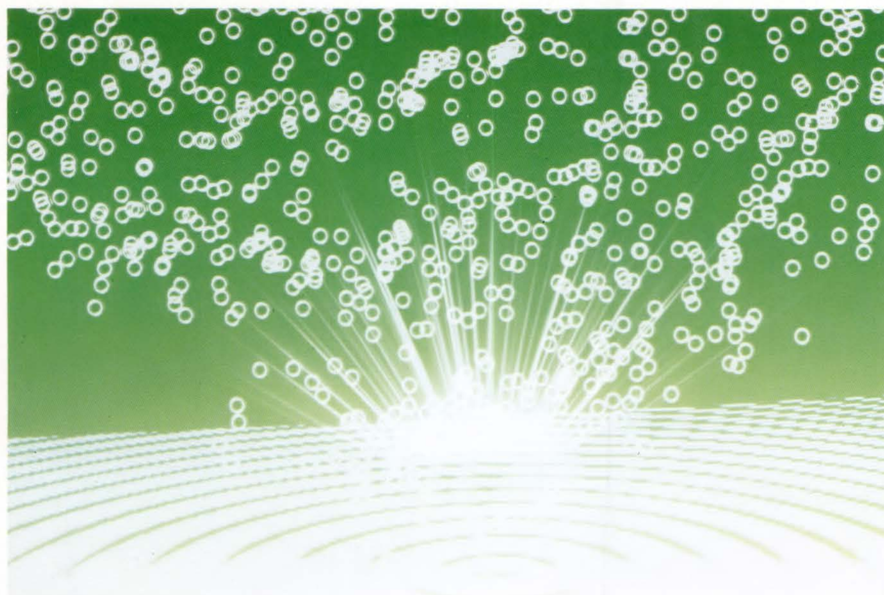
ISSN 0016-2523

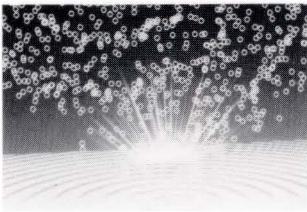
FUJITSU SCIENTIFIC & TECHNICAL JOURNAL

Spring 1988 VOL.24, NO.1



FUJITSU

FUJITSU Sci. Tech. J., **24**, 1, pp. 1-69, Kawasaki, March, 1988



This Issue's Cover

The remarkable advances being made in semiconductor development are having great impact on our daily lives.

Fujitsu is aggressively pursuing the research and development of new-dimensional semiconductors to achieve higher levels of integration for semiconductor devices based on crystal growth technology, which controls the flow of electrons, and to reduce the number of required integrated circuits.

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 Togei, 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 (March 30, 1988)

FUJITSU SCIENTIFIC & TECHNICAL JOURNAL

Spring 1988 VOL.24, NO.1

CONTENTS

Review

1 **Automated Visual Inspection for Printed Circuit Boards**

● Moritoshi Ando ● Masato Nakashima ● Takefumi Inagaki

Papers

24 **Development of Advanced Solid CAD System ICAD/SOLID**

● Yutaka Tanaka ● Takahiro Ueda ● Koichi Akatsuka

35 **A Real-Time Hough Transform Processor**

● Keishi Hanahara ● Tsugito Maruyama ● Takashi Uchiyama

47 **Reproducible Nb/AlO$_x$/Nb Josephson Junctions**

● Shiro Ohara ● Takeshi Imamura ● Shinya Hasuo

54 **Resonant-Tunneling Hot Electron Transistors (RHETs) Using an InGaAs/In (A1Ga) As Heterostructure**

● Kenichi Imamura ● Shun-ichi Muto ● Naoki Yokoyama

60 **Selenium-Indium-Antimony Alloy Film for Erasable Optical Disks**

● Naga-aki Koshino ● Ken-ichi Utsumi ● Yasuyuki Goto

UDC 620.179:621.3.049.73

Automated Visual Inspection for Printed Circuit Boards

● Moritoshi Ando ● Masato Nakashima ● Takefumi Inagaki

(Manuscript received March 4, 1988)

Fujitsu has been developing automatic visual inspection systems for printed circuit boards (PCBs) since 1979. This paper describes newly developed inspection technologies for intermediate products of PCBs such as photomask patterns, inner layer copper patterns, plated through holes of multilayer boards, and mounted components. New techniques were developed in such areas as sensing optics and inspection algorithms.
The systems developed have been installed at Fujitsu's plants and have contributed to the high reliability of PCBs.

1. Introduction

Production lines for manufacturing printed circuit boards (PCBs) have mostly been automated except for visual inspection. This inspection has conventionally been done with the naked eyes using a magnifier, and it has been very difficult to automate.

Human inspectors have difficulties in keeping stable and reliable inspection. The miniaturization of printed wiring board (PWB) patterns used in computers has made it difficult for humans to do an efficient job of inspecting these products. It is thus necessary to develop automatic inspection techniques to insure product reliability.

Image data processing has been undergoing dramatical developments over last ten years, and some automated inspection systems have been developed. But there are still some remaining difficulties in image data processing when used in practical systems. The reason is that it is difficult for image data processing to be accurate when evaluating a wide variety of criteria such as color, surface roughness, and light reflectance of the inspection targets. Therefore, automization of inspection systems depends heavily on image capture techniques that are able to accomplish this.

This paper describes newly developed sensing methods and inspection algorithms, that make automatic systems possible.

2. Overviews of development

2.1 Roles of inspection for PCB production

Figure 1 shows the process of PCB production. Visual inspection is performed at each step. In the production line, photomasks may have several defects, and inner layer sheets may also have some defects. These intermediate products must be inspected and repaired before proceeding to the next process. Here, the PCBs are finished products where components are mounted on PWBs.

Figure 2 shows an evaluation of the inspection cost at each stage of PCB production. As with any manufacturing process, the cost of defects increases as items get closer to the finished product. In the case of PCB production, it is also remarkably cost effective to inspect at early stages and remove the defective pieces as soon as possible.

2.2 History of development

Fujitsu developed an automatic photomask inspection system in 1979, the first of its kind. This system used laser scanning optics

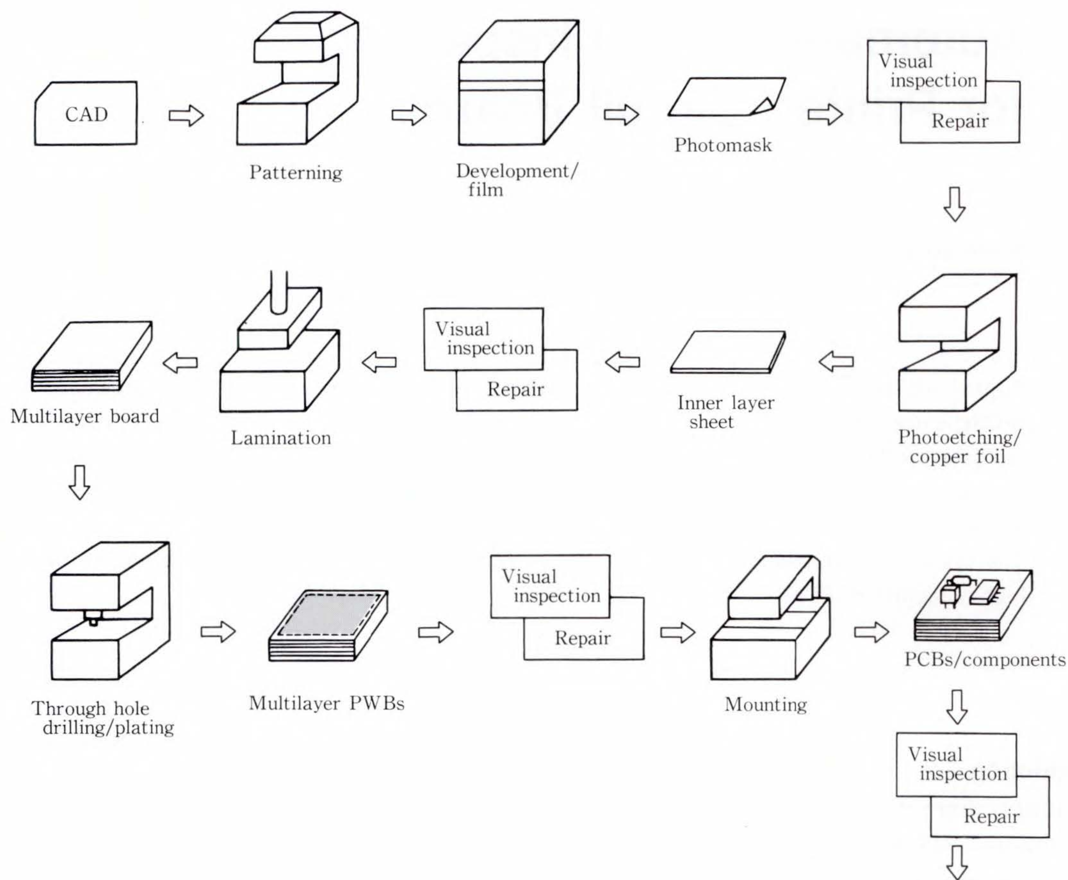


Fig. 1—PCB production.

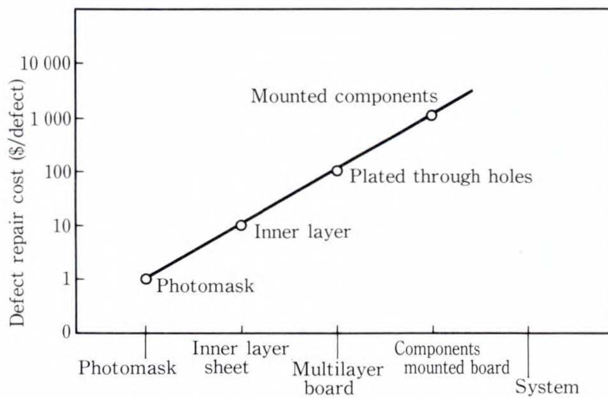


Fig. 2—Economical evaluation of inspection.

and a specialized algorithm to detect non-permissible defects only.

Figure 3 shows the history of development as it progressed from inspecting photomasks, until it reached the final stage of inspecting mounted components on the PCBs. The inspection technology becomes more complex as the objects approach the finished products. For example, photomask inspection handles two-dimensional binary image data, and inspection for mounted components uses three-dimensional data.

3. Key technologies

3.1 Sensing and optical processing

Image capture of PCBs should be performed at a speed of more than 40 MHz, and at a contrast greater than three.

Table 1 lists the sensing types used. Light sources and detectors are chosen according to the object. Special optical processing, such as filtering, masking, and combination optics, have been developed for high contrast image capture.

