

Figure 5.7 統計在不同的長度範圍之下，從小星星主題變奏曲裡找出的 nontrivial and approximate repeating pattern 數量。因爲我們衡量 substrings 之間相似與否的 distance threshold 是一變動值（$\varepsilon \times L$），所以當長度範圍變的越大，會找到更多 approximate repeating pattern，但是 nontrivial and approximate repeating pattern 的數量卻未必如此，因爲可能包含很多 trivial patterns 被刪除掉，例如在 Figure 5.7 中 pattern 的長度範圍是 31~125 之時。

Figure 5.8 展示 curve-based method 挑出的 candidates 與 approximate repeating patterns（without removal）的數量依長度作統計。Curve-based method 在長度是 31 時無法有效的挑選較少的 candidates，這是因爲我們以長度範圍的 upper bound 來計算 distance threshold，但是隨著長度越長留下的 candidates 越少，這是 curve-based method 在效率上能勝過 Naïve method 的關鍵因素所在，因爲長度越長的 interval substrings 需要花費

更多時間計算 Euclidean distance，而 curve-based method 可以有效的減少這個部分的計算量。

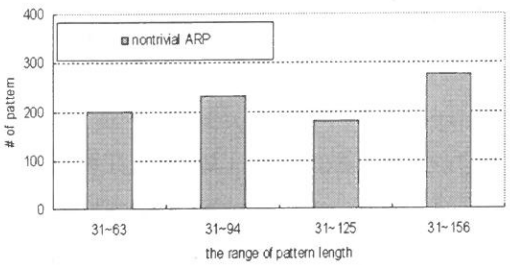


Fig. 5.7. # of pattern vs. the range of pattern length（ε = 0.14、f = 20、$31 \le L \le 63$、m = 12）

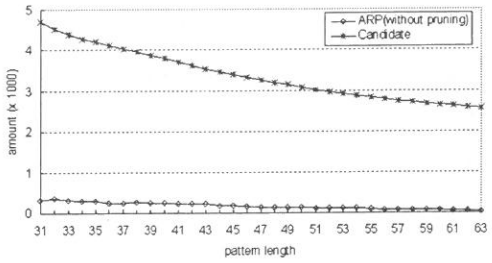


Fig. 5.8. Amount vs. pattern length（ε = 0.14、f = 20、$31 \le L \le 63$、m = 12）

5.2.3.刪除的效果

從我們的實驗結果得知，從音樂物件裡找出的 approximate repeating pattern（without removal）數量太多，所以 curve-based method 有「Removal of fragmental patterns」與「Removal of trivial patterns」兩個階段，最後找出符合音樂特性的 nontrivial and approximate repeating patterns。Figure 5.9、Figure 5.10、Figure 5.11 展示在三種參數條件下，「all approximate repeating patterns（all ARPs）」、「Removal of fragmental patterns」、「Removal of trivial patterns」三個階段裡保留的平均 patterns 數量。

在 Figure 5.9（f = 20、$31 \le L \le 63$、m = 12，在 $\varepsilon = 0.1$ 至 $\varepsilon = 0.24$ 八個 dissimilarity ratio 下的 removal 效果），與「all ARPs」階段相比，「Removal of fragmental patterns」階段只保留了 18%~26%的 patterns；最後在「Removal of trivial patterns」階段，除了 $\varepsilon = 0.1$ 之外，只保留 1%~5%的 patterns。

在 Figure 5.10（$\varepsilon = 0.14$、$31 \le L \le 63$、m = 12，在 f = 10 至 f = 80 八個 frequency threshold 下的 removal 效果），與「all ARPs」階段相比，「Removal of fragmental patterns」階段只保留了 8%~18%的 patterns（f = 70、f = 80 例外）；最後在「Removal of trivial patterns」階段，只保留 1%~4%的 patterns。

在 Figure 5.11（$\varepsilon = 0.14$、f = 20、m = 12，在 $31 \le L \le 63$、$31 \le L \le 94$、$31 \le L \le 125$、$31 \le L \le 156$ 四個長度範圍下的 removal 效果），與「all ARPs」階段相比，「Removal of fragmental patterns」階段只保留了 15%~20%的 patterns；最後在「Removal of trivial patterns」階段，只保留 0.5%~4%的 patterns。

由實驗結果得知，這二個 removal 階段有效地刪除 95%以上較不重要的 patterns，從小星星主題變奏曲裡找出的 nontrivial and approximate repeating pattern 只佔所有 patterns 的 5%。

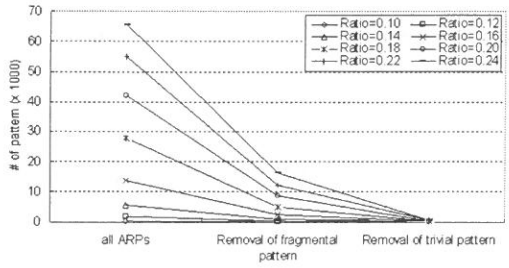


Fig. 5.9. The removal effect—frequency threshold

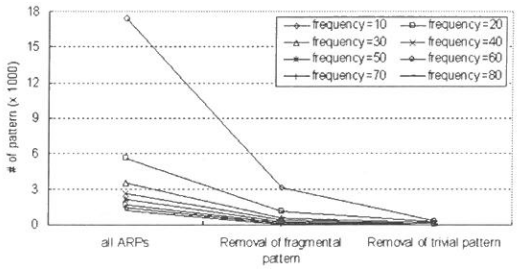


Fig. 5.10. The removal effect—dissimilarity ratio

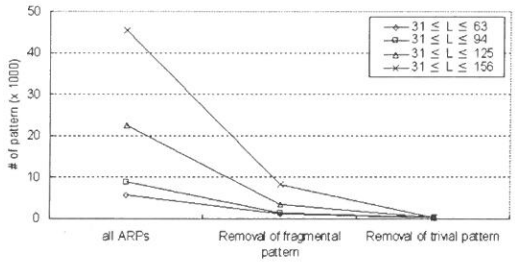


Fig. 5.11. The removal effect—the range of pattern length

6.結論與未來目標

我們提出一個 curve-based method，從音樂物件裡快速地找尋出重要而近似的重複樣型（nontrivial and approximate repeating pattern）。我們以 frame-based 方式從 melody 中擷取 interval sequence，pattern 與每一個 occurrence 會有相似的音樂輪廓（music contour），相似度表現在一 dissimilarity ratio 上。我們把資料量龐大音程序列 mapping 至低維度的 feature space 裡的點，一個點保有一段 substring 絕大部分的資訊，這使 curve-based method 能快速地先過濾絕不可能的 candidates。爲了進一步改善效果，我們刪除不符合樂理的 fragmental patterns，並提出依 patterns 的「長度」與「重現度」作排序，這有助於挑選出較少也較重要的 patterns。

Curve-based method 可以有效地解決主題在旋律上的變奏，但是尚無法解決節奏上的變奏，因爲我們以間隔固定的時間擷取一音高值，所以需要均勻的拉長或壓縮主題

的時間長度才可以找到主題的節奏變奏，未來的工作我們希望能利用 dynamic time warping 的技巧，來有效地解決此問題。

目前我們的實驗只展示時間效率與 patterns 數量，這些 patterns 之中有哪些是屬於樂曲的主題，也是我們關注的問題，未來我們將探討 curve-based method 在準確度上的效果。在本論文中，只用了少數的資料量，顯示我們初步的成果，我們正在執行更多的例子，整理成更完整的實驗結果。

此外，我們目前的方法是從 midi 擷取單音的資料，再以 curve-based method 找出 nontrivial and approximate repeating patterns，但是主題可能跨聲部地重現，這使我們的方法無法找到出現在不同聲部的變奏，所以未來的工作希望能直接處理複音的資料。

參考文獻

(1) Barthélemy, J., and A. Bonardi, "Figured Bass and Tonality Recognition," in *Proc. of 2nd Annual International Symposium on Music Information Retrieval (ISMIR)*, 2001.

(2) Dovey, Mathew J., "A Technique for Regular Expression Style Searching in Polyphonic Music," in *Proc. of 2nd Annual International Symposium on Music Information Retrieval (ISMIR)*, 2001.

(3) Faloutsos, C., "Fast Subsequence Matching in Time-Series Databases," in *Proc. of the ACM SIGMOD International Conference on Management of Data*, 1994.

(4) Faloutsos, C., *Searching Multimedia Databases by Content*. Kluwer Academic Publishers, 1996.

(5) Harada, L., "An Efficient Sliding Window Algorithm for Detection of Sequential Patterns," in *Proc. of Eighth International Conference on Database Systems for Advanced Applications (DASFAA '03)*, 2003.

(6) Hsu, Jia-Lien, Chih-Chin Liu, and Arbee L. P. Chen, "Discovering Nontrivial Repeating Patterns in Music Data," *IEEE Transactions on Multimedia*, Vol. 3, No. 3, September 2001.

(7) Hsu, Jia-Lien, Arbee L. P. Chen, and Hung-Chen Chen, "Finding Approximate Repeating Patterns from Sequence Data," in *Proc. of 5th International Conference on Music Information Retrieval (ISMIR)*, 2004.

(8) Kadota, T., Masahiro Hirao, Akira Ishino, Masayuki Takeda, Ayumi Shinohara,

Fumihiro Matsuo, "Musical Sequence Comparison for Melodic and Rhythmic Similarities," in *Proc. of Eighth International Symposium on String Processing and Information Retrieval (SPIRE)*, 2001.

(9) Kosugi, N., Yuichi Nishihara, Tetsuo Sakata, "A Practical Query-By-Humming System for Large Music Database," in *Proc. of the Eighth ACM International Conference on Multimedia*, 2000.

(10) Lartillot , O., "Discovering Musical Patterns through Perceptive Heuristics," in *Proc. of International Symposium on Music Information Retrieval (ISMIR)*, 2003.

(11) Liu, C. C., and Chuan-Sung Huang, "A Singer Identification Technique for Content-Based Classification of MP3 Music Objects," in *Proc. of the Conference on Information and Knowledge Management (CIKM)*, 2002.

(12) Logan, B., and S. Chu, "Music Summarization Using Key Phrase," in *Proc. of International Conference of Acoustic, Speech, Signal Processing*, 2000.

(13) Medina, Richard A., Lloyd A. Smith, and Deborah R. Wagner, "Content-based Indexing of Musical Scores," in *Proc. of the ACM/IEEE-CS Joint Conference on Digital Libraries*, 2003.

(14) Alan Victor Oppenheim and Ronald W. Schafer. *Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, N.J, 1975.

(15) Pienimäki, A., "Indexing Music Databases Using Automatic Extraction of Frequent Phrases," in *Proc. of Third International Conference on Music Information Retrieval (ISMIR)*, 2002 .

(16) Sun, X., E. Orlowska, and M. Xue Li, "Introducing Uncertainty into Pattern Discovery in Temporal Event Sequences," in *Proc. of Third IEEE International Conference on Data Mining (ICDM 2003)*, 2003.

(17) Xu, C., Yongwei Zhu, and Qi Tian, "Automatic Music Summarization based on Temporal,Spectral and Cepstral Features," in *Proc. of IEEE International Conference on Multimedia and Expo (ICME '02)*, 2002.

received October 21, 2004
revised December 08, 2004
accepted December 18, 2004

A Curve-based Algorithm for Finding Approximate Repeating Patterns from Music Data[3]

Yi-Chu Liu, and Jia-Lien Hsu[4]

Department of Computer Science and Information Engineering
Fu Jen Catholic University
Taipei, Taiwan 242, R.O.C.

Abstract

Finding approximate repeating patterns from music data is one of the key issues in the research field of music information retrieval. In this paper, we propose a curve-based algorithm to efficiently find nontrivial and approximate repeating patterns from music data. First, the given interval sequence is cut into interval substrings by the sliding window. By applying DCT on interval substrings, each substring will be mapped into the corresponding point in the feature space. Those points, which are near to each other in Euclidean distance, will be *self-joined* together into a *trial*. Therefore, similar trials correspond to similar interval substrings which are considered as candidates for approximate repeating patterns. By validation process, the final result set of nontrivial repeating patterns will be confirmed. Experiments are also performed to show the efficient and effectiveness of our approach.

Keywords: nontrivial and approximate repeating patterns, similar trails, DCT

---

[3] This research was partially sponsored by the Republic of China, National Science Council under Contract No. NSC-93-2213-E-030-005.

[4] Corresponding author. Tel: +886-2-29053817; fax: +886-2-29023550; E-mail: alien@csie.fju.edu.tw

具可擴充性之高效能 AES 加解密晶片之設計與實現

顏水鈿 高淑卿 呂學坤*
輔仁大學電子工程學系
台北縣新莊市中正路 510 號

摘 要

本篇論文中我們提出一個高效能且可擴充的 AES 加解密晶片之設計與實現，由於在整個 AES 硬體實現中影響執行速度及使用面積的最主要關鍵是 S-Box 的設計，因此我們利用模組化觀念將 S-BOX 之演算法轉換成基本模組，藉由基本模組的組合可產生三種型式的 S-Box (Type-I, Type-II 與 Type-III)，此三種型式的 S-Box 在執行速度及面積資源上各有優缺點，我們可依照實際應用來選擇適當的硬體架構，而整個 AES 晶片是由 0.18μm CMOS 製程實現，我們並與目前最常使用的查表法 (lookup table) 及其他相關的論文做比較，在我們所設計的 S-Box 中，Type-I 的硬體資源比查表法減少了 57%，執行的頻率爲 80 MHz，在 AES-128 的資料處理速度約爲 930 Mb/s，整個 AES 所使用的硬體資源的邏輯閘數爲 38.5K。

關鍵詞 ：Advanced Encryption Standard (AES)、S-Box、lookup table

1. 簡 介

在目前快速成長的網際網路及無線通訊上其使用的傳送頻道都爲開放式的，使用者對於資訊的安全越來越重視，其資料的傳送過程就顯得更重要，如何確保資料的安全呢? 最好的方式便是在資料傳送前將資料加密成爲密文 (Ciphertext)，接收後再將加密的密文解密成明文 (Plaintext)，目前加密及解密系統大致可分爲兩大類：對稱式 (symmetric) 及非稱式 (asymmetric) 密碼系統，對稱式密碼系統目前最常用的有 Data

*Corresponding author. Tel.:+886-2-29053800; Fax: +886-2-29042638
E-mail address : sklu@ee.fju.edu.tw

Encryption Standard (DES)， 3DES 及 Advanced Encryption Standard (AES) [1]，其傳送端與接收端是用相同的私鑰 (Private Key) 來做加密及解密，而非對稱密碼系統則有 Rivest-Shamir-Adleman (RSA) [2]及 Elliptic Curve [3] 等演算法，其傳送端與接收端是用不同的鑰匙來做加密及解密。

目前美國政府所採用的最新加解密標準爲進階加解密標準 (AES)，也是本篇論文所要討論的主題，它是利用 Rijndael [4] 演算法所推演出來的，其輸入資料方塊爲 128 位元，使用加解密鑰匙的長度可爲 128、192 或 256 位元，目前使用最廣泛的加解密鑰匙長度爲 128 位元，此加解密演算法被稱爲 AES-128，在本篇論文中我們針對 AES-128 做實現，AES-192 及 AES-256 演算法只有 RoundKey 的長度及執行的迴圈次數不同，在我們所提出的架構中也能直接擴充至不同的 RoundKey 長度及執行的迴圈以實現 AES-192 及 AES-256。

AES 加解密的實現方式有硬體實現及軟體實現兩種，而軟體的可靠度較差，故目前 AES 實現大部分都爲硬體實現，其中有的設計使用 FPGA [5-8] 架構，也有的設計成 ASIC [9-10] 型式，在整個 AES 硬體實現中影響執行速度及使用面積的最主要關鍵是 S-Box 的設計，因此我們針對 S-Box 的硬體實現做出可調整的機制，在執行速度及面積資源上依照實際應用來選擇適當的硬體。

在 S-Box 設計方面，我們利用模組化觀念將 S-BOX 之演算法轉換成兩種基本模組，藉由基本模組的組合可產生三種型式的 S-Box (Type-I, Type-II 與 Type-III)，此三種型式的 S-Box 在執行速度及面積資源上各有優缺點，我們可依照實際應用來選擇適當的硬體架構，最後再與目前最常使用的查表法及其他相關的論文做比較，而整個 AES 晶片是由 0.18μm CMOS 製程實現，我們所設計的 S-Box 中，Type-I 的硬體資源比查表法減少了 57%，執行的頻率爲 80 MHz，在 AES-128 的資料處理速度約爲 930 Mb/s，整個 AES 所使用的硬體資源的邏輯閘數爲 38.5K。

整篇論文的結構分爲六部分，在第二段我們簡單描述 AES 的演算法，第三段爲整個 AES 加解密的硬體實現，第四段爲 S-Box 的改善方法，第五段爲實驗結果，第六段爲結論。

2. AES 演算法

AES 的演算法 [1] 是利用多次的重複運算後達到加解密的效果，它是一個對稱式的方塊結構並以運算的回合做基礎，在此演算法中定義了 *state* 及 *Cipher Key* 二個名

詞。State (見圖 2.1) 的定義是在 AES 運算過程中所產生的暫存值，並以 *4* × *Nb* 爲矩陣大小，*Nb* 是在 *State* 中行的數目，而每一行爲 32 位元。*Cipher Key* 的定義是用來做加解密運算的 key，並以 *4* × *Nk* 爲矩陣大小，Nk 是由 key 的長度除以 32 位元而求得。其中，運算回合數 *(Nr)* 是由 *Nb* 及 *Nk* 所決定的 (表 2.1) ，其 AES 的 *cipher key* 共分爲 128、192 及 256 位元 (*Nk* = *4, 6* 或 *8*) 三種長度，並分別以爲 AES-128、AES-192 及 AES-256 爲簡稱。

$S_{0,0}$	$S_{0,1}$	$S_{0,2}$	$S_{0,3}$
$S_{1,0}$	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$
$S_{2,0}$	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$
$S_{3,0}$	$S_{3,1}$	$S_{3,2}$	$S_{3,3}$

圖 2.1 State 陣列表示法

表 2.1: Key-Block-Round 組合表

	Key Length (Nk words)	Block Size (Nb words)	Number of Rounds(Nr)
AES-128	4	4	10
AES-192	6	4	12
AES-256	8	4	14

在圖 2.2 中描述 AES 加解密的轉換步驟，加密包含四種不同的轉換過程- *SubBytes()* 、*ShiftRows()* 、*MixColumns()* 及 *AddRoundKey()* ，其重複執行的次數 (*Nr*) 由 cipher key 的長度所決定，值得注意的是在最後一個回合中並沒有執行 *MixColumns()* ，而解密的執行步驟正好與加密相反，只要執行四種不同的反轉換- *InvSubBytes()* 、*InvShiftRows()* 、*InvMixColumns()* 及 *AddRoundKey()* ，即可將密文轉換成明文，以下爲四種主要轉換的簡單描述 [1]：

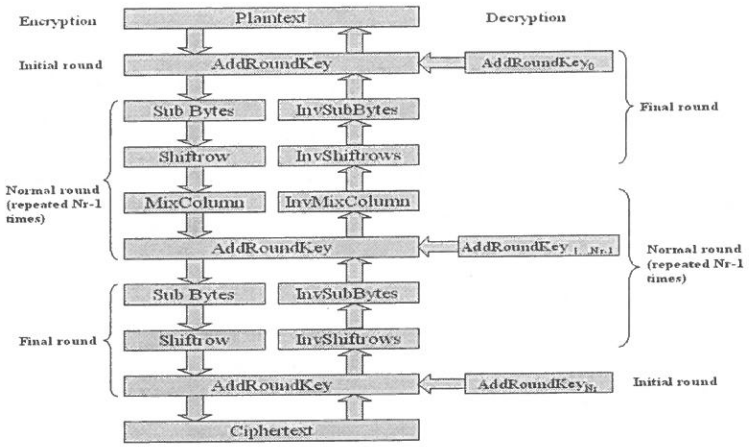


圖 2.2. AES 加解密的轉換步驟 (AES-128: *Nr = 10*, AES-192: *Nr = 12*, AES-256: *Nr = 14*) .

- 位元取代轉換 (SubBytes Transformation): 是一個以位元組爲單位的非線性取代運算，輸入的每一個位元組由相同的替代函數 (S-Box) 所取代，S-Box 是經過兩個運算過程而產生，並且是可逆的轉換，首先、第一個步驟先找出每個位元組在 $GF(2^8)$ 中的乘法反元素，並以不可化簡的多項式 $m(x) = x^8 + x^4 + x^3 + x + 1$ 爲模數，而 *{00}* 的乘法反元素並不運算，第二個步驟爲仿射 (Affine) 轉換其直接以矩陣的方式運算。
- 移列轉換 (ShiftRows Transformation): 是對 *State* 裡的每一列以不同的偏移量做環狀位移。第零列不動，第一列左移 C1 個位元組，第二列左移 C2 個位元組，第三列左移 C3 個位元組。位移的偏移量 C1、C2、C3 與區塊的數量 (*Nb*) 有關，定義如表 2.2 所示。

表 2.2 位移偏移量表

Nb	C1	C2	C3
4	1	2	3
6	1	2	3
8	1	3	4

- 混行轉換 (MixColumns Transformation): 在 State 陣列中以行對行的方式作運算。其中每個行視爲是一個存在 $GF(2^8)$ 中的多項式。對一固定的多項式 $c(x)$ 作乘法，若發生溢位時，則需再 *modulo* $x^4 + 1$。這裡的 $c(x) = \{03\}x^3 + \{01\}x^2 + \{01\}x + \{02\}$。

- AddRoundKey 轉換 (AddRoundKey Transformation) ：把混行轉換後的 State 與每一個 Round Key 做 bitwise XOR 運算然後存入 State 中 。在 round = 0 時直接與 initial Round Key 做 bitwise XOR 運算，而在 1≤ round ≤ Nr 時，AddRoundKey 轉換是與 Key Expansion 做 bitwise XOR 做運算。

在 AES 解密演算法中有四種轉換的運算過程: InvSubBytes()、InvShiftRows()、InvMixColumns()、及 AddRoundKey()，圖 2.2 描述 AES 解密的轉換步驟，以下將爲此四種主要反轉換做簡單描述 [1]：

- 位元取代反轉換 (InvSubBytes Transformation) ： 是位元組取代轉換(SubBytes Transformation) 的反運算，The inverse substitution table (S^{-1}-box) 是由二個運算過程所建立的。首先將執行仿射 (Affine) 轉換的反運算，然後再取位元組 $GF(2^8)$ 中的乘法反元素即可得到 S^{-1}-box。
- 移列反轉換 (InvShiftRows Transformation) ：是 ShiftRows 轉換的反運算，對 State 裡的每一列以不同的偏移量做環狀位移，第零列不動，第一列右移 *Nb - C1* 個位元組，第二列右移 Nb - C2 個位元組，第三列右移 *Nb - C3* 個位元組，定義如表 2.2 所示。
- 混行反轉換 (InvMixColumns Transformation) ：是 MixColumns 轉換的反運算，也是以 column-by-column 的方式作運算，其中每行視爲是一個存在 $GF(2^8)$ 中的多項式。對一固定的多項式 $c^{-1}(x)$ 作乘法，若發生溢位時，則需再 *modulo* $x^4 + 1$，這裡的 $c^{-1}(x) = \{0b\}x^3 + \{0d\}x^2 + \{09\}x + \{0e\}$。
- AddRoundKey 反轉換: 在解密過程中，直接取 AddRoundKey Transformation 的反運算。

RoundKey [1] 是從 Cipher Key 經過運算所產生的。由於演算法需要一個 *Nb* 個 words 的 initial set，而每個 *Nr* rounds 也需要 *Nb* 個 words，所以，Key Expansion 總共須產生 *Nb(Nr+1)*個 words。因此 Cipher Key 須展開爲擴充金鑰 (Expanded Key) 。擴充後的 Key 是一個 4-byte words 的線性陣列，表示爲 $[w_i]$，i 的範圍爲 $0 \le i < Nb(Nr+1)$。RoundKey 的反轉換與 RoundKey 轉換方式一樣，唯有送出子密鑰的順序不同。

3. AES 之硬體實現

圖 3.1 描述整個 AES 在硬體實現的路徑及架構，其中 Integrated (inv) SubBytes、Integrated (inv) MixColumns 整合加密及解密在同一方塊中共用其硬體資源，因此加密

及解密只須利用多工器切換，這裡以 En/De 信號線來控制，在加密 (En/De 爲 Low、Final 爲 Low) 步驟一開始 Initial/Loop 爲 High 時，先將輸入資料與 Initial RoundKey 做 XOR 運算後接著執行 SubBytes()、ShiftRow()、MixColumn()最後與 RoundKey 做 XOR 運算後第一個迴圈結束，然後進入第二個迴圈，此時 Initial/Loop 爲 Low，開始執行與第一個迴圈相同的步驟，一直執行到第十個迴圈時 Final 信號爲 High，將不執行 MixColumn()轉換，直接與 RoundKey 做 XOR 運算，此步驟結束爲資料加密眞正的完成，在解密的運算過程其步驟正好與加密步驟相反，以下小節將針對 Integrated (inv) SubBytes 、Integrated (inv) MixColumn 及 AddRoundKey 方塊做說明。

