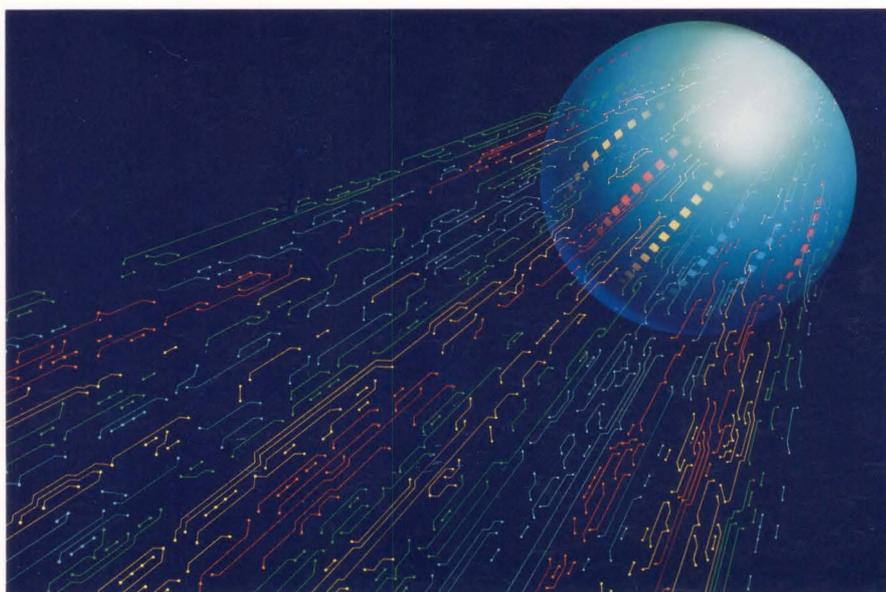
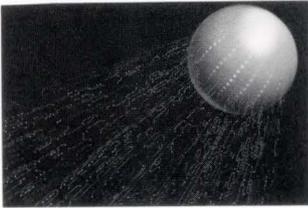


FUJITSU SCIENTIFIC & TECHNICAL JOURNAL

Summer 1988 VOL.24, NO.2





This Issue's Cover

Highly integrated LSIs are becoming more advanced, and the ULSI having 100 000 000 or more integrated devices per chip will soon be realized. Recently, attention has focused on the structure and movement of DNA molecules which store the genetic code (information) found in nature.

This concept would directly use the molecules as various devices based on the structure of the DNA molecules.

This issue's cover illustrates a view to express the wiring patterns and movement of the DNA molecules.

FUJITSU Scientific & Technical Journal is published quarterly by FUJITSU LIMITED of Japan to report the results of research conducted by FUJITSU LIMITED, FUJITSU LABORATORIES LTD., and their associated companies in communications, electronics and related fields. It is the publisher's intent that FSTJ serve to promote the international exchange of such information, and we encourage the distribution of FSTJ on an exchange basis. All correspondence concerning the exchange of periodicals should be addressed to the editor.

FSTJ can be purchased through KINOKUNIYA COMPANY LTD., 3-17-7 Shinjuku, Shinjuku-ku, Tokyo 160-91, (Telex No. 2424344, answerback KINOKUNI J).

The price is US \$5.00 per copy, excluding postage.

FUJITSU LIMITED reserves all rights concerning the republication and publication after translation into other languages of articles appearing herein.

Permission to publish these articles may be obtained by contacting the editor.

FUJITSU LIMITED

Takuma Yamamoto, *President*

FUJITSU LABORATORIES LTD.

Dr. Bun-ichi Oguchi, *President*

Editorial Board

Editor Hiroshi Yamada

Associate Editor Takahiko Misugi

Editorial Representatives

Koichi Dazai	Toshito Hara	Masao Hiyane
Noriaki Fujimura	Kazuhisa Kobayashi	Yoshihiro Kosaka
Ken-ichi Murata	Seiya Ogawa	Shinji Ohkawa
Eisaku Ohshima	Shigeru Sato	Tohru Sato
Shozo Taguchi	Hideo Takahashi	Hirobumi Takanashi
Mitsuhiko Toda	Ryoiku Toge	Toru Tsuda
Akira Yoshida		

Editorial Coordinator

Kazuo Yono

FUJITSU LIMITED 1015 Kamikodanaka, Nakahara-ku,
Kawasaki 211, Japan

Cable Address: FUJITSULIMITED KAWASAKI

Telephone: +81-44-777-1111

Telex: 3842122, 3842221, 3842802, 3842803,

Answerback FTWKAW J

Printed by Mizuno Art Printing Co., Ltd. In Japan

© 1988 FUJITSU LIMITED (June 15, 1988)

FUJITSU SCIENTIFIC & TECHNICAL JOURNAL

Summer 1988 VOL.24, NO.2

CONTENTS

Review

71 Devices Using High-Coupling Piezoelectric Crystals

●Hideki Tominaga ●Masaaki Ono ●Yoshiro Fujiwara

Papers

100 An Expert System Based on Qualitative Knowledge

●Hiroyuki Yoshida ●Hiroyuki Izumi ●Ryoichi Narita

109 A Design Method of Reflector Antennas with Arbitrary Configuration

●Yoshihiko Asano ●Hiroshi Kurihara

118 LISP-PAL: An Approach to Natural Language Consultation in a Programming Environment

●Sanya Uehara ●Rieko Nishioka ●Tomoya Ogawa

126 A System for Protein-Crystal Growth Experiment in a Microgravity Environment

●Shozo Fujita ●Koji Asano ●Takafumi Hashitani

133 Liquid Phase Epitaxial Growth of Buried Heterostructure DFB Lasers

●Toshihiro Kusunoki ●Nirou Okazaki ●Toshiyuki Tanahashi

143 DX-Center-Free GaAs/N-AlGaAs HEMT Structures

●Tomonori Ishikawa ●Kazuo Kondo

150 Fast Ray Tracing

●Koichi Murakami ●Katsuhiko Hirota ●Mitsuo Ishii

160 Simulation Processor: SP

●Fumiyasu Hirose ●Hirofumi Hamamura ●Mitsuo Ishii
●Hiroshi Yamada

166 Full-Color Image Printer

●Tomohisa Mikami ●Tsugio Noda ●Toshio Konaka

UDC 53.082.73:548.55:621.373.1
FUJITSU Sci. Tech. J., **24**, 2, pp. 71-99(1988)

Devices Using High-Coupling Piezoelectric Crystals

• Hideki Tominaga • Masaaki Ono • Yoshiro Fujiwara

This paper reviews the development and application piezoelectric devices using high-coupling piezoelectric crystals, such as LiTaO_3 and LiNbO_3 , and describes some products that incorporate these devices.

This paper describes in detail the technological development of a substrate in the form of a large wafer, the process development of device manufacturing, device design technology, mass-production technology, and device applications in relation to market trends. Fujitsu is researching and developing the materials, processes, and design to market piezoelectric devices that meet today's demands for large capacity, high speed, and high quality in information communications.

UDC 520.6.05:548.52
FUJITSU Sci. Tech. J., **24**, 2, pp. 126-132(1988)

A System for Protein-Crystal Growth Experiment in a Microgravity Environment

• Shozo Fujita • Koji Asano • Takafumi Hashitani

To evaluate the effect of microgravity on protein-crystal growth, a system performing experiments automatically in the Small Self-Contained Payload container offered by NASA has been developed. The system can simultaneously perform 16 independent experiments which apply three crystallization methods. Crystal growth is started by sliding out plates separating protein and ammonium sulfate solutions, and the state of the growth is periodically recorded by taking photographs for up to 120 hours. The temperature is passively controlled by a heat regenerator. Fujitsu can thereby evaluate the effectiveness of microgravity by comparing the protein-crystal growth in space with that on the ground.

UDC 001.81:681.3
FUJITSU Sci. Tech. J., **24**, 2, pp. 100-108(1988)

An Expert System Based on Qualitative Knowledge

• Hiroyuki Yoshida • Hiroyuki Izumi • Ryoichi Narita

This paper describes the implementation of an expert system that is based on deep knowledge. This system uses a qualitative model of the target object to predict the behavior of it through qualitative simulation, diagnose multiple faults in the object, and create a decision tree for diagnosing typical symptoms. Qualitative modeling has been extended for practical target objects in which discontinuous state transitions occur, and the system uses a new simulation method based on extended models. Modal operators are used to represent time-varying symptoms for diagnosing multiple faults. The system also uses an algorithm to create a decision tree from a deep knowledge base.

UDC 548.524:621.375.826
FUJITSU Sci. Tech. J., **24**, 2, pp. 133-142(1988)

Liquid Phase Epitaxial Growth of Buried Heterostructure DFB Lasers

• Toshihiro Kusunoki • Nirou Okazaki • Toshiyuki Tanahashi

Flat-surface buried-heterostructure distributed-feedback laser diodes were fabricated by liquid-phase epitaxy. There were three difficulties to be solved for improving the characteristics of the laser. The first difficulty was the formation of pits in the InGaAsP optical guide layer on the corrugations. The other difficulties were a thermal deformation of the corrugations and leakage current flowing along the sides of the mesa. The first difficulty was solved by using an oxygen plasma process for cleaning the corrugations. The thermal deformation of the corrugations was reduced by covering the corrugations with a GaAs plate that had a rough surface. The leakage current was reduced by using new etchant for etching the mesa.

UDC 519.6:621.396.67
FUJITSU Sci. Tech. J., **24**, 2, pp. 109-117(1988)

A Design Method of Reflector Antennas with Arbitrary Configuration

• Yoshihiko Asano • Hiroshi Kurihara

A computer program is developed for designing reflector antennas with arbitrary configurations for use in arbitrary alignments between primary feeds and reflectors of arbitrary shapes.

Recent progress in satellite communications has increased the importance of efficient use of the geostationary orbit together with importance of efficient use of the radio spectrum. Highly efficient use of the geostationary orbit is ensured by giving an accurate design of a low sidelobe antenna which requires rigorous estimation of antenna radiation characteristics. Since newly developed computer program calculates radiation characteristics for arbitrary antenna configurations, the best configuration to minimize the sidelobe can be chosen from the variety of antenna configurations. The estimated radiation characteristics have been experimentally confirmed.

UDC 621.315.592:621.382.3
FUJITSU Sci. Tech. J., **24**, 2, pp. 143-149(1988)

DX-Center-Free GaAs/N-AlGaAs HEMT Structures

• Tomonori Ishikawa • Kazuo Kondo

The DX centers in $\text{N-Al}_x\text{Ga}_{1-x}\text{As}$ layers grown by MBE were investigated. The concentration ratio of DX centers to total donors was determined for various compositions of AlGaAs and doping concentrations of Si. Based on these results, DX-center-free selectively doped GaAs/N-AlGaAs heterostructures were proposed and successfully applied to HEMTs.

UDC 519.6:681.326
FUJITSU Sci. Tech. J., **24**, 2, pp. 118-125(1988)

LISP-PAL: An Approach to Natural Language Consultation in a Programming Environment

• Sanya Uehara • Rieko Nishioka • Tomoya Ogawa

The LISP-PAL approach is designed to assist programmers by supplying them with experienced expert programmer knowledge in response to their questions. For example, when a user asks in natural Japanese what functions or programming schemes would be useful for a particular programming purpose, the system displays appropriate usage guides and programming examples.

The purpose of this research is to build a knowledge-based software development environment and to develop technologies for utilizing knowledge bases. The main contributions of this work are the development of a consultation system that stores and retrieves programming knowledge in text form and a frame-based knowledge representation system. A technique to handle frame-based representation through pattern matching in an object-oriented programming language has also been developed.

UDC 681.32:681.332.3:76
FUJITSU Sci. Tech. J., **24**, 2, pp. 150-159(1988)

Fast Ray Tracing

• Koichi Murakami • Katsuhiko Hirota • Mitsuo Ishii

Ray tracing produces images of excellent quality, but it requires lengthy computations that limit its practical use. This paper discusses two approaches to shortening ray tracing computation time. The first approach involves partitioning the environment with voxels to reduce the number of ray-object intersections. Proposed here is a new traversal algorithm that efficiently traverses the voxel data structure. Experimental data demonstrate that the resulting traversal overhead is small. The second approach involves implementing a ray tracing algorithm on a CAP (Cellular Array Processor) parallel processor. The parallelization of ray tracing takes advantage of the computational independency of individual rays. The results demonstrate that, when using the proposed load distribution scheme, performance increases in proportion to the number of cells used.

UDC 681.325.65

FUJITSU Sci. Tech. J., 24, 2, pp. 160-165(1988)

Simulation Processor: SP

- Fumiyasu Hirose • Hirofumi Hamamura • Mitsuo Ishii
- Hiroshi Yamada

A high-speed, large-capacity logic simulation engine, the simulation processor SP, has been developed for efficient logic verification of Fujitsu's large digital system. The SP implements an event-driven simulation algorithm in a multi-processor configuration. It utilizes a multi-stage switching network with an attached buffer-memory so that inter-processor communications do not cause a bottleneck. The SP executes 4-input-1-output logic primitive and 1-bit-per-word memory primitive level simulation with an accuracy of 16 signal values and unit/zero delay simulation. It can simulate 4 mega primitives and 32 Mbytes of memory as its maximum speed of 0.8 giga active primitive evaluations per second.

UDC 681.327.5:681.621.7

FUJITSU Sci. Tech. J., 24, 2, pp. 166-176(1988)

Full-Color Image Printer

- Tomohisa Mikami • Tsugio Noda • Toshio Konaka

This paper describes the print data control circuit developed for a new full-color thermal ink-transfer printer that uses a unique thermofusible ink sheet with gray scale reproducibility. This printer can print twice as fast using this sheet as when using a conventional sublimable ink sheet, and offers high image quality by incorporating the precision color conversion circuit and temperature drift compensation circuit.

Precise color reproduction was enabled by interpolating conversion tables instead of calculating color-conversion equations. Moreover, significant print density variations due to medium-speed temperature drift were reduced to about one quarter by calculating the temperature drift line-by-line.

Devices Using High-Coupling Piezoelectric Crystals

• Hideki Tominaga • Masaaki Ono • Yoshiro Fujiwara

(Manuscript received June 15, 1988)

This paper reviews the development and application piezoelectric devices using high-coupling piezoelectric crystals, such as LiTaO_3 and LiNbO_3 , and describes some products that incorporate these devices.

This paper describes in detail the technological development of a substrate in the form of a large wafer, the process development of device manufacturing, device design technology, mass-production technology, and device applications in relation to market trends. Fujitsu is researching and developing the materials, processes, and design to market piezoelectric devices that meet today's demands for large capacity, high speed, and high quality in information communications.

1. Introduction

Piezoelectric single crystals, such as LiTaO_3 (LT) and LiNbO_3 (LN), are called high-coupling piezoelectric crystals because they have a high electromechanical coupling coefficient.

Research into the physical properties of LT and LN has been carried out since Matthias and Remeika¹⁾ discovered the ferroelectric characteristics of these materials at AT&T Bell Laboratory in the U.S.A. in 1949. The technology to grow single crystals of LT and LN was developed using the Czochralski (CZ) method developed by Ballman²⁾ at AT&T Bell Laboratory in 1965. In 1974, Brandle and Miller³⁾, also at AT&T Bell, reported a way to grow 40-mm diameter \times 180 mm single crystal of LT. Valentino and Brandle⁴⁾ followed, developing a simple automatic growth technology for oxide single crystals. These developments were the beginning. The enlargement of piezoelectric single crystals and computer-controlled automatic growth has developed further⁵⁾, and research, development, and marketing has rushed to find applications for devices using this material.

Table 1 shows the electromechanical

coupling coefficients of piezoelectric crystals. The high coupling factors of LT and LN indicate their great potential for efficient piezoelectric devices. Figure 1 shows the LT and LN wafers mass-produced by Fujitsu. These wafers, up to four inches in diameter, are used as piezoelectric devices in many applications. The major use of the high-coupling piezoelectric crystals is as the SAW filter for color television IF circuit, shown in Fig. 2. Figure 3 shows the patterning of a SAW filter for television on a four-inch-diameter LN wafer. They are manufactured in batches of approximately 190 chips.

In the second half of the 1970s, major large-scale integration in color television circuits was planned. The IF circuit was the main hindrance in the quest to make the circuits more compact and eliminate the need for tuning. As shown in Fig. 4, the filter for the IF circuit was conventionally a combination of coils, capacitors, and resistors. It had to be tuned by a skilled worker. The SAW filter did not require tuning and it was compact, which meant that the demand for LT and LN wafers rose rapidly. Some of the substrate materials used in the SAW filter for the television IF

Table 1. Electrical characteristics of piezoelectric single crystals

Crystal material	Direction of crystal	Electromech. coupling factor ($K^2\%$)	Dielectric const. () ^{*1}	Temp. characteristics (ppm/°C)	Application to devices
Quartz	42.75° Y (ST) AT-cut	0.14 0.79	4-5	To 0	SAW ^{*2} BAW ^{*3}
AlPO ₄ ⁶⁾ (AP)	170 X 28° X	0.49 2.25		To 0	SAW BAW
Li ₂ B ₄ O ₇ ^{7), 8)} (LB)	X 49.23° Y	1.0 6.7	82-85	To 0 -25	SAW BAW
LiTaO ₃ (LT)	X-cut 163° X	0.7 16.6	41-53	25	SAW BAW
LiNbO ₃ (LN)	131° Y 163° X	5.5 37.8	29-84	-72.2	SAW BAW
Pb ₂ KNb ₅ O ₁₅ ⁹⁾ (PKN)	X	1.98	81-810	To 0	SAW
Bi ₁₂ GeO ₂₀ ¹⁰⁾ (BGO)	(110)	1.4	-	-100	SAW
Bi ₁₂ SiO ₂₀ ¹¹⁾ (BSO)	X	1.4	56	-118	SAW

*1: Depend on direction of crystal

*2: SAW: Surface Acoustic Wave

*3: BAW: Bulk Acoustic Wave

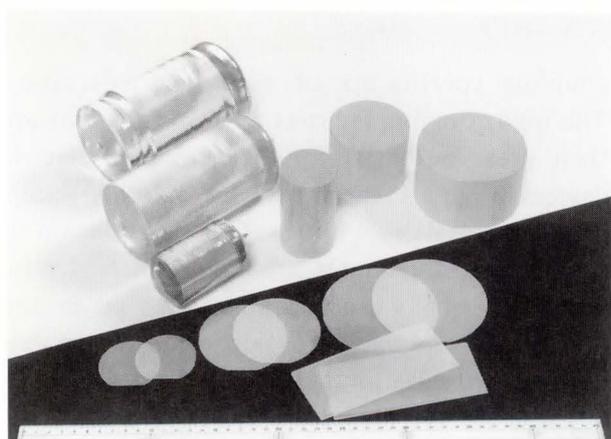


Fig. 1—LiTaO₃ and LiNbO₃ single crystals grown by automatic crystal growth (as grown crystals and cut boules with flat orientations, and wafers).

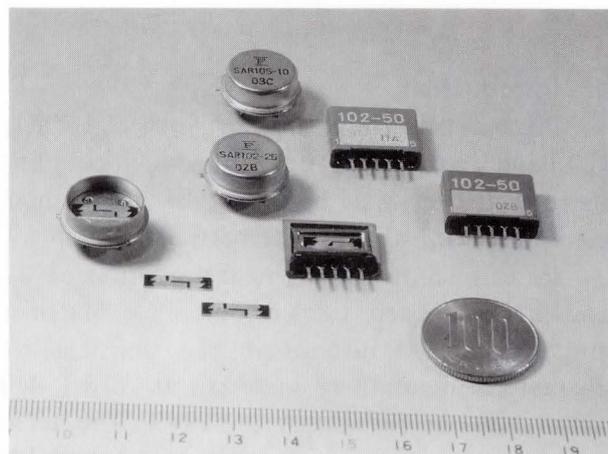


Fig. 2—Samples of a SAW filter for color televisions.

circuit are quartz, piezoelectric ceramics, ZnO thin film, AlPO₄ (AP)⁶⁾, Li₂B₄O₇ (LB)^{7), 8)}, LT, LN, Pb₂KNb₅O₁₅ (PKN)⁹⁾, Bi₁₂GeO₂₀ (BGO)¹⁰⁾, and Bi₁₂SiO₂₀ (BSO)¹¹⁾. LT and LN single crystal substrates are most frequently used today because they make it easy to obtain large substrates, and to manufacture substrates. They also enable device manufacturing reproducibility, and long-term reliability. The demand to reduce the cost of devices has led

to bigger single crystal wafers being made each year¹²⁾, as shown in Fig. 5. At present, the device manufacturing process is the main factor in determining the wafer size. In addition to SAW devices, there is great promise in piezoelectric devices such as very small resonators and filters, which use the high-coupling piezoelectric crystal characteristics. Research, development^{13), 14)}, and marketing is being conducted in these areas.

The main application of LT and LN wafers

is in television SAW filters, and the second main application is in very small size resonators¹⁵⁾ developed and marketed by Fujitsu.

As Table 1 shows, the relative dielectric constant of LT and LN is much smaller than that of piezoelectric ceramics. Resonator elements are small, and the parallel capacitance of resonators (C_0) is only a few pF. As a result, resonators can be designed to have a low capacitance ratio (γ) ($= C_0/C_s$) of parallel capacitance with equivalent series capacitance (C_s). Voltage controlled oscillators (VCOs) constructed using these resonators can obtain a frequency variable range (0.4 percent) about 20 times that of VCOs using quartz resonators.

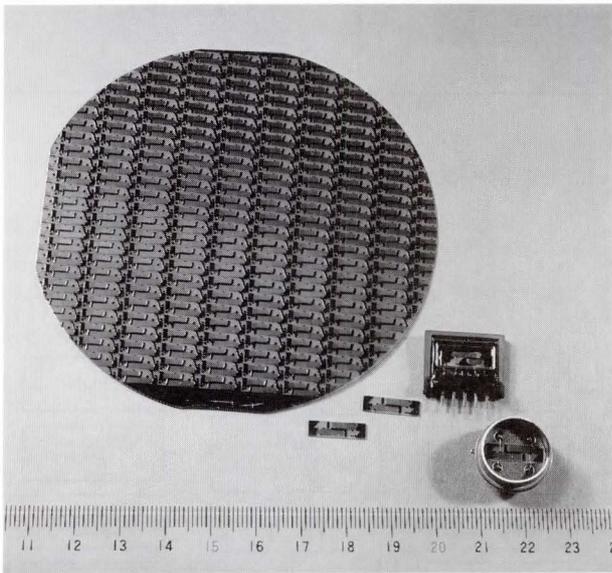


Fig. 3—Patterned four-inch LN wafer for SAW filters.

Details will be given in the main body of this paper. These VCOs are optimum for synchronization between digital equipment because of their high stability and wide variable range.

Filters using these resonators can be designed with a wide band and high stability. These filters give optimum performance for regeneration of signals in the digital equipment. Figure 6¹⁶⁾ compares the appearance of these filters with traditional LC filters using coils and capacitors. Figure 7 compares the sizes and shows the characteristics of LT filters. Resonators using high-coupling piezoelectric crystals can be designed to have compact resonator elements. For example, the typical element size of a 4-MHz resonator is

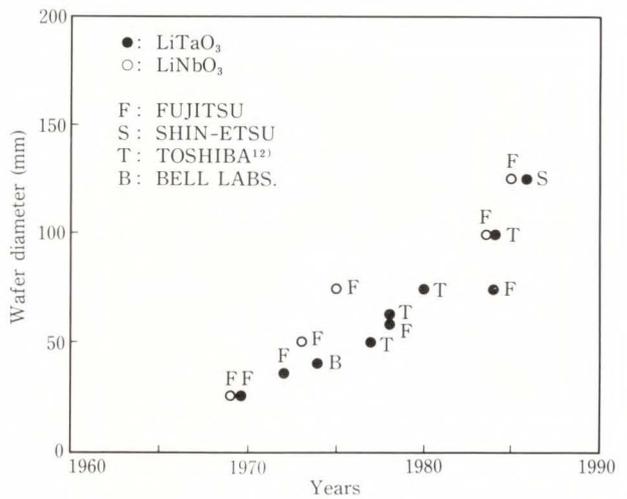
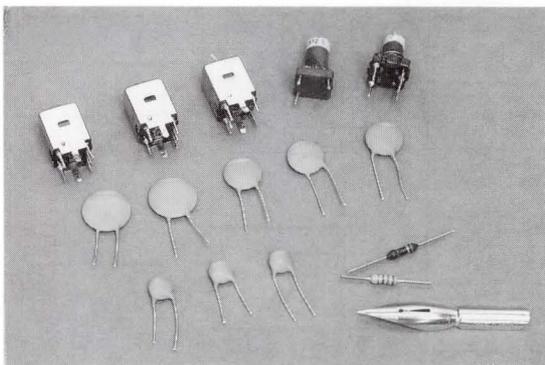
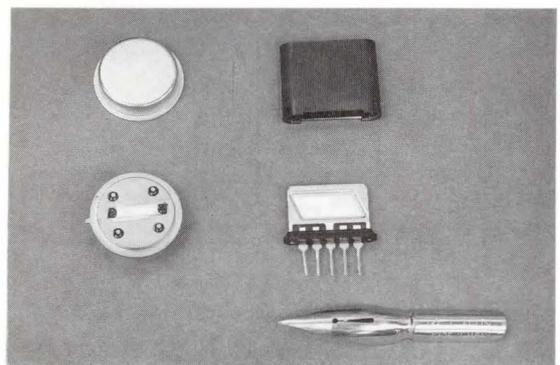
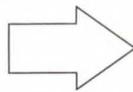


Fig. 5—Trend in maximum LT and LN wafer size.



a) Components of LC filters for color televisions



b) SAW filter for color televisions

Fig. 4—Comparison of LC filter and SAW filter.

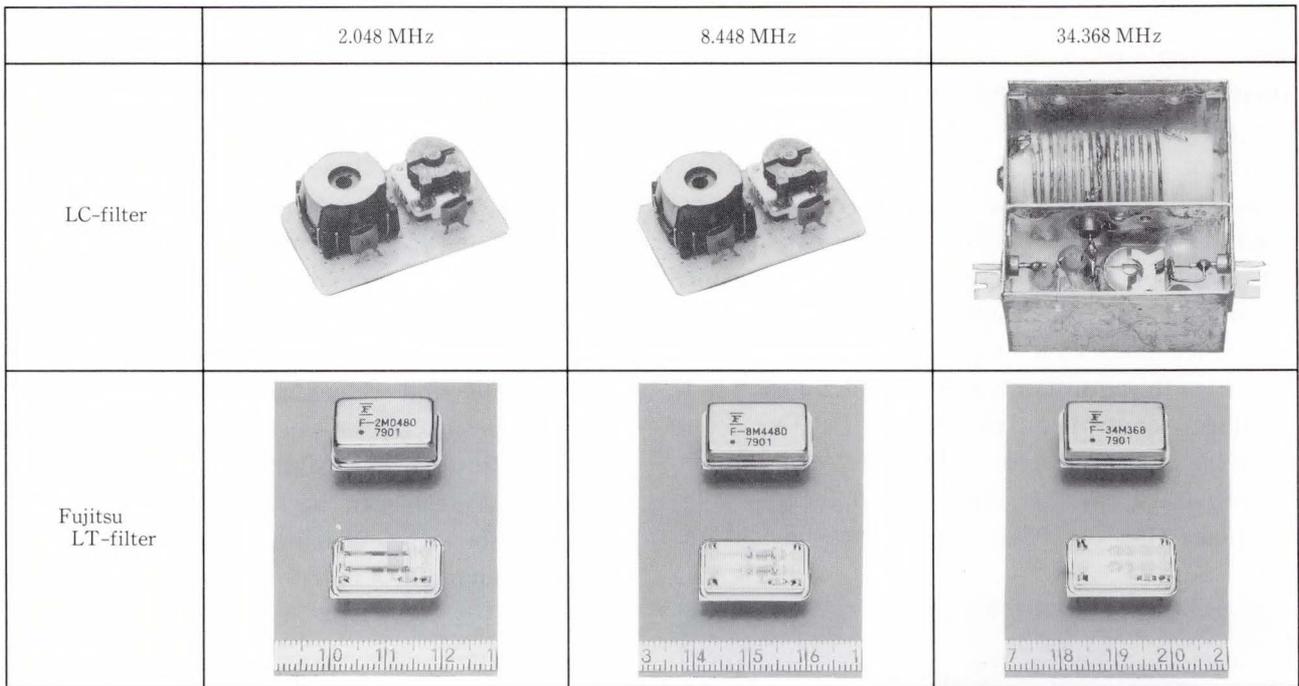


Fig. 6—Comparison of LC filter and LT filter dimensions.

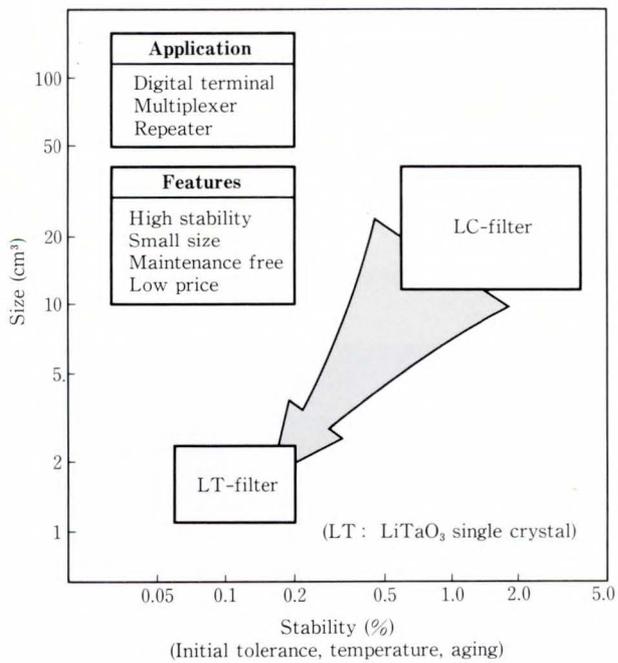
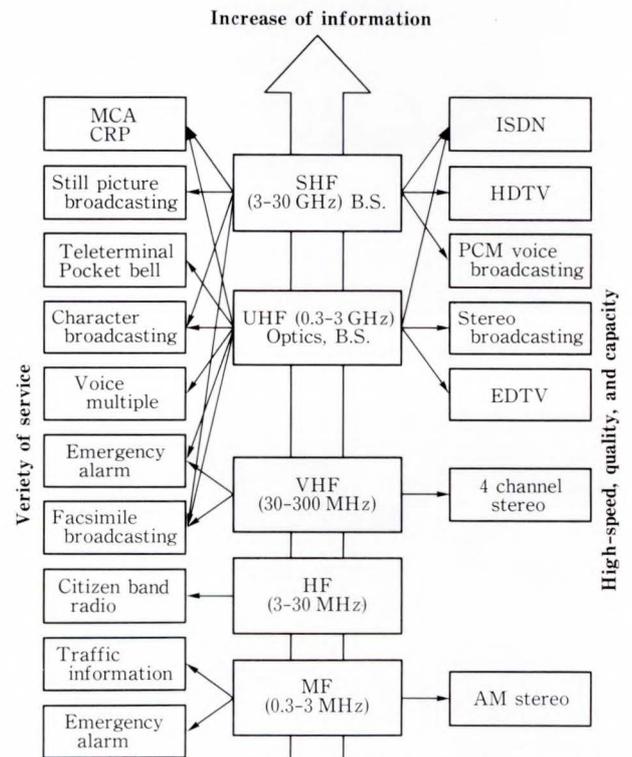


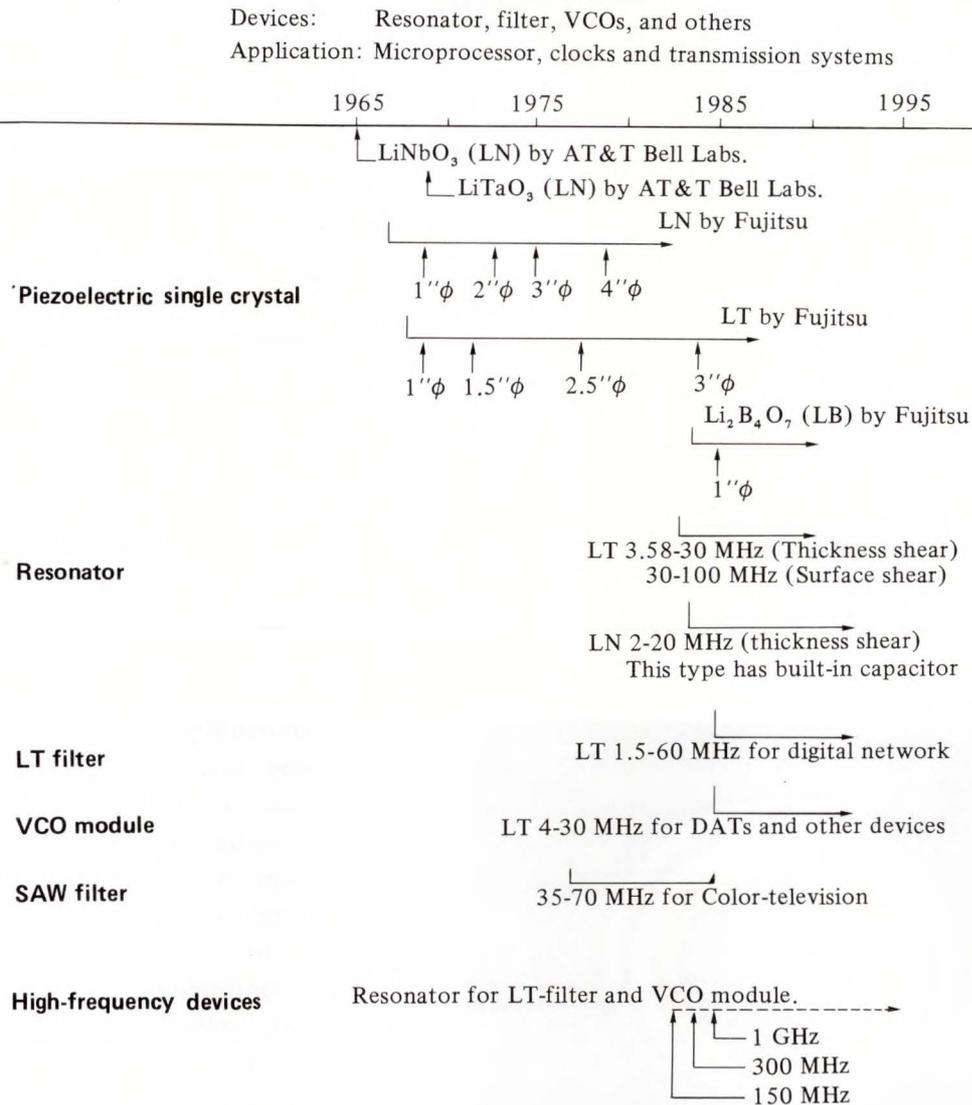
Fig. 7—LT filter features.



MCA : Multiple Channel Access CRP : Combinience Radiophone
 HDTV : High Definition TV EDTV : Extended Definition TV

Fig. 8—Trend of high-frequency services and technology.

Table 2. Development of high-coupling piezoelectric devices in Fujitsu



0.6 mm × 0.5 mm × 7 mm, and that of a 12-MHz resonator is 0.3 mm × 0.15 mm × 4 mm.

The thermal expansion coefficient in the longitudinal direction of the resonator element has been matched with that of the ceramic substrate for packaging. This enables the resonator to be stuck directly onto the ceramic substrate, using conductive adhesives. This packaging has achieved a resonator with better resistance to heat, cold, sudden temperature changes, and mechanical shocks. It is so robust that it can be used in cars, even in the engine compartment. This paper gives details of typical piezoelectric devices, with the emphasis on resonators, filters, and VCOs. Table 2 shows the history of Fujitsu's research and development of

high-coupling piezoelectric devices.

In the near future, data processing and transmission will be done better and faster by more powerful electrical equipment. This equipment will be set up in wider areas and used as one's personal effects. The electronic devices within electrical equipment must be compact and light; they must operate at high frequencies, and use SMT.

must operate at high frequencies, and use SMT.

Signal processing is already changing from analog to digital. The demand is increasing for new, compact devices in ISDN equipment; devices such as wide-bandwidth, high-stability filters, and wide-variable range VCOs for timing tank circuits.

Table 3. Growth method of piezoelectric single crystals

Crystal material	Growth method	Growth condition	Material of crucible	Max X'tal size (mm)	Growth period of largest X'tal
Quartz	Hydrothermal	200-400 °C (1 000-2 000 atm)	Ag	50 × 60 × 190	40-60 days
AlPO ₄ (AP)		To 200 °C (10 atm)	Pt	—	—
Li ₂ B ₄ O ₇ (LB)	Czochralski (CZ)	920 °C (Air-atom)	Pt	50ϕ × 120	6-8 days
LiTaO ₃ (LT)		1 660 °C (N ₂ -atm)	Ir or PtRh	125ϕ × 150	3-5 days
LiNbO ₃ (LN)		1 260 °C (Air-atom)	Pt	100ϕ × 150	
Pb ₂ KNb ₅ O ₁₅ (PKN)		1 310 °C (Air-atom)		10 × 10 × 10	1 day
Bi ₁₂ GeO ₂₀ (BGO)		— (Air-atom)		75ϕ × 150	4-6 days
Bi ₁₂ SiO ₂₀ (BSO)	— (Air-atom)				

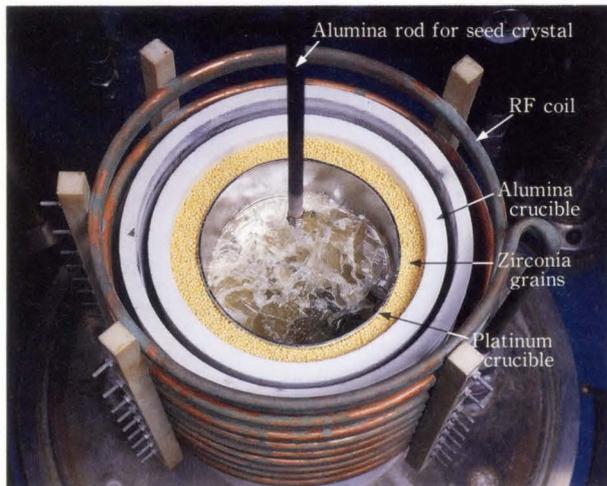


Fig. 9—Internal view of the furnace used for crystal growth.

Figure 8 shows how higher and higher frequencies are being used: the 400-MHz to 900-MHz band for car telephones and teleterminals, and the 400-MHz to 2.4-GHz band for ISDN trunk lines. There is an ever-increasing demand for compact, lightweight, high-frequency key devices, such as oscillators, filters, and VCOs.

This paper reports on these piezoelectric devices.

2. High-coupling piezoelectric crystals

2.1 Single crystal growth technique

Quartz is a typical piezoelectric crystal. Devices using quartz and single crystal growth technology of quartz had histories of research and development for about 50 years, but research and development into new applications is still continuing. An example is the GT-cut resonator, which has a better temperature characteristic than the AT-cut resonator. Another example is tuning fork type very small resonator used in electronic watches. In the manufacturing process of quartz resonator, new technology such as photolithography is applied. Many elements can now be manufactured at one time by photolithographic batch processing.

High-coupling devices using LT and LN are still only ten years old, but further research and development will move quickly. Table 3 shows the types of piezoelectric single crystals and growth methods. LT and LN crystals up to five inches in diameter can be grown by the Czochralski (CZ) method. Three- and four-inch crystals are used in device mass production. Figure 9 shows the internal view of a CZ furnace. Figure 10 shows its internal structure.

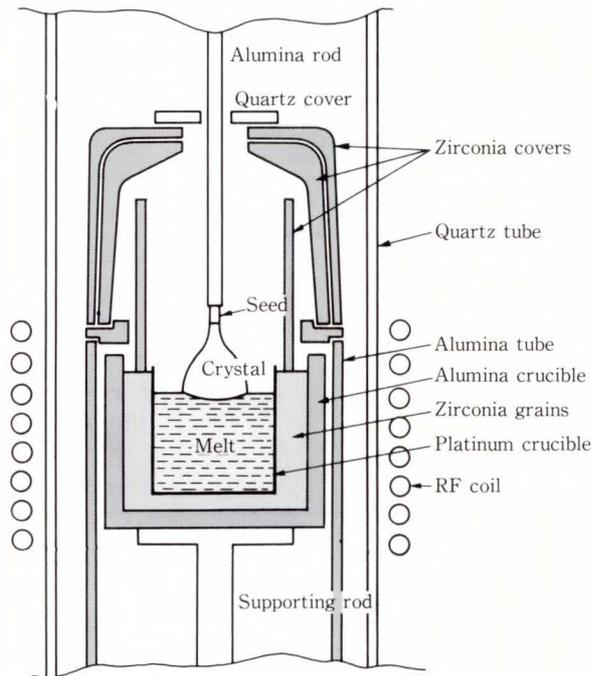


Fig. 10—Furnace structure.

LT and LN are oxide single crystals. When they are melted at high temperatures, they react very easily with other metal oxides. For this reason, single crystal growth crucibles must be matched with the growth atmosphere shown in Table 3. These crucibles are usually made of precious metal with a high melting point, such as Ir, PtRh, or Pt. The crucible is heated by high frequency induction through the working coil. The frequency depends on the oscillation circuit: 9-kHz, 200-kHz, and 400-kHz frequencies are often used. The material composition is usually congruent melt¹⁷⁾ so that it does not change in the crucible as the single crystal is pulled out.

Single crystals are grown by homoepitaxial growth on seed crystals tied on an alumina seed shaft set at the center of the top of the crucible, under the conditions shown in Table 3. When the seeds crystal is inserted into the melt, at the melt temperature of which the seed size becomes smaller, the crystal growth is started. The temperature is gradually reduced, and controlled so that the diameter of the crystals reaches the required size.

The growth rate of single crystals depends on the type of material, the direction of growth, and the structure of the heat shield material of the furnace, but is about 1-10 mm/h. This is about one-tenth the growth rate of semiconductor crystals such as Si and Ge, and is one reason why the cost of piezoelectric crystals is so high.

Piezoelectric crystals are not very resistant to thermal stress, and current technology limits the maximum growth length to about 200 mm. For practical reasons, the growth length is often limited to about 150 mm.

After growth, single crystals must be gradually cooled from the growth temperature to room temperature for about 20 h to 30 h. It takes about 80 h from heat up to cooling, so the growth cycle is in units of one week. Because growth requires 60 h to 120 h of continuous operation, it is usually controlled by computer.

Figure 11 is a block diagram of automatically controlled crystal growth. Figure 12 shows Fujitsu's mass-production growth furnaces. Figure 13 is an example of an LN bowl of four inches in diameter that has been pulled out with 128 °Y axis.

2.2 Wafer finishing technology

When the LT and LN crystals have grown, they have a single crystal structure, but they cannot function as dielectric or piezoelectric cells yet because they are in the multi-polarization state. For use in piezoelectric devices, they must be in the single polarization state.

To change their polarization, the temperature of the crystals is raised above the Curie point, and DC voltage is applied in the direction of self polarization to generate single polarization.

LN has a Curie point of about $1130 \text{ }^\circ\text{C} \pm 70 \text{ }^\circ\text{C}$, depending on the composition. Therefore, the poling furnace is heated to about $1200 \text{ }^\circ\text{C}$. Crystals are inserted between two pairs of LN ceramics on platinum sheets, and voltage is applied. To discharge the charge caused by pyroelectric effects, single polarized

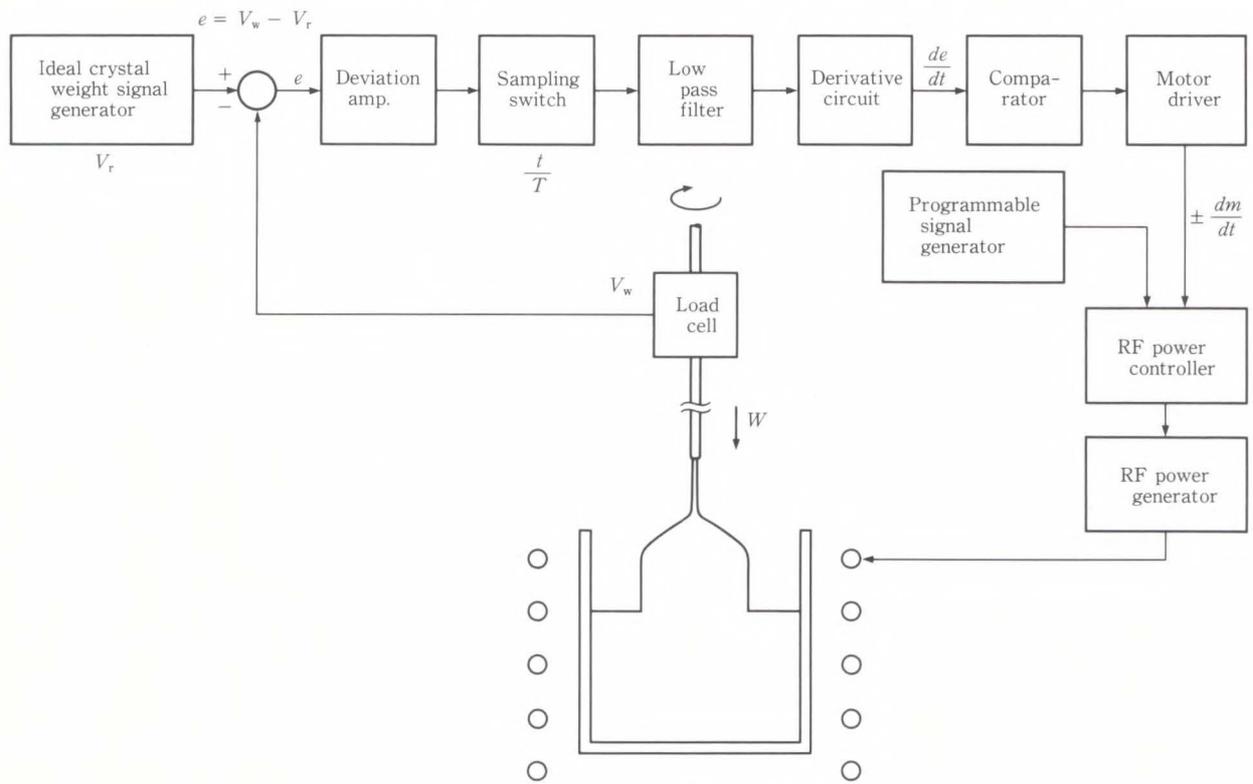


Fig. 11—Block diagram of automatically controlled crystal growth.



Fig. 12—Crystal growth stations at Fujitsu.



Fig. 13—As grown four-inch LN crystal under production at Fujitsu.

crystals are linked to a power supply while being cooled to room temperature.

LT has a Curie point of about $618 \text{ }^\circ\text{C} \pm 20 \text{ }^\circ\text{C}$, depending on its composition. Therefore the poling furnace is heated to $640 \text{ }^\circ\text{C}$. Voltage is applied to the electrode through

the platinum sheet, using PtPd or AgPd paste. The piezoelectric wafer finishing process is continuous cycles of increasing and decreasing both the temperature and the mechanical stress.

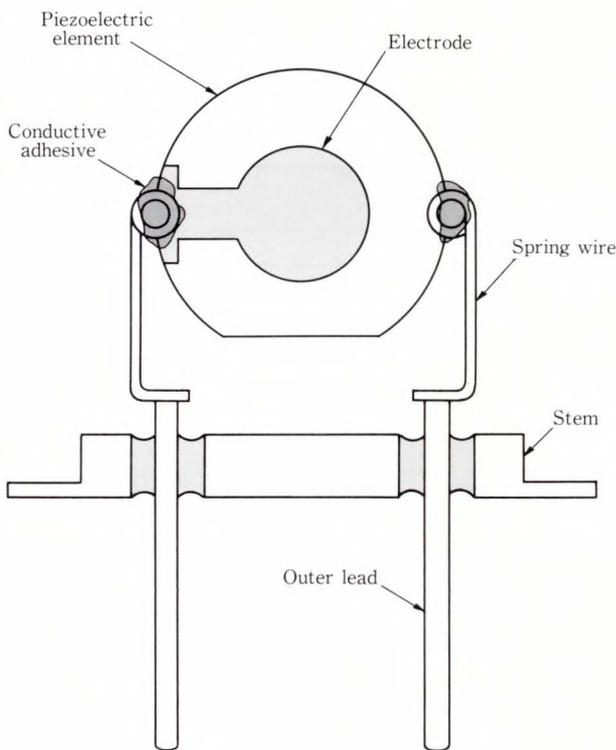


Fig. 14—Method of supporting conventional piezoelectric elements.

This must be done to discharge the charge generated by each process.

Piezoelectric crystals are easily cracked by a slight sonic shock wave caused by discharge. Single crystals are grown in the crystal direction that is best for the intended purpose. However, the crystal cut direction must be determined accurately by the X-ray raue method or by X-ray cut surface inspection so that it meets the specifications. Normally, the direction is controlled to within $\pm 0.1^\circ$ to $\pm 0.5^\circ$, according to the purpose. The crystals are ground to cylinders for easy wafer patterning, a post-process. Cylindrical crystals are attached to the orientation or sub-orientation flat cut surface in the specified direction so that the direction in the surface can be identified in the wafer state.

Crystals are then sliced into wafers with a multi-wire saw, and are rough-lapped and fine-lapped with SiC abrasive to the specified thickness.

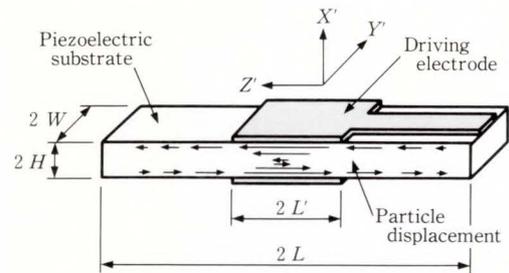


Fig. 15—Structure of piezoelectric strip.

The use frequency and plate thickness are in inverse proportion in wafers for thickness shear resonators. Because of the difficulties of lapping, there are several LN wafer sizes: 75-mm diameter for 4-6 MHz, 50-mm diameter for 6-9 MHz, 30-mm diameter for 9-13 MHz, 18-mm diameter for 14-20 MHz, and 15-mm diameter for 20-30 MHz.

One side of wafers for surface wave devices is polished by the same process as semiconductor silicon wafers. The HF and HNO₃ mixed acid boiling etching¹⁸⁾ process is used to remove mechanical stress under the wafer surface and to trim the wafer thickness finely. When piezoelectric wafers are used in thickness shear resonators, any variation in thickness directly causes variation in the device frequency. For this reason, wafers are finished with a thickness precision of about ± 0.15 percent by repeated etching.

In LN wafers with a 163° rotation Y-cut for thickness shear resonators, the absolute value of the thickness is $480 \mu\text{m} \pm 0.6 \mu\text{m}$ at 4 MHz and $160 \mu\text{m} \pm 0.27 \mu\text{m}$ at 12 MHz.

3. High-coupling piezoelectric resonator

3.1 Strip-type resonator

3.1.1 Resonator design

Resonators applying to a few MHz to several tens of MHz use the thickness wave where the elastic wave is propagated in the normal direction of the piezoelectric plate. The resonant frequency is determined by thickness of the piezoelectric plate.

As Fig. 14 shows, the form of the resonator elements in conventional resonators had to be

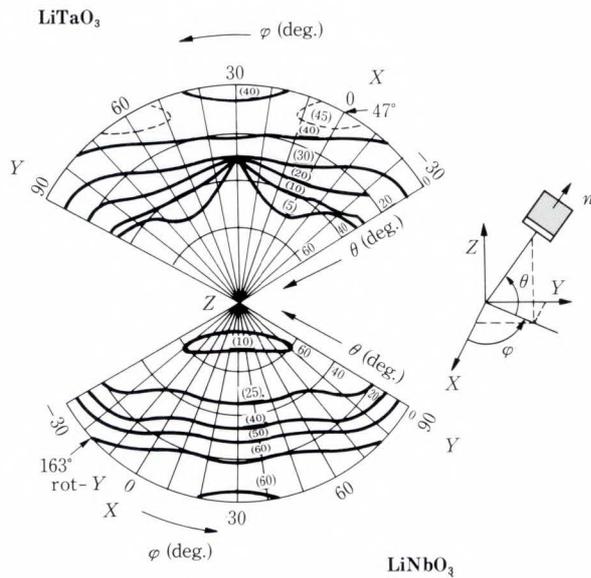


Fig. 16—Distributions of coupling coefficient K^2 (%) of LiTaO_3 and LiNbO_3 .