Table 1. Sensing types

Objects	Sensing data types	Light source	Detector, process
Photomask	Spot light transmittance	He-Ne laser	Photomultiplier
Inner layer	Reflectance (3-D SCAN)	He-Ne laser	Photomultiplier + spatialfilter
Plated through hole	Leakage-light transmittance	Halogen lamp	CCD linesensor + roller mask
Mounted components	Structured light reflectance	He-Ne laser	CCD areasensor
	Diffused-light reflectance	Halogen lamp	CCD areasensor

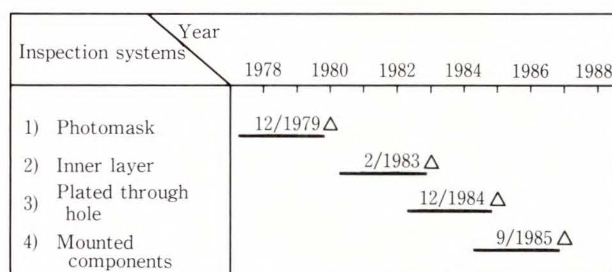


Fig. 3—History of development.

3.2 Inspection algorithms

Inspection algorithms have two requirements, which are necessary to make the systems practical.

1) Non-reference inspection
2) Non-permissible defect detection

The former is for easy inspection preparation. Algorithms for non-reference is based on pattern rule judgement and pattern feature extraction.

The latter is necessary for easy repairs of defects. Algorithms for non-permissible defect detection is based on the combination of pattern width measurement, which detects pattern defects, and feature extraction, which detects random defects, and together they detect non-permissible defects selectively.

4. Photomask pattern inspection

If a photomask has any defects, they are transcribed onto the PWBs during the photo-etching process. Production of the photomask free of defects is an important requirement for improving the quality and production efficiency of PWBs.

Inspections and repairs are necessary steps in the production process, because each photo-mask may have several defects. The pattern is generated with a computer-aided-design system (CAD), and then developed on photographic film. Scratches on the film or dust in the solution can cause most of the pattern defects with sizes ranging from ten to several hundred microns. Sometimes there are several hundreds of defects in a single mask.

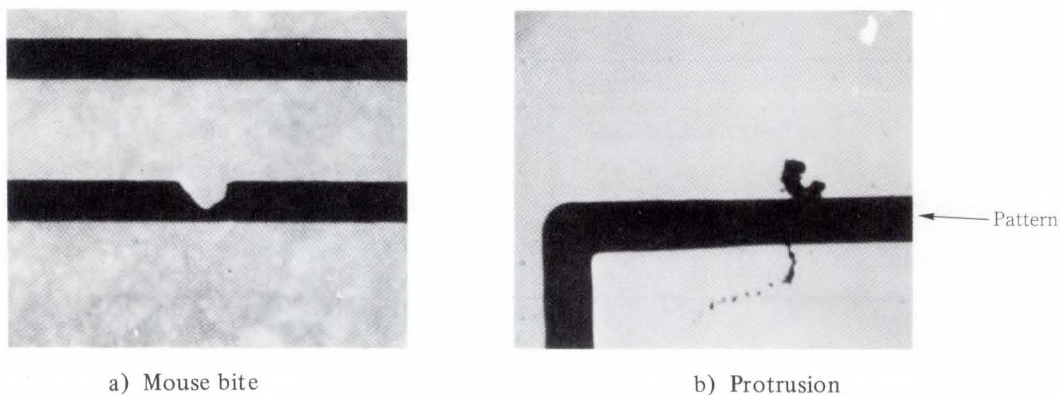
These patterns have always been inspected and repaired by human inspectors. However, as photomask patterns have become finer and more complex, it has become difficult for inspectors to inspect them. Therefore, an automatic inspection method is required.

4.1 Requirements

An automatic photomask pattern inspection system must be capable of sensing the pattern with a high *S/N* ratio, and discriminating between permissible and non-permissible defects. Details of these requirements are described in the following section.

4.1.1 Requirements for sensing

Only patterns that cause defective copper patterns need be detected. For example, dust on a black pattern does not cause any defects because the photoetching process uses transmitted light. However, dust on a transparent area causes excessive defects on the PWBs. Scratches do not result in defects when they are on the transparent part. However, scratches in a black pattern cause open areas on the PWBs. Therefore, the sensing optics should be able to detect dust located on transparent areas and a scratches located on black patterns.



a) Mouse bite b) Protrusion

Fig. 4—Non-permissible defects.

4.1.2 Requirements for inspection algorithms

The inspection algorithms should detect only the non-permissible defects. Non-permissible defects are defined as those that damage the performance or quality of the PWBs. Permissible defects are ones that do not effect the PWB quality. In other words, permissible defects are either ones that are located in free space areas where defects do not damage the PWB patterns or those that are so small in size that they disappear after the photoetching process is completed.

Figure 4 shows examples of non-permissible defects. Figure 4-a) shows a mouse bite with a depth of 100 μm, and Fig. 4-b) shows a protrusion with a height of 80 μm.

Figure 5 shows the critical areas for judging whether a defect is permissible. The mask surfaces are divided into four inspection areas; lead area, pad area, inside pad area and free area. The critical areas are those within the pattern and up to 100 μm away. Table 2 indicates defect sizes and types in the critical area. The sizes of non-permissible defects are defined according to their locations and types. Smaller defects have to be defected in the lead area, rather than other areas. Their permissible sizes also depend on their types. For example, shorts and breaks are always non-permissible, as are protrusions and mouse bites more than 30 μm wide. In the free area and inside pad area, however, large defects are permissible.

Thus, the sizes of non-permissible defects are determined according to the pattern area and the types of defects.

4.2 Sensing method

A laser beam scanning method is used to detect the pattern. Transmitted light intensity is sensed by a photodetector. The configuration of the system is shown in Fig. 6. A He-Ne laser beam with a 633 nm wavelength is focused by a scanning lens and scans over a mask by means of a rotating polygon mirror. The laser beam diameter on the mask is 10 μm at half the maximum intensity of the Gaussian distribution. The laser beam scans 50 mm in the direction of the x-axis. The effective read length is 41 mm. The mask travels in the direction of the y-axis on an X-Y table in 10 μm steps and is continuously scanned.

A photodetector receives the beam coming through the mask to obtain pattern signals. The laser beam also scans over a grating mask in order to detect the location of the scanning laser beam. These signals are used to control the scanning speed.

The pattern signals are sampled at 10 μm per bit (size of picture element) with the aid of the grating signals.

The laser beam scanning system provides the following advantages:

Table 2. Defect sizes and types in critical area

Locations \ Defect types		Pattern width	Random	
		Mouse-bite Protrusion Excess	Shorts Breaks	Black spots Pinholes
Critical area	Lead area	≧30 μm	≧10 μm	≧30 μm
	Pad area	≧80 μm	≧10 μm	≧30 μm
	Inside pad area		≧20 μm	≧80 μm
Free area		–	–	≧80 μm

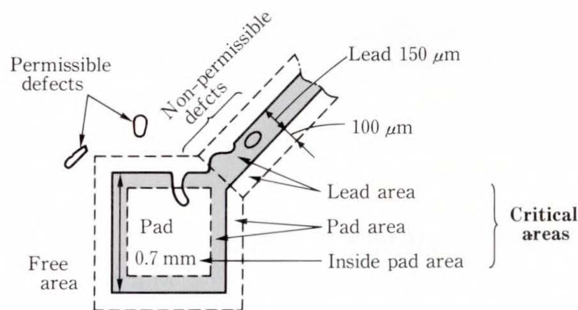


Fig. 5–Critical areas and defects.

1) Spot illumination enables pattern shapes to be read with ease, that is, the S/N ratio is high

Using transmitted light for the sensing system has the advantage that dust on black patterns and scratches on transparent film do not affect the pattern shapes sensed.

2) Long focus enables pattern to be read clearly

A focus depth of ±300 μm is obtained. This gives sufficient depth for the inspection system.

4.3 Inspection algorithms

4.3.1 Algorithm to detect non-permissible defects

The detection logic consists of pattern width measurement and the Sensing Matrix Region Tracing (SMART) method. The pattern width measurement detects the widths of the patterns and SMART detects random defects such as pinholes and black dots. It is necessary to combine these methods to detect most of the non-permissible defects.

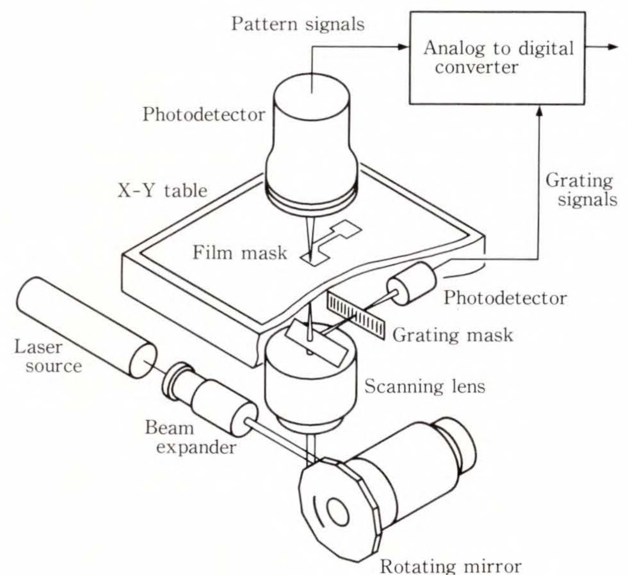


Fig. 6–Sensing optics.

4.3.2 Pattern width measurement

Automatic measurement of pattern width without artwork data for comparison requires the directions and the areas to be measured.

An algorithm using the parallel edge detection method has been devised to obtain this information directly from the patterns to be inspected.

Figure 7 shows how pattern width measurement is done.

1) The edge signals in four directions (0°, 45°, 90°, and 135°) are extracted from the pattern signals.
2) After detecting some sets of parallel edges from the patterns in the recognition windows which have 64 x 64 bit sensors,

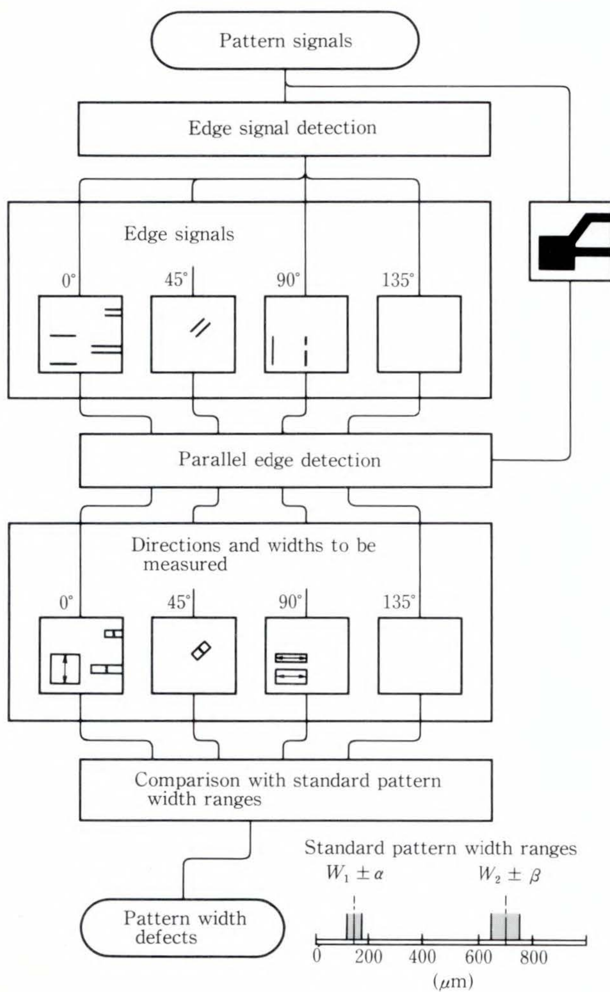


Fig. 7—Pattern width measurement procedure.

measure the pattern width perpendicular to the edges.