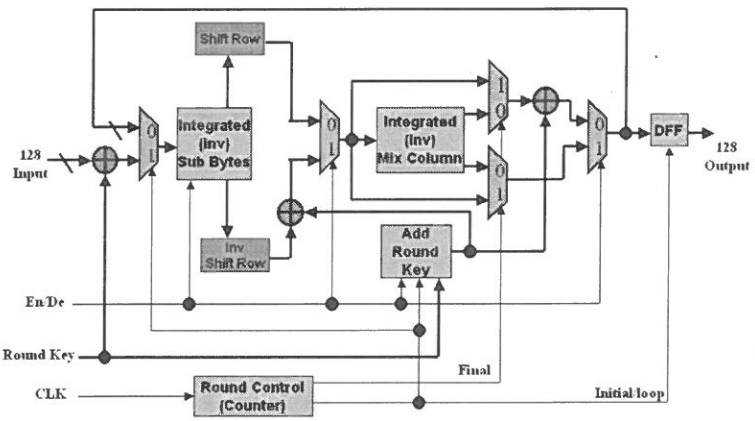


圖 3.1. 整合 AES 加密及解密之硬體架構

3.1 整合的 S-Box and Inverse S-Box 架構

S-Box 的運算先後順序爲乘法反元素 (Inverter) 然後仿射轉換，而解密使用的 Inverse S-Box 是先執行反仿射，然後才是乘法反元素轉換，它們使用的乘法反元素運算是相同的，所以利用多工器切換來共用其硬體資源，圖 3.2 爲結合 S-Box 與 Inverse S-Box 的硬體設計圖，當 En/De 信號爲 Low 時，執行加密的 S-Box 運算，此時 SubOut 輸出才是有效的輸出，反之解密的 Inverse S-Box 輸出則爲 InvSubOut 輸出，其中乘法反元素 GF (256) 將在下章節討論並且提出我們的設計方法。

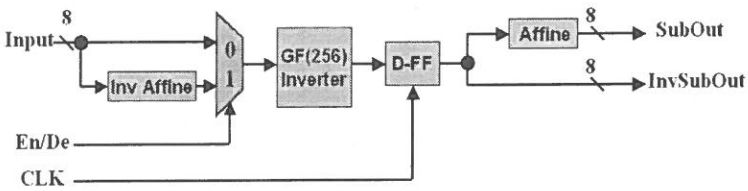


圖 3.2. 整合 S-Box 與 Inverse S-Box 硬體實現

3.2 整合的 MixColumn() and InvMixColumn()實現

在第二章節中我們發現加密及解密的混行轉換只有係數的差別，依據此特性我們可以結合 MixColumn()與 InvMixColumn()轉換來節省硬體空間，圖 3.3 爲 MixColumn() 與 InvMixColumn() 結合後的線路圖 [11]，圖中 Xtime 的運算式爲輸入 MSB(bit7) = 0 時將輸入左移一個位元，當輸入 MSB(bit7) = 1 時將輸入左移一個位元後與"1b"做 XOR 即可，此 Xtime()的布林方程式如下:

$$Out[7:0] = \{in[6:4], in[3:0] \wedge \{in[7], in[7], 1'b0, in[7]\}\} \tag{3.1}$$

在圖 3.4 中爲結合 4-bytes MixColumn() 與 InvMixColumn() 硬體線路圖，其中四個方塊爲結合後的 MixColumn() 與 InvMixColumn()，因轉換的係數不同所以其四個方塊的輸入信號順序亦不同，此輸入爲單一 Column 的 MixColumn 轉換，故在整個 AES 硬體實現中共需要十六個 1-bytes MixColumn() 與 InvMixColumn()的硬體。

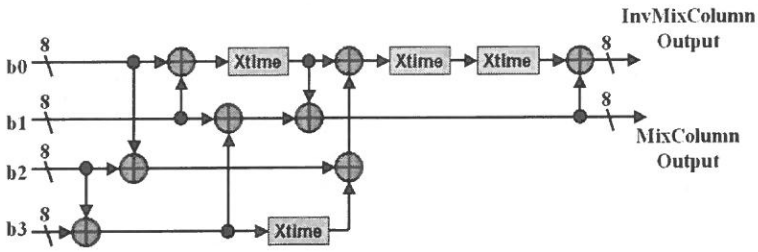


圖 3.3. 結合 1-bytes MixColumn() and InvMixColumn() 硬體線路圖

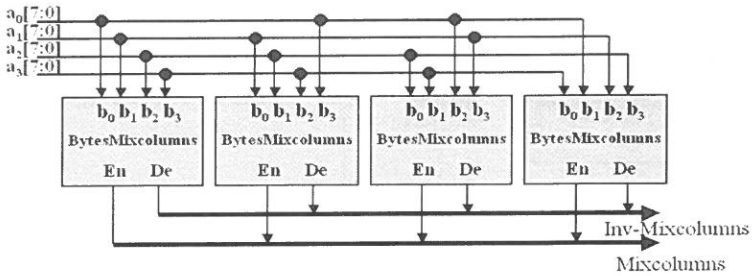


圖 3.4. 結合 4-bytes MixColumn() 與 InvMixColumn() 之架構

3.Key Expansion 的實現

依據 Key Expansion 的演算法我們依照步驟可以實現如圖 3.5 的硬體線路圖，其中 W0、W1、W2、W3 爲 32 位元的暫存器，在 SubBytes() 運算前必須將輸入資料左移 8 位元，然後在完成 SubBytes() 後與 Rcon[i] 做 XOR 運算，在此線路圖中已結合加密及解密的功能，只要控制 En/De 信號線即可執行加密或解密的 RoundKey 輸出。

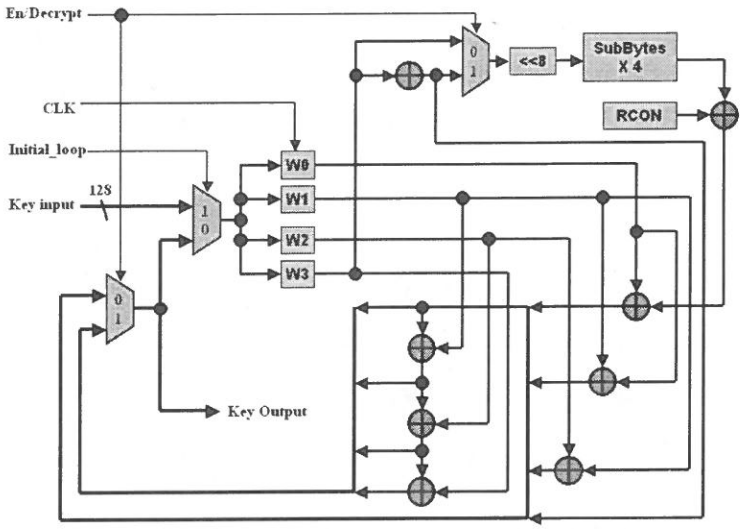


圖 3.5: AES-128 key 產生器

4. S-Box 架構設計

在 AES 整個演算法中，SubBytes() 及 Key Expansion 都會使用到 S-Box，且 S-Box

在整個演算法過程中相當複雜，它足以影響晶片的使用面積及執行速度，往往成爲研究者研究的重點。目前 S-Box 的硬體實現分爲兩類，第一類爲查表法，第二類爲直接算運法 (On the fly)，前者佔用的面積資源最大但是執行速度最快，反之後者的使用面積資源最少但執行速度較慢，如能將 S-Box 妥善處理，將能獲得最小的面積資源及最佳的處理速度。

在第二節中我們已經介紹過 S-Box 的運算分爲兩個步驟，其中的最爲關鍵的步驟爲乘法反元素的運算，以下我們簡單推導乘法反元素的數學公式:

$$A \cdot A^{-1} = 1 = A^{2m-1} = A \cdot A^{N} = A^{N+1},\ \text{N 爲正整數} \tag{4.1}$$

從 (4.1) 式中可得到 $N + 1 = 2^m - 1$，因此 $N = (2^{m-1} + 2^{m-2} + \cdots + 2^1)$，所以 A 的乘法反元素可被表示爲 $A^{-1} = (A^{2^{m-1}}) \cdot (A^{2^{m-2}}) \cdots (A^{2^1})$，在 AES 架構中 SubBytes() 以一個位元組爲單位，因此將 m 帶入 8 可得到 (4.2) 式，再將此式以硬體實現如圖 4.1 所示。

$$A^{-1} = A^{128} \bullet A^{64} \bullet A^{32} \bullet A^{16} \bullet A^{8} \bullet A^{4} \bullet A^{2} = A^{254} \tag{4.2}$$

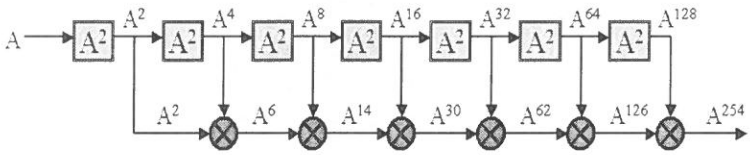


圖 4.1. 標準型乘法反元素的架構圖 [10]

在這裡使用的 A^2 或 $A \times B$ 必須以不可化簡的多項式 $m(x) = x^8 + x^4 + x^3 + x + 1$ 爲 modulo，其中 A^2 經計算後可以被化簡成如表 4.1 所示結果，而 $A \times B$ 爲一般的 8 位元並列乘法器 [13]。

表 4.1. $C = A^2$ 的係數

C_7	$a_7 + a_6$
C_6	$a_5 + a_3$
C_5	$a_6 + a_5$
C_4	$a_7 + a_4 + a_2$
C_3	$a_7 + a_6 + a_5 + a_4$
C_2	$a_5 + a_1$
C_1	$a_7 + a_6 + a_4$
C_0	$a_6 + a_4 + a_0$

在圖 4.1 中，我們得知其最長的 Critical Path 經過兩個 A^2 運算及六個乘法器，此方式已嚴重影響 S-Box 的執行速度，如果將它在每一級結束時加上 Pipeline 處理約可增加執行速度六倍，如 圖 4.2 所示架構圖，但是此方式將增加整個 AES 硬體暫存器的數量約 2240 ((16+4) × 16 × 7) 個，對 Throughput/Gate count 的增加並沒有太大助益。

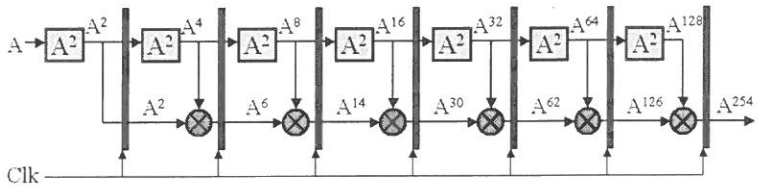


圖 4.2. 加上 Pipeline 的乘法反元素架構圖

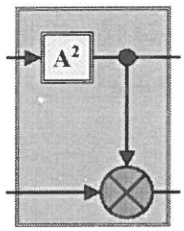


圖 4.3. 單一模組設計架構

在圖 4.2 中我們發現此硬體架構是有規則性及重複性的，故可以將此硬體做單一模組設計來減少硬體花費如圖 4.3 所示，此單一模組包含一個 A^2 和一個乘法器，而我們依乘法反元素架構加入單一模組設計成三個不同的型式以實現乘法反元素，再將三個不同的模組分別加入 AES 系統中實驗，可以獲得執行的速度及使用的面積，以下爲三種型式的介紹：

(1) Type-I

Type-I 是由一個單一模組所設計，如圖 4. 4 所示，其中在第一個輸入並沒有乘法器，而我們將它輸入爲 "1" 即接 VCC，我們以圖 4.5 做爲說明，整個運算是由 Initial/Loop 信號來控制初始值輸入或是內部迴圈，當 Initial/Loop 信號爲 High 時，單一模組接收外部輸入的初始值 A 後執行運算再將運算值存入暫存器，然後連續以 Clk 執行，在第八個 Clk 中產生 A^{254} 後以 Initial/Loop 信號上緣觸發擷取此乘法反元素，這裡增加的硬體有兩個八位元多工器及 24 個暫存器，而 Initial/Loop 信號是由簡單的三位元計數器及少許邏輯閘產生的，佔用的資源非常少，在 Critical Path 方面由原來的兩

個 A^2 運算及六個乘法器變成一個 A^2 運算、一個乘法器及一個多工器，而我們知道乘法器的運算時間是相當長的，所以我們不但減少硬體面積也增加了執行速度。

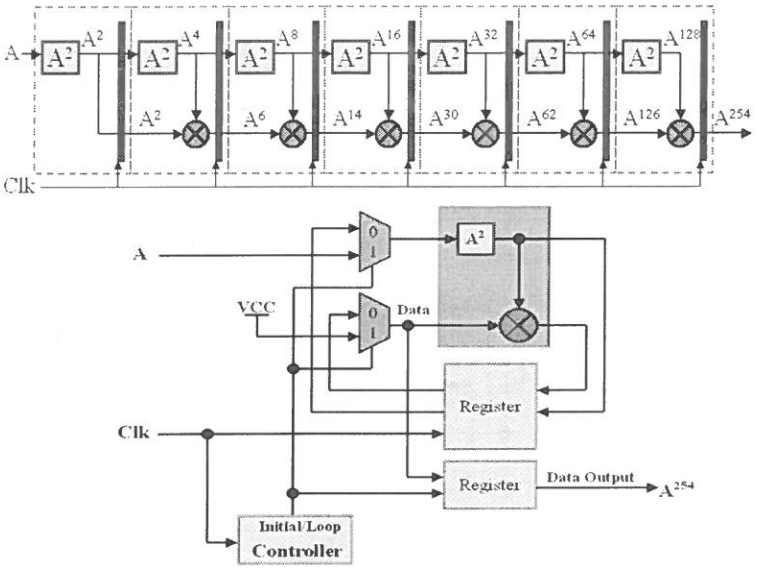


圖 4.4. Type-I 乘法反元素的架構圖

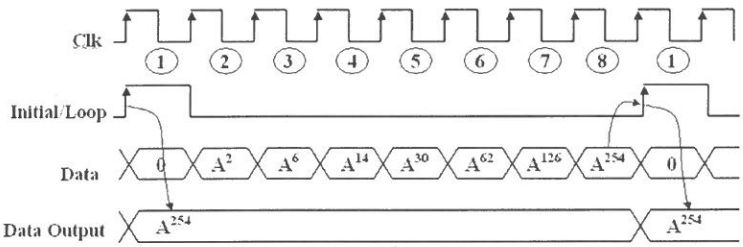


圖 4.5. Type-I 乘法反元素的波形圖

(2)Type-II

Type-II 是由二個單一模組所設計的如圖 4.6 所示，而第一個輸入先執行一個 A^2 後再進入單一模組，我們以圖 4.7 做爲說明，整個運算是由 Initial/Loop 信號來控制初始值輸入或是內部迴圈，當 Initial/Loop 信號爲 High 時，單一模組接收 A^2 運算後執行兩個單一模組的運算再將運算值存入暫存器，然後連續以 Clk 執行，在第四個 Clk 中產生 A^{254} 後以 Initial/Loop 信號上緣觸發擷取此乘法反元素，這裡增加的硬體有兩個八

位元多工器、24 個暫存器及一個 A^2 運算，而 Initial/Loop 信號是由簡單的兩位元計數器及少許邏輯閘產生的，在 Critical Path 方面由原來的兩個 A^2 運算及六個乘法器變成兩個 A^2 運算、兩個乘法器及一個多工器。

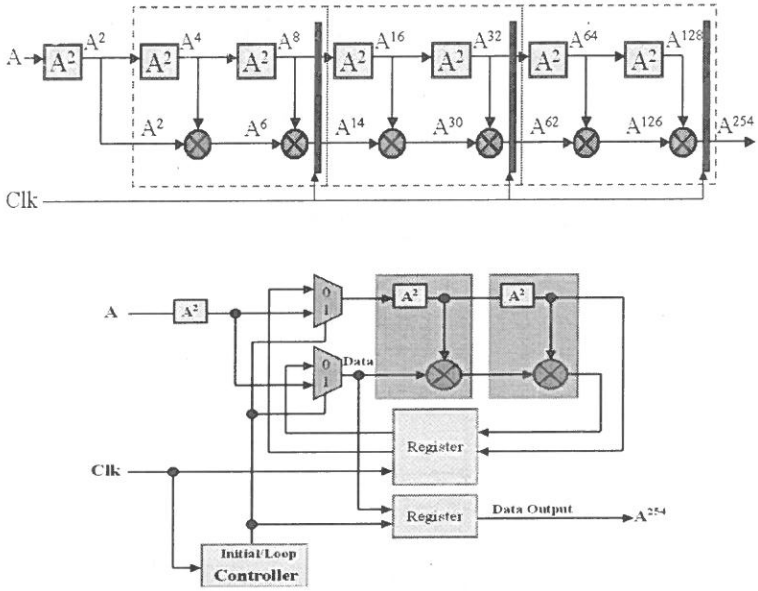


圖 4.6. Type-II 乘法反元素的架構圖

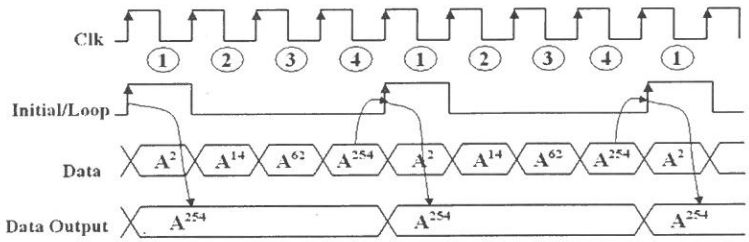


圖 4.7. Type-II 乘法反元素的波形圖

(3)Type-III

Type-III 是由三個單一模組所設計的如圖 4.8 所示，而第一個輸入先執行一個 A^2 後再進入單一模組，我們以圖 4.9 做爲說明，整個運算是由 Initial/Loop 信號來控制初始值輸入或是內部迴圈，當 Initial/Loop 信號爲 High 時，單一模組接收 A^2 運算後執行三個單一模組的運算再將運算值存入暫存器，然後連續以 Clk 執行，在第三個 Clk 中

產生 A^{254} 後以 Initial/Loop 信號上緣觸發擷取此乘法反元素，這裡增加的硬體有兩個八位元多工器、24 個暫存器及一個 A^2 運算，而 Initial/Loop 信號是由簡單的兩位元計數器及少許邏輯閘產生的，在 Critical Path 方面由原來的兩個 A^2 運算及六個乘法器變成兩個 A^2 運算、三個乘法器及一個多工器。

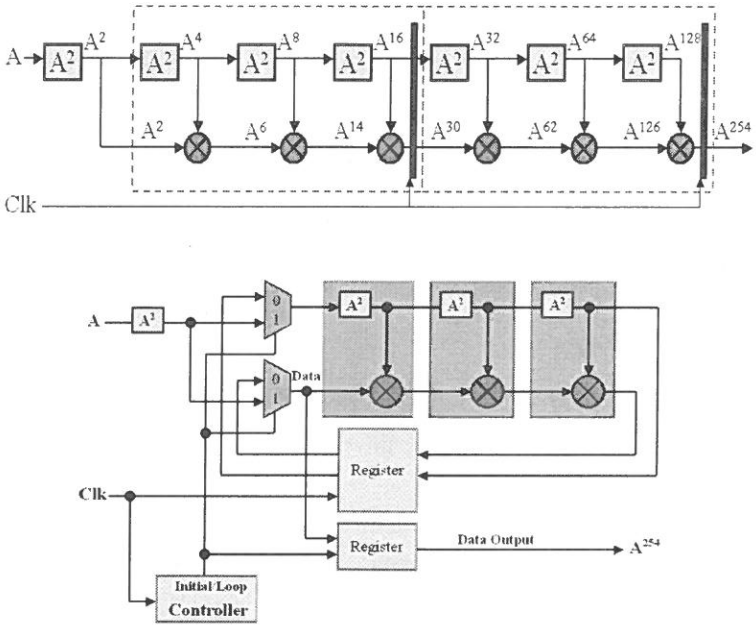


圖 4.8. Type-III 乘法反元素的架構圖

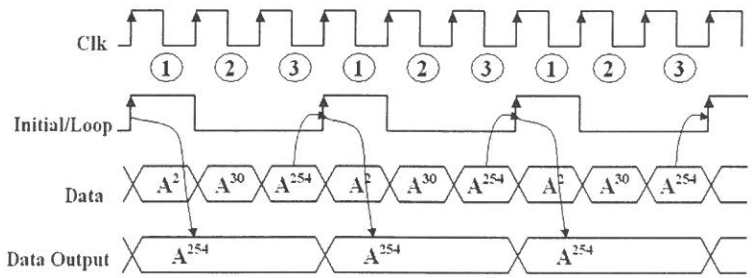


圖 4.9. Type-III 乘法反元素的波形圖

5. 實驗結果

我們設計的 AES 加解密系統是採用資料方塊爲 128 位元，金鑰長度爲 128 位元的架構，此 AES 晶片是使用 Verilog HDL 及標準 0.18μm CMOS Cell Library 製程實現。我們設計的三個 S-Box 中，型式一使用的面積最小約爲 571 gate counts 比 Look-up table 方式減少 57%，它在 AES chip 的工作頻率爲 80MHz，其 throughput 爲 930 Mb/s， 而整個 AES chip 的硬體花費爲 38.5K gate count。

表 5.1 爲 Type-I 與 LUT-based、SRAM-based、ROM-based、Inv-optional S-box 及台大 [14] 的 S-Box 做比較，而我們設計的 Type-I 比 Look-up table 方式減少 57%。表 5.2 爲提出的三種 S-Box 的 Gate count 與 Throughput 做比較，由於較少的硬體需重複運算的次數較多，但是執行的頻率也同時增加。表 5.3 爲目前已知設計的 AES chip 比較表，我們以 Type-I 來做比較。

由實驗結果得知，我們將 S-Box 增加 Pipeline 處理後並利用多工器做循環執行，可以減少使用面積的花費，在執行速度方面，由於加上 Pipeline 處理使得執行的頻率也同時提升。

表 5.1. S-box 比較表

	S-box + InvS-Box (LUT)	SRAM 256x8	ROM 2x256x8	Inv-optional S-box [11]	NTU[14]	Ours (Type-I)
Gate Count	662+662=1324	2138	1866	789	612	571

表 5.2. 三種型式之 S-box 比較表

	Type-I	Type-II	Type-III
Gate Count	571	948	1282
Throughput	930 Mb/s	465 Mb/s	232 Mb/s

表 5.3. AES chip 比較表

	Ichika[15]	Kuo[16]	Li[17]	Lu[11]	IBM[18]	NTU[14]	Ours (Type-I))
Process(um)	0.35	0.18	0.35	0.25	N/A	0.35	0.18
Clock(MHz)	N/A	N/A	166	100	32	70	80
Throughput	1.95Gb	1.82Gb	1.328Gb	609Mb	7.5Gb	831Mb	930Mb
Gate Count	612K	173K	120K	31.9K	256K	31.1K	38.5 K
Throughput /Gate Count	3.18	10.52	11.07	19.09	29.3	26.72	24.15

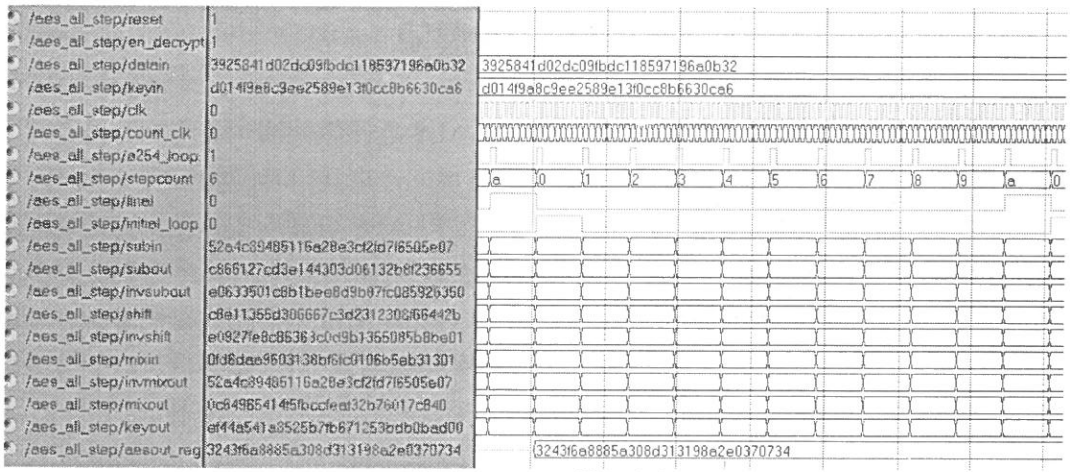


圖 5.1. Type-I 模擬時序圖

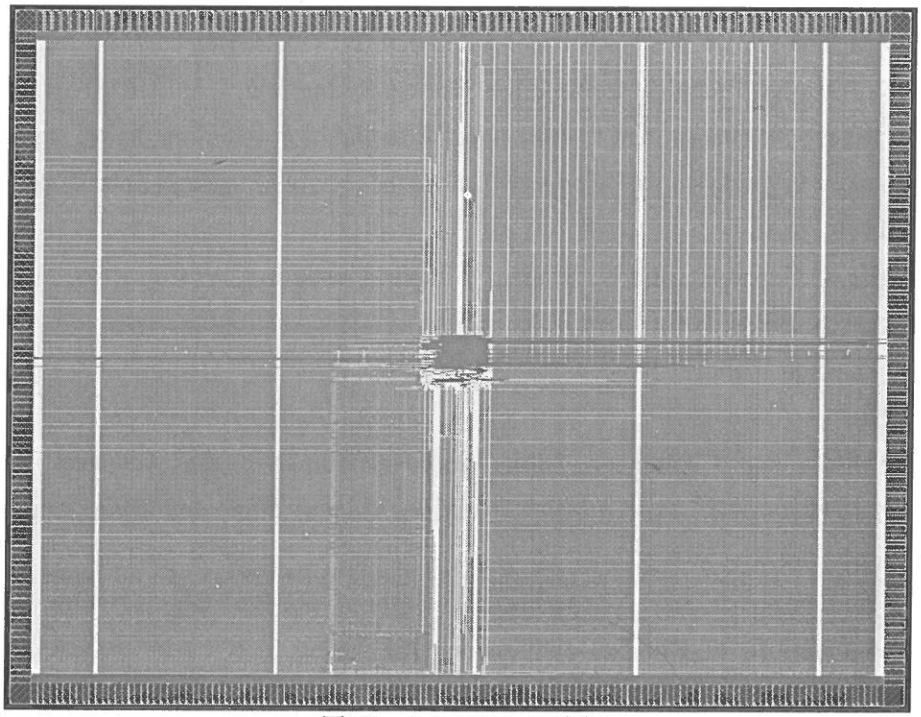


圖 5.2. Type-I Layout 圖

6. 結論與未來展望

在本篇論文中，我們提出的是一個可擴充式 AES 之硬體架構，其中的可擴充性是針對 S-Box 的設計，主要是利用運算的重複性及規則性來共用其硬體資源，達到降低成本的目的，我們將 S-Box 延伸爲三種型式，而使得 AES 可實現多種執行速度及面積的模式，在我們提出的三種 S-Box 中以 Type-I 的使用面積最小，其比查表法減少了 57%，在 AES-128 的資料處理速度約爲 930 Mb/s，整個 AES 所使用的硬體資源在標準合成約爲 38.5K gate count。在未來我們以 AMBA 架構結合 AES 做整合，以使其成爲可掛在 AMBA 上之 IP。

REFERENCE

(1) National Institute of Standards and Technology (NIST), *Advanced Encryption Standard (AES)*, National Technical Information Service, Springfield, VA 22161, Nov.2001

(2) R.L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120-126, Feb.1978.