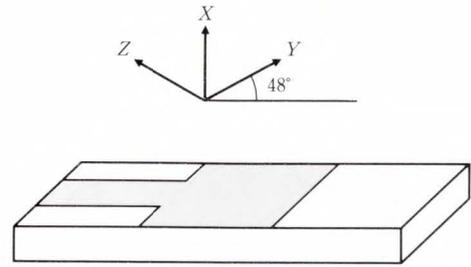
large enough to avoid resonance wave dumping and reflection from supporting region of resonator element. But now, other electronic devices have become more compact and are surface mounted.

We developed a strip-type resonator element that will realize very small, inexpensive resonators for surface mounting. The resonator element in Fig. 15 has been designed so that the longitudinal direction of the element parallels to the direction of particle displacement of the thickness shear wave. The strip-type resonator shows a single mode resonance, and the element size is less than one-third of the size of conventional elements^{15),19)}. The following section describes the design of the strip-type resonator.

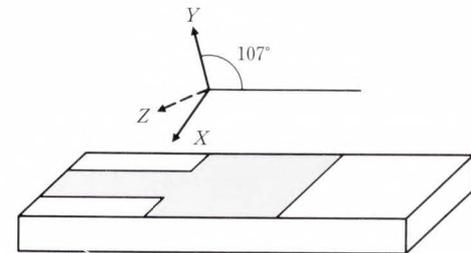
1) Crystal orientation of the resonator element

The temperature coefficient and electro-mechanical coupling coefficient (or simply, coupling factor) are important factors for piezoelectric substrate in resonators. These factors depend on the kind of crystal material and crystal orientation. We studied optimum crystal orientation of the LN and LT to realize following four requirements^{13),20)-22)}.

- i) Direction that has a large coupling factor



a) X-cut LT resonator element



b) 163° rot-Y LN resonator element

Fig. 17—LT and LN high-coupling piezoelectric elements.

for waves that are used and a small coupling factor for waves that are not required.

- ii) Direction in which the wave velocity that are not required is at least 20 percent more or less than the wave velocity that are used.
- iii) Direction in which items i) and ii) are satisfied, and the temperature coefficient is small.
- iv) Direction in which items i) to iii) are satisfied, and the thermal expansion coefficient in the longitudinal direction of the element is close to that of alumina ceramics.

Figure 16 shows the coupling factor (K^2) for thickness shear fast mode of LT and LN crystals. Only one-third of all directions is shown because LT and LN crystals belong to the trigonal crystal group. The coupling factor of LT reaches maximum value of 47 percent near X-cut. The coupling factors of thickness expansion mode and thickness shear slow mode which are unneeded modes are 0.6 percent or less. The coupling factor of LN reaches

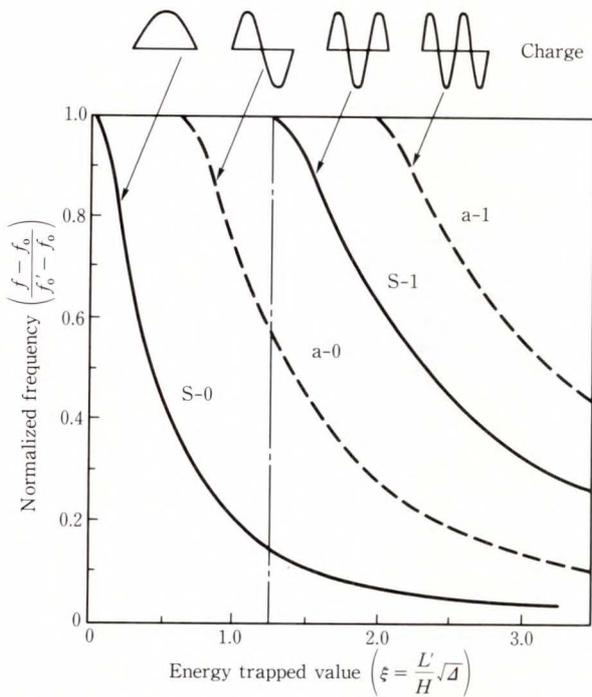


Fig. 18—Normalized frequency for harmonic and anharmonic modes of LT X-cut plate against trapped energy value.

62 percent near the 163° Y-cut, and the coupling factor for unneeded modes is zero.

Figure 17 shows the strip direction of the resonator element for the above crystal orientations. If the orientation is not the specified design value, unneeded spurious responses will be generated by mode conversion due to reflection in the side edge of resonator element.

2) Trapped energy

The trapped energy effect is where resonance energy is concentrated around the electrode because of the mass loading effect of the electrode and plate back.

Single-mode resonators that are not affected by the piezoelectric plate edges can be made by using the trapped energy effect. The trapped energy can be controlled by adjusting the electrode size and mass, and the resonant frequency can be finely adjusted.

The resonance spectrum produced by the trapped energy value can be expressed by the wave equation²³⁾ below. The frequency change (Δ) when the electrode is attached to the piezoelectric resonator surface is expressed

by the plate back frequency (Δ') and the mass loading (Δ''), as shown by Equation (1):

$$\Delta = \Delta' + \Delta'', \dots\dots\dots (1)$$

where

$$\Delta' = (f'_0 - f_0)/f'_0, \Delta'' = 2\rho'H'/\rho H$$

- f_0 : Cut-off frequency of electrode region
- f'_0 : Cut-off frequency of non-electrode region
- ρ : Piezoelectric material density
- ρ' : Electrode material density
- $2H$: Piezoelectric material thickness
- H' : Electrode material thickness.

Δ'' is 0.2 percent (Δ' is about 10 percent) when 200 nm of gold is attached as the electrode in a 6-MHz resonator using the X-cut LT. For 163° rotated Y-cut LN, Δ'' is 0.6 percent (Δ' is about 19 percent). So only Δ' need be considered when appraising the general trend. Equations (2) and (3) are the wave equations for trapped energy resonators. Figure 18 shows that the overtone in symmetrical mode must be the trapped energy value that does not become excited.

$$(K_0/K'_0) \tan(K_0 L') = 1 \text{ (symmetrical mode), } \dots\dots\dots (2)$$

$$(K_0/K'_0) \cot(K_0 L') = 1 \text{ (asymmetrical mode), } \dots\dots\dots (3)$$

where

$$K_0 = H/(2\pi) \{ (f/f_0)^2 - 1 \}^{1/2},$$

$$K'_0 = H/(2\pi) \{ 1 - (f/f_0)^2 \}^{1/2},$$

L' : Electrode length.

The dimension realized the single mode resonator is $L' < 4H$ in LT and $L' < 3.2H$ in LN. The frequency temperature characteristic changes according to the kind of crystal material and trapped energy value (ξ)²⁴⁾.

Figure 19 shows the temperature characteristic of the resonant frequency for the X-cut LT resonator when the trapped energy value is changed. When the trapped energy value (ξ) is infinite (infinite electrode), the turnaround temperature is about -40°C. Together with the reduction in ξ , it shifts to the higher temperature. When ξ is 0.8, the turnaround temperature is at room temperature and the first

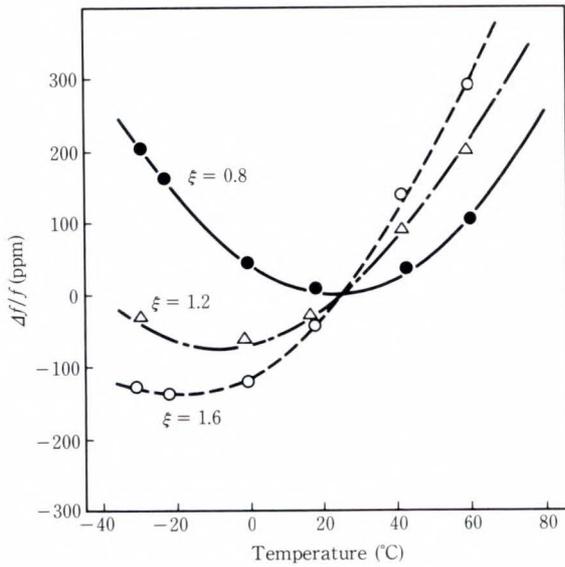


Fig. 19—Temperature characteristics of X-cut LT strip-type resonator with finite electrode.

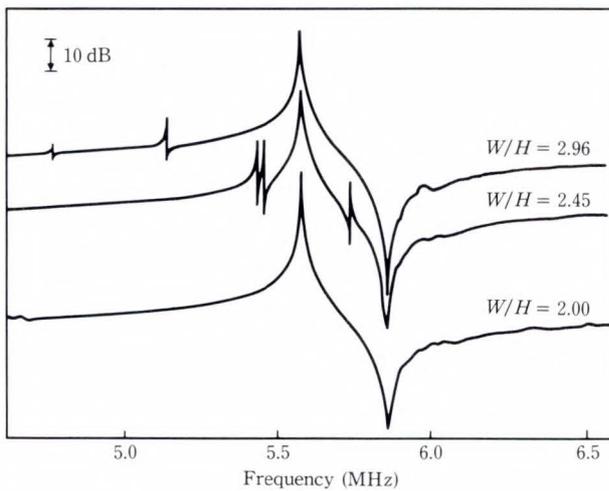


Fig. 20—Spurious responses caused by width of LT strip resonator.

order temperature coefficient is zero. On the other hand, the temperature coefficient for all crystal directions of LN is $-70 \text{ ppm}/^\circ\text{C}$ to $-100 \text{ ppm}/^\circ\text{C}$, and the turnaround temperature is quite different from the room temperature. The turnaround temperature cannot be shifted to near room temperature even if controlling the trapped energy value.

3) Optimization of strip dimensions²⁵⁾

The resonator element dimensions must

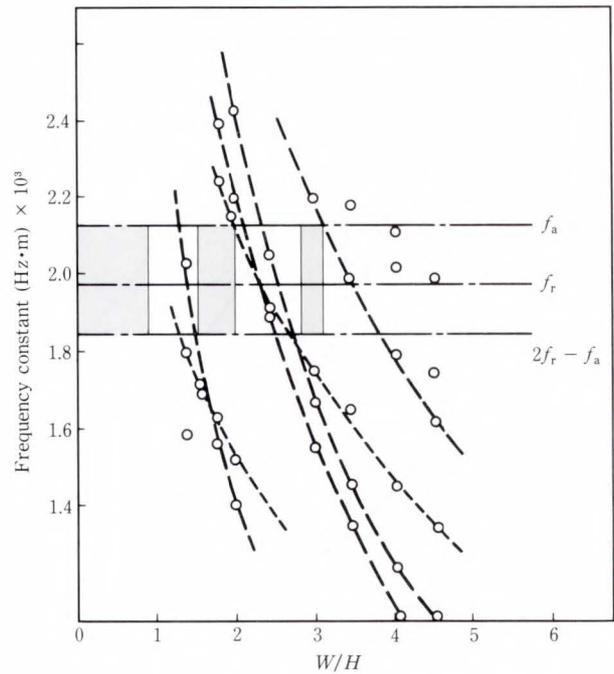


Fig. 21—Frequency spectrum of spurious responses plotted against the W/H ratio of the LiTaO_3 resonator.

determine the optimum ratio of strip width to thickness to separate the main mode and spurious mode due to strip width. To select a suitable strip width ($2W$), as shown in Fig. 20, to avoid generating spurious responses near the main resonance. A ratio of strip length to thickness must be determined so that the resonance energy dumping by supporting the element is sufficiently small.

Figure 21 shows the optimum strip width for the LT resonator. As shows by the area with hatched lines, spurious responses due to strip width where W/H is 1 or less, or near 2 and 3 do not affect to main mode. There is no resonance energy dumping if the resonance element is long enough, because of concentrating the resonant energy around center of elements. However, the resonator element should be as short as possible to be economic. To keep the size of the resonator element to a minimum, the parameter of resonance resistance was determined by experiment to be length L .

As Fig. 22 shows, L/H should be set to at least 10 for LT and at least 8 for LN. After

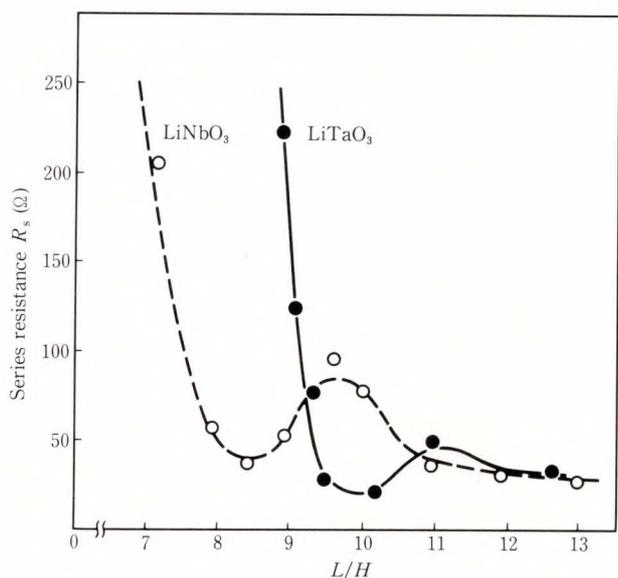


Fig. 22—Equivalent series resistance R_s plotted as a function of the ratio of the strip thickness on the LiTaO₃ and LiNbO₃ strip resonators.

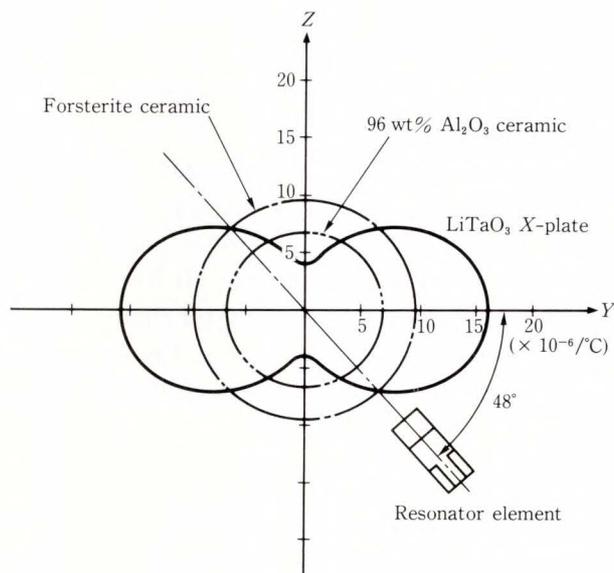


Fig. 24—Thermal expansion of LT X-cut plate and ceramic substrates.

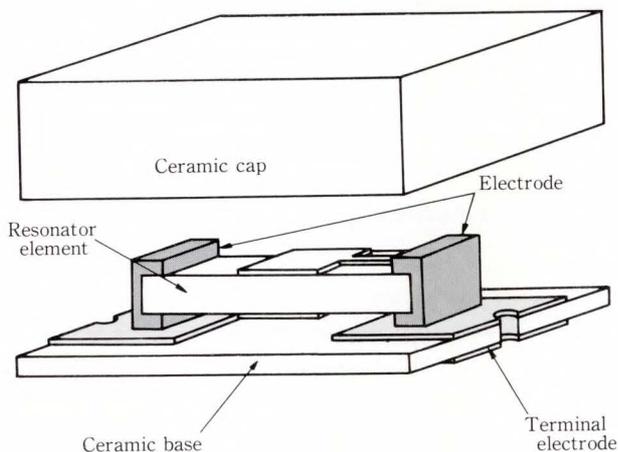


Fig. 23—Structure of the piezoelectric chip resonator.

investigation, an element one-third to one-fifth the size of a quartz or ceramic resonator was designed.

3.1.2 Resonator element packaging method²⁶⁾

Figure 23 shows the resonator packaging structure. An electrode pattern is formed by thick-film printing of AgPd conductive paste on the both side of a ceramic substrate. Connections between the both side are formed

when the thick-film is printed, by through-holes in the ceramic substrate edges. Using conductive epoxy resin, the resonator element is connected to the edge in the longitudinal direction of the element on the AgPd pattern of the ceramic substrate.

As stated in item 1) of Subsec. 3.1.1, the thermal expansion coefficient of the longitudinal direction of the resonator element is made to match the thermal expansion coefficient of the ceramic substrate. Figure 24 shows the angle dependence of the thermal expansion coefficient for the longitudinal direction of the X-cut LT resonator element with as function of crystal orientations.

The thermal expansion coefficient in an optimum orientation (-48° from the y axis) for the strip-type resonator is $9.7 \times 10^{-6}/^{\circ}\text{C}$. Forsterite ceramic, with a thermal expansion coefficient of $8 \times 10^{-6}/^{\circ}\text{C}$, makes the suitable substrate. Since $8 \times 10^{-6}/^{\circ}\text{C}$ is the thermal expansion coefficient for a 163° rotation Y-cut LN resonator element, a 96 wt% alumina substrate with a coefficient of $6.7 \times 10^{-6}/^{\circ}\text{C}$ is the suitable substrate in consideration of cost. The thermal stress dumping effect caused by the epoxy conductive resin is also related, and the

packaging structure withstands the thermal shock of a 260 °C solder dip.

Figure 25 shows the packaging structure of the resonator with built-in load capacitance for microprocessors. The two capacitors for load capacitance are formed by thick-film printing using hybrid technology. Stable oscillator can be obtained at a capacitance of 10-30 pF.

3.1.3 High-coupling resonator characteristics

The equivalent circuit in Fig. 26 shows the electrical characteristics near the resonance

frequency of a resonator using a piezoelectric crystal.

Figure 27 shows the admittance characteristics of a 4-MHz high-coupling resonator using LT and LN. Since it has a high coupling factor, it also has a wide inductive area.

Table 4 compares the 4-MHz resonator equivalent circuit constants for LT, LN, and quartz. Since the coupling factor is large, the LT and LN resonators have small capacitance ratio (C_0/C_s) which value is about 1/30 compared

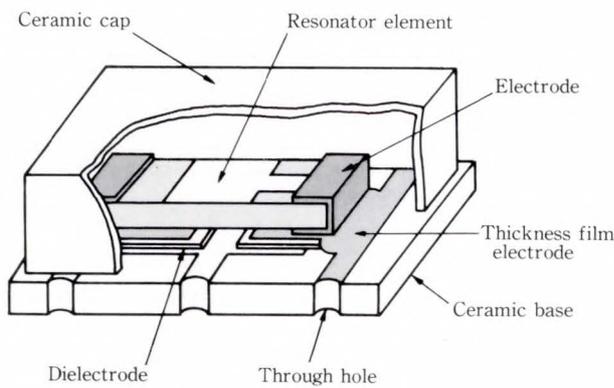


Fig. 25—Structure of LN resonator with built-in capacitors.

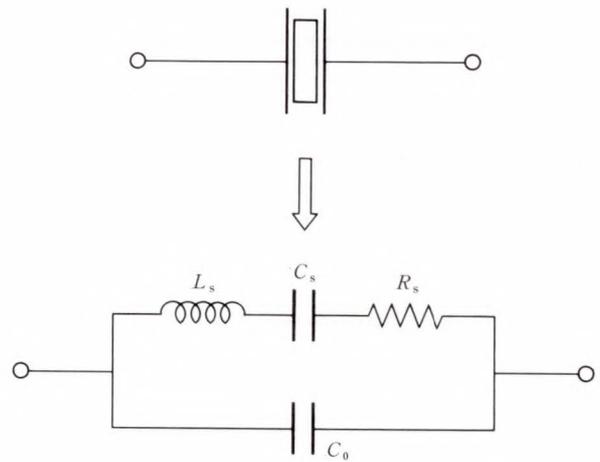


Fig. 26—Equivalent circuit of piezoelectric resonator.

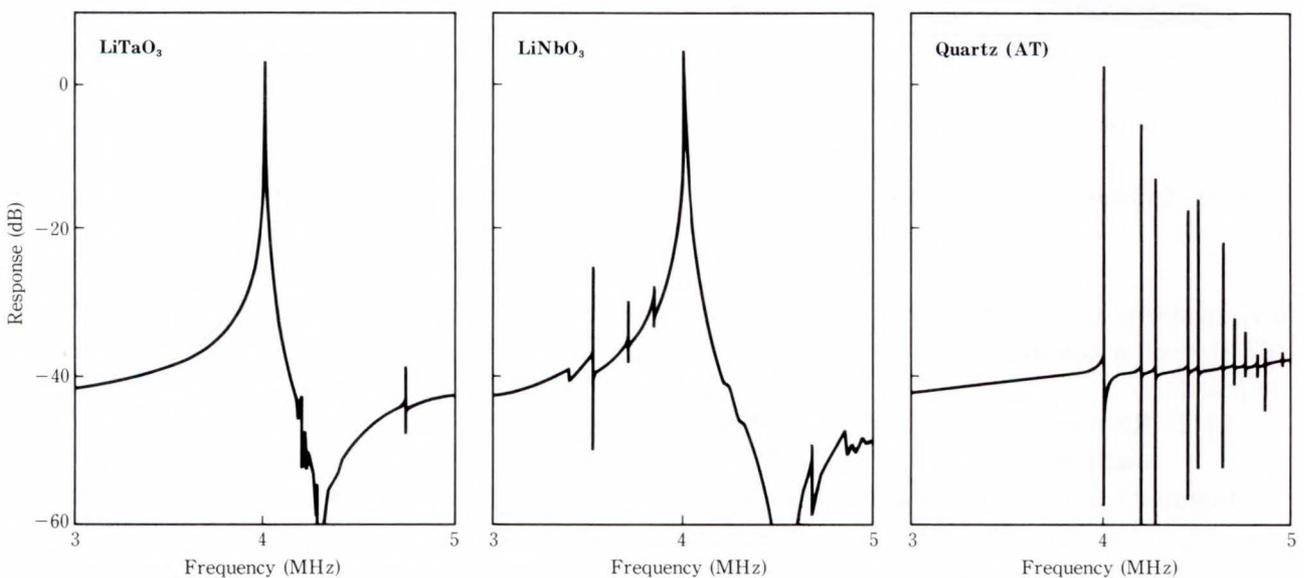


Fig. 27—Comparison of admittance characteristics of LT and LN strip-type resonators and AT-cut quartz resonator.

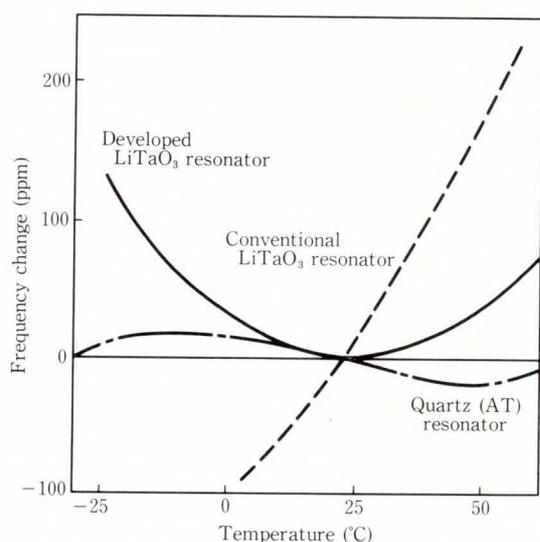


Fig. 28—Temperature characteristics of LT resonator.

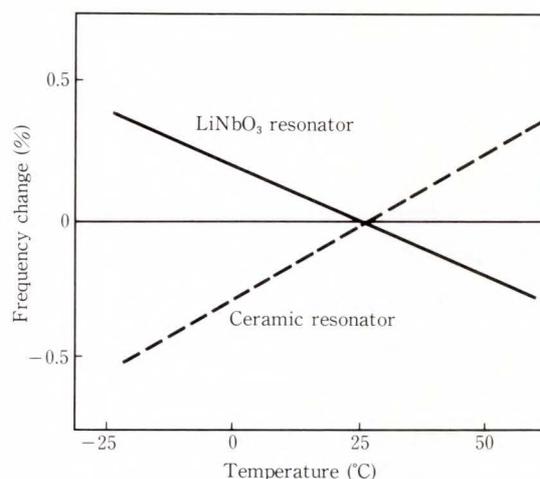


Fig. 29—Temperature characteristics of LN resonator.

with the quartz resonator. The inductive areas of LT and LN are about 20 times as large as those of quartz resonators. This shows that oscillation can occur over a wide frequency range.

Compared to the quartz resonator, the LT and LN resonators have few spurious response in frequencies near the main mode, and have small spurious response level. This means that when these high-coupling resonators are used in oscillation circuits, there are few oscillation frequency jumps or quenching accidents. As stated above, the temperature characteristic of resonance frequency is basically determined by the piezoelectric substrate and form of the drive electrode.

Table 4. Electrical characteristics of LiTaO₃, LiNbO₃ and AT-cut quartz 4 MHz resonators

	LiTaO ₃	LiNbO ₃	Quartz
Electromech. coupling coefficient	0.43	0.62	0.07
<i>Q</i> factor	5 000	3 000	200 000
Parallel capacitance (<i>C</i> _o) (PF)	3	3	4
Series capacitance (<i>C</i> _s) (PF)	0.39	0.51	0.012
Series inductance (<i>L</i> _s) (mH)	4.1	2.6	132
Capacitance ratio <i>r</i> (= <i>C</i> _o / <i>C</i> _s)	7.7	5.9	333
Temp. characteristics (-20 °C to +70 °C, ppm)	200	6 000	80

For the LT resonator, the relation between the crystal orientation and temperature characteristic was investigated with (element thickness)/(electrode length)(*H/L'*) as a parameter. This realized a resonator with a zero temperature coefficient near room temperature. LT resonator have a stable temperature characteristics with a frequency change of 150 ppm or less at -10 °C to +60 °C. Figure 28 shows that temperature characteristics of LT resonator (compared with AT-cut quartz resonator).

On investigating the crystal orientation of the LN resonator, a first order temperature coefficient of -100 ppm/°C was obtained, as Fig. 29 shows. This value is the same as that of a ceramic resonator with a stability of about 0.4 percent at of -10 °C to +60 °C. This is a satisfactory performance for clock signal generation in a single-chip microcomputer used in general mechanical control.

3.2 Edge reflection type resonator

The resonance frequency of the thickness shear mode is in inverse proportion to the thickness of piezoelectric elements. As the frequency increases, the thickness of elements must be processed so that it is thinner. This processing of about 70 μm is necessary for 30-MHz LT resonators. It is extremely difficult to manufacture wafers that have a stable yield. To manufacture resonators of more than

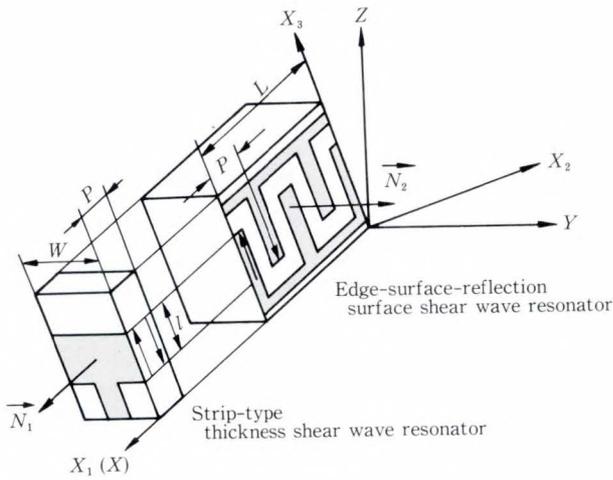


Fig. 30—Schematic of edge reflection type LT resonator.

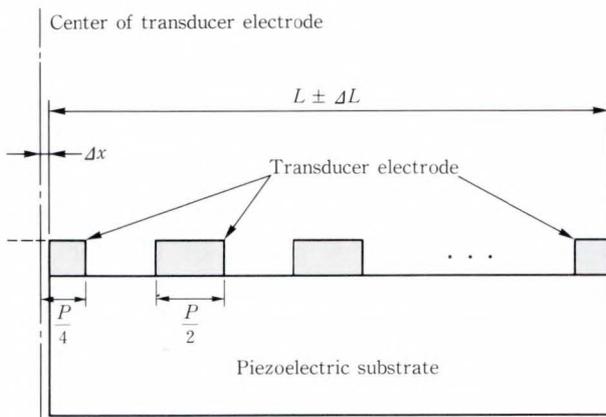


Fig. 31—Cross section of resonator element.

30 MHz, we developed a high-frequency resonator²⁷⁾ using surface shear wave excited by a interdigital transducer formed on one surface of the piezoelectric substrate.

Figure 30 shows schematic construction of edge reflection type resonator. This structure uses only the edge of many thickness shear resonator elements. The edges of an element are used as the reflectors to the wave to make the element smaller and to reduce the capacitance ratio. The direction of reflecting edges is matched to the surface shear wave particle displacement direction in the same way as for thickness shear resonance. As a result, resonance was obtained without spurious

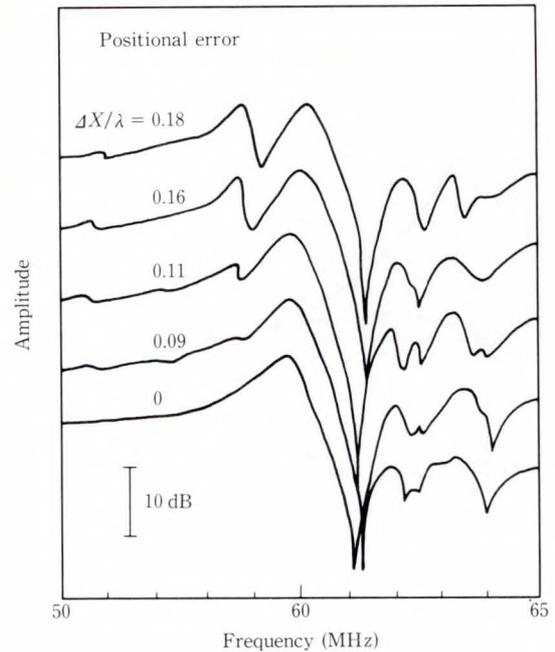


Fig. 32—Cut position error and resonance.

modes.

The admittance characteristic of this resonator is expressed by Equation (4).

$$Y = j\omega C_0 + 1 / \left\{ R_s + \frac{1}{2W_R C_0 K^2 \tan W/W_R} \right\} \dots (4)$$

Here,

$$C_0 = nWC_1, \text{ and } W_R = 2\pi v/L,$$

n : Number of finger pairs

W : Overlap length of fingers

C_1 : Capacitance of each finger unit

v : Velocity of surface shear wave

L : Distance between reflecting edges

K^2 : Surface shear wave coupling factor.

To prevent loss caused by electrode resistance or spreading sound field, the overlap length of fingers (W) should be 5λ to 15λ (λ : wavelength).

Figure 31 shows the relationship of the finger position and reflecting edge position. Ideally, the reflecting edge should be in the center of the finger. Δx is the distance from the center of the finger to the reflecting edge. When there is a reflection position error (Δx), the resonance frequency may be different from the designed frequency or high-order spurious responses may occur near the resonance fre-

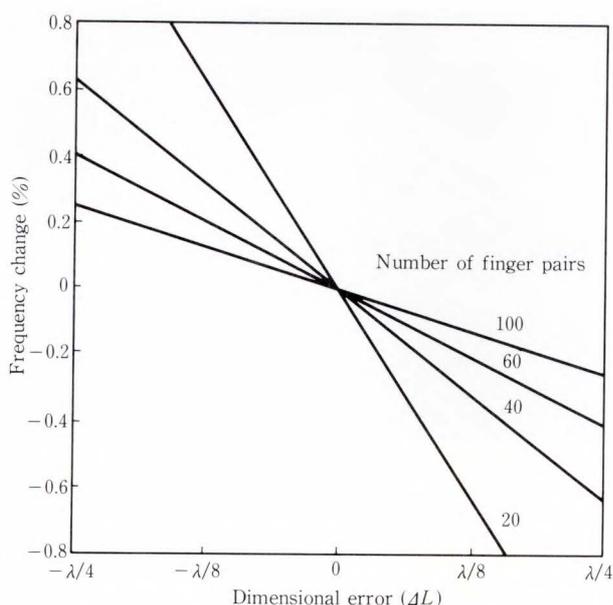


Fig. 33—Dimensional error of cavity length and frequency change.

quency.

Figure 32 shows the frequency characteristics when a positional error occurs on the reflecting edge.

Figure 33 shows the relationship between the number of finger pairs and the change in resonance frequency. The more pairs there are, the smaller the manufacturing variance in frequency caused by the position error (ΔL). However, a small number of finger pairs is effective in diverting the high-order spurious responses caused by a frequency pitch of $n/2W_R$ from the resonance point. The number of finger pairs is decided by the manufacturing frequency deviation and spurious response trade-off. We selected 20 to 40 pairs. We inferred the direction of the surface shear wave substrate from X-cut LT where the coupling factor for thickness shear resonance is large and the frequency temperature coefficient is low.

Figure 34 shows the characteristics of a 92-MHz resonator using a 42° rotation Y LT substrate. The capacitance ratio (γ) of this resonator is 20. SAW resonators using grating-type reflectors with conventional metal patterns have a capacitance ratio of 250. Consequently,

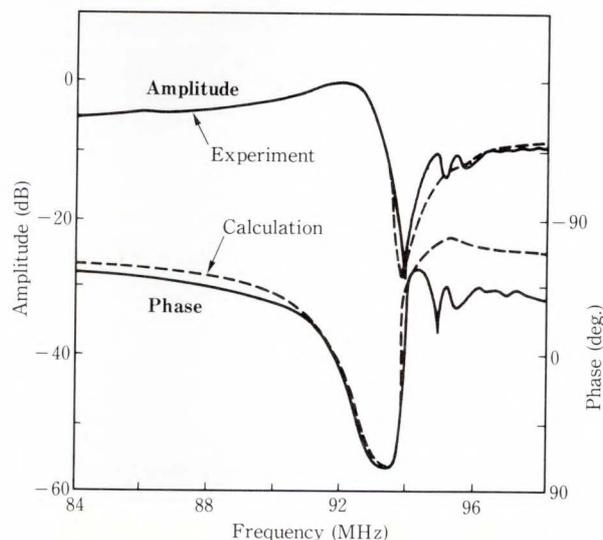


Fig. 34—Amplitude and phase characteristics.

this resonator is extremely effective in wide-band filter or wide variable range VCO applications because its capacitance ratio is about one-tenth that of conventional resonators.

3.3 Production method

About two million piezoelectric resonators are manufactured each month throughout the world. They must be more reliable than almost any other component because they are used for the microcomputer clock. There are many frequencies for resonators because of their many uses in electronic equipment. In the technological revolution, the extremely short life of electronic products is reflected in the very strong demand for inexpensive resonators that can be delivered quickly.

After investigating manufacturing lines to meet these customer demands, the following three targets were established.

- 1) To build large mass production manufacturing lines using large piezoelectric wafers and large ceramic substrates.
- 2) To build flexible manufacturing lines to manufacture resonators of the many required frequencies.
- 3) To build automated manufacturing lines by standardizing components and materials to make resonators.

A manufacturing process where 6-mm

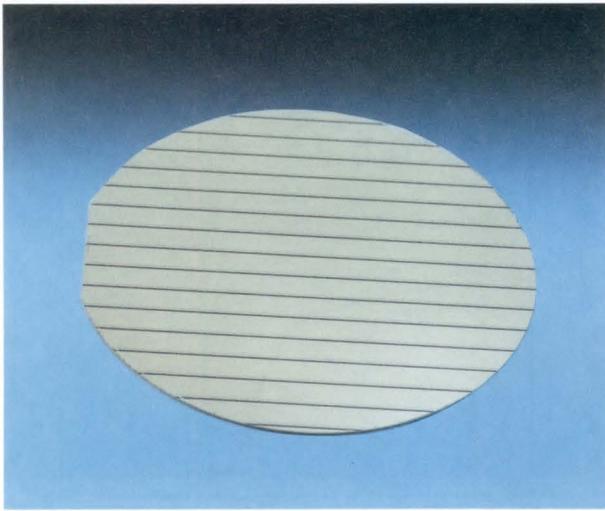


Fig. 35—Pattern-formed three-inch LN wafer (1 018 4-MHz elements/wafer).

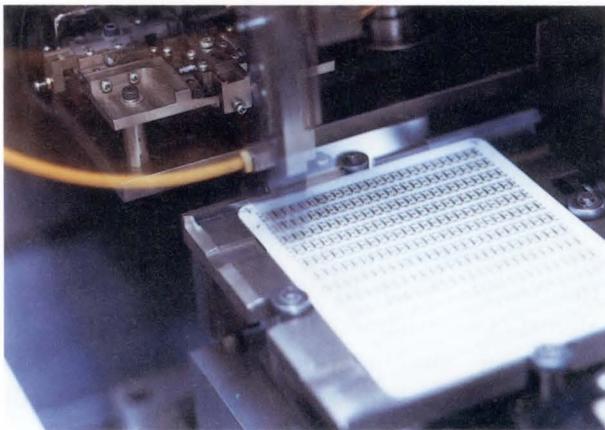


Fig. 36—Automatic mounting of resonator elements (338 6-MHz resonator elements/substrate).

and 8-mm diameter resonator elements are processed individually is used in manufacturing MHz-band quartz resonators, whereas the wafer process is used in manufacturing resonator elements for high-coupling piezoelectric resonators using LT and LN. As stated in section 2.2, the wafer size is determined by the frequency of the required resonator.

After the driving electrode is formed by bothside photolithography, the wafer is sliced with dicing saw and wire saw. Figure 35 shows the pattern-formed three-inch LN wafer. The resonant energy of strip-type resonator elements

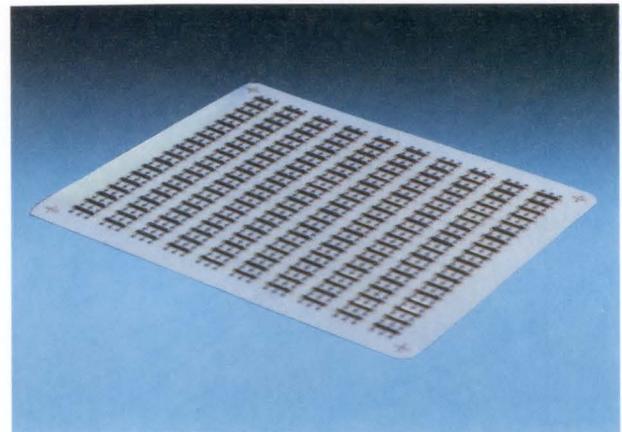


Fig. 37—Substrate with resonator elements mounted (180 3.58-MHz resonator elements/substrate).

is distributed all along the element width direction. For this reason, chipping at element edge during slicing will affect resonance characteristics such as generating spurious response and increasing equivalent series resistance (R_s). Consequently, a wire saw is used to cut longitudinal direction of the element. This wafer process efficiently mass-produces 1 468 elements on each 6-MHz resonator from 75-mm diameter LN wafer, 436 elements on each 9-MHz resonator from 50-mm diameter LN wafer, and 195 elements on each 12-MHz resonator from 30-mm diameter LN wafer.

Space must be formed around the central part of the resonator elements driven electrically because the center causes mechanical vibration. Resonator elements mounted on ceramic substrates must be sealed with a ceramic cap. In high-precision quartz resonators, elements are sometimes vacuum sealed to prevent elastic vibration dumping by air. Since quartz has a low coupling factor, it is not very resistant to dumping. The ceramic substrates are 144 mm \times 94 mm. 180 pieces of 4-MHz resonator element or 338 pieces of 6-MHz element may be mounted per substrate. These ceramic substrates have through-holes for electrical conduction between the both side, and break grooves for separating of ceramic substrate after cap sealing.

Figure 36 shows where the resonator elements are automatically mounted. Even if

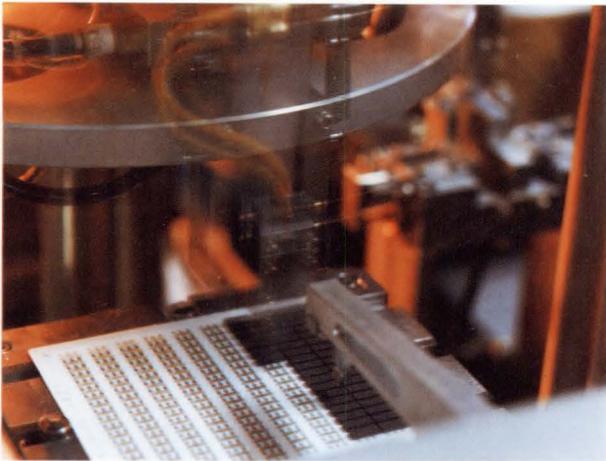


Fig. 38—Ceramic cap encapsulation by automatic cap sealer.

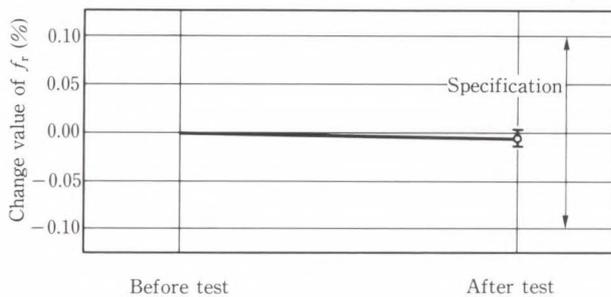


Fig. 39—Frequency change after heat stress test at 260°C for 10 s.

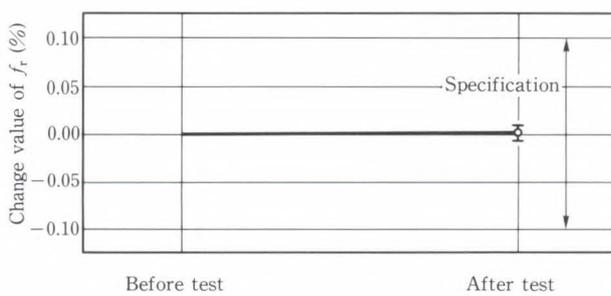


Fig. 40—Frequency change after mechanical shock test of 3 000 G three times in each of three directions.

the element size is different, the same mounting unit can be used by changing the vacuum tip of the element.

Figure 37 shows a substrate with elements

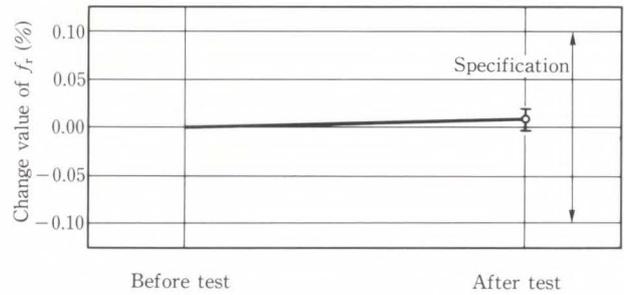


Fig. 41—Frequency change after vibration test at 10-55 Hz 1.5 mm amplitude three directions for 2 h each.

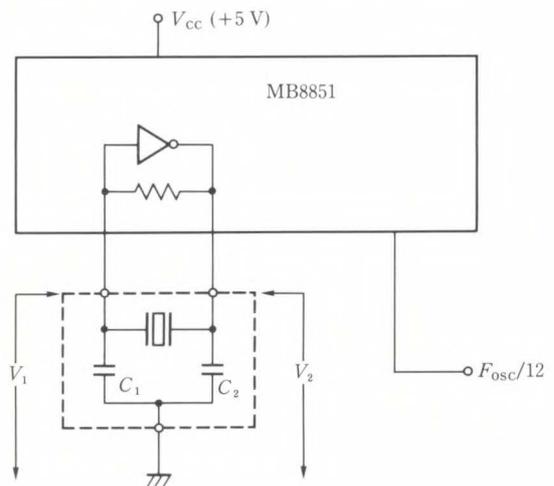


Fig. 42—Single-chip microprocessor and clock oscillator.

mounted. Most of this manufacturing and assembly process uses mass production, because the automation is used to mount the elements, trim the frequency by laser, and seal the caps on large ceramic substrates. Figure 38 shows where the ceramic caps are automatically sealed.

If caps of different sizes are to be used, the automatic mounting unit can be used by changing the automatic ceramic cap feeder and the cap seal pitch.

Resonators separated along the grooves to automatic testing for SMT packaging, then to lead frame connection for SIP types. Figures 39, 40, and 41 show environmental characteristics of high-coupling resonators that have passed through these processes.

These resonators can withstand severe

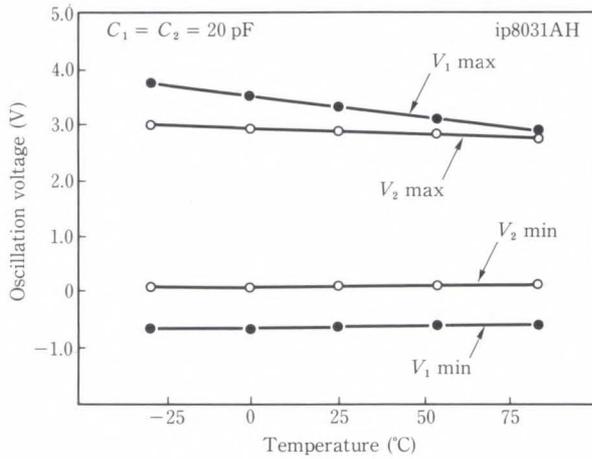


Fig. 43—Oscillation voltage vs. temperature.

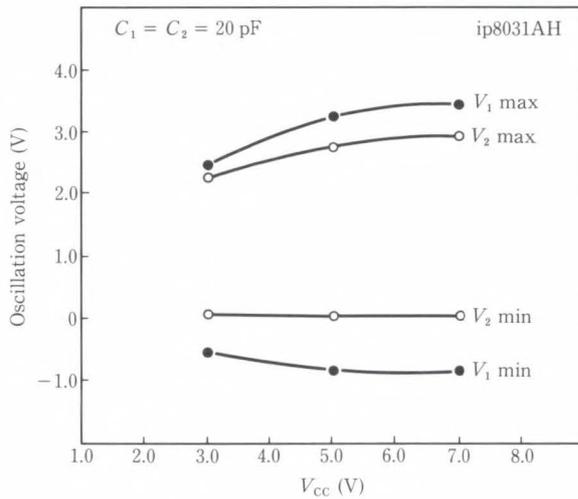


Fig. 44—Oscillation voltage vs. V_{CC} .

thermal and mechanical shocks, and can be mounted in car engines room. They have a good performance of $\Delta f/f < \pm 0.1\%$ for 1000 h at storage temperatures of $-55\text{ }^\circ\text{C}$ to $+100\text{ }^\circ\text{C}$, and also have the superior reliability not possible in piezoelectric ceramic resonators in THB tests ($85\text{ }^\circ\text{C}$, 95% RH, 12 VDC).

3.4 Applications²⁸⁾

Resonators are used in clock signal generators of microprocessor. Figure 42 shows a resonator connected to a single-chip microprocessor. Conventionally, two capacitors adjust the gain and phase of the driver IC and the

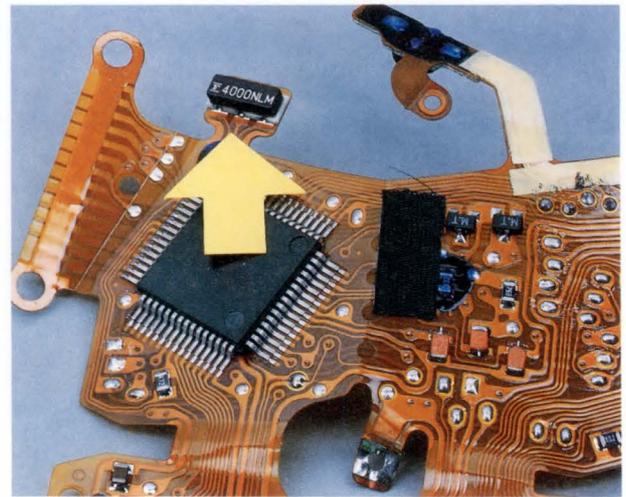


Fig. 45—Chip resonator mounted on a flexible PC board in an electronic camera.

resonator. These capacitors are usually connected to the Xtal and EXtal terminals of the microprocessor. Our new resonators have built-in capacitors. We investigated many microprocessors on the market and now offer resonators with three sizes of built-in capacitors (10 pF, 20 pF, and 30 pF).

Figure 43 shows temperature characteristics for different oscillator output voltages. It shows stable oscillation over a wide temperature range. Figure 44 shows the oscillator output voltage over a range of power supply voltages. The clock signal assures a $\pm 0.8\text{ V}$ margin even for wide variations of about 20 percent in the 5 V power supply.

Building the capacitors into the resonator chip has reduced the number of components to one third. The three-terminal chip has one-third the mounting area and one-third the assembly cost.

The resonator is very small and can be surface-mounted, so it is ideal for equipment that must be compact. Video cameras, electronic cameras, washing machines, and hybrid ICs are just some of the applications. Many more are expected in the near future.

Figure 45 shows a chip resonator mounted on a flexible printed circuit board in an electronic camera. Figure 46 shows a chip resonator mounted on a hybrid IC for micro-

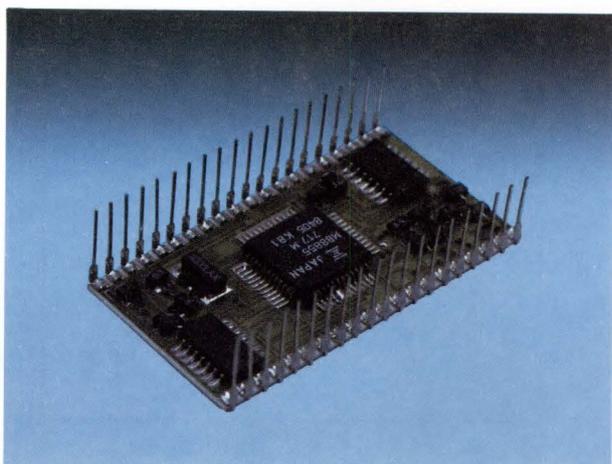


Fig. 46—Chip resonator mounted on a hybrid IC for micro-floppy disk drive control.

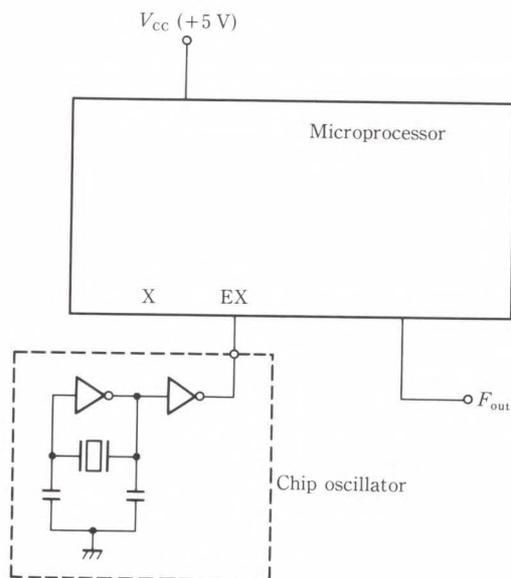


Fig. 47—Chip oscillator structure.

floppy disk drive control.

Figure 47 shows the structure of a chip oscillator. An oscillator is usually connected to the EXtal terminal of a microprocessor to generate the clock signal. Because the accuracy of the clock signals depends on the accuracy of the oscillator, the oscillator frequency must be stable with changes in power supply voltage.