3) Compare the measured results with the standard ranges of pattern width, for exmple, $W_1 \pm \alpha$ and $W_2 \pm \beta$. Where W_1 and W_2 represent standard widths and α and β the tolerant ranges. A pattern width defect is detected when the measured result is outside these ranges.

This method can find the directions and areas to be measured directly from the inspected patterns. Consequently, when two or three sets of standard widths and their tolerant ranges are input, the inspection system automatically sets the standard ranges for pattern width. Figure 8 shows the areas inspected by the pattern width measurement method.

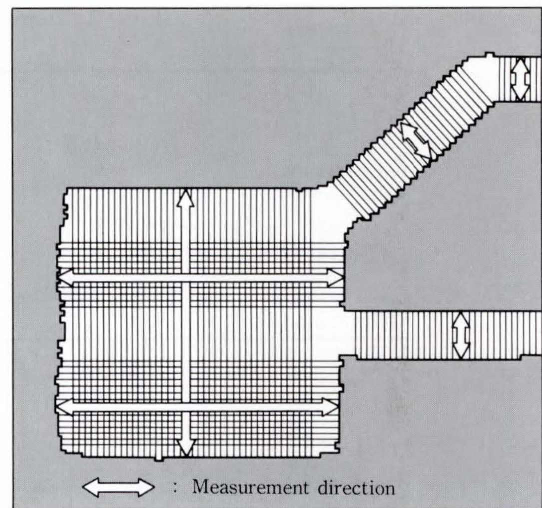


Fig. 8—Areas inspected by pattern width measurement.

4.3.3 SMART detection

The SMART (Sensing Matrix Region Tracing) method was developed to detect defects within a certain distance inside or outside pattern edges.

Figure 9 shows the inspection areas and matrices of detection bits of SMART. As specified number of detection bits are lined up, beginning from the pattern edge inside pattern area, or outside of the pattern within the critical areas. This SMART sensor is set in eight directions, each separated by 45 degrees.

The SMART method detects random defects along normal pattern edges and inspects clearances between neighboring patterns. The pattern detection window is 16 bits square. Figure 9-b) shows four groups of the sensors employed.

This method can also be applied to detect the space between adjacent patterns.

4.4 System configuration

Figure 10 shows the automatic mask inspection system for the PWB photomask. The system consists of an inspection optical unit, a detection logic unit, an inspection result output unit, and a system control unit. Figure 11 shows a photograph of the system. The main specifications are listed in Table 3.

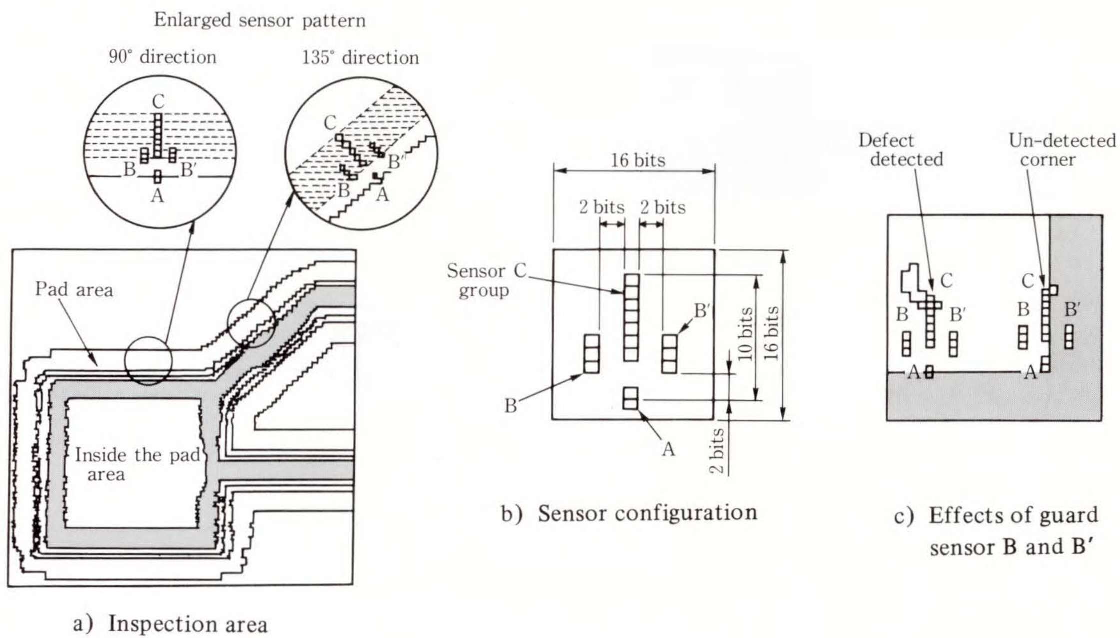


Fig. 9—Areas inspected by SMART detection and SMART matrices.

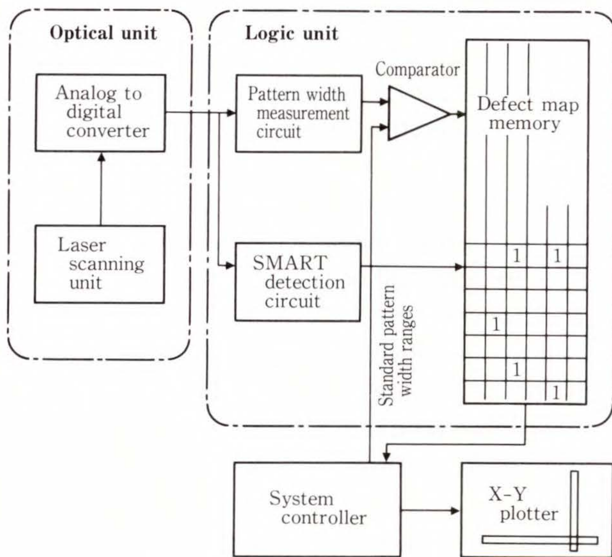


Fig. 10—Block diagram of the inspection system.

Table 3. Main specification of the photo mask inspection system

Item	Specifications
Resolution	10 μm
Inspection time	3 min/300 mm square
Output	Defect map (With 2.5 mm square marks)
Size	1.5 x 1.5 x 1.2 mm ($l \times b \times h$)

Fig. 11—Automatic mask pattern inspection system.

4.5 Inspection characteristics

Figure 12 shows the detection rate for non-permissible defects comparing conventional inspection and automatic inspection. It shows that SMART detects 74 percent of the non-permissible defects and pattern width measurement detects 57 percent of them.

By using the pattern width measurement method and the SMART method together, all the critical areas of the mask can be inspected. Pattern width measurement used alone detected 100 percent of the pattern

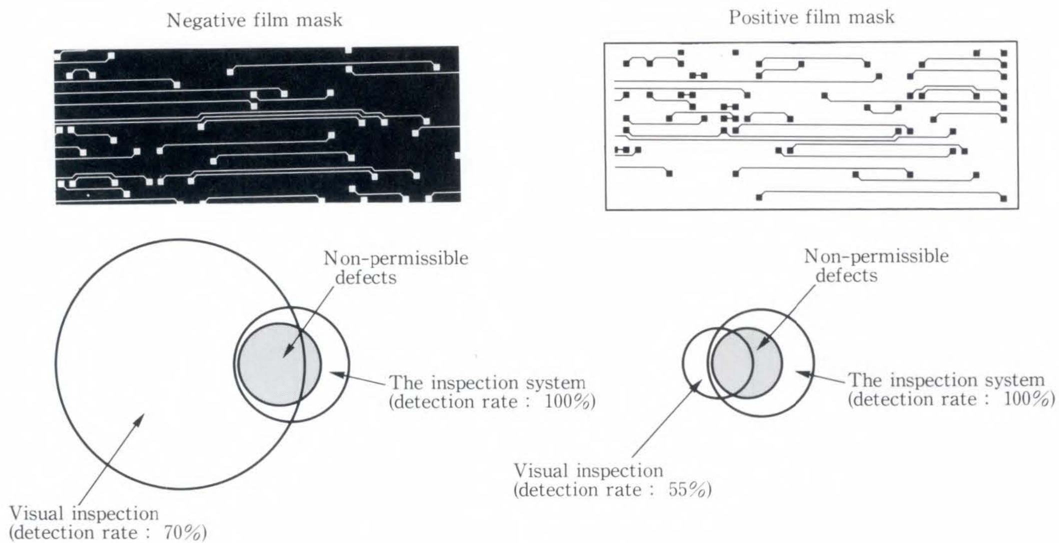


Fig. 12–Detection efficiency for non-permissible detects of the conventional inspection and automatic inspection.

width defects, and SMART used alone detected 100 percent of the random defects. Combining these methods allowed all the non-permissible defects to be detected.

Figure 12 shows a comparison between detection efficiencies of non-permissible defects for visual inspection and the new inspection system. The detection rate of visual inspection is about 70 percent for negative masks and 55 percent for positive masks. These results indicate that visual inspection alone is not sufficient; two or three repeated inspections are required to obtain a detection rate of at least 90 percent. Figure 12 shows these results. Visual inspection detects many permissible defects; this is shown by the large circles.

The automatic inspection system is capable of detecting 100 percent of non-permissible defects for both negative and positive film masks.

5. PWB inner layer pattern inspection

High reliability is required for the inner layers because they are laminated to make multilayer boards. Defects on the inner layer patterns must be detected and repaired before the lamination process.

The inner layers are usually inspected visually and with an electrical continuity tester. The tester can only detect shorts and breaks in the conductive pattern, but cannot detect mouse bites and protrusions which affect reliability. Conventionally, visual inspection with an optical magnifier is combined with the tester to detect all of the defects.

However, the widths of the conductive patterns in newly designed boards are often less than 100 μm, and a minimum resolution of less than 10 μm is needed. It is very difficult for humans to inspect such fine patterns effectively.

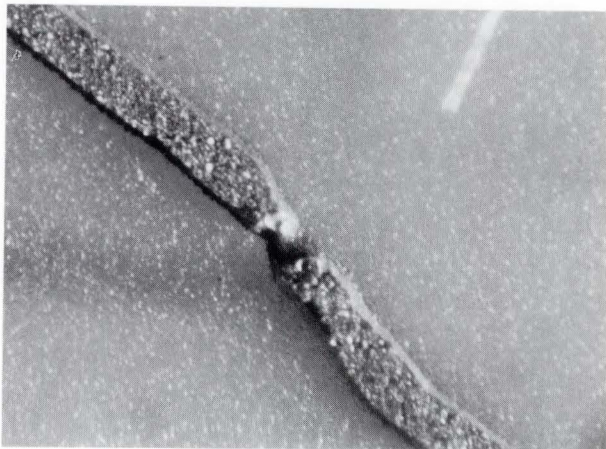
To solve these problems, a new automatic inspection system has been developed that can inspect inner layer conductive patterns using an optical method.

5.1 Requirements

The following three requirements must be met before the inspection system can become practical.