(3) I.Blake, G.Seroussi, and N. Smart, "Elliptic Curve in Cryptography," Cambridge University Press, New York, NY, 1999.

(4) Joan Daemen, Vincent Rijmen, "*AES Proposal : Rijndael*," Document Version 2, Mar, 9, 1999.

(5) S.McMillan and C.Patterson, "Jbits Implementations of the Advanced Encryption Standard (Rijndael)," *Proc. 11th Int'l Conf. Field-Prog. Logic and Apps.*, Aug. 2001,vol. 2147, pp. 162-71.

(6) P. Chodowiec et al., "Experimental Testing of the Gigabit IPSes-Compliant Implementations of Rijndael and Triple DES Using SLAAC-1V FPGA Accelerator Board," *Proc. Info. Sec. Conf.*, Oct. 2001, LNCS, vol. 2200, pp. 220-34.

(7) K. U. Jarvinen, M. T. Tommiska, and J. O. Skytta," A fully Pipelined Memoryless 17.8Gb/s AES-128 Encryptor," *Proc. Int'l Symp. FPGA*, Monterey, CA, 2003, pp. 207-15.

(8) I. Verbauwhede, P. Schaumont, and H. Kuo, "Design and Performance Testing of a 2.29GB/s Rijndael Processor," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, Mar. 2003, pp.569-72.

(9) T. F. Lin et al.," A High-throughput Low-cost AES Cipher Chip," *Proc. 3rd IEEE Asia-Pacific ASIC Conf.*, Taipei, Taiwan, Aug. 2002, pp. 85-88.

(10) S. Mangard, M. Aigner, and S. Dominikus, "A Highly Regular and Scalable AES Hardware Architecture," *IEEE Trans.Comp.*, vol. 52, no.4, Apr. 2003, pp.483-91.

(11) C.C. Lu and S.Y. Tseng, "Integration of AES(Advanced Encryption Standard) encryptor and decrypter," in *Proc. Application-Specific System, Architecture and Processor*, pp. 277-285, 2002.

(12) C. C.Wang, T. K. Truong, H. M. Shao, L. J. Deutsch, J. K. Omura, and I. S. Reed," VLSI architecture for computing multiplications and inverses in $GF(2^m)$," *IEEE Trans. Computers*, vol. C-34, no.6, Aug. 1985.

(13) M. H. Jing, Y. W. Chen, Y. C. Chang, T. K. Troung," A new VLSI for Implementing the Multiplication and Inverse in the RS-Code," *4^{th} multimedia technology and applications*, Taiwan, Apr. 16, 1999, pp. 304.

(14) 王信中、林志修、吳安宇, "具成本效益的 AES 加密引擎之設計與實現," Master Thesis, Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan, pp51-60, 2003.

(15) T. Ichikawa, T. Kasuya and M. Matsui, "Hardware evaluation of the AES finalists," *Proc. 3rd AES Candidate Conf.*, 2000.

(16) H. Kuo and I. Verbauwhede, "Architectural optimization for a 1.82 Gbits/sec VLSI implementation of the AES Rijndael algorithm," *Proceedings CHES 2001*, Paris, France, May 2001, pp. 51.64.

(17) M. H. Li, "A Gbps AES cipher," Master Thesis, Dept. of Computer Science, National Tsing Hua University, Hsin-Chu, Taiwan, June 2001.

(18) A. Rudra, P. K. Dubey, C. S. Jutla, V. Kumar, J. R. Rao and P. Rohatgi, "Efficient implementation of Rijndael encryption with composite field arithmetic," *Proc. CHES 2001*, Paris, France, May 2001, pp. 171-184.

Design and Implementation of a Scalable High-Performance AES Cipher Chip

Shoei-Jea Yan、Chu-Ching Kao、Shyue-Kung Lu
Department of Electronic Engineering
Fu Jen Catholic University
510 Chung Cheng Rd, Hsinchuang, Taipei Hsien 24205 Taiwan

Abstract

A scalable high-performance AES cipher processor is proposed in this paper. Since the design and implementation of the S-Box is the critical step for implementing the entire AES encipher/decipher, a modularized methodology is adopted to implement it. According to the algorithm of the S-box, two types of basic modules can be derived. These two types of basic modules can be combined to form three types of S-box. The user can select among these types according to performance and area specifications. In order to verify our approach, an AES-128 processor is implemented with 0.18 μm CMOS technology. The gate count is 38.5 K. The operating frequency is 80 MHz and the throughput is 930 Mb/s. The gate count of our approach is 57% lower then the lookup table approach.

Keywords : Advanced Encryption Standard (AES)、S-Box、lookup table

綠藻 FJ03 培養基營養組成之研究

林泰宏 李雨薇 黃資螢 呂誌翼*
輔仁大學生命科學系

謝子陽
中國技術學院通識中心

林燕輝
中台醫護技術學院環境與安全衛生工程系

摘 要

本研究利用實驗設計 fractional factorial design 篩選出影響綠藻 FJ03 生長之主要營養因子，進一步藉由逼近法找出主要營養因子之最適合濃度範圍，最後確定培養基 BGII 各組成成份之濃度。由實驗結果顯示，增加 $MgSO_4 \cdot 7H_2O$ 濃度是明顯正向影響藻類 FJ03 生長之營養因子；而 $MgSO_4 \cdot 7H_2O$ 與 ferric ammonium citrate 對藻類 FJ03 生長有交互作用之影響，當 $MgSO_4 \cdot 7H_2O$ 採用較高濃度(1.0 g/l)，ferric ammonium citrate 使用高濃度(0.02 g/l)時，可以促使藻類 FJ03 有較高的生長速率，相對地，當 $MgSO_4 \cdot 7H_2O$ 採用較低濃度(0.05 g/l)，ferric ammonium citrate 使用高濃度時，卻對藻類 FJ03 生長有抑制作用。逼近法的結果顯示，培養基 BGII 中 $MgSO_4 \cdot 7H_2O$ 濃度爲 1.5~2.5 g/l 以及 ferric ammonium citrate 濃度爲 0.15~0.35 g/l 時，綠藻 FJ03 具有最高之比生長速率。綜合研究結果，藻類 FJ03 生長之最適化培養基各成份爲 $NaNO_3$ 0.5 g/l、K_2HPO_4 0.06 g/l、$MgSO_4 \cdot 7H_2O$ 2.0 g/l、$CaCl_2 \cdot 2H_2O$ 0.06 g/l、citric acid 0.06 g/l、ferric ammonium citrate 0.25 g/l、EDTA 0.001 g/l、Na_2CO_3 0.02 g/l 和 trace metal solution 1 ml。

關鍵詞：二氧化碳固定、藻類、最適化

* Corresponding author. Tel: +886-2-29052120; Fax: +886-2-29052193
E-mail: yihleu@yahoo.com.tw

前　言

近年來藻類在生技產業上的應用已逐漸受到重視，藻類的應用非常廣泛，包括從最簡單整個細胞體作爲食用或動物飼料用，甚至醫藥工業上所需之特殊天然物的提供，藻類產品的市場顯著逐年擴大，根據調查近幾年全球藻類生物技術產品每年的產值大約有 6×10^9 美元（Pulz and Gross, 2004）。

綠藻是生技產業傳統上使用的藻類，目前大部份產業仍以室外培養方式或反應槽中混營培養方式做大規模的養殖，主要作爲水產生物餌料（De Pauw and Persoone, 1988；Laing and Ayala, 1990），或用於萃取高價值化合物，例如綠藻生長因子、免疫調節因子（Belay, 1993；Osinga *et al.*, 1999）、蝦紅素（Lorenz and Cysewski, 2000）、β-胡蘿蔔素（Borowitzka, 1998）、多元不飽和脂肪酸（Hayashi *et al.*, 2001）、具有生理活性化合物（Muller-Feuga, 2003）等。由於工業發展快速進步，近年來全球溫室效應問題持續嚴重，環境永續發展已爲人類所重視，研究人員亦計畫如何利用綠藻來固定工業排放過多之溫室氣體二氧化碳，將工業來源之二氧化碳做爲營養源利用來培養大量的綠藻細胞（Hirata *et al.*, 1996；Kurano *et al.*, 1995），同時減少地球大氣環境中二氧化碳的量。

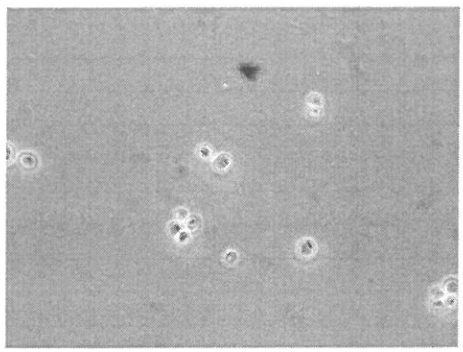
由以上敘述可知綠藻細胞的利用性很多，但是如何能大量且高密度的培養綠藻是生技產業應用上一個很重要的議題（Riesenberg and Guthke, 1999），而解決藻類大量且高密度培養問題，最開始可從培養基的設計與其組成成份及濃度著手。培養基中的營養成份及其使用濃度深深影響到藻類的生長速率，例如藻類生長時所需之主要營養素中的氮爲蛋白質、核酸、葉綠素和某些輔酶之關鍵元素；磷爲能量代謝所需的元素；鎂影響到光合作用時利用二氧化碳的葉綠體合成和 carbonic anhydrase 酵素的功能；鈣爲細胞組織結構所需要；鐵與電子轉移時之電子攜帶蛋白(如硫鐵蛋白)有關；氯化鈉與細胞生長之滲透壓及離子運輸有關。

本實驗室從事藻類研究多年，2004 年 6 月由輔仁大學荷花池中篩選出容易培養且生長速率高的綠藻 FJ03，綠藻 FJ03 的特性適合用於固定二氧化碳的研究和應用。爲了提高綠藻 FJ03 培養時之生長速率，進一步應用於固定二氧化碳之研究，本研究針對綠藻 FJ03 之培養基 BGII 進行其組成成份的探討，利用實驗設計方法 factional factorial design 找尋出影響綠藻 FJ03 生長之主要營養因子，接著探討出綠藻 FJ03 有較佳生長速率時培養基 BGII 之組成成份最適合之濃度範圍。

材料與方法

實驗用微藻類及培養基

本研究用爲二氧化碳固定培養之微藻類爲綠藻 FJ03（圖一），綠藻 FJ03 於 2004 年 6 月分離自輔仁大學荷花池，爲球形細胞，具有杯狀葉綠體構造，非常容易培養，生長速率快，因此適合作爲二氧化碳固定研究之藻類。FJ03 所使用之培養基爲 BGII，所含之營養成份爲 $NaNO_3$ 1.5 g/l、K_2HPO_4 0.04 g/l、$MgSO_4•7H_2O$ 0.075 g/l、$CaCl_2•2H_2O$ 0.036 g/l、citric acid 0.006 g/l、ferric ammonium citrate 0.006 g/l、EDTA 0.001 g/l、Na_2CO_3 0.02 g/l 及 trace metal solution 1ml；其中 trace metal solution 所含之成份爲 $FeCl_3•6H_2O$ 3.150 g/l、$MnCl_2•4H_2O$ 0.180 g/l、$ZnSO_4•7H_2O$ 0.022 g/l、$CoCl_2•6H_2O$ 0.010 g/l、$CuSO_4•5H_2O$ 0.010 g/l 和 $Na_2MoO4•2H_2O$ 0.006 g/l。綠藻生長情形之測量乃利用分光光度計（spectrophotometer）以波長 600 nm 量測其吸光值（optic density，OD_{600}）。



圖一　本研究所使用之綠藻 FJ03。

影響藻類固定二氧化碳生長之主要營養因子分析

光合作用固定二氧化碳之藻類生長受到培養基成份及其濃度影響很大，爲了瞭解培養基成份及其濃度之影響程度，並找出主要影響營養因子較佳之濃度範圍，本研究利用實驗設計的方法進行探討。本研究採用 fractional factorial design 探討氯化鈉（NaCl）、氮源（$NaNO_3$）、鐵（ferric ammonium citrate）、鎂（$MgSO_4•7H_2O$）、磷（K_2HPO_4）和鈣（$CaCl_2•2H_2O$）等 6 個因子對藻類生長之影響，鑑定出哪幾個因子爲主要影響藻類生長之主要因子。二階層 6 個因子之 fractional factorial design 將進行 32

表一 二階層 6 個因子之 fractional factorial design 之 32 組試驗。

試驗因子 / 實驗組	$NaNO_3$	K_2HPO_4	$MgSO_4 \cdot 7H_2O$	$CaCl_2 \cdot 2H_2O$	ferric ammonium citrate	NaCl
1	-	-	-	-	-	-
2	-	-	-	-	+	+
3	-	-	-	+	-	+
4	-	-	-	+	+	-
5	-	-	+	-	-	+
6	-	-	+	-	+	-
7	-	-	+	+	-	-
8	-	-	+	+	+	+
9	-	+	-	-	-	+
10	-	+	-	-	+	-
11	-	+	-	+	-	-
12	-	+	-	+	+	+
13	-	+	+	-	-	-
14	-	+	+	-	+	+
15	-	+	+	+	-	+
16	-	+	+	+	+	-
17	+	-	-	-	-	+
18	+	-	-	-	+	-
19	+	-	-	+	-	-
20	+	-	-	+	+	+
21	+	-	+	-	-	-
22	+	-	+	-	+	+
23	+	-	+	+	-	+
24	+	-	+	+	+	-
25	+	+	-	-	-	-
26	+.	+	-	-	+	+
27	+	+	-	+	-	+
28	+	+	-	+	+	-
29	+	+	+	-	-	+
30	+	+	+	-	+	-
31	+	+	+	+	-	-
32	+	+	+	+	+	+

組試驗，32 組試驗之設計如表一所示。表一中之 6 個試驗因子即爲本此實驗所選定之 6 個營養因子，包括 NaCl、$NaNO_3$、ferric ammonium citrate、$MgSO_4 \cdot 7H_2O$、K_2HPO_4 和 $CaCl_2 \cdot 2H_2O$，而表中之"+" 和 "-" 分別爲該營養因子之較高濃度值和較低濃度值，各因子所設計之濃度範圍爲 $NaNO_3$ 之 "+" 爲 3.0 g/l；"-" 爲 0.5 g/l。K_2HPO_4 之 "+" 爲

0.06 g/l；"-" 爲 0.01 g/l。$MgSO_4•7H_2O$ 之 "+" 爲 1.0 g/l；"-" 爲 0.05 g/l。$CaCl_2•2H_2O$ 之 "+" 爲 0.06 g/l；"-" 爲 0.01 g/l。ferric ammonium citrate 之 "+" 爲 0.02 g/l；"-" 爲 0.002 g/l。NaCl 之 "+" 爲 5.0 g/l；"-" 爲 0.5 /l。此 32 組實驗結果利用統計電腦軟體 Minitab（Minitab Inc., 3081 Enterprise, State Colledge, PA 16801, USA）之 one-way 和 two-way Analysis Of Variance（ANOVA）進行分析，*P* 值小於 0.05 視爲統計上有顯著性，鑑定出影響藻類固定二氧化碳效率之主要因子。進一步利用逼近法針對主要影響營養因子進行較適合使用濃度之實驗，找出此主要營養因子之較適合濃度值範圍。

結果與討論

影響綠藻 FJ03 生長之主要營養因子分析

本研究採用二階式之 fractional factorial design 探討 $NaNO_3$、K_2HPO_4、$MgSO_4•7H_2O$、$CaCl_2•2H_2O$、ferric ammonium citrate 和 NaCl 等 6 個營養因子對綠藻 FJ03 生長之影響。各實驗組綠藻 FJ03 之比生長速率結果列於表二，實驗數據經由 ANOVA 分析結果如表三所示，分析結果顯示在探討之 6 個營養因子中，只有 $MgSO_4•7H_2O$ 顯著影響藻類 FJ03 生長，並且是濃度愈高藻類比生長速率越快。Price 等人對於 *Synechococcus* PCC7942 之 carboxysomes 中 carbonic anhydrase 活性研究指出，20 mM 鎂離子（Mg^{2+}）使 carbonic anhydrase 有最高的活性（Price *et al.*, 1992）。藻類是光合性生物，能固定氣體二氧化碳以生長繁殖，氣體二氧化碳在藻類細胞質中會形成碳酸氫根（HCO_3^-），碳酸氫根必須進入 carboxysomes 中經由酵素 carbonic anhydrase 催化成爲氣體二氧化碳，氣體二氧化碳才進一步爲藻類細胞所利用。因此 $MgSO_4•7H_2O$ 爲影響綠藻 FJ03 生長之主要因子，其中原因可能與影響 carbonic anhydrase 活性表現有關。

另外，$MgSO_4•7H_2O$ 與 ferric ammonium citrate 對藻類 FJ03 生長有交互作用之影響，其影響情形如圖二所示，由圖二結果明顯發現當 $MgSO_4•7H_2O$ 採用較高濃度，ferric ammonium citrate 使用高濃度時，可以促使藻類 FJ03 有較高的比生長速率；相對地，當 $MgSO_4•7H_2O$ 採用較低濃度，ferric ammonium citrate 使用高濃度時，卻對藻類 FJ03 生長有抑制作用。$MgSO_4•7H_2O$ 與 ferric ammonium citrate 對藻類生長之交互影響作用極少人進行探討過，經過文獻搜尋並沒有發現類似之研究結果。鎂離子對藻類生長之影響已於上述討論過，而有關鐵離子應用於增加藻類生長繁殖之研究，近年來很

表二　使用 fractional factorial design 探討影響綠藻 FJ03 生長之主要營養因子各實驗組結果。

試驗因子 / 實驗組	$NaNO_3$	K_2HPO_4	$MgSO_4 \cdot 7H_2O$	$CaCl_2 \cdot 2H_2O$	ferric ammonium citrate	NaCl	比生長速率(μ_{max}) (h^{-1})
1	-	-	-	-	-	-	0.00475
2	-	-	-	-	+	+	0.00437
3	-	-	-	+	-	+	0.00418
4	-	-	-	+	+	-	0.00458
5	-	-	+	-	-	+	0.00424
6	-	-	+	-	+	-	0.00531
7	-	-	+	+	-	-	0.00377
8	-	-	+	+	+	+	0.00529
9	-	+	-	-	-	+	0.00350
10	-	+	-	-	+	-	0.00305
11	-	+	-	+	-	-	0.00550
12	-	+	-	+	+	+	0.00437
13	-	+	+	-	-	-	0.00467
14	-	+	+	-	+	+	0.00568
15	-	+	+	+	-	+	0.00534
16	-	+	+	+	+	-	0.00575
17	+	-	-	-	-	+	0.00370
18	+	-	-	-	+	-	0.00420
19	+	-	-	+	-	-	0.00399
20	+	-	-	+	+	+	0.00341
21	+	-	+	-	-	-	0.00425
22	+	-	+	-	+	+	0.00451
23	+	-	+	+	-	+	0.00480
24	+	-	+	+	+	-	0.00486
25	+	+	-	-	-	-	0.00444
26	+	+	-	-	+	+	0.00405
27	+	+	-	+	-	+	0.00506
28	+	+	-	+	+	-	0.00474
29	+	+	+	-	-	+	0.00456
30	+	+	+	-	+	-	0.0055
31	+	+	+	+	-	-	0.00426
32	+	+	+	+	+	+	0.00493

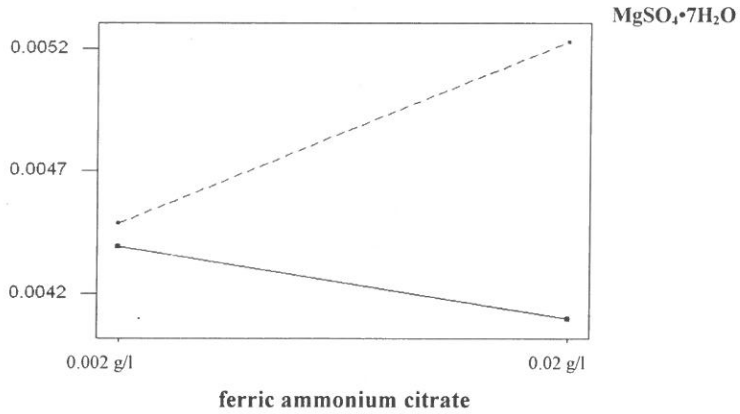
註：本實驗共進行 126 個小時，比生長速率(μ_{max})=[(126 小時之 OD_{600} 值) – (0 小時之 OD_{600} 值)] /126。

大部份著重於將鐵離子播施於海洋或水域中，探討水體中藻類生長增殖的情形或營養源的利用情形（Buesseler *et al.*, 2004；Coale *et al.*, 2004；Strzepek and Harrison, 2004；Tsuda *et al.*, 2003），這些研究中皆指出增加水體中鐵離子會增加藻類之增殖率及其對營養源之利用率。

表三　綠藻 FJ03 固定二氧化碳生長之 fractional factorial design 影響因子之 *P* 值。

影響因子	*P* 值
$NaNO_3$	0.427
K_2HPO_4	0.177
$MgSO_4•7H_2O$	0.008*
$CaCl_2•2H_2O$	0.296
ferric ammonium citrate	0.355
NaCl	0.676
$MgSO_4•7H_2O$ × ferric ammonium citrate	0.013*

*表示有顯著差異



圖二　培養基中 $MgSO_4•7H_2O$ 與 ferric ammonium citrate 對藻類 FJ03 比生長速率之交互作用影響。實線（—）：$MgSO_4•7H_2O$ 採用 0.05 g/l；虛線（- - -）：$MgSO_4•7H_2O$ 採用 1.0 g/l。

由此實驗之結果可以推論出影響藻類 FJ03 生長固定二氧化碳的主要營養因子爲 $MgSO_4•7H_2O$，次要因子爲 $MgSO_4•7H_2O$ 與 ferric ammonium citrate 之交互作用（$MgSO_4•7H_2O$ ×ferric ammonium citrate），並且可推導出進一步研究藻類 FJ03 生長培養之培養基各營養成份的濃度（表四；ANOVA 分析各營養因子之正向或負向影響未列於本文中），以便進一步用爲探討藻類 FJ03 生長之 $MgSO_4•7H_2O$ 與 ferric ammonium citrate 較適合之濃度範圍。

綠藻 FJ03 生長之培養基成份較適合之濃度

綠藻 FJ03 生長固定二氧化碳之培養基成份較適合濃度之研究，實驗時所採用之培養基如表四所示，本實驗利用漸進法找出綠藻 FJ03 之較大比生長速率範圍，藉以推導

表四　綠藻 FJ03 進行最適化生長培養研究使用之修正培養基各成份濃度。

成份	原來培養基的成份濃度（g/l）	修正後培養基的成份濃度（g/l）
$NaNO_3$	1.5	1.5
K_2HPO_4	0.04	0.06
$MgSO_4•7H_2O$	0.075	進一步試驗
$CaCl_2•2H_2O$	0.036	0.06
Citric acid	0.006	0.06
ferric ammonium citrate	0.006	進一步試驗
EDTA（Na_2Mg salt）	0.001	0.001
Na_2CO_3	0.02	0.02
trace metal solution	1 ml	1 ml
deionized water	1000 ml	1000 ml
pH	7.4	7.4