4. Wide-band, high-stability filter

4.1 Electrical characteristics

There is a wide frequency range between

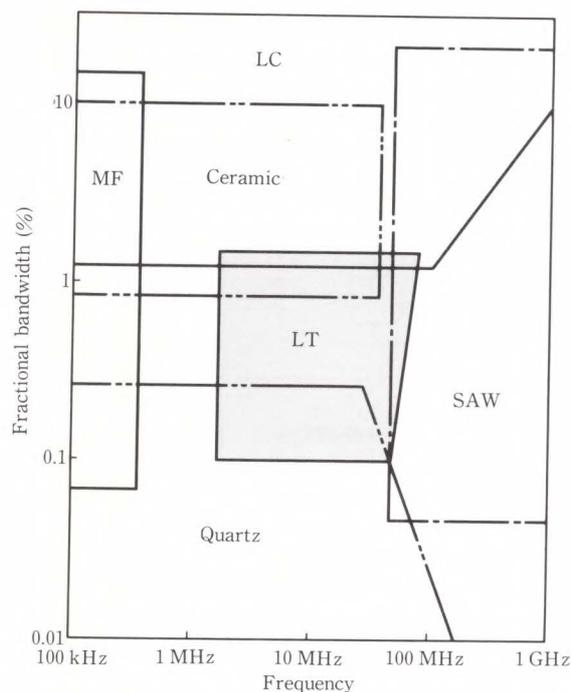
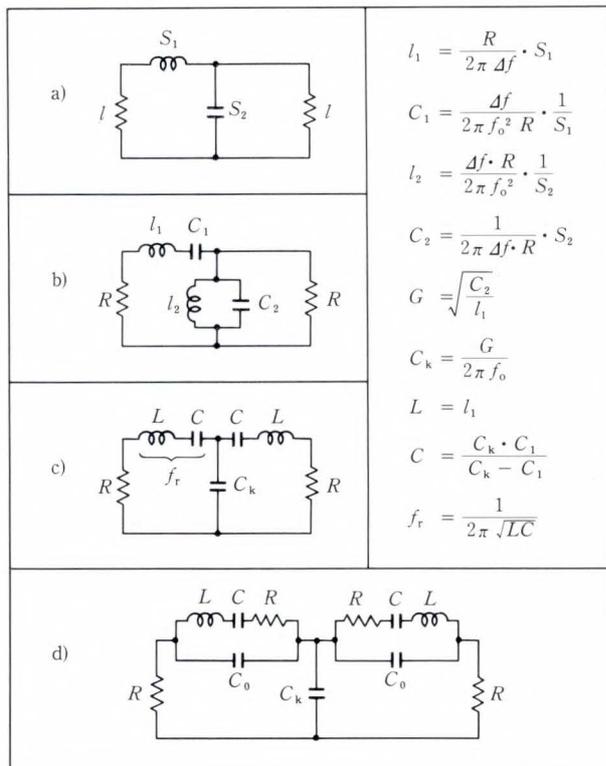


Fig. 48—Application area of filters.

the resonance and antiresonance of high-coupling piezoelectric resonators. Consequently, a wide-band and high-stability bandpass filter can be obtained by LT resonator. Bands that could conventionally only be reached with LC filters using coils and capacitors can now be reached by an LT filter. As a result, a filter can be realized with a stability that cannot be obtained by LC filters. Figure 48 compares the fractional bandwidth that can be manufactured for different filters used in the 1-100 MHz band. The LT filter can cover bands that are difficult to reach with conventional LC, quartz, ceramic, or SAW filters.

This filter has excellent characteristics as a bandpass filter for a fractional bandwidth of about one percent required for timing extraction in digital transmission system. The image parameter and effective parameter methods are used to design electromechanical filters. The LT filter is designed by the effective parameter method, which allows high-level design. This method follows the procedure shown in Fig. 49.

Frequency conversion from the non-loss



f_0 : Center frequency Δf : Bandwidth

Fig. 49—Filter design procedure.

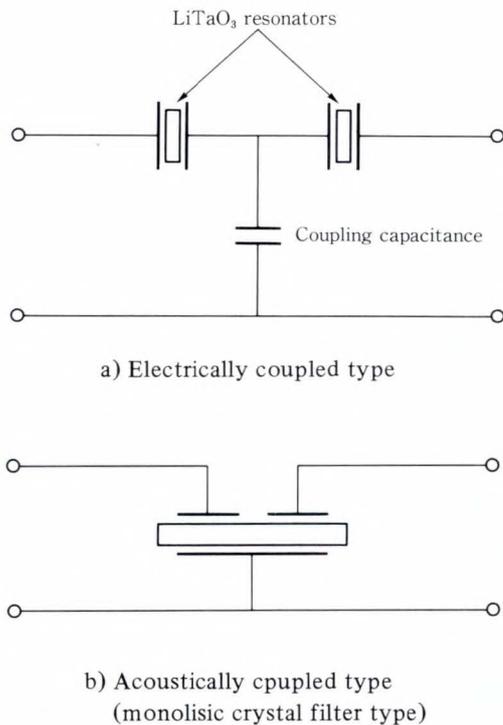


Fig. 50—Filter circuit structure.

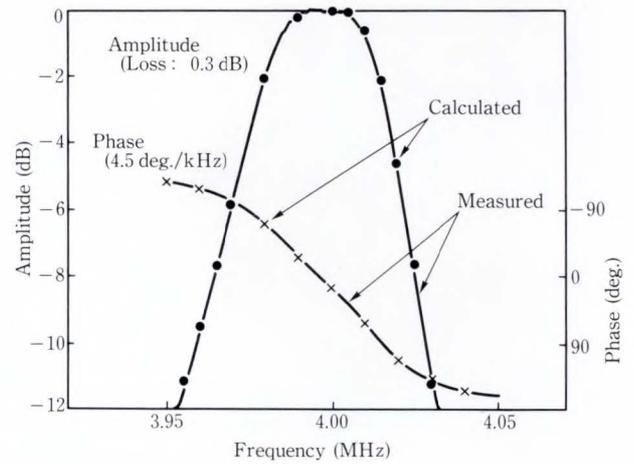


Fig. 51—Frequency characteristics of a bandpass filter using 4-MHz LT resonators.

reference low-pass filter and gyrator conversion are performed, and a ladder circuit connecting the resonance circuit in series is obtained, as -c) of Fig. 49 shows.

Normally, a transformer is used to match the impedance with the external circuit. In LT filter design, impedance is matched with external circuits, premising that a matching transformer is not used.

Resonance coupling for a pair of resonance circuits related to the filter bandwidth can be performed electrically or acoustically. The coupling capacitor (C_k) in Fig. 49-c) performs the coupling electrically. To perform coupling acoustically, two resonators are built into the same resonator element. Acoustic coupling (called monolithic crystal filter) does not require a coupling capacitor, is small, and has a good stop band characteristic. However, 1-MHz to 3-MHz filters have large elements, and it is difficult to trim the degree of coupling compared with the electrical coupling type. Figure 50 shows the electrical coupling and acoustic coupling filter structures.

The timing extraction LT filter has been designed by the electrical coupling method. The resonator in the filter circuit has the parallel capacitance (C_0) and resonance resistance (R_s) as shown in Fig. 26. This C_0 acts as the dumped capacitance. The circuit shown in d) of Fig. 49 is simulated, and the capacitance ratio (γ) and

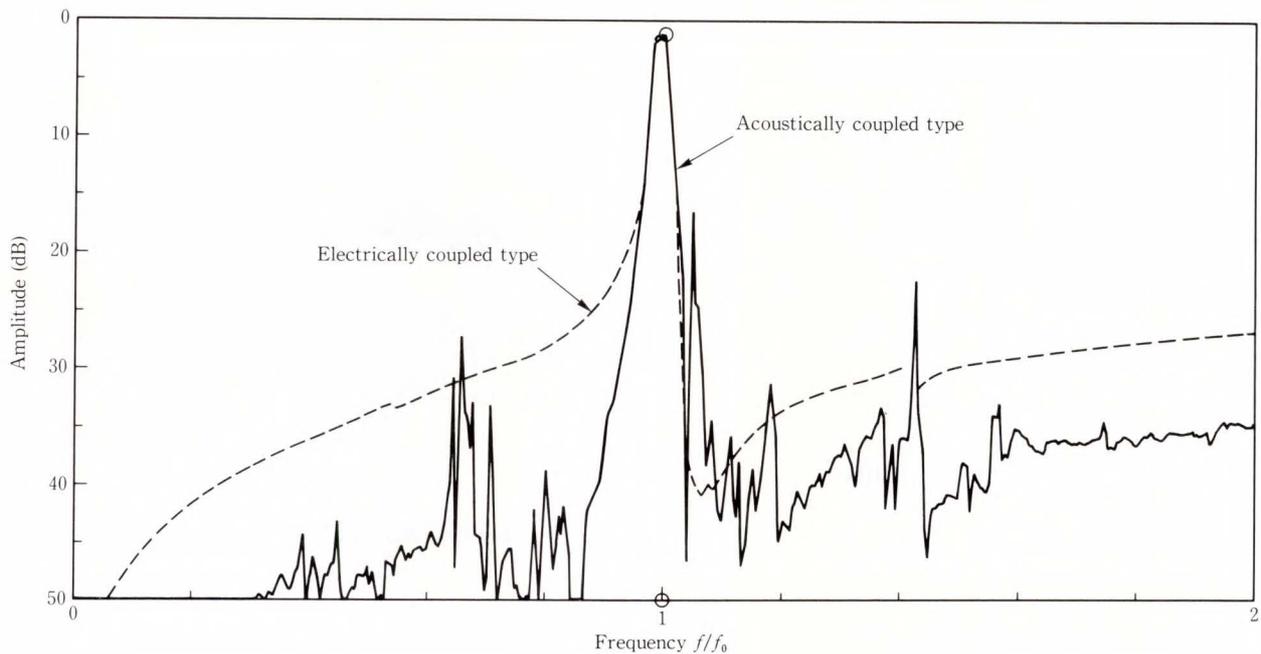


Fig. 52—Stop band characteristics of LT filters.

effect of Q are feed back to the design. Figure 51 shows the pass band characteristics of an LT filter with electrical coupling. The center frequency is 4 MHz, and the fractional bandwidth with attenuation of 3 dB is one percent. This value is matched with the target fractional bandwidth required by the timing extraction filter in PCM communications. The LT resonator has a resonant frequency of 3.983 5 MHz, a resonant resistance of 30 Ω , inductance of 5.1 mH, parallel (dumped) capacitance of 2.3 pF, and Q of 4 200. The coupling capacitance is 41 pF, and the terminal resistance is 1 k Ω . An insertion loss of 0.3 dB and phase linearity in the pass band are obtained. The phase inclination is 4.5 $^\circ$ /kHz. The amplitude and phase characteristic match the simulation result of the lumped constant approximation, showing that the design method has been established.

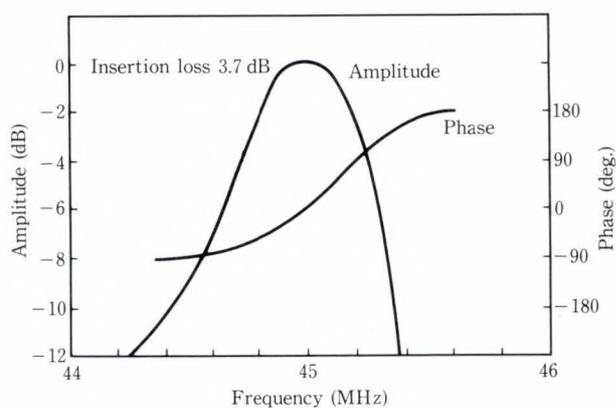
The temperature characteristic of the center frequency of the filter using the thickness shear LT resonator is 200 ppm or less at temperatures of -10°C to $+60^\circ\text{C}$. This value is approximately the same as the temperature characteristic of the resonance frequency of the resonator.

Figure 52 shows the stop band characteristics of the LT filter. The attenuation of electrically coupled filters is bad at higher frequency because of the dumped capacitance, but the characteristics are good enough for timing extraction. The stop band characteristics of acoustically coupled filters has a good attenuation of 5-35 dB, better than that of electrically coupled filters. Figure 53 shows the frequency characteristic of a 45-MHz filter using an edge reflection type LT resonator, with a fractional bandwidth of about one percent. It has an insertion loss of 3.7 dB, about 15 dB better than that of a normal SAW filter.

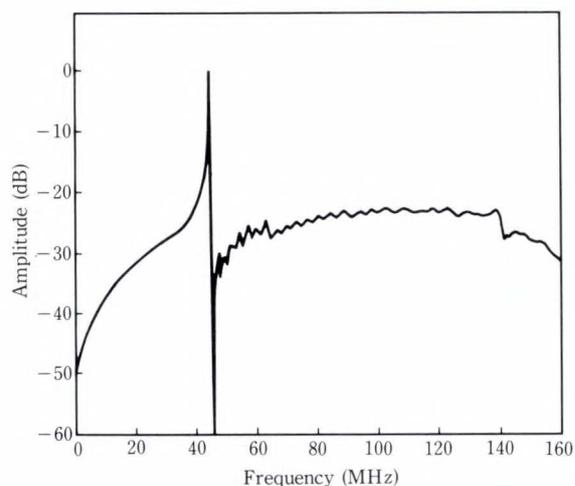
4.2 Production method

Filters used in transmission equipment require high reliability. Severe specifications are set for the aging characteristic of frequency, in particular. The packaging structure of LT filters is designed with it in mind.

Figure 54 shows the packaging structure of an LT filter. Resonator elements are packaged on a ceramic substrate through thin pieces of metal to avoid frequency fluctuation caused by irreversible stress due to mechanical shock



a) Pass-band characteristics



b) Stop-band characteristics

Fig. 53—Frequency characteristics of a bandpass filter using 45-MHz LT resonators.

or thermal expansion. This structure suppresses the frequency fluctuation to 10 ppm or less¹⁶⁾. These metal pieces are nickel-silver, 80 μm thick, cut and bent by high-precision press processing. To increase the reliability of electrical connections, copper and gold are plated electrically onto the pieces of metal. A reel-to-reel process is used from processing to plating to make a good manufacturing line. The resonator elements are packaged by batch processing using 14 connected lead frames. The metal pieces are stuck to the resonator elements with conductive silver epoxy adhesive, and 40-60 soldering is used to attach the metal pieces to the AgPd

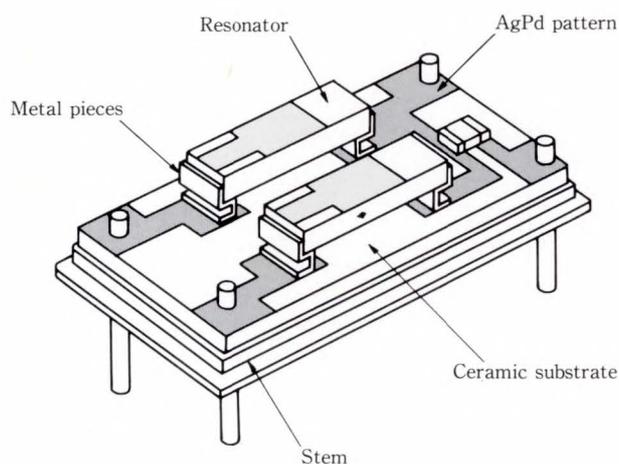


Fig. 54—Packaging structure of an LT filter.

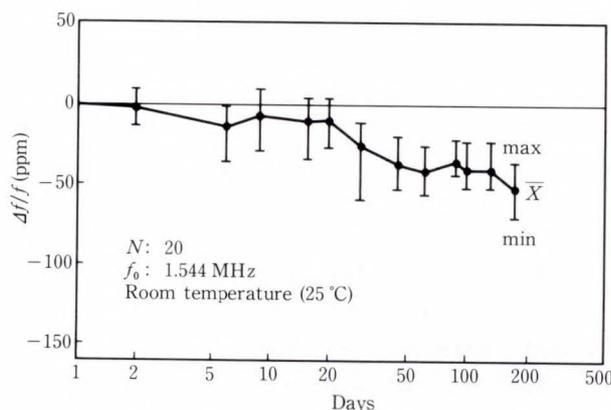


Fig. 55—Aging characteristic at room temperature¹⁶⁾.

pattern on the ceramic substrate. Edge reflection type LT resonators are used in filters of 30 MHz or more. In the same way as surface wave devices, the resonance energy is concentrated around the piezoelectric element surface. Consequently, epoxy adhesives are used to stick the back of the resonator to the ceramic substrate. Ceramic substrates with LT resonators and coupling capacitors are hermetically sealed in a dry N₂ gas using metal stems and caps. Figure 55 shows the aging characteristic of LT filters. An aging characteristic of only -35 ppm/decade is obtained, allowing the filters to be used for a long time.

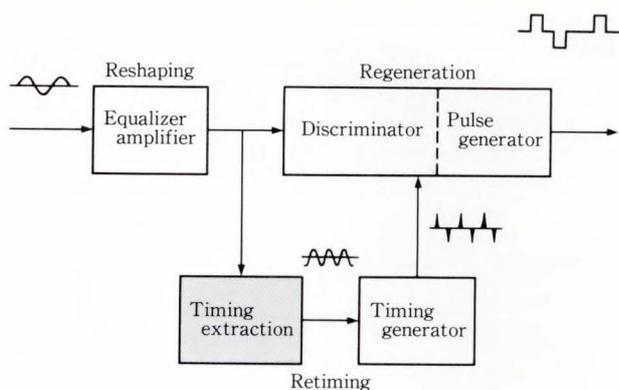


Fig. 56—Block diagram of repeater.

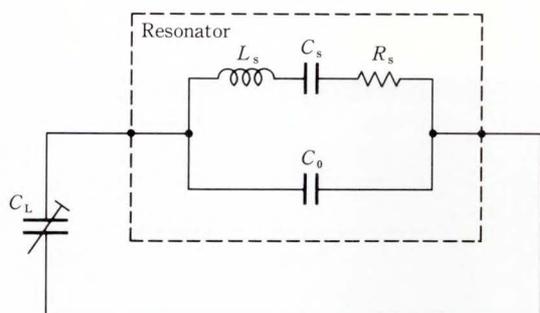


Fig. 57—Oscillation circuit.

4.3 Applications

With the worldwide expansion in digital networks, there is an increasing demand for more compact, lighter PCM multiplexers and repeaters that do not require tuning. As stated in the introduction, LT filters are used because they have a wide band and are very stable. When transmitted over long distances, the waveform of digital signals may become distorted and attenuated. Repeaters are needed to boost the signals. Figure 56 is a block diagram of the timing extraction circuit in the repeater. The hatched timing extraction circuit generates time reference clocks from input signals. The circuit requires a bandpass filter with a fractional bandwidth of 0.1-1 percent. A wide-band filter is especially necessary to extract clocks from signals with a lot of jitter.

5. Voltage controlled oscillator (VCO)

Phase locked loop (PLL) circuits that syn-

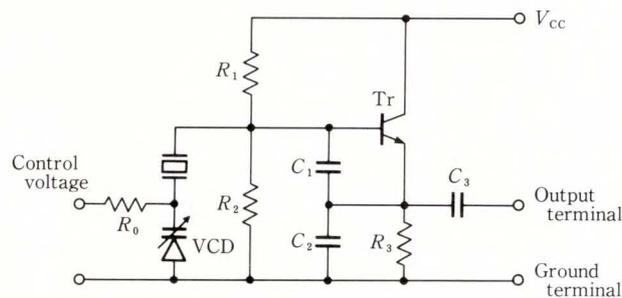


Fig. 58—Colpitts type VCO circuit.

chronize digital signals between electronics equipment use voltage controlled oscillators (VCOs) that can change the oscillation frequency in DC voltage. A VCO for this purpose is required to have high frequency stability and a wide frequency variable range.

5.1 Electrical characteristics

The load capacitance (C_L) and change in oscillation frequency in the oscillation circuit shown in Fig. 57 are expressed by Equation (5).

$$\Delta f/f = (2\gamma)^{-1} \{1/(1 + C_L/C_0)\}, \quad \dots (5)$$

Δf : Frequency change

f : Oscillation frequency

γ : Capacitance ratio (C_0/C_s)

C_0 : Parallel capacitance

C_s : Series capacitance

C_L : Load capacitance of resonator in oscillator circuit.

From Equation (5), the figure of merit²⁹ (M) of the resonator for a VCO is expressed as shown in Equation (6).

$$M = 1/(\gamma \cdot \Delta\tau). \quad \dots (6)$$

Here,

$$\Delta\tau = (f_2 - f_1)/f_0,$$

f_0 : Oscillation frequency at reference temperature

f_1 and f_2 : Minimum and maximum frequencies in operating temperature range.

A resonator where the product of γ and $\Delta\tau$ is small is excellent for a VCO. Table 4 shows that a high-coupling resonator has a small value of γ compared to a quartz resonator.

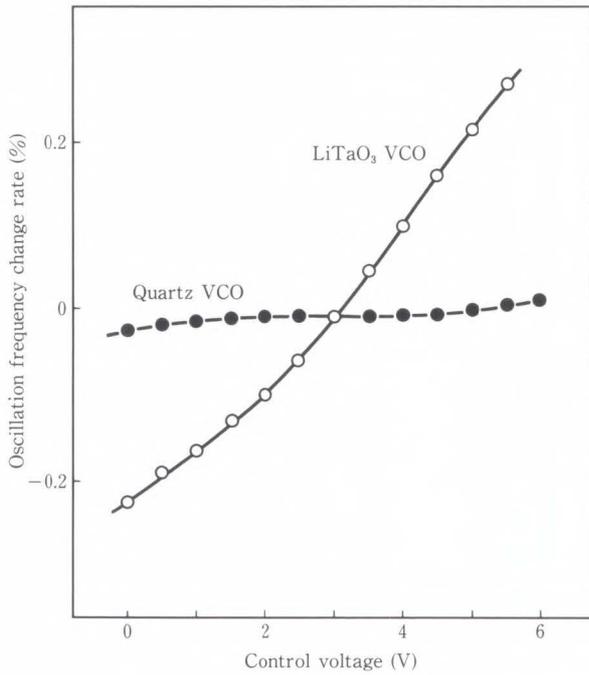
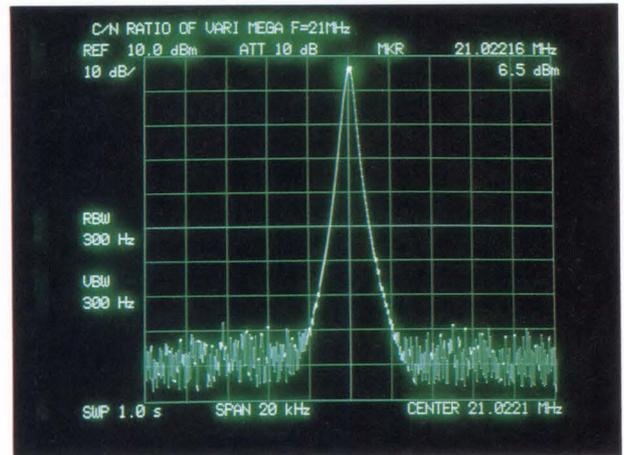


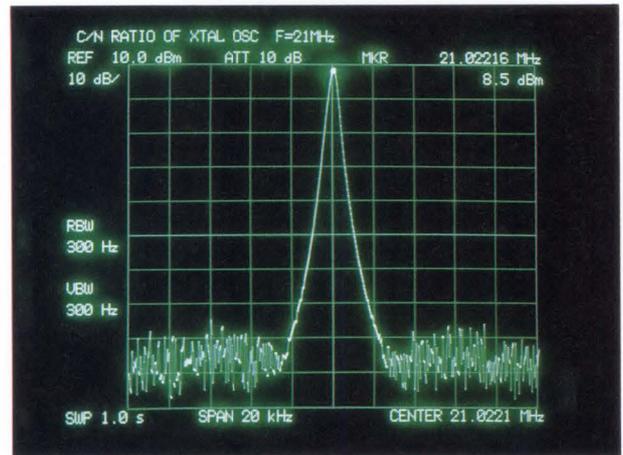
Fig. 59—Characteristics of voltage-controlled oscillators.

The figures of merit are 650 ($\Delta\tau = 200$ ppm) for an LT resonator and 40 ($\Delta\tau = 80$ ppm) for a quartz resonator. The LT resonator therefore has a wider frequency variable range than the quartz resonator has, as a VCO.

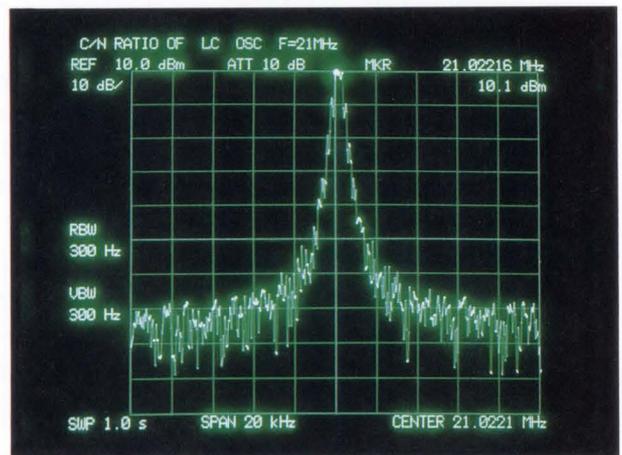
When DC voltage is applied to the variable capacitance diodes inserted in series in the resonators in a Colpitts circuit, the oscillation frequency is changed continuously. Good VCO characteristics are obtained in the circuit shown in Fig. 58. Figure 59 shows the relationship between the control voltage and oscillation frequency obtained in this circuit. Given the same load capacitance change, the difference between frequency changes in LT resonators is about 20 times wider than that in quartz resonators. Table 4 shows that the LT resonator has a small γ , and a relatively large Q ($= 5000$), so the carrier/noise (C/N) ratio of the output signal is extremely good. Figure 60 shows the output spectra when LC, LT, and quartz resonators are used in the VCO circuit. An LT-VCO circuit with a C/N ratio roughly equal to that of a quartz-VCO circuit was obtained. It had a frequency variable range 20 times wider than that of a quartz VCO circuit.



a) Using an LT resonator



b) Using a quartz resonator



c) Using an LC circuit

Fig. 60— C/N ratio of VCO.

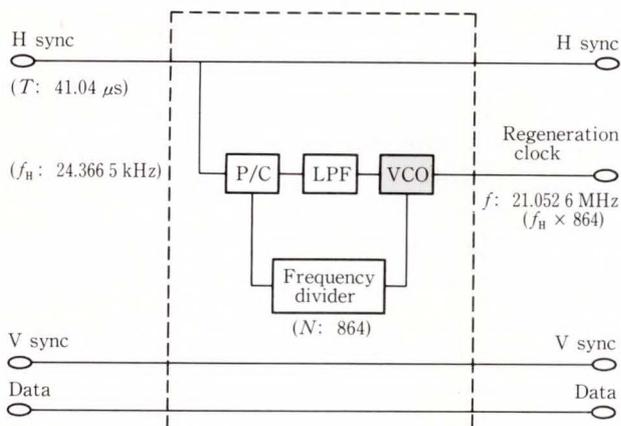


Fig. 61—Block diagram of bit clock regeneration from horizontal synchronized signals.

5.2 Applications

VCOs using LT resonators have a frequency variable range 20 times as wide as VCOs using quartz resonators, and better stability than VCOs using LC. Therefore potentially, there are many applications for them. Digital technology used in electronic equipment in data processing and communication is extending to other fields. The latest signal processing technology has begun to be used in home audiovisual equipment, such as VTRs with digital memory, and DAT (digital audio tape recorder) that uses PCM signal processing. Signal phase synchronization is necessary for error-free, efficient signal communication within and between electronics equipment.

As digital technology applications in electronic equipment have increased, the synchronization range has had to expand, and very stable phase synchronization circuits are now required. For example, PDP displays and LCDs are now used instead of CRT displays on OA equipment. In this display application, stable and high-speed bit clock signals corresponding to individual dots on the display must be generated using horizontal synchronized signals. Figure 61 is a block diagram of the circuit regenerating bit clocks from horizontal synchronized signals. The connected personal computer supplies horizontal synchronized signals for reference. These signals for different frequencies among the computers, individually and the

display must correspond to the difference. This kind of circuit, that is, one that obtains different frequencies synchronized with signals of a certain frequency, is required by many electronic equipment. DATs, digital VTRs, and optical disk drives, all expected to grow rapidly, will require very stable VCOs with wide frequency variable ranges.

6. Conclusion

In data processing and transmission, the key devices today are oscillators to generate the basic digital signals, filters to suppress unwanted signals and let the required signals pass through, and VCOs for synchronization of digital electronic equipment. Piezoelectric devices using quartz and high-coupling piezoelectric devices using LiTaO_3 (LT) and LiNbO_3 (LN) will become increasingly important.

Large-scale integration of high-speed semiconductor microchips using GaAs is being announced for oscillators and digital filters. To improve cost performance at high frequencies, different devices use different frequencies. Thickness shear resonators, edge reflection surface shear wave resonators, and pattern grating SAW resonators using LT and LN are the basic 1-100 MHz devices. SAW and surface transverse wave (STW) devices using LT, LN, LB, and quartz are the basic 100-MHz to 1-GHz devices. Magnetic static wave (MSW) devices using YIG/GGG and STW devices using LT, LB, and quartz are the basic 1-100 GHz devices.

In future, in addition to resonators, filters, and VCOs, the application range will extend to realtime signal processing (for example, signal compression, expansion, and confidential signals) in the high-frequency band, based on convolution functions using elastic waves.

We plan to increase our market competitiveness by making key devices for systems. To do so, we must develop new materials, process technologies, and design technologies.

References

- 1) Matthias, B.T., and Remeika, J.P.: Ferroelectricity in the Ilmenite structure. *Phys. Rev.*, **76**, p. 1886 (1949).

- 2) Ballman, A.A.: Growth of Piezoelectric and Ferroelectric Materials by the Czochralski Technique. *J. Am. Ceram. Soc.*, **48**, 2, pp. 112-113 (1965).
- 3) Brandle, C.D., and Miller, D.C.: Czochralski Growth of Large Diameter LiTaO₃ Crystals. *J. Cryst. Growth*, **24/25**, pp. 432-436 (1974).
- 4) Valentino, A.J., and Brandle, C.D.: Diameter Control of Czochralski Growth Crystals. *J. Cryst. Growth*, **26**, pp. 1-5 (1974).
- 5) Satoh, T., Inui, J., and Iwamoto, H.: Automatic Control System for Czochralski Growth of Large Diameter LiNbO₃ Crystals. *FUJITSU Sci. Tech. J.*, **12**, 1, pp. 93-113 (1976).
- 6) Kolb, E.D., Glass, A.M., Rosenberg, R.L., Grenier, J.C., and Laudise, R.A.: Dielectric and Piezoelectric Properties of Aluminum Phosphate. ULTRASONIC SYMP., 1981, pp. 332-336.
- 7) Shorrocks, N.M., Whatmore, R.W., Ainger, F.W., and Young, I.M.: Lithium Tetraborate — A New Temperature Compensated Piezoelectric Substrate Material for Surface Acoustic Wave Devices. ULTRASONIC SYMP., 1981, pp. 337-340.
- 8) Peach, R.C., Emin, C.D.J., Werner, J.F., and Doherty, S.P.: High Coupling Piezoelectric Resonators using Lithium Tetraborate. ULTRASONIC SYMP., 1983, pp. 521-526.
- 9) O'Connell, R.M.: Cuts of Lead Potassium Niobate, Pb₂KNb₅O₁₅, for Surface Acoustic Wave (SAW) Applications. *J. Appl. Phys.*, **49**, 6, pp. 3324-3327 (1978).
- 10) Slobodnik Jr., A.J., and Szabo, T.L.: Minimal Diffraction Cuts for Acoustic Surface Wave Propagation on Bi₁₂GeO₂₀. *J. Appl. Phys.*, **44**, 7, pp. 2937-2941 (1973).
- 11) Rehwald, W.: Anomalous Ultrasonic Attenuation in Bi₁₂GeO₂₀, Bi₁₂SiO₂₀, and Bi₁₂(Ge_{0.5}Si_{0.5})O₂₀. *J. Appl. Phys.*, **44**, 7, pp. 3017-3021 (1973).
- 12) Yamada, K., Omi, T., Matsumura, S., and Nishimura, T.: Growth and Fabrication of High-Quality 4-inch Large LiTaO₃ SAW Substrate. *Toshiba Rev.*, **152**, pp. 30-33 (1985).
- 13) Détaint, J., and Lançon, R.: Temperature Characteristics of High Frequency Lithium Tantalate Plates. Proc. 30th Freq. Control Symp., 1976, pp. 132-140.
- 14) Vale, C.R., and Weinert, R.W.: Ladder Filter using Temperature Stable LiTaO₃ Resonators in Strip Form. Proc. IEEE Ultrasonic Symp. 1986, pp. 103-107.
- 15) Wakatsuki, N., Fujiwara, Y., and Ono, M.: Chip Resonators. (in Japanese), *FUJITSU*, **34**, 7, pp. 1179-1186 (1983).
- 16) Yamamoto, S., Gounji, T., and Shimizu, J.: Wide-band Timing Tank Filters for Digital Transmission Systems. 40th Annual Freq. Cont. Symp., 1986, pp. 275-284.
- 17) Carruthers, J.R., Peterson, G.E., and Grasso, M.: Nonstoichiometry and Crystal Growth of Lithium Niobate. *J. Appl. Phys.*, **42**, 5, pp. 1846-1851 (1971).
- 18) Nassau, K., Levinstein, H.J., and Loiacono, G.M.: Ferroelectric Lithium Niobate. 1. Growth, Domain Structure, Dislocations and Etching. *J. Phys. & Chem. Solids*, **27**, pp. 983-988 (1966).
- 19) Nakamura, K., and Shimizu, H.: On the Equivalent Circuit for Thickness Twist Modes and Thickness Shear Modes in Piezoelectric Plates. (in Japanese), *Paper IECE, Jpn.*, **US-69**- 21, pp. 1-17 (1969).
- 20) Warner, A.W., Onoe, M., and Coquin, G.A.: Determination of Elastic and Piezoelectric Constants for Crystals in Class (3 m). *J. Acoust. Soc. Am.*, **42**, 6, pp. 1223-1231 (1967).
- 21) Onoe, M.: General Equivalent Circuit of a Piezoelectric Transducer Vibrating in Thickness Modes. (in Japanese), *Trans IECE, Jpn.*, **55-A**, 5, pp. 239-244 (1972).
- 22) Smith, R.T., and Welsh, F.S.: Temperature Dependence of the Elastic, Piezoelectric Constants of Lithium Tantalate and Lithium Niobate. *J. Appl. Phys.*, **42**, 6, pp. 2219-2230 (1971).
- 23) Onoe, M., and Jumonji, H.: Analysis of Piezoelectric Resonators Vibrating in Trapped Energy Modes. (in Japanese), *J. IECE, Jpn.*, **48**, 9, pp. 1574-1581 (1965).
- 24) Fujiwara, Y., Yamada, S., and Wakatsuki, N.: Strip Type LiTaO₃ Resonator with Zero Temperature Coefficient. (in Japanese), *Paper IECE, Jpn.*, **US-82**-46, pp. 1-6 (1982).
- 25) Fujiwara, Y., and Wakatsuki, N.: LiTaO₃ and LiNbO₃ Strip Type Resonators. *IEEE Trans. Ultrasonic, Ferroelectric, Frequency Control*, **UFFC-34**, 1, pp. 39-44 (1987).
- 26) Fujiwara, Y., Inui, J., Gounji, T., and Wakatsuki, N.: Fabrication and Applications of Chip-Type Resonators. Proc. IEEE CHMT Symp., 1984, pp. 104-109.
- 27) Ono, M., Fujiwara, Y., and Wakatsuki, N.: Surface Shear Wave Resonators for VHF Range Utilizing Reflection of Edge Surfaces. ULTRASONIC SYMP., 1984, pp. 218-223.
- 28) Shimizu, N., Fujiwara, Y., and Wakatsuki, N.: Examination of Oscillator Circuit Conditions.

ULTRASONIC SYMP., 1984, pp. 400-404.

- 29) Urabe, S.: Consideration of SAW Resonator Characteristics for Voltage Controlled Oscillator. (in Japanese), *Trans. IECE, Jpn.*, **J66-C**, 1, pp. 110-111 (1983).



Hideki Tominaga

Electronic Components Marketing Dept.
FUJITSU LIMITED
Bachelor of Physics
Hokkaido University 1964
Specializing in Piezoelectric Devices
and Piezoelectric Single Crystals



Yoshiro Fujiwara

Solid State Devices Division
Process Engineering Dept.
FUJITSU LIMITED
Bachelor of Electronics Eng.
Kobe University 1976
Specializing in Piezoelectric Devices



Masaaki Ono

Solid State Devices Division
Circuit Devices & Bubble Memory Dept.
FUJITSU LIMITED
Bachelor of Electronic Eng.
Yamagata University 1971
Master of Electronic Eng.
Yamagata University 1973
Specializing in Piezoelectric Devices

An Expert System Based on Qualitative Knowledge

• Hiroyuki Yoshida • Hiroyuki Izumi • Ryoichi Narita

(Manuscript received December 1, 1987)

This paper describes the implementation of an expert system that is based on deep knowledge. This system uses a qualitative model of the target object to predict the behavior of it through qualitative simulation, diagnose multiple faults in the object, and create a decision tree for diagnosing typical symptoms. Qualitative modeling has been extended for practical target objects in which discontinuous state transitions occur, and the system uses a new simulation method based on extended models. Modal operators are used to represent time-varying symptoms for diagnosing multiple faults. The system also uses an algorithm to create a decision tree from a deep knowledge base.

1. Introduction

The next generation of expert systems will use deep knowledge bases to understand the principles of target objects, and so solve problems using these knowledge bases. Several theoretical methods of using deep knowledge are now being studied. *Qualitative Reasoning* is a technique used to predict and explain the motions of an object. It uses qualitatively abstracted (or simplified) physical equations as deep knowledge of the target object to produce sequences of qualitative state transitions of the target object. De Kleer and Brown¹⁾ and Kuipers²⁾ developed methods for qualitative modeling and simulation of objects in which only continuous state transitions occur. Nishida³⁾ recently assumed that discontinuous transition is a chain of microscopic continuous transitions, and proposed a method by which such a chain can be determined using heuristics. The *General Diagnostic Engine* is a general framework for solving problems using deep knowledge. De Kleer and Williams⁴⁾ showed how to apply this framework for diagnosing digital circuits that have no time-varying properties.

We have expanded on the previously

mentioned methods for more practical objects; such as automobile engines, and have systematized and incorporated them into a deep knowledge-based expert system. Figure 1 shows the configuration of our system. It uses a qualitative model of the target object to predict the behavior of the object by using a qualitative simulator, and to diagnose multiple faults in the object. This system also creates a decision

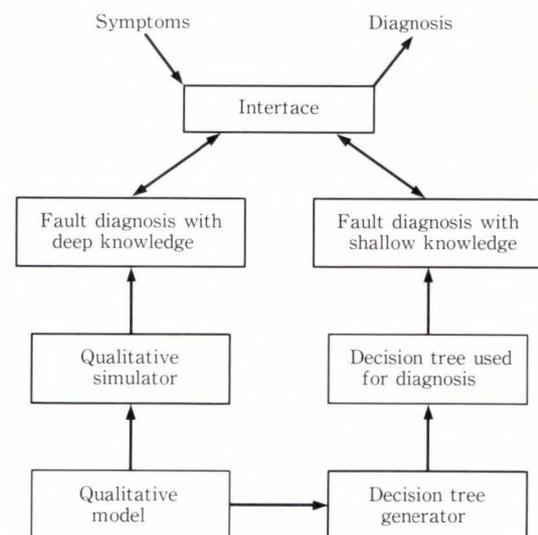


Fig. 1—System configuration.

tree for quickly diagnosing typical symptoms. To represent practical objects, we have proposed a modeling method to cope with discontinuity and have showed how to analyze discontinuity without using heuristics. To apply the general diagnostic engine to our model, we have introduced modal operators to represent time-varying symptoms. Because diagnosis based on deep knowledge is a time-consuming task, however, it is essential that shallow or more efficient knowledge be obtained from the task. Therefore, we have used an algorithm to create a decision tree (a type of shallow knowledge) from a deep knowledge base.

The following chapters describe the four major features of our system: extensions of qualitative modeling for practical target objects, simulation method for extended models, method of diagnosis based on deep knowledge, and algorithm used to create a decision tree.

2. Extensions of the qualitative model

2.1 Primitives for specifying constraints

In general, a qualitative model of a target object is a set of constraints. Kuipers proposed six operators for specifying qualitative constraints²⁾; ADD (f, g, h), MULT (f, g, h), MINUS (f, g), DERIV (f, g), M^+ (f, g), and M^- (f, g), which specify arithmetic, derivative, and monotonic relations between parameters. Consider the valve shown in Fig. 2.

The conventional physical model of this valve is as follows:

$$Q = CA \sqrt{\frac{2P}{\rho}}, P > 0, \dots (1)$$

where Q is the flow rate through the valve, C is the discharge coefficient of the valve, A is the

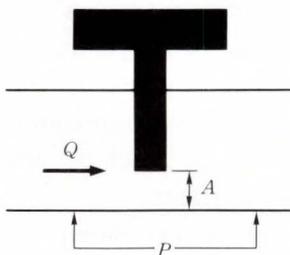


Fig. 2—Valve.

area available for flow, P is the pressure across the valve, and ρ is the mass density of the fluid.

By using Kuipers' operators, a qualitative model of the valve can be written as follows:

$$M^+(A, f_1), M^+(P, f_2), \text{MULT}(f_1, f_2, Q), \dots (2)$$

To construct Kuipers' type of qualitative model, the quantitative model of valve (1) would be necessary.

De Kleer and Brown have showed that the qualitative constraints of a valve are written as a *confluence*¹⁾:

$$\partial P + \partial A - \partial Q = 0. \dots (3)$$

They also described how confluence (3) can be derived from quantitative model (1).

When considering an object that includes a component valve, however, we only have to know that Q increases when A or P increases, without understanding the exact function (1). When considering partial derivatives, we can easily obtain:

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial A} \times \frac{dA}{dt} + \frac{\partial Q}{\partial P} \times \frac{dP}{dt}. \dots (4)$$

To obtain (3) from (4), we simply have to know the signs positive or negative (plus or minus) of the partial derivatives ($\frac{\partial Q}{\partial A}$ and $\frac{\partial Q}{\partial P}$) that our common sense tell us.

In our formalism, the sets of signs of partial derivatives and the operators proposed by Kuipers can be used as primitives to specify constraints.

2.2 Knowledge about definite integrals

In Kuiper's formalization, one can specify more detailed information on constraints, such as *correspondences* between values. For example, $M_0^+(X, Y)$ means that a monotonic increasing relation exists between X and Y , and that X always becomes 0 when Y becomes 0.

However, there is no existing qualitative modeling that can handle a definite integral, which can be considered a correspondence between values constrained by a derivative

relation. For example, after fully opening a water valve above a tank, and then returning some time later when the tank should be full, one may find that the tank is not yet full. One may assume that the flow of water was less than usual for some reason.

To extend the formalism used to describe this type of knowledge, a method is required to specify the time. We assumed that the model (or target object) has a parameter that serves as a "clock". Knowledge about definite integrals can then be specified by using the parameter. An example is given in Equation (5), which specifies the correspondence between increasing rate f_0 and net increase of fuel q_0 in the cylinders of an engine.

$$\int_{c=0}^{c=\pi} f_0 dt = q_0 . \quad \dots \dots \dots (5)$$

In this case, the angle of crankshaft C serves as a clock. From this, our simulator infers $\int_{c=0}^{c=\pi} F dt \geq q_0$ when the increasing rate of fuel $F \geq f_0$ is maintained during $0 < C < \pi$.

2.3 State transition rules

When we deal with practical objects, we usually encounter discontinuous changes in parameters. One such discontinuity arises when some state transitions occur during a much shorter time than others. When considering the motion of an engine, for example, one can normally assume that the ignition in a cylinder is completed instantaneously. Although one could describe this process as a sequence of continuous state transitions from a microscopic point of view, commonsense requires that the ignition be considered a discontinuous transition of states from the macroscopic point of view.

To handle these discontinuous transitions, we have introduced state transition rules as a form by which to describe them.

A state transition rule consists of a trigger and the result of the discontinuous transition:

if *trigger*, then *result*. $\dots \dots \dots (6)$

When the object enters the state that satisfies the trigger, the state is assumed to

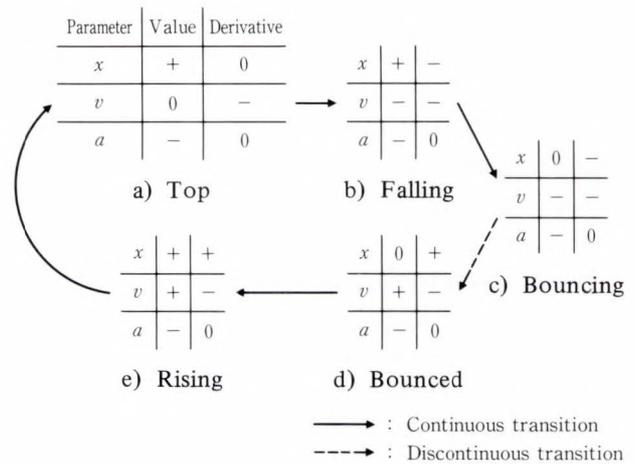


Fig. 3—Behavior of a ball.

discontinuously change into the state specified by the result.

Consider a ball bouncing on a floor. Let x be the position, v the velocity, and a the acceleration of the ball. Because the bounce is much quicker than the time during which the ball is in the air, we can define the following state transition rule:

if $x = 0 \wedge v < 0$, then $v > 0$. $\dots \dots (7)$

Figure 3 shows the behavior of the ball.

2.4 Switches

Another type of discontinuity arises when an object includes a switch. A switch is a component that has several modes with different constraints placed on the parameters in different modes. When a fuel tank is not empty, for example, the flow through a fuel pipe depends on the power of a fuel pump. But when the tank is empty, the flow is always 0. In this case, the tank can be considered a switch that has two modes: EMPTY and NOT-EMPTY. We can also assume that the mode of the switch is dominated by one of its parameters. Thus, each mode is defined by using the dominating parameter. A parameter constraint can also have a condition, that is, a set of switching modes in which it has an effect.

Consider the buzzer shown in Fig. 4. Let x be the position, v the velocity, and a the acceleration of the clapper. Clapper switch X has two

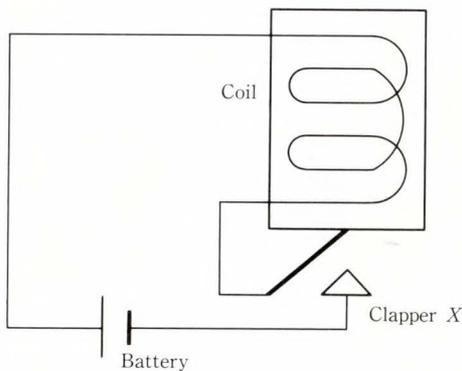


Fig. 4—Buzzer.

modes (CLOSED and OPEN) that are dominated by x .

$$X: \text{ OPEN} \equiv x > 0, \quad \dots \dots \dots (8)$$

$$X: \text{ CLOSED} \equiv x = 0, \quad \dots \dots \dots (9)$$

The five constraints and a transition rule are as follows:

$$x \geq 0, \quad \dots \dots \dots (10)$$

$$\partial x = v, \quad \dots \dots \dots (11)$$

$$\partial v = a, \quad \dots \dots \dots (12)$$

$$a > 0 \text{ when } X: \text{ CLOSED}, \quad \dots \dots \dots (13)$$

$$a < 0 \text{ when } X: \text{ OPEN}, \quad \dots \dots \dots (14)$$

$$\text{if } x = 0 \wedge v < 0, \text{ then } v = 0. \quad \dots \dots (15)$$

Figure 5 shows the behavior of the buzzer.

3. Qualitative simulation

3.1 Intra-state analysis

Qualitative reasoning includes intra-state analysis and inter-state analysis. Kuipers has devised an efficient algorithm to create state transitions beginning from a given initial state without using intra-state analysis²⁾. Intra-state analysis is not needed because the next possible states are generated according to the continuity of parameters.

While the simulator in our system is based on Kuipers' algorithm, the next values or derivatives of some parameters cannot always be determined because of discontinuities. Therefore, in case of discontinuous changes, intra-state analysis is required and the next values are determined by the propagation of parameter constraints.

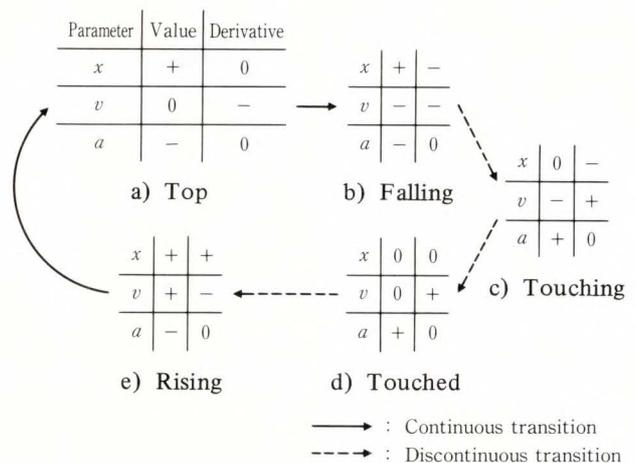


Fig. 5—Behavior of a buzzer.

3.2 Inter-state analysis

Inter-state analysis determines the next states of a given state. We give a non-heuristic definition of discontinuous state transitions.

3.2.1 Assumptions

- 1) We can assume that each state transition rule is to be applied whenever its trigger is satisfied by the current state of the object. Therefore, the next states of such a current state are the only ones specified by the result part of the rule. We also assume that such discontinuous transition occurs without any time elapse.
- 2) We assume that switching occurs only when continuous transition of the dominating parameter occurs or when a state transition rule forces switching. Because the continuous transition of another parameter may occur before that of the dominating parameter, the next states are not necessarily those caused by switching. This assumption means that switching cannot be caused by other switching.

3.2.2 Extent of influence of discontinuous transitions

A state is a set of values and derivatives of parameters. For subset s of state S and a set of constraints R , any subset of S is called a *seed* if and only if it includes s and is consistent with R . A *maximal seed* is a seed that has no proper super-set seed.

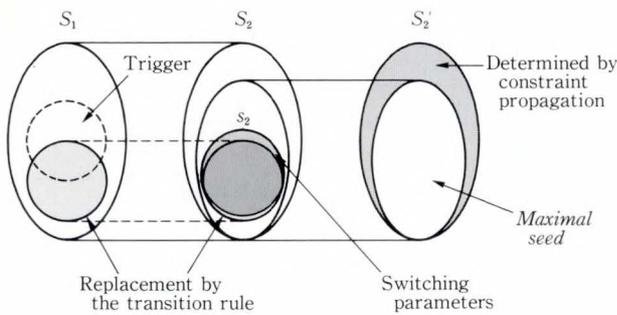


Fig. 6—Discontinuous transition using the state transition rule.

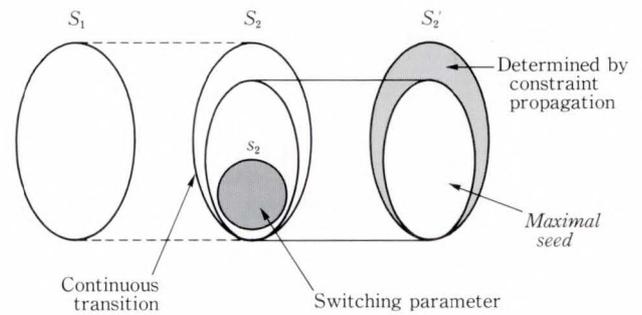


Fig. 7—Discontinuous transition by switching.

1) Discontinuous transition using state transition rules

Let current state S_1 satisfy a state transition rule. S_2 is the next possible state obtained by replacing values and/or derivatives of several parameters in S_1 according to the result part of the rule. Let R_2 be a set of constraints that have effects (when their conditions are satisfied) on S_2 . Note that S_2 may be inconsistent with R_2 . Let s_2 be a subset of S_2 which includes the part replaced by the rule and all switch-dominating parameters. Next state S_2' is one which is consistent with R_2 and includes the maximal seed for s_2 and R_2 as a subset. Figure 6 shows this situation. Consider the example of the bouncing ball given in section 2.3. Let S_1 be state (c) in Fig. 3.

$$S_1 = \{x = 0, \partial x < 0, v < 0, \partial v < 0, a < 0, \partial a = 0\}. \dots\dots\dots (16)$$

Because this satisfies state transition rule (7),

$$S_2 = \{x = 0, \partial x < 0, v > 0, \partial v < 0, a < 0, \partial a = 0\}, \dots\dots\dots (17)$$

$$s_2 = \{v > 0\}. \dots\dots\dots (18)$$

Because there is no conditioned constraint, R_2 includes all constraints.

$$R_2 = \{x \geq 0, \partial x = v, \partial v = a, \partial a = 0\}. \dots\dots\dots (19)$$

Thus, the maximal seed for s_2 and R_2 is as follows:

$$\{x = 0, v > 0, \partial v < 0, a < 0, \partial a = 0\}. \dots (20)$$

The next state is

$$S_2' = \{x = 0, \partial x > 0, v > 0, \partial v < 0, a < 0, \partial a = 0\}. \dots\dots\dots (21)$$

2) Discontinuous transition by switching

Let current state S_1 satisfy no state transition rule, and let R_1 be a set of constraints that have effects on this state. The next possible states are constructed by combining the results of continuous transitions of parameters from S_1 . If a next possible state includes no switched parameters and is consistent with R_1 , then it is a next state by continuous transition. Let S_2 be a next possible state that includes switched parameters, and let R_2 be a set of constraints that have effects on S_2 . Note that S_2 may be inconsistent with R_2 . Let s_2 be a subset of S_2 , which includes all switch-dominating parameters. Next state S_2' is one which is consistent with R_2 and includes the maximal seed for s_2 and R_2 as a subset. Figure 7 shows this situation. Consider the example of the buzzer in given section 2.4. Let S_1 be state (d) in Fig. 5.

$$S_1 = \{x = 0, \partial x = 0, v = 0, \partial v > 0, a > 0, \partial a = 0\}. \dots\dots\dots (22)$$

Parameter x can change continuously to $\{x > 0, \partial x > 0\}$, $\{x = 0, \partial x = 0\}$, or $\{x < 0, \partial x < 0\}$. v can only change continuously to $\{v > 0, \partial v > 0\}$ while a can only change continuously to $\{a > 0, \partial a = 0\}$.