1) Detection of both the top and bottom width

The conductive patterns are formed on epoxy or polymide resin substrate with the etching process. The pattern has a 100 μm width



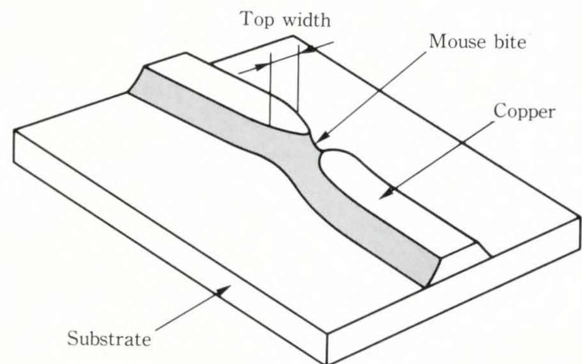
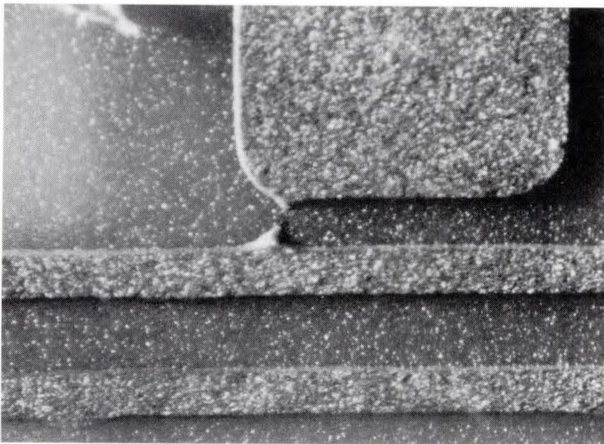


Fig. 13–Copper mouse bite.



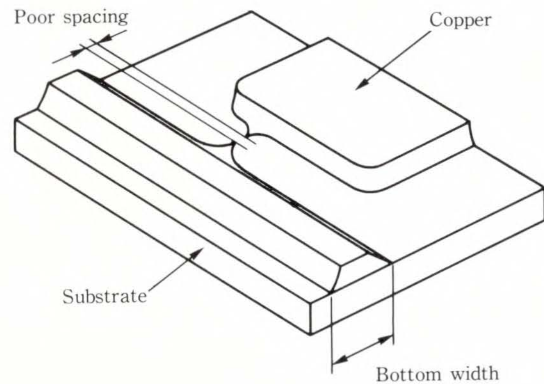


Fig. 14–Poor spacing.

and 35 μm thickness. The cross section is a trapezoid. To assure conductivity of the pattern, mouse bites, poor spacing and protrusions must be detected in addition to shorts and breaks. The inspection of both the top and bottom width of the conductive pattern is required.

Mouse bites usually have a narrower pattern width, but they sometimes have the correct bottom width and too thin a thickness. To detect this type of mouse bite, as shown in Fig. 13, it is necessary to measure the top width of the conductor.

Poor spacing sometimes has the correct top width, but a defective bottom width as shown in Fig. 14. In this case, it is necessary to measure the bottom width of the conductor.

A CCD line sensor is used as the sensing device in conventional systems. The top width of the conductive pattern can be detected, but the bottom width of the pattern cannot because the side wall reflects poorly.

2) Detection of all defects

Defects are divided into two types. The first type of defect is found in wide areas such as incorrect width, and the second type is random defects such as mouse bites, protrusions, and opens. The algorithm must be able to extract all of these defect types. It must also be able to filter out permissible defects such as small isolated patterns or dust patterns that do not affect continuity. This is important because if these permissible defects were detected, the yield would decrease.

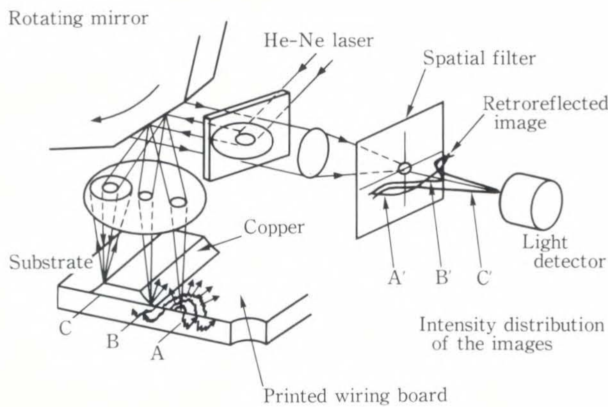
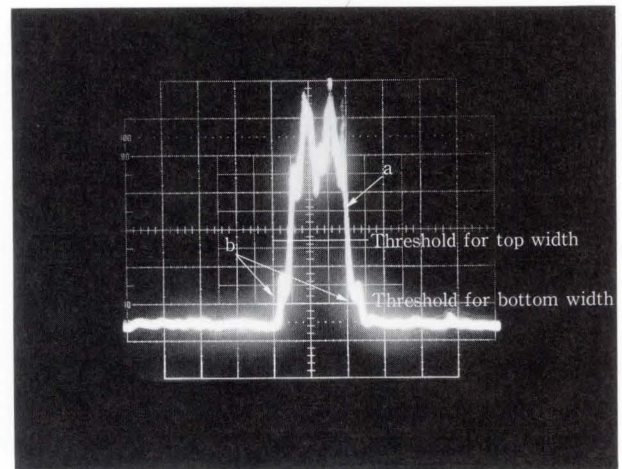


Fig. 15–3D SCAN detection.



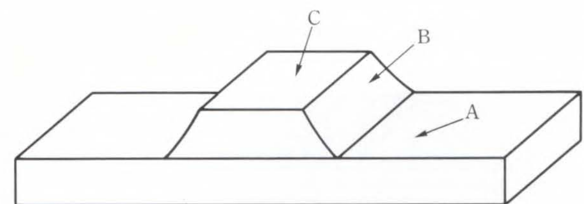


Fig. 16–Copper pattern signal.

3) Easy indication of defect positions

Defects such as excess copper can be repaired by a skilled worker. Easy-to-use output is required so that the defects can be easily found and repaired.

5.2 Sensing method

5.2.1 Conditions of pattern detection

The following conditions are required.

1) Top and bottom width of copper pattern should be detected.
2) Accuracy should be better than 10 μm.
3) Image contrast should be larger than three.
4) Depth of focus should be larger than ±300 μm.

5.2.2 3D-SCAN detection

3D-SCAN detection method was developed to meet the above requirements. A laser scanner is used and both the top and the bottom width of the copper line can be detected. The configuration is shown in Fig. 15. The He-Ne laser beam is deflected with a rotating polygon mirror and is scanned on the inner layer pattern. Deflected light from the inner layer is retroreflected and imaged on a plane by a lens. A spatial filter with a pinhole of diameter 100 μm in the center is set on an imaging plane. The transmitted light through the pinhole is detected with a photomultiplier tube.

When the beam strikes position A, it diffuses into substrate material before it is reflected. The image of beam A′ on the image plane increases in size and has a flat distribution. When the diameter of the filter is set to the same size as the scanning beam, the filter blocks the reflected light beam from the substrate. However, when the laser beam strikes the top of conductor C or side wall B, it does not diffuse and the image sizes of laser beam B′ and C′ are the same as the scanning laser beam. When the filter size is larger than the scanning laser beam, most of the light energy passes through the filter. This method enables distinguishing the substrate from the top of the copper line and side wall line with a sufficient *S*/*N* ratio.

The result is shown in Fig. 16. Large signal "a" at the center corresponds to the reflected light signal from the top of the copper line. Small signals "b" at the sides of main signal "a" correspond to the signals from the side walls of the copper pattern. Here, two threshold levels are used to measure the top and bottom width of the conductive pattern simultaneously.

Defects detected using this method are shown in Fig. 17. Both the top width area

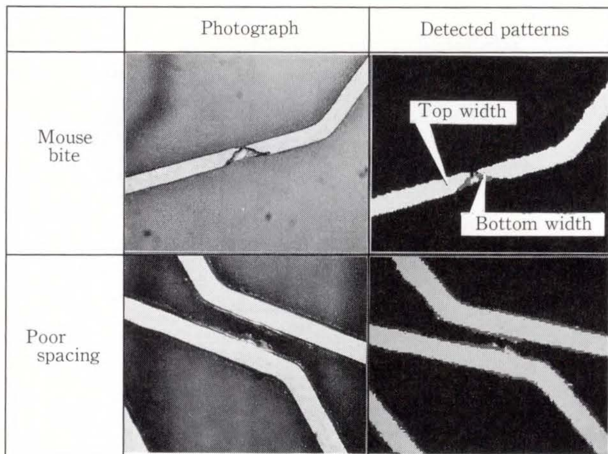


Fig. 17–Examples of defects.

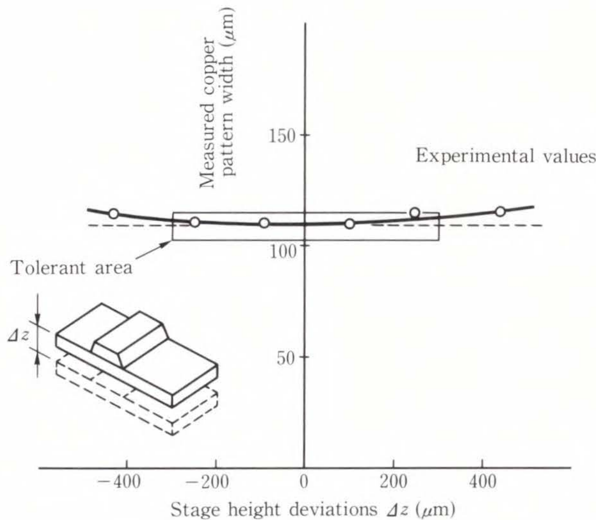


Fig. 18–Stage height deviations vs. measured copper pattern width.

(white) and bottom width area (gray) can be detected.

Figure 18 shows the focal depth obtained which was ±300 μm with an accuracy of ±5 μm.

This system can obtain a deeper focal depth than conventional CCD camera systems.

5.3 Inspection algorithms

The algorithms are a combination of the width measurement and feature extraction methods. They can detect both pattern width defects and random defects.

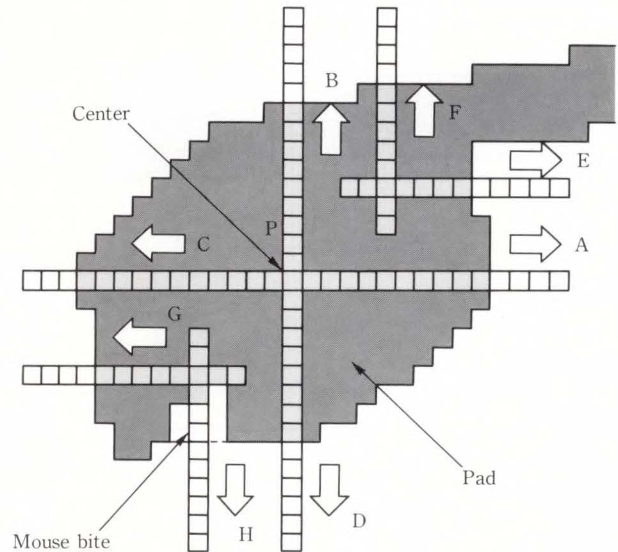


Fig. 19–Pad width measurement method.