表五　最適化生長培養研究中，培養基中 MgSO4•7H2O 與 ferric ammonium citrate 濃度對藻類 FJ03 比生長速率之影響。

		比生長速率（μ_{max}）（h^{-1}）					
$MgSO_4•7H_2O$ (g/l)	**3.0**	0.0072	**0.0086**	**0.0086**	**0.0081**	**0.0087**	0.0069
	2.5	0.0067	**0.0088**	**0.0085**	**0.0086**	**0.0084**	**0.0089**
	2.0	0.0048	0.0073	**0.0088**	**0.0088**	**0.0089**	**0.0086**
	1.5	0.0032	0.0078	**0.0088**	**0.0087**	**0.0085**	0.0078
	1.0	0.0058	0.0078	**0.0085**	0.0079	0.0072	0.0072
	0.5	0.0065	**0.0081**	0.0073	0.0075	0.0075	0.0071
		0.02	**0.08**	**0.15**	**0.25**	**0.35**	**0.50**
		ferric ammonium citrate (g/l)					

出生長培養時最適化培養基中 $MgSO_4•7H_2O$ 與 ferric ammonium citrate 較佳濃度範圍。進行研究時 $MgSO_4•7H_2O$ 之實驗濃度範圍分別爲 0.5、1.0、1.5、2.0、2.5 和 3.0 g/l；ferric ammonium citrate 之實驗濃度範圍分別爲 0.02、0.08、0.15、0.25、0.35 和 0.50 g/l，實驗結果發現綠藻 FJ03 之較高比生長速率範圍位在於 $MgSO_4•7H_2O$ 濃度 1.5~2.5 g/l，ferric ammonium citrate 濃度 0.15~0.35 g/l（表五）。由以上結果可以得到一個結論：綠藻 FJ03 生長培養及固定二氧化碳之最適化培養基各營養成份爲 $NaNO_3$ 0.5 g/l、K_2HPO_4 0.06 g/l、$MgSO_4•7H_2O$ 2.0 g/l、$CaCl_2•2H_2O$ 0.06 g/l、citric acid 0.06 g/l、ferric ammonium citrate 0.25 g/l、EDTA 0.001 /l、Na_2CO_3 0.02 g/l 和 trace metal solution 1 ml。

結　論

藻類生長時會固定二氧化碳，是光合性生物，藻類生長之速率明顯受到培養基中各組成成份及其濃度之影響。本研究利用實驗設計方法 factional factorial design 進行綠藻 FJ03 培養時所用培養基 BGII 之改良，以提高藻類生長速率，增加藻類生質體，實驗結果證明利用 factional factorial design 確實能發現影響綠藻 FJ03 生長之主要營養因子，並進一步確定培養基 BGII 中各組成成份之最佳濃度。在本研究中所得到的結論如下所述：

1.本研究所探討之 6 個營養因子中，只有 $MgSO_4 \cdot 7H_2O$ 顯著影響藻類 FJ03 生長，並且其濃度愈高藻類比生長速率越高。

2.$MgSO_4 \cdot 7H_2O$ 與 ferric ammonium citrate 對藻類 FJ03 生長有交互作用之影響，結果明顯發現當 $MgSO_4 \cdot 7H_2O$ 採用較高濃度，ferric ammonium citrate 使用高濃度時，可以促使藻類 FJ03 有較高的比生長速率；相對地，當 $MgSO_4 \cdot 7H_2O$ 採用較低濃度，ferric ammonium citrate 使用高濃度時，卻對藻類 FJ03 生長有抑制作用（圖二）。

3.綠藻 FJ03 生長及固定二氧化碳之最適化培養基各組成成份爲 $NaNO_3$ 0.5 g/l、K_2HPO_4 0.06 g/l、$MgSO_4 \cdot 7H_2O$ 2.0 g/l、$CaCl_2 \cdot 2H_2O$ 0.06 g/l、citric acid 0.06 g/l、ferric ammonium citrate 0.25 g/l、EDTA 0.001 /l、Na_2CO_3 0.02 g/l 和 trace metal solution 1 ml。

誌　謝

本研究非常感謝中華民國行政院環保署之九十三年環保科技創新育成中心計畫經費補助，以及人宇生物科技股份有限公司在研究上之技術討論協助。

參考文獻

(1) A. Belay, "Current knowledge on potential health benefits of *Spirulina platensis*", *J. Appl. Phycol.*, 5: 235–240, 1993.

(2) M. A. Borowitzka, "Company news", *J. Appl. Phycol.*, 10: 417, 1998.

(3) K. O. Buesseler, J. E. Andrews, S. M. Pike, and M. A. Charette, "The effects of iron fertilization on carbon sequestration in the Southern Ocean", *Science*, 304 (5669): 396-

397, 2004.

(4) K. H. Coale, K. S. Johnson, F. P. Chavez, K. O. Buesseler, R. T. Barber, M. A. Brzezinski, W. P. Cochlan, F. J. Millero, P.G. Falkowski, J. E. Bauer, R. H. Wanninkhof, R. M. Kudela, M. A. Altabet, B. E. Hales, T. Takahashi, M. R. Landry, R. R. Bidigare, X. Wang, Z. Chase, P. G. Strutton, G. E. Friederich, M. Y. Gorbunov, V. P. Lance, A. K. Hilting, M. R. Hiscock, M. Demarest, W. T. Hiscock, K. F. Sullivan, S. J. Tanner, R. M. Gordon, C. N. Hunter, V A. Elrod, S. E. Fitzwater, J. L. Jones, S. Tozzi, M. Koblizek, A. E. Roberts, J. Herndon, J. Brewster, N. Ladizinsky, G. Smith, D. Cooper, D. Timothy, S. L. Brown, K. E. Selph, C. C. Sheridan, B. S. Twining, and Z. I. Johnson, "Southern Ocean iron enrichment experiment: carbon cycling in high- and low-Si waters", *Science*, 304 (5669): 396-397, 2004.

(5) N. De Pauw and G. Persoone, "Micro-algae for aquaculture", In: Microalgal biotechnology, M. A. Borowitzka and L. J. Borowitzka (eds), Cambridge University Press, Cambridge, pp 197–221, 1988.

(6) M. Hayashi, T. Yukino, I. Maruyama, S. Kido, and S. Kitaoka, "Uptake and accumulation of exogenous docosahexaenoic acid by *Chlorella*", *Biosci. Biotechnol. Biochem.*, 65 (1): 202-204, 2001.

(7) S. Hirata, M. Hayashitani, M. Taya, and S. Tone, "Carbon dioxide fixation in batch culture of *Chlorella* sp. using a photobioreactor with a sunlight-collection device", *J. Fermen. Bioeng.*, 81 (5): 470-472, 1996.

(8) N. Kurano, H. Ikemoto, H. Miyashiya, T. Hasegawa, and S. Miyachi, "Carbon dioxide uptake rate of *Chlorococcum littorale*", *J. Mar. Biotechnol.*, 3: 108-110, 1995.

(9) A. Muller-Feuga, J. Moal, and R. Kaas, "The microalgae for aquaculture", In: Life feeds in marine aquaculture, J. G. Stottrup and L. A. McEvoy (eds), Blackwell, Oxford, 2003.

(11) I. Laing and F. Ayala, "Commercial mass culture techniques for producing microalgae", In: Introduction to applied phycology, I. Akatsuka (ed), SPB, The Hague, pp 447–477, 1990.

(12) R. T. Lorenz and G. R. Cysewski, "Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin", *Trends Biotechnol.*, 18: 160–167, 2000.

(13) R. Osinga, J. Tramper, J. G. Burgess, and R. H. Wijffels, "Marine bioprocess engineering", *Proc. Prog. Ind. Microbiol.*, 35:1–413, 1999.

(14) G. D. Price, J. R. Coleman, and M. R. Badger, "Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942", *Plant Physiol.*, 100: 784-793, 1992.

(15) O. Pulz and W. Gross, "Valuable products from biotechnology of microalgae", *Appl. Microbiol. Biotechnol.*, 65: 635 -648, 2004.

(17) D. Riesenberg and R. Guthke, "High-cell-density cultivation of microorganisms", *Appl. Microbiol. Biotechnol.*, 51: 422-430, 1999.

(18) R. F. Strzepek and P. J. Harrison, "Photosynthetic architecture differs in coastal and oceanic diatoms", *Nature*, 431 (7009): 689-692, 2004.

(19) A. Tsuda, S. Takeda, H. Saito, J. Nishioka, Y. Nojiri, I. Kudo, H. Kiyosawa, A. Shiomoto, K. Imai, T. Ono, A. Shimamoto, D. Tsumune, T. Yoshimura, T. Aono, A. Hinuma, M. Kinugasa, K. Suzuki, Y. Sohrin, Y. Noiri, H. Tani, Y. Deguchi, N. Tsurushima, H. Ogawa, K. Fukami, K. Kuma, and T. Saino, "A mesoscale iron enrichment in the western subarctic Pacific induces a large centric diatom bloom", *Science*, 300 (5621): 958-961, 2003.

received October 08, 2004
revised November 16, 2004
accepted November27, 2004

Nutritional Compositions of Growth Medium for *Chlorella* FJ03

Tai-Hung Lin Yu Wei Lee Tzu Ying Huang Jyh-Yih Leu*

Department of Life Science
Fu Jen Catholic University
Taipei, Taiwan, 242, R.O.C.

Tzu-Yang Hsien

General Education Center
Chung Kuo Institute of Technology
Taipei, Taiwan, 116, R.O.C.

Yen-Hui Lin

Department of Health and Safety and Environmental Engineering
Chungtai Institute of Health Sciences and Technology,
Pei-tun District, Taichung, Taiwan, 406, R.O.C.

Abstract

Here we report the successful implementation of the fractional factorial design to screen the limiting components for the growth of *Chlorella* FJ03 and subsequent use of step-by-step approaching to design medium BGII that supported higher specific growth rate of *Chlorella* FJ03. The results showed that $MgSO_4 \cdot 7H_2O$ is a significant factor positively affecting the growth of *Chlorella* FJ03. Also factors $MgSO_4 \cdot 7H_2O$ and ferric ammonium citrate have a significant interaction effect on the growth of strain FJ03. The interaction study indicated the positive effect of ferric ammonium citrate, since the higher level (0.02 g/l) results in a higher specific growth rate at the high $MgSO_4 \cdot 7H_2O$ setting (1.0 g/l). The interaction also showed the negative effect of ferric ammonium citrate, but only at the lower $MgSO_4 \cdot 7H_2O$ setting (0.05 g/l). Furthermore, the results showed the highest specific growth rate of *Chlorella* FJ03 grown in the medium BGII

with the concentrations of 1.5~2.5 g/l of $MgSO_4 \cdot 7H_2O$ and 0.15~0.35 g/l of ferric ammonium citrate by step-by-step approaching. Combining the results in the study, the components of the optimal medium BGII for *Chlorella* FJ03 are $NaNO_3$ 0.5 g/l, K_2HPO_4 0.06 g/l, $MgSO_4 \cdot 7H_2O$ 2.0 g/l, $CaCl_2 \cdot 2H_2O$ 0.06 g/l, citric acid 0.06 g/l, ferric ammonium citrate 0.25 g/l, EDTA 0.001 g/l, Na_2CO_3 0.02 g/l and trace metal solution 1 ml.

Keywords: culture media, microalgae, optimization

新的圖像矩為基礎的繪圖成像演算法

王宗銘* 楊龍鴻**
圖學、影像、虛擬實境實驗室
國立中興大學資訊科學系
台中市 402 南區國光路 250 號

摘 要

非擬眞成像演算法(non-photorealistic rendering)爲近期電腦圖學的熱門研究領域，繪圖成像(painterly rendering)更是其中重要的研究主題。雖然圖像矩(image moment)爲研究學者使用的重要技巧之一；然而，以圖像矩爲基礎的演算法並不具曲線筆觸的特徵，此有違畫家繪畫之特質。本論文提出了一個基於圖像矩、提供曲線筆觸、具繪畫風格的新穎演算法。此演算法利用線積分迴旋(line integral convolution)來產生曲線筆觸路徑，並以統計的方法來自動產生長度、寬度不同大小、方向多樣化的局部筆觸。演算法也使用影像分類(image segmentation)技術來區分繪畫內容之邊界，確保作畫筆觸不致橫跨邊界，保持圖像細緻資訊。演算法更利用高斯柔化技巧，模擬畫布對繪圖風格處理的影響。本論文展示數張以我們演算法所繪製而成的圖像，並與相關學者之成果做比較。結果顯示：我們的演算法確實能達到具有曲線筆觸效果、畫家風格的圖像，這是其他相關學者所無法做到的。

關鍵詞：圖像矩、曲線筆觸、繪圖成像、邊界測定

一、前 言

非擬眞成像演算法(non-photorealistic rendering)爲近代電腦圖學中一項重要的研究

* Corresponding author. Tel: +886-4-22840497 ext 915
E-mail: cmwang@cs.nchu.edu.tw
** E-mail: batylh@gmail.com

範疇，其優點爲可省略瑣碎細節、誇張的強調重點、吸引觀察者之目光、傳達特定之意念等。其中繪圖成像(painterly rendering)爲一種將寫實影像轉化成爲具有畫家風格影像的一種成像方法。繪圖成像演算法大致分爲兩種形式：一種爲以電腦模擬某種作畫的技巧，例如考慮物理的特性，模擬水彩畫；另一種演算法則強調自動繪圖，亦即以電腦自動產生類似畫家的筆觸，然後根據這些筆觸來繪畫(Haeberli,1990)。此也被通稱爲基於筆觸的成圖(stroke-based rendering, SBR) (Hertzmann,2003)。

基於筆觸的成圖是本文所要探討的主題。在此主題中，使用圖像矩技巧(image moment) (Freeman, 1998)來研究此主題更是熱門的方法(Shiraishi,2000; Nehab, 2002)。此乃由於圖像矩常被利用於電腦視覺(computer vision)或機器人視覺(robot vision)來快速的計算出影像的概括位置與方向。因其爲一個近似矩形，是故被稱爲圖像矩。圖像矩提供各種有效資訊，包括矩形中心、矩形長寬、矩形走向等，因此，利用這些特性，我們除可逼近出一個影像外，也可以用來模擬筆刷的筆觸(strokes)在畫布上的繪圖過程。

Shiraishi 等人據此提出基於圖像矩的繪圖成像演算法，展現出圖像矩在非擬眞成像的應用(Shiraishi,2000)。然而，他們的演算法採用固定的取樣方格(stamp)，以致筆觸的大小受到限制。Nehab 等人將圖像矩的觀念加以延伸(Nehab, 2002)。他們提出了多重比例(multiscale)的方法，突破 Shiraishi 等人的論文中，取樣方格固定性的限制，產生具有畫家風格的繪圖影像。

綜觀 Shiraishi、Nehab 等人的研究，我們發現他們均使用直線、單一的筆觸來作畫，並不考慮曲線、連續筆觸的表徵。然而，在眞實世界中，畫家筆下的筆觸，除了大小不一外，更含有曲線、連續筆觸的特質。因此，我們認爲有必要研究具有曲線、連續筆觸的演算法，以避免因爲僅使用單一直線筆觸作畫，導致產生死板、僵硬的圖像。這個論點也符合 Hertzmann 等人之呼籲 (Hertzmann,1998)。雖然他們不使用圖像矩，但卻使用隨機、連續的方法來達成具有曲線、連續筆觸的繪圖成像。圖像矩在繪圖成像演算法有其優異之特性，故乃成爲一種熱門的研究方法；然而，相關學者所提出的演算法，表現出的特徵並不具曲線、連續筆觸的風格。我們必須研究改進基於圖像矩的繪圖成像演算法，方能使繪圖的結果更臻生動、更添美感。

本論文據此提出了一個基於圖像矩、提供曲線連續筆觸、具繪畫風格的新穎演算法。圖像矩在本論文中主要是用來計算局部筆觸的構成要素，其中包括長(length)、寬(width)、旋轉中心、旋轉角度等。我們利用線積分迴旋(line Integral convolution)演算法(Cabral, 1993)的概念來模擬出曲線筆觸路徑，並利用其線路徑來做爲曲線筆觸的塗抹

軌跡。這個曲線路徑配合上述的局部筆觸，即可構成連續的曲線筆觸，可以實際模擬出畫家作畫過程。此外，我們也改進了傳統圖像矩的繪圖成像演算法；我們以統計的方式自動的變更其取樣方格大小，以獲得更具彈性的局部筆觸長度、寬度、與方向大小。再者，我們更使用影像分類(image segmentation)來偵測邊界，致使在作畫時，連續的曲線筆觸不會橫跨邊界，而是會被裁切其多餘的筆觸。此一方面可確保筆觸在邊界內能展現出特徵效果，另一方面也可保存原始細緻的圖像資訊。最後，我們利用高斯柔化方法，模擬在畫布的作畫，來達成視覺上的柔順與美感。

本論文結構如下：第二章說明相關工作；第三章敘述我們所提新的演算法；第四章節說明實驗結果，並與相關研究學者之結果作客觀比較。最後，第五章提出結論與未來工作。

二、相關工作

非擬眞成像演算法是近年來在電腦繪圖範疇中發展相當快速之一項研究課題，主要目的是以電腦演算法自動產生類似素描、水彩及油畫之類，輸出與一般照片不同類型的圖像(Hertzmann,2003)。非擬眞成像演算法可進一步根據實體模型（physical model）之有無來細分。若有實體模型，則依據實體模型特性，以電腦程式模擬，其結果會較爲眞實；反之，若無實體模型，則需憑藉研究者思考與觀察，盡可能找出更多正確之資訊，來模擬求得最後結果。非擬眞成像演算法可參考 Gooch & Gooch 的相關著作「non-photorealistic rendering」(Gooch,2001)來獲得更詳盡的介紹與論述。目前非擬眞成像演算法的研究相當盛行，研究領域多樣。本文列出 2004 年最新的論文，以供參考(Chen,2004; Hays,2004; Selle,2004; Sykora,2004)。值得注意的是，以圖像矩做繪圖成像的近期的文獻中，以 Nehab (Nehab,2002) 爲最新的代表。

畫家認爲繪畫即是光與陰影之表現；以科學角度而言，其實光與影之表現就是亮度(luminance)強弱之展現。非擬眞成像演算法可以視爲以電腦的科技逐步模擬、實現畫家作畫之研究，此亦即爲繪圖成像名稱之由來。繪圖成像演算法可分爲使用圖像矩(image moment based)與非使用圖像矩(non-image moment based)兩大類。前者，可由 Shiraishi (Shiraishi,2000)及 Nehab (Nehab,2002)等所發表的論文中窺得；而在後者，可參考 Hertzmann(Hertzmann,1998)於畫家風格處理上所致力的研究發展。以下，我們將簡略的介紹各相關研究與貢獻。

Litwinowicz (Litwinowicz,1997)已嘗試在影片(video)上以自動的方式去達成印象派

處理的效果。在處理時，筆刷被隨機地分布在輸入影像上，且切割其影像邊緣，以確保畫面的精細度；而其筆觸之方位走向，是以梯度去作爲計算。而在筆觸分布上，其利用擾動方格(jittered grid)來進行初始化。另外光流場(optical flow field)被用來在畫面上移動筆刷，以避免畫面的繪圖過於突兀。

Hertzmann (Hertzmann,1998) 以鍵槽曲線(spline curves)來表現筆觸。其方法是合成多個圖層於一個圖像，且每個圖層示以其對應的固定筆寬繪製而成；由底層之對應筆寬向上逐漸減小，最上層則爲最細微的筆寬所構成。 如此，所合成的圖像相較於原始影像，其特徵便可逐一被保留。因此，其合成影像每增加一層，便更加逼近於原始圖面。其中筆觸之走向是以梯度來控制。基於以上技巧，所產生的結果即是包含著數種不同筆寬筆觸的繪圖圖像。

Shiraishi (Shiraishi,2000) 提出了一個自動的繪圖成像演算法，主要是以圖像矩作爲基礎的手法，產生局部近似的矩形，來模擬筆刷。除了利用顏色差別影像(color different image)來計算強度外，他們並配合空間填補曲線(space filling curve)來找出筆觸落點。當圖面所有筆觸落點被決定、各相對應的筆刷參數已被取得後，演算法便依各筆觸大小，由大到小進行塗抹上色。如此小筆觸便不會被大筆觸覆蓋，而確保圖像精細度。

Nehab (Nehab,2002) 參考了 Shiraishi 的手法，提出了另一種繪圖成像的方法。Nehab 將圖像矩的筆觸演算法做兩個方面的延伸：採用多重比例(multiscale)方法去計算筆觸；提供一種參數結構去控制筆觸分布。Nehab 導入圖層(layer)的觀念，將每一層的筆觸做不同的解析度縮放，再以大筆觸作爲底層，小筆觸作上層，進行重疊。在筆觸分佈方面，是採用抖動演算法(dithering algorithm)進行計算之。

表 1 比較本論文演算法與目前我們所知，以圖像矩處理繪圖成像兩篇論文之異同(Shiraishi,2000) (Nehab,2002)。從筆觸分佈之計算方法來看，Shiraishi 與 Nehab 所提出的方法均無法達成曲線筆觸的效果，而我們是採用線積分迴旋的概念，進而促使筆刷形狀有曲線、連續的特徵；此外他們的演算法並未考慮邊界偵測之重要性，而我們是以影像分類來達成保持圖像細微特徵的目的；再進一步探討，Shiraishi 使用固定取樣方格， Nehab 使用多重比例概念，於各圖層上做出數種不同的取樣大小，兩者都需要使用者預先設定取樣方格的大小。相反的，我們以統計的方式，無須使用者之介入，就可自動達成取樣方格大小變動的效果；再者，由於我們的筆刷已經涵蓋曲線形狀，故筆刷長度也已長於其他兩者的成果，且長度也不受限於取樣方格內；最後，我們別於其他兩者都未考慮畫布之影響；我們使用高斯柔化技巧，作爲畫布的底層，以達成

視覺上的柔順美感。

表 1 Moment Based painterly rendering 之比較表

比較 / 項目	Shiraishi (Shiraishi,2000)	Nehab (Nehab,2002)	Our method
筆觸分布之計算方法	Space-Filling Curve algorithm	Dithering algorithm	Line Integral Convolution
筆刷形狀	單一矩形	單一矩形	曲線
圖形邊界偵測	無	無	Image segmentation
取樣方格大小	固定	Multiscale	以統計學自動計算變動量
筆刷路徑長度	小於取樣方格大小	小於取樣方格大小	LIC 路徑長度
畫布處理	無	無	高斯濾鏡

三、新的曲線筆觸演算法

本論文提出一個基於圖像矩、具有曲線連續筆觸的繪圖成像演算法。我們的演算法是利用線積分迴旋演算法的概念(line integral convolution algorithm) (Cabral,1993)，來模擬出曲線筆觸路徑，並以統計的方式變更其取樣方格大小，進而使用圖像矩來計算局部筆觸的構成要素；緊接著使用影像分類來測定邊界，確保筆觸能展現出特徵效果；最後，我們再利用高斯柔化，產生畫布，以達成視覺上的柔順。

我們的演算法大致上區分爲成六個步驟，其流程圖如圖 1 所示，分別爲 1. 變動取樣計算（variable stamp calculation）、2. 線積分迴旋路徑計算（line integral convolution path calculation）、3. 局部筆觸計算（local source image approximation）、4. 邊界測定（image segmentation computation）、5. 畫布高斯柔化（gaussian filter）、6. 繪圖（drawing）。在過程上，首先我們計算出筆觸路徑與計算筆觸大小所需之相關參數，並沿著線積分迴旋路徑做筆刷模擬；接著，我們導入邊界測定，避免筆觸顏色跨越邊界；最後，我們透過高斯濾鏡，先柔化原始影像來作爲最底層之畫布，然後於其上將筆觸進行繪製塗抹，即可完成一幅具有畫家風格的圖像。以下我們在各節中，分別詳細說明此六個步驟的作法。

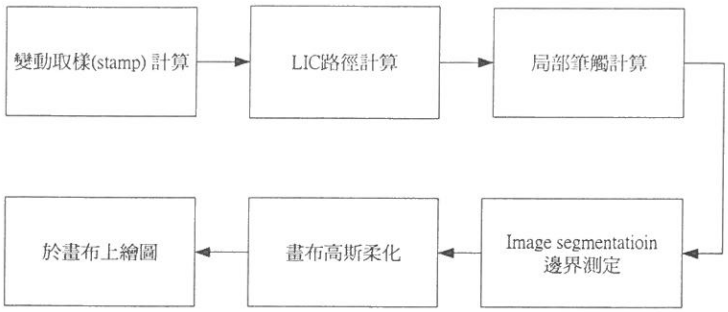


圖 1 圖像矩爲基礎的曲線筆觸 painterly rendering 演算法之流程圖

3.1 變動取樣計算（Variable Stamp Calculation）

Shiraishi (Shiraishi,2000)所提出的方法中，取樣方格尺寸爲固定，造成圖像中筆觸大小受限；爾後的 Nehab (Nehab,2002)爲了改善這個問題，提出了多重比例概念。其方法是於每個圖層上做有限度的筆觸放大或縮小，故變化量會隨各圖層多寡而有所不同。我們認爲影像的來源有多種，不能一概使用同一種固定的模式來決定筆觸的大小，應該以更自動的方式來處理。據此，我們採取統計的方法來達成上述目的。