By combining these, the next possible states are:

$$S_{2_1} = \{x > 0, \partial x > 0, v > 0, \partial v > 0, a > 0, \partial a = 0\}, \dots\dots\dots (23)$$

$$S_{2_2} = \{x = 0, \partial x = 0, v > 0, \partial v > 0, a > 0, \partial a = 0\}, \dots\dots\dots (24)$$

$$S_{2_3} = \{x < 0, \partial x < 0, v > 0, \partial v > 0, a > 0, \partial a = 0\}. \dots\dots\dots (25)$$

Here, both S_{2_2} and S_{2_3} include no switching. However, neither is a next state because they are inconsistent with constraint $\partial x = v$ of Equation (11). In S_{2_1} , switch-dominating parameter x changes from 0 to positive (or from CLOSED to OPEN). Then

$$s_2 = \{x > 0\}, \dots\dots\dots (26)$$

$$R_2 = \{(10), (11), (12), (14)\} \dots\dots (27)$$

The maximal seed for s_2 and R_2 is as follows:

$$\{x > 0, \partial x > 0, v > 0, \partial a = 0\} \dots\dots (28)$$

The next state is

$$S_{2'} = \{x > 0, \partial x > 0, v > 0, \partial v < 0, a < 0, \partial a = 0\} \dots\dots\dots (29)$$

3.3 Definite integrals

Now we'll describe how to use knowledge about definite integrals. Let parameter F be the derivative of parameter Q , and let f_0 be a typical value of F specified by:

$$\int_{C=c_2}^{C=c_1} f_0 dt = q_0. \dots\dots\dots (30)$$

Each state must be checked by the following algorithm.

- 1) If the value of C in the state is not c_1 , then stop with answer OK.
- 2) Let the value of Q be q_1 .
- 3) Tracing back through the chain of states, find a state where the value of C is c_2 , and let it be S_2 . If no such state exists, then stop with answer OK.
- 4) Let the value of Q at S_2 be q_2 .
- 5) Check the value of F in all states from that following S_2 to the current one. (In these states $c_2 < C < c_1$ holds necessarily.)
 - i) If $F < f_0$ holds in these states, $q_2 + q_0 < q_1$ must be qualitatively satisfied. Return with an answer regarding whether the condition is satisfied.
 - ii) If $F = f_0$ holds in these states, $q_2 + q_0 = q_1$ must be qualitatively satisfied. Return with an answer regarding whether the condition is satisfied.
 - iii) If $F > f_0$ holds in these states, $q_2 + q_0 > q_1$ must be qualitatively satisfied. Return with an answer regarding

whether the condition is satisfied.

- iv) In other cases, stop with answer OK.

4. Diagnosis based on deep knowledge

De Kleer and Williams proposed a general schema for model-based diagnosis called the general diagnostic engine (GDE)⁴. We have summarized it as follows:

A model M is a set of component descriptions.

A conflict X for symptom s is a subset of M , which is inconsistent with s . This means that behavior s' is derived from the elements of X , and s' is inconsistent with s . A conflict is *minimal* when it has no proper subset that is also a conflict. Because a super-set of a conflict is also a conflict, we need only consider minimal conflicts. For such minimal conflicts, at least one of its elements fails to hold true; the corresponding component is out of order. For example, when a set of component descriptions $\{C_1, C_2, C_3\}$ is the minimal conflict for symptom s , it logically means:

$$C_1 \wedge C_2 \wedge C_3 \rightarrow \neg s. \dots\dots\dots (31)$$

Reversing this implication,

$$s \rightarrow \neg C_1 \vee \neg C_2 \vee \neg C_3. \dots\dots\dots (32)$$

Thus, if we find symptom s , we can conclude that at least one of the components $\{C_1, C_2, \text{ or } C_3\}$ is out of order.

A subset of M is a *candidate* for a set of symptoms S if it shares at least one element with every conflict for each symptom s in S . A candidate is *minimal* if it has no proper subset that is also a candidate. Because a super-set of a candidate is also a candidate, we need only consider minimal candidates. The goal of diagnosis is to identify a set of candidates for all noted symptoms. For example, when $\{C_1, C_2, C_3\}$ is the minimal conflict for s_1 and $\{C_1, C_4\}$ is that for s_2 , it logically means:

$$s_1 \rightarrow \neg C_1 \vee \neg C_2 \vee \neg C_3, \dots\dots\dots (33)$$

$$s_2 \rightarrow \neg C_1 \vee \neg C_4. \dots\dots\dots (34)$$

Thus, we can conclude that:

$$s_1 \wedge s_2 \rightarrow \neg C_1 \vee (\neg C_2 \wedge \neg C_4) \vee (\neg C_3 \wedge \neg C_4). \dots\dots\dots (35)$$

In this case, the minimal candidates for $\{s_1, s_2\}$ are $\{C_1\}$, $\{C_2, C_4\}$, and $\{C_3, C_4\}$. If we assume that there is only one fault, we conclude that component C_1 is out of order.

The following is an incremental diagnostic algorithm.

- 1) Let the current minimal candidate set Z be ϕ .
- 2) Obtain newly observed symptom s . If no such symptom exists, then stop.
- 3) Generate the minimal conflicts for s .
- 4) Compute the minimal candidate set using Z and the minimal conflicts for s , (let it be Z).
- 5) Return to step 2.

De Kleer and Williams only described the troubleshooting of digital circuits without inter-state transitions. Here, we will describe how to apply the general diagnostic engine to target objects that cause inter-state transitions. For this purpose, the terms "component description", "symptom", and "be inconsistent with" must be defined.

According to our formalism, constraints (conditional or unconditional) are naturally regarded as component descriptions. When the diagnostic algorithm concludes that a constraint between battery and starter voltages is a final candidate, for example, it means that the cable between the battery and starter may be disconnected, or the cable contacts may be bad.

Moreover, a state transition rule is also a component description. When a rule on engine ignition is a candidate, it means, for example, that the spark plugs may be sub-standard because they are too dirty or have incorrectly adjusted gaps.

Considering that the lack of fuel is a possible cause of an engine fault, any description about the initial state is also a component description.

For objects that cause inter-state transitions, symptoms are usually observed over a time progression. Therefore, we have introduced modal operator \square to represent "forever" and \diamond to represent "sometimes" to describe the symptoms. For example, $\square (R = 0)$ means that rotation speed R of the starter is always 0 (i.e. the starter never

works). $\diamond (F = 0)$ means that fuel quantity F of the tank eventually becomes 0.

The qualitative simulation result is a transition tree of states whose root is the initial state. Here, when we say a sequence, we mean a sequence of state transitions from the initial state in the tree. A sequence represents one possible behavior of the target object. Therefore, the set of component descriptions is defined to be inconsistent with a given symptom when there is no sequence representing the behavior of the symptom in the transition tree created by simulating the set. By using the above example, when the transition tree has no sequence that includes a state where $F = 0$, the set of component descriptions is inconsistent with symptom $\diamond (F = 0)$.

5. Diagnostic tree

A decision tree is one of the most compact and efficient forms of shallow knowledge, although it is less adaptable. This chapter describes how to make a decision tree for diagnosing typical symptoms from deep knowledge.

A decision tree consists of nodes that are quadruple in form: *check*, *candidate*, *positive-child*, and *negative-child*. A check is a typical symptom. A candidate is a set of component descriptions. The positive and negative children are other nodes or end markers. A leaf is a node for which positive and negative children are end markers. A root is a node that is not the child of another node. There is only one root in a decision tree.

Here we assume that:

- 1) there is only one fault in the target object, and
- 2) when a component is out of order, the symptoms whose minimal conflict includes the component description are necessarily observed. This means that the component is not out of order when the symptom is not observed.

Under these assumptions, the following algorithm forms a decision tree that never asks unnecessary questions.

- 1) If there is more than one symptom whose minimal conflicts are the same, then discard all but one of these symptoms.
- 2) If there is a symptom that has more than one minimal conflict, then regard the intersection of those conflicts as the minimal conflict. The validity of this operation is supported by the above two assumptions.
- 3) Create a root node whose candidate is the union of the minimal conflicts of all symptoms. Initialize a set of nodes E to only include the root node.
- 4) Select a node from E and let it be N . If E is empty, then stop.
- 5) Select symptom s that satisfies the following conditions and let it be the check for N .

Conditions: s is not the check for any node in the path from the root to N , and the candidate of N has a non-empty intersection with the minimal conflict of s .

If there is no such symptom, let the positive and negative children of N be end markers, and return to step 4.

- 6) Create a node whose candidate is the intersection of the candidate of N and the minimal conflict of s . Let the node be the positive-child of N and add it to E .
- 7) Create a node whose candidate is the set of remaining elements of N . Let the node be the negative-child of N and add it to E .
- 8) Return to step 4.

Using this decision tree for diagnosis is explained as follows:

- 9) Let the root be the current node.
- 10) If the current node is a leaf, then output its candidate and stop.
- 11) Ask the observer whether the check of the current node can be observed.
- 12) If the answer is yes, then let the positive child be the current node and return to step 10.
- 13) Otherwise, let the negative-child be the current node and return to step 10.

Consider a case in which the typical symptoms are $\{s_1, s_2, s_3\}$, and the minimal conflicts of those are $\{a, b, c\}$, $\{a\}$, and $\{b, d\}$

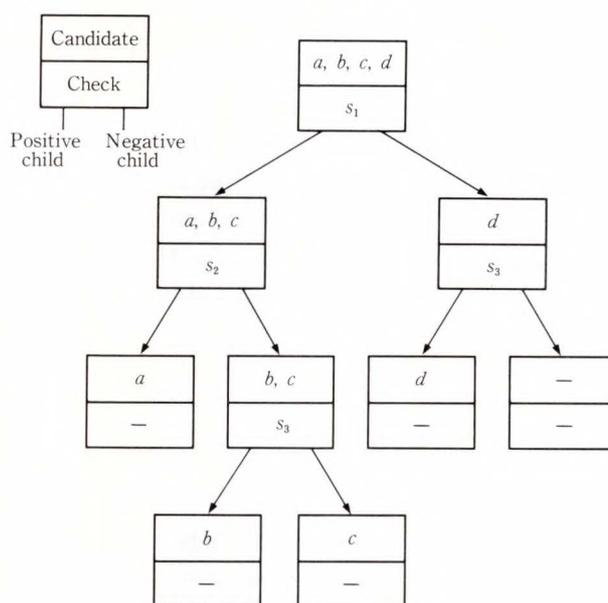


Fig. 8—A decision tree.

respectively. Figure 8 shows the decision tree of this case.

6. Conclusion

This paper has presented our extensions to existing qualitative reasoning techniques and the general diagnostic engine.

We introduced four extensions to the formalism of qualitative models to use knowledge about definite integrals and monotonic constraints, and to cope with discontinuous state transitions.

We showed how the qualitative simulator copes with those extensions, the most important technique being the introduced concept of a maximal seed to formalize the minimal influence of discontinuous transitions. We also presented a non-heuristic method to determine the next states after discontinuous transitions. We also applied the general diagnostic engine to objects that cause inter-state transitions. Finally, we described how to create a decision tree from deep knowledge. While the algorithm results in a maximum tree depth of $\log_2 n$ (where n is the total number of component descriptions), further analysis using information theory is required for selecting the symptoms for each node to minimize the average costs of obser-

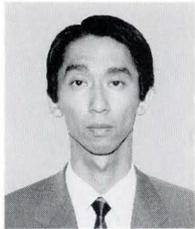
vation.

7. Acknowledgement

The authors would like to express our appreciation to Messrs. Y. Iijima and H. Hara for their helpful suggestions and comments. The authors would also like to thank Mr. H. Hayashi for his valuable assistance and encouragement during this project.

References

1. De Kleer, J., and Brown, J.S.: A Qualitative Physics Based on Confluences. *Artificial Intelligence*, **24**, pp. 7-83 (1984).
2. Kuipers, B.: Qualitative Simulation. *Artificial Intelligence*, **29**, pp. 289-338 (1986).
3. Nishida, T., and Doushita, S.: Reasoning about Discontinuous Change. Proc. Sixth Nat. Conf. Artificial Intelligence, Seattle, WA, 1987, pp. 643-648.
4. De Kleer, J., and Williams, B.C.: Diagnosing Multiple Faults. *Artificial Intelligence*, **32**, pp. 97-130 (1987).



Hiroyuki Yoshida

Artificial Intelligence Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Information Sciences
Tokyo Institute of Technology 1980
Master of System Sciences
Tokyo Institute of Technology 1982
Specializing in Artificial Intelligence



Ryoichi Narita

Artificial Intelligence Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Liberal Arts
The University of Tokyo 1979
Master of Science
The University of Tokyo 1981
Specializing in Artificial Intelligence



Hiroyuki Izumi

Artificial Intelligence Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Electro-Communications
The University of Electro-Communications
1978
Dr. of Information Eng.
Nagoya University 1983
Specializing in Artificial Intelligence

A Design Method of Reflector Antennas with Arbitrary Configuration

• Yoshihiko Asano • Hiroshi Kurihara

(Manuscript received December 1, 1987)

A computer program is developed for designing reflector antennas with arbitrary configurations for use in arbitrary alignments between primary feeds and reflectors of arbitrary shapes.

Recent progress in satellite communications has increased the importance of efficient use of the geostationary orbit together with importance of efficient use of the radio spectrum. Highly efficient use of the geostationary orbit is ensured by giving an accurate design of a low sidelobe antenna which requires rigorous estimation of antenna radiation characteristics. Since newly developed computer program calculates radiation characteristics for arbitrary antenna configurations, the best configuration to minimize the sidelobe can be chosen from the variety of antenna configurations. The estimated radiation characteristics have been experimentally confirmed.

1. Introduction

Due to the current increase of information transmitted via satellite, the efficient use of resources such as the geostationary orbit and the radio spectrum has become more important. The recent trend toward larger satellites is favorable for both frequency reuse and efficient use of the geostationary orbit because the satellite can be equipped with a multibeam antenna with a large reflector^{1), 2)}. Since several primary feeds illuminate a large reflector from different positions, the antenna radiates multiple beams which propagate toward different directions. Frequency reuse is possible for beams which illuminate different terrestrial areas. In order to reduce interference between beams, it is necessary to estimate the antenna radiation precisely.

small diameters in earth stations was enabled by enlarged on-board antennas together with the improved performance of on-board transponders³⁾. Convenience in installation of small antennas reduces cost for installing the earth station and reduces the cost of the total system.

However, the broader beam of small antennas increases interference between the communication paths. Observing efficient use of the radio spectrum and the geostationary orbit, FCC specifies strict requirements on the radiation patterns of earth stations⁴⁾. Correspondingly, CCIR recommends similar characteristics for the radiation patterns⁵⁾. The strict requirements for the radiation patterns makes it more important to calculate the near-axis radiation characteristics precisely, taking misalignment between reflectors and primary feeds into account.

In view of this, a computer program was developed for analyzing antenna radiation characteristics quantitatively. To ensure accuracy, the current-distribution method and the aperture-field method were used to analyze reflectors and horn antennas respectively, without applying the far-field approximation.

The program can be widely applied to calculate radiation characteristics of various reflector antennas including on-board multi-beam antennas and small antennas for earth

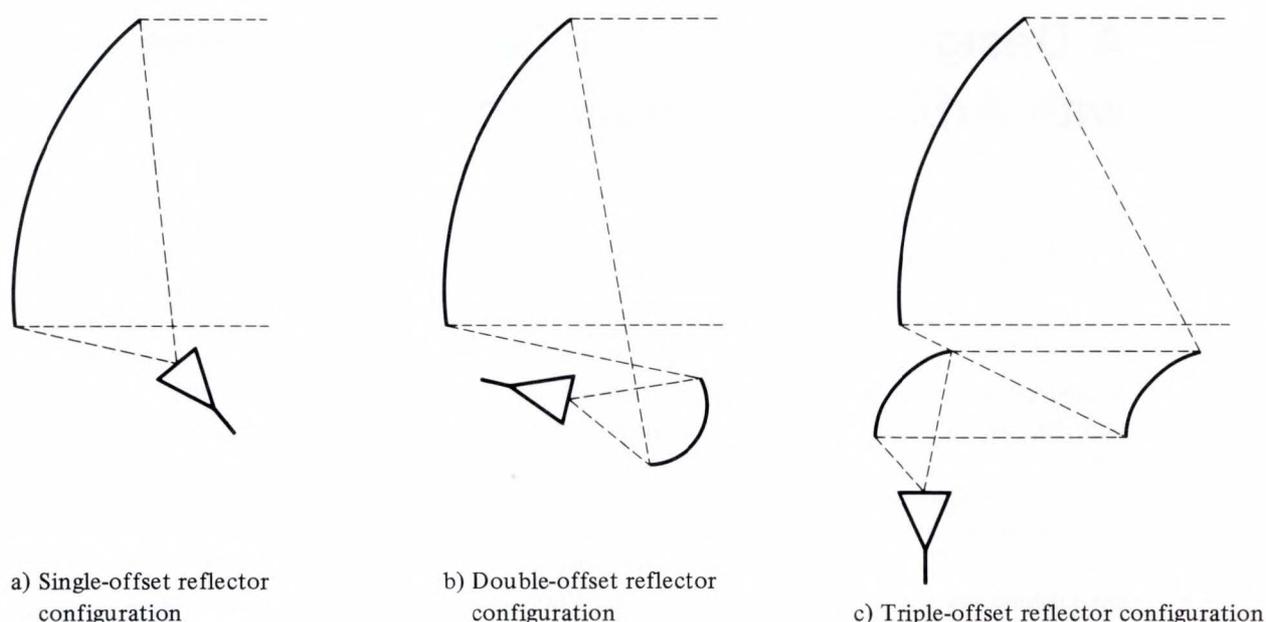


Fig. 1—Reflector antenna geometry.

stations. Since geometrically asymmetric antennas such as offset antennas with parabolic reflectors or offset Gregorian antennas are frequently used for earth stations, the application to arbitrary antenna configuration including geometrical asymmetry was considered in the development of the program.

First, a background explanation will be given for configuration of reflector antennas followed by a method for estimating antenna radiation characteristics. The various functions of the developed computer program will then be explained in detail. The calculated results for radiation characteristics will be shown, including misalignment between reflectors and primary feeds, and finally, the estimated results verified by the experiment will be shown.

2. Configuration of reflector antennas

Figure 1 shows a cross section of an offset antenna configuration using one to three reflectors. The offset reflector antenna reduces aperture-blocking effects due to a primary feed, subreflector, or their supporting struts and its sidelobe characteristics are excellent compared with the axis-symmetric antenna. Furthermore, the sidelobe characteristics of the offset type in

the wide-angle region can be easily improved by installing absorbers around the antenna.

The electromagnetic wave radiated from the primary feed, after being reflected by one or more reflectors, are radiated forward from the antenna.

The shape of the primary feed differs depending on the frequency band in use. In Ku-band for example, conical or pyramidal horn antennas have been widely put in practical use. Various shapes of reflectors, such as paraboloid, ellipsoid, and hyperboloid, are used.

When two or more reflectors are used, the antenna characteristics, including antenna efficiency and sidelobe characteristics, can be improved by slightly deforming the reflector from being paraboloid, ellipsoid, etc., to match the radiation characteristics of the primary feed⁶⁾. Various methods of shaping offset reflectors are investigated^{7), 8)}.

3. Analytical method

3.1 Radiation analysis

The aperture-field method⁹⁾, the current-distribution method¹⁰⁾ and geometrical theory of diffraction (GTD)¹¹⁾ are well known analytical methods for aperture antennas. The

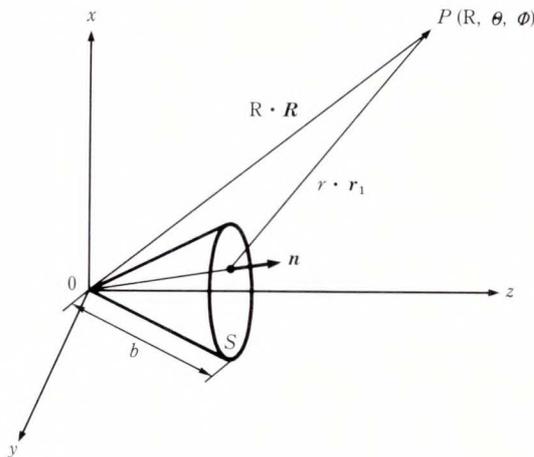


Fig. 2—Coordinate system for conical horn.

radiation characteristics of horn antennas are accurately calculated using the aperture-field method when proper aperture field is assumed as a wave source. The current-distribution method enables estimation of the near-axis radiation characteristics of reflector antennas including the cross-polarization components generated at the reflector surface.

The aperture-field method was used for the analysis of primary-feed radiation characteristics while the current-distribution method was used for the reflector analysis.

3.1.1 Aperture-field method

This method assumes an antenna aperture electromagnetic field using the geometrical optics to obtain the radiation characteristics. When applied to the horn antennas, the following assumptions are made¹²⁾:

- 1) The field distribution at the horn aperture completely coincides with that at the same cross section of an infinite horn.
- 2) The field does not spill around the rim of horn's mouth.

These assumption are justified if the flare angle of the horn is not too large and if the horn is not too short in terms of wavelength. The aperture is taken to be spherical over the mouth of the horn. Figure 2 shows the coordinate system.

In this case, the electric and magnetic fields are expressed by the following equations:

$$E_p = \frac{-jk}{4\pi} \int_s r_1 \times [n \times E_a - \sqrt{\frac{\mu}{\epsilon}} r_1 \times (n \times H_a)] \frac{e^{-jkr}}{r} dS, \dots\dots (1)$$

$$H_p = \frac{-jk}{4\pi} \int_s r_1 \times [n \times H_a + \sqrt{\frac{\epsilon}{\mu}} r_1 \times (n \times E_a)] \frac{e^{-jkr}}{r} dS, \dots\dots (2)$$

E_p, H_p : Electric and magnetic field at P
 E_a, H_a : Electric and magnetic field over the aperture S .

- n : Unit vector normal to S
- r : Distance from source point on S to the field point P .
- r_1 : Unit vector along r
- S : Horn aperture – Spherical surface with radius b
- k : Propagation constant in free space
- ϵ : Permittivity
- μ : Permeability.

In the present analysis, parameters such as the shape of in-wall of the corrugated conical horn, and the mode ratio of the multimode horn, are specified to calculate the electromagnetic field distribution at the horn aperture and the horn radiation characteristics using the above equations.

3.1.2 Current-distribution method

When the electromagnetic wave is incident to a perfectly conducting surface, a surface current flows on the surface proportional to the strength of the magnetic field. Assuming that the electromagnetic wave is reflected in the same way as the plane wave when it is locally incident to the infinite plane, the surface current K is expressed by the following equation, where the incident magnetic field is H_i :

$$K = 2(n \times H_i). \dots\dots (3)$$

Considering the case when the electromagnetic wave is incident on reflector, the area illuminated is the range S surrounded by curve Γ . Figure 3 shows the coordinate system. For the equation of continuity to be satisfied at the

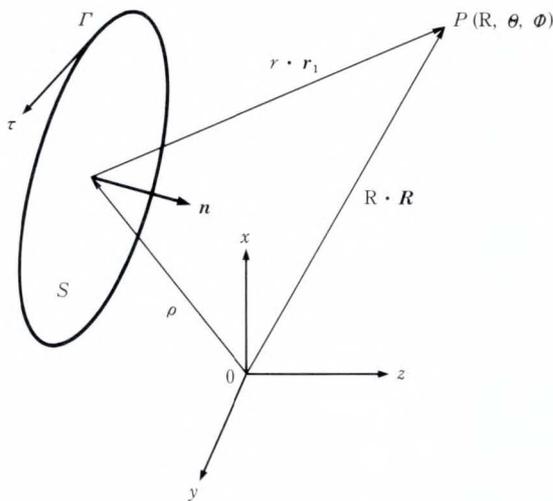


Fig. 3—Coordinate system for reflector antenna.

boundary, the line charge σ should be considered on Γ .

$$\sigma = -\frac{2}{j\omega} \Gamma \cdot \mathbf{H}_i \quad \dots \dots \dots (4)$$

Γ is the unit tangent vector of Γ shown in Fig. 3. Putting these vectors as wave source, the radiation electric field is expressed by the following equations.

$$\mathbf{E}_p = \frac{-j\omega\mu}{2\pi} \int_s [(\mathbf{n} \times \mathbf{H}_i) - \{(\mathbf{n} \times \mathbf{H}_i) \cdot \mathbf{r}_1\} \mathbf{r}_1] \frac{e^{-jkr}}{r} dS, \quad \dots \dots \dots (5)$$

$$\mathbf{H}_p = \frac{-jk}{2\pi} \int_s [(\mathbf{n} \times \mathbf{H}_i) \times \mathbf{r}_1] \frac{e^{-jkr}}{r} dS, \quad \dots \dots \dots (6)$$

- $\mathbf{E}_p, \mathbf{H}_p$: Electric and magnetic field at P .
- \mathbf{H}_i : Incident magnetic field
- \mathbf{n} : Unit vector normal to S
- r : Distance from source point on S to the field point P
- \mathbf{r}_1 : Unit vector along r
- S : Reflector surface
- k : Propagation constant in free space
- ω : Angular frequency
- μ : Permeability.

The magnetic field radiated from the primary feed or the front stage reflector is used

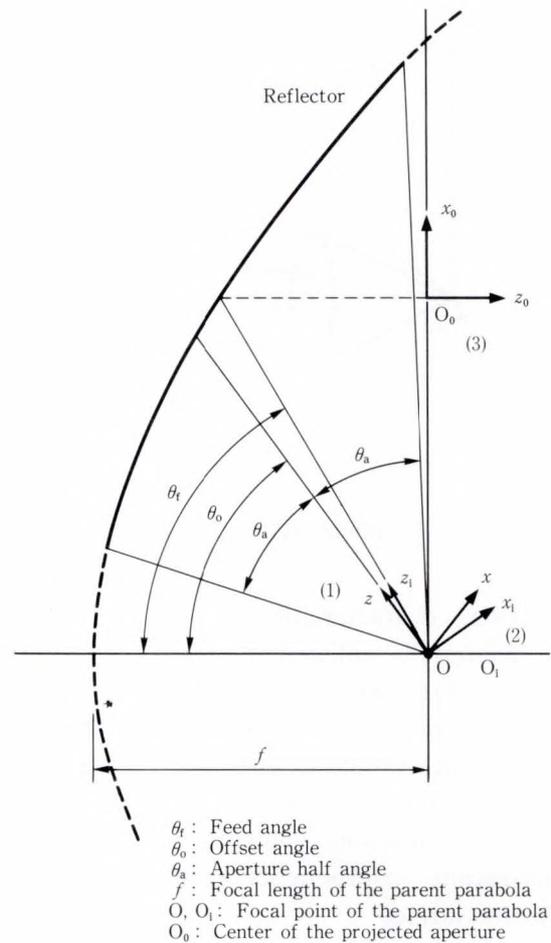


Fig. 4—Offset parabolic reflector coordinate system.

as the incident magnetic field. Any interference between the reflectors is ignored.

3.2 Reflector coordination method

The analytical program should be usable for antennas consisting of various kinds of reflectors. Therefore, a unified coordination method for processing each kind of reflector uniformly is necessary. For this purpose, the following three coordinates are introduced for each reflector:

The coordinates (1) define the shape of the reflector surface.

The coordinates (2) indicate the incident direction.

The coordinates (3) indicate the emitting direction.

These three coordinates were defined independently for respective reflectors.

Figure 4 indicates these coordinates for the offset-parabolic reflector.

For the coordinates (1), the most appropriate coordinates to represent the reflector shape are set. For the offset-parabolic reflector in Fig. 4, the z-axis of coordinates (1) is defined to agree with the offset angle direction.

The coordinates (2) and (3) indicate the incident and emitting direction of the reflector in the transmission mode. When several reflectors are combined, the emitting coordinates of the front stage reflector are set to coincide with the incident coordinates of the post stage reflector. For the primary feed, only coordinates (3) are defined and they are set to coincide with the incident coordinates of the first-stage reflector.

Various kinds of reflectors can be handled in the same way using these three coordinates. For an antenna consisting of a combination of several reflectors, the standard alignment position can be determined in only one way.

Figure 5 indicates the coordinates of an antenna consisting of two reflectors.

When this method is used, it is not necessary to calculate the absolute position of each reflector to analyze the antenna radiation characteristics. The relative location is easily calculated using the three coordinates and the radiation characteristics are calculated in the order of primary feed, reflector 1, and reflector 2.

3.3 Displacement handling

In the multibeam antenna, two or more primary-feeds are used to form two or more beams. Therefore, primary feeds are not located at a specific position such as the focus of the reflector. When the interference between beams is investigated, the characteristic degradation due to this displacement should be evaluated.

The characteristic degradations due to mechanical misalignment are serious problems for the design and manufacture of low sidelobe antennas. Therefore, it is necessary to evaluate the influence of misalignment quantitatively.

For these reasons, in the analysis program,

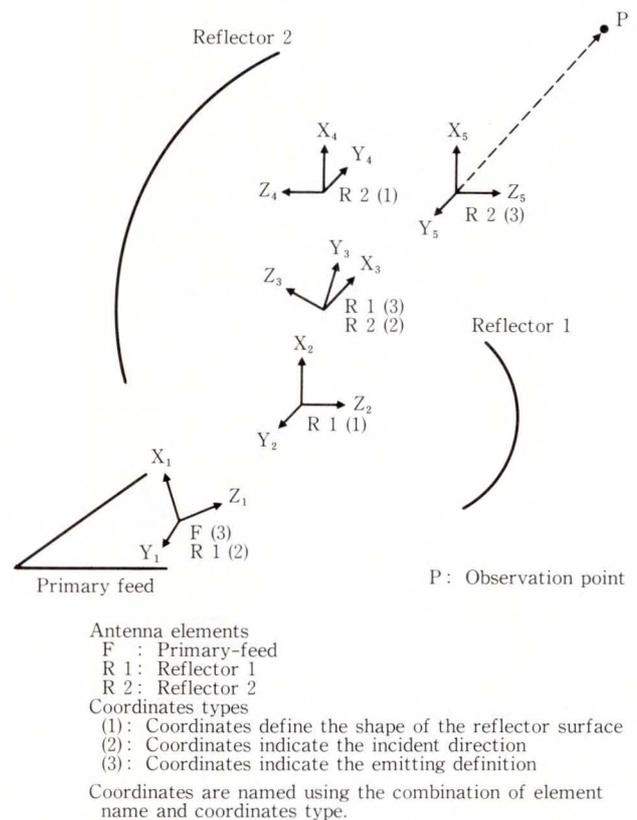
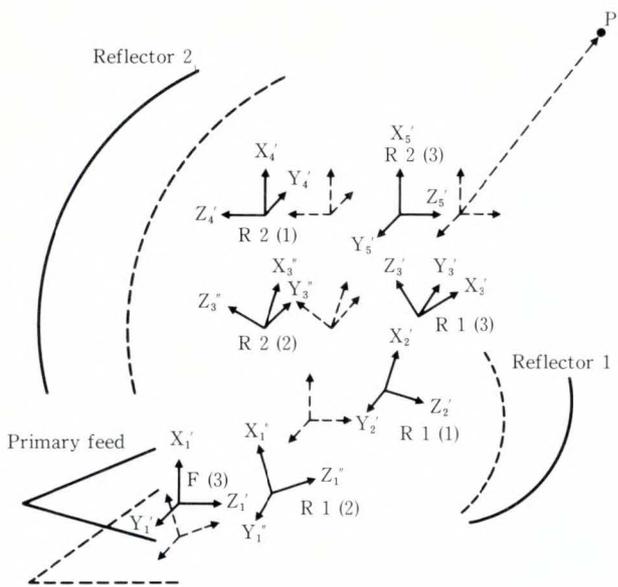


Fig. 5—Coordinate system for double-offset antenna.

the alignment displacement of each reflector is quantitatively represented using the coordinates (1) mentioned in the previous section and the radiation characteristics are analyzed for the antenna system with displacement.

The displacement is evaluated by six quantities: movement parallel to the X, Y, and Z axes and rotation around the X, Y and Z axes. Thus, it is possible to set the reflector at an arbitrary position. And since the displacement quantities are expressed by local coordinates defined for each reflector, they can be set independently for each reflector and it is easy to make these displacement quantities correspond to the quantities that can be measured in a real antenna system.

Figure 6 shows the relationship of the coordinates when each antenna element has alignment errors in an antenna consisting of two reflectors. The coordinates represented by dotted lines indicate the positions without displacement.



Dashed lines show feed, reflector and coordinates position without displacement.

Fig. 6—Coordinate system for double-offset antenna with displacement.

Table 1. Analysis objectives and methods

Kinds of primary feeds	Corrugated conical horn Multimode horn Pyramidal horn, etc.: 6 kinds
Kinds of reflectors	Parabolic Ellipsoid Hyperbolic, etc.: 6 kinds
Antenna configuration	Primary feed: 1 Reflectors: 1 to 3
Analytical method	Primary feed: Aperture field method Reflector: Current-distribution method

4. Program functions

The program was developed using the above-mentioned analytical methods. It can be used to calculate the radiation characteristics of the reflector antenna shown in Table 1.

The program has the following features:

- 1) The radiation characteristics of horn antennas are calculated using the aperture-field method. And models such as corrugated conical horn antenna, multimode, conical horn antenna, etc., widely used in satellite communication, are built into the

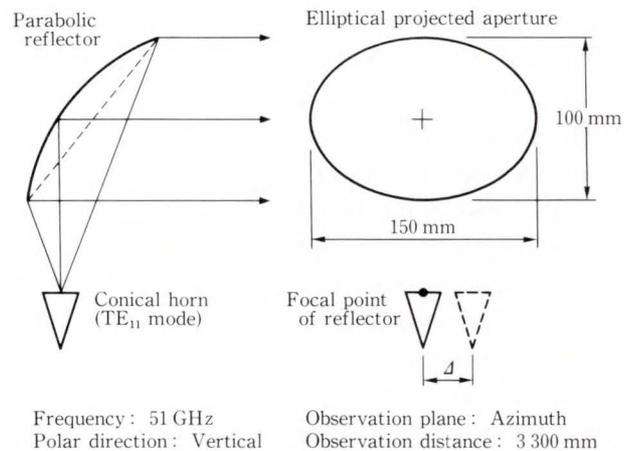


Fig. 7—Geometry of offset parabolic antenna with elliptical aperture.

program.

- 2) The current-distribution method was employed for reflectors and the analysis sections for reflectors such as paraboloid, hyperboloid and ellipsoid, are also built into the program. The radiation characteristics for antennas consisting of arbitrary combinations of these reflectors and primary feeds can be calculated.
- 3) The position of reflectors and primary feeds can be set arbitrarily and quantitative analysis for the misalignment and beam tilt characteristics due to the off-focus feed can be calculated.
- 4) Far-field approximation is not included in the analysis. Both the near field and far field characteristics of antennas can be calculated.
- 5) As the radiation electromagnetic field is handled using vectors, the cross-polarization components generated at reflector can be accurately calculated.
- 6) The data interface with GRAPHMAN¹³⁾ and JEF KING¹⁴⁾, which are application programs offered by Fujitsu, enables their graphical processing functions to be used.

5. Analytical results

Figure 7 shows the offset parabolic antenna configuration for the millimeter wave. The projected aperture shape is an ellipse and a

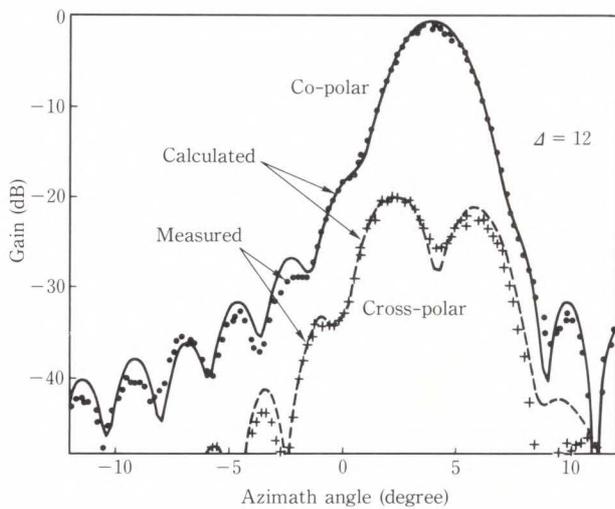


Fig. 8—Radiation field from offset parabolic antenna with elliptical aperture.

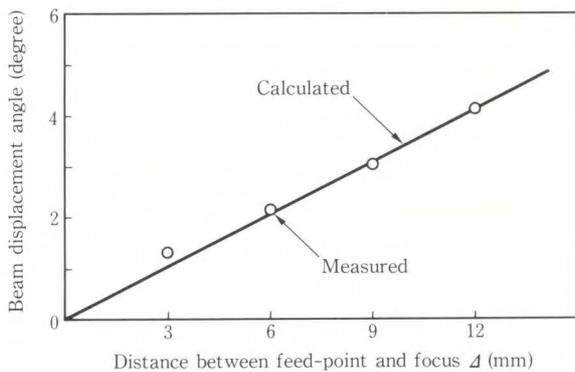


Fig. 9—Beam displacement as a function of primary-feed deviation.

conical horn antenna excited in TE^{11} mode is used as the primary feed. Figure 8 indicates radiation characteristics of this antenna when the primary feed of this antenna is shifted 12 mm from the regular position (see Fig. 7).

Since the feed point is moved, the direction of the main beam of the antenna changes about four degrees and the co-polarized and cross-polarized sidelobe levels of one side have increased. In this case, the observation points are located near the antenna. Nevertheless the analytical results agree well with the experimental results. Figure 9 shows the relationship between the primary feed deviation and beam tilt angle.



Fig. 10—Ku-band offset Gregorian antenna.

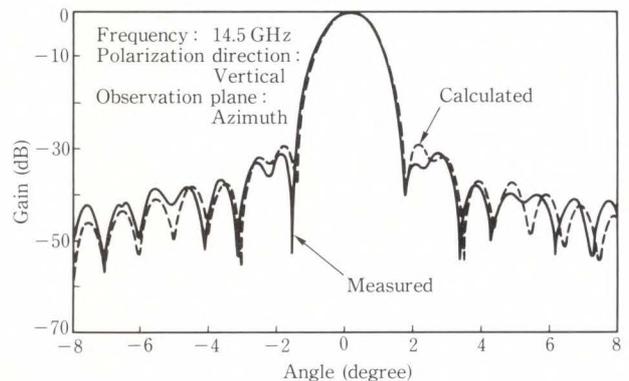


Fig. 11—Radiation field from offset Gregorian antenna.

Figure 10 shows the offset Gregorian antenna for a Ku-band earth station. The main reflector is paraboloid and the diameter of its aperture is 1.2 m. The subreflector is ellipsoid and a corrugated conical horn antenna excited in hybrid mode is used as a primary feed. The radiation characteristics of this antenna are shown in Fig. 11. The calculated results agree well with the experimental results.

In the analysis of multibeam antennas,

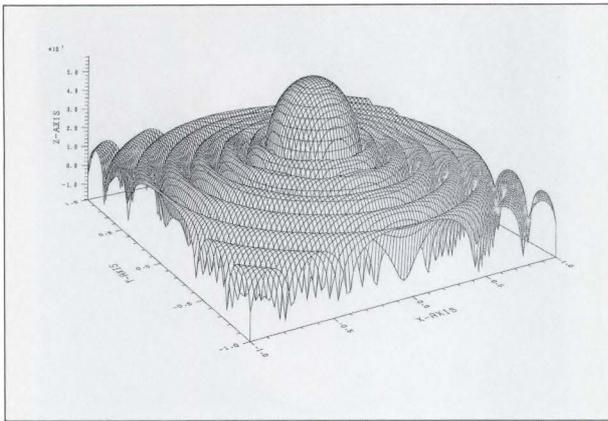


Fig. 12—Bird-eyes view of antenna pattern.

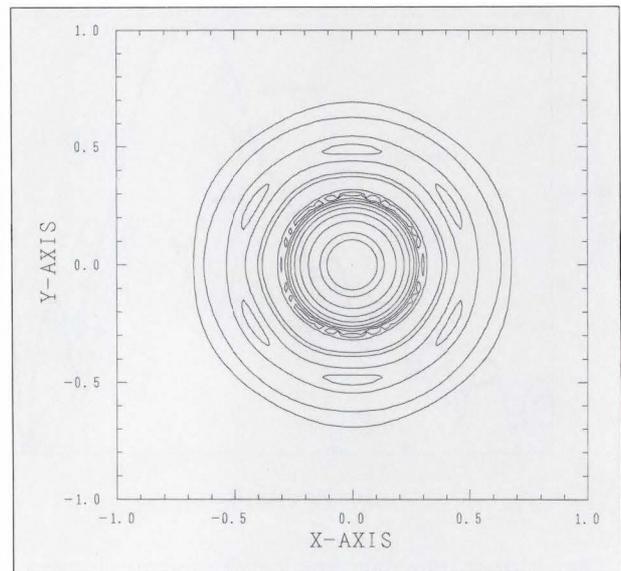


Fig. 13—Contour map of antenna pattern.

calculation of main beam direction and calculation of interference between beams are important. For these calculations, the antenna characteristics can be also expressed as a bird-eye view and/or a contour map as shown in Figs. 12 and 13. These patterns show the antenna characteristics calculated for investigating the mobile satellite communication system using a large geostationary platform that has been investigated for development at the beginning of the twenty-first century.

The time taken for the analysis of the offset Gregorian antenna was about 10 minutes, using a FACOM M-380.

6. Conclusion

Methods for estimating antenna radiation characteristics have been described. Both asymmetrical antenna configurations and misalignment between reflectors and primary feeds are taken into account. A computer program was developed using the estimation methods. Various functions of the program were explained in detail.

Radiation characteristics were estimated by the program. The influence of misalignment between reflectors and primary feeds on the radiation characteristics were also estimated to determine the required accuracy of antenna fabrication. The estimated results were verified experimentally.

The newly developed program gives sufficient information to determine design

accuracy for various antenna configurations. The estimated radiation characteristics is inevitable for the design of satellite communication systems.

7. Acknowledgement

The authors are grateful to Messrs. K. Yamada, S. Takenaka, and N. Okubo for their encouragement and helpful advices.

The authors would also like to thank Mr. S. Fukui for his useful discussions.

References

- 1) Bennet, S.B. and Braverman, D.J.: INTELSAT VI-A Continuing Evolution. *Proc. IEEE*; **72**, 11, pp. 1457-1468 (Nov. 1984).
- 2) Hagiwara, E., Yasuda, Y., and Mishima, H.: Optimized Equivalent-Weight Budget Design of Communication Satellite in Multibeam Mobile Satellite Communication System. *Trans. IECE, Jpn.*, **J69-B**, 3, pp. 237-245 (March 1986).
- 3) Parker, E.B.: Micro Earth Station as Personal Computer Accessories. *Proc. IEEE*; **72**, 11, pp. 1526-1531 (Nov. 1984).
- 4) FCC: Antenna Performance Standards. Regulation 25-209, (19xx).
- 5) CCIR: Radiation Diagram for Use as Design Objectives for Antennas of Earth Station Operating with Geostationary Satellites. Recommendation 580-1, (1986).

- 6) Galindo, V.: Design of Dual-Reflector Antennas with Arbitrary Phase and Amplitude Distributions. *IEEE Trans. Antennas & Propag.*, **AP-12**, 4, pp. 403-408 (July 1964).
- 7) Lee, J.J., Parad, L.I. and Chu, R.S.: A Shaped Offset-Reflector Antenna. *IEEE Trans. Antennas & Propag.*, **AP-27**, 2, pp. 165-171 (March 1979).
- 8) Mittra, R., Hyjazie, F. and Galindo-Israel, V.: Synthesis of Offset Dual Reflector Antennas Transforming a Given Feed Illumination Pattern into Specified Aperture Distribution. *IEEE Trans. Antennas & Propag.*, **AP-30**, 2, pp. 251-259 (March 1982).
- 9) Silver, S.: Microwave Antenna Theory and Design. New York, McGraw-Hill Book, 1st ed., pp. 158-160 (1949).
- 10) Silver, S.: Microwave Antenna Theory and Design. New York, McGraw-Hill Book, 1st ed., pp. 144-146 (1949).
- 11) Keller, J.B.: Geometrical Theory of Diffraction. *J. Opt. Soc. Am.*, **52**, pp. 116-130 (1962).
- 12) Li, T. and Turrin, R.H.: Near-Zone Field of the Conical Horn. *IEEE Trans. Antennas & Propag.*, **AP-12**, 8, pp. 800-802 (Nov. 1964).
- 13) FUJITSU Manual: FACOM OSIV/F4 MSP GRAPH-MAN HANDBOOK (in Japanese), 78AR-5170 (Jan. 1984).
- 14) FUJITSU Manual: FACOM OSIV/F4 MSP JEF/KING HANDBOOK (in Japanese), 70SP-7100, (March 1984).



Yoshihiko Asano

Radio & satellite Communication
Systems Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Electrical Communication
Eng.
Waseda University 1982
Specializing in Satellite
Communication Systems
and Equipments



Hiroshi Kurihara

Radio & Satellite Communication
Systems Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Electronics Eng.
Tokyo Institute of Technology 1969
Master of Electronics Eng.
Tokyo Institute of Technology 1971
Specializing in Radio Communication
Systems

LISP-PAL: An Approach to Natural Language Consultation in a Programming Environment

• Sanya Uehara • Rieko Nishioka • Tomoya Ogawa

(Manuscript received December 11, 1987)

The LISP-PAL approach is designed to assist programmers by supplying them with experienced expert programmer knowledge in response to their questions. For example, when a user asks in natural Japanese what functions or programming schemes would be useful for a particular programming purpose, the system displays appropriate usage guides and programming examples.

The purpose of this research is to build a knowledge-based software development environment and to develop technologies for utilizing knowledge bases. The main contributions of this work are the development of a consultation system that stores and retrieves programming knowledge in text form and a frame-based knowledge representation system. A technique to handle frame-based representation through pattern matching in an object-oriented programming language has also been developed.

1. Introduction

One of the tasks of software engineering is to develop the means to apply artificial intelligence to software engineering to improve programmer productivity and program reliability. One of the several approaches being taken is to assist programmers rather than replace them by fully automating the task. The advantage of this approach is that it leads to the construction of useful systems in the near future, while fully automated programming is still a long term goal.

The Programmer's Apprentice¹⁾ for example, assists with the routine aspects of the programming process by making use of a part of the programming knowledge called a "plan", thus allowing the programmer to concentrate on more difficult tasks.

Lisp Tutor²⁾ is a programming tutor that examines programs written by novice programmers and explains errors in a way that helps the programmer learn. It uses correct and buggy models of the programmer's programming

knowledge in the form of "if-then" rules.

Although these systems utilize programming knowledge to provide automatic assistance, they both require the system maintenance engineer to identify and input intrinsically complex programming knowledge in the required forms of representation, which is not a straightforward task. The current emphasis of LISP-PAL³⁾ is not to automate assistance, but to make available practical and extensive programming knowledge in text form. The knowledge supplied is typically a text consisting of explanations, a program scheme or example, and directions to help programmers apply the knowledge properly. The system targets novice and intermediate-level programmers.

LISP-PAL adopted a natural language interface because natural language allows even casual users to express themselves easily.

The goals in constructing the system were:

- 1) that the natural language interface be able to handle sentences with unknown words and ellipses,

- 2) that the system be able to lead the user to appropriate information even when the user's questions were not specific enough or not matched to an expression in the knowledge base, and
- 3) that the system make it easy for programmers to add information about their newly discovered programming knowledge to the knowledge base.

The purpose of the research was to build a knowledge-based software development environment and to develop technologies for utilizing a knowledge base. The main contributions of this work are the development of item 1) a consultation system that stores and retrieves programming knowledge in text form and of item 2) a frame-based knowledge representation system and a technique to handle it with pattern matching in an object-oriented programming language. This paper also discusses several features designed to cope with problems raised by maintenance of the knowledge base and the use of unknown words in user's questions.

The following sections describe the main features of LISP-PAL, its frame-based knowledge representation and its handling of complex frame structures with pattern matching. The conclusion reports on the current status of the system.

2. System structure

The system design is based on the following implementation decision. Throughout the system, knowledge is uniformly described in SEM (semantic representation) form, which is designed to express networked structures of knowledge bases, and most programming is done using pattern matching in SEM. This makes the program much easier to understand and write because it reduces the amount of programming code required to traverse the networked structure. This is discussed in detail in section 3.

Figure 1 shows the system structure. The system consists of two major parts, a translator and a question-answer module, which are joined by a SEM interface. Pattern matching is intensively used in both the translation rules and the question-answer rules.

User questions are translated by the translator into the internal representation SEM, which expresses the essential meaning of the input sentence. The translator is implemented using case-frame grammar definitions and heuristic parsing techniques that have been greatly extended from those of the ELI system⁴⁾.

A pattern-matching technique in SEM triggers question-answer rules to fetch appropriate knowledge from a LISP knowledge-base and to generate an answer in SEM. During this process, the question and the answer, both in SEM, are added to the context database. This context database is later used by the translator to decide, for example, what a particular pronoun denotes. The process of adding new knowledge to the LISP knowledge base progresses in much the same way as the question-answer process except that knowledge is added instead of being retrieved.

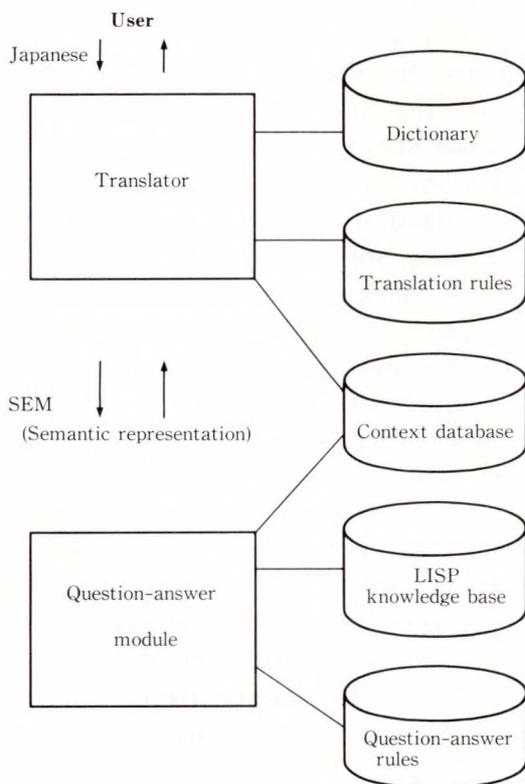


Fig. 1—System structure.

```
USER: How can I process all elements in a list?
PAL: Which skill do you want? Please select a number.
      1 process a list with MAP functions.
      2 process a list with DOLIST function.
USER: 2
PAL: The following is a skill to process all elements in a list.

      With the DOLIST function, you can handle all elements of a list very
      easily. For example, to apply the FOO function to elements of the list L,
      (DOLIST (elm L)
              (FOO elm))
      is sufficient.
      Although the value of the above expression is NIL, you can set it
      differently (ask for an example of DOLIST).

      It is simpler than MAP functions, but cannot treat a set of lists in a
      complex way as MAP functions can do.
```

Fig. 2—Question-answer session.

2.1 Question-answer

The system is most often asked questions concerning programming skills, LISP functions, and LISP technical terms. One important type of user question is the “purpose” question: What functions or schemes are available for specific programming purposes. Examples of this type of question are “(I want to know) how to write a program to select elements from a list”, and “I get the message ‘assumed special.’ What should I do?”