5.3.1 Width measurement

The width of the conductor lines are measured to detect such items as defective pattern width and poor pattern spacing.

PWB patterns consist of both leads of 100 μm width and pad patterns of 800 μm size. Two algorithms have been developed for both pad and lead inspection.

5.3.2 SMART detection

The SMART (Sensing MAtrix Region Tracing) can inspect whether the pattern width satisfies the tolerance. The working principle is the same as that developed for photomask inspection as shown in Fig. 9. This SMART sensor is set in eight directions, each separated by 45 degrees. All pattern gap defects can be also detected when the same sensors are applied to the pattern space.

5.3.3 Pad measurement

This algorithm is used to measure the diameter of the pad. The pads are usually rectangular or circular and are sometimes hexagonal. When rectangular shape is inspected, the spacing between two pairs of parallel edges are detected and measured. When a hexagonal shape is inspected, the spacing between three pairs of parallel lines are measured. An example of the sensor for inspecting a hexagonal shape is shown in Fig. 19. A to H indicate line measurement sensors.

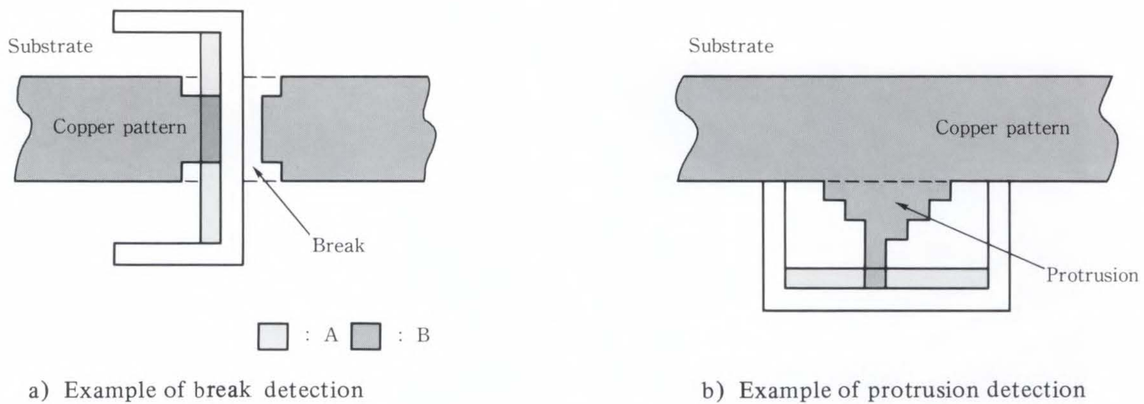


Fig. 20—Local defect detection algorithm.

Fig. 21—Automatic PWB pattern inspection system.

The inspection procedure is as follows.

1) Pad center detection

Pad center is detected when the results of sensor A to C and B to D are equal.

2) Parallel edge measurement

The results of sensor E and G, and F and H are added. These pairs are parallel to A and B. The same value as ones at the center would be obtained if the hexagonal shape is normal.

3) Defect detection

Measured results are compared to the tolerance value. It is determined whether the results are within the tolerance range. Measurements are done only at the center area. The four sensor pair measure the whole area and the results are then compared to the tolerance values.

In the example of Fig. 19, the pad pattern has a mouse bite. When added values F and H become less than the standard values, the pad is determined as defective.

The measurements of a rectangular and a circle pad can be done using these sensors.

5.3.4 Local defect detection

The local defect detection algorithm can detect random defects such as breaks, protrusions and extraneous coppers. It detects abrupt changes in edge orientation which occur at the defect. Figure 20 shows a diagram of the algorithm. Sensor A is a copper pattern detection sensor and sensor B is a substrate detection sensor. When the each sensor detects the corresponding pattern, the pattern is determined as defective. For example, Fig. 20-a) shows an example of a break detection and Fig. 20-b) shows an example of a protrusion pattern. Breaks and protrusions in all direction can be detected by using eight sensors each 45 degrees apart. These sensors inspect both the top and the bottom width pattern. They can detect not only breaks and protrusions of the bottom width pattern, but can also detect mouse bites in the top width.

5.4 System configuration

Figure 21 shows the automatic inspection system, and Fig. 22 shows its configuration.

The system is made up of the following components.

1) System controller

The system controller controls X-Y stage movements, memorizes the positions of defects and controls the threshold values that distinguish between defective and normal patterns.

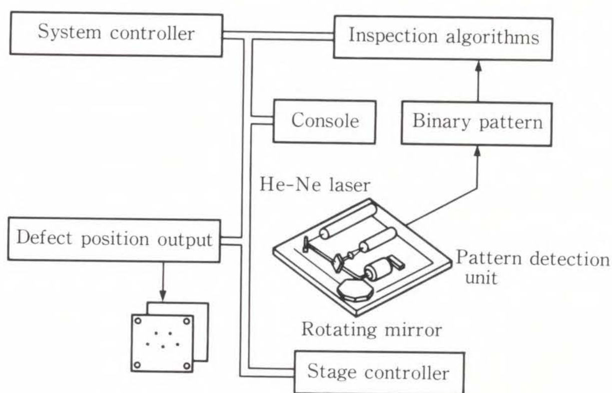


Fig. 22–Schematic diagram of inspection system.

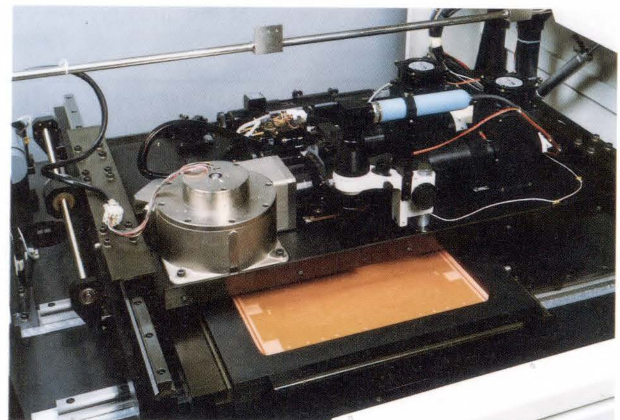


Fig. 23–Optical apparatus.

2) Pattern detector

The pattern detector consists of a laser scanner and an X-Y table as shown in Fig. 23. The laser beam is guided to a twelve faceted mirror rotating at 7 200 rpm. The resolution is set to 10 μm, and the scanning speed is 100 m/s. Reflected light is detected with the 3D-SCAN detector, and both the top and bottom widths are detected. The PWB is firmly attached to the X-Y table with a vacuum. A PWB 300 x 360 mm in size can be inspected in three minutes.

3) Defect detector

The detector to detect defects consists of two width measurement algorithms, one for line patterns and the other for pads, and two feature extraction algorithms, one for a SMART sensor and the other for a protrusion sensor. In this system, all inspection algorithms are implemented in hardware for high speed.

4) Detected results display

The system has two output devices, a CRT display and a hole punch. The CRT displays defect shapes. The defective positions are punched on a paper sheet the same size as the inspected PWB. This is shown in Fig. 24. The holes are 2.56 mm in diameter. The paper is laid over the PWB, and the defects can be seen through the punched holes. This allows the defects to be easily found. The defects are examined to determine whether they can be retouched by a human worker. The repairable defects are retouched through the punched holes. The inspection results for both sides of a PWB are punched out on two consecutive pages. The pages are then wrapped around the PWB, and also protect the PWB from scraches.

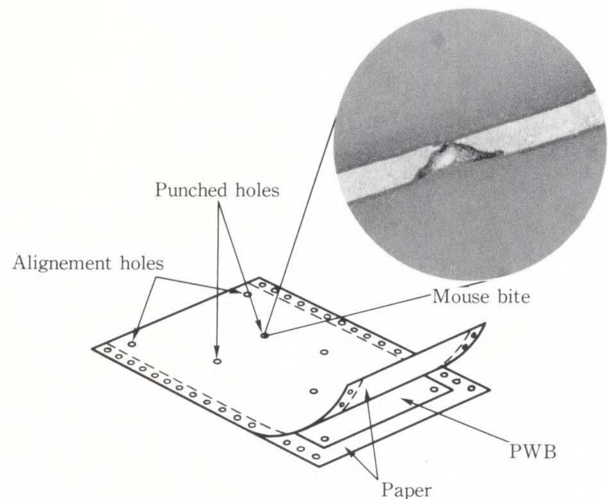


Fig. 24–Defect location reporting method by punched holes.

5.5 Inspection characteristics

The main characteristics of the system are listed in Table 4. This system was implemented at a plant and evaluated. Figure 25 shows the results. PWBs were inspected by both the automatic inspection system and the conventional visual inspection. It shows that the system can detect all types of defects. Visual inspection yielded relatively good results when detecting breaks, shorts, and extraneous cop-

Table 4. Main characteristics of the inner layer inspection system

Items	Specifications
Resolution	10 μm
Inspection time	3 min/300 mm square
Criteria values	Independently selectable
Output	Punched hole sheet (dia 2.56 mm)

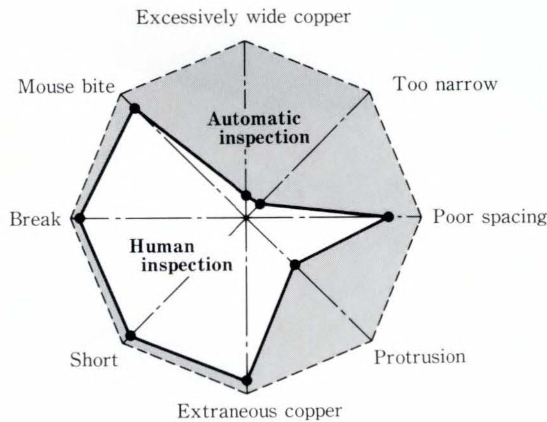


Fig. 25—Comparision of performance between automatic and visual inspection.

pers, however, missed patterns that were excessively wide or too narrow and also missed poorly spaced pattern. Human inspectors also missed protrusions and small mouse bites.

The system shortened the inspection time down to 4 min from what took human inspectors 20 min to 40 min, and could detect not only bottom width defective patterns but also top width defective patterns.

As for inspection speed and resolution, further development has been continued by increasing image processing speed and using high accuracy optics. Up to now, inspection speed has been increased ten fold with 5 μm resolution on the newly developed system.

6. Through hole inspection

PWB technology is getting more complex as LSI technology developes. As a result, the number of inner layers of the PWB are increasing and the diameter of the through hole, smaller. The PWB used for the FACOM M-380 for example, has through holes with diameters

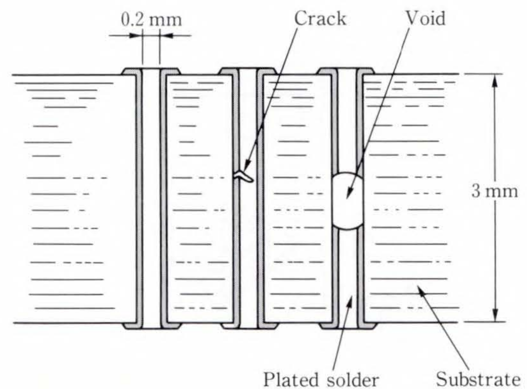


Fig. 26—Cross section of PWB.

as small as 0.2 mm, and aspect ratios as large as 15. Plated through holes connect electrical circuits of inner layers vertically through the substrate with electrodepositioning copper on the hole wall.