我們如能計算出圖像中每一個像素點所對應的取樣方格值，則筆觸大小將因爲不同之輸入圖像而有所不同，且筆觸大小也將更具多樣化、自由化、自動化，此也將使繪製的結果更爲生動、有趣、多變。

首先，我們先給定一個取樣方格上界值 S_{max}，作爲取樣方格(stamp)邊長的最大上限值。這個數值代表在來源影像中，所有的取樣方格的最大範圍數值。緊接著我們根據這個上限數值，計算每個像素點所應該對應的取樣數值。假設以某個像素點爲中心，對某個取樣方格，其範圍內有 SxS 個像素點。我們的方法如下：

1.計算此取樣方格內，每一個像素的強度標準差(standard deviation) σ_{xy} (Treavett, 1997)。強度 $I(x,y) = f(d(C,C_w(x,y)))$，C 爲筆刷顏色；而 C_w 爲來源影像的顏色值。

$\sigma_{xy} = \sqrt{\frac{1}{S^2-1}\sum_{x=1}^{S}\sum_{y=1}^{S}\left(I(x,y)-\mu_{xy}\right)}$ ，其中 μ 爲每個取樣方格內的強度平均(mean)。

$$\mu = \frac{1}{S}\frac{1}{S}\sum_{x=1}^{S}\sum_{y=1}^{S}\left(I(x,y)\right)$$

根據上述公式，我們可以找出每個像素點的強度標準差 σ_{xy}。

2.找出來源影像中的最大強度標準差 $\sigma_{\max}$。計算完每個像素的強度標準差後，此資訊僅需將所有的像素標準差加以排序即可求出。

3.利用像素點的強度標準差 σ_{xy} 與來源影像的最大強度標準差 $\sigma_{\max}$，依照線性差補(interpolation)技巧，計算出每一個畫素(x,y)所相對應的取樣方格大小 S_{xy}。

$$S_{xy}=\begin{cases}=\dfrac{S_{\max}-S_{\min}}{\sigma_{\max}}\cdot\left(\sigma_{\max}-\sigma_{(x,y)}\right)+S_{\min} & if\ \sigma_{(x,y)}\le\sigma_{\max}\\ =S_{\min} & if\ \sigma_{(x,y)}>\sigma_{\max}\end{cases}$$

經過上述的方法，我們就可以達到變動取樣計算的效果。輸入任何的來源影像後，使用者僅需指定樣方格上界值 S_{max}，我們的方法會根據此數值，自動的變更每個像素點所對應的取樣方格數值，達成動態的變動大小效果。

3.2 Line Integral Convolution 路徑計算

Shiraishi (Shiraishi,2000)與 Nehab (Nehab,2002)於手法上，其筆觸之形成，均取決取樣方格，是故圖像中，每一個筆觸都是一個獨立的小片段，無法模擬出實際上畫家作畫時的風格，此有違畫家實際之作畫技巧。爲謀改善此缺失，我們利用線積分迴旋(Line Integral Convolution)的原理(Cabral,1993)，計算其曲線軌跡，來作爲筆刷的塗抹路徑，藉此達成模擬曲線筆觸，促成擬眞的效果。

圖 2 (a)爲原始影像，圖 2 (b)爲使用 LIC 長度爲 125 作展現出來的 LIC 路徑，在往後章節的步驟，其將作爲筆刷行走的路線。每一條 LIC 曲線路徑全長爲 125 個畫素，並如圖 3 所示，以點(x,y)爲中心，向前延伸與向後延伸。

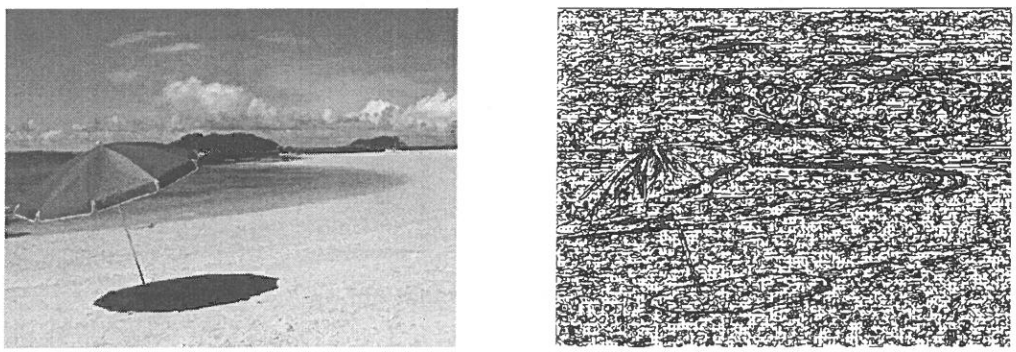


圖 2　原始影像(左圖)與線積分迴旋路徑圖(右圖)

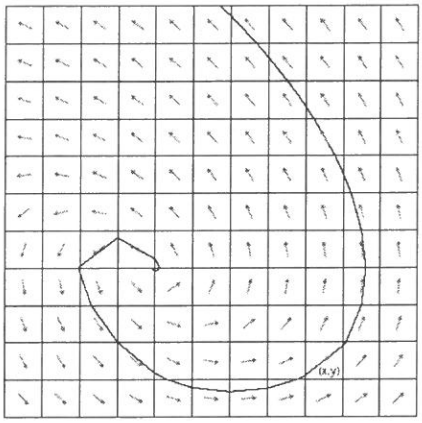


圖 3 線積分迴旋簡示圖。(x,y)代表來源圖像中的某個像素點，各個箭頭代表在該像素點的向量方向

3.3 局部筆觸計算（Local Source Image Approximation）

我們於前節已使用線積分迴旋概念，決定了筆刷行走的路線。接下來，我們將以圖像矩的手法找出筆刷路徑上的局部筆觸形狀。每條線積分迴旋路徑代表每一條曲線筆觸的路徑，其筆刷顏色是使用每個線積分迴旋路徑中心點(x,y)的顏色來塗抹，並且依路線向前、後延伸，於路徑上取經過的像素點做局部筆觸計算。因此，一條曲線筆觸的構成要素可分爲下列數個資訊；

1.筆觸路線；爲線積分迴旋路徑所形成的軌跡，如圖 4 (c)。

2.筆刷顏色；爲線積分迴旋的中心點之顏色 $C_{(x,y)}$ 。

3.局部筆觸計算；計算線積分迴旋的曲線上之各個局部筆觸。

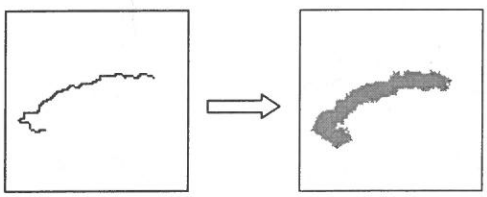


圖 4 以線積分迴旋路徑爲基礎的單一曲線筆觸

曲線筆觸演算法係用來繪製畫家風格的圖像，如圖 5 所示。在此演算法中，計算局部筆觸爲其首要的步驟。首先，我們必須決定取樣方格的大小(stamp)。我們的方法是採用變動性的取樣方格，其值 S_{xy} 已於前節中敘述。利用每個計算出來的 S_{xy} 值，我們界定出一個取樣範圍後，便可以計算出各點(x,y)的圖像矩 M_{xy} 。然後，我們再利用 M_{xy} 計算出局部筆觸的各項參數，包括：長度(length)、寬度(width)、旋轉中心、旋轉角度等。得到各項參數後，我們即可以起始點(亦即線積分迴旋線段的中心點)的顏色值作爲筆刷顏色，再沿著線積分迴旋線段上之各個像素點，做局部筆觸計算。每個局部筆觸均可獲得其相對應之筆觸參數值；依照這些數值，我們便可以沿著線積分迴旋線段，以圖像矩的手法做繪製具有畫家風格的圖像。

1.取得線積分迴旋中心點 (i,j) 顏色 C_LIC_{ij} 。
2.沿著線積分迴旋線段取得各點 (x,y) 之原始顏色值 C_{xy} ，並計算強度值 $I(x,y)=f(d(C\_LIC_{ij},C_{xy}))$ ，若強度值符合條件，則繼續計算，否則跳到下一點。
3.於點(x,y)以 C_LIC_{ij} 及變動取樣方塊值 S_{xy} 計算圖像矩 M_{xy} ，並計算其對應之長 L_{xy} 、寬 W_{xy} 、旋轉角度 θ_{xy} 、旋轉中心 (X_c,Y_c) 。
4.每計算一個點的局部筆觸參數，隨即以 C_LIC_{ij} 及其參數進行塗抹。
5.整個線積分迴旋線段內的像素點均塗抹後，即構成一條曲線筆觸。

圖 5 曲線筆觸演算法

圖 6 顯示曲線補處演算法的塗抹過程。此圖中，(a)~(c) 爲曲線筆觸塗抹過程，黑線部分爲線積分迴旋的向量曲線。作畫時係由左下角的海灘部分開始，依序爲陽傘部分、最後是其背景部分。(d)爲完成之圖，此圖已作邊界測定處理，故筆觸不致橫跨邊界。但此圖並未考慮畫布的因素，故尚未使用高斯濾鏡，也尚未置於畫布上。因此，左側邊緣天空部分仍可見未被塗抹的空白像素點。我們將在下節敘述如何達邊界測定，以及可達成之效果。

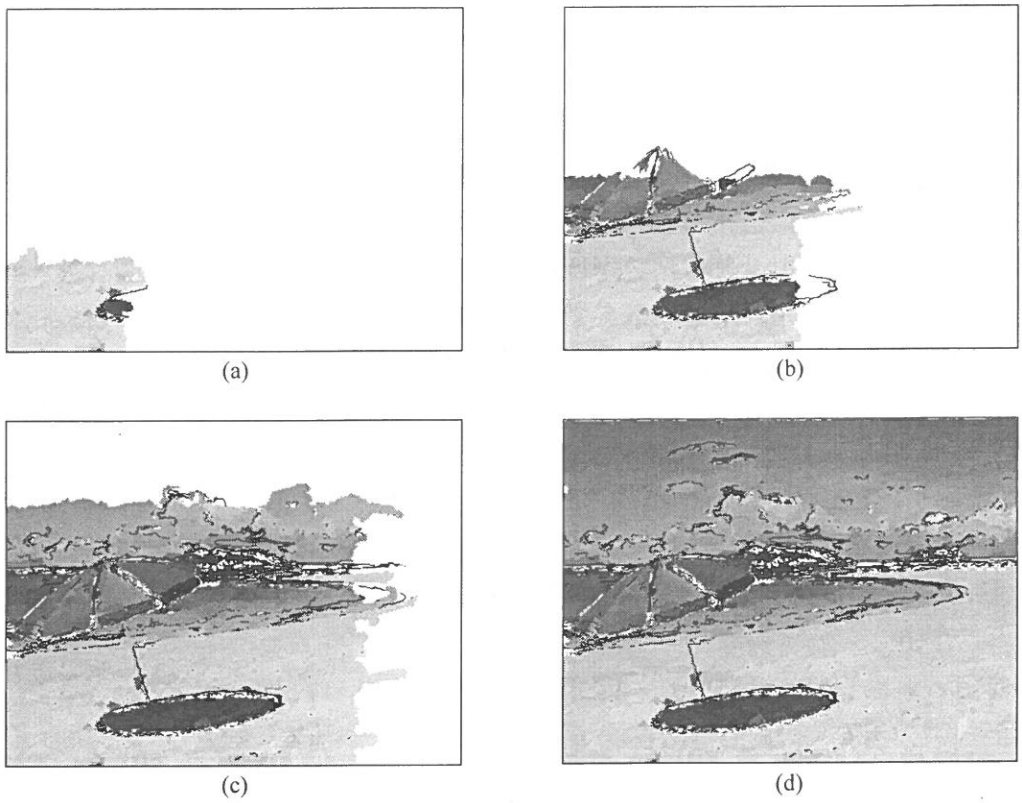


圖 6、曲線筆觸塗抹過程示意圖。依照順時針順序由上而下，(a)開始塗抹左下角，(b)塗抹陽傘部分、(c)繼續爲塗抹其餘背景部分、(d)爲完成圖

3.4 邊界測定（Image segmentation Computation）

爲了防止不同顏色的筆觸跨越到其他部分，而造成結果過於繁亂，我們必須對於邊緣的部份特別加以處理。在此採取影像分類的方法，進行邊界偵測。此法之優點爲可避免圖像中的筆觸跨越塗抹邊界，以確保圖像特徵的保有。

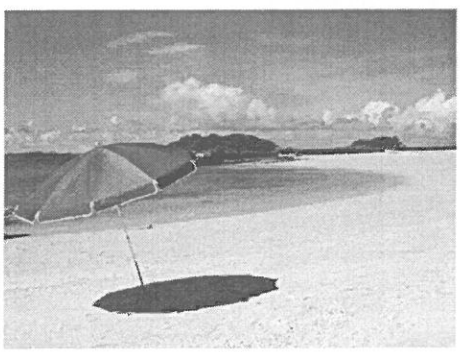
圖 7 說明邊界測定與塗抹演算法的過程與判斷。首先我們將原始圖以影像分類計算出各個區域分野並記錄之。我們以各區域之顏色值其作爲邊界測定之數據。我們由前節得知，每條曲線筆觸，其顏色爲該線積分迴旋線段的中心點顏色值(C_LIC_{ij})。因此，線積分迴旋線段的中心點(i, j)所對應的影像分類顏色值(C_SEG_{ij})，便是作爲邊界偵測的重要依據。每一條曲線筆觸，我們有必要比對其影像分類顏色值

(C_SEG_{xy})與中心點顏色值(C_SEG_{ij})是否相同。若相等，代表於屬於同一區域；若不相等，則不屬於同等區域。當不爲同等區域時，其筆觸之塗抹即偵測是否超過邊界，是否需進行截除的處理。

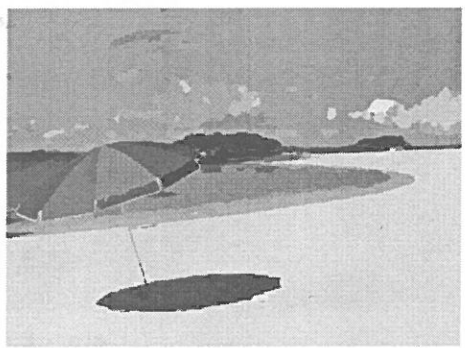
1.以原始圖面計算影像分類，並記錄每一個畫素所對應的分群顏色值C_SEG_{xy}。
2.取線積分迴旋 線段的中心點(i, j)對應之顏色值C_SEG_{ij}，與線積分迴旋 線段上的所繪出之筆觸之所有點所對應的顏色值C_SEG_{xy}。
3.比對C_SEG_{ij}與C_SEG_{xy}，若相等，則該點(x, y)進行筆觸顏色C_LIC_{ij}塗抹，若不等，則該點不塗抹。

圖 7 邊界測定與塗抹演算法

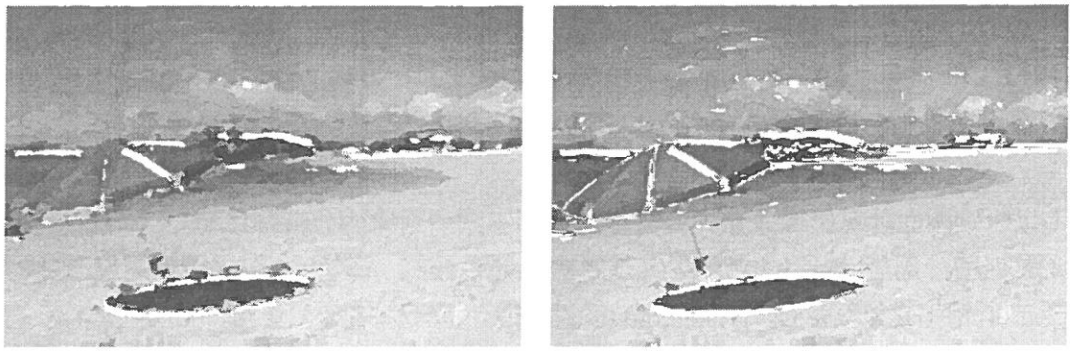
圖 8 顯示影像分類與邊界測定圖。在圖中，(a)爲原始來源影像，(b)爲經過影像分類之圖像。可以明顯看出影像已被歸類在不同的區域內，陽傘部分因爲顏色相異，故視覺觀之尤爲明顯。(c)爲不考慮影像分類的邊界；(d)爲經過邊界測定之處理後之結果，此代表塗抹過程中，控制筆觸侷限於同一個區域內，不橫跨到另一個區域。我們可以明顯看出(c)圖顯得雜亂不一，但是(d)圖則保持細部特徵的效果，其邊界也較有規則。仔細觀看陽傘的紅色部分，我們發現(c)圖的紅色筆觸，橫跨了隔鄰的陽傘藍色部分，但是在(d)圖中，則未見此情形。整體觀之，(d)圖相較於(c)圖保存了更多的原始資訊。



(a)



(b)



(c) (d)

圖 8 影像分類及邊界測定

3.5 畫布高斯柔化（Gaussian Filter）與繪圖(Drawing)

我們是以圖像矩局部筆觸計算來作爲其近似矩形，故在圖像的塗抹過程中，有可能出現未被塗抹的像素點。因此，我們需要一個以原始圖面作爲畫布的底層，來彌補這些空缺。然而，在繪圖成像的風格上，展現出的結果並不會與原始圖面完全相同。是故，在填補這些縫隙時，我們需要作進一步的處理，來達成在視覺觀感的平衡。在此，我們使用高斯濾鏡先處理原始圖面，並以此作爲畫布。我們所使用的高斯濾鏡半徑值爲 1 個像素寬度，此可將原始圖面作些許之柔焦化，但又不致變的太模糊不可辨認。圖 9 爲經過高斯柔化後的影像。

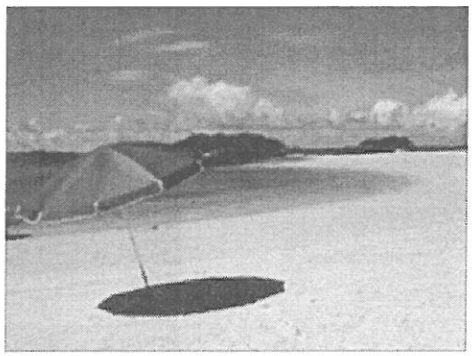


圖 9 高斯柔化後之影像

我們可以觀察出其柔焦後的效果，圖像中的精細度與銳利程度均已降低，包括圖像邊緣，例如：洋傘的邊緣等。而待畫布經過處理形成後，我們便可以其作爲底層，將各個曲線筆觸疊到畫布上，完成繪圖動作。以圖 2 的原始圖像爲例，經過我們的演

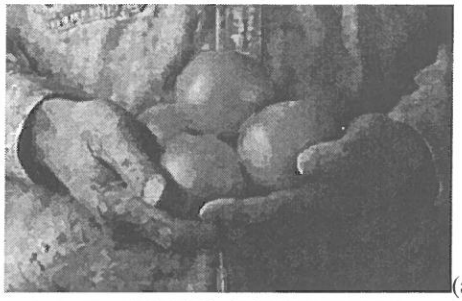
算法處理過後，完成結果如圖 11 (a)所示。

四、執行結果

本篇論文所使用的實驗平台爲個人電腦，配備 2.6 GHz 處理器，1.0 GB DDR RAM，顯示卡爲 RADEON 9600 Pro，程式撰寫工具爲 Visual C++。在展示結果中，我們取用 Hertzmann (Hertzmann,1998)於論文中所使用的影像，以便能與其他學者之結果做客觀的比較。值的一提的是，Hertzmann 並非以圖像矩來達成繪圖成像的效果，但是其他的學者普遍使用他論文中所提出的影像作爲展示的來源影像。使用 Hertzmann 論文圖像所處理的結果如圖 10 (a)所示。此外，我們也自行取用兩個圖像做處理，以便作更進一步的展示，其結果圖 11 (a)、圖 12 (a) 與圖 14 (a)。以上所展現的圖像，均是以取樣方格大小最大值 15，線積分迴旋曲線長度爲 125 的長度，所處理而成的。

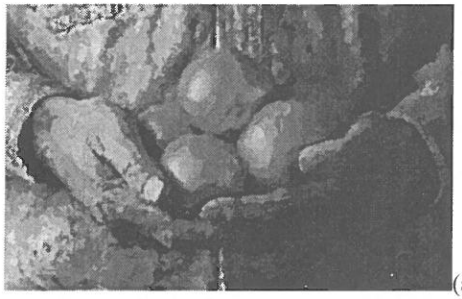
我們可以明顯看出圖 10 (a)在人物的手與水果的部分，均展現出其曲線效果。其 rms (Root Mean Square)數值爲與 Hertzmann 原始影像之比較結果。而圖 11 (a)於雲與海洋的部分，亦有曲線筆觸的表現。我們也列出與來源圖像比較的 rms 數值，以供參考。同樣的，在圖 12 (a)的天空部分，曲線特徵表現相當突出。我們在此也列出各個處理結果的相關參數分布圖。圖 10 (b)表示圖 10 (a)中於局部單一筆觸的角度偏轉分佈情形，大致上是分布在-45 度到+45 度之間。圖 10 (c)是表示該圖像中於局部單一筆觸的矩形長度之分佈情況，大多集中在 6~10 個像素長度之間，且於 0~15 的長度間均有分佈。這些數據顯示圖像中的局部筆觸大小，是多樣化變動的；此也表示我們演算法已能藉由統計的方法自動計算出取樣方格，達成我們所期望的多樣化筆觸的目標。

同樣的，圖 10 (d) 是展示出該圖像中於局部單一筆觸的矩形寬度分佈，此也佐證我們的演算法確能達成多樣化筆觸的目標。圖 11 (b)、圖 11 (c)、圖 12 (b)與圖 12 (c)亦展示出圖 11 (a)與圖 12 (a)的各項參數分布情況。如前所述，這些都足以證明我們的演算法能達成曲線、連續筆觸、多樣化之效果。



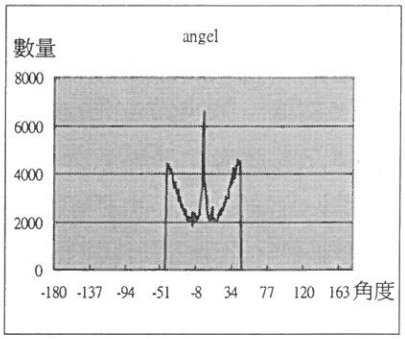
(a-1)

圖 10 (a) 我們演算法之結果 rms=11.096612

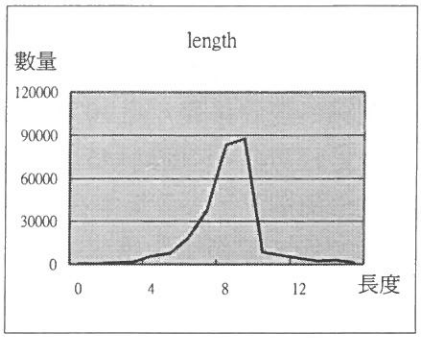


(a-2)

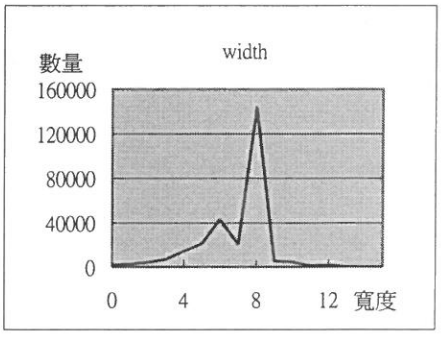
Hertzmann 論文之結果



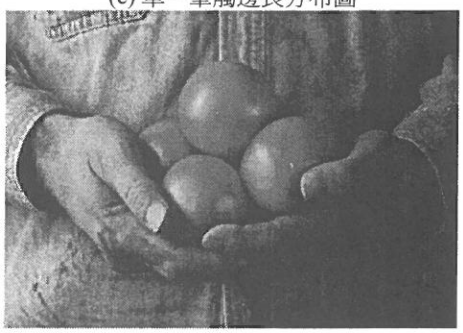
(b) 單一筆觸旋轉角度分布圖



(c) 單一筆觸邊長分布圖



(d) 單一筆觸邊寬分布圖



(e) Herzmann 原始圖像

圖 10 使用我們的演算法，利用 Herzmann 論文中的圖像所繪製出來的圖像與數據分析。

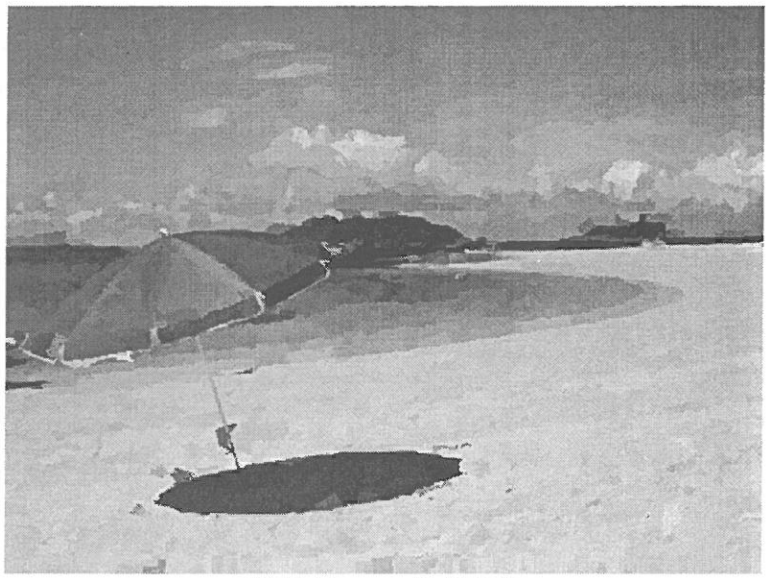
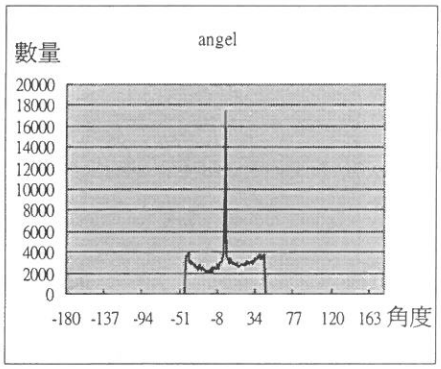
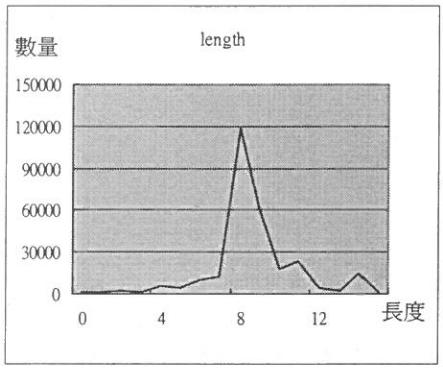