Figure 2 illustrates a question-answer session. (The sentences have been translated into English from Japanese, the system language.) When the system finds more than one piece of programming knowledge related to the question, it displays a menu of this knowledge to help the user select the most appropriate course of action. The user may then copy and modify the displayed program scheme in accordance with his program in the editing window.

The user may also resort to a variety of expressions that all have the same meaning, or he may pose questions that do not fully express his purpose in the existing situation. The features introduced to handle these problems are described in sections 2.2 and 2.4.

Other common types of questions are about technical terms and LISP function purposes, syntax, and usage. Examples of this type of question are “What is a ‘macro’?” and “How do I use a ‘macro’?”

Pronouns and ellipses are common in conversation. For example, consider this interaction:

USER: What function puts lists together?

PAL: Functions APPEND and NCONC do.

USER: (Give me) an example?

To deal with the second question, the system records the questions and answers in the context database in SEM form. It then searches the context database for the words that have something to do with “example,” namely, APPEND and NCONC. The information that LISP functions have something to do with “example” is found in the dictionary.

2.2 Varying expressions

Different expressions can be used that have the same meaning, such as “make a vector” and “create a vector.” Also, different choices of level lead to different expressions that are close in meaning, such as “process a vector” and “make a vector”, or “process an array” and “process a one-dimensional array.” The last example results from the user failing to reason out and specify his purpose clearly.

In LISP-PAL, the first type of problem is handled by the natural language interface, that is, the translator translates synonyms into single SEM primitives (symbols). Sentences whose word order is different or which include other syntactical differences that do not change the meaning are also translated into the same SEM by making use of the translation knowledge.

The second type of problem is handled by a hierarchical model for the meaning of sentences. When a user asks for a skill, the system also searches for more specific skills than that demanded by the question because therein may lie the skill that the user really needs. For example, if a user asks for information on how to “process an array,” the system also includes information on how to “process a vector” on the menu. (Note, “vector” is defined as a “one-dimensional array” in Common LISP⁵). With the model, “make a vector” is recognized as more specific than

“process a vector,” and process a “one-dimensional array” is more specific than “process an array”.

This level of specificity is decided based on a hierarchical model of words, which is composed of nouns and verbs. For instance, “make” is more specific than “process” in the model, “make a vector” is more specific than “process a vector”. In addition, a modifier increases specificity, so “a one-dimensional array” is more specific than “an array”. This model is domain (LISP) dependent. The implementation of this model is described in section 3.

2.3 Defining knowledge

The ease with which knowledge can be defined is a crucial factor in the maintenance of any type of knowledge base. Because the LISP-PAL knowledge base is organized in a complex manner with “frames,” as will be discussed in section 3, constructing and debugging the knowledge base are difficult

tasks.

Because communication between the user and the system is in Japanese, it is both more natural and easier to understand if the knowledge is defined in Japanese rather than in terms of frames.

Figure 3 shows how to use a template to add knowledge about the DOLIST LISP function. The system uses this description to create the corresponding frame structure and then adds it to the knowledge base. The system uses the translator to translate the key sentence led by “purpose is” and decides the proper place to put it in the meaning hierarchy discussed in section 2.2. The entire LISP knowledge base of LISP-PAL is constructed using three templates: a skill template, a function template, and a term template.

A knowledge definition interface is also provided to help the programmer while he is using the system. When LISP-PAL receives a sentence like “add new knowledge,” it creates a window showing the three templates. By simply filling in the appropriate template, the user can add his knowledge without starting another editor. The added knowledge is also output to the user’s private file so that the system knowledge base manager can subsequently include the information in the common knowledge base in a coherent form.

2.4 Unknown words

Dictionary maintenance is one of the most important tasks in managing systems that have a natural language interface.

When the system cannot understand words in a user’s question, it asks the user to teach it what they mean. The user is then asked to input a set of “synonyms” of each unknown word. If the system recognizes at least one of the synonyms, it registers all new words as synonyms in the “private” word dictionary. If none of the synonyms are recognized, the system first asks the user for a part of speech and a usage pattern before registering the synonym set. In this way, the system immediately enables a user to use new vocabulary in

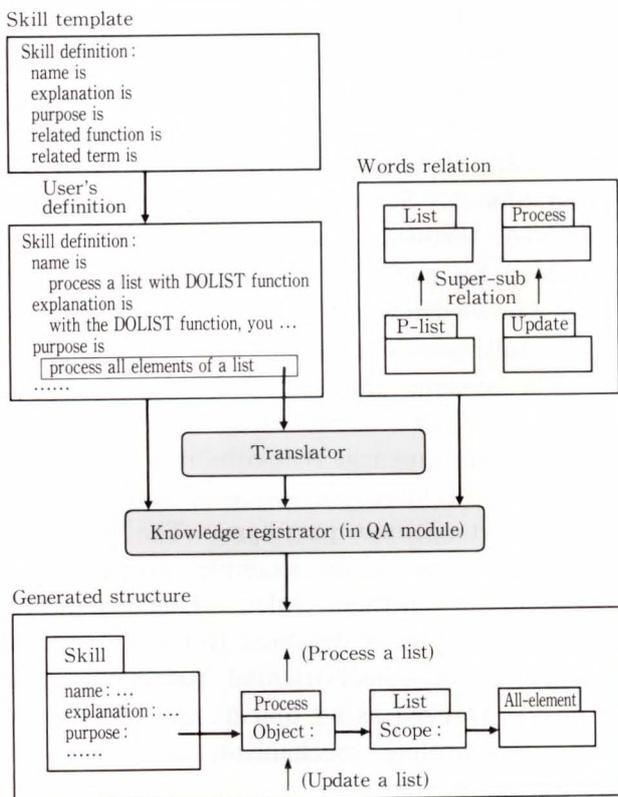


Fig. 3—Defining knowledge with a template.

questioning and adding new skills. The contents of private dictionaries should also be added to the common dictionary in a coherent manner by a dictionary maintenance technician.

3. Knowledge representation

The central issue in knowledge representation is how to organize an inclusion-relation of phrasal meaning in a consistent manner that enables the system to handle a variety of Japanese phrases in user's questions and knowledge definitions. Suppose, for example, that the meaning of "a computer" includes "a microcomputer". A frame system was adopted to achieve this type of relationship because frame hierarchy is suited to capturing relationships at the conceptual level. To implement the relationship, however, we must construct the phrasal inclusion-relation from the relationships already defined using words. For example, we can define that "He buys a computer" includes "He buys a micro-

computer" provided that "a computer" is defined to include "a microcomputer". We currently define the phrasal relationship based on the relationship defined for nouns, verbs, prepositional phrases, and adjectives.

In the LISP knowledge base, a LISP function includes the following information: function name, syntax, purpose, usage examples, and side effects. The important point of our approach is that the purpose slot of the LISP function or the skill contains information linked to the verb frame, like "make a vector" for the function VECTOR, as shown in Fig. 4.

The "purpose" of every LISP function or skill is represented by a verb, a LISP term, and other kinds of frames based on the results of natural language analysis. With this representation, the relationships between "purposes" can be defined in terms of the hierarchical relationships between individual frames. For instance, "make a vector" is regarded as having a more specific meaning than "make an array" because "vector" has a more specific meaning than "array". ("Vector" is defined as subtype of "array" in Common LISP 5).

Fundamental knowledge for the knowledge base can be acquired from LISP textbooks and manuals, but more practical and useful knowledge can be obtained from experienced programmers. Some of this knowledge was obtained through interviews and some was obtained by analyzing programming processes. Most knowledge was compiled in the form "how a particular goal is achieved".

The system and knowledge base are implemented in LISP-based object-oriented programming language that combines functional-procedural-oriented, data-oriented, rule-oriented, and object-oriented paradigms. The question-answer strategy is, for example, properly represented as if-then rules. This language, developed by us, is designed to realize a frame system in an object-oriented paradigm. Each frame is realized as an object, and inheritance and slot-handling mechanisms are supported by message passing.

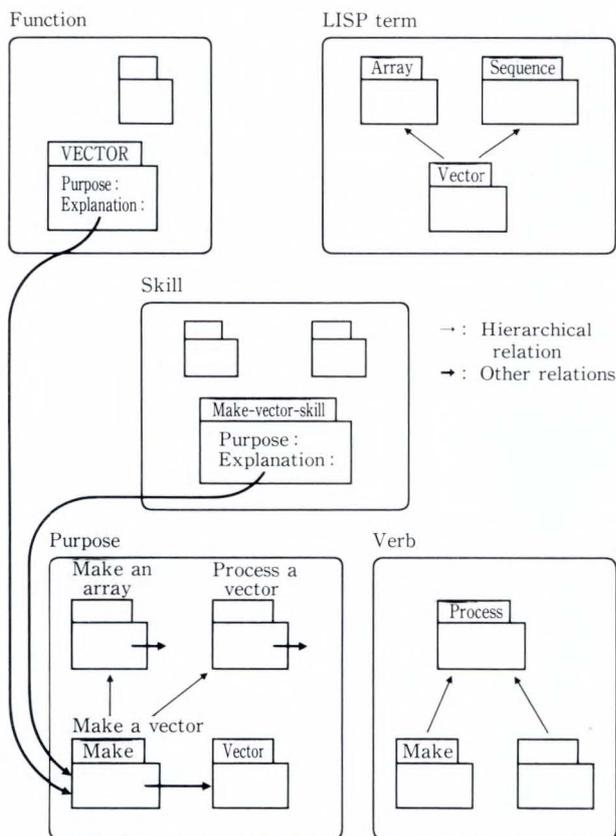


Fig. 4—Frame representation of purpose in LISP knowledge base.

3.1 Problems in handling knowledge bases

Because a frame is realized with an object in the programming language, the term “networked object” is used to denote the implementation of a set of interrelated frames such as those shown in Fig. 4. (One such frame set consists of the function “vector,” the verb “make”, and the term “vector”.) The main feature of object-oriented languages is the definition of methods for either single objects or a hierarchy of objects. In other words, they only allow us to traverse each object in a networked-object in sequence.

When the knowledge base is searched, the target is usually a set of networked-objects rather than a single object. This requires a new type of framework, one that improves the writability of networked objects.

3.2 Implementation framework

Our approach to the problem of handling networked objects was two-fold:

- 1) to implement “SEM” (semantic representation) to allow nested frame descriptions, and
- 2) to introduce knowledge-based operations associated with SEM. The design of SEM representation is based on frame-theory and predicate calculus⁶⁾. This framework enables the direct inclusion of networked objects in programs.

The principle idea in designing the consultation system was to provide the freedom to choose the most appropriate paradigm or notation for different problems. Figure 5 shows different representations of the concept: “make a list from elements of list A that satisfy a certain condition.” The frame-based knowledge representation at the bottom of the figure is the target we want to describe in programming; there is one instance-frame of “make-from” and three instance-frames of “list” and “element”. These frames are linked by an attribute relationship.

In common object-oriented programming languages, each frame is generally described separately. In SEM, however, such knowledge is described in the compact frame-based form shown in the figure. In this nested frame rep-

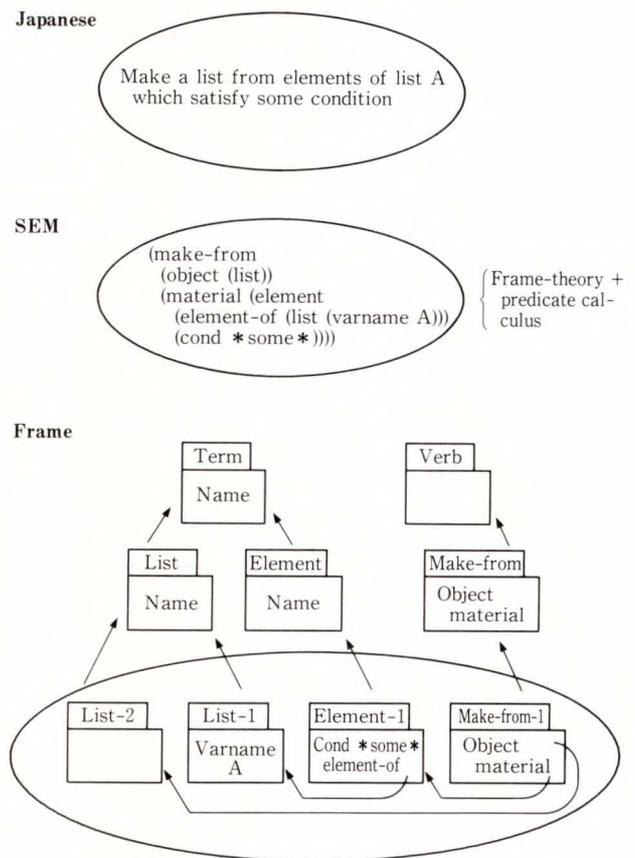


Fig. 5—Representations for Japanese sentence in LISP-PAL.

resentation, the frame/object name and a slot name appear alternately, that is, (make-from (object (list)) ...). No instance name appears because SEM represents the meaning of sentences rather than an internal representation. In general, SEM is the appropriate representation for networked objects to be written in programs.

Figure 6 shows how SEM is utilized in a question-answer session. The Japanese translator converts an input sentence into SEM. An appropriate question-answer rule performs the next four-stage task: understanding the user’s question by pattern matching, fetching the desired skill from the knowledge base using a SEM operation, multi-fetch, and creating a response in SEM with the retrieved knowledge. The multi-fetch argument is a pattern for networked-objects, which allows us to handle a set of objects at one time.

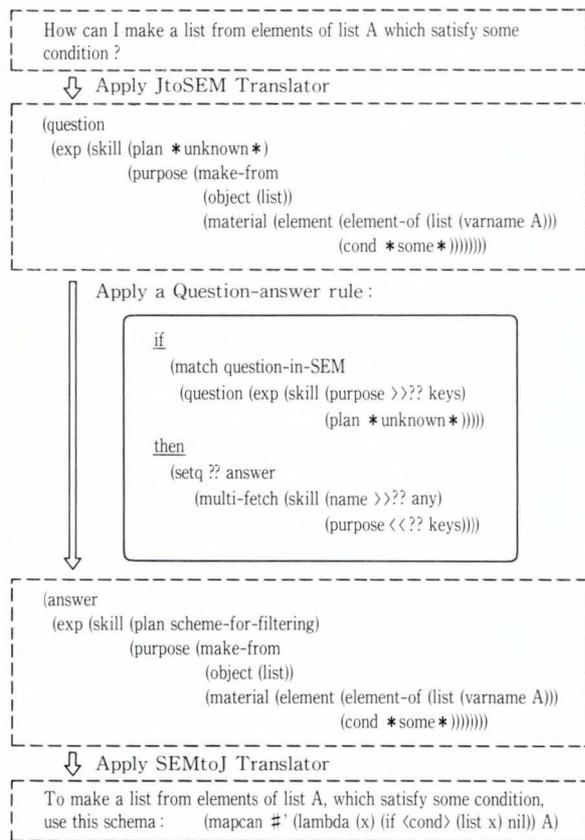


Fig. 6—SEM representation in the question-answer process.

The variable beginning with “??” represents a set of objects, “>>” is an operator that assigns a set of objects to the variable, and “<<” is an operator that refers to a set of objects from the variable. This assignment step also eliminates the need to handle each object separately. The key advantage of this if-then program is that a pattern of networked objects can play an important role as a meaningful piece of knowledge in this process. Finally, the response in SEM is translated into Japanese by applying the Japanese generation rules used by the translator.

4. Conclusion

In this paper, we described a system for teaching LISP programming skills using a natural language interface. We discussed frame-based knowledge representation, which expresses the phrasal inclusion-relation, an implementation technique to handle knowledge using pattern

matching, and several strategies to cope with other issues associated with the system. We also described the use of a menu-driven interface in cooperation with the natural language interface; a menu is created when the user has a choice of many things to view. We also observed that additional strategies must be introduced to compensate for the shortcomings of natural language analysis. An example of this is that when the system does not understand a sentence completely, it shows the user typical sentence patterns that it can accept.

Through our experimental implementation of LISP-PAL, we observed that nested frame expression with its associated operations is a powerful tool capable of handling a knowledge base in the frame model. This implies that if we were to restrict ourselves to handling single objects rather than a structural set of objects, we would not explore the full power of object-oriented paradigms. We applied a pattern-matching technique to an object-oriented paradigm to handle a structural set of objects, thereby improving program writability and understandability to some extent. For example, it is easy to modify our question-answer rules. This approach is an extension of the work done in Unix Consultant⁷⁾ towards the application of object-oriented paradigms.

A prototype system was implemented on a FACOM M-series computer. The enhanced system is currently running on an Apollo Domain workstation DN3000 and a FACOM FMR-series personal computer. The knowledge base contains more than 300 programming examples and explanations of programming skills and LISP functions.

We are currently evaluating the system from three viewpoints: the effectiveness of the user interface, including the natural language interface, the quality of the knowledge base, and the usefulness of the system for educational and programming assistance purposes. Important improvements in the future include an improved collection of programming examples that would also enable the system to show different sets of program examples depending on the user's skill level. Skill-differentiated assistance would

require that the system automatically recognize the user's skill level and that the knowledge base contents be organized by user skill levels. We are also looking at a formalized extension of the object-oriented paradigm to handle structural set of objects based on our experimentation.

5. Acknowledgement

The authors would like to thank Messrs. Y. Soma, N. Murakami, M. Sugimoto, T. Uehara, and T. Mori for their support and encouragement during this work.

References

- 1) Waters, R.C.: The Programmer's Apprentice: A Session with KBEmacs. *IEEE Trans. Soft. Eng.*, **SE-11**, 11, pp. 1296-1320 (Nov. 1985).
- 2) Anderson, J.R. and Skwarechi, E.: The Automated Tutoring of Introductory Computer Programming. *Commun. ACM*, **29**, 9, pp. 842-849 (Sept. 1986).
- 3) Uehara, S., Nishioka, R., Ogawa, T., and Mohri, T.: Implementing Programming Consultation System LISP-PAL: Framework for Handling Frame-Based Knowledge in Object-Oriented Paradigm. CIAM 86, Second Inter. AI Conf., 1986, pp. 409-423.
- 4) Schank, R.C. and Riesbeck, C.K.: Inside Computer Understanding. Lawrence Erlbaum Associates, (1981).
- 5) Steele Jr., G.L.: Common LISP: The Language. Digital Press, (1984).
- 6) Charniak, E. and McDermott, D.: Introduction to Artificial Intelligence. Addison-Wesley, (1985).
- 7) Chin, D.N.: Case Study of Knowledge Representation in UC. IJCAI, pp. 388-390 (1983).



Sanya Uehara

Software Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Information Science
Tokyo Institute of Technology 1977
Master of Information Science
Tokyo Institute of Technology 1980
Specializing in Computer Science



Tomoya Ogawa

Software Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Control Eng.
Tokyo Institute of Technology 1983
Master of System Science
Tokyo Institute of Technology 1985
Specializing in Computer Science



Rieko Nishioka

Software Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Electronics and
Communications Eng.
Waseda University 1983
Specializing in Computer Science

A System for Protein-Crystal Growth Experiment in a Microgravity Environment

• Shozo Fujita • Koji Asano • Takafumi Hashitani

(Manuscript received December 3, 1987)

To evaluate the effect of microgravity on protein-crystal growth, a system performing experiments automatically in the Small Self-Contained Payload container offered by NASA has been developed. The system can simultaneously perform 16 independent experiments which apply three crystallization methods. Crystal growth is started by sliding out plates separating protein and ammonium sulfate solutions, and the state of the growth is periodically recorded by taking photographs for up to 120 hours. The temperature is passively controlled by a heat regenerator. Fujitsu can thereby evaluate the effectiveness of microgravity by comparing the protein-crystal growth in space with that on the ground.

1. Introduction

Recent advances in space transportation systems have opened the way to using unique qualities of space, such as microgravity, ultra-high vacuum and cosmic radiation. Of the many proposals for space experiments, the processing of biochemical material would be one of the most fruitful, because the high value per unit weight offsets the cost of transportation.

Another new technology, protein engineering, enables us to modify parts of proteins to improve their functions, such as thermal stability and substrate specificity. Because the protein's function depends on its molecular structure, the protein's three-dimensional structure must be determined for this improvement to be efficient.

Of several methods for determining the structure, X-ray diffraction analysis of protein single-crystals, which enables us to determine the protein's absolute atomic coordinates, is now the most reliable means. Although it is important, the determination of protein structures is a slow job. The major bottle neck is the difficulty to make well-ordered and large single-

crystals.

To reduce the bottle neck, there have been several attempts to make protein crystals in space. Littke and John¹⁾ recently reported that a microgravity environment accelerates crystal nucleation and growth based on their observations of the Spacelab 1 mission in 1983. DeLucas et al. on the other hand carried out several experiments on the Mid-deck of the Space Shuttle and reported that disorders caused by convection can conceivably be eliminated in a microgravity environment²⁾.

Although these experiments point to the advantages of space, they were operated manually by payload specialists. The resulting vibrations could conceivably have reduced the beneficial effects of microgravity. Furthermore, the structural restrictions of their hardware allowed only one or two crystallization methods to be investigated.

We have since developed fully automated system to investigate three different crystallization methods at the same time. Using this system, crystal growth in space will be compared with that on the ground. Since the thermal

conditions in space greatly depend on the altitude of the particular Shuttle flight, the simulations conducted under the same conditions on the ground as those recorded in the system will be carried out after the recovery of the system.

This paper describes the structure of the hardware, and the results of experiments on the ground carried out to test the system.

2. Structure of hardware for the GAS system

2.1 GAS system

The small self-contained payload (the so-called Get Away Special or GAS) system of the National Aeronautics and Space Administration (NASA) is an inexpensive and convenient means of accessing a space environment³⁾. NASA provides a canister with a thermal insulator and board which interfaces the system with the Space Shuttle, enabling the crew members to switch the GAS experiment on and off up to six times according to a predetermined time schedule.

In order to fit into the standard GAS canister, the system must occupy five cubic feet (19.75 inches in diameter and 28.25 inches high) and weigh less than 200 pounds. Since the system must be self-contained, it should have its own power source, control and safety circuits, data recording unit and so on.

2.2 Hardware

To efficiently use the system for various experiments in future with minimum modifications, it has been separated into two parts, a module for experiment and a support system. The former is actually responsible for performing the experiment and the latter controls the module and does housekeeping.

2.2.1 Experimental module

1) Requirements

The protein-crystal growth experiment and its hardware, the experimental module, are required to have the following features.

- i) Adaptability to suit several crystallization methods
- ii) Minimum vibration during crystal growth
- iii) Long experimental duration to complete crystal growth
- iv) Efficient means of recording the crystal growth process in space
- v) Endurance to acceleration and vibration during launch.

2) The crystallization unit

Protein can be crystallized from a moderately supersaturated aqueous solution, which can be prepared in various ways⁴⁾. Three crystallization methods were chosen: free-interface diffusion, batch, and vapor diffusion.

Three crystallization units corresponding to each method were designed (see Fig. 1).

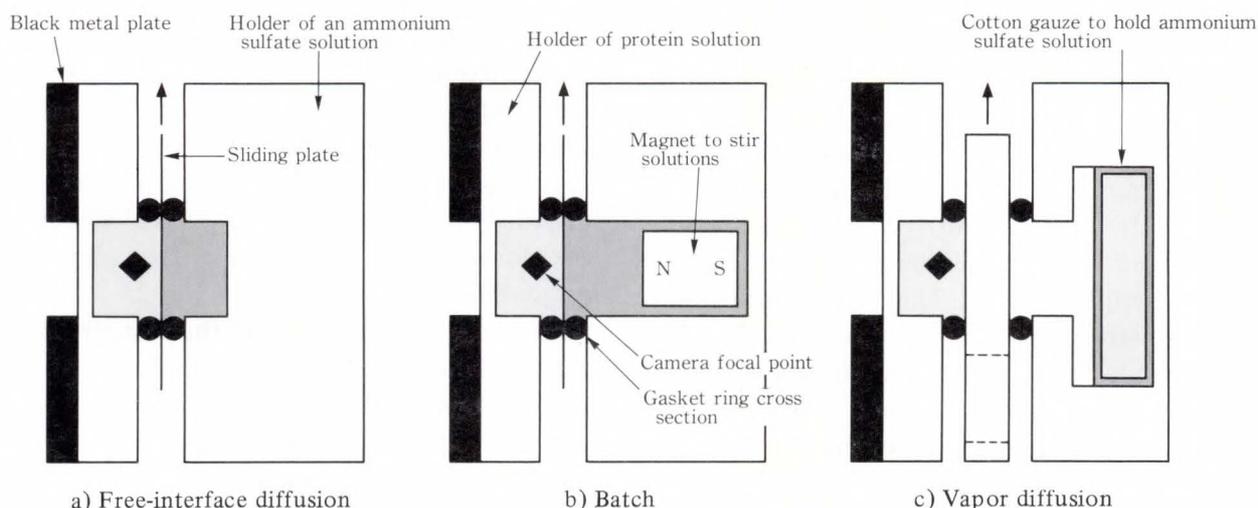


Fig. 1—Scheme of three types of crystallization unit.

They are the same size but have different internal structures. The batch unit contains a small magnet which is used to mix the solutions. The internal cavity of each unit is separated into two compartments by a thin stainless steel or thick plastic plate. One side holds the protein solution and the other an ammonium sulfate solution. Each unit made of acrylic resin is required to have a tight seal to prevent evaporation of the solvent during the several months of stand-by. Therefore, each joint has a gasket made of Viton or Teflon.

The experiments are initiated by sliding out the plates. In free-interface diffusion, ammonium sulfate ions diffuse into the protein solution via the free interface, forming crystal nuclei near the interface. In the batch method, the two solutions are thoroughly mixed by the magnet to obtain supersaturation at the start of the experiment. In vapor diffusion, the two solutions, a highly concentrated solution of ammonium sulfate and a protein solution containing a low concentration of ammonium sulfate, come into contact with each other through an air gap between them. The former gradually absorbs water, while the latter gradually becomes supersaturated.

3) Assembly of the experimental module

Sixteen crystallization units are mounted radially in the module, with their sliding plates linked to a triangular plate just above them (see Fig. 2). Once the Space Shuttle is in orbit, the experiments are initiated by raising the triangular plate by means of the stepping motor and ball-screws. The plate is lifted at a rate of 6 mm/min.

The crystal growth is photographed by a still camera with a power film winder and a 30-foot film attachment. A lens is fitted in the cover plate and a mirror is rotated by another stepping motor. They allow each unit to be observed in turn without movement of the units and crystals. A ring-flash lamp flashes each time camera shutter is released.

The assembled experimental module withstood a random vibration test 9 g(rms) which was carried out according to the NASA manual³⁾.

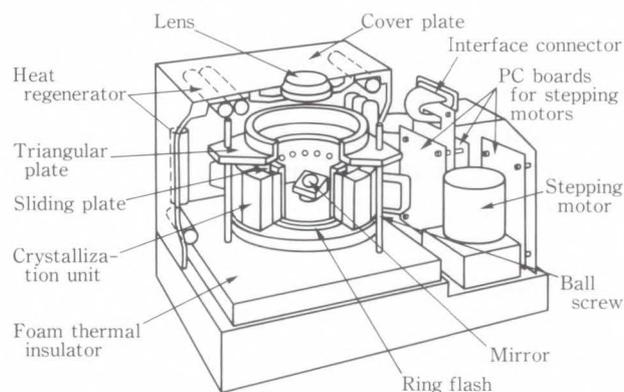


Fig. 2—Scheme and photograph of the experimental module.

4) Thermal control

Predicting and controlling the thermal conditions during the experiments are the most important and most difficult problems to be overcome when designing a GAS payload⁵⁾. Since precise thermal control is difficult because of restrictions of power, a baroswitch and a heat regenerator are used to regulate the temperature within a certain range.

During ascent, the baroswitch activates the system at 70 000 feet, prior to the opening of the doors of the shuttle cargo bay. Since the cargo bay is air-conditioned to 21 °C before launch, the system can start the experiments at that temperature before exposure to the extreme temperatures of space.

The heat regenerator acts as a buffer, preventing thermal fluctuation by means of exothermic and endothermic phase transition reactions of inorganic hydrate. The regenerator

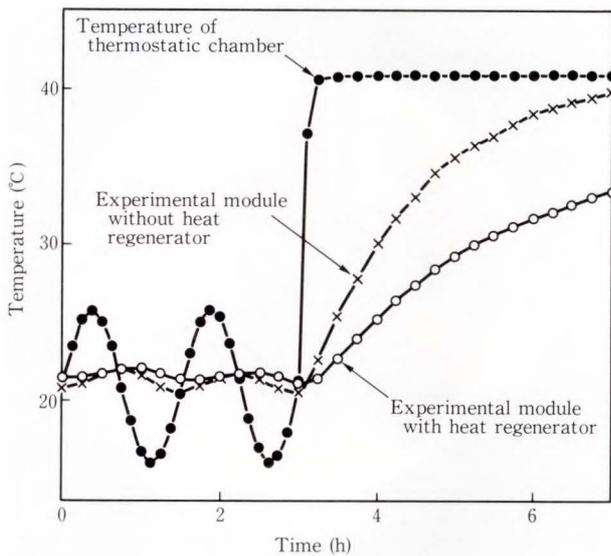


Fig. 3—Effect of the heat regenerator on the temperature of a crystallization unit.

consists of 44 tubes of 12 ml each, each containing a mixture of nine parts of $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$ and one part of $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$. Its transition temperature is 25 °C, which is dependent on the composition of the hydrates. These tubes are installed inside the cover plate (see Fig. 2).

The temperature change of the crystallization unit in the module is indicated in Fig. 3, where the first three hours simulate cyclic fluctuations originating from the orbital period (90 min) of the Shuttle and the latter four hours provide a constant heat flux. The temperature of the unit equipped with the regenerator changes much more slowly than that without the regenerator, although the regenerator is unable to keep the temperature constant.

2.2.2 Support system

Figure 4 shows the appearance of the whole system, while Fig. 5 is its block diagram. Three disks and four struts provide mechanical support. Lead-acid batteries are mounted at the top of the system and the module on the bottom plate. The camera is mounted in the central cavity of the system. All other subsystems are mounted on the struts. A detailed description of the support system has been published previously⁶⁾.

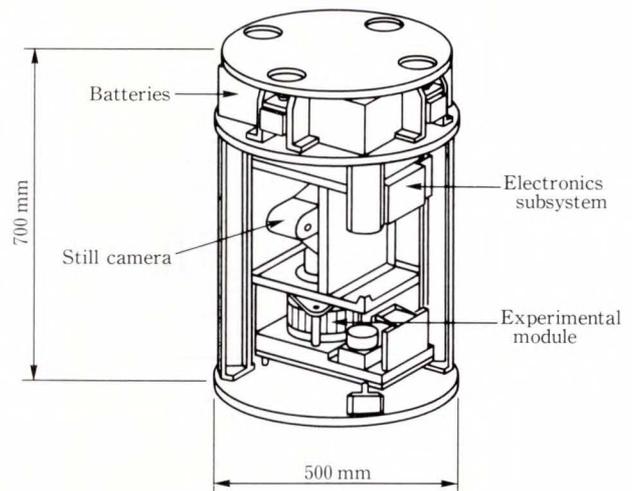


Fig. 4—Scheme of the whole system.

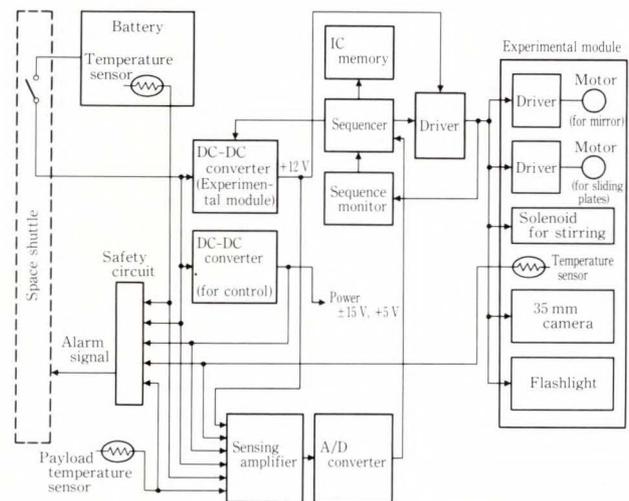


Fig. 5—Block diagram of the system.

3. Operational scenario

After being activated by the baroswitch, the system waits three hours to allow the Shuttle to maneuver, then starts the experiments (see Fig. 6).

First, each unit is photographed as a control. Then, the triangular plate is lifted to remove the sliding plates and the magnets in the batch units are operated to mix the solutions. Each unit is then photographed periodically every 6, 12, or 24 hours for up to 120 hours.

The housekeeping functions keep a record

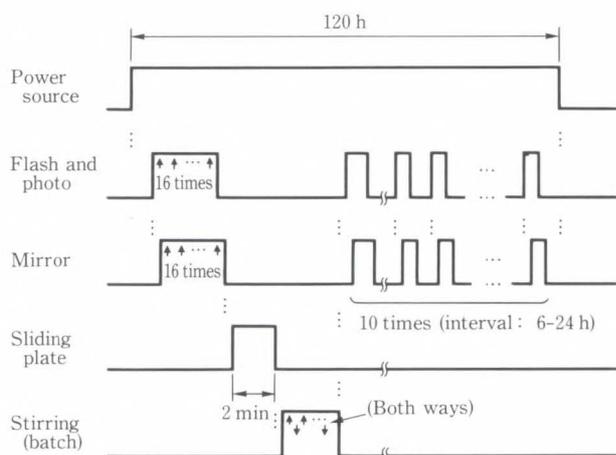


Fig. 6—Control sequence for the system.

Table 1. Composition of the solutions used in the experiment on the ground

Solution Methods	Protein solution		Salt solution
	Myoglobin ¹⁾	(NH ₄) ₂ SO ₄ ²⁾	(NH ₄) ₂ SO ₄ ²⁾
Free-Interface diffusion ³⁾	2.0	55	97
	1.5	70	82
	1.0	55	97
Batch	4.7	50	88
	3.2	50	88
	3.2	50	91
	3.2	50	94
Vapor diffusion ³⁾	1.0	60	90
	1.0	60	80
	1.0	60	70

1): concentration of sperm whale myoglobin (weight %)
 2): concentration of ammonium sulfate (% of saturation)
 3): experiment was duplicated under these conditions

of such items as the temperature in the experimental module and support system, by storing the data in the IC memory every three minutes. Thus, the thermal conditions in space can be reproduced when the comparable experiments on the ground are performed subsequently.

4. Experiment on the ground

The system has been evaluated on the ground by running the experiments under conditions as close as possible to those in space. The experiments are performed according to the sequence shown in Fig. 6, except for the influence of normal gravity and that the system



Fig. 7—Photograph of Myoglobin crystals taken during the experiment on the ground (One of the free-interface diffusion units at 120 h).

is activated by a dummy switch instead of the baroswitch.

Sperm whale myoglobin was chosen as the sample protein because it remains stable for several months in aqueous solution, it easily and stably crystallizes, and has a brownish color that makes it easy to observe. Furthermore, its three dimensional structure is understood.

Composition of the solutions used is summarized in Table 1. Fig. 7 shows an example of crystals grown in the units, one of free-interface diffusion units at 120 hours. The largest crystal in the photograph is 0.9 mm long.

In the free-interface diffusion and batch experiments, crystals were formed after 36 and 48 hours, respectively (see Fig. 8), and continued to grow, while crystals in the vapor diffusion units had not formed by the end of the experiments (120 hours), but formed the next day (data not shown). This delay suggests that crystal nucleation requires some temperature variation or vibration.

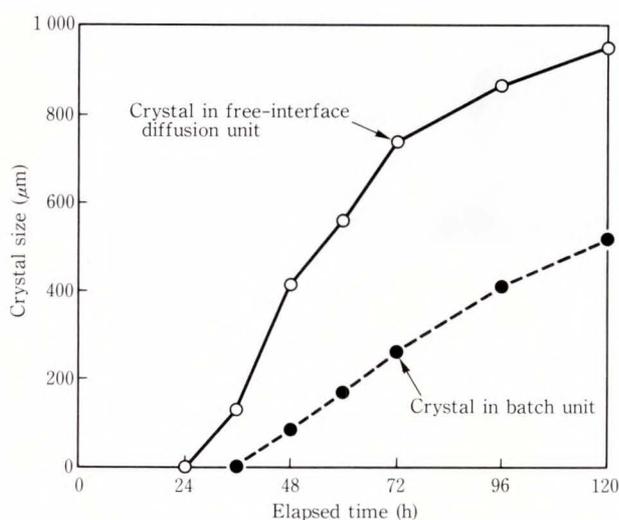


Fig. 8—Size of the largest Myoglobin crystals grown in the experiment on the ground

5. Conclusion

An automated system for a protein-crystal growth experiment in space has been developed to enable 16 experiments using three crystallization methods for 120 hours, photographing periodically. It has been repeatedly tested on the ground and accomplished experiments satisfactorily and large crystals of myoglobin could be grown by two of the three methods.

6. Acknowledgement

The authors would like to thank Messrs. S. Kitamura and H. Nemoto of Fujitsu Limited for their activity on the development of the entire system and Messrs. H. Okuyama and T. Kitakohji of Fujitsu Laboratories Limited for their encouraging advice.

This project is supported by the Mechanical Social Systems Foundation and the Society of Japanese Aerospace Companies.

Part of this work was presented at the 1987 Get Away Special Experimenter's Symposium⁷⁾, at Goddard Space Flight Center of NASA, Maryland USA, on October 27, 1987 and the third meeting of the Japan Society of Microgravity Application, at the National Research Institute for Metals, Tokyo Japan, on November 13, 1987.

References

- 1) Littke, W., and John, C.: Protein single crystal growth under microgravity. *J. Cryst. Growth*, **76**, 3, pp. 663-632 (1986).
- 2) DeLucas, L.J., Suddath, F.L., Snyder, R., Naumann, R., Broom, M.B., Pusey, M., Yost, V., Herren, B., Carter, D., Nelson, B., Meehan, E.J., McPherson, A., and Bugg, C.B.: Preliminary investigations of protein crystal growth using the space shuttle. *J. Cryst. Growth*, **76**, 3, pp. 681-693 (1986).
- 3) NASA: *Get Away Special (GAS) Small Self-contained Payloads Experimenter Handbook*. Greenbelt Maryland, NASA Goddard Space Flight Center, 1985, p. 91.
- 4) McPherson, A.: *Preparation and Analysis of Protein Crystal*. New York, John Wiley & Sons, 1982, pp. 82-159.
- 5) Cronise, R.J.: Electrochemical Growth of Linear Conducting Crystals in Microgravity. *Proc. 1987 Get Away Special Experimenter's Sympo.*, NASA Goddard Space Flight Center, 1987-10, Greenbelt Maryland, 1987, pp. 47-54.
- 6) Kitamura, S., Nemoto, H., and Kodama, S.: Development of GAS Small Self-Contained Payload. (in Japanese), *FUJITSU*, **38**, 1, pp. 57-66, (1987).
- 7) Fujita, S., Asano, K., Hashitani, T., Kitakohji, T., Nemoto, H., and Kitamura, S.: Protein-Crystal Growth Experiment (Planned). *Proc. 1987 Get Away Special Experimenter's Sympo.*, **6**, NASA Goddard Space Flight Center, 1987-10, Greenbelt Maryland, 1987, pp. 63-68.



Shozo Fujita

Organic Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Science
Niigata University 1975
Dr. of Medical Science (Neurochemistry)
Niigata University 1981
Specializing in Biochemistry and
Development of Equipments for
Space Experiments



Takafumi Hashitani

Organic Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Education
Kanazawa University 1984
Master of Education (Scientific
Education)
Kanazawa University 1986
Specializing in Molecular Biology and
Development of Equipments for Space
Experiments



Koji Asano

Organic Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Applied Chemistry
Nagoya Institute of Technology 1982
Master of Synthetic Chemistry
Nagoya Institute of Technology 1984
Specializing in Chemistry and
Development of Equipments for
Space Experiments

Liquid Phase Epitaxial Growth of Buried Heterostructure DFB Lasers

• Toshihiro Kusunoki • Nirou Okazaki • Toshiyuki Tanahashi

(Manuscript received November 30, 1987)

Flat-surface buried-heterostructure distributed-feedback laser diodes were fabricated by liquid-phase epitaxy. There were three difficulties to be solved for improving the characteristics of the laser. The first difficulty was the formation of pits in the InGaAsP optical guide layer on the corrugations. The other difficulties were a thermal deformation of the corrugations and leakage current flowing along the sides of the mesa. The first difficulty was solved by using an oxygen plasma process for cleaning the corrugations. The thermal deformation of the corrugations was reduced by covering the corrugations with a GaAs plate that had a rough surface. The leakage current was reduced by using a new etchant for etching the mesa.

1. Introduction

Distributed-feedback (DFB) laser diodes are used as light sources for high-speed and high-capacity optical communication systems. We constructed flat-surface buried-heterostructure (FBH) DFB lasers with an original structure, but encountered three difficulties. The first difficulty was that pits were formed in the InGaAsP epitaxial layer caused by small particles of resist remaining on the surface of the corrugations. The second difficulty was that the corrugation covered by an InP plate was thermally deformed by the heating process before the LPE growth. The third difficulty was that a leakage current flowed along the sides of the mesa with the (111)A plane because the (111)A plane was easily degraded in a thermal ambient¹⁾ at the growth temperature.

In the following sections, the production of FBH-DFB lasers is described. The solutions to the difficulties mentioned above will also be described.

2. The production of FBH-DFB lasers

As shown in Fig. 1-a), a guide layer (n-InGaAsP), an active layer (InGaAsP), a cladding layer (p-InP), and a cap layer (p-InGaAsP) are

grown in that order on a (100) InP substrate with corrugations. A SiO₂ mask is patterned on this wafer, and the wafer is then etched with a new type of etchant which is a mixture of HBr, H₂O₂, and H₂O to form the mesa structure, as shown in Fig. 1-b). The mesa is buried in the p-InP layer and then in the n-InP layer {see Fig. 1-c)}. The SiO₂ mask and the cap layer are removed, and the p-InP and p-InGaAsP contact layers are grown on the wafer {see Fig. 1-d)}. After growth, the p and n electrodes are patterned on both surfaces of the wafer. Al₂O₃ is then coated on one of the cleaved facets of the wafer. The resultant FBH-DFB laser is shown in Fig. 2.

As shown in Fig. 2, the active layer is buried in a current-confinement layer (p-n-p InP), and the surface of the wafer buried by the LPE growth is flat.

3. Improving the difficulties

3.1 Improving the cleaning process of the corrugated substrate

3.1.1 Experiment

InGaAsP ($\lambda_{Eg} = 1.3 \mu\text{m}$) was grown by LPE on a (100) InP corrugated substrate. The growth temperature was 585 °C and the growth time

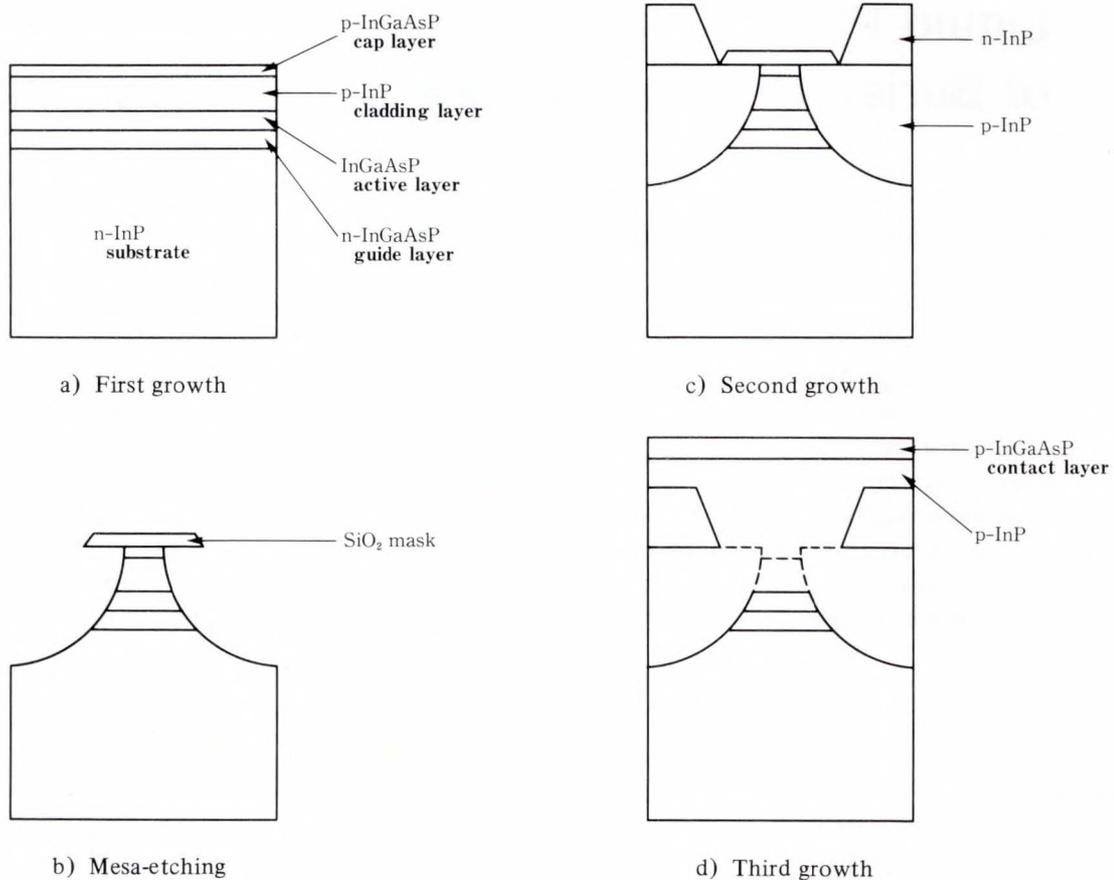


Fig. 1—The production of a flat-surfaced buried heterostructure DFB laser.

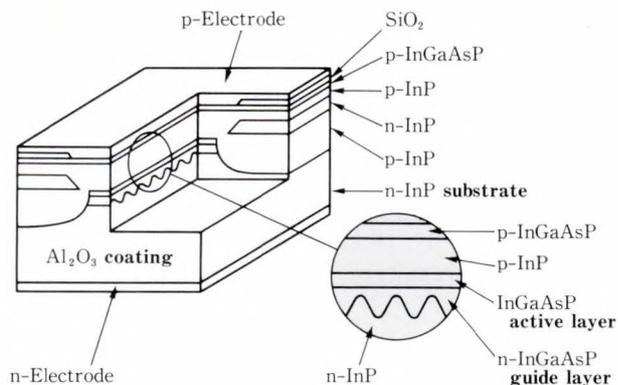


Fig. 2—The schematic structure of a flat-surfaced buried heterostructure DFB laser.

was 20 s.

Before growth, the corrugations were cleaned by one of three methods:

- 1) The corrugated substrate was ultrasonically cleaned in an organic solvent, rinsed in water, dipped in H_2SO_4 , and again rinsed in

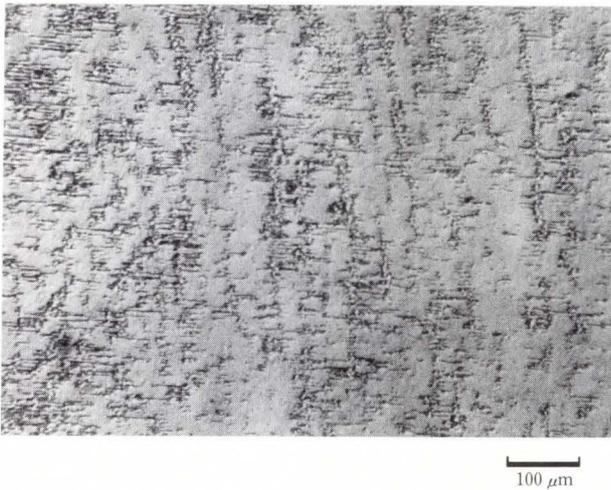
water;

- 2) Corrugations were formed on only half the substrate which was cleaned by the same procedure as in method 1); or
- 3) The corrugated surface was cleaned by the oxygen plasma process. The substrate was then cleaned by the same procedure as in the above mentioned method 1).

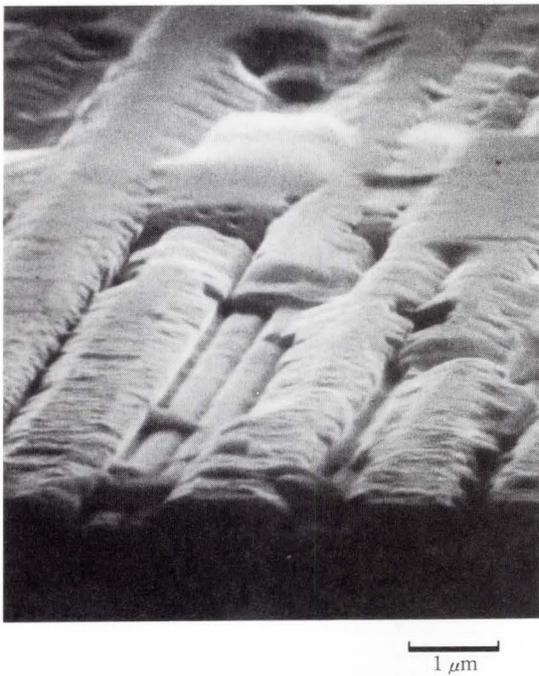
The surface morphology of the epitaxial layer on the corrugations was observed using a Nomarski interference microscope and SEM.

3.1.2 Results and discussion

The surface morphologies of InGaAsP epitaxial layers grown on the corrugations that were cleaned by using the above three methods are shown in Figs. 3, 4, and 5. Figures 3-a), 4 and 5 are observations using the Nomarski interference microscope. Figure 3-a) shows the surface of the epitaxial layer on the corrugations cleaned by method 1). Many pits were observed. Figure 3-b) observed by SEM shows



a) Photograph taken with the Nomarski interference microscope



b) Photograph taken with SEM

Fig. 3—The surface morphology of the epitaxial layer grown on the corrugated substrate prepared by using method 1).

the detail of the regions where smooth growth was inhibited. Figure 4 shows the surface of the epitaxial layer on the (100) InP substrate prepared by method 2). The pits appeared on the epitaxial layer only in the corrugated region, but did not appear on the flat region. Figure 5 shows the surface of the epitaxial layer on the corrugations cleaned by the method 3). It was

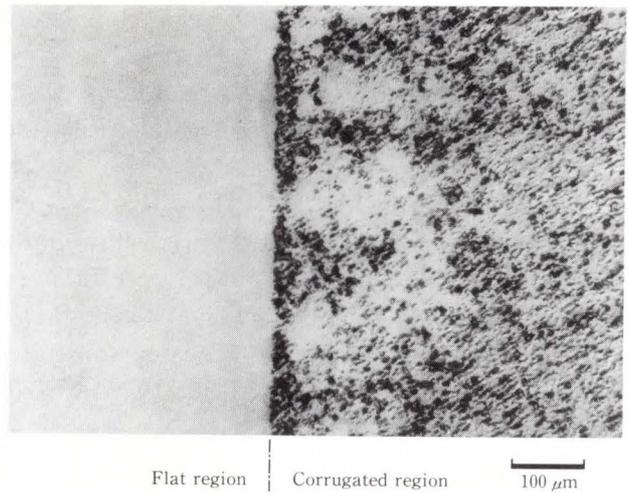


Fig. 4—The surface morphology of the epitaxial layer grown on the flat region and on the corrugated region of the substrate prepared by using method 2).

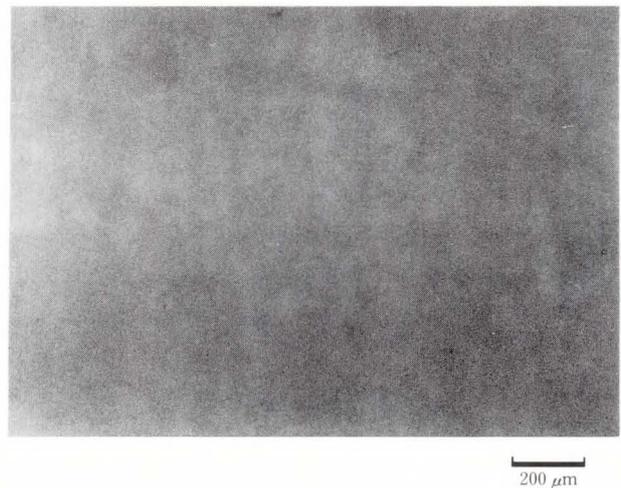


Fig. 5—The surface morphology of the epitaxial layer grown on the corrugated substrate prepared by using method 3).

found that no pits existed on the surface.