Through holes sometimes have defects such as voids and cracks as shown in Fig. 26. Electrical continuity testers and visual inspection are the conventional means of inspecting through holes. The electrical continuity tester can detect plating voids that break the continuity of the circuit, but they cannot detect cracks. Visual inspection is effective in detecting cracks only in holes with low aspect ratio PWBs. It is not efficient for inspecting high aspect ratio through holes. The automatic inspection system for the high aspect ratio through hole is very much required.

6.1 Requirements

The requirements for inspecting plated through holes are high speed and non-reference inspection. The system must be capable of inspecting a 300 mm square PWB within a few minutes. Non-reference inspection is also required because large scale memory is required to store the hole position data.

6.2 Sensing method

6.2.1. Principle of sensing method

A new method called leakage light detection has been developed using the optical methods as shown in Fig. 27.

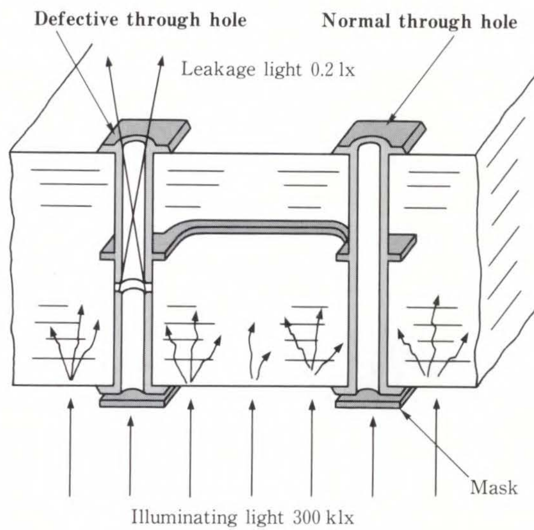


Fig. 27—Leakage light detection.

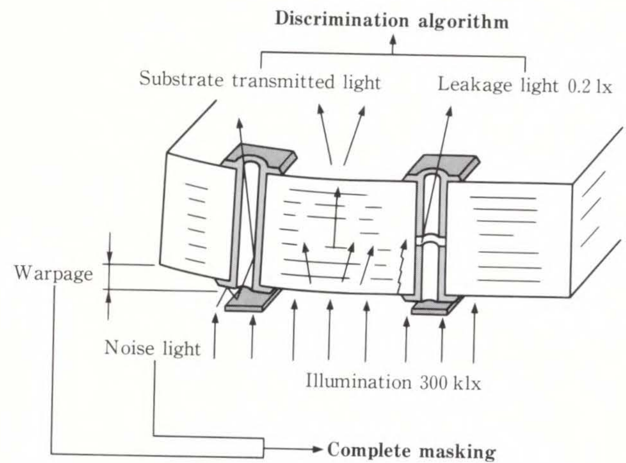


Fig. 29—Technical requirements.

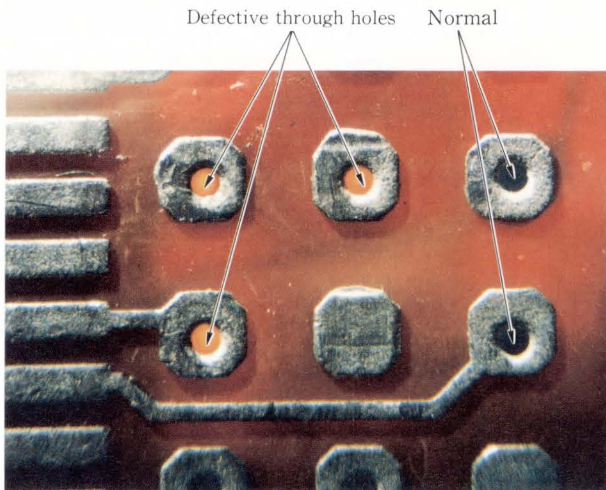


Fig. 28—Leakage light from defective through hole.

The method uses following four facts.

1) Plated through holes must be made of metal cylinders.
2) The substrate is exposed to the inside of the plated through hole.
3) The substrate diffuses and transmits some light.
4) The light can pass through the defects.

The detection process is explained as follows.

One side of the hole is covered with a mask. When the hole is illuminated with intense light, the light diffuses into the substrate. When it reaches a defect, the light is transmitted through the defect, and leakes into the through hole. When the PWB is observed from the other side, only defective through holes can be seen, while normal through holes remain dark. Then this is the "Leakage light detection" method. Figure 28 shows the experimental results. The three bright holes at the left indicate defective through holes, and the two dark holes at the right show the normal through holes.

6.2.2 Leakage light detection

There are three technical requirements to make the method practical as shown in Fig. 29.

1) Detection of weak light

Leakage light intensity from detects is low. For example, when PWBs are illuminated with 300 klx, the leakage light is only 0.2 lx. The system has to be able to detect this weak light.

2) Complete masking

The through hole being inspected must be completely masked. This is so the illuminating light will not leak into the through holes. If it does leak, the *S/N* ratio becomes worse. Therefore the masking has to be complete in spite of PWB warpage as large as 1 mm.

3) Discrimination between defective leakage light and substrate transmitted light

The illuminating light diffuses into the substrate and shines through the substrate. The system must be able to distinguish the transmitted light from the defective leakage light.

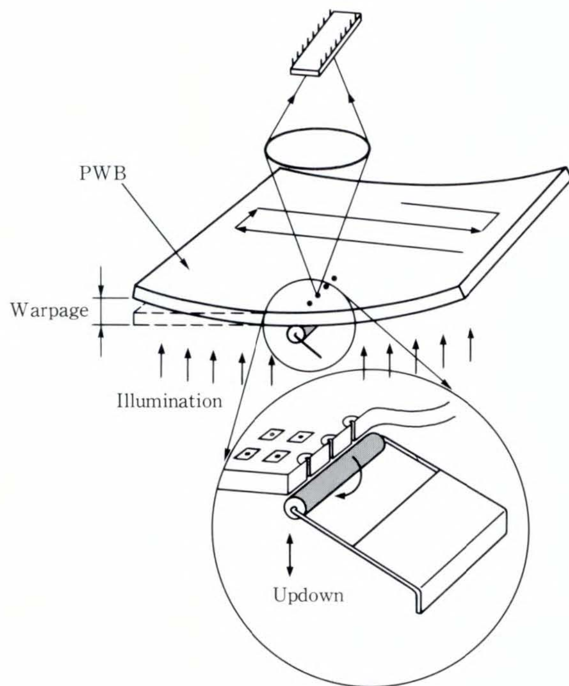


Fig. 30–Structure of roller mask.

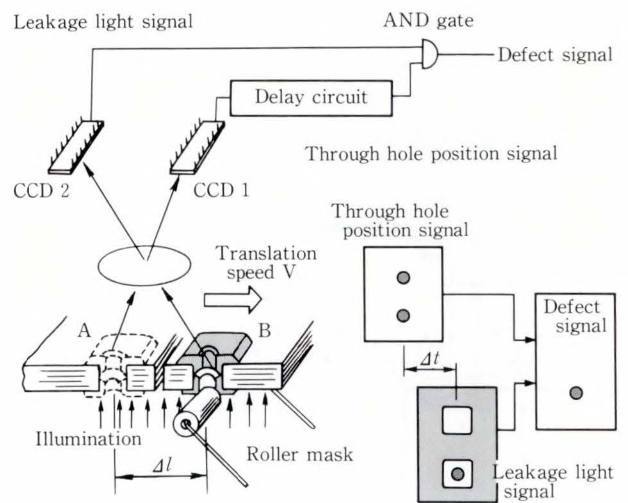


Fig. 31–Principle of self strobing.

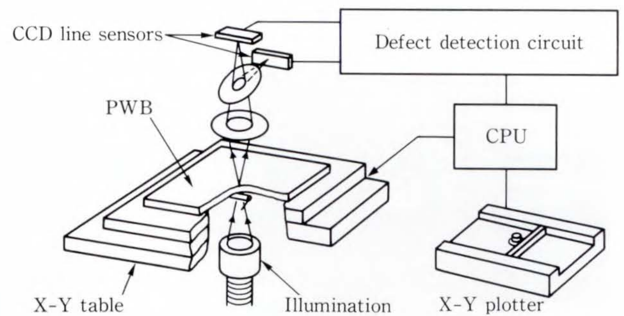


Fig. 32–System configuration.

The first two requirements were solved by using a roller mask and the third requirement by using the self strobing technique.

6.2.3 Complete masking

The mask must be able to block the illuminating light below the 10^{-7} level. A roller mask was developed for the masking. Figure 30 shows the configuration of the roller mask. The roller is a cylinder with a diameter of 2-3 mm. The surface is covered with a black light absorbing material. It contacts the PWB firmly with spring action, and rolls when the PWB moves. The roller mask blocks the illuminating light completely and is capable of following the warpage of PWB even as large as 1.5 mm.

6.2.4 Self strobing

Leakage light detection sensors cannot distinguish between defective leakage light and PWB transmitted light. Therefore the self strobing method was developed to distinguish these two types of light. The principle of the method is shown in Fig. 31. CCD sensor 1 detects the light transmitted directly through the hole. The signal is delayed for Δt by the delay circuit. The delay time is set to the time required for the PWB to travel from point A to point B. When the PWB reaches point B, CCD sensor 2 detects leakage light from the defective through hole. The delayed hole position signal and the leakage light signal are input to an AND gate. The output of the AND gate indicates the presence of a defect. The method can inspect only through hole positions automatically, and has the advantage that any PWB can be inspected without prior knowledge of the through hole positions.

6.3 System configuration

The system configuration is shown in Fig. 32. The line sensor is made up of 1 024 elements. The PWB travels 10 mm/s on an X-Y table. A 300 by 300 mm area can be inspected

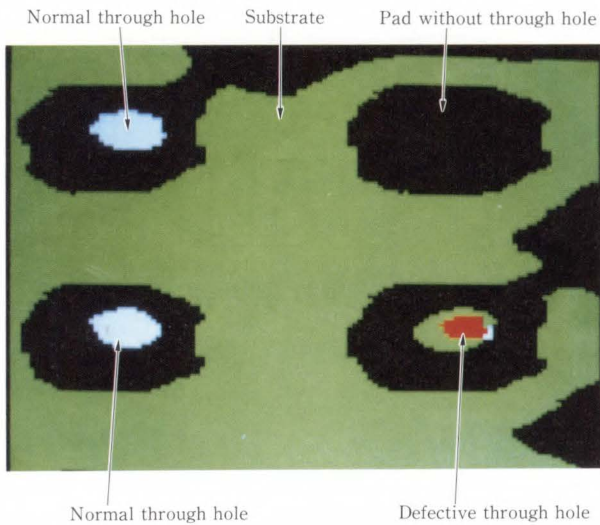


Fig. 33—Output of defective through hole.