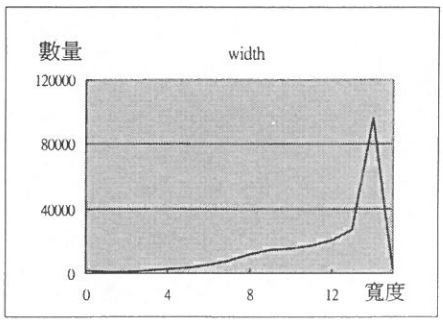
圖 11 (a) rms=10.330070



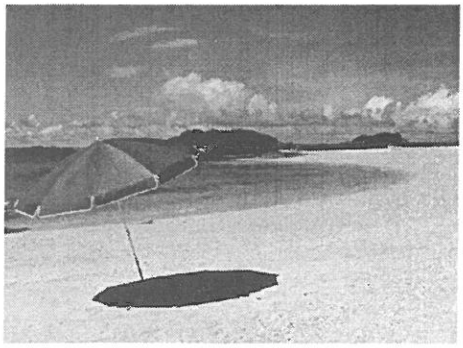
(b) 單一筆觸旋轉角度分布圖



(c) 單一筆觸邊長分布圖



(d) 單一筆觸邊寬分布圖



(e) 原始影像圖

圖 11 海灘圖像之結果與分析

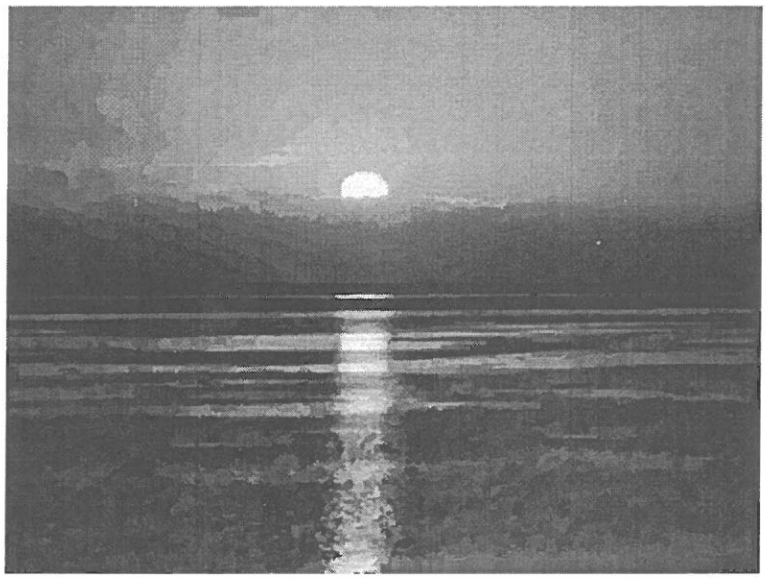
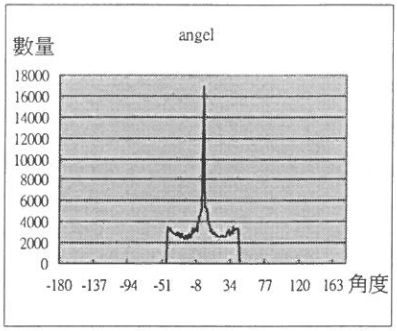
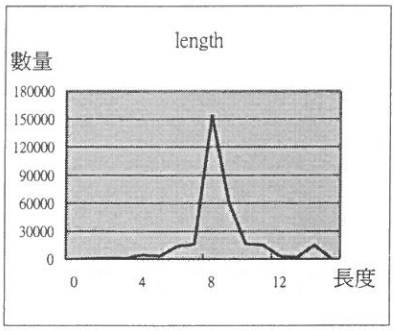


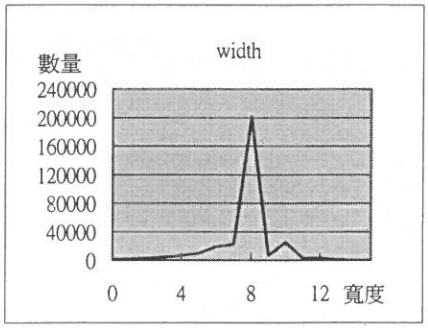
圖 12 (a) rms=8.930853



(b) 單一筆觸旋轉角度分布圖



(c) 單一筆觸邊長分布圖



(d) 單一筆觸邊寬分布圖

(e) 原始影像圖

圖 12 夕陽圖像之結果與分析

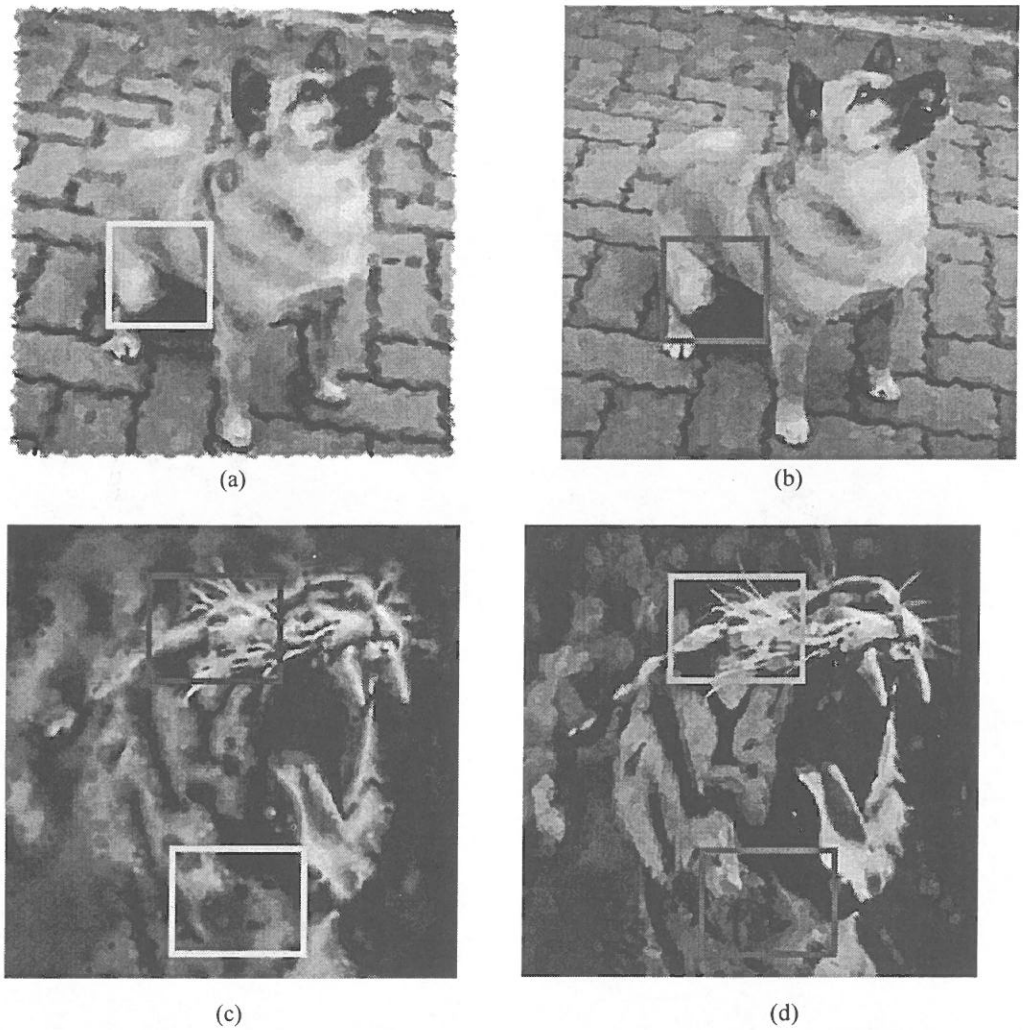
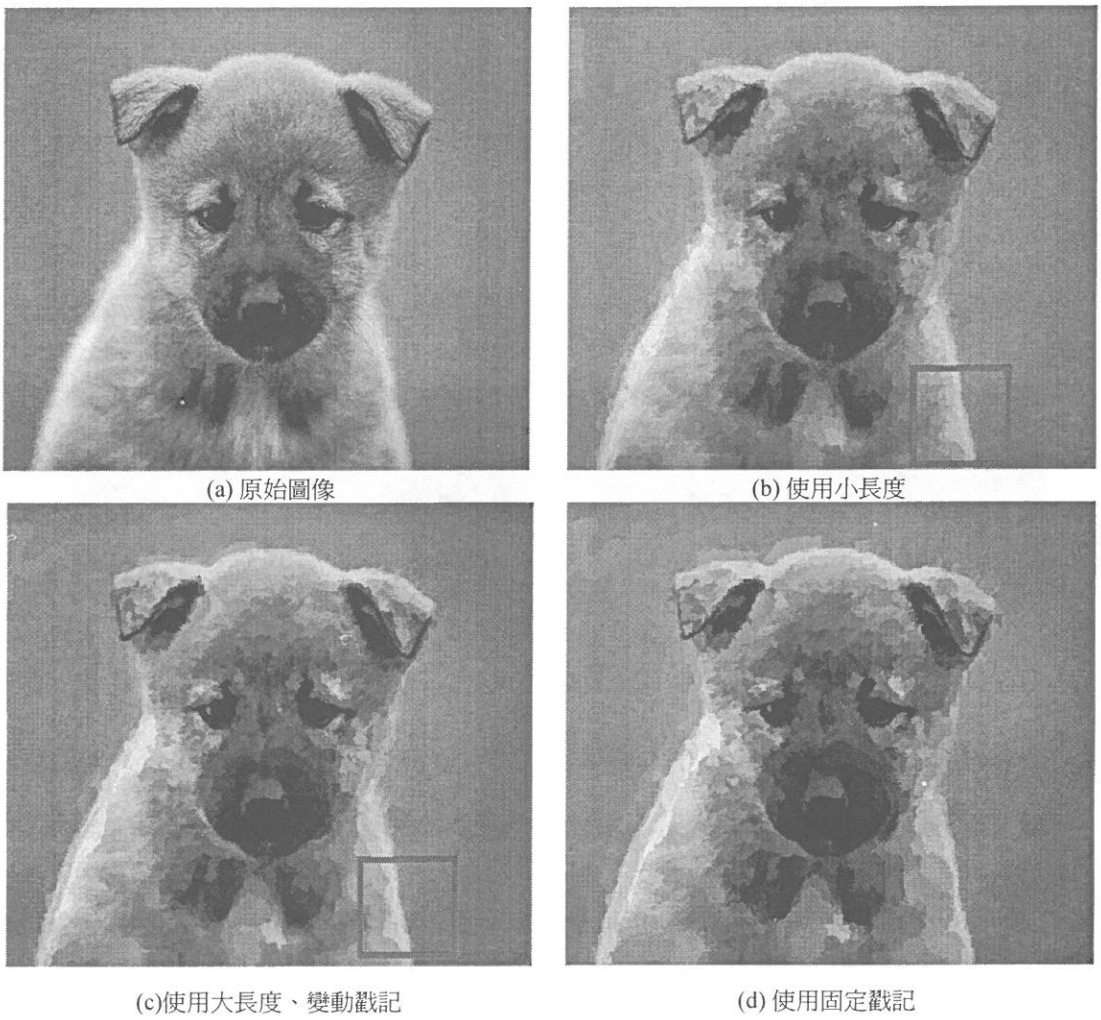


圖 13　Shiraishi 結果（左上(a)圖）及 Nehab 結果（左下(c)圖）與我們的演算法結果（右上(b)圖、右下(d)圖）之比較。

圖 13(a)爲 Shiraishi 於論文中使用的方法所展現效果，圖 13(c)爲 Nehab 使用其多重比例的方法所形成的結果；圖 13(b)與圖 13(d)是以本論文中所提出的演算法處理過後的影像。仔細比對圖 13(c)與圖 13(d)的紅色方框部位與圖 13(a)與圖 13(b)的黃色方框部位，會發現 Shiraishi 與 Nehab 所使用的手法爲片狀筆觸；而使用我們的方法，則有明顯的曲線筆觸效果。此外，我們的方法可以保留更多的影像細節，但是其他兩人的方法則可能會因爲筆觸大小控制不良，故無法保留細節部分。例如，Shiraishi 結果中

主題犬的眼睛、腳爪部分已不甚明顯。Nehab 結果中主題老虎的觸鬚、門牙都比我們演算法所處理出來的圖像更不清楚，失去纖細之輪廓特徵。



(a) 原始圖像　(b) 使用小長度

(c)使用大長度、變動戳記　(d) 使用固定戳記

圖 14　我們所提的演算法中，使用小長度(b)、大長度(c)筆觸及使用變動戳記(c)、固定戳記(d)之結果比較

我們是以線積分迴旋的長度作爲曲線筆觸的長度。圖 14 比較使用不同長度筆觸之處理效果。圖 14(a)爲原始圖像；圖 14(b)爲經長度 30 處理，圖 14(c)則是爲長度 125 處理之結果。我們可以觀察出，使用較大長度筆觸，其曲線筆觸也越長，呈現連續的型態。圖 14(c)是變動戳記(variable stamp)的結果，圖 14(d)是以固定戳記(fixed stamp)的

方式處理。很明顯的，我們可以看出以變動戳記方式所產出的圖片較爲細膩，其微小部分的表現比較細緻，且較爲寫實；而利用固定戳記技巧所展現的圖片，較具印象派風格，但表現較爲粗糙，其細微部分均無法表現得宜。此也適足證明我們所提的演算法使用變動戳記有其優異特性。

五、結論與未來工作

本論文提出一個新穎的以圖像矩爲基礎的繪圖成像演算法。我們以線積分迴旋的線段長度做爲筆觸的長度，並以其路徑作爲筆刷塗抹路線，來達成連續、曲線筆觸的效果。我們更改圖像矩的方法，使用變動的局部取樣大小，加入影像分類邊界測定、高斯濾鏡等，展現出更令人激賞的圖畫風格。

我們方法具有下列優點：第一、連續、長曲線筆觸：具有連續、長曲線筆觸的特徵，表現上較接近於實際畫家作畫筆觸風格。第二、可變動的取樣：改進固定取樣的效果，使其具有大筆觸與小筆觸之多重風格。第三、邊界測定：避免筆觸跨過邊界，可確保筆刷顏色能正確的落在著色部位，保持影像細緻資訊。第四、自動演算及繪圖：我們的程式允許使用者於開端輸入相關參數後，進行自動繪圖，無須中途進行調校或修改。第五、畫布高斯柔化：於底層我們採用原始影像之高斯柔化爲襯底，以達成更爲自然的效果。第六、更佳的視覺效果；由於曲線筆觸的表現得宜，使得處理後的影像，更近似於畫家風格，而展現出更令人讚賞的效果。綜合以上，我們於表 1 對目前以圖像矩處理繪圖成像之兩篇論文與本論文的演算法做比較後發現，本論文所提出的新穎的繪圖成像演算法，達成了曲線筆觸風格的效果，並展現出比其它兩者更佳的視覺效果與風采。

在未來工作方面，我們建議朝向數個方向繼續研究：第一、套用到其他格式影像上。目前繪圖成像的相關研究繁多，可將本演算法與運用到其他種類影像上。第二、縮短處理時間：我們的方法在處理時間上仍有改善空間，故可於處理過程上作最佳化的改進。

六、參考文獻

(1) A. Hertzmann, “Painterly Rendering with Curved Brush Strokes of Multiple Sizes,” in *SIGGRAPH 98 Conference Proceedings*, Michael F. Cohen, editor, pp. 453-460, 1998.

(2) A. Hertzmann, "A Survey of Stroke-Based Rendering," *IEEE Computer Graphics and Applications*, 23(4):70-81, July/August 2003.

(3) B. Cabral and C. Leedom, "Imaging Vector Fields Using Line Integral Convolution," *SIGGRAPH 93 Conference Proceedings*, pp. 263-270, ACM SIGGRAPH, 1993.

(4) B. Gooch and A. Gooch, *Non-Photorealistic Rendering*, A.K. Peters, New York, 2001.

(5) D. Comaniciu and P. Meer, "Mean Shift: a Robust Approach toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603-619, 2002.

(6) D. Nehab and L. Velho, "Multiscale Moment-based Painterly Rendering," in *Proceedings of 15th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'02),* pp. 244-251, IEEE Press, New York, 2002.

(7) J. Hays and I. Essa, "Image and Video Based Painterly Animation," in *Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering (NPAR '04)*, pp. 113-120, Annecy, France, June 7-9, 2004.

(8) M. Shiraishi and Y. Yamaguchi, "An Algorithm for Automatic Painterly Rendering Based on Local Source Image Approximation," in *Proceedings of the 1st International Symposium on Non-photorealistic Animation and Rendering (NPAR '00)*, pp. 53-58, ACM Press, New York, 2000.

(9) P. Haeberli, "Paint by Numbers: Abstract Image Representations," *Computer Graphics (SIGGRAPH 90 Conference Proceedings)*, 24(4):207-214, August 1990.

(10) P. Litwinowicz, "Processing Images and Video for an Impressionist Effect," in *SIGGRAPH 97 Conference Proceedings*, Turner Whitted editor, pp. 407-414, August 1997.

(11) S. M. F. Treavett and M. Chen., "Statistical Techniques for the Automated Synthesis of Non-Photorealistic Images," in *Proceedings of 15th Eurographics UK Conference*, pp. 15-21, March 1997.

(12) W. T. Freeman, D. B. Anderson, P. A. Beardsley, C. N. Dodge, M. Roth, C. D. Weissman, W. S. Yerazunis, H. Kage, K. Kyuma, Y. Miyake, and K. Tanaka, "Computer Vision for Interactive Computer Graphics," *IEEE Computer Graphics and Applications*, 18(3):42-53, May/June 1998.

(13) H. Chen, Z. Liu, C. Rose, Y. Q. Xu, H. Shum, and D. Salesin, "Example-Based

Composite Sketching of Human Portraits," in *Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering (NPAR '04)*, pp. 95-102, Annecy, France, June 7-9, 2004.

(14) A. Selle, A. Mohr, and S. Chenney, "Cartoon Rendering of Smoke Animations," *Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering (NPAR '04)*, pp. 57-60, Annecy, France, June 7-9, 2004.

(15) D. Sykora, J. Zara, and J. Burianek, "Unsupervised Colorization of Black-and-White Cartoons", *Proceedings of the 3rd International Symposium on Non-photorealistic Animation and Rendering (NPAR '04)*, pp. 121-137, Annecy, France, June 7-9, 2004.

received October 20, 2004
revised December 01, 2004
accepted December17, 2004

A Novel Moment Based Painterly Rendering Algorithm

Chung-Ming Wang Lung-Hung Yang
Department of Computer Science
National Chung Hsing University
Taichung 402, Taiwan, R.O.C.

Abstract

Painterly rendering is a popular research topic for Non-Photorealistic Rendering in computer graphics communities. Though moment-based painterly rendering algorithms have been presented with some success, they did not take into consideration the curved strokes encountered when artists are performing the painting works. In this paper, we present a novel moment-based painterly rendering algorithm. In particular, our algorithm employs the line integral convolution technique to construct the curved brush strokes. We also adopt a statistics method so that it is possible to automatically vary the stroke sizes corresponding to the local details in the image. In addition, we make use of the image segmentation technique to detect the edges in the contents of the image, ensuring the stroke not to across the boundaries of the objects. Finally, our algorithm adopts the Gaussian filter and considers the effect of canvas, demonstrating rendered images with more visual appealing features. Images thus rendered illustrate painterly rendering styles with curved strokes, versatile stroke sizes, and better visual appearance than our counterparts.

Keywords: Image Moment, Line Integral Convolution, Curved Brush Strokes, Painterly Rendering, Edge Detection

Design of Class E Resonant Inverter Incorporating Piezoelectric Transformer

Chang-Hua Lin*
Dept. of Computer and Communication Engineering
St. John's & St. Mary's Institute of Technology
Taipei 251, Taiwan, R.O.C.

Ying-Chi Chen
Dept. of Electronic Engineering
St. John's & St. Mary's Institute of Technology
Taipei 251, Taiwan, R.O.C.

Abstract

A novel single-stage backlight inverter constituted by class E topology for simplifying the circuit structure and promoting the system efficiency is proposed in this paper. Moreover, in order to overcome the drawbacks of the conventional electromagnetic transformer and miniaturize the backlight module, the piezoelectric transformer (PT) is adopted for driving the cold cathode fluorescent lamp. A simple model and design procedure are constructed by reasonable parameter simplification and combination schemes. Complete analysis and design considerations are discussed in detail. Experimental results agree with the theoretical prediction.

Keywords: backlight, resonant inverter, CCFL, LCD, piezoelectric Transformer

* Corresponding author. Tel.:+886-2-28013131 ext. 6388; fax: +886-2-28013413 ext. 6391
E-mail address : ljh@mail.sjsmit.edu.tw

I. INTRODUCTION

Due to the popularity of information apparatus, the requirements for thin and flat display panels are increasing drastically to replace conventional CRT displays used in many products with different panel sizes. The backlight module is a crucial component for driving light source in flat display panel (FDP) technologies, and its performance will influence the display quality of FDP. At present, there are two major backlight sources available, one is white LED, and the other is cold cathode fluorescent lamp (CCFL). Because of the high cost and small lighting area, the white LED is usually used in small-sized devices, such as PDA, GPS, and portable instruments. The CCFL is normally suitable for the larger display panels. The LCD combined with CCFL satisfies the increasing demand on display performance, size, and efficiency [1]. In the past, the conventional electromagnetic transformers were generally adopted in the backlight module design to boost output voltage for driving CCFL. However, the conventional transformer will cause the magnetic hysteresis loss and electromagnetic interference (EMI) and thus decrease the efficiency of energy conversion. In 1956, Rosen presented the direct and converse piezoelectric effect of piezoelectric components as the mechanism of mechanical and electrical energy interchange, which provides high conversion efficiency by transferring electrical energy via the polarized and mechanical vibration medium. Therefore, piezoelectric components can be employed in promoting and demoting power voltages as well. As a sinusoidal alternating voltage with frequency close to the resonant frequency is applied to the input terminal of a piezoelectric component, because the converse piezoelectric effect causes the piezoelectric component to yield the mechanical strain resonance and then the direct piezoelectric effect causes stress deformation of the resonance, the mechanical energy is transformed to electrical energy and outputted from the piezoelectric component, and hence the voltage transformation is achieved [3, 6]. Since PT can sustain its inherent high voltage gain while matching high impedance load with optimal efficiency, it is applicable to driving high-voltage high-impedance lamps, such as CCFL. In addition, PT possesses many advantages, namely, compact size, high efficiency, low power losses and free of EMI. Moreover, PT not only can provide high voltage gain but also is easy to realize the miniature design of FDP, in which the backlight module is joined with PT [5-6]. Section II depicts and analyzes the main circuit configuration employed in this paper. Section III introduces the model integration of class E resonant inverter with piezoelectric transformer. Sec-

tion IV discusses the design considerations and demonstrates the experimental results. Finally, Section V is conclusions.

II. ANALYSIS OF CLASS E RESONANT INVERTER

Fig. 1 shows the basic configuration of class E resonant inverter, which consists of choke inductor L_1, a power switch S, a shunt capacitor C_1, and a L_x-C_s-R_L series–resonant circuit. The following assumptions are essential to simplify the analysis procedures: 1) The power switch and diode are both ideal. 2) The choke inductor is high enough so that its ac component is much lower than the dc component of the input current. 3) The quality factor Q_L of the L_x-C_s-R_L series–resonant circuit is high enough so that the current i through the resonant circuit is sinusoidal.