From these results, we deduced that the formation of the pits is related to the surface of the corrugations. Since the corrugations were formed by photo-lithography, it was considered that the origin of the pits was due to the particles of resist remaining on the corrugated surface (the resist is difficult to remove). Figure 5 shows the sample treated with the oxygen plasma process. A uniform epitaxial layer with no pits

indicates that the resist has been completely removed.

3.2 Reducing the degree of thermal deformation of the corrugations

3.2.1 Experiment

The InGaAsP ($\lambda_{Eg} \cong 1.3 \mu\text{m}$) followed by InP layers, or InGaAsP ($\lambda_{Eg} \cong 1.3 \mu\text{m}$) followed by InGaAsP ($\lambda_{Eg} \cong 1.55 \mu\text{m}$) layers were grown by LPE on the (100) InP corrugated substrate. The growth temperature was 585 °C, the soak temperature was 600 °C, the soak time was 30 min, and the cooling rate was 0.7 °C/min. This wafer was cleaned by the same method as in 3) of section 3.1.1.

To prevent thermal deformation of the corrugations during the growth, the following five kinds of cover plates were used:

- 1) An InP plate,
- 2) a GaAs plate with a surface roughened with an abrasive,
- 3) a GaAs plate with a mirror surface,
- 4) a GaAs plate that had already been used in the growth of a surface roughened with an abrasive,
- 5) a GaAs plate that had already been used as a cover plate with a rough surface and that was then further roughened with an abrasive.

The InGaAsP ($\lambda_{Eg} \cong 1.3 \mu\text{m}$) followed by InGaAsP ($\lambda_{Eg} \cong 1.55 \mu\text{m}$) layers were grown on

the corrugations by using the cover plates 1) and 2) above. The InGaAsP ($\lambda_{Eg} \cong 1.3 \mu\text{m}$) followed by InP layers were grown on the corrugations by using the cover plates 3), 4), and 5) above. Only the heat treatment (at 600 °C, for 30 min) was performed using the cover plate 1).

After LPE growth, the wafers were cleaved. The cleaved sections were stain-etched in a $\text{K}_3\text{Fe}(\text{CN})_6$, NaOH, and H_2O solution for SEM observation.

Table 1. Corrugation height before and after the LPE growth for several kinds of covers

	The corrugation height before growth (nm)	The corrugation height after growth (nm)
(1) an InP cover plate	200	< 50
(2) a GaAs cover plate with a rough surface	200	200
(3) a GaAs cover plate with a mirror surface	200	100
(4) a once used GaAs cover plate with a rough surface	200	80–100
(5) a roughened again GaAs cover plate which was once used	200	200

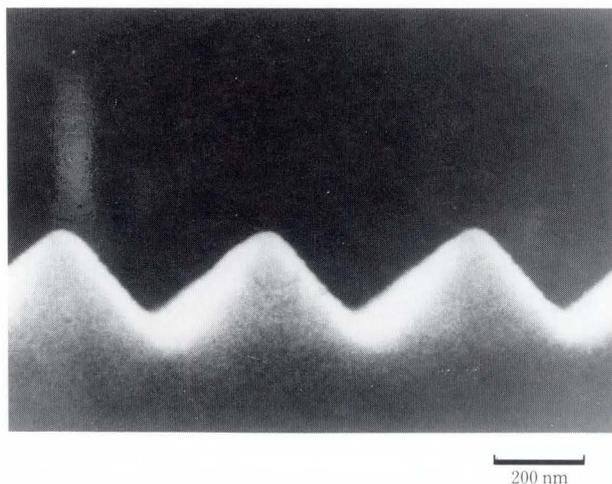


Fig. 6—SEM photograph of the corrugation before LPE growth.

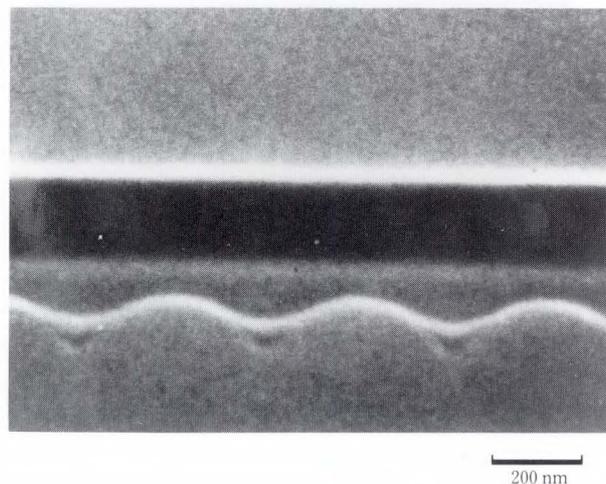


Fig. 7—SEM photograph of the corrugation after LPE growth for using the cover plate (1).

3.2.2 Results and discussion

Figure 6 shows an SEM photograph of a section of the corrugated wafer. Figures 7 and 9-12 show the SEM photographs of sections of the wafers after LPE growth. Table 1 lists the height of the corrugation before and after LPE growth for all cover plates. Using the cover plate (1), the corrugation height decreased from 200 nm to 50 nm after LPE growth. Figures 7 and 8 show the effect of thermal deformation. In Fig. 6 (before the growth) the shape of the corrugations is clearly triangular. It was known from Fig. 8 that the corrugations were deformed during the heating process before bringing the melt into contact with the corrugations. Similar thermal deformation was also observed in the V-grooves formed on an InP substrate²⁾. In that case, it was confirmed by energy dispersion X-ray spectroscopy that the deposited material in the bottom of the corrugations was also InP.

When the GaAs cover plate was used in LPE growth, the corrugations deformed very little as shown in Figs. 9-12. Thermal deformation using a GaAs cover plate was reported to be less than that using an InP cover plate^{3),4)}, where it was found that the GaAs cover plate with a rough surface could almost completely prevent thermal deformation of the corrugations.

Tanahashi et al. suggested²⁾ the following

model for thermal deformation of a V-groove. During etching in HCl to form the V-groove, chloride remained in the V-groove. Because of this chloride, InP was etched at the shoulder of the V-groove, and then was deposited at the bottom of the V-groove. However, as to cleaning with the oxygen plasma process, no residuums (e.g. chloride) remain on the corrugations.

Nagai et al. suggested⁵⁾ that P dissociated more easily at the top of the corrugation than at the bottom because the equilibrium vapor pressure of P was greater at the top than at the bottom. Therefore, the droplets of In were

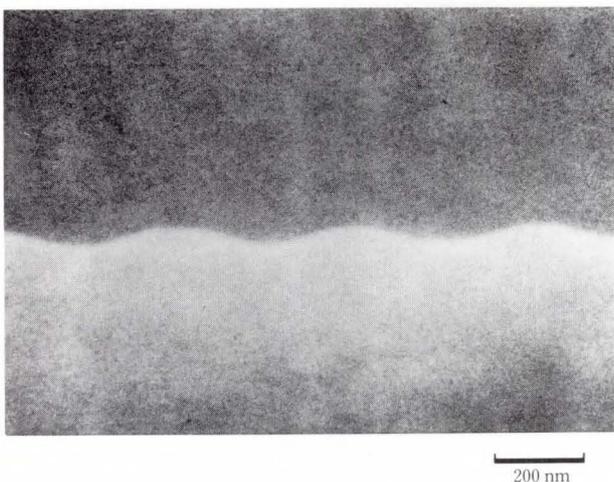


Fig. 8—SEM photograph of the corrugation after the heat treatment using the cover plate (1).

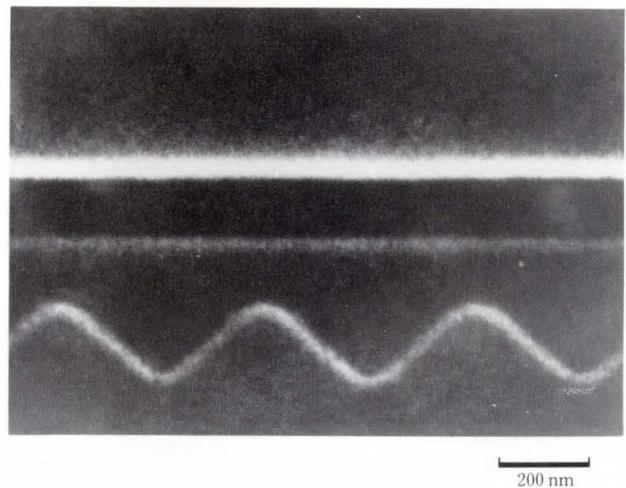


Fig. 9—SEM photograph of the corrugation after LPE growth for using the cover plate (2).

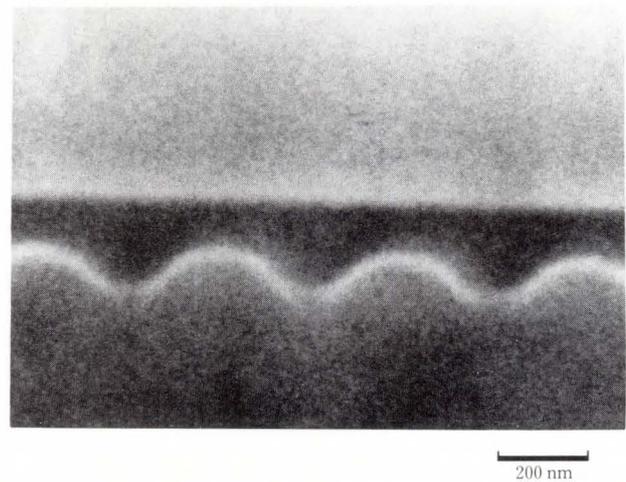


Fig. 10—SEM photograph of the corrugation after LPE growth using the cover plate (3).

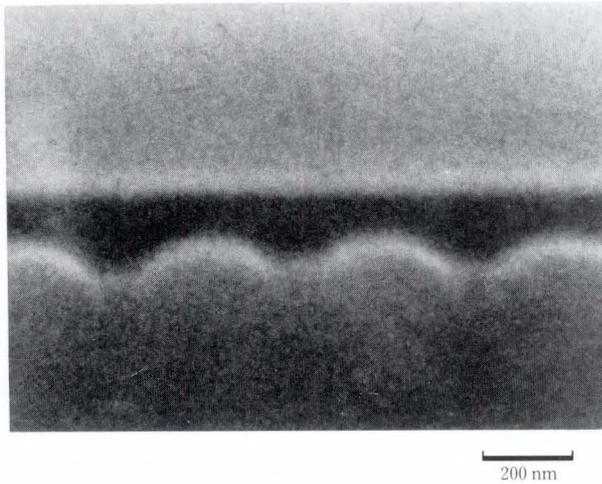


Fig. 11—SEM photograph of the corrugation after LPE growth using the cover plate (4).

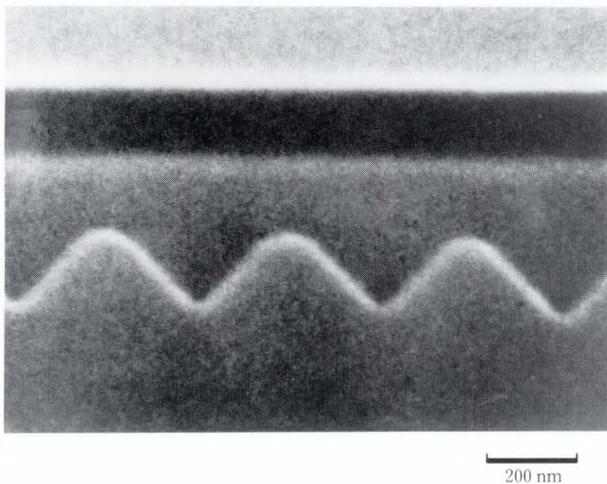


Fig. 12—SEM photograph of the corrugation after LPE growth using the cover plate (5).

generated at the top and migrated to the bottom. The In droplets combined with the P vapor to form InP at the bottom.

In this experiment the dissociation of P at the top of the corrugation was one of the causes of thermal deformation of the corrugations.

To confirm that the degree of thermal deformation was less when using a GaAs cover plate than when using an InP cover plate, a (100) InP substrate covered by a GaAs plate was thermally treated at 600 °C for 30 min. The surface of the InP substrate was analyzed by SIMS. Ga and As were detected in a thin

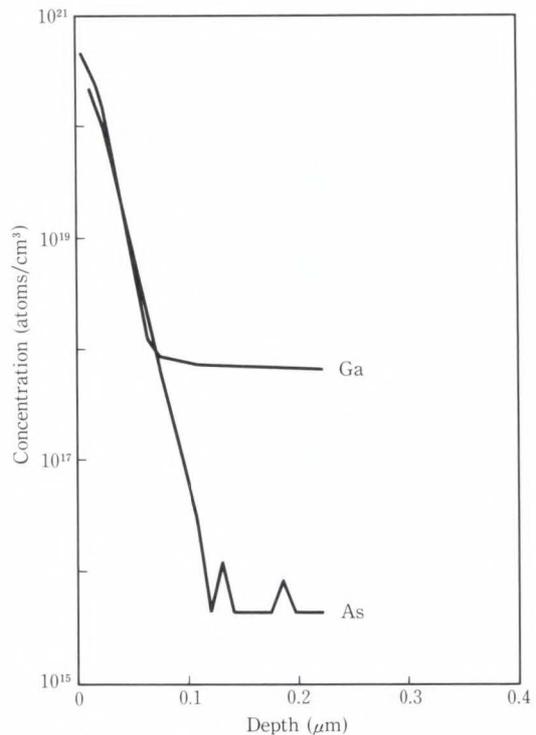


Fig. 13—SIMS analysis at the surface of an InP substrate after the heat treatment using the GaAs cover plate.

layer 0.1 μm under the surface, as shown in Fig. 13. These elements were also detected by X-ray fluorescence analysis. It was also reported⁴⁾ that a thin InGaAsP layer forming on the surface of the InP substrate was measured by XPS and Raman spectroscopic analyses. In our experiment we believed that the equilibrium vapor pressure of P decreased due to the formation of the thin InGaAsP layer on the surface of the corrugations covered by a GaAs plate so the thermal deformation of the corrugations decreased.

The corrugations were more deformed using a GaAs cover plate with a mirror surface (3) than using a roughened GaAs cover plate (2), as shown in Figs. 9 and 10. The reason why easy thermal deformation of the corrugations was prevented when using (2) was that the P vapor pressure at the surface of the corrugations decreased less than the P vapor pressure at the surface of the corrugation when using (3). This decrease of the P vapor pressure was caused by an amount of InGaAsP formed on the surface

of the corrugation when using (2) was more than that when using (3) because the surface area of GaAs cover (2) was more than that of GaAs cover (3). The corrugations were more deformed when using (4) a roughened GaAs plate that had already been used once than when using a roughened GaAs plate (2) (as shown in Figs. 9 and 11), because the As vapor pressure was lower at the surface of GaAs plate (2). The roughened GaAs plate after being used once could be made into (5), a useful cover plate by roughening it again (as shown in Fig. 12). The important point is that the cover plate must have a large As vapor pressure at the surface of the GaAs plate.

3.3 Mesa-etching

The novel etchants different from the reported ones⁶⁾ were developed for mesa-etching. They were mixtures of HBr (47 wt%), H₂O₂ (31 wt%), and H₂O. The InP etching rate was measured using this etchant. Three compositions of the etchant were made (4HBr/2H₂O₂/20H₂O), (3HBr/1.5H₂O₂/20H₂O), and (3HBr/1.5H₂O₂/30H₂O). These etchants were made by mixing HBr and H₂O, then adding H₂O₂ to this solution and allowing the solution to stand for a set period of time. A wax mask was then

patterned on a (100) InP substrate. The solutions were allowed to stand for 1, 5, 10, 15 and 20 minutes after H₂O₂ was added. The etching time in each case was one minutes. The wax was removed by trichlorethylene. The height of the step formed by etching the wafer was measured by an interference microscope.

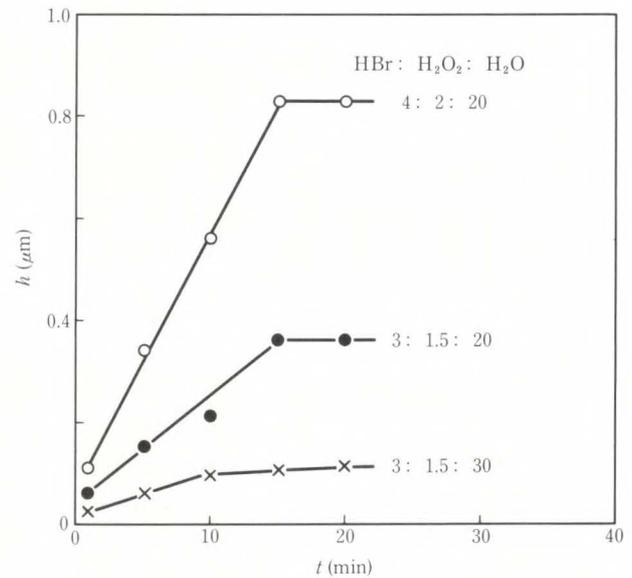


Fig. 14—Etching characteristics for a (100) InP substrate etched with HBr, H₂O₂, and H₂O etchant.

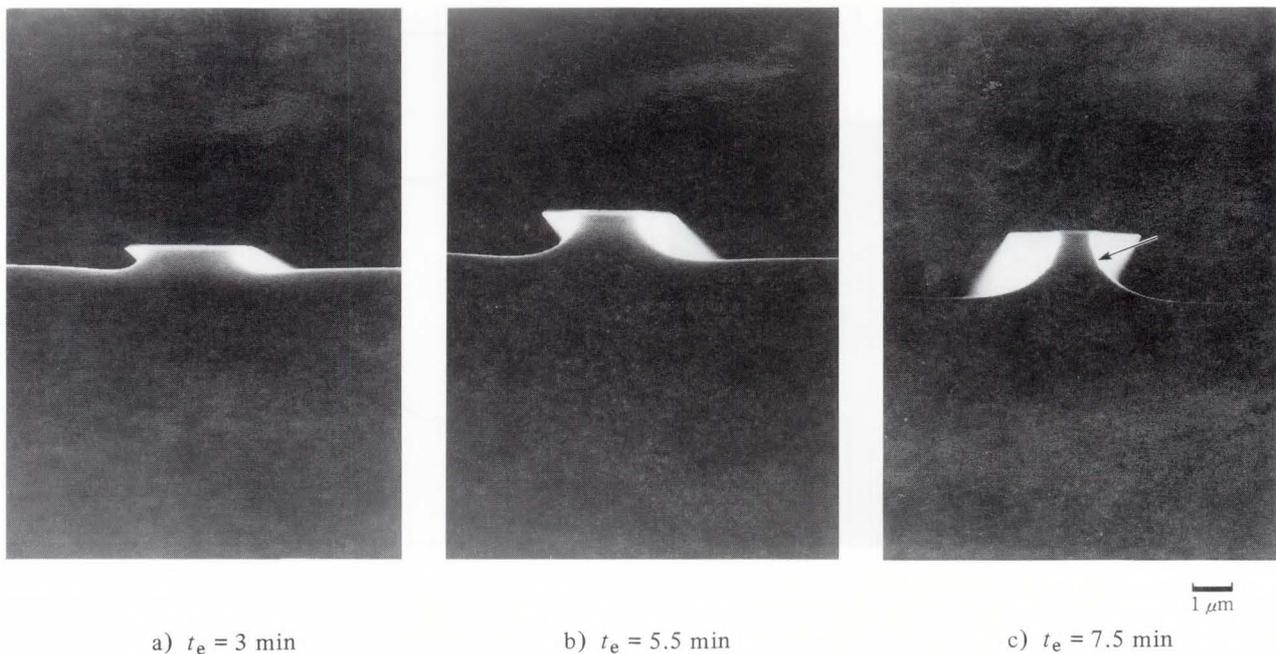
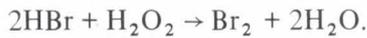


Fig. 15—Sections of mesa-etched wafers of the FBH-DFB lasers.

Figure 14) shows the height of step h , for one minute of etching as a function of t (the time allowed to stand before inserting the wafer into the solution). For the three kinds of etchants, h was zero at $t = 0$ min, and increased with t . However, h was constant after $t = 15$ min. The temperature of the etchant increased gradually. The etchants were colorless when H_2O_2 was added and the color then changed from yellow to orange. Finally the color approached that of Br_2 .

From these results the following heating reaction is considered:



This reaction generates heat and InP and InGaAsP are etched by produced Br_2 . This reaction slowly progresses and the concentration of Br_2 slowly increases. Therefore h gradually increases.

An approximately $3 \mu m$ wide SiO_2 mask was patterned in the $\langle 110 \rangle$ direction on the wafer of Fig. 1-a). This wafer was etched with the HBr, H_2O_2 , and H_2O etchant with a volume ratio of 4:2:20. Three etching times of 3, 5.5, and 7.5 minutes were used. The sections of the etched wafers were observed by SEM.

Figures 15-a), -b), and -c) show sections of the wafer observed by SEM for etching times, (t_e), of 3, 5.5, and 7.5 minutes. Figure 16 is an

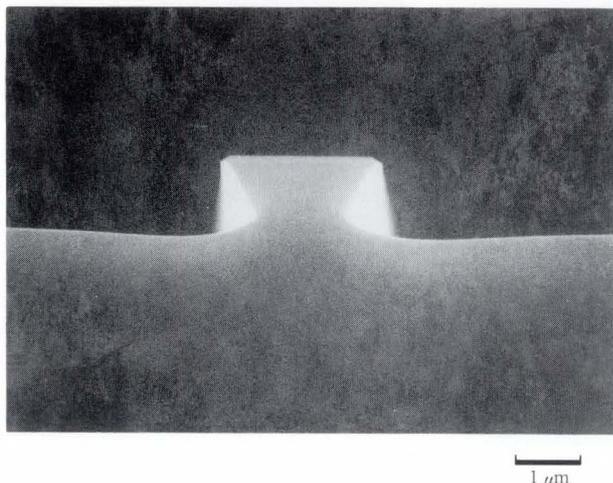


Fig. 16—Section of the mesa that is formed by etching with the Br_2 and CH_3OH solution.

SEM photograph of a section of a wafer etched with a Br_2 and CH_3OH solution. Figure 16 shows the (111)A plane on the sides of the etched mesa structure. However, the (111)A plane was not formed using this etchant as shown in Figs. 15-a), -b), and -c). Figures 15-a) and -b) don't show the notches on the sides of the mesa structure, but the notches appear on the active layer on the sides of the mesa structure { indicated by the arrow in Fig. 15-c) }. The etching rate for InGaAsP appears to be greater than that for InP.

The diagram in Fig. 17 shows the etched depth, t_e the width of the SiO_2 mask, W_{SiO_2} and the width of the mesa at the top, W_{mt} . The length of the eaves of the SiO_2 mask L_u , was calculated from the following equation:

$$L_u = (W_{SiO_2} - W_{mt})/2.$$

Figure 17 shows L_u versus d_e . The relationship between L_u and d_e is linear.

To define the mesa structure, the origin of the d - L coordinate system was placed at the tip of the SiO_2 mask as shown in the diagram in Fig. 18. The dimensions d and L were measured by SEM analysis. The center of the three circular arcs shown in Fig. 18 is $d = 0 \mu m$ and $L = 0 \mu m$; i.e. the tip of the SiO_2 mask. The side of the

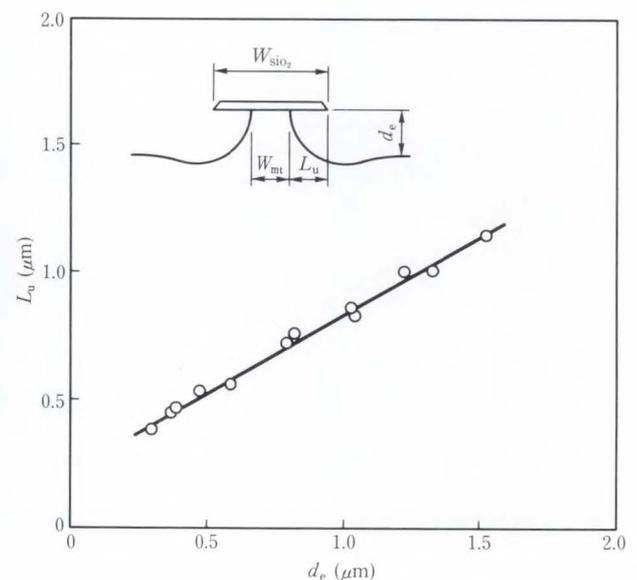


Fig. 17—Mesa characteristics of the FBH-DFB laser, L_u versus d_e .

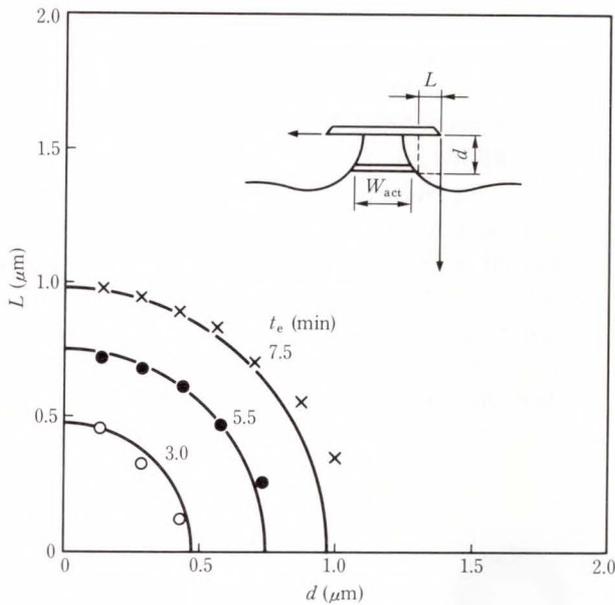


Fig. 18—Mesa characteristics, L versus d .

mesa is circular, therefore, the width of the active layer, W_{act} as shown in Fig. 18 is given by the following equation:

$$W_{act} = W_{SiO_2} - 2\sqrt{L_u^2 - (d_c + d_{cap})^2}.$$

Where d_c is the thickness of the p-InP cladding layer and d_{cap} is the thickness of the p-InGaAsP cap layer in Fig. 1-a). For example, for $d_c + d_{cap} = 0.4 \mu\text{m}$, $L_u = 0.9 \mu\text{m}$ and $W_{SiO_2} = 2.61 \mu\text{m}$, W_{act} is $1.0 \mu\text{m}$. Since the relationship between L_u and d_e is defined in Fig. 17, W_{act} can be controlled by adjusting W_{SiO_2} , $d_c + d_{cap}$, and d_e .

The wafer was etched until the active layer became $1 \mu\text{m}$ wide. The etched wafer was buried by the LPE growth according to the procedure in Fig. 1 and was made into the FBH-DFB laser as shown in Fig. 2.

Figure 19 shows a section of the wafer buried by the procedure shown in Fig. 1. W_{act} was about $1.0 \mu\text{m}$. The threshold current, (I_{th}), of these FBH-DFB lasers developed was 19 mA to 37 mA (25 mA average). I_{th} of the mesa etched with the Br_2 and CH_3OH etchant was more than 40 mA. The reason why I_{th} was lower because the (111)A plane was not formed on the mesa sides.

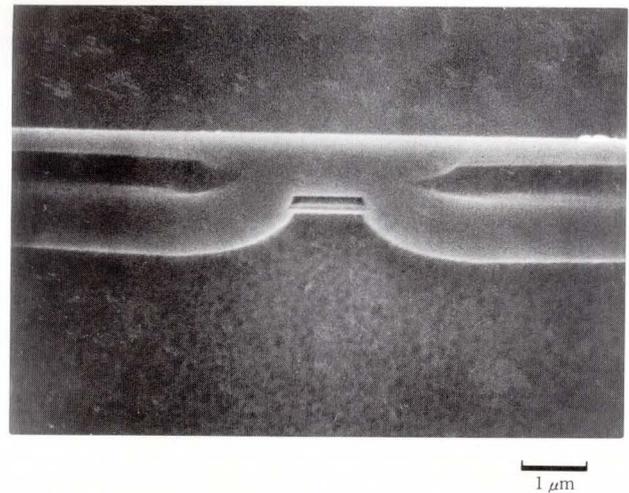


Fig. 19—Section of the FBH-DFB laser after the third growth.

4. Conclusions

It was found that the amount of resist remaining on the corrugated substrate depends on the method used to prepare the surface of the corrugated substrate before growth. The resist residue was almost completely removed by the oxygen plasma process.

The thermal deformation of the corrugated substrate can be prevented by using an appropriate cover plate. The best cover plate is the GaAs plate with a surface roughened with an abrasive.

It was found that the HBr , H_2O_2 , and H_2O etchant is useful for making the mesa structure of the FBH-DFB lasers. The (111)A plane is not formed on the sides of the mesa using this etchant. The width of the active layer can be controlled by adjusting the width of the SiO_2 mask, the thicknesses of the p-InP cladding, the p-InGaAsP cap layers, and the etched depth. The average threshold current, I_{th} of our FBH-DFB lasers were down to a 25 mA.

5. Acknowledgement

We would like to thank the many people of the Optical Semiconductor Devices Laboratory who assisted us with our work.

References

- 1) Logan, R.A., Temkin, H., Merritt, F.R. and Mahajan,

- S.: GaInAsP/InP Buried Heterostructure Formation by Liquid Phase Epitaxy. *Appl. Phys. Lett.*, **45**, 12, pp. 1275-1277 (1984).
- 2) Tanahashi, T., Ushijima, I., Umebu, I., Nakai, S. and Akita, K.: Deformation of V-groove on InP Substrate by Heat Treatment. *J. Cryst. Growth*, **64**, 3, pp. 492-498 (1983).
 - 3) Kinoshita, J., Okuda, H. and Uematsu, Y.: Preserving InP Surface Corrugations for 1.3 μm GaInAsP/InP DFB Lasers from Thermal Deformation during LPE Process. *Electron. Lett.*, **19**, 6, pp. 215-216 (1983).
 - 4) Kubo, M., Koga, K., Kohiki, S., and Ogura, M.: Characterization of InGaAsP Surface Corrugation by Means of Raman Spectroscopy. (in Japanese), *Papers Tech. Group IECE, Jpn., Solid State Devices*, **SSD 85-164**, pp. 73-79 (1986).
 - 5) Nagai, H., Noguchi, Y., and Matsuoka, T.: Thermal Deformation of Surface Corrugation on InGaAsP Crystal. *J. Cryst. Growth*, **71**, 1, pp. 225-231 (1985).
 - 6) Adachi, S., Noguchi, Y., and Kawaguchi, H.: Chemical Etching of InGaAsP/InP DH Wafer. *J. Electrochem. Soc.: SOLID-STATE-SCIENCE AND TECHNOLOGY*. **129**, 5, pp. 1053-1062 (1982).



Toshihiro Kusunoki

Optical Semiconductor Devices
Laboratory
FUJITSU LABORATORIES, ATSUGI
FUJITSU Technical School 1976
Specializing in Optical Semiconductor
Devices



Toshiyuki Tanahashi

Optical Semiconductor Devices
Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
The University of Tokyo 1976
Master of Physics
The University of Tokyo 1978
Specializing in Optical Semiconductor
Devices



Nirou Okazaki

Optical Semiconductor Devices
Laboratory
FUJITSU LABORATORIES, ATSUGI
Hyogo Technical High School 1970
Specializing in Optical Semiconductor
Devices

DX-Center-Free GaAs/N-AlGaAs HEMT Structures

• Tomonori Ishikawa • Kazuo Kondo

(Manuscript received December 1, 1987)

The DX centers in N-Al_xGa_{1-x}As layers grown by MBE were investigated. The concentration ratio of DX centers to total donors was determined for various compositions of AlGaAs and doping concentrations of Si. Based on these results, DX-center-free selectively doped GaAs/N-AlGaAs heterostructures were proposed and successfully applied to HEMTs.

1. Introduction

N-type dopants in Al_xGa_{1-x}As layers form two types of donors, shallow donors mainly in the composition region below $x = 0.2$ and deep donors, called DX centers, above $x = 0.2$ ¹⁾⁻³⁾. The DX centers exhibit peculiar properties such as a persistent photo-conductivity effect, and has been believed to be the complexes of donors (D) and unknown defects (X) because of a large lattice relaxation upon charge capture⁴⁾. Since the advent of high electron mobility transistors (HEMTs) in 1980⁵⁾, much attention has been paid to the DX centers and it has become apparent that the DX centers cause some serious device instabilities such as drain current collapse^{6),7)}.

To overcome such undesirable effects, a more detailed study on the nature and origin of DX centers must be performed. In section 2, the DX center concentrations in N-Al_xGa_{1-x}As layers are determined in detail and the origin is discussed. Based on these results, DX-center-free selectively doped GaAs/N-AlGaAs heterostructures are proposed in section 3.

2. DX center in AlGaAs

To determine the DX center concentration in the N-Al_xGa_{1-x}As layers, 1- μ m-thick Si-doped Al_xGa_{1-x}As layers were grown by conventional MBE on semi-insulating GaAs substrates at a substrate temperature of 680 °C. They in-

cluded 0.2- μ m-thick undoped GaAs and 0.3- μ m-thick undoped Al_xGa_{1-x}As buffer layers to prevent two-dimensional-electron-gas (2DEG) formation. The doping concentration was varied from $8 \times 10^{16} \text{ cm}^{-3}$ to $3 \times 10^{18} \text{ cm}^{-3}$. The AlAs mole fraction was determined from the thickness of the AlAs and GaAs epitaxial layers grown just before the sample growth. It was also determined by X-ray diffraction and the results proved to be consistent. The accuracy was within ± 2.5 percent.

For N-Al_xGa_{1-x}As layers with various compositions and doping concentrations, the concentrations of shallow donors and DX centers were estimated by Hall measurement at 77 K. Figure 1 shows electron concentration in darkness (solid circles) and after five minutes of turning off exposing light (open circles) as a function of reciprocal temperature of the Si effusion cell during MBE growth for $x = 0.22$ and $x = 0.19$. For $x = 0.22$, a considerable persistent-photoconductivity (PPC) effect was observed for all doping concentrations. The electron concentration increased considerably after being exposed to light and persisted after the light source was removed. The electron concentration in darkness implies a shallow donor concentration which can ionize at 77 K. On the other hand, the persistent increase in the electron concentration after light exposure is thought to indicate the DX center

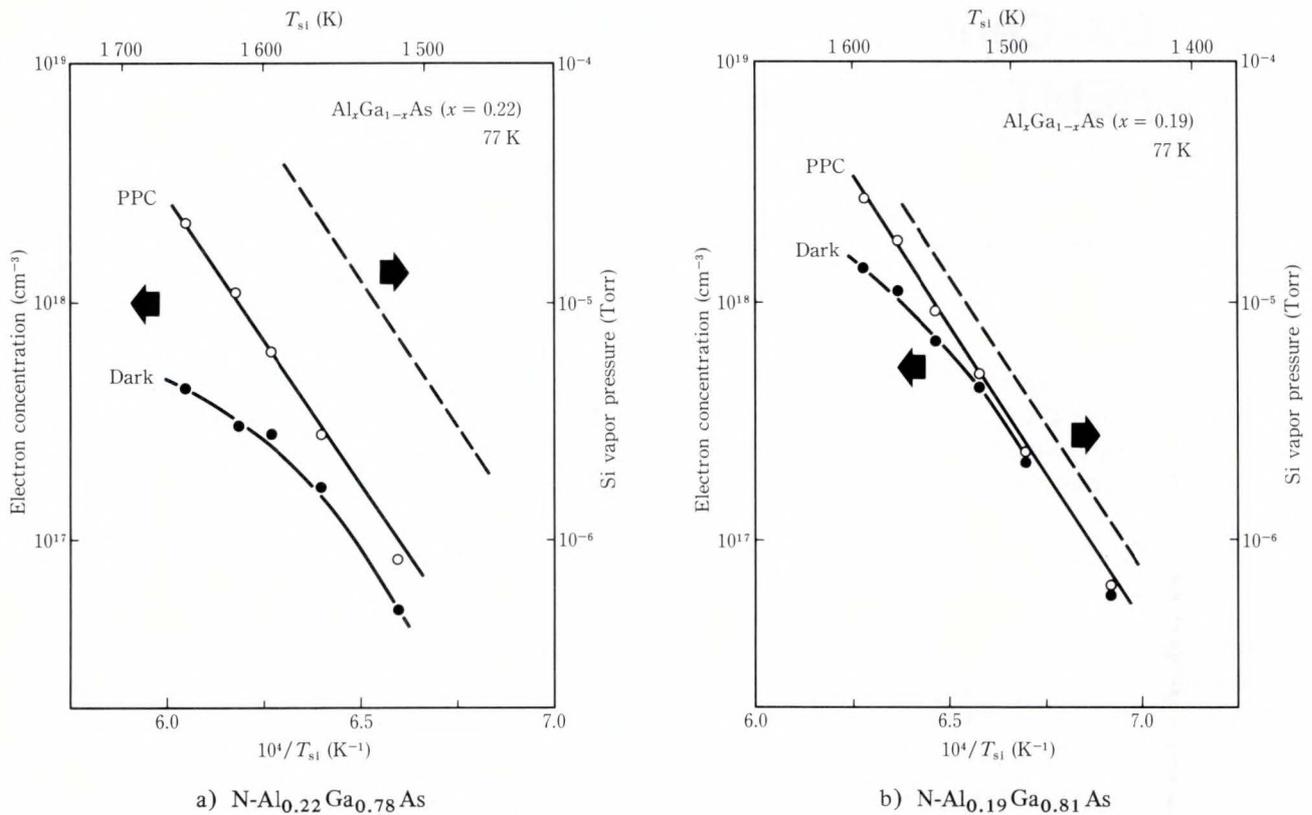


Fig. 1—Electron concentration at 77 K in darkness (solid circles) and after light exposure (open circles) as a function of reciprocal temperature of the Si effusion cell.

concentration, because the DX centers have a very small capture cross section at 77 K⁸). For $x = 0.19$, persistent increase in electron concentration was not observed for the samples with a doping concentration of less than $3 \times 10^{17} \text{ cm}^{-3}$, indicating no DX centers. By means of heavy Si doping to a concentration of more than $5 \times 10^{17} \text{ cm}^{-3}$, however, the PPC effect due to DX centers was observed. Around $x = 0.2$, therefore, the formation of DX centers is thought to depend on the doping concentration, and a considerable amount of DX centers is formed by heavy doping, even for compositions of less than $x = 0.2$. It should also be noted that the electron concentration after light exposure is proportional to the Si vapor pressure (broken line) both for $x = 0.22$ and $x = 0.19$. This indicates that all incorporated Si atoms act as either kind of donor and that the auto-compensation effect does not seem to be important.

To assure DX center formation in the

samples of $x = 0.19$, DLTS measurement was performed. Figure 2 shows DLTS spectra for four samples with a doping concentration, N_D , of less than $9 \times 10^{17} \text{ cm}^{-3}$, which had sufficiently low leakage current for Schottky contacts. While no DLTS signal for the electron trap was observed in the lightly doped sample of $8 \times 10^{16} \text{ cm}^{-3}$, a small signal appeared at $3 \times 10^{17} \text{ cm}^{-3}$ and increased significantly as the doping concentration was increased. This signal had the characteristic two-peak feature often observed for DX centers, and the activation energy of electron emission obtained from this signal is 0.47 eV, which is almost equal to that of the DX centers in the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ layer (0.43-0.45 eV^{2),8),9}). Additionally, the electron trap concentration estimated from this DLTS signal was proportional to the persistent increase in the electron concentration as shown in Fig. 1. Therefore, it can be concluded that DX centers are formed by heavy Si doping in the $\text{N-Al}_{0.19}\text{Ga}_{0.81}\text{As}$ layer and

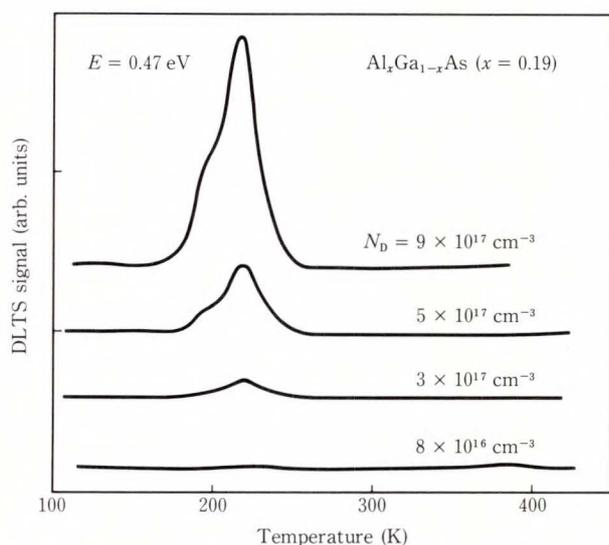


Fig. 2—DLTS spectra for N-Al_{0.19}Ga_{0.81}As layers with various doping concentrations.

causes the PPC effect.

Also for Al_{0.3}Ga_{0.7}As and Al_{0.15}Ga_{0.85}As layers with the various Si concentrations, N_{Si} , the concentration of DX center, N_{DX} , was estimated from the Hall measurement at 77 K in the same manner as in Fig. 1. Figure 3 shows the concentration ratio of DX centers to total donors as a function of the total donor concentration. Both for $x = 0.22$ and 0.15 , the ratio was also found to be dependent on the doping concentration as with $x = 0.19$, but was higher for $x = 0.22$ and lower for $x = 0.15$. It should be noted that even in the N-Al_{0.15}Ga_{0.85}As layer, which has been believed to include no DX centers, the PPC effect was observed in the heavily doped samples of more than $2 \times 10^{18} \text{ cm}^{-3}$. Recently, Mooney et al. reported the observation of DX centers even in GaAs with very heavily doped Si¹⁰⁾.

Since the DX center concentration ratio was proved to be independent of the MBE growth conditions¹¹⁾, the DX centers do not seem to include any of the native defects suggested by Lang et al.⁴⁾, and are thought to be intrinsic to N-Al_{*x*}Ga_{1-*x*}As. As seen in Fig. 3, moreover, the concentration ratio of DX centers to the total donors is determined both by the composition and the doping concentration. These results suggest the important role of the band structure

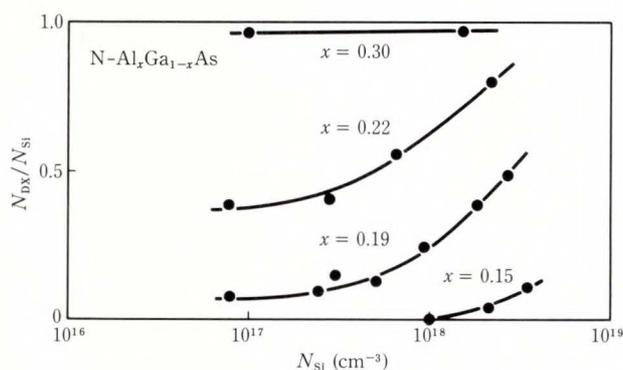


Fig. 3—Concentration ratio of DX centers to total donors as a function of the total donor concentration.

in the formation of DX centers. Thus we try to explain our results as follows. The DX centers in AlGaAs are believed to be deep levels associated with the L-valley^{12),13)} or X-valley⁴⁾. Around $x = 0.2$, the energy level of the DX centers associated with such valleys is expected to be just above the bottom of the Γ -valley. Therefore, when the doping concentration is light, electrons mainly occupy the Γ -valley. By increasing the doping concentration, however, the Fermi level rises and electrons will begin to populate the DX center levels. For the doping concentration of more than $1 \times 10^{18} \text{ cm}^{-3}$, the DX centers will become dominant. For $x = 0.15$, the energy level of the DX centers is expected to be further above the Γ -minimum, so the DX centers appear only at higher doping concentrations. For Al_{*x*}Ga_{1-*x*}As above $x = 0.3$, in contrast, the DX centers are dominant independently of doping concentration because its energy level is considered to be much lower than Γ -minimum. Based on these interpretations, DX center formation dependent on doping concentration can be observed in other alloy systems like the candidates suggested by Tachikawa et al.¹³⁾, where the DX center level is very close to Γ -minimum. In S-doped Ga_{0.52}In_{0.48}P grown by chloride VPE, Kitahara et al.¹⁴⁾ reported donor related deep centers which have very similar characteristics to those of the DX centers in AlGaAs. The concentration ratio of these centers to the donor total in GaInP also proved

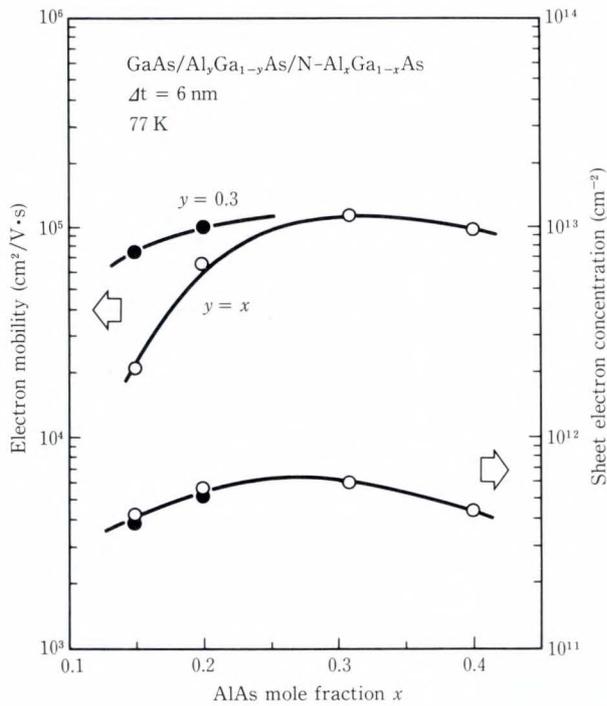


Fig. 4—2DEG characteristics at 77 K in a conventional selectively doped heterostructure (open circle) and in the modified structure (solid circles) as a function of the AlAs mole fraction in the Si-doped layers.

to increase with the doping concentration, and can be explained in the same way as the present case.

3. DX-center-free GaAs/N-AlGaAs heterostructures

As discussed in the previous section, a considerable number of DX centers were formed even in N-Al_xGa_{1-x}As layers with the composition of less than $x = 0.2$ by heavy doping above $1 \times 10^{18} \text{ cm}^{-3}$, which is common in HEMTs. To avoid the influence of DX centers, therefore, the AlAs mole fraction x in the Si-doped layers of HEMTs must be reduced to less than 0.15, where DX center concentration becomes negligible. In this section, we examine the two-dimensional electron gas (2DEG) characteristics of the modified heterostructures with N-Al_{0.15}Ga_{0.85}As layers and thin undoped AlGaAs barrier layers, and also the characteristics of HEMTs fabricated on such heterostructures.

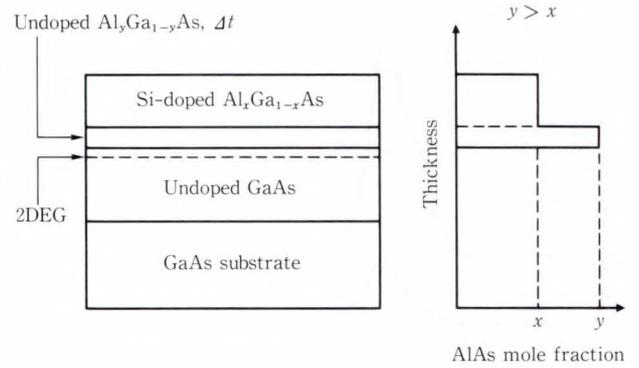


Fig. 5—Structure of the modified selectively doped GaAs/N-AlGaAs heterostructure.

First, the dependence of 2DEG characteristics in the conventional selectively doped heterostructures on the AlAs mole fractions of the Si-doped layers is discussed. The result is shown by the open circles in Fig. 4. These samples were also grown by conventional MBE. Undoped spacer layers are 6 nm thick. With a reduced AlAs mole fraction of less than 0.2, the mobility was greatly reduced and the electron concentration was decreased slightly. This reduction in mobility is due to the increased penetration of the 2DEG wave function into the low mobility AlGaAs layers, because of a low barrier height at the heterointerface. For a DX-center-free composition of $x = 0.15$, the mobility was as low as $20\,000 \text{ cm}^2/\text{V}\cdot\text{s}$, even with a 6-nm spacer.

To prevent the penetration of 2DEG into AlGaAs layers having a low Al content, thin undoped AlGaAs barrier layers were introduced at the heterointerface¹⁵⁾. Figure 5 shows the modified structure used in this experiment. The mole fraction of the Si-doped layers, x , was decreased to reduce the DX center concentration. The mole fraction of the undoped barrier layers, y , was larger than x to prevent the penetration of the 2DEG wave function. The solid circles in Fig. 4 show the mobility and sheet concentration of 2DEG in such modified structures, in which the mole fraction of the barrier layers was fixed to 0.3 and the thickness to 6 nm. By introducing an AlGaAs barrier layer, the 2DEG mobility was improved. This

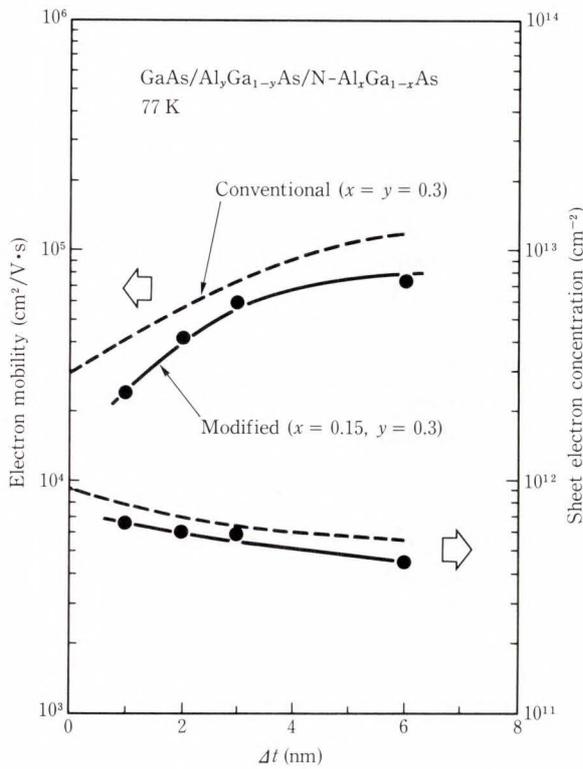


Fig. 6—2DEG characteristics at 77 K in the modified structure as a function of barrier thickness.

indicated that barrier layers are very effective in preventing penetration of 2DEG into the AlGaAs layers. Mobility was especially improved for $x = 0.15$, which gave a value as high as $70\,000\text{ cm}^2/\text{V}\cdot\text{s}$ as shown in Fig. 4.

Since a high carrier concentration is desirable for a practical application of HEMTs, the 2DEG characteristics in the modified structure of $x = 0.15$ with a thinner barrier was examined. Figure 6 shows the mobility and sheet concentration of 2DEG at 77 K in the modified structure as a function of the barrier layer thickness. For comparison, the data for a conventional structure with an AlAs mole fraction of $x = y = 0.30$ is shown by the broken lines. Although the electron concentration in the modified structure is less than that in a conventional structure, a relatively high mobility was obtained. For a thickness of 2 nm or 3 nm, the barrier layers seem to be effective, and an electron mobility of more than $40\,000\text{ cm}^2/\text{V}\cdot\text{s}$ with a sheet concentration of more than $6 \times 10^{11}\text{ cm}^{-2}$ was obtained.

Table 1. HEMT characteristics of the modified and conventional structures

	Structural parameters			300 K		77 K	
	Δt (nm)	x	y	g_m (mA/mm)	K ($\text{mA}/\text{V}^2 \cdot \text{mm}$)	g_m (mA/mm)	K ($\text{mA}/\text{V}^2 \cdot \text{mm}$)
Modified	3	0.15	0.30	170	170	230	540
Conventional	3	0.30	0.30	180	170	260	570

$$L_G = 1.2\ \mu\text{m}, L_W = 50\ \mu\text{m}$$

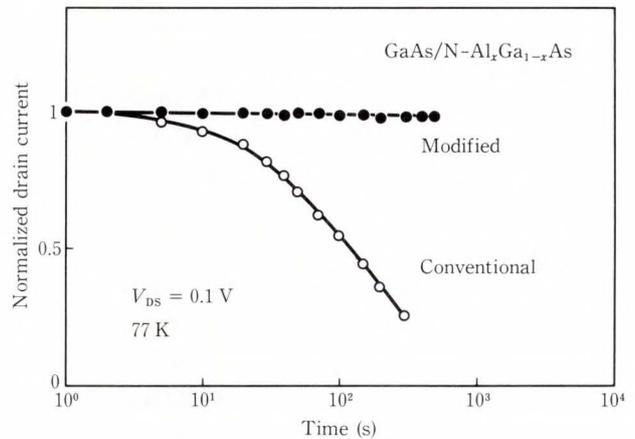


Fig. 7—Normalized drain current in the modified and conventional HEMT at 77 K as a function of operating time.