Fig. 34—Inspection system.

in 3 min. The results are shown in Fig. 33. The green areas show leakage light signal and the white areas show through hole to location signal. The two white spots at the left show normal through holes, and the black square at the top right shows the pad without a through hole. The red spot under the right shows a defective through hole where the through hole position signal and leakage signal coincide.

The outside view of the inspection system is shown in Fig. 34, and the main specifications of the system are listed in Table 5.

The system has been installed at a Fujitsu plant.

Table 5. Main specification of the through hole inspection system

Items	Specification
PWB	Multi-layer PWB
Detectable defects	Void, Crack, Pinhole
Area	300 mm square
Speed	20 s/100 mm square
Size	1.4 x 1.6 x 1.3 m ($l \times b \times h$)

7. Mounted components inspection

Boards with the components mounted on PWB are conventionally inspected by using electrical function testers and human inspectors. The electrical testers can detect defects such as incorrect orientation and missing ICs. However, a tester cannot detect missing components, incorrect orientation, and inadequate insertion of capacitors and resistors. The defects do not affect functionality, but they decrease the reliability of the PCBs. Incorrect orientation of capacitors can significantly effect the system, that is, the capacitor may break down after some usage. The break causes a short circuit and can cause the whole system to malfunction. The orientation of capacitors must thus be inspected accurately. This is accomplished by visual inspection, and has been indispensable to assure the reliability of PCBs. Visual inspection by humans requires much time to achieve the necessary reliability. The method cannot adapt to the rapid expansion of production quantity. An automated inspection method has been developed to inspect capacitors.

7.1 Requirements

There are two requirements for implementing this system.

The first requirement is to test the components with high speed. A new component must be inserted by automatic inserting machine every 0.5 s. Inspection must be done within this time frame.

The second requirement is high speed image processing with a microprocessor to make the system cost efficient. Data of a 3D-image is much larger than for a 2D-image. For example,

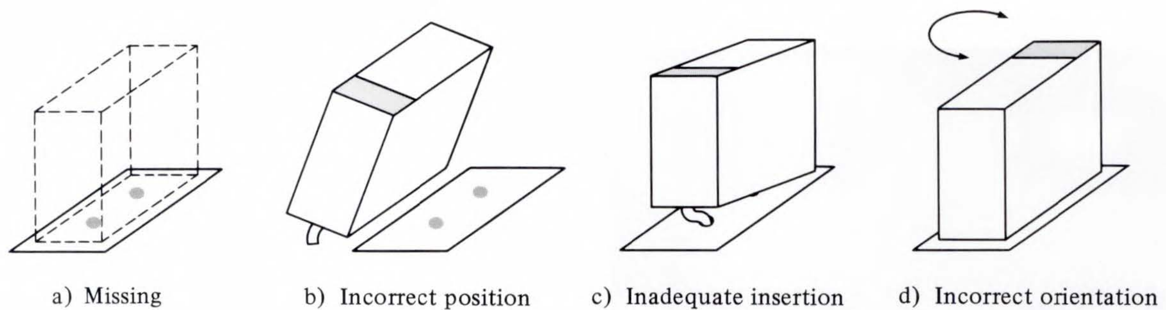


Fig. 35–Example of defects.

a 3D-image with a 256 by 256 pixels resolution, and 100 pixels in the vertical direction requires 4 Mbits, while a 2D-image requires only 64 kbits. It is assumed, however, that a microprocessor can handle image data less than 0.5 kbits within the required 0.5 s.

High speed posture detection algorithm using a microprocessor is required.

7.2 Sensing method

7.2.1 Requirements for sensing method

There are three requirements.

1) Three dimensional posture measurement of black components.
2) Verification mark printed on the components.
3) Inspection time of less than 0.5 s per component.

Figure 35 shows the types of several defects that must be detected.

Incorrect orientation and position are detected by making use of the marks painted on components that denote polarity and specifications. The marks are usually painted with a bright color such as white, and can easily be detected with the gray scale images sensed with a TV camera.

Insufficient insertion occurs when the lead pin is inserted in the through hole as shown in Fig. 35-c). However, gray level images taken by TV camera cannot detect the defect. This type of defect can only be detected by measuring the height of the components.

Incorrect position can be detected when the mark position is seen exceeding the mid position between adjacent insertion holes. However, components are sometimes inserted at an angle tilting as great as 20 degrees. In this case, the mark exceeds the limit position. Therefore, detecting the only mark position does not determine whether the components are inserted correctly. When components are inserted in adjacent positions and overlap the correct position, they may seem correct when viewed from above. This means that the components posture must be measured to detect wrong position.

Two sensing methods, reading marks painted and detecting 3D posture of the components are required.

Conventional 3D posture measurement use such techniques as stereo vision and grid pattern projection.

Stereo vision is similar to human 3D perception. Two TV cameras are set at different positions and are aimed at the same object. There are difficulties in detecting black capacitors because little light is reflected. PCBs use many black components, so this method is not very effective. The processing is also quite slow.

The pattern projection method uses many sheet-type lights and projects them on the object. The image obtained shows the structure of the object. It takes a long time to reconstruct the posture of the object, because there are several lines in the image.

7.2.2 Principle of 3D posture detection

A new sensing method called the "Two step structured light method" has been developed. This method can measure 3D posture using the

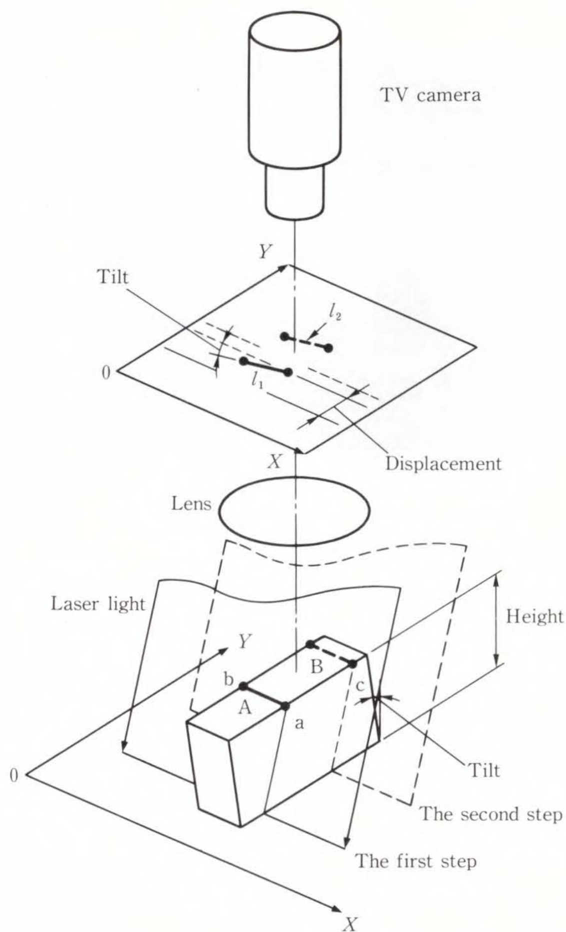


Fig. 36–Principle of Two-step laser light structured method.

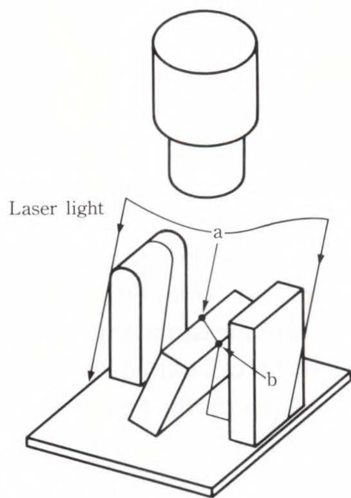
least amount of information without being affected by colors. Figure 36 shows the configuration of the sensing system. Two laser lines are projected on the components with a sequentially tilting angle. The laser lines are detected with a TV camera. The intensity of the laser light is high enough so that even black components can be detected. This structured light method reduces the number of pixels from 4 Mbits to 2 kbits.

The advantage of this method is acquiring a 3D posture of the component with the least amount of information being required.

The principle of the method is described as follows.

First, a laser beam strikes position A on the components, and the edge position of line l_1 can be detected. After calculating the displacement and tilting angle of line l_1, the 3D coordinate of edge points "a" and "b" can be calculated.

Next, a second beam is projected to position B which is located a known distance from position A. The second line determines the 3D coordinate of positions "c" and "d". After detecting the 3D coordinate of four positions, "a", "b", "c", and "d" of the top surface, the orientation and position of the plane can be



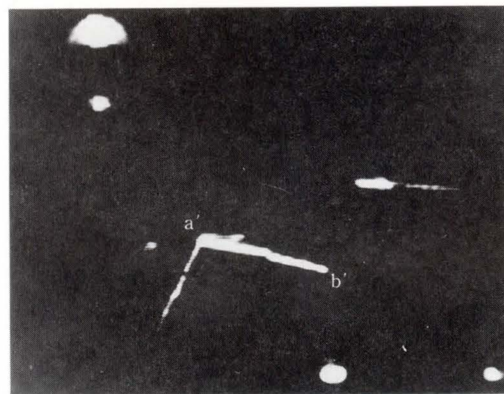
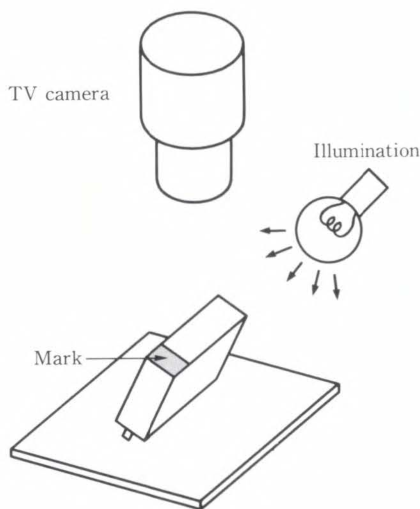


Fig. 37–Laser line image.



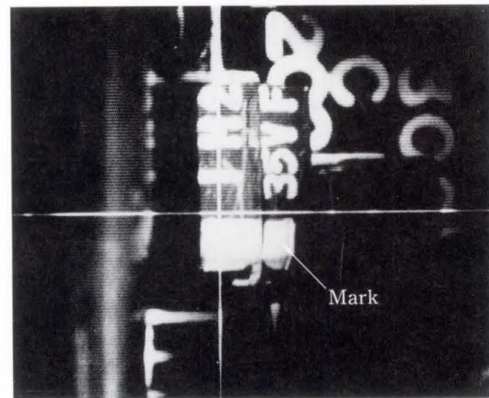


Fig. 38–Observed image with scattered light.