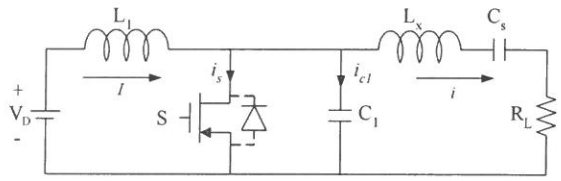


Fig. 1. Basic circuit scheme of class E converter

Fig. 2 depicts the equivalent circuits of class E resonant inverter. When the switch is on as Fig. 2a, the resonant circuit consists of L_x, C_s and R_L because C_1 is short-circuited by the switch. Hence, the resonant frequency f_{o1} is obtained as follows:

$$f_{01} = \frac{1}{(2\pi\sqrt{L_x C_s})} \tag{1}$$

However, when the switch is off as Fig. 2b, the resonant circuit consists of C_1, L_x, C_s and R_L connected in series. The resonant frequency f_{o2} can be expressed as below

$$f_{02} = \frac{1}{2\pi\sqrt{L_x C_s C_1 /(C_s + C_1)}} \tag{2}$$

From the preceding statement, f_{o1} is lower than f_{o2}, i.e., $f_{01} < f_{02}$. In general, the operating frequency f is selected between f_{01} and f_{02} to acquire an approximate sinusoidal load current, i.e.

$$f_{01} < f < f_{02} \tag{3}$$

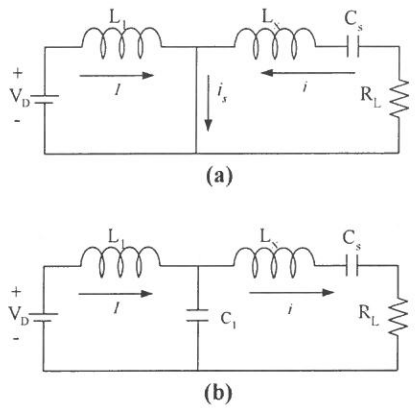


Fig. 2. Equivalent circuit models of class E inverter, (a) S on, (b)S off

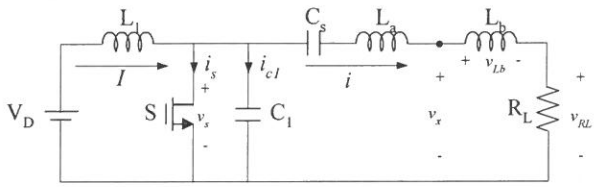


Fig. 3. Equivalent circuit of Class E resonant inverter

For $f > f_{01}$, the series-resonant circuit represents an inductive load shown as Fig. 3, where L_x can be divided into L_a and L_b. Assuming that the resonant frequency formed with L_a and C_s is equal to the operating frequency f, that is

$$\omega = \frac{1}{\sqrt{L_a C_s}} \tag{4}$$

The quality factor Q_L is

$$Q_L = \frac{\omega L_x}{R_L} = \frac{\omega(L_a + L_b)}{R_L} = \frac{1}{\omega C_s R_L} + \frac{\omega L_b}{R_L} \tag{5}$$

Then the resonant current i, which through the resonant circuit, can be represented as sinusoidal form.

$$i = I_m \sin(\omega t + \phi) \tag{6}$$

where I_m and ϕ are amplitude and phase angle of the resonant current i, respectively.

Therefore, the currents in switch S and shunt capacitor C_1 can be expressed as

$$i_s + i_{c1} = I - i = I - I_m \sin(\omega t + \phi) \tag{7}$$

For the time interval $0<\omega t\leq 2\pi D$, the switch is on and therefore $i_{c1}=0$. Oppositely, for the time interval $2\pi D <\omega t\leq 2\pi$, the switch is off, which implies that $i_s=0$. Hence, the current through the shunt capacitor C_1 is given by

$$i_{C1} = \begin{cases} 0 & \text{for } 0 < \omega t \leq 2\pi D \\ I - I_m \sin(\omega t + \phi) & \text{for } 2\pi D < \omega t \leq 2\pi \end{cases} \tag{8}$$

The voltage across the shunt capacitor C_1 and the switch is found as

$$\begin{aligned} v_S &= \frac{1}{\omega C_1} \int_{2\pi D}^{\omega t} i_{C1} d(\omega t) \\ &= \begin{cases} 0 & \text{for } 0 < \omega t \leq 2\pi D \\ \dfrac{1}{\omega C_1}\{I(\omega t - 2\pi D) + I_m[\cos(\omega t + \phi) - \cos(2\pi D + \phi)]\} & \text{for } 2\pi D < \omega t \leq 2\pi \end{cases} \end{aligned} \tag{9}$$

In optimum operation [12], substituting the boundary condition $v_s=0$ at $\omega t=2\pi$ into (9) yields the amplitude of resonant current

$$I_m = I \frac{2\pi(1-D)}{\cos(2\pi D + \phi) - \cos\phi} \tag{10}$$

Substitution of (10) into (8) and (9), we can obtain i_{c1} and v_s as follows:

$$\frac{i_{C1}}{I} = \begin{cases} 0 & \text{for } 0 < \omega t \leq 2\pi D \\ 1 - \dfrac{2\pi(1-D)\sin(\omega t + \phi)}{\cos(2\pi D + \phi) - \cos\phi} & \text{for } 2\pi D < \omega t \leq 2\pi \end{cases} \tag{11}$$

$$v_S = \begin{cases} 0 & \text{for } 0 < \omega t \leq 2\pi D \\ \dfrac{I}{\omega C_1}\{\omega t - 2\pi D + \dfrac{2\pi(1-D)[\cos(\omega t + \phi) - \cos(2\pi D + \phi)]}{\cos(2\pi D + \phi) - \cos\phi}\} & \text{for } 2\pi D < \omega t \leq 2\pi \end{cases} \tag{12}$$

Rearranging (12), the input voltage V_D is

$$V_D = \frac{1}{2\pi} \int_{2\pi D}^{2\pi} v_S d(\omega t) = \frac{I}{\omega C_1}\left\{\frac{(1-D)[\pi(1-D)\cos\pi D + \sin\pi D]}{\tan(\pi D + \phi)\sin\pi D}\right\} \tag{13}$$

Substituting the boundary condition for optimum operation [12], $dv/d(\omega t)=0$ at $\omega t=2\pi$ into (13) yields the relationship between phase ϕ and duty cycle D.

$$\phi = \pi + \arctan\left\{\frac{\cos 2\pi D - 1}{2\pi(1-D) + \sin 2\pi D}\right\} \tag{14}$$

Since the resonant current i can be regarded as sinusoidal form, the fundamental component of v_x at operating frequency is

$$v_x = V_{RL}\sin(\omega t + \phi) + V_{Lb}\cos(\omega t + \phi) \tag{15}$$

Combing V_{RL} in (15) with (12) by Fourier formula yields

$$V_{RL} = \frac{1}{\pi}\int_{2\pi D}^{2\pi} v_s \sin(\omega t + \phi)d(\omega t) = -\frac{2\sin\pi D\sin(\pi D + \phi)}{\pi(1-D)}V_D \tag{16}$$

And combing V_{Lb} in (15) with (13) by Fourier formula yields

$$\begin{aligned} V_{Lb} &= \omega L_b I_m = \frac{1}{\pi}\int_{2\pi D}^{2\pi} v_S \cos(\omega t + \phi)d(\omega t) \\ &= \frac{1 - 2(1-D)^2\pi^2 - 2\cos\phi\cos(2\pi D + \phi)[\cos 2\pi D - \pi(1-D)son2\pi D]}{2(1-D)\pi\cos(\pi D + \phi)[(1-D)\pi\cos\pi D + \sin\pi D]}V_D \end{aligned} \tag{17}$$

Combing (10), (13) and (16) yields

$$C_1 = \frac{2\sin\pi D\cos(\pi D + \phi)\sin(\pi D + \phi)[(1-D)\pi\cos\pi D + \sin\pi D]}{\pi^2(1-D)\omega R_L} \tag{18}$$

By combing (10), (13), (17) and (18) yields

$$\frac{\omega L_b}{R_L} = \frac{2(1-D)^2\pi^2 - 1 + 2\cos\phi\cos(2\pi D + \phi) - \cos 2(\pi D + \phi)[\cos 2\pi D - \pi(1-D)\sin 2\pi D]}{4\sin\pi D\cos(\pi D + \phi)\sin(\pi D + \phi)[(1-D)\pi\cos\pi D + \sin\pi D]} \tag{19}$$

Assuming that the main circuit operates in optimum operation, that is, duty cycle D=0.5 [12], one can obtain the phase $\phi=147.52^\circ$ from (14). Substituting above-mentioned parameters into (18) and (19), then (18) can be simplified as

$$C_1 = \frac{8}{\pi(\pi^2 + 4)\omega R_L} \tag{20}$$

And (19) is rearranged as

$$\frac{\omega L_b}{R_L} = \frac{\pi(\pi^2 - 4)}{16} \tag{21}$$

From (5) and (21), we can derive C_s as follows:

$$C_S = \frac{1}{\omega R_L \left[Q_L - \frac{\pi(\pi^2 - 4)}{16} \right]} \tag{22}$$

The resonant inductor L_x is calculated from (5).

$$L_X = \frac{Q_L R_L}{\omega} \tag{23}$$

III. MODEL INTEGRATION OF CLASS E INVERTER WITH PIEZOELECTRIC TRANSFORMER

A Rosen type PT, as shown in Fig. 4, is composed of two poled piezo-ceramic plates with equal cross-sections, or a single piezo-ceramic plate with both ends poled separately. Moreover, PT is based on the use of both piezoelectric effects, in which an applied electric energy is transformed into mechanical vibrations and then these vibrations are transformed back into electric energy [13]. In other words, the input terminals are supplied with an alternating electric field that produces a time varying strain in the ceramic material. Then this strain is electromechanically coupled along the length of the plate, and induced a polarization field between the two output terminals [14]. In this paper, PT is employed to integrate class E resonant inverter for simplifying the circuit structure and promoting the system efficiency. Fig. 5 shows the employed class E resonant inverter, which incorporates a piezoelectric transformer, for driving CCFL. When the operating frequency of the circuit is close to the resonant frequency of the employed inverter, the harmonics and DC component of the current flowing in the circuit will be filtered out due to the high quality factor QL of the main circuit [3-4]. Therefore, sinusoidal voltage and current can be produced to start and drive CCFL. Fig. 5a illustrates the equivalent main circuit of Fig. 1 in steady state, in which a PT equivalent model, external resonant components, and a secondary reflected impedance are included to form a

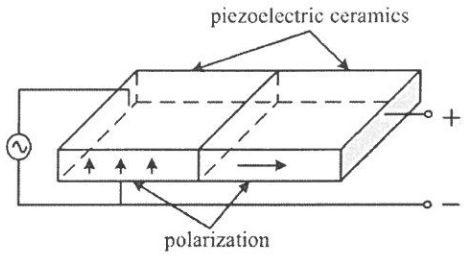


Fig. 4. Schematic of Rosen type piezoelectric transformer

resonant tank [1, 8]. The secondary reflected impedance comprises lamp impedance and parasitic capacitance. The mathematical models of derived circuits presented in [4-5] are too complicated due to numerous internal equivalent parameters of PT. Therefore, reasonable parameter simplification and combination schemes are utilized to decide the desired parameters of circuit components and to derive a simple mathematical model. To simplify the circuit model, the parasitic capacitance is neglected and the CCFL is replaced by RCCFL. The secondary components of PT are converted equivalently to the primary ones of PT and the resultant circuit is shown in Fig. 5b. When PT is operating in its resonant frequency, R, L, and C elements in the circuit model are considered as short-circuited due to the voltages of inductor L and capacitor C having the same magnitudes but opposite polarities. Thus, we can further simplify the circuit to be equivalent to a resonant circuit structure as shown in Fig. 5c. Fig. 5d illustrates the reduction circuit model, where

$$C_K = C_{01} + N^2 C_{02} \tag{24}$$

Fig. 5d also can be converted into Figs. 5e or 5f by circuit transformation technique. From the (24), R_{CCFL} and N in Fig. 5c, the equivalent load R_L is found as

$$R_L = \frac{\frac{R_{CCFL}}{N^2}}{1+\left(\frac{R_{CCFL}}{N^2}\Big/X_{CK}\right)^2} \tag{25}$$

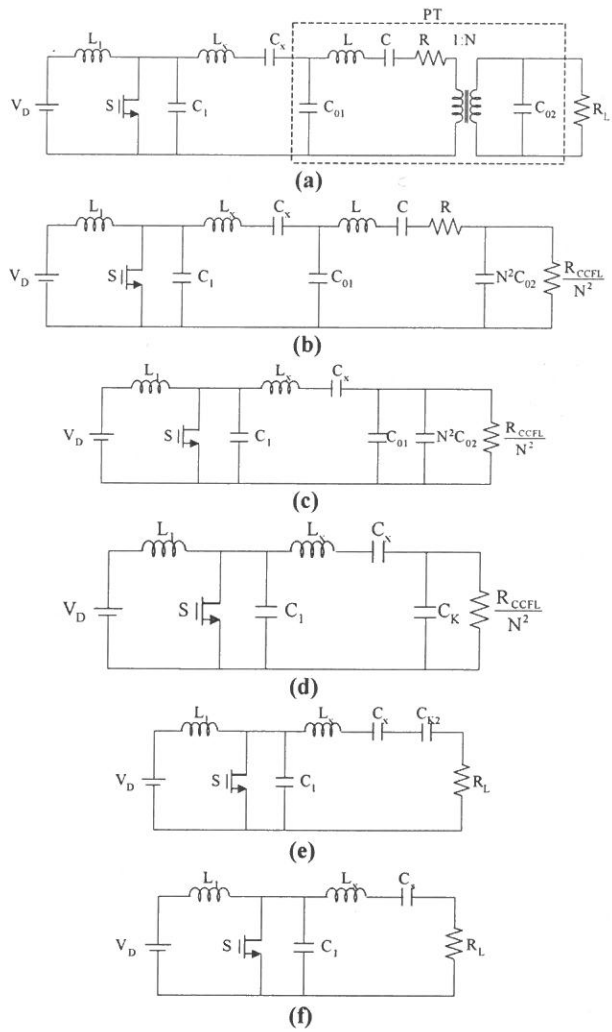


Fig. 5. (a) Equivalent circuit model of class E resonant inverter with PT, (b) equivalent circuit of 5a, (c) simplified circuit of 5b, (d) simplified circuit of 5c, (e) simplified circuit of 5d . (f) simplified circuit of 5e .

Substituting R_L into (20), (22) and (23), we can acquire C_1, C_s and L_x, respectively. From Figs. 5d and 5e, the impedance of C_{K2} can be expressed as

$$X_{CK2} = \frac{X_{CK}}{1+\left(X_{CK}\Big/\frac{R_{CCFL}}{N^2}\right)^2} \tag{26}$$

Owing to the C_{K2} can be expressed as

$$C_{K2} = \frac{1}{\omega X_{CK2}} \tag{27}$$

Hence, the equivalent capacitor C_s can be calculated from Figs. 5e and 5f.

$$C_s = \frac{C_x C_{K2}}{C_x + C_{K2}} \tag{28}$$

Therefore, the resonant capacitor is obtained from (28) as follows:

$$C_x = \frac{C_s C_{K2}}{C_{K2} - C_s} \tag{29}$$

The quality factor Q_L is an important parameter to determine the resonant components in this paper. The reasonable value of Q_L will be decided in the following discussion. First, due to $C_x > 0$ in (29), the equivalent capacitor C_{K2} should higher than C_s. Substituting C_{K2} and R_L into (22) yield Q_{LA}, which is the lower boundary value.

$$Q_{LA} = \frac{1}{\omega R_L C_{K2}} + 1.1525 \tag{30}$$

Moreover, quality factor Q_L will also directly affect the resonant components L_x and C_s. Hence, we can define a parameter γ by combing (22) and (23) as follows:

$$\gamma = \frac{1}{LxCs\omega^2} = \frac{Q_L - 1.1525}{Q_L} \tag{31}$$

The relationship of γ and Q_L can be depicted by Matlab simulation.

From Fig. 6, it is noted that γ seems unaffected by Q_L when $\gamma > 0.9$, in other words, γ is almost at saturation state; However, it is apparent that γ is susceptible to Q_L when $\gamma < 0.9$. Then we can set the corresponding value of quality factor (Q_{LB}) at $\gamma = 0.9$ as the higher boundary value. Therefore, we can acquire a reasonable range of quality factor, that is, $Q_{LA} < Q_L < Q_{LB}$.

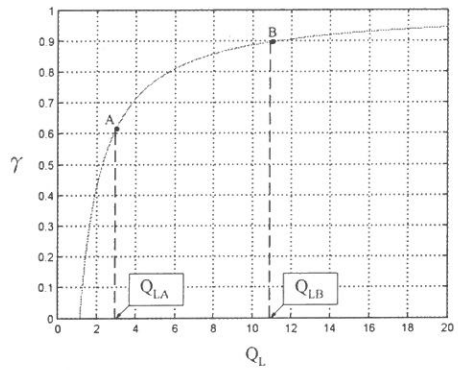


Fig. 6. Relationship of γ and Q_L.

IV. DESIGN CONSIDERATION AND EXPERIMENT RESULTS

CCFL, which is characterized by its length, diameter, and structure, is inherently a nonlinear load and hence will influence the design of driving circuit. In general, the best lamp current for driving CCFL is sinusoidal, i.e., its crest factor near 1.414, which dose not only reduce EMI but also raises efficiency [1]. Although other wave shapes may provide higher luminance, nevertheless the lamp life will be shortened [10]. Additionally, the conventional electromagnetic transformer has not only bulky volume, which is not adequate for designing thin and flat display panel, but also causes magnetic hysteresis and electromagnetic interference, which decrease the efficiency of energy conversion [8]. Therefore, a PT is employed to overcome the preceding disadvantages. Moreover, PT possesses high efficiency, high gain of conversion performance and low power dissipation [5, 6, 8, 11]. The PT utilized in the paper is EFTU14R0M02, which has rated power of 4W, resonant frequency of 57 kHz, input voltage of 22 $V_{rms(max)}$, input current of 500 $mA_{rms(max)}$, output voltage of 820 $V_{rms(max)}$, output current of 7 $mA_{rms(max)}$, and operating temperature range of -10℃~60℃. The CCFL used is FL-30266AE, which has rated power of 3.4 W, operating voltage v_ℓ of 620 V_{rms}, operating current i_ℓ of 5±1 mA, and starting voltage v_{start} of 930 V_{rms}. The operating frequency of CCFL is generally within the range of 20~80 kHz. In this paper, we select the operating frequency f_o of 57 kHz and input voltage $V_D = 12V_{DC}$. The most important parameters in the proposed circuit structure are the internal equivalent impedance R_L of PT as well as resonant components C_1, L_x, C_x of class E resonant inverter. The design considerations and steps of these parameters are discussed below

1) Measuring the parameters of the equivalent PT model

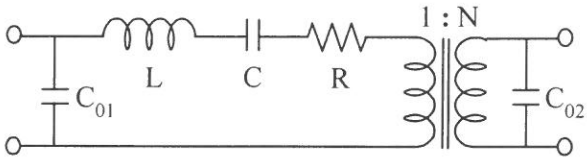


Fig. 7 Equivalent internal model of PT.

Fig. 7 shows the equivalent circuit model of PT operating in its resonant frequency. The parameters of the employed PT are measured with an impedance analyzer HP-4194A. The related parameter values measured are shown below

C_{01} = 99.66 nF L = 1.22 mH C = 6.21 nF
R = 1.26 Ω C_{02} = 9.09 pF N = 55

2) Calculating proper values of R_L and C_{K2}

In steady state, the lamp voltage and lamp current employed in this paper are 620 v_{rms} and 5 mA_{rms}, respectively. Hence, the equivalent lamp impedance is R_{CCFL}= $620v_{rms}/5mA_{rms}$ =124 kΩ. In addition, f = 57kHz is a better operating frequency of the employed PT under room temperature. According to (24)-(27) and above-mentioned parameters, we can obtain R_L=9.17 Ω, and C_{K2}=163.5 nF.

3) Estimating the quality factor Q_L and C_s:

From (30) and C_{K2} estimated in previous step, we can obtain the lower boundary value Q_{LA}=3.0148. According to the previous statement, the higher boundary value Q_{LB}=11.525 is found by substituting γ=0.9 into (31). Therefore, the reasonable value of Q_L is in the range of 3.0148< Q_L <11.525, we choose Q_L=8 in the practical design. Substituting R_L and Q_L into (22) yields C_s=45.8 nF.

4) Determining parameters of resonant circuit

Substituting Q_L, C_s, R_L and parameters of PT into (20), (23) and (29), we can calculate the shunt capacitor C_1=55.9 nF (use 57 nF), the resonant components L_x=204.8 μH (use 201 μH) and C_x=61 nF (use 68 nF), respectively. All the estimated parameters are substituted into (1) and (2) to yields f_{01}=52.72 kHz and f_{02}=71.11 kHz. Hence, the operating frequency f=57 kHz employed in this paper is satisfied the requirement of (3).

Fig. 8 demonstrates the measured waveforms of gating signal v_{gs} and the switch voltage v_{ds} under ideal operating mode (D=0.5) with zero voltage switching (ZVS) [12]. Fig. 9 shows the measured waveforms of switch current i_s and the shunt capacitor current i_{c1}, which forms an approximate sinusoidal current to drive the load. The choke inductor current I and resonant current i are illustrated in Fig. 10, the measured phase difference is ϕ=139° , which is close to the derived phase difference ϕ=147.52° , with error of 8° (below 6%). From the experimental results, the measured waveforms of the input current of PT and lamp current i_{CCFL} are both sinusoidal type as shown in Fig. 11.

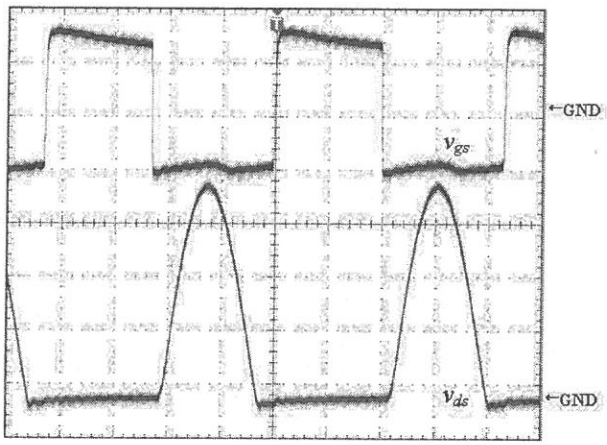


Fig. 8.Measured waveforms of v_{gs}, v_{ds} (Ver: 5V/div for v_{gs}; Ver: 10V/div for v_{ds}; Hor: 4μs/div)

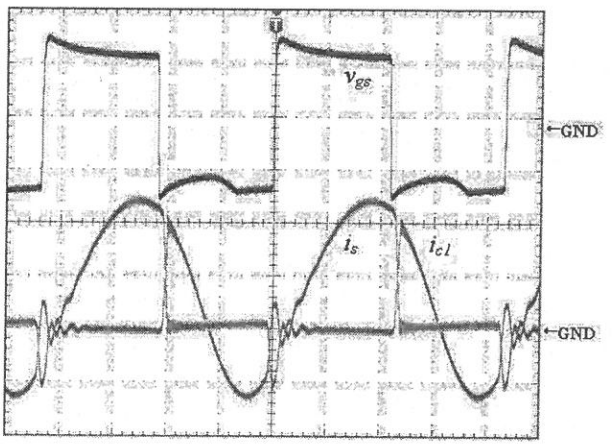


Fig. 9. Measured waveforms of v_{gs}, i_s and i_{c1} (Ver: 5V/div for v_{gs}; 1A/div for i_s; Ver: 500mA/div for i_{c1}; Hor: 4μs/div)

V. CONCLUSIONS

In this paper, a class E resonant inverter is incorporated with a piezoelectric transformer to drive the cold-cathode fluorescent lamp with zero voltage switching. The model integration and simplification are both adopted to derive the optimal parameters of backlight system, so as to promote the system efficiency and stability. In addition, the overall volume is reduced and thin-shaped to miniature the backlight module. The overall system efficiencies are all above 91% in all test conditions.