The HEMTs fabricated using the modified structure with 3 nm-thick barrier layer had good performance. Table 1 lists the performance values at room temperature and at 77 K without light for conventional and modified HEMTs. Their structural parameters are also given. The gate was $1.2\ \mu\text{m}$ long and $50\ \mu\text{m}$ wide. Good performance, comparable to that of the conventional structure, was obtained for the modified structure. With decreased temperature, the K-value showed excellent enhancement for both structures; this may be due to the enhanced mobility at low temperatures. This modified structure also prevents drain current collapse because of a low concentration of DX centers in the AlGaAs layer. The time dependence of the normalized drain current at 77 K is shown by the solid circles in Fig. 7. For comparison, the result for the HEMT

of conventional structure with $x = 0.19$ is also shown by the open circles. The drain current at $V_{DS} = 0.1$ V was measured since collapse is most remarkable in the small drain voltage region. While the drain current of the conventional structure collapsed after several seconds, that of the modified structure was stable even after 500 s of operation. Thus, this structure is promising for application to HEMT ICs with high reliability at low temperatures.

4. Conclusion

The DX center concentration in N-Al_xGa_{1-x}As layers were investigated in detail by Hall measurement and DLTS measurement. For the composition above $x = 0.3$, all the Si donors form DX centers. Around $x = 0.2$, on the other hand, the DX-center-to-total-donor concentration ratio was found to increase with increasing doping concentration of Si. These results suggest the DX centers originate from the deep levels associated with the L- or X-valleys.

The 2DEG mobility in selectively doped heterostructures with an Al content of less than $x = 0.2$ was found to be improved by introducing AlGaAs barrier layers at the hetero-interface. Applying these modified structures with a Si-doped layer of $x = 0.15$ to HEMTs, a good performance without drain current collapse was obtained.

5. Acknowledgement

The authors thank Messrs. T. Yokoyama, M. Suzuki and T. Mimura for their valuable discussions and Messrs. T. Maeda and T. Yamamoto for their work on MBE growth. They also thank Messrs. M. Abe, J. Komeno, A. Shibatomi and M. Kobayashi for their encouragement and support. The present research effort is the National Research and Development Program on the "Scientific Computing System", conducted under a program set by the Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Japan.

References

1) Ishikawa, T., Saito, J., Sasa, S., and Hiyamizu, S.: Electrical Properties of Si-Doped Al_xGa_{1-x}As

Layers Grown by MBE. *Jpn. J. Appl. Phys.*, **21**, 11, pp. L675-L676 (1982).

2) Watanabe, M.O., Morizuka, K., Mashita, M., Ashizawa, Y., and Zohta, Y.: Donor Levels in Si-Doped AlGaAs Grown by MBE. *Jpn. J. Appl. Phys.*, **23**, 2, pp. L103-L105 (1984).

3) Schubert, E.F., and Ploog, K.: Shallow and Deep Donors in Direct-gap N-type Al_xGa_{1-x}As: Si Grown by Molecular Beam Epitaxy. *Phys. Rev. B*, **30**, 12, pp. 7021-7029 (1984).

4) Lang, D.V., Logan, R.A., and Jaros, M.: Trapping characteristics and a donor-complex (DX) model for the persistent-photoconductivity trapping center in Te-doped Al_xGa_{1-x}As. *Phys. Rev. B*, **19**, 2, pp. 1015-1030 (1979).

5) Mimura, T., Hiyamizu, S., Fujii, T., and Nanbu, K.: A new field-effect transistor with selectively doped GaAs/n-Al_xGa_{1-x}As heterojunction. *Jpn. J. Appl. Phys.*, **19**, 5, pp. L225-L227 (1980).

6) Fischer, R., Drummond, T.J., Kopp, W., Morkoc, H., Lee, K., and Shur, M.S.: Instabilities in modulation doped field-effect transistors (MODFETs) at 77 K. *Electron Lett.*, **19**, 19, pp. 789-791 (1983).

7) Kastalsky, A., and Kiehl, R.A.: On the low-temperature degradation of (AlGa) As/GaAs modulation-doped field-effect transistors. *IEEE Trans. Electron Devices*, **ED-33**, 3, pp. 414-423 (1986).

8) Zhou, B.L., Ploog, K., Gmelin, E., Zheng, X.Q., and Schulz, M.: Assessment of persistent-photoconductivity centers in MBE grown Al_xGa_{1-x}As using capacitance spectroscopy measurements. *Appl. Phys. A*, **28**, pp. 223-227 (1982).

9) Kumagai, O., Kawai, H., Mori, Y., and Kaneko, K.: Chemical trends in the activation energies of DX centers. *Appl. Phys. Lett.*, **45**, 12, pp. 1322-1323 (1984).

10) Mooney, P.M., Theis, T.N., and Wright, S.L.: The role of DX centers in limiting the free carrier density in GaAs. to be published in the Proceeding of 14th International Symposium on GaAs and Related Compound, (1987).

11) Ishikawa, T., Kondo, K., Hiyamizu, S., and Shibatomi, A.: Influence of MBE growth conditions on persistent photoconductivity in N-AlGaAs and selectively doped GaAs/N-AlGaAs heterostructures. *Jpn. J. Appl. Phys.*, **24**, 6, pp. L408-L410 (1985).

12) Mizuta, M., Tachikawa, M., Kukimoto, H., and Minomura, S.: Direct evidence for the DX center being a substitutional donor in AlGaAs Alloy system. *Jpn. J. Appl. Phys.*, **24**, 2, pp. L143-L146 (1985).

- 13) Tachikawa, M., Mizuta, M., Kukimoto, H., and Minomura, S.: A simple calculation of the DX center concentration based on an L-donor model. *Jpn. J. Appl. Phys.*, **24**, 10, pp. L821-L823 (1985).
- 14) Kitahara, K., Hoshino, M., Kodama, K., and Ozeki, M.: Dnor-related deep level in S-doped Ga_{0.52}In_{0.48}P grown by chloride VPE. *Jpn. J. Appl. Phys.*, **25**, 7, pp. L534-L536 (1986).
- 15) Ishikawa, T., Saito, J., Sasa, S., and Hiyamizu, S.: The electrical effect of AlGaAs barrier at the hetero-interface of selectively doped GaAs/n-AlGaAs heterostructure. Extended Abstracts of the 16th Conference on Solid State Devices and Materials, Kobe, 1984, pp. 603-606.



Tomonori Ishikawa

Compound Semiconductor Devices
Division
FUJITSU LIMITED
Bachelor of Physics
Kanazawa University 1974
Dr. of Electronic Eng.
Shizuoka University 1980
Specializing in Semiconductor Materials



Kazuo Kondo

Compound Semiconductor Devices
Division
FUJITSU LIMITED
Bachelor of Electronic Eng.
Osaka University 1972
Dr. of Electronic Eng.
Osaka University 1977
Specializing in Semiconductor Materials

Fast Ray Tracing

• Koichi Murakami • Katsuhiko Hirota • Mitsuo Ishii

(Manuscript received November 5, 1987)

Ray tracing produces images of excellent quality, but it requires lengthy computations that limit its practical use. This paper discusses two approaches to shortening ray tracing computation time.

The first approach involves partitioning the environment with voxels to reduce the number of ray-object intersections. Proposed here is a new traversal algorithm that efficiently traverses the voxel data structure.

Experimental data demonstrate that the resulting traversal overhead is small.

The second approach involves implementing a ray tracing algorithm on a CAP (Cellular Array Processor) parallel processor. The parallelization of ray tracing takes advantage of the computational independency of individual rays. The results demonstrate that, when using the proposed load distribution scheme, performance increases in proportion to the number of cells used.

1. Introduction

Recent CAD systems such as those for industrial design require the generation of high-quality images, with optical effects such as reflection, refraction and shadow becoming very important in the modeling of reality. The necessary quality can be achieved by ray tracing, developed by Whitted¹⁾, as well as by an extension of the ray casting process used by Appel²⁾ and Goldstein and Nagel³⁾. Ray tracing can generate images of excellent quality, but it requires lengthy computation times that limit its practical use. The image generation time of conventional ray tracing can be several hours¹⁾.

A number of studies have been made to accelerate ray tracing. One group of studies uses coherence^{4), 5)} to reduce computation time, especially for ray-object intersection calculations, which are the most computationally intensive part of ray tracing. Ray coherence, which is the tendency of some collections of ray beams to behave similarly, is used in beam tracing⁶⁾ exclusively for polygonal objects.

Speer⁷⁾ also uses this form of coherence in a method that deals with a variety of objects.

Another application of coherence is space partition techniques, which use object-space coherence that capitalizes on the relative spatial position of the traced objects. There are two types of space partition schemes, octree and voxel. The space partition technique was first proposed by us⁸⁾ Matsumoto and Murakami⁸⁾ and Glassner⁹⁾ use the octree scheme for spatial representation. The octree is a tree structure in which each node has eight sub nodes corresponding to eight subspaces. As long as an object occupies one of these subspaces, that subspace is subdivided repetitively until no object is involved or until the subspace reaches a predetermined size. In the voxel partition scheme^{10), 11)}, the entire space is uniformly divided into congruent cubes called voxels (volume cells.)

Glassner presented an algorithm that traverses octree data structures. His method improved processing time, but by less than a factor of ten. There are two reasons for the octree scheme not yielding much improvement. First, traversal of the octree data adds to the computation overhead of the traversal itself.

Backtracking is necessary whenever traversal fails in one subnode. Second, because the space is dynamically divided as the octree is built, only the size of the current octant (cell of octree) can be known, which prevents the algorithm from incrementally calculating the next voxel element. For these reasons, we chose voxel data structure representation as the space partition scheme. Later in this paper we will present a voxel traversal algorithm that efficiently determines the next voxel element to which the tracing must move.

Another group of studies involves parallel computation of the ray tracing algorithm. The parallelization of ray tracing capitalizes on the computational independence of individual rays that pass through the different screen pixels. This method requires effective computation load balancing to achieve performance increases that are in line with the number of cell used. In Links¹²⁾, the task of each processor is allocated dynamically during processing. When a processor finishes a task, it is allocated another one by the host processor. The problem with this method is that it adds to communication overhead. The solution we propose here is to use two static load distribution schemes, which enables processor load balancing without adding communication overhead.

Chapter 3 discusses implementation of the parallel ray tracing algorithm on the CAP that we developed. The results, shown in chapter 4, demonstrate performance improvements over conventional ray tracing.

2. Traversal algorithm

2.1 Concept of the space partition method

Conventional ray tracing tests each object in the environment for intersections. Most of these intersection calculations are useless because only a few objects intersect with a given ray. The space partition approach is more efficient because intersection calculations are only done for what are called candidates.

The key idea behind the space partition method is to divide and conquer. Intersection calculation time is significantly reduced by subdividing the environment into small

orthogonal volumes (voxels) and performing intersection calculations for only those objects that are likely to intersect the ray. First, object identifiers are recoded for each object and voxel within which the object exists. As a ray propagates from one voxel to the next, the objects in each voxel become candidates for ray-intersection calculation. By rejecting objects that are not likely to intersect with the ray, the algorithm avoids many ray-object intersection calculations. This method can be applied not only to primary rays (the original rays from the view point) but to shadows and reflected and refracted rays as well.

2.2 Performance model

The movement operation that finds succeeding voxel elements for the ray to trace must be fast so that its calculation does not add overhead. Here we will analyze the performance obtained when using the voxel partition method with a cost evaluation model compared with that obtained using conventional ray tracing (Equations 1 and 2). We assume that objects are uniformly distributed in space and that the objects are smaller than the voxel elements. These assumptions imply that the ray will traverse several voxel elements before hitting an object. Under these conditions, the computational cost of conventional ray tracing, C_{conv} , and that of the ray tracing using the voxel partition technique, C_{vox} are represented in the following equations.

$$C_{\text{conv}} = R(RO + C_c \cdot P) + SO, \quad \dots \dots (1)$$

$$C_{\text{vox}} = R \cdot m \cdot N \left(RO + \frac{C_c \cdot P_s}{N^3} + C_t \right) + SO, \quad \dots \dots (2)$$

R : number of rays

N : number of partitions along the axis

P : number of primitives

C_c : cost of ray-object intersection calculation

C_t : cost of voxel traversal (one movement)

SO : overhead

RO : other calculations such as for shading

m : average traversal length.

The effective number of objects in a voxel element is inversely proportional to N^3 so the cost of the intersection calculation for a ray in conventional ray tracing, $C_c \cdot P$, is reduced by $1/N^3$ when using the voxel element scheme. The cost for a voxel element is the sum of the above cost and the cost of voxel traversal for one movement, C_t . Since the number of voxels the ray traverses until it hits the object is $m \cdot N$, this factor is multiplied to estimate the cost for the ray to find the intersection. The performance with respect to the number of partitions for the voxel partition method is as follows:

$$C_{\text{vox}} \approx \frac{C_1}{N^3} + C_2 N. \quad \dots \dots \dots (3)$$

The performance curve (cost vs. the number of partitions) falls abruptly and then rises gradually after the optimal partition (minimum processing time). To obtain good performance, the position of the turning point must be shifted in the direction of an increased number partitions. To reduce overhead, the relation between the two computational costs, C_t , and C_c , must be held as ($C_t \ll C_c$).

We present here a new traversal algorithm that efficiently determines, from the current voxel element, the next voxel element for the ray to enter. The method is called the Parametric 3DDDA (3 Dimensional Digital Differential Analyzer) since it uses the parameter of the ray (t in $\bar{X} = \bar{\alpha} t + \bar{\beta}$). The DDA algorithm¹³⁾ was originally used to obtain a rasterized straight line in two dimensions, where successive pixels along the line are efficiently determined using only incremental addition. This concept can be applied to the traversal algorithm over the voxel data structure with little computational cost. The 3DDDA algorithm performs as well as the original DDA algorithm in determining the next voxel element with respect to the current element using only incremental addition. The difference, of course, is that the 3DDDA algorithm performs in three dimensions rather than two.

The key idea behind this algorithm is to retain the points on the boundary planes of

the voxel element. These points are then used to guide the ray to the next voxel element. We use parameter t to refer to points on the boundary along the ray. The value of t increases as we move away from the origin, where t is 0. Three t parameters, t_x, t_y, t_z , that indicate the points X_x, X_y, X_z on the boundary with which the ray will intersect are provided as shown in Equation 4.

$$X_i = \alpha_i \cdot t_i + \beta_i, (i = x, y, z). \quad \dots \dots \dots (4)$$

After finding the minimum t parameters, the next voxel element to be penetrated by the ray is determined by adding or subtracting 1 to the component of the current voxel indices (v_x, v_y, v_z). There are six candidates, three for each component times two for each direction. The sign of the step (+1 or -1) for the current voxel index, however, can be determined for each component from the sign of the direction of the ray. For example, for a ray with the direction $\alpha_x > 0$, the sign of the step must be positive. Here, the coordinate of the voxel is aligned with the coordinate axis of the environment. The algorithm can only select the component to change from a field of three.

The selection of the component is based on the following observation. We define t_x, t_y , and t_z as the value of the ray parameter at the point on the boundary planes (Y-Z, X-Z, and X-Y) that the ray will penetrate.

These values determine the next voxel

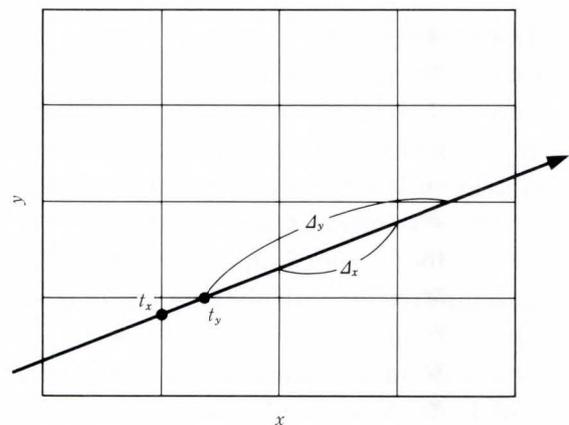


Fig. 1—Scheme of voxel traverse.

element with respect to the current one. The smallest of the three values causes the component to change. Figure 1 illustrates this in two dimensions. Parameters t_x and t_y indicate the points on the boundary at which the ray will be intersected. When t_x is smaller than t_y , the ray traverses the boundary of x ahead of y , that is the ray crosses the line ($y = \text{const}$) first. This corresponds to the x component of the voxel index change, which is +1 or -1 according to the sign of the direction, α_x . In this way, component changes are determined simply by examining the smallest of the t parameter values. This requires only two floating comparisons.

After determining the minimum value, the parameter has to be updated for the next evaluation. The precomputed step, $\Delta_x \cdot \Delta_y$, and Δ_z , that represents the length between the boundary planes along the ray is added to the t parameter just selected. Note that these step values are computed once during preprocessing by

$$\Delta_x = \frac{\text{units size of voxel}}{\alpha_x}.$$

The operation cost for one voxel movement requires one integer incremental addition operation for updating the voxel index, one floating incremental addition operation for updating the t parameter, and two floating comparisons for selecting the smallest of three values.

An exception is that the ray may not intersect any of the boundary planes. For example, the ray for which $\alpha_x = 0$ does not intersect YZ-planes that make the voxel boundary. The initial value $t_x = \infty$ can automatically avoid having the ray intersect with YZ-planes because it can never be chosen as the minimum value.

The following algorithm is used for the traversal of voxel data structures.

Algorithm -1: 3DDDA

```
int vx, vy, vz;           /*voxel indices */
int incre_x, incre_y, incre_z; /* +1 or -1 */
float tx, ty, tz;        /* t parameters */
```

```
float delta_x, delta_y, delta_z /* step */
int component;

while (in the voxel space) {

/* get component that corresponds to the
smallest value among t_x, t_y, t_z; */
component = GetMinimumComponent
(tx, ty, tz);
switch (component) {

case X:
    vx += incre_x;
    tx += delta_tx;
    break
case Y:
    vy += incre_y;
    ty += delta_ty;
    break;
case Z
    vz += incre_z;
    tz += delta_tz;
    break;
}
}
```

Kay¹⁴⁾ and Glassner pointed out that any particular object might end up in more than one voxel, potentially requiring a ray to encounter the same object more than once. But this drawback of the space partitioning approach can be overcome with an algorithm we will propose.

To provide an array corresponding to the object, we begin with a ray number (ensured to be unique by incrementing for an additional ray), which is passed as an argument to the voxel traversal procedure. This number is used to identify the ray and is recorded in an array, which is used to determine whether the ray-intersection calculation for a given object was done with the specific ray. Before the intersection calculation, a check is made to avoid duplicating the test. This check involves comparing the current ray number with the ray number in the array corresponding to the object to be tested. If the two numbers match, indicating that the ray object intersection calculations have already been done with the current ray, the intersection test can be omitted.

Array values must be updated when new intersection calculations are made. Every value in an array is initialized to some special value, for example -1.

The following algorithm is used to avoid repeating intersection calculations

Algorithm: Find intersections without the duplicate calculation
 (this portion of code is in the traversal algorithm)

```

input: ray_no
into RayNo [MAXOBJECT];

objects =
  GetObjectsWithinVoxel (voxel_index);
  for (i = 0; Objects is not empty; i++) {
    if (RayNo [Object [i]] != ray_no)
      Intersection =
        FindIntersection (Object [i]);
    RayNo [Object [i]] = ray_no;
  }
  }
    
```

3. Implementing ray tracing on the CAP

3.1 Parallel ray tracing

The ray tracing algorithm is analyzed to determine the course of parallel ray tracing. The first step is to fire a ray from the eye through the pixel into the environment. The closest intersection between this ray and the objects in the environment determines a visible point. Shadows are determined by firing rays from the visible point toward the light sources to determine whether any object blocks the ray. Depending on material properties, this surface may spawn reflected and refracted rays, which are recursively traced in the same fashion.

The process is continued until none of the rays intersect any objects in the environment or until the branch levels reach a predetermined number.

Because the ray is traced backward from the view point to the light source, computations for rays that pass through different pixels can be done independently. Although this limits

optical effects to specular reflection (such as with a mirror), it allows calculations to be simplified and done in parallel. The parallel ray tracing algorithm can capitalize on this computational independence of individual rays.

Each processor has a copy of the object database and processes rays that pass through the pixel element assigned to the cell in parallel.

3.2 CAP architecture

This algorithm is implemented on a parallel processor called the CAP (Cellular Array Processor¹⁵), which we developed. CAP is a MIND-type parallel computer consisting of 64 processors called cells configured in an 8 × 8 two-dimensional array. Each cell consists of an i80186 MPU, an i8087 NDP, and 2M-bytes of local memory. Figure 2 shows the CAP system. Every cell is connected by two buses, with the command bus used to download data and programs from the host computer. This bus can also be used for broadcast communication from the host computer to the cells or from a particular cell to the host computer or other cells. Each cell is connected to four neighboring cells for local communication. Cells at array boundaries are connected to cells at the opposite sides so that the network topology resembles the surface of a torus.

The CAP graphical facility is provided by giving each cell an integrated video memory instead of a frame buffer.

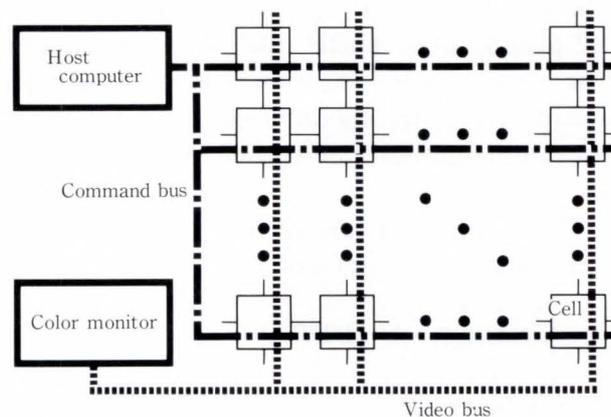


Fig. 2—CAP configuration.

The images in video memory are displayed on a monitor through the video bus. A significant feature of CAP's display capability is its variety of screen partitioning modes, in which any cell can output its result, usually an intensity value, to any pixel on the screen. Each cell knows its identifier and the pixel location on the screen that the cell handles. This facility makes it possible to achieve optimal load balancing of ray tracing. The task is distributed to many cells and the corresponding image that the cell produces is displayed.

3.3 Load balancing

Computation loads must be balanced if performance is to increase as the number of cells rises. Our method uses static load distribution to allocate part of the screen to each cell instead of dynamic load distribution. Static load distribution eliminates the communication of dynamic distribution. Further, since the amount of reallocation and communication increases as the number of processors increases, dynamic allocation limits the number of processors per array.

Two types of static load distribution mode are possible. With dot mode load balancing, each cell performs ray tracing on every eighth pixel horizontally and vertically. Cell-0 handles the top right pixel addressed (0, 0), the eighth right neighboring pixel addressed (8, 0), and so on.

Cell-1 processes the right neighboring pixels that cell-0 handles. In this way, the whole screen is processed by the 64 cells. In Fig. 3-b), the pixels marked with white boxes are handled by cell-0, for example. In line mode, each cell is allocated every 64th scan line. Cell-0 starts the 0th scan line and cell-1 starts the 1st one. The white boxes shown in Fig. 3-a) are handled by cell-0. This static distribution algorithm demonstrates that the load can be evenly balanced without communication, since the task load allocated to each cell can be considered to be averaged statistically. The performances of dot mode and line mode are compared in the next chapter.

4. Results and discussion

4.1 The results of the voxel partition scheme

Figure 4 shows the SPHERES model with 125 spheres used to demonstrate the performance of the space partition method. The test image was calculated at a resolution of 512×384 pixels. No antialiasing was done for the tests. Table 1 shows the processing statistics with respect to the number of partitions along each axis. This method reduces the computation time significantly in comparison with standard ray tracing that does not use any accelerating techniques. Note that the performance of conventional ray tracing is equal to the performance of one voxel partition in Table 1. For the

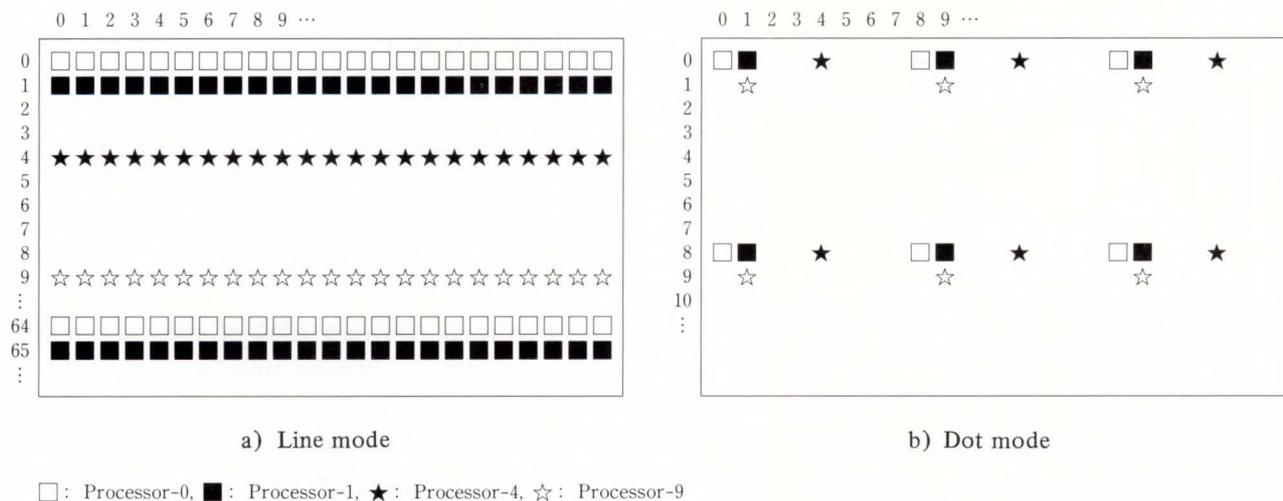


Fig. 3—Line and dot distribution modes.



Fig. 4—Model spheres: 125 objects.

Table 1. Processing time with respect to the number of voxel partitions

Model	1	2	4	8	16	32	48
16	102	39	23	20	21	25	30
512	825	212	68	34	28	31	36
1000	1537	450	126	59	38	38	43
4096	13094	2856	678	211	104	96	101
6858	32621	7306	1557	482	219	162	163

The model is SPHERES and the Time is in seconds.

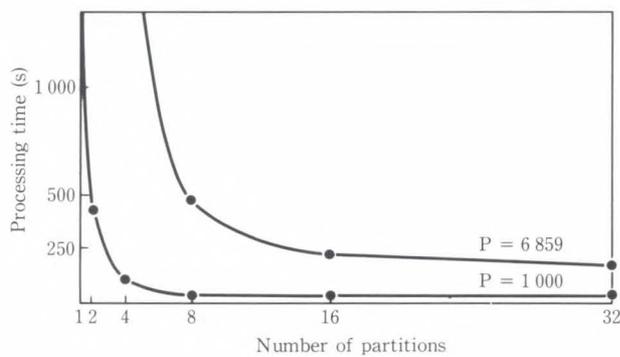


Fig. 5—Processing time.

SPHERES model with 6 859 spheres, this method is approximately two hundred times faster. Figure 5 is a graph of the statistics for models with 1 000 and 6 859 spheres. While the number of partitions is small, the performance increases significantly. The processing time decreases from 32 621 seconds to 7 306 seconds when the number of partitions along each axis

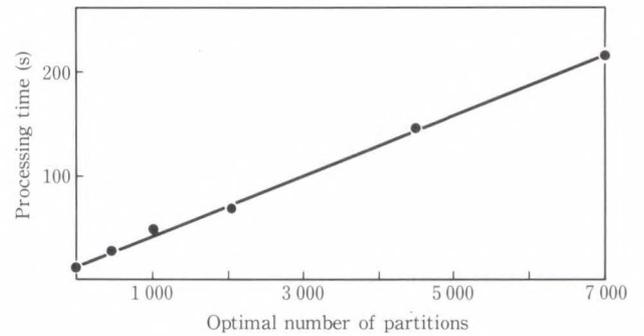


Fig. 6—Optimal ratio with respect to the number of objects.

Table 2. Costs of major algorithm components

Algorithm component	2 (%)	16 (%)	32 (%)
Intersection	108.5 (84.4%)	11.4 (34.2%)	7.8 (23.0%)
3DDDA	3.4 (2.7%)	5.1 (15.4%)	6.9 (20.4%)
Voxel generation	0.2 (0.1%)	0.5 (1.6%)	2.3 (6.6%)
Others (shading)	10.4 (8.0%)	10.3 (30.8%)	11.1 (32.4%)
Total	128.5	33.4	34.1

The model is SPHERES and the Time is in seconds.

changes from 1 to 2. However, as the number of partitions increases, the amount of performance increase decreases due to the additional overhead for traversing voxels. If the number of partitions becomes large enough, processing time actually increases. For example, when the number of partitions for 1 000 spheres changes from 32 to 48, the processing time increases slightly as shown in Table 1. Note that the optimal number of space partitions depends on the model and the number of objects. For the model with 16 spheres, the optimal voxel resolution is 8, while that for the model with 6 859 spheres is 32.

Figure 6 shows a plot of the performance ratio for the processing time with the optimal number of partitions to that without partitions. The performance ratio is proportional to the number of objects in the environment. Thus, performance effectiveness increases along with the number of objects. This implies that this

Table 3. Load balance statistics between the cells

Model	Line mode min-max difference ratio(%)	Dot mode min-max difference ratio(%)
SPHERES with 125 spheres	84-93 9.7%	88-86 2.3%
Chess	259-324 20.0%	281-294 4.4%
CAD	637-792 19.6%	667-691 3.5%

Difference ratio = (max - min)/max × 100.

method is effective for models involving large numbers of objects, which is, in general, a complex scene.

More detailed observations were made for internal processing times. Table 2 shows the processing time for different numbers of voxel partitions for the major components of the algorithm, including the ray-intersection calculation, 3DDDA, shading calculation, and the generation of the voxel data structure. The processing time for intersection calculations falls when the number of voxels in the partitions increases, while that for 3DDDA and the voxel data generation increase. The result is that overhead increases total processing time slightly when moving from 16 to 32 partitions for this model.

4.2 Load balancing

Table 3 shows figures obtained from the line and dot distribution modes. The difference ratio is defined as the difference between the maximum and minimum processing time divided by the maximum processing time times 100. The difference ratio in dot mode is only a few percent while that of the line distribution mode ranges up to 20 percent. This implies that the dot mode distribution scheme works effectively. Figure 7 shows how performance increases with respect to the number of cells in both dot and line load distribution modes. The graph is normalized by the processing time when using 16 cells. In dot mode, the performance factor of 64 cells compared with that of 16 cells is 3.96, which is almost proportional to the number of cells. With line mode, this ratio is

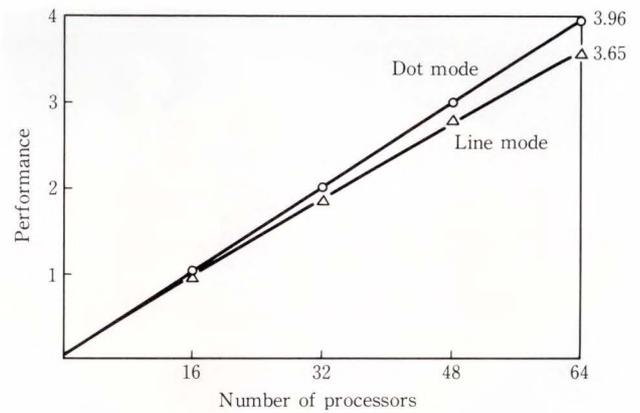


Fig. 7—Performance with respect to number of processor cells.

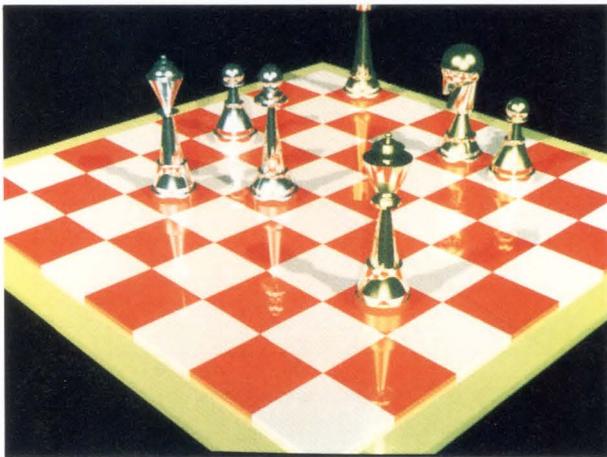
Table 4. Comparison of CAP and FACOM M-380 performance (seconds)

Model (number of objects)	CAP	M-380
SPHERES (125)	72	60
Chess (144)	155	128
CAD (178)	20	20
Piston (179)	40	33

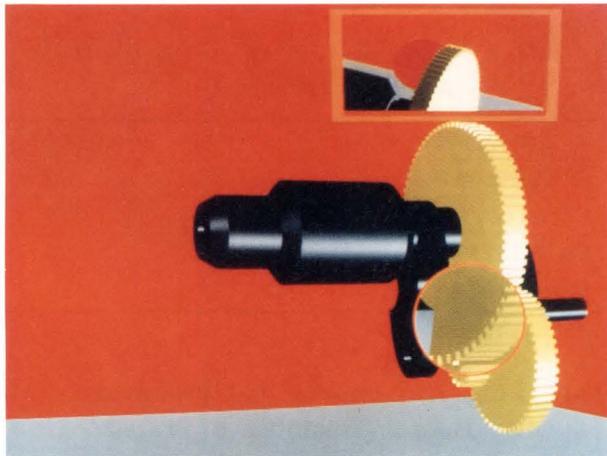
only 3.65. This implies that performance will fall when using large numbers of processors.

4.3 Comparison with large-scale computers

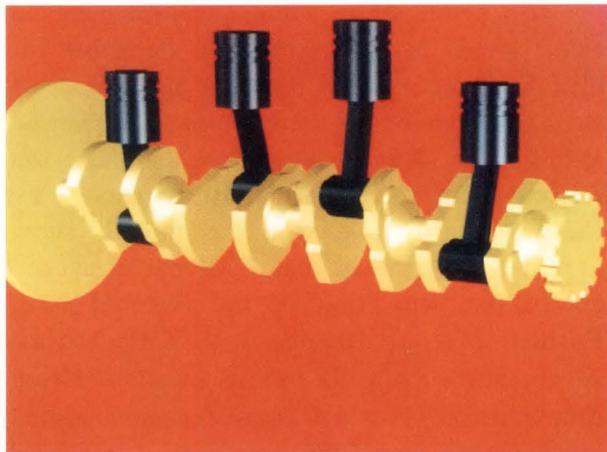
To demonstrate the performance and effectiveness of the ray tracing algorithm runs on CAP, we compared the CAP performance to that of a large scale computer, the FACOM M-380 (15 MISP). Processing times were measured for the same algorithm. Table 4 lists the processing times for the models shown in Fig. 8 and shows that CAP performance is comparable with that of the FACOM M-380. Furthermore, ray tracing with CAP has two advantages. First CPU time is equal to elapsed time with CAP while elapsed time is usually much longer than CPU time with a large scale computer. Since interactive image generation requires high-speed processing, this characteristic is very important. Second, as CAP performance increases with the number of processors, much higher performance can be achieved by a large number of processors.



1) Chess



b) CAD



c) Piston

Fig. 8—Models.

5. Conclusion

We have developed a fast ray tracing algorithm using two techniques. First, the ray-

object intersection calculation is reduced using space partition techniques and an efficient traversal algorithm. Significant improvements in processing time were obtained, from several dozen times to 200 times faster than conventional ray tracing. Second, this algorithm was implemented on a parallel processor, the CAP. We also proposed an efficient static load balancing method, which, when using a dot mode static distribution algorithm, demonstrates the ability to balance loads evenly while eliminating communication overhead.

It was shown that, in some ranges, performance increases with the number of processors. We have tried up to 64 processors. We compared the processing time of CAP with that of a large scale computer, the FACOM M-380 and showed that CAP performance with 64 cells is comparable to that of a main frame computer.

As mentioned in the section on implementation, each cell contains the same data. This limits the number of objects the system can handle. Data can be distributed between every cell and accessed through local communication.

6. Acknowledgement

The authors would like to thank Messrs. J. Tanahashi, H. Shiraishi for promoting this research project, and the staff members of CAP hardware group for their valuable discussions.

References

- 1) Whitted, T.: An Improved Illumination Model for Shaded Display. *Commun. ACM.*, **23**, 6, pp. 343-349 (1980).
- 2) Appel, A.: Some Techniques for Shading Machine Rendering of Solids. *Proc. AFIPS JSCC*, **32**, pp. 37-45 (1968).
- 3) Goldstein, R., and Nagel, R.: 3-D Visual Simulation. *Simulation*, 25-31, (1971).
- 4) Atheron, P.: A Scan-line Hidden Surface Removal Procedure for Constructive Solid Geometry. *Comput. Graphics*, **17**, 3, pp. 73-82 (1983).
- 5) Sutherland, I., Sproull, R., and Shumacker, R.: A Characterization of Ten Hidden Surface Algorithms. *Comput. Surv.*, **6**, 1, p. 1 (1974).
- 6) Heckbert, S., and Hanrahan, P.: Beam Tracing

- Polygonal Objects. *Comput. Graphics*, **18**, 3, pp. 119-127 (1984).
- 7) Speer, L., and Barsky, B.: A Theoretical and Empirical Analysis of Coherent Ray-tracing. Proc. Graphics Interface 1985, pp. 1-8.
 - 8) Matsumoto, H., and Murakami, K.: Fast Ray Tracing using the octree partitioning. (in Japanese), 27th Inform. Proc. Conf. Proc., 1983, pp. 15-22.
 - 9) Glassner, A.: Space Subdivision for Fast Ray Tracing. *IEEE CG & A*, **4**, 10, pp. 15-22 (1984).
 - 10) Murakami, K., Hirota, K., and Sato, H.: Ray Tracing System using the voxel partitioning on a Cellular Array Processor. (in Japanese), *Graphics and CAD*, **22**, 2, pp. 1537-1538 (1986).
 - 11) Fujimoto, A., Tanaka, T., and Iwata, K.: ARTS: Accelerated Ray Tracing System. *IEEE CG & A*, **6**, 4, pp. 16-26 (1986).
 - 12) Nishimura, H., Ohno, H., Kawata, T., Shirakawa, I., and Omura, K.: LINKS-1: A Parallel Pipelined Multimicrocomputer System for Image Creation. Proc. 10th Sympo. Comput. Architec., SIGARCH, 1983, pp. 387-394.
 - 13) Bresenham, J.E.: Algorithm for Computer Control of a Digital Plotter. *IBM Sys. J.*, **4**, X, pp. 25-30 (1965).
 - 14) Kay, T., and Kajiya, J.: Ray Tracing Complex Scenes. *Comput. Graphics*, **20**, 4, pp. 269-278 (1986).
 - 15) Sato, H. et al.: Fast Image Generation of Constructive Solid Geometry Using a Cellular Array Processor. *Comput. Graphics*, **22**, 2, pp. 95-102 (1985).



Kouichi Murakami

Computer-Based Systems Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Nuclear Eng.
Hokkaido University 1978
Master of Nuclear Eng.
Hokkaido University 1980
Specializing in Computer Graphics



Mitsuo Ishii

Computer-Based Systems Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Electrical Eng.
Nagoya Institute of Technology 1966
Dr. of Electrical Eng.
Tokyo Institute of Technology 1988
Specializing in Computer Aided Design



Katsuhiko Hirota

Computer-Based Systems Laboratory
FUJITSU LABORATORIES,
KAWASAKI
Bachelor of Computer Informations Eng.
Toyohashi University of Technology 1981
Master of Computer Informations Eng.
Toyohashi University of Technology 1983
Specializing in Computer Graphics

Simulation Processor : SP

• Fumiyasu Hirose • Hirofumi Hamamura • Mitsuo Ishii
• Hiroshi Yamada

(Manuscript received March 31, 1987)

A high-speed, large-capacity logic simulation engine, the simulation processor SP, has been developed for efficient logic verification of Fujitsu's large digital system. The SP implements an event-driven simulation algorithm in a multi-processor configuration. It utilizes a multi-stage switching network with an attached buffer-memory so that inter-processor communications do not cause a bottleneck. The SP executes 4-input-1-output logic primitive and 1-bit-per-word memory primitive level simulation with an accuracy of 16 signal values and unit/zero delay simulation. It can simulate 4 mega primitives and 32 Mbytes of memory at its maximum speed of 0.8 giga active primitive evaluations per second.

1. Introduction

The logic simulator plays an important role in the development of large digital systems. Because logic designs are implemented as VLSIs, the importance of complete logic verification before manufacture of VLSIs is increasing. However, since digital systems today are getting larger and more complex, the time and the number of functions required for verification has increased considerably. It is said that several tens of years would be required to simulate a large mainframe design with the test program for the manufactured hardware, though it takes only several minutes when it is run on the real mainframe¹⁾. Some general-purpose large computers are already being used as special-purpose machines for logic simulation.

With this background, a special purpose simulation processor SP²⁾⁻⁴⁾, has been developed for logic verification of Fujitsu's large digital systems, and has improved the performance over the conventional simulator by up to three orders of magnitude. Difficult simulations normally done on the large mainframes are eliminated. Since the SP is capable of simulating several hundred clock cycles of a very large digital system in a moment, the interactive simulation of an entire large mainframe system is also possible.

2. Simulation algorithm

Figure 1 shows the SP itself. Figure 2 shows a circuit example and Table 1 shows the simulation output. As in the table, the net status of all gates is initially set to X (unknown status) at time 00. At time 10, an external input D changes from X to 1. Then the simulator changes the net status of D to 1. This process is called updating. A net status change is called an event. At time 11, gate g_1 , which is the gate whose input changed during the previous time cycle, is fetched. This is called the fanout-fetch process.

The gate g_1 is evaluated with the new input status, $D = 1$ and $g_0 = x$, to check if the input event on D causes the output event of g_1 . This process is called the evaluation process. Then it is found that the output g_1 remains at X . Since there are no events at time 11, the circuit becomes stable. At time 20, an external input HOLD changes from X to 0. Fanout gates g_0 and g_2 of the external input HOLD are fetched and the new output states are evaluated. It is found that g_0 changes from X to 1 and g_2 changes from X to 0 at the next time, 21. At time 21, the net status of gates g_0 and g_2 is updated and the fanout gates g_1 and g_3 of g_0 and g_2 , respectively, are evaluated. In this manner, the following facts are simulated: the



Fig. 1—Simulation processor: SP

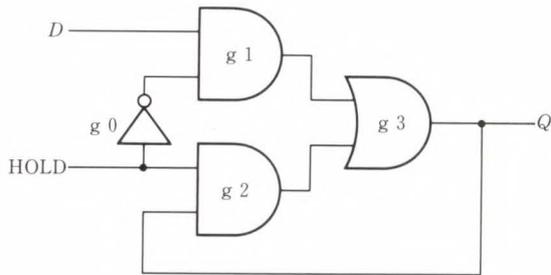


Fig. 2—Logic design example.

circuit takes input *D* into the feedback loop and becomes 1 at time 24 input *HOLD*, which changes from 0 to 1 at time 30, causes input *D* to be held in the feedback loop as an internal state *Q*, and state *Q* does not change even when input *D* changes at time 40. As is shown in the example, the method of simulating a circuit by evaluating gates whose inputs changed during the previous cycle is called an event-driven algorithm.

Since an event-driven simulator handles combined, synchronous, and asynchronous circuits with equal ease, and delays can be easily processed, most software simulators today employ the event-driven algorithm. Several simulation machines have already been developed^{1),5)}. The compiler driven algorithms⁶⁾ and the level sort & event-driven

Table 1. Simulation output list

Net	Time	0	1	1	2	2	2	2	2	3	3	3	3	4	4
<i>D</i>		X	①	1	1	1	1	1	1	1	1	1	1	1	0
<i>HOLD</i>		X	X	X	⊙	0	0	0	0	①	1	1	1	1	1
<i>g 0</i>		X	X	X	X	①	1	1	1	1	⊙	0	0	0	0
<i>g 1</i>		X	X	X	X	X	①	1	1	1	1	⊙	0	0	0
<i>g 2</i>		X	X	X	X	⊙	0	0	0	0	1	①	1	1	1
<i>Q=g 3</i>		X	X	X	X	X	①	1	1	1	1	1	1	1	1

①: Event→ 1 ▲: Stable state
 ⊙: Event→ 0 —: Evaluated but no event

algorithms⁷⁾ employed especially for simulating large digital systems do not allow the flexibility of an event-driven algorithm. Fujitsu's automation systems can offer consistent modeling and accuracy in simulation by using both conventional software and SPs.

3. Hardware implementation

Figure 3 shows the block diagram of a gate-processor (GP). The GP executes the simulation of a partitioned subcircuit of an entire digital system. The GPs send and receive the next-time events on the boundary nets of subcircuits to each other through the event transmission network (ET) shown at the left side of Fig. 3. Evaluation gate memory (EGM), shown at the right side of Fig. 3, is a pair of buffers for storing the primitives (logic and memory primitives) to be evaluated. One buffer holds the primitive to be evaluated at present, and the other holds the primitive to be evaluated next. Net status memory (NSM) at the bottom of Fig. 3 holds the present-time net status of the circuit. The evaluation pipeline checks if the output status of an evaluated gate changes at the present time. It does this by reading inputs and the current output status of the gate from the NSM and evaluating it according to the gate type. New event memory (NEM) holds the next-time events, and the NSM is updated according to the contents of the NEM when the time advances. The fanout pipeline fetches gates whose input changes at the next time.

Suppose the time is *T* minus 1. The NEM holds the events of time *T*. The NSM holds the net status at time *T* minus 1. EGM 1 contains the gates whose input changes at time *T*, and

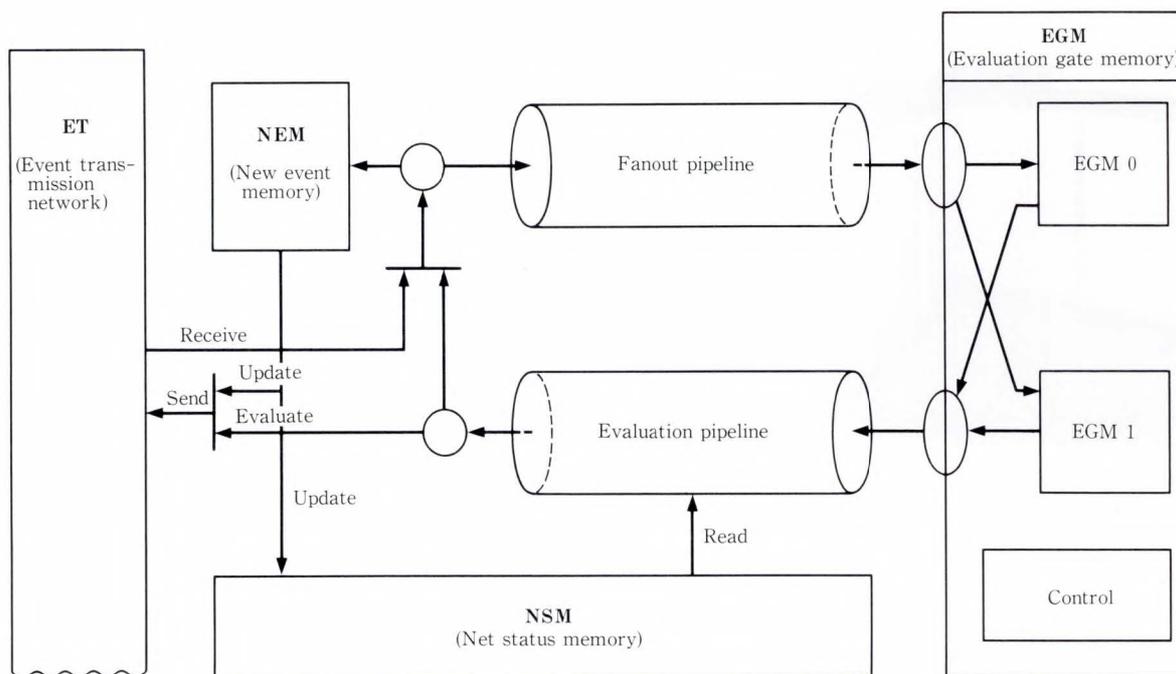


Fig. 3—GP architecture.

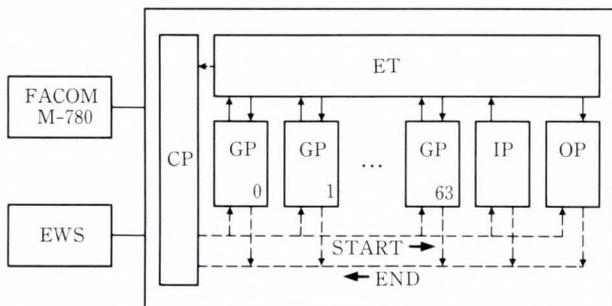


Fig. 4—SP processor configuration.

EGM 0 is empty. The GP has finished its task at time T minus 1. A control processor (CP), an internal processor of the SP, advances the time by one and commands GP to start the operations of time T . Then, in the GP, the NEM updates the NSM and changes the net status of the circuit to that for time T . If there is an NEM event of which the designer wants to monitor the gate activity, the event is sent to an output processor (OP), one of the SP processors, through the ET. The connections among the EGM buffers and the two pipelines are then switched, and EGM 1 begins to send gates to the evaluation pipeline one by one. The evaluation

pipeline extracts the gates whose outputs will change at time T plus 1, and registers the events in the NEM. It also requests the fanout pipeline to fetch the gates whose inputs change at time T plus 1, if a fanout gate fetch is necessary. The outputs of the fanout pipeline are then registered in EGM 0. When an event occurs on the boundary net, the event is sent to the other GPs through the ET. The event from the ET is then received at a time when the evaluation pipeline is not registering internal events.

4. SP processor configuration

Figure 4 shows the configuration of the SP processors. A large digital system to be verified is partitioned into several subcircuits and each subcircuit is loaded to a GP of multi-processor configuration. When a GP detects an event on the boundary net among subcircuits, it sends the event to some of the other GPs through the ET. The GP handles a 4-input-1-output logic primitive and a 1-bit-per-word memory primitive as an object to evaluate. A single GP can simulate a subcircuit with a maximum capacity of 64 kilo primitives and 512 Kbytes of memory. Up to 64 GPs can be connected to the ET.