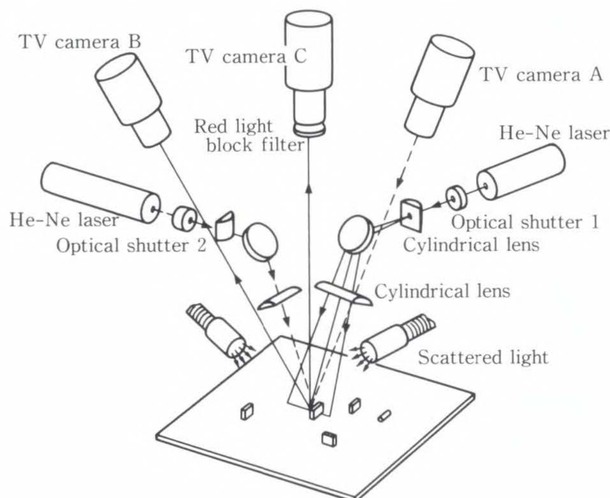


Fig. 39–Configuration of optical system.

calculated. Figure 37 shows the detected structured light on the surface points between a and b, and corresponded to the structured line image of point a and b shown with a′ and b′.

The structured light projection method enables detection of the 3D posture of components without being affected by color. However, it cannot detect marks painted on the components.

7.2.3 Scattered light illumination

Scattered light illumination is used to detect these marks. The components are sometimes tilted at angles up to 20 degrees, and the flat illumination is needed to illuminate the surface without causing shade. Scattered light is used to illuminate the surface. The color of the scattered light is blue because it can easily be discriminated from the red He-Ne laser light. The detected surface is shown in Fig. 38.

Figure 39 shows the configuration of the detection system. The optics for two laser beams are arranged to illuminate the components so that the lines are perpendicular to each other. One of the beams is selected depending on the orientation of the components. Two shutters are used to select these beams. Each line is detected with corresponding TV camera A or B. The mark image is sensed by a TV camera C mounted above the components.

7.3 Inspection algorithms

The algorithms detect component orientation, mark position, and determine whether the components are correctly inserted.

There is an algorithm for posture measurement and one for orientation measurement.

7.3.1 Posture measurement

Posture can be calculated from the two step structured light method. This method has the following advantages.

1) Processing time of less than 0.5 s/component
2) Segmentation of the line image on the components from the surrounding components images

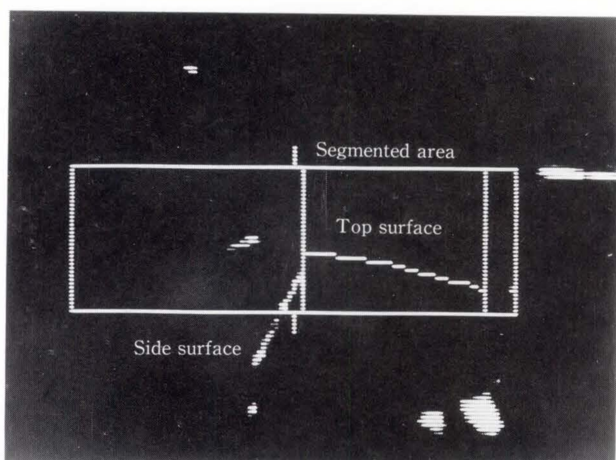


Fig. 40—Segmented line.

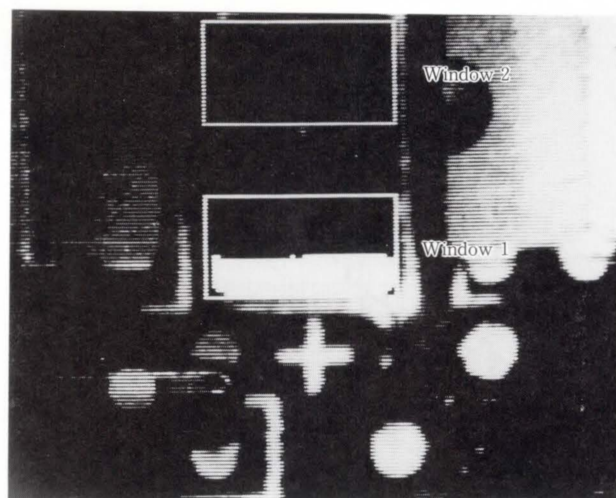


Fig. 41—Double window.

3) Wide dynamic range

Reflected light intensity can vary as much as ten times depending on the optical characteristics of the surface such as color and roughness. The inspection algorithms should not be affected by this.

To meet the first requirement, a window is set at the line image of the laser light to limit the detection area. The window position can be calculated from the size of the components being inspected that are known prior to the inspection. The window reduces the image data from four million pixels down to two thousand pixels to shorten the processing time to 0.5 s.

For the second requirement, the continuity of the line is calculated to segment the line image. The continuity is detected by calculating the angle of tilt and the length of the same angle region. The length of each segment is compared with the tolerance. Only the one segment nearest the correct one is selected as the image of the component that is inspected. Figure 40 shows the segmented results of the line image.

The third requirement is satisfied by detecting the brightest points of the line pattern. The line intensity is checked perpendicular to its direction and the center of the line is selected. The peak detecting method widens the dynamic range to ten times of that without it. This process works well even when the line intensity exceeds the limit of the sensor's dynamic range.

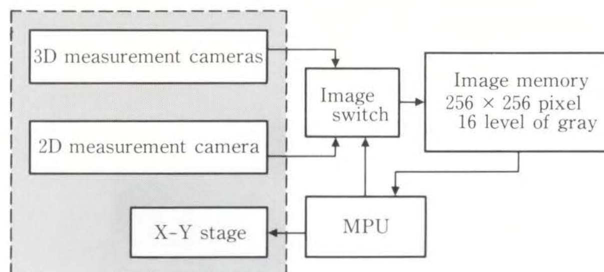


Fig. 42—System configuration.

7.3.2 Polarity detection

This algorithm detects whether the marks are seen at the correct positions. This method checks the mark orientations using two variable windows. One window is set where the correct position mark may appear and the other window is set where the wrong mark will appear. The window positions are displaced to the angle of tilt as measured by the structured light method. In each window, areas brighter than the threshold are detected and compared between the two windows. When the bright area in the correct window is wider than the wrong one, then the orientation is judged to be correct because the mark area is usually wider than that of noise patterns. Figure 41 shows an example of a detected image in the windows.

7.4 System configuration

Figure 42 shows the system configuration.

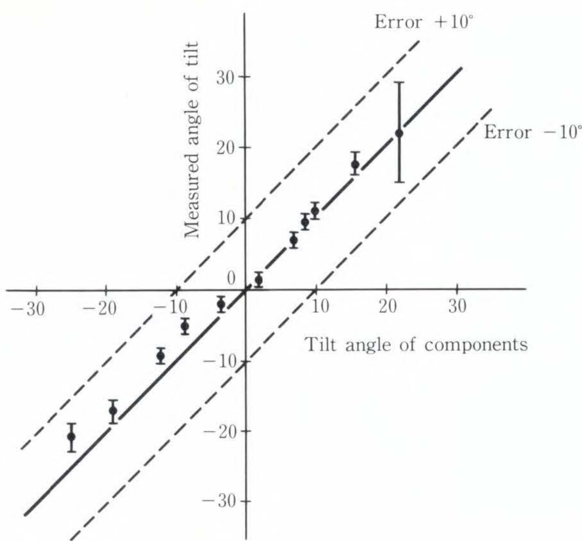


Fig. 43—Experimental results.

Fig. 44—Inspection system.

The images that are taken from the three TV cameras are stored in frame memories. The frame size is 256 by 256 pixels with 16 gray levels. A frame is taken in 33 ms. An 8 088 microprocessor is used for calculating the component orientation, and the angle of tilt. An Intel 8087 numeric coprocessor is used for triangular calculation.

7.5 Inspection characteristics

Figure 43 shows the accuracy of calculating the angle of tilt. These results were obtained after 20 measurements. They show the mean and standard deviation, and the dotted lines are at the 3σ level. Samples are tilted from −25 degrees to 25 degrees. The required accuracy is ±10 degrees. Figure 43 shows that accuracy requirement is satisfied. There are some errors because the shape of the components are not rectangular cubics.

A practical inspection system has been constructed as shown in Fig. 44. The missing rate is zero percent, and the detection rate is 99.7 percent. The inspection time is 0.5 s per component. The system is now on the production line at a Fujitsu plant.

The main specifications of the system are listed in Table 6.

Table 6. Main specification of the component inspection system

Item	Specifications
Components	Rectangular capacitors
Size of PCB	300 x 400 mm, t = 1.2-2.5 mm
Resolution	Horizontal: 0.1 mm, Vertical: 0.2 mm
Defects	Missing, incorrect position, inadequate insertion, incorrect orientation
Inspection speed	0.5 s/component
Size	1.5 x 1.5 x 1.2 m ($l \times b \times h$)

8. Conclusion

The technologies for automatic PWB and PCB visual inspection have been developed. Based on these technologies, automatic inspection systems had been developed by Fujitsu, and working systems have been installed at the Nagano, Akashi, and Oyama plants since 1979. These systems have been effective in achieving high reliability of PWBs and PCBs which are used in large scale computers, etc.

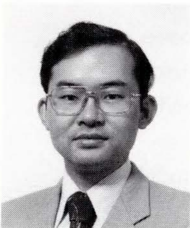
Up to now, further developments in automatic inspection technologies have been continued until the inspection speed increased to ten times, and the resolution as fine as 5 μm. The technologies are being implemented in the latest inner layer pattern inspection system.

In the future, PWBs will have finer patterns on a larger scale, and will demand further high reliability. Inspection technologies will play

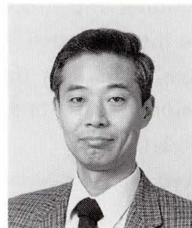
a more important role in PWB and PCB production.

References

1) Nakashima, M., Fujihara, K., and Inagaki, T.: Automatic Mask Pattern Inspection for Printed Circuits Based on Pattern Width Measurement. *Proc. SPIE, Opt. Sys. Eng., III.*, **182**, pp. 38-48 (1979).
2) Ando, M., Mita, K., Kakinoki, Y., and Inagaki, T.: 3-D Profile Detection of Etched Patterns Using a Laser Scanner (3D-Scan Detection) for Automatic Inspection of Printed Circuit Boards. *Proc. SPIE, Opt. Sys. Eng., III.*, **389**, pp. 44-52 (1983).
3) Ando, M., Mita, K., and Inagaki, T.: Automatic Through-Hole Inspection Method for Printed Wiring Boards Using Leaking Light Detection. Proc. IEEE Int. Conf. Robotics and Automation, St. Louis, pp. 481-486 (1985).
4) Inagaki, T., Kakinoki, Y., and Ando, M.: Loaded Printed Circuit Board Inspection by 3-D Measurement of Component Orientation. (in Japanese), *Trans. IEE Jpn.*, **106-C**, 4, pp. 49-56 (1986).



Moritoshi Ando
Sensing Technology Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physical Electronic Eng.
Tokyo Institute of Technology 1971
Specializing in Sensing Technology and Laser Application Systems



Takefumi Inagaki
Electromechanics Division
FUJITSU LABORATORIES, ATSUGI
Bachelor of Electronics Eng.
Nagoya University 1963
Dr. of Electrical Eng.
Nagoya University 1980
Specializing in Electro-Optical Systems, Holography and Hologram Application, Electron Beam Technology, and Inspection and Sensing Technology

Masato Nakashima
Sensing Technology Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Chemistry
Tohoku University 1969
Specializing in Sensing Technology, Image Processing, and Inspection Systems

UDC 62-11:681.32

Development of Advanced Solid CAD System ICAD