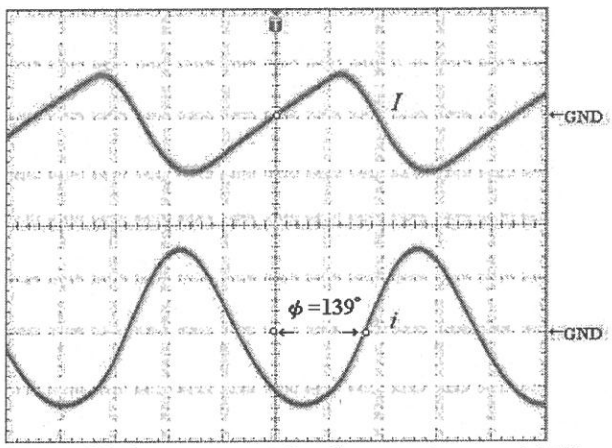


Fig. 10. Measured waveforms of inductor current *I* and resonant current *i* (Ver: 100mA/div for *I* ; Ver: 500mA/div for *i* ; Hor: 4μs/div)

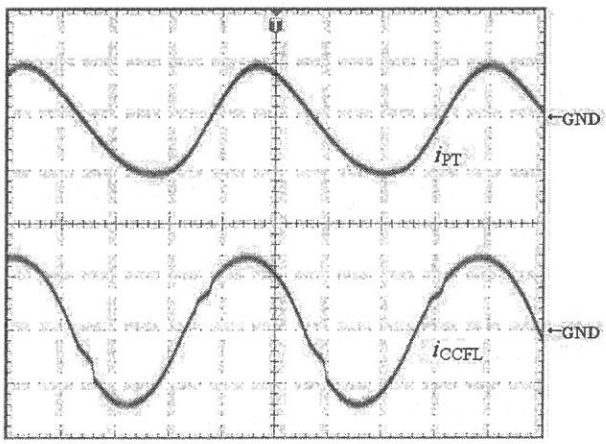


Fig. 11. Measured waveforms of input current of piezoelectric transformer i_{PT} and lamp current i_{CCFL} (Ver: 1A/div for i_{PT} ; Ver: 5mA/div for i_{CCFL} ; Hor: 4μs/div)

ACKNOWLEDGMENT

This work was sponsored by the National Science Council, Grant NO. NSC93-2213-E-129-014.

REFERENCES

(1) M. Jordan and J. A. O'Connor, "Resonant fluorescent lamp converter provides efficient and compact solution," *IEEE-APEC'93 Conference Record*, pp. 424-431, 1993.

(2) N. O. Sokal and A. D. Sokal, "Class E-a New Class of High-Efficiency Tuned Single-Ended Switching Power Amplifiers," IEEE J. of Sol. Circ., vol. 10,no.3,pp.168-176, 1975.

(3) D. O. Hur, T. K. Kang, C. H. Cho, H. M. Lee, H. K. Ahn and D. Y. Han, "Design and fabrication of piezoelectric ceramic transformers for the LCD backlight," *IEEE Proc. ICPAD'97,* vol.2, pp. 843 -846, 1997.

(4) C. S. Moo, W. M. Chen and H. K. Hsien, "Electronic ballast with piezoelectric transformer for cold cathode fluorescent lamps," *IEE Proc. Electric Power Appli.*, vol.150, pp. 278 -282, 2003.

(5) C. D. Wey, T. L. Jong and C. T. Pan, "Design and analysis of an SLPT-based CCFL driver," *IEEE Trans. Ind. Electron.,* vol.50, pp. 208 -217, 2003.

(6) S. J. Choi, K. C. Lee and Bo H. Cho, "Design of fluorescent lamp ballast with PFC using a power piezoelectric transformer," *IEEE-APEC'98 Conference Record*, vol.2, pp. 1135 -1139, 1998.

(7) *Kazimierczuk, M.K.; Bui, X.T.;"* Class E AC/DC converters with an inductive impedance inverter" IEEE Transactions on , Volume: 4 Issue: 1, pp. 124 –135, , Jan. 1989.

(8) Ray L. Lin, Fred C. Lee, Eric M. Baker and Dan Y. Chen "Inductor-less piezoelectric transformer electronic ballast for linear fluorescent lamp," *IEEE-APEC'01 Conference Record*, vol. 2, pp. 664 –669, 2001.

(9) G. C. Chryssis, *High-Frequency Switching Power Supplies*, McGraw-Hill Inc., 1995.

(10) *Applications Handbook*, Micro Linear, AB-8, Mar. 1998.

(11) B. Y. Sam and Simon Lineykin "Frequency tracking to maximum power of

piezoelectric transformer HV converters under load variations," *IEEE-PESC'02 Conference Record*, pp. 657 –662, 2002.

(12) M. K. Kazimierczuke, and D. Czarkowski, "Resonant Power Converter" New York: Wiley, 1995.

(13) T. Hemsel, W. Littmann, and J. Wallaschek, "Piezoelectric transformers-state of the art and development trends," *IEEE-ULTRASONICS Symposium'02 Conference Record*, vol. 2, pp. 645 –648, 2002.

(14) E. M. Syed, and F. P. Dawson, "Analysis and modeling of a Rosen type piezoelectric transformer," *IEEE-PESC'01 Conference Record*, vol. 3, pp. 1761–1766, 2001.

received October 21, 2004
revised November 28, 2004
accepted December 27, 2004

E 類諧振換流器結合壓電變壓器之設計

林 長 華
電腦與通訊工程系
聖約翰技術學院

陳 英 碩
電子工程系
聖約翰技術學院

摘　要

本文中提出一種由 E 類架構所組成之新型的單級背光換流器，可簡化電路架構及提升系統效率。其次，爲了克服傳統電磁式變壓器之諸多缺點，並達到背光模組小型化的目的，採用壓電變壓器來驅動冷陰極管。再者，本文中利用合理的參數簡化及合併的技巧來建構簡單的系統模型，而完整的分析及設計準則也將詳細地討論。最後，以實際量測結果來驗證理論的正確性。

關鍵詞：背光、諧振換流器、冷陰極管、液晶顯示器、壓電變壓器

Application of Wavelet Methods to the Detection of Galactic Central Structures

Chien-Chang Yen*
Department of Mathematics,
Fu Jen Catholic University,
Hsinchuang, Taipei Hsien

Chi Yuan
Institute of Astronomy and Astrophysics,
Academia Sinica,
Taipei

I-Hui Li
Department of Earth Science,
National Taiwan Normal University,
Taipei

Abstract

This paper presents results of application of wavelet methods to the galactic central regions. Most of the nearby galaxies are found to have a gas-dust disk in the central regions. However, the details structures of the disk are often obscured by the background luminous star lights. Wavelet methods can extract these hidden structures from the observed data. Atrous wavelet method is simple and extremely useful for such purpose. We analyze the NICMOS and WFPC (WFPC2) data from HST for certain disk galaxies. The results show that some major spiral arms outside can go all the way into the central regions (eg. NGC 5383, NGC 4321) and some galaxies have bars in the nuclear center (eg. NGC 1068, NGC 1097).

* Corresponding author. Tel: 2-29053547; Fax:2-29044509
E-mail:yen@math.fju.edu.tw

These results are important for us to understand the starburst and AGN phenomena.

Keywords: wavelet, atrous algorithm, galaxy, spiral, bar, ring

1. INTRODUCTION

Recent high-resolution observations of galactic central region in disk galaxies have revealed nuclear structures on scales of a few hundred parsecs. Most of the nearby galaxies are found to have a central gas-dust disk. These nuclear structures could be bars, spirals and rings. However, they are often obscured by the background luminous star lights. People have tried different methods to extract these hidden structures from central regions, for instance, by unshape mask [Malin, 1977], and NIR color index images [Knapen 1995b]. In this paper, we use wavelet methods to analyze data for galaxies NGC 5383, NGC 1097, NGC 1068 and NGC 4321.

Density wave theory developed by Lin & Shu [Lin & Shu, 1964] has successfully explained the spiral structure in disk galaxies. We will employ this theory to explain the detected structures in the galactic center. The shapes of nuclear spirals can be inferred from the density wave theory. We show easily that these spirals are waves, since material arms in disk galaxies cannot last after many rotations due to the differential rotation in the disk. Any material arms would be quickly sheared and become tightly wound with the pitch angle tending to zero, and no steady state can be achieved.

Our studies focus on probing the galactic central regions with the help of wavelet methods. The atrous wavelet method is simple and extremely useful. We analyze the NICMOS and WFPC (WFPC2) data from HST for a few nearby disk galaxies. The results show that some major spiral arms can go deep into the central regions (eg. NGC 5383, NGC 4321) and others have bar structure (eg. NGC 1068, NGC 1097).

2. WAVELET METHODS

Wavelet has been developed in the past thirty years and it is applied on many fields in the past few years. At first it was only used in limited fields for special purposes: engineering

(subband coding), physics (coherent states, renormalization group), and pure mathematics (study of Calderon-Zygmund operators). In recent years, wavelet methods have carved out a practical niche and had a tremendous impact on signal and image processing, multi-fractal analysis, statistics, etc[Bacry et al,1993; Papoulis 1991].

It is important to point out that wavelet methods can be applied on analyzing image because they provide a multi-scale analysis [Daubechies 1992]. A signal u_0 which is the information at the finest scale 0 can be decomposed into the coarsest averaging data u_J plus various differencing data w_j, where $j = 1,\ldots,J$, at various scales. The above process is called wavelet transformation and it can be achieved by Daubechies' orthogonal wavelets (Daub) [Daubechies 1988], Choen-Daubechies-Feauveau wavelets (CDF) [Cohen et al 1992, Daubechies 1992], Chui-Wang wavelets (CW) [Chui 1992], difference wavelet (DW) [Chern & Yen 2002] or atrous algorithm. For data compression (reducing size of images but preserving the quality to make no difference between the original and reduced one), Daub, CDF, CW and DW will be good candidates. For imaging analysis, significant feature appears at certain scale that the atrous wavelet is better choice than others in our experience.

Wavelet methods allow us to decompose any image into its constituent parts based on their size scales. Some of the most important advantages are (1) the ability to remove smoother structures which are more intense and which therefore obscure the underlying smaller scale features, (2) the ability to detect sharp boundaries which will define interfaces between structures so that it can be used for recognition purposes in different images, and (3) the ability to enhance contrast by suppressing or removing various wavelet components. The following experiment shows the fact that wavelet is useful to detect the structures of galactic central regions.

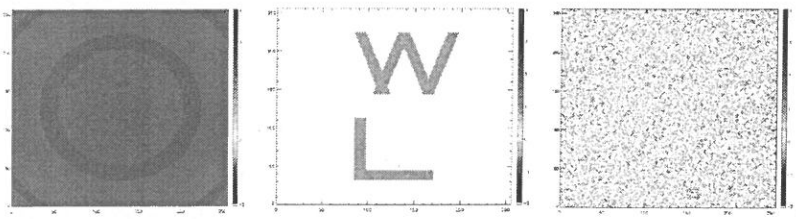


Fig. 1. The above images from left to right are background, true signal and noise, respectively. Their magnitudes are of the ratio 100:10:1. They are correspondent to background luminous star lights, structure of galactic central regions and noise.

Those images in Fig. 1 from left to right are background, true signal and noise, respectively. Their magnitudes are of the ratio 100:10:1. They are correspondent to background luminous star lights, structure of galactic central regions and noise. All of these three images are combined into the left-hand-side image of Fig. 2. The true signal is unclear due to the strong background sources. The right-hand-side image of Fig. 2 revealed its structure by employing wavelet method to extract the true signals. Among many wavelets, the atrous algorithm is better than others. For the reason of completion, the atrous algorithm is given by

$$I_0 = ORIGIN \quad IMAGE$$

$$I_k = I_{k-1} * B_3(\frac{x}{2^k}, \frac{y}{2^k})$$

$$w_k = I_k - I_{k-1}, \quad k \geq 1$$

where B_3 is the cubic B-spline function and we call w_k the k th-wavelet plane. In short, the information between level k th and m th wavelet planes helps us to analyze the signal (observed data).

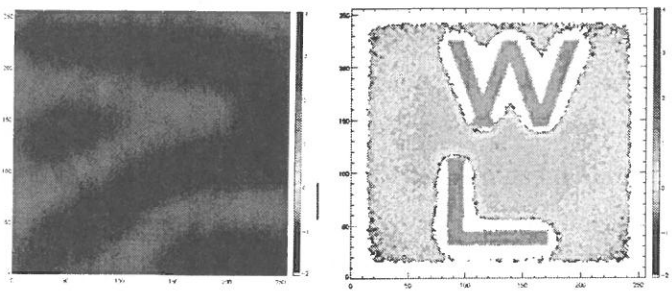


Fig. 2. These images demonstrate the useful wavelet methods. Wavelets can extract the hidden structure from a real signal coupled with strong background sources. The left image is the test signal (observed data) and the right image has been detected by wavelet methods and the structures are revealed.

3. RESULTS

We present our results in two categories: one for galaxies with a major bar and one without a major bar.

3.1 Galaxies with major bar

3.1.1 NGC 5383

NGC 5383 is a barred spiral galaxy of type SBb in the constellation Canes Venatici. This galaxy has a bright nucleus and weak arms. The coarsest resolutoin (i.e., ground-based image) can produce artifacts that resemble bars or rings. Preliminary analysis of the rotation curve according to the BIMA observations (Sheth et al 2000) suggests that NGC 5383 has an inner inner Lindblad resonance (IILR). Sheth et al (2000) use unsharp-mask method on HST-NICMOS image and reveal a pair of trailing spirals in the very center but not a bar. This result overthrows their previous suggestion. The HST-NICMOS observation by Sheth et all (2000) show two phenomena: (1) The main bar dust lane can be seen in absorption as it travels inward the nucleus. (2) Collection of star formation regions is at the north-northwest of the relatively smooth nucleus. They also claim that a pair of trailing spirals which starts from the main northwestern dust lane winds its way down to an unresolved core. In their observation, there is no evidence of a nuclear bar. The result in Fig. 3 by wavelet method has confirmed the unsharp mask work of Sheth et al. Both methods are applied on HST-NICMOS data. The conclusion is that a pair of trailing spirals connects to dust lanes at one end and the other end goes all the way into the central region.

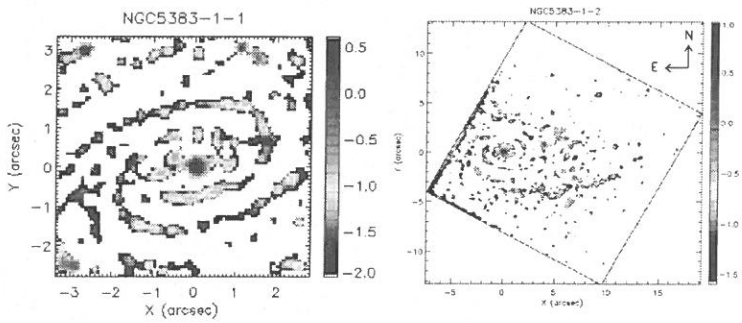


Fig. 3. These images demonstrate that a pair of trailing spirals connects to dust lanes at one end and the other end goes all the way into the central region in different wavelet planes, 1-1 and 1-2.

3.1.2 NGC 1097

Classified as a SB-type active galaxy, NGC 1097 has been the target of extensive multi-wavelength studies. This galaxy consists of a mini nuclear bar and star formation ring which

is actually two tightly wound trailing spiral arms. The mini bar is almost perpendicular to the primary bar and is of size approximately to the diameter of the ring (Quillen et al., 1995). Intriguingly, the role of the mini bar is to drive spirals.

The most remarkable feature is the bright, almost circular star formation (SF) ring (diameter $\approx 18^{\circ} = 1.6$ kpc), which is typical for the nuclear rings surrounding galactic nuclei (Kotilainen et al. 2000). In the near IR observations, the ring can be resolved into tightly wound spiral arms (Forbes et al., 1992). The clumpy morphology that varies strongly with wavelength is due to a combination of extinction, emission from host dust and red supergiants, and the age of the stellar population in the rings (Kotilainen et al. 2000). In the result of wavelet methods as shown in Fig. 4, this star formation ring is actually a pair of spiral arms extending all the way to the center. These two spiral arms are trailing, tightly wound, and clumpy.

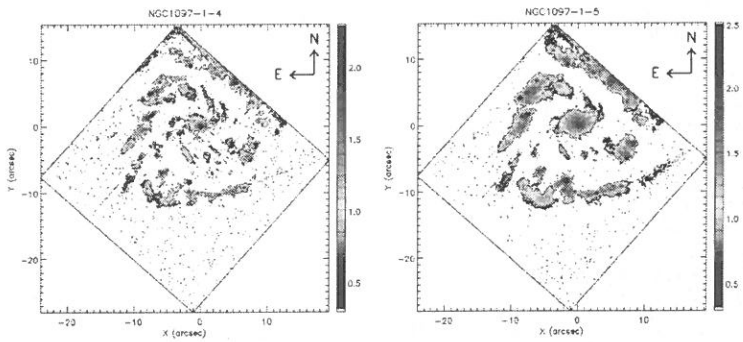


Fig. 4. In the result of wavelet methods, this star formation ring is actually a pair of spiral arms extending all the way to the center. These two spiral arms are trailing, tightly wound, and clumpy.

3.2 Galaxies without major bar

3.2.1 NGC 1068

Being the nearest Seyfert galaxy, NGC 1068 (M77) is one of the so-called Seyfert 2 galaxy which have broad and strong radio sources in the galaxy's inner regions. The detailed structures do not come visible with large telescopes. Morphologically it is classified as a Sab-type galaxy but it is more often considered as a barred galaxy. Although originally classified as a Seyfert 2 galaxy based on its direct-path, NGC 1068 may have a hidden Seyfert 1 nucleus. From the recent observations, most of the ^{12}CO(2-1) emission is in two spiral arms

with diameter $\approx 40^{'}$. A stellar bar of diamenter $\approx 20^{'}$ is seen in the inner NIR, which ends join the spiral arms in the molecular gas (Schinnerer et al., 2000).

The wavelet technique on the HST-NICMOS reveals that in the central region, there are two spiral arms with a diameter of $\approx 40^{'}$ (Schinnerer et al., 2000). They are consistent with the CO observation and there is a bar structure (the right-hand-size image of Fig. 5) which is connected to two spiral arms at nearly 90 degree. The WFPC data (the left-hand-size image of Fig. 5) shows a more complex bar-spiral structure in the central region.

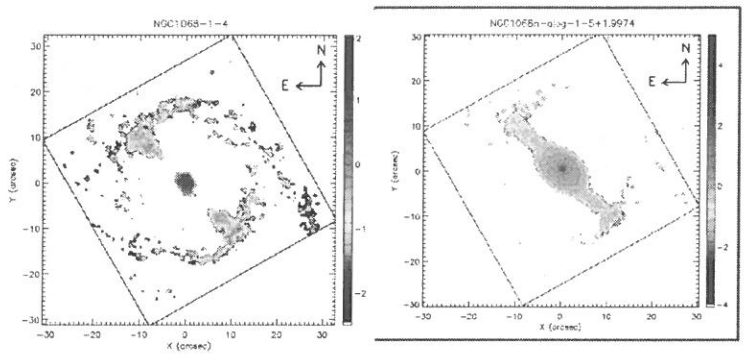


Fig. 5. The left image corresponds to wavelet plane 1-4 shows the two arms spiral structure. The right image corresponds to wavelet plane 1-5 demonstrates the bar in large scale.

3.2.2 NGC 4321

NGC 4321 (M100) is a two-armed grand design spiral galaxy in the Virgo Cluster, classified as a SAB(s)bc type galaxy with a bar in size $\approx 60^{''}$ bar (Knapen et al, 1995a). As a grand-design spiral structure, its gas in the spiral arms contributes significantly to star formation and its inner region has been extensively studied at several wavelengths. It was claimed by Knapen (1995a, 1995b) that the central regions have four arms spiral structure and they are leading, but the leading spirals do not appear in wavelet methods. A nuclear bar almost perfectly aligns with the large-scale bar and occupies the region inside the ring. Furthermore, the morphological features in the 2.2 μm image reveals an inner bar aligned with the 5 kpc stellar bar and a pair of leading arms emerging from its ends (Knapen et al., 1995b). The large-scale trailing shocks represent the offset dust lanes in the bar.

A late-type barred galaxy with a prominent nuclear ring of star formation is claimed to lie in the vicinity of the ILR(s) (Pierce, 1986; Arsenault et al., 1988), and two armed molecu-

lar spiral arms starting at the ends of the 1.3 kpc long nuclear stellar bar have been in the NIR images (Sakamoto et al., 1995). Optically, the light is dominated by two spiral arms in a central zone of enhanced star formation (Knapen et al., 1995b). This nuclear starburst is induced and maintained by a global bar-driven density wave (Shlosman et al., 1989; Yuan & Kuo, 1998). However, the results in Fig. 6 by wavelet method demonstrate that (1) there seems to have a small bar in the center that drives two spiral arms into the center. (2) The dust spiral pattern shown in R-band is consistent with CO observation (Sakamoto et al., 1995) , and (3) there are no leading spirals as claimed by Knapen (1995a).

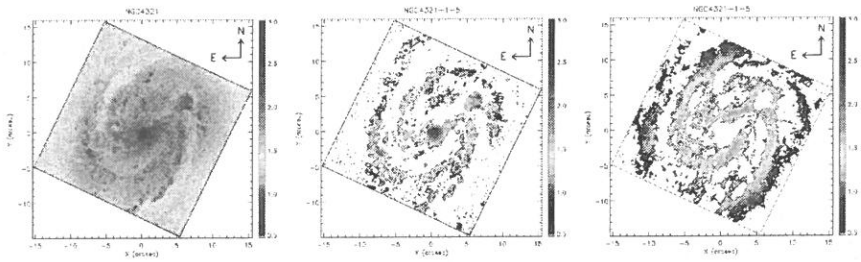


Fig. 6. The left image is the original observed data. The middle image is detected by wavelet plan 1-5 and shows that there is a small bar which drives two spiral arms into the center. The right image is also detected by the wavelet plane 1-5 but it detects the dust spiral pattern which is consistent with CO observation.

4. CONCLUSIONS AND DISCUSSIONS

We have shown that the wavelet technique provides a powerful means to probe the central structure of galaxies. The results of detections of certain galaxies, NGC 5383, NGC 1097, NGC 1068 and NGC 4321, show us that spiral pattern can be traced all the way to the nucleus and galaxies which could have a mini-bar in the central region.

The other wavelets, for instance, CDF, CW and DF have been also used to detect the galactic central region. Clump and fluctuation appear in the processed images. Therefore, we suggest that the atrous algorithm is the best for such purpose. Besides, the targets near the boundary of image require boundary wavelet approach. It should improve the detections.

More results on galaxies would be required before the construction of a model. Furthermore, numerical approach is also useful to understand the dynamics of galaxies. This approach could also help us understand the observations better.

REFERENCES

(1) Arsenault R., Boulesteix J., Georgelin Y., and Roy J.R., A&A, 200, 29, 1988

(2) Bacry E., Muzy J., and Arneodo A., *Singularity spectrum of fractal signals from wavelet analysis: Exact results,* J. Stat. Phys. 70, 635, 1993

(3) Cohen A., Daubechies I. and Feauveau J., Biorthogonal bases of compactly supported wavelets, Comm. Pure Appl. Math., 45,485, 1992.

(4) Chui C.K., *An Introduction of Wavelets, Academia Press*, San Diego, CA, 1992

(5) Chern I.-L. and Yen C.C., *Difference Wavelet—Theory and A Comparison Study*, Methods and Appl. of Analysis, 9, No. 4, 469, 2002

(6) Daubechies I., *Ten Lectures on Wavelets, CBMS-NSF Regional Conf. Series in Appl. Math.* Vol. 61, SIAM, Philadelphia, PA, 1992.

(7) Daubechies I., *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41, 909, 1988.

(8) Forbes D.A., Wada M.J., DePoy D.L., Boisson C.B., and Smith M.S., MNRAS, 254, 509, 1992.

(9) Malin D.F., *American Astronomical Society Photo-Bulletin*, 16, 10, 1977.

(10) Knapen J.H., and Beckman J.E. MNRAS 283, 251, 1996

(11) Knapen J.H., Beckman J.E., Heller C.H., Shlosman I., and de Jong R.S., ApJ, 454, 623, 1995a

(12) Knapen J.H., Beckman J.E., Shlosman I., Peletier R.F., Heller C.H., and de Jong R.S., ApJ, 443, L73, 1995b

(13) Kotilainen J.K., Renunanen J., Laine S. and Ryder S.D., A&A 353, 834, 2000

(14) Lin C.C. and Shu F.H,ApJ, 140, 646, 1964.

(15) Papoulis A., Probability, *Random Variables, and Stochastic Processes*, Mc Graw Hill (Ed.), 1991

(16) Pierce M.J., AJ, 92, 285, 1986

(17) Quillen A.C., Frogel J.A., Kuchinski L.E. and Terndrup D.M., AJ, vol. 110, 156,1995.

(18) Sakamoto K., Okumura S., Minexaki T., Kobayashi Y., and Wada K., AJ, 110 No. 5, 275, 1995

(19) Sheth K., Regan M.W., Vogel S.N. and Teuben P.J., ApJ, 532, 221, 2000

(20) Shlosman I., Frank J., and Begelman M.C., Nature, 338, 45, 1989

(21) Yuan C., and Kuo C.L., ApJ, 497, 689, 1998

小波應用於探測星系中心結構

嚴健彰
輔仁大學數學系

袁　旂
中央研究院天文所

李怡慧
台灣師範大學地科系

摘　要

本論文的目的是小波應用於探測星系中心結構。大部份靠近我們的星系，都發現它們中心有一個氣體盤，然而氣體盤的結構經常因爲背景的星光太亮而被遮掩。小波提供一個有效的方法可以從觀測的資料中，獲得那些被隱藏的結構。

Atrous 小波是一個簡單快速，且針對以上目的是非常有用的。我們分析從哈伯望遠鏡的 NICMOS 和 WFPC (WFPC2) 的星系資料，結果顯示某些星系的螺旋結構可延深到星系中心(例如 NGC 5383,NGC 4321)及中心可能有棒狀結構 (例如 NGC 1068, NGC 1097)。這些結果對瞭解星爆及活躍星系核是重要的。

關鍵詞：小波理論， atrous 小波，星系， 螺旋， 棒，環