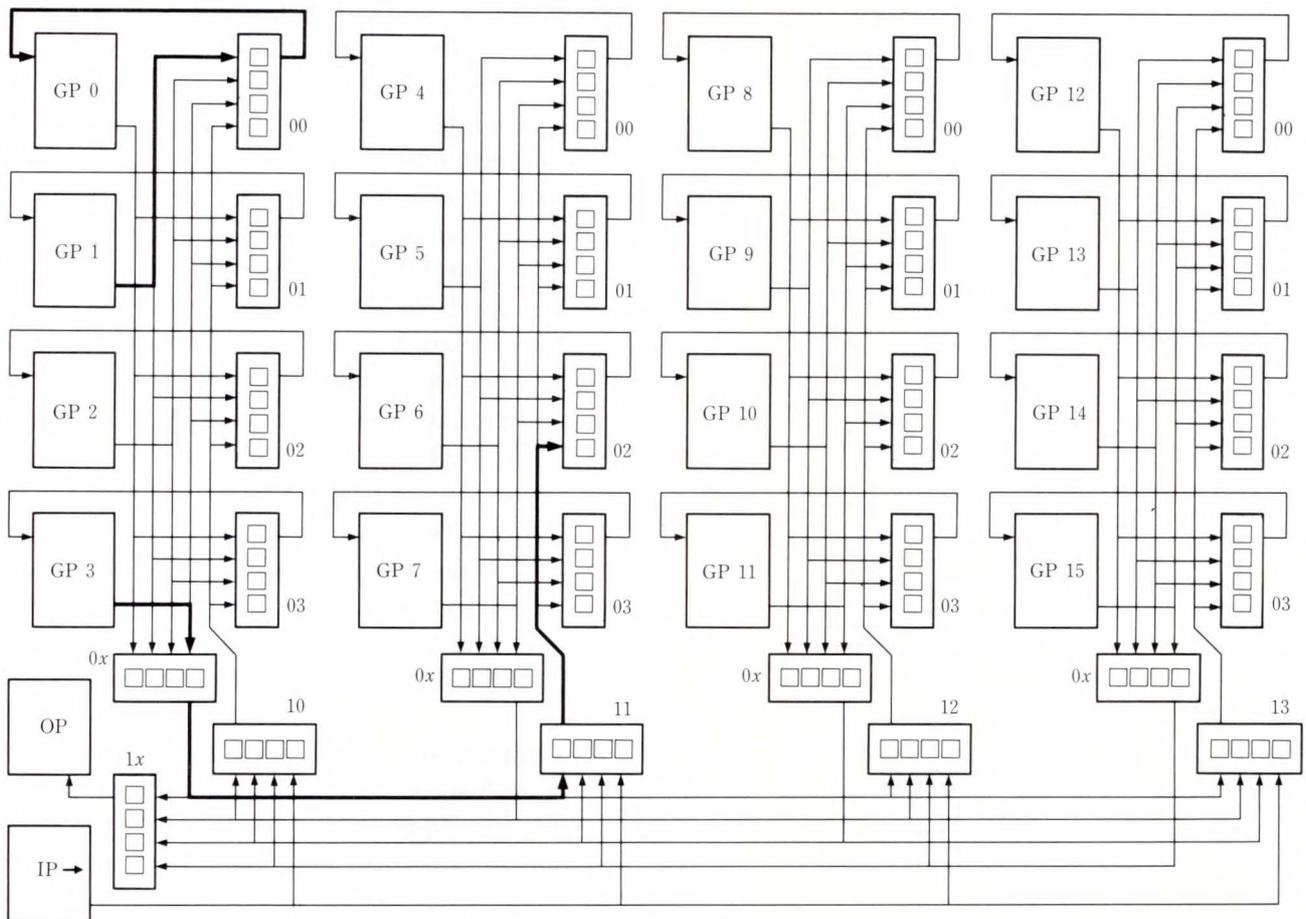


Fig. 5—Event transmission network ET for 16 gate-processors.

The SP can simulate a digital system up to a maximum of 4 mega primitives and 32 Mbytes of memory. This capacity is sufficient to simulate an entire large mainframe or supercomputer. Figure 4 shows how an input processor (IP) sends external input stimuli to the GPs. An output processor (OP) receives the activity of the gates that the user wants to monitor. The control processor (CP) synchronizes all the processors, and controls the SP stop and start according to the simulation controls.

5. High-speed switching network ET

Reducing inter-processor communication overhead to achieve a processing speed commensurate with the increased number of processors is one of the most important goals in parallel processing. Common bus-structured communication systems require a lot of over-

head⁸⁾; the SP copes with this by means of a high-speed multi-stage switching network with an attached buffer memory.

Figure 5 shows an ET for 16 GPs. The GPs are partitioned into four groups of four members each. Group 1 consists of GP 0 to GP 3, group 2 of GP 4 to GP 7, group 3 of GP 8 to GP 11, and group 4 of GP 12 to GP 15. In the figure, shaded rectangles with 4-input-1-output interfaces are called ET bases (ETB). Five ETBs form an ET unit (ETU), a unit in the hierarchical structure of the ET. There are five ETUs in the figure. For example, the four sets of ETBs labeled 00, 01, 02, 03, and 0x form four ETUs, and each ETU corresponds to the GP group. The ETBs labeled 10, 11, 12, 13 and 1x also form a ETU that switches the events among the GP groups, IP and OP. An ETU exchanges events among four GPs in the same

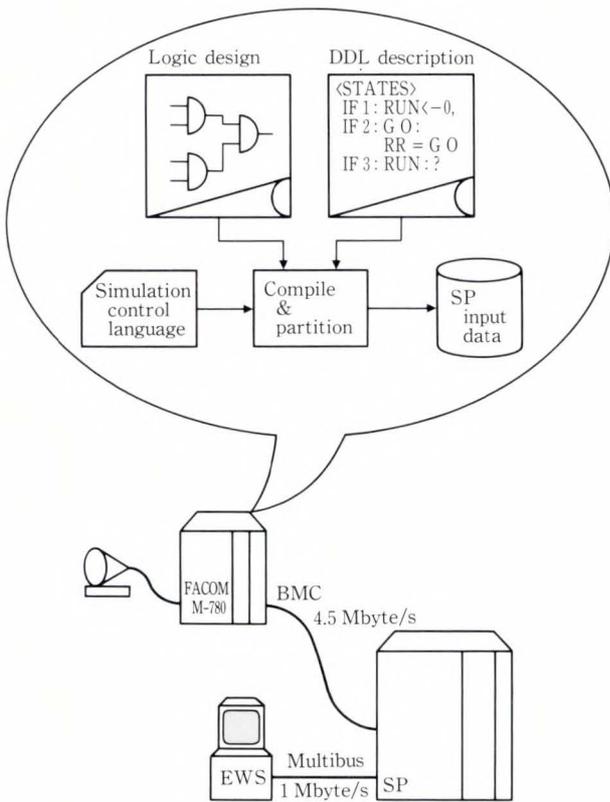


Fig. 6—SP system configuration.

group, or among four ETUs one level lower in the hierarchy. At the same time an ETU sends events to another ETU one level higher in the hierarchy.

In Fig. 5, two event transmission routes are indicated. The first route (communication within the same group) starts from GP 1 and ends at GP 0 via the ETB labeled 00. The second route (communication with another group) starts from GP 3 and ends at GP 6 via the ETB labeled 0x in GP group 1, and the ETBs labeled 11 and 02 in GP group 2.

Any GP can send an event to the ET independent of the communication requests of other GPs and receive an event from the ET when the GP is not busy.

6. System configuration

Figure 6 shows the SP system configuration. A large digital system to be verified is compiled and partitioned by the Fujitsu design automation system^{9),10)} on the FACOM M-780. Circuit

units whose detailed logic design is not yet completed, and circuits that do not need to be verified, are described using the register transfer-level language DDL and are connected to the simulation model^{11),12)}. The simulation control language is also translated for the SP, and the input data for the SP is prepared. The data is loaded to the SP through the BMC channel with an I/O speed of 4.5 Mbyte/s. When the SP is used in the batch mode, certain stop conditions are set before simulation starts. When one of the conditions is satisfied, the CP notifies the host. The host then reads out the gate activities of the monitoring points, and the user checks and adjusts the circuit activity to the time the condition was satisfied. When the SP is used in the interactive mode, the user debugs the circuit in the interactive command base after the FACOM M-780 loads the input data.

7. Conclusion

A high-speed, large-capacity logic simulation engine, the Simulation Processor SP, has been developed for efficient logic verification of Fujitsu's large digital systems. The SP can simulate an entire digital system with its a maximum capacity of 4 mega primitives and 32 Mbytes of memory. It operates more than three thousand times faster than a conventional software simulator using a large general purpose computer. In conclusion, difficult simulation operations previously done on large mainframes can be eliminated. Since the SP can simulate several hundred clock cycles of a very large digital system in a moment, the interactive simulation of an entire large mainframe is also possible.

8. Acknowledgement

The authors would like to thank the staff members of Computer-based Systems Laboratory, DA Development Dept., Computer Engineering Dept., 1st Main Frame Div. and Fujitsu Automation for their continued interest and valuable discussions throughout this project.

References

- 1) Ohmori, K., and Koike, N.: CAD Machines. (in Japanese), *J. IPS, Jpn.*, **25**, 10, pp. 1144-1152 (1984).
- 2) Yamada, H., Hirose, F., Niitsuma, J., and Sindo, T.: Simulation Processor SP. (in Japanese), *Trans. IEICE, Jpn.*, **J71-D**, 4, pp. 644-651 (1988).
- 3) Yamada, H., Shindo, T., Hirose, F., and Niitsuma, J.: Memory simulation part of gate processor and input, output processor used in a simulation processor SP. (in Japanese), *Trans. IEICE, Jpn.*, **J71-D**, 2, pp. 222-228 (1988).
- 4) Hirose, F., Ishii, M., Niitsuma, J., Shindo, T., Kawato, N., Hamamura, H., Uchida, K., and Yamada, H.: Simulation Processor SP. Proc. IEEE Inter. Conf. Computer-Aided Design, pp. 484-487, (1987).
- 5) Blank, T.: A survey of hardware accelerators used in computer-aided design. *IEEE Design Test Computers*, **1**, 4, pp. 21-39 (1984).
- 6) Denneau, M.M.: The Yorktown Simulation Engine. Proc. 19th Design Autom. Conf., pp. 55-59 (1982).
- 7) Koike, N., Ohmori, K., Kondo, H., and Sasaki, T.: A High-Speed Logic Simulation Machine. Proc. COMPCON, Spring '83, pp. 446-451 (1983).
- 8) Erickson, B.: Automation and Simulation in large system design. *VLSI SYS. DESIGN*, pp. 42-48 (1986).
- 9) Nagamine, M., and Hamamura, H.: CAD System for Digital Systems. (in Japanese), *J. IECE, Jpn.*, **61**, 7, pp. 744-749 (1978).
- 10) Hamaguchi, T., Hamamura, H., and Horii, Y.: Design Automation System for FACOM M-780. (in Japanese), *FUJITSU*, **37**, 2, pp. 135-139 (1986).
- 11) Kawato, N., Saito, T., Maruyama, F., and Uehara, T.: Design and Verification of Large Scale Computer by Using DDL. Proc. 16th Design Autom. Conf., pp. 360-366 (1979).
- 12) Sato, S., and Sugimoto, M.: Artificial Intelligence. *FUJITSU Sci. Tech. J.*, **22**, 3, pp. 139-181 (1986).



Fumiyasu Hirose

Computer-Based Systems Laboratory
 FUJITSU LABORATORIES,
 KAWASAKI
 Bachelor of Electrical Eng.
 Tokyo Institute of Technology 1976
 Master of Electrical Eng.
 Tokyo Institute of Technology 1978
 Specializing in CAD Systems



Mitsuo Ishii

Computer-Based Systems Laboratory
 FUJITSU LABORATORIES,
 KAWASAKI
 Bachelor of Electrical Eng.
 Nagoya Institute of Technology 1966
 Dr. of Electrical Eng.
 Tokyo Institute of Technology 1988
 Specializing in Computer Aided Design



Hirofumi Hamamura

DA Development Dept.
 FUJITSU LIMITED
 Bachelor of Electrical Eng.
 Keio University 1970
 Specializing in Electrical Design
 Automation



Hiroshi Yamada

Managing Director
 FUJITSU LABORATORIES
 Bachelor of Physics
 The University of Tokyo 1952
 Master of Physics
 The University of Tokyo 1956
 Specializing in Computer Architecture

Full-Color Image Printer

• Tomohisa Mikami • Tsugio Noda • Toshio Konaka

(Manuscript received November 30, 1987)

This paper describes the print data control circuit developed for a new full-color thermal ink-transfer printer that uses a unique thermofusible ink sheet with gray scale reproducibility. This printer can print twice as fast using this sheet as when using a conventional sublimable ink sheet, and offers high image quality by incorporating the precision color conversion circuit and temperature drift compensation circuit.

Precise color reproduction was enabled by interpolating conversion tables instead of calculating color-conversion equations. Moreover, significant print density variations due to medium-speed temperature drift were reduced to about one quarter by calculating the temperature drift line-by-line.

1. Introduction

Practical and simplified full-color image processing has been achieved by recent advances made in computer technology. This has caused a surge of research and development efforts in the field of color hard copy applications, especially those based on ink-jet and thermal ink-transfer printing technologies. Until now, most high-quality image printers have used a thermal ink-transfer sheet with sublimable dye¹⁾ due to its excellent gray scale reproducibility. This ink sheet can reproduce 32 to 64 gray scale levels per dot, which provides the high print quality necessary to reproduce pictorial images. Many commercially available color video printers use the sublimable ink sheet. Conversely, this sheet requires much more printing energy and time than thermofusible sheets, which makes its use difficult in high-speed printers or those that use large paper sizes.

A thermofusible ink sheet with gray scale reproducibility²⁾ has been developed to solve these problems. At the same time, the use of this ink sheet in high-speed, gray-scale printers required the development of precise and compact color conversion, and precise temperature drift compensation for the thermal print head.

Study was consequently made on a print

data control circuit for gray scale printers that offer satisfactory image quality and printing speeds twice as fast as those printers that use sublimable ink sheets. By developing a color conversion circuit interpolating conversion table, and a temperature drift compensation circuit to handle temperature drifts ranging from low (page-by-page) to high (line-by-line) speeds, such a printer has been made practical. We have developed a prototype printer that reproduces high-quality pictorial images at the targeted printing speed of 16 ms/line.

This paper first describes the specifications and structure of this prototype printer. Then, the theory of operation, theoretical precision, and testing results of the print data control circuit, which consists of an image data control circuit and gray scale level control circuit, are described. The description of the image data control circuit focuses on the color conversion circuit, and that of the gray scale level control circuit focuses on the temperature drift compensation circuit.

2. Printer specifications and structure

Table 1 lists the specifications of the prototype printer. The printer uses a thermal print head with a printing width of 216 mm and a

Table 1. Specifications of full-color image printer

Item	Specifications
Printing method	Thermofusible ink transfer
Print speed	165 s/A4-size
Print resolution	8 dot/mm
Gradation levels	64 (16-level dither with 2 × 2 matrix)
Positional deviation	100 μm
Paper size	A4-size (210 × 297 mm)
Recording size	190 × 266 mm

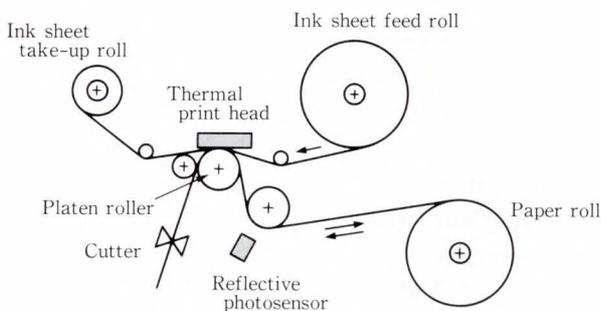


Fig. 1—Structure of printer mechanism.

resolution of 8 dots/mm. Yellow (Y), magenta (M), cyan (C) and black (K) images are printed sequentially page-by-page. For a line cycle time of 16 ms, a full-color A4-size image with up to 64 gradation levels can be printed in 165 s. The printer uses density modulation with 16 gray scale levels and multi-level dithering using a two-by-two matrix to improve the gray scale level accuracy and printing speed for achieving our design objectives.

Figure 1 shows the structure of the printer. The paper is fed from a paper roll, and is then cut into an A4-size sheet after each sheet is printed. The paper is fed three times forward and backward to overlap the Y, M, C and K images. Because a precision paper-feed mechanism is required to obtain high-quality print images, the printer mechanism shown in Fig. 1 attained an average printing dot deviation of 40 μm in Y, M, C and K.

Figure 2 shows a cross section of the ink sheet. In addition to the wax and dye, the solid ink layer includes fine carbon powder as filler.

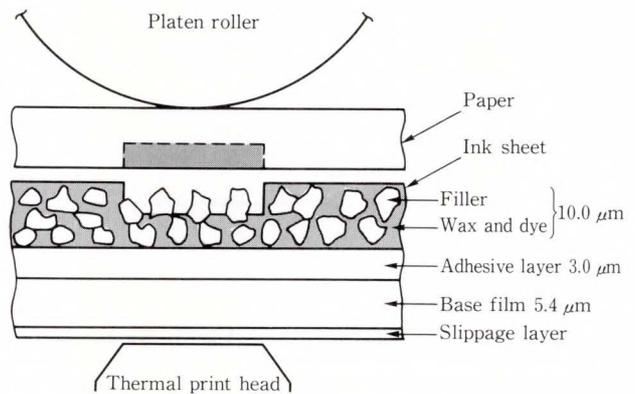


Fig. 2—Printing mechanism and cross section of ink sheet.

The print density is controlled by regulating the printing power being applied to the thermal print head because melted ink gradually oozes out through the porous stone-like wall formed by the filler. High print density and low graininess were achieved by using coated paper having a very smooth surface. The print density is 1.2 to 1.3 (O.D.) for a printing power of 0.24 W at a pulse width of 4.5 ms.

Our objective of printing highly defined, full-color images at high speeds required the development of print data control circuits, especially the following color conversion and temperature drift compensation circuits. A print data control circuit drives the thermal print head corresponding to the input NTSC-RGB signal. As shown in Fig. 3, this circuit consists of two sub-circuits: an image data control circuit that calculates the gray scale level for the YMCK ink based on the input RGB signal, and a gray scale level control circuit that drives the thermal print head according to the corresponding print density and gray scale level set by the image data control circuit.

The following sections describe these sub-circuits in detail, and focus on color conversion and temperature drift compensation.

3. Image data control circuit

3.1 Structure

The image data control circuit controls the printed image quality, and consists of sub-circuits for color conversion, gray scale trans-

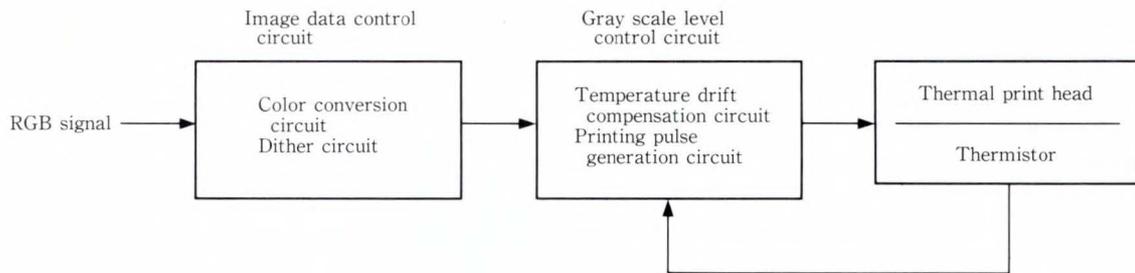


Fig. 3—Block diagram of print data control circuit.

formation, image sharpening, and multi-level dithering. Figure 4 shows a block diagram of the image data control circuit. The accuracy and speed of the color conversion circuit that calculates the gray scale level of the YMCK ink from the RGB signal are important factors.

The image data control circuit inputs RGB (eight bits per color) and converts it into an $L^*a^*b^*$ value defined by the CIE 1976 $L^*a^*b^*$ uniform color space. Image processing of gray scale transformation and image sharpening is executed in this color space. The gray scale level for the YMCK ink (six bits per color) is then calculated based on the spectral reflectance of the YMCK ink for all gray scale levels. The value of the gray scale level is then dithered at the 16 gray scale levels to obtain 4-bit gray scale level values for the YMCK ink.

3.1.1 Gray scale transformation and image sharpening

The highly defined printing of pictorial images requires gray scale transformation and image sharpening, for which color spaces must be evaluated. A subjective evaluation of the image quality indicated that the $L^*a^*b^*$ uniform color space was ideally suited for image processing³⁾. Accordingly, the image data control circuit converts the RGB signal into $L^*a^*b^*$, and then gray scale transformation and image sharpening are done. $L^*a^*b^*$ is then converted into the gray scale level value of the YMCK ink. Thus, the processing of image data in a uniform color space facilitated optimum image enhancement.

3.1.2 Color conversion

Precise color reproduction was made possible by predicting colors obtained from subtrac-

tive color mixtures. The print density of the YMCK ink corresponding to the input RGB signal was calculated from such color prediction. Because this calculation (called color conversion) is usually very complicated, a high-speed, precise, and compact color conversion circuit was required. This was one of the keys to achieving our printing objectives.

Generally, the relationship between the amount of YMC ink (Y, M, C) and the RGB density (D_r, D_g, D_b) is measured by using color patches, which is approximated by the following second-order polynomial:

$$\begin{pmatrix} Y \\ M \\ C \end{pmatrix} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,10} \\ a_{2,1} & \cdots & a_{2,10} \\ a_{3,1} & \cdots & a_{3,10} \end{pmatrix} \begin{pmatrix} D_r^2 \\ D_g^2 \\ D_b^2 \\ D_{rg} \\ D_{gb} \\ D_{br} \\ D_r \\ D_g \\ D_b \\ 1 \end{pmatrix} \dots \dots \dots (1)$$

The equation for color conversion includes considerable margin for error. Moreover, the color conversion circuit is usually complex and conversion is slow because the equation includes at least ten multiplication and addition operations.

The color conversion circuit shown in Fig. 4 simply interpolates the conversion tables stored in Read Only Memory (ROM), instead of actually calculating the complex color conversion

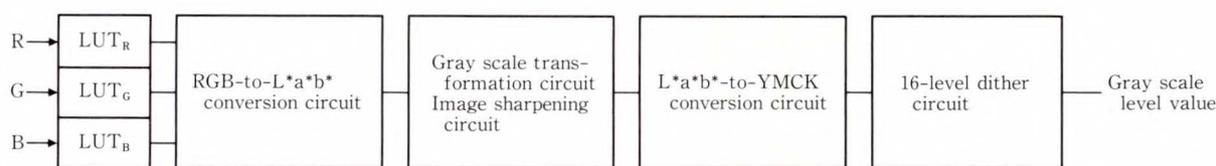


Fig. 4—Block diagram of image data control circuit.

equations. Because conversion table data can be generated without restrictions placed on circuit fabrication or calculation times, precise color conversion using an interpolation circuit offers a practical solution. This is explained later in more detail.

3.1.3 Multi-level dithering

As the printing speed is increased, the printing pulse width becomes more narrow with a greater margin for quantization error in the printing pulse width. This means that printing image quality cannot be improved by simply increasing the number of gray scale levels, since this results in an increased quantization error at each gray scale level. We determined the optimum number of gray scale levels by subjectively assessing the image quality. We selected multi-level dithering with 16 gray scale levels using a two-by-two matrix. The use of 16-level dithering is expected to reduce the color differences caused by quantization error to about half of that for density modulation using 64 gray scale levels. Reducing the number of gray scale levels also helped increase the printing speed and simplify the circuits. Any deterioration in printing image quality due to dithering is negligible because the density difference between adjacent dots in a dither matrix at an interval of 125 μm is very small (less than 0.1 O.D.).

3.2 Color conversion circuit

Converting RGB into L*a*b* and L*a*b* into YMCK was done by using conversion tables (stored in ROM) and the interpolation circuits used for these tables. Data for the RGB-to-L*a*b* conversion table was generated from the following theoretical equations:

$$\left. \begin{aligned} L^* &= 116 \left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} - 16 && \left(\frac{Y}{Y_0} > 0.008856\right), \\ L^* &= 903.25 \left(\frac{Y}{Y_0}\right) && \left(\frac{Y}{Y_0} < 0.008856\right), \\ a^* &= 500 \left\{ \left(\frac{X}{X_0}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} \right\}, \\ b^* &= 200 \left\{ \left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_0}\right)^{\frac{1}{3}} \right\}, \end{aligned} \right\} \dots\dots\dots (2)$$

where X, Y, Z, X₀, Y₀ and Z₀ are given by

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.6067 & 0.1736 & 0.2001 \\ 0.2988 & 0.5868 & 0.1144 \\ 0 & 0.0661 & 1.1150 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix},$$

$$\begin{aligned} X_0 &= 0.9804, \\ Y_0 &= 1.0000, \\ Z_0 &= 1.1812. \end{aligned}$$

Calculating the L*a*b*-to-YMCK conversion table is very complicated. The spectral reflectances of YMCK color patches for all gray scale levels must first be measured, then L*a*b* must be calculated by substituting after X, Y, and Z into Equation (2).

$$\left. \begin{aligned} X &= k \sum_{\lambda} \beta_{\lambda} H_{\lambda} \bar{x}(\lambda) \Delta\lambda, \\ Y &= k \sum_{\lambda} \beta_{\lambda} H_{\lambda} \bar{y}(\lambda) \Delta\lambda, \\ Z &= k \sum_{\lambda} \beta_{\lambda} H_{\lambda} \bar{z}(\lambda) \Delta\lambda, \end{aligned} \right\} \dots\dots\dots (3)$$

where,

$$k = \frac{100}{\sum_{\lambda} H_{\lambda} \bar{y}(\lambda) \Delta\lambda},$$

β_λ: spectral reflectance factor
H_λ Δλ: spectral luminous flux
x̄(λ), ȳ(λ), z̄(λ): color matching functions.

The data for the L*a*b*-to-YMCK conversion table can then be generated.

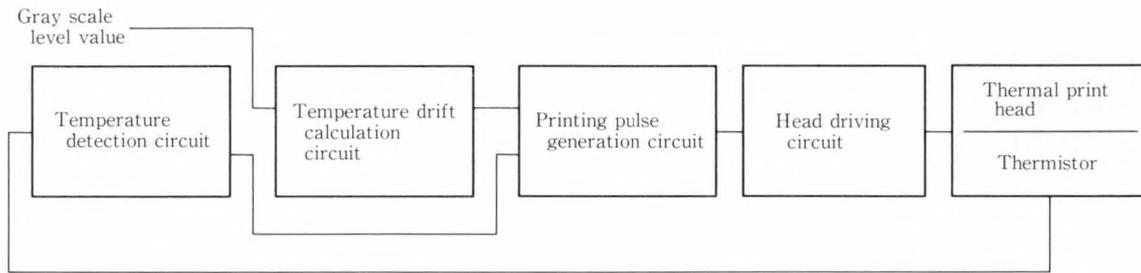


Fig. 5—Block diagram of gray scale level control circuit.

Because these calculations are somewhat complex, the time required to generate table data was increased only slightly. The tables for RGB-to-L*a*b* and L*a*b*-to-YMCK conversion contain data in the five high-order bits of each color (32 Kbytes each).

The interpolation circuit interpolates conversion function *f* for the remaining three low-order bits per color as follows:

$$\begin{aligned}
 & f(i + \Delta i, j + \Delta j, k + \Delta k) \\
 &= \{ f(i + h, j, k) - f(i, j, k) \} \cdot \Delta i/h \\
 &+ \{ f(i, j + h, k) - f(i, j, k) \} \cdot \Delta j/h \\
 &+ \{ f(i, j, k + h) - f(i, j, k) \} \cdot \Delta k/h \\
 &+ f(i, j, k), \quad \dots \dots \dots (4)
 \end{aligned}$$

where,

$$\begin{aligned}
 & 0 \leq \Delta i, \Delta j, \Delta k < h, \\
 & i, j, k: \text{ input signal.}
 \end{aligned}$$

The overall precision of RGB-to-YMCK conversion was thus estimated as approximately six bits per color.

The accuracy of RGB-to-L*a*b* conversion was improved by using separate lookup tables (LUT) for R, G and B (256 bytes each). The color conversion circuit with a 0.75 Mbits table is comparable to a 6 Mbits table. By using this color conversion circuit, an adequate level of accuracy was achieved for reproducing full-color pictorial images using a compact circuit.

4. Gray scale level control circuit

4.1 Structure

The gray scale level control circuit is used to control the printing pulse width to obtain an accurate YMCK density, regardless of the room temperature or temperature drift of the thermal

Table 2. Major causes of and compensation for temperature drift

Type	Phenomenon	Compensation
Cross talk	Effect on adjacent printing dots	Interlaced driving
High-speed temperature drift	Effect of previous print data	Cancellation by post heating
Medium-speed temperature drift	Temperature drift within a page	Prediction of temperature drift from print data
Low-speed temperature drift	Temperature drift of head heat sink	Detection of head heat sink temperature

print head. This circuit consists of sub-circuits for temperature drift compensation and printing pulse generation as shown in Fig. 5. The accuracy of predicting heating element temperature is a very important factor.

The gray scale level control circuit is used to predict the temperature of each heating element based on the room temperature and printing pulse width history data. The printing pulse width is calculated from the gray scale level value and predicted temperature, and is then translated into a series of binary data items that are sent to the thermal print head.

4.1.1 Temperature drift compensation

The print density of the thermal ink-transfer printer depends primarily on the temperature of the thermal print head, where highly defined color reproduction requires precise compensation for temperature drift. Because temperature drift increases rapidly at faster printing speeds, a temperature drift compensation circuit for

various temperature drifts having high (line-by-line) to low (page-by-page) temperature drift speeds is another key to precise and high-speed image printing.

We studied the causes of temperature drift in heating elements. Table 2 lists the major causes⁴⁾ of and compensations for temperature drift. By using interlaced driving and compensation for high, medium and low-speed temperature drifts, respectively⁵⁾, precise print densities were consistently obtained for any given printing patterns at any given room temperature.

An analysis of temperature fluctuations indicated a medium-speed temperature drift with a time constant of a few seconds, which is very important at the targeted printing speed. When the printing speed was doubled, the amount of temperature drift was quadrupled. However, compensation for this phenomenon was previously unknown. According to our analysis, a temperature drift compensation circuit that suppresses print density variations to less than one quarter had to be developed. Through simulation, we discovered that three-quarters of the temperature drift can be compensated for by calculating the amount of heating and cooling for each line using a dominant time constant for medium-speed temperature drift.

The following section describes this compensation process in detail.

4.1.2 Printing pulse width control

Precise compensation for temperature drift

in thermal print heads requires fine control of the printing pulse width. A printing pulse generation circuit (for 16 gray scale levels) and a thermal print head with 27 parallel input lines for each driving LSI were developed for this purpose⁶⁾. As a result, we achieved a printing pulse width control of 17.5 μ s, which satisfied the targeted precision of temperature drift compensation.

4.1.3 Relationship between gray scale level value and print density

Because the relationship between the gray scale level value and print density for Y, M, C and K can be set as required through density modulation, we studied ways to determine the optimum relationship.

The print density for each gray scale level (0 to 15) was specified so that the mixture of Y, M and C for the same gray scale level produced neutral gray, and so that the L^* values for the resulting neutral gray lines on the L^* axis were set at equal distances. The distribution of 4096 ($= 16^3$) colors produced by subtractive color mixtures in the $L^*a^*b^*$ color space proved more uniform than that for a print density for each gray scale level linear to the optical density⁷⁾. This decreased false contours, and improved image quality.

4.2 Compensation circuit for medium-speed temperature drift

The amount of medium-speed temperature drift depends on the printing pulse history data,

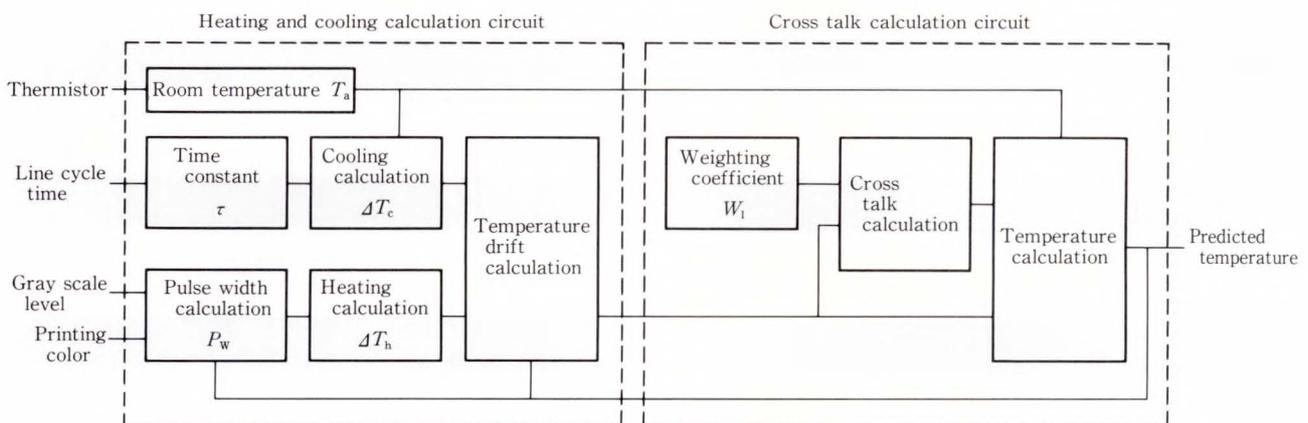


Fig. 6—Block diagram of temperature drift compensation circuit.

the thermal print head structure and heat sink, and the room temperature. We studied how to properly calculate the medium-speed temperature drift (line-by-line) based on the heat sink temperature of the thermal print head and the printing pulse width. Figure 6 shows a block diagram of the compensation circuit. The heating and cooling calculation circuit is used to predict the amount of temperature drift based on the gray scale level, printing color, line cycle time and heat sink temperature.

The temperature drift of the m -th heating element at n -th line due to heating $\Delta Th_{n,m}$ and cooling $\Delta Tc_{n,m}$ are given respectively by Equations (5) and (6) as follows⁸⁾:

$$\Delta Th_{n,m} = \frac{W_{n,m}R}{\sqrt{2\tau}}(\sqrt{t} - \sqrt{t - t_w}), \dots\dots\dots (5)$$

where,

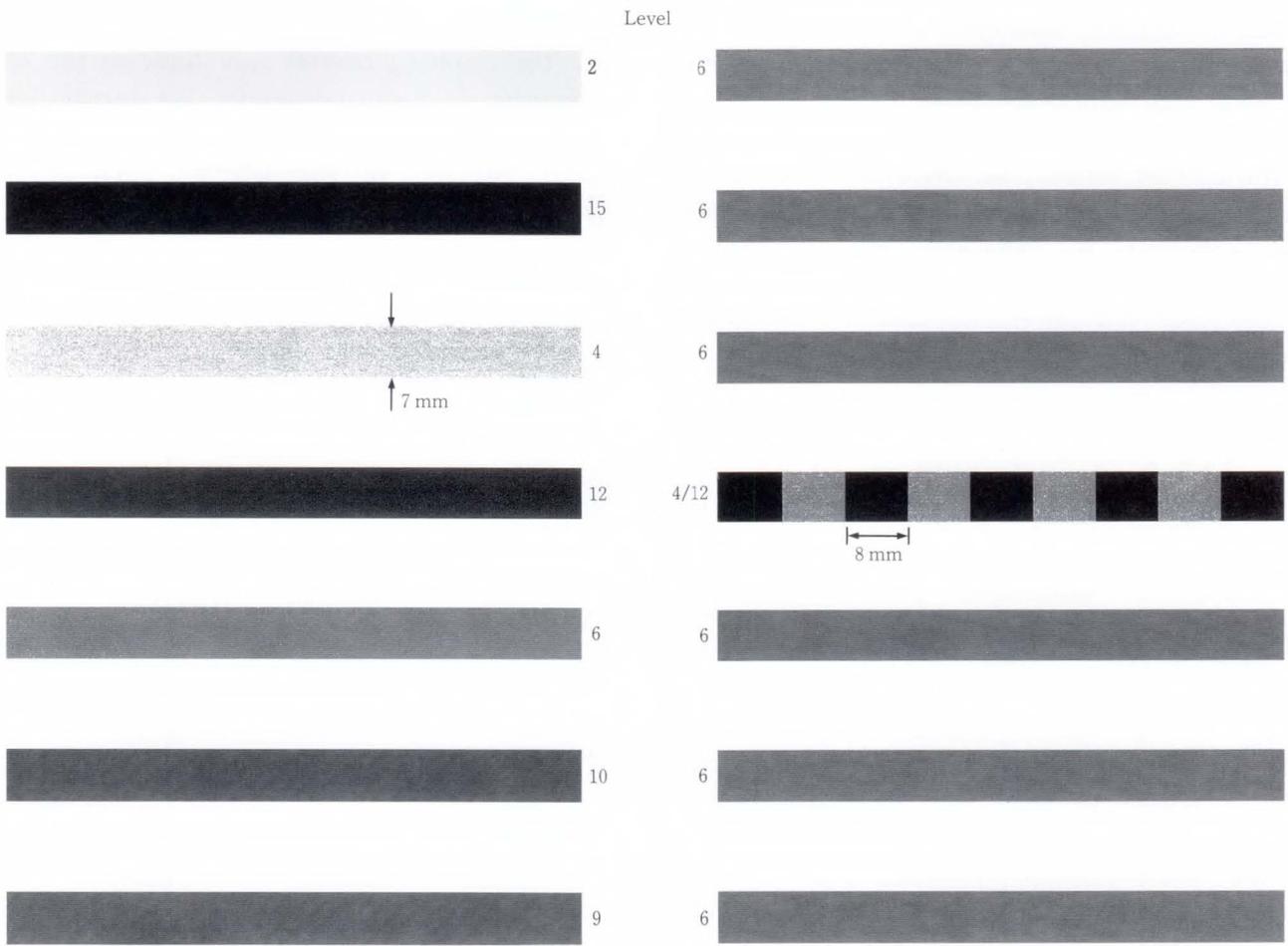
- $W_{n,m}$: heater power
- t_w : pulse width
- τ : time constant
- R : heat resistance
- t : line cycle time.

$$\Delta Tc_{n,m} = (T_{n-1,m} - T_a) \exp\left(\frac{-t}{\tau}\right), \dots\dots\dots (6)$$

where,

- $T_{n,m}$: heating element temperature
- T_a : room temperature.

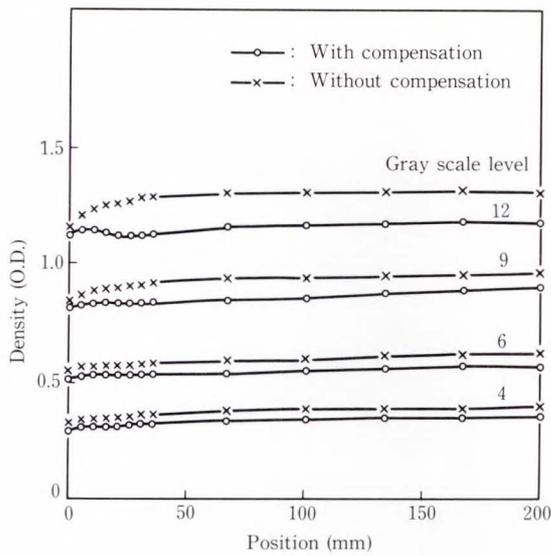
The cross talk calculation circuit is used to calculate the amount of cross talk and predict the heating element temperature $T_{n,m}$ according to the following equation:



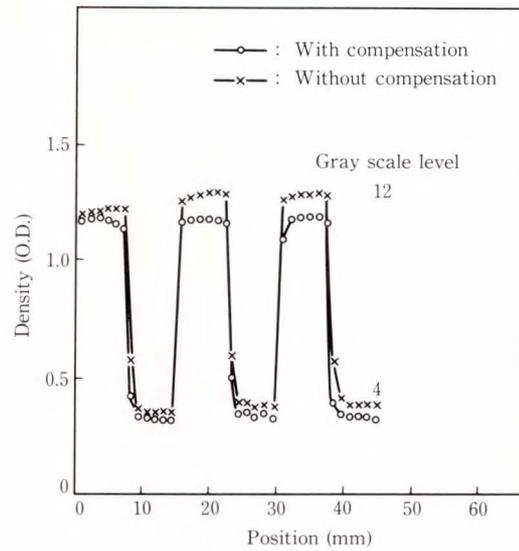
a) Constant gray scale level pattern

b) Regularly modulated gray scale level pattern

Fig. 7—Test pattern used in temperature drift compensation experiment (full size).



a) Print density for constant gray scale level pattern



b) Print density for rectangularly modulated gray scale level pattern

Fig. 8—Characteristics of medium-speed temperature drift.

$$T_{n,m} = T_{n-1,m} + \Delta T h_{n,m} + \Delta T c_{n,m} + \sum_{l \neq 0} W_l \cdot (T_{n-1,m-l} - T_a), \dots \dots \dots (7)$$

where,

W_l : weighting coefficient for cross talk between m -th and $(m \pm 1)$ -th heating elements.

Weighting coefficient, W_l , in Equation (7) is given on an experimental basis.

We experimented with the proper compensation for medium-speed temperature drift. The test patterns shown in Items a) and b) of Fig. 7 were printed, and the variations in print density were measured. Items a) and b) in Fig. 8 show the measurement results. In this experiment, we were able to compensate for about three-quarters of the print density variations caused by medium-speed temperature drift. This indicated that the temperature drift compensation circuit provided sufficient precision for the targeted printer.

By using the compensation circuit for medium-speed temperature drift with those for low and high-speed temperature drifts, we were able to achieve accurate color reproductions of printing patterns, regardless of a room temperature from 16 °C to 40 °C.

5. Conclusion

Figure 9 shows a printing sample produced by this printer. The development of print control circuits for a thermofusible ink sheet with gray scale reproducibility has made it possible to achieve high image quality at high-printing speeds. Color conversion and temperature drift compensation circuits, the keys to our targeted printing objectives, made it possible to reproduce highly defined, full-color images at a printing speed of 16 ms/line, which is double that of printers that use a sublimable ink sheet.

By using an interpolation circuit for the color conversion table stored in ROM instead of a calculation circuit for color conversion equations, we were able to obtain high precision using a compact circuit. The conversion table memory space was also reduced from 6 Mbits to 0.75 Mbits by interpolating the conversion tables.

The temperature drift compensation circuit was able to handle temperature drifts at various time constants. We consider medium-speed temperature drift at a time constant of a few seconds to be a remarkable accomplishment at the targeted printing speed. We successfully compensated for three-quarters of print density variations caused by temperature drift within a



Fig. 9—Printing sample (full size).

page by calculating the temperature drift line-by-line for all heating elements.

6. Acknowledgement

The authors would like to express their thanks to Mr. T. Matsuda of Fujitsu for his encouragement and assistance.

References

- 1) Masuda, S., Kihara, N., and Majima, O.: Color Video Picture Printer. *IEEE Trans. Consum. Electron.*, **CE-28**, 3, pp. 226-232 (1982).
- 2) Kutami, M., Ueda, H., Kiyota, K., and Matsuda, T.: A New Thermal Transfer Ink Sheet for Continuous-tone Full Color Printer. IGC Annual Conf. Electronic Imaging, **1**, pp. 575-580 (Nov. 1987).
- 3) Moroo, J., Mikami, T., and Mori, M.: An estimation of color space for the Laplacian. (in Japanese), IEICE Jpn. National Conf. Inf, Sys., **1**, p. 223 (Nov. 1987).
- 4) Hatabe, E., Hida, T., Yamashita, H., Endo, T., and Uchida, A.: A Thick-Film Thermal Print Head with Driver Circuit. (in Japanese), *Mitsubishi Denki Giho*, **57**, 6, pp. 428-431 (June 1983).
- 5) Noda, T., Mikami, T., and Satoh, T.: Temperature control for high-speed gray scale thermal printing. (in Japanese), IECE Jpn. Genral National Conf., **5**, p. 89 (March 1984).
- 6) Noda, T., Mikami, T., and Abe, F.: Ptint Control Subsystem for Full-Color Thermal Ink-Transfer Printer-2, (in Japanese), IEICE Jpn. General National Conf., **5**, p. 155 (March 1987).
- 7) Mikami, T., Inoue, H., Mori, M., and Abe, F.: Full-Color Thermal Ink-Transfer Printer. (in Japanese), Proc. 17th Joint Conf. Image Tech., pp. 165-168 (Dec. 1986).
- 8) Mortenson, K.E.: Transistor Junction Temperature as a Function of Time. *Proc. IRE*, **45**, p. 504 (Apr. 1957).



Tomohisa Mikami

Color Imaging Devices Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Electronic Eng.
Hokkaido University 1974
Master of Electronic Eng.
Hokkaido University 1976
Specializing in Imaging Devices Control



Toshio Konaka

Color Imaging Devices Laboratory
FUJITSU LABORATORIES, ATSUGI
FUJITSU Technical School 1975
Specializing in Imaging Devices Control



Tsugio Noda

Color Imaging Devices Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Instrumentation
Keio University 1978
Specializing in Image Processing
Hardware Design

Overseas Offices

Algiers Office

6, Boulevard Colonel, Amirouche,
Alger, Algeria
Telephone: (213)-63-42-55
Telex: 408-67522

Amman Office

P.O. Box 5420, Ammán, Jordan
Telephone: (962-6)-662417

Bangkok Office

Room 405, 4th Floor, Dusit Thani Bldg.,
1-3, Rama IV, Bangkok, Thailand
Telephone: (66-2)-236-7930
Telex: 86-87956

Beijing Office

Room 1601, Lubao Dasha
Jianguomen-wai Dajie, Beijing, China
Telephone: (86-1)-512-3610
Telex: 85-22405

Bogotá Office

Cra. 13 No. 27-50,
Edificio Centro Internacional Tequendama,
Oficina No. 326, Bogotá, D.E. Colombia
Telephone: (57-1)-296-8076
Telex: 35-45408

Hawaii Branch

6660 Hawaii Kai Drive, Honolulu,
HI 96825, U.S.A.
Telephone: (1-808)-395-2314
Telex: 705-7430204

Jakarta Office

16th Floor, Skyline Bldg.,
Jalan M.H. Thamrin No. 9, Jakarta, Indonesia
Telephone: (62-21)-333245
Telex: 73-61144

Kuala Lumpur Office

Letter Box No. 47, 22nd Floor,
UBN Tower No. 10, Jalan P. Ramlee,
50250, Kuala Lumpur, Malaysia
Telephone: (60-3)-238-4870
Telex: 84-31005

London Office

2, Longwalk Road, Stockley Park,
Uxbridge, Middlesex, UB11 1AB,
England
Telephone: (44-1)-573-4444
Telex: 51-263871

Munich Office

c/o D V 4 Siemens AG, Otto-Hahn-Ring 6,
D-8000, München 83,
F.R. Germany
Telephone: (49-89)-636-3244
Telex: 41-5210935

New Delhi Office

Flat No. 517-B, 5th Floor,
World Trade Centre, Barakhamba Lane,
New Delhi, 110001, India
Telephone: (91-11)-331-1311
Telex: 31-65271

New York Office

680 Fifth Avenue, New York,
N.Y. 10019, U.S.A.
Telephone: (1-212)-265-5360
Telex: 23-234969

Paris Office

Batiment Aristote,
Rue Olof Palme 94006,
Creteil Cedex, France
Telephone: (33-1)-4-399-0897
Telex: 42-262661

Shanghai Office

Room 1504 Ruijin Bldg.,
205 Maoming Road South,
Shanghai, China
Telephone: (86-21) 336-462
Telex: 30180

Taipei Office

Sunglow Bldg. 66, Sung Chiang Road,
Taipei, Taiwan
Telephone: (886-2)-561-7715
Telex: 785-26773

Overseas Subsidiaries

		Telephone	Telex (1-408)-562-1000
FKL Dong-Hwa Ltd.	338-13, Dae Hong-Ri, Sunghwan-Eub, Chunwon-Gun, Chungnam, Republic of Korea	(82-417)-4-3660	—
Fujian Fujitsu Communications Software Limited	Wuliting Fuma Road, Fuzhou, Fujian, China	(86-591)-556280	92147 PTAJ CN
Fujitsu America, Inc.	3055 Orchard Drive, San Jose, CA 95134-2017, USA	(1-408)-432-1300	230-176207
Fujitsu Australia Ltd.	475 Victoria Ave., Chatswood, NSW 2067, Australia	(61-2)-410-4555	71-25233
Fujitsu Business Communications of America, Inc.	3190 Mira Loma Ave., Anaheim, CA 92806, USA	(1-714)-630-7721	230-685571
Fujitsu Canada, Inc.	6280 Northwest Drive, Mississauga, Ontario, L4V 1J7 Canada	(1-416)-673-8666	21-6-968132
Fujitsu Component (Malaysia) SDN. BHD.	No. 1 Lorong Satu Kawasan Perindustrian Parit Raja, 86400 Batu Pahat, Johor, Malaysia	(60-7)-481392	84-60577
Fujitsu Component of America, Inc.	3320 Scott Boulevard, Santa Clara, CA 95054-3197, USA	(1-408)-562-1000	25-910-338-0190
Fujitsu Customer Service of America, Inc.	12670 High Bluff Drive, San Diego, CA 92130-2013, USA	(1-619)-481-4004	25-910-337-1127
Fujitsu Deutschland GmbH	Rosenheimer Strasse 145, D-8000, München 80, F.R. Germany	(49-89)-413010	41-17-897106
Fujitsu do Brasil Comunicação Eletrônica Máquinas e Serviços Ltda.	Rua Manoel da Nóbrega, 1280-2º Andar, C.E.P. 04001, São Paulo, SP, Brazil	(55-11)-885-7099	38-1121777
Fujitsu España, S.A.	Edificio Torre Europa 5a, Paseo de la Castellana, 95, 28046 Madrid, Spain	(34-1)-581-8000	52-23887
Fujitsu Europe Ltd.	2 Longwalk Road, Stockley Park, Uxbridge, Middlesex, UB11 1AB, England	(44-1)-573-4444	51-263871
Fujitsu GTE Business Systems, Inc.	2411 West 14th Street, Tempe, AZ 85281, USA	(1-602)-921-5900	—
Fujitsu Hong Kong Ltd.	Room 2521, Sun Hung Kai Centre, 30 Harbour Road, Hong Kong	(852-5)-742917	802-62667
Fujitsu Imaging Systems of America, Inc.	Corporate Drive, Commerce Park, Danbury, CT 06810, USA	(1-203)-796-5400	—
Fujitsu Italia S.p.A.	Via Melchiorre, Gioia No. 8, 20124 Milano, Italy	(39-2)-657-2741	43-350142
Fujitsu Korea Ltd.	9th Floor, Korean Reinsurance Bldg., 80, Susong-Dong, Chongro-Gu, Seoul Special City, Republic of Korea	(82-2)-739-3281	801-24553
Fujitsu Microelectronics Asia Pte. Ltd.	No. 2 Second Chin Bee Road, Jurong Town, Singapore 2261, Singapore	(65)-265-6511	87-22194
Fujitsu Microelectronics Inc.	3545 North First Street, San Jose, CA 95134-1804, USA	(1-408)-922-9000	25-910-671-4915
Fujitsu Microelectronics Ireland Ltd.	Unit D Greenhills Centre, Greenhills Rd., Tallaght, Dublin 24, Ireland	(353-1)-520-744	500-30152
Fujitsu Microelectronics Italia S.r.l.	Centro Direzionale, Milanofiori, Palazzo A2, 1-20090 Assago-Milano, Italy	(39-2)-824-6170	—
Fujitsu Microelectronics Ltd.	Vaughan St., West Gorton, Manchester M12 5DU, England	(44-61)-230-6262	51-667403
Fujitsu Microelectronics Pacific Asia Ltd.	8th Floor, Tsim Sha Tsui Centre, 66 Mody Road, Kowloon, Hong Kong	(852-3)-7320100	802-31959
Fujitsu Microsystems of America, Inc.	3025 Orchard Parkway, San Jose, CA 95134-2017, USA	(1-408)-434-1160	230-176207
Fujitsu Mikroelektronik GmbH	Lyoner Strasse 44-48, Arabella Centre 9, OG./A, D-6000 Frankfurt am Main 71, F.R. Germany	(49-69)-66320	41-411963
Fujitsu Network Switching of America, Inc.	4403 Bland Road, Somerest Park, Raleigh, N.C. 27609	(1-919)-790-2211	—
Fujitsu New Zealand Ltd.	6th Floor, National Insurance House, 119-123 Featherston Street, Wellington, New Zealand	(64-4) 733-420	74-30133
Fujitsu Nordic AB	Torggatan 8, S-171 54 Solna, Sweden	(46-8)-764-7690	54-13411
Fujitsu Philippines, Inc.	2nd Floor, United Life Bldg., Pasay Road, Legaspi Village, Makati, Metro Manila, Philippines	(63-2)-85-4951	75-22288
Fujitsu (Singapore) Pte. Ltd.	200, Cantonment Road, #11-01 South Point, Singapore 0208, Singapore	(65)-22240159	87-20138
Fujitsu Systems of America, Inc.	12670 High Bluff Drive, San Diego, CA 92130-2013, USA	(1-619)-481-4004	25-910-337-1127
Fujitsu Vitória Computadores e Serviços Ltda.	Avenida Nossa, Senhora da Penha, 570-8º Andar, Praia do Canto-Vitória-Espirito Santo, Brazil	(55-27)-225-0355	38-272218
Intellistor, Inc.	2120 Miller Drive, Longmont, Colo 80501, USA	(1-303)-678-0697	—
Tatung-Fujitsu Co., Ltd.	5 Floor Tatung Bldg., 225, Nanking East Road 3rd Section, Taipei, Taiwan	(886-2)-713-5396	785-26209

(The information here as of May 31, 1988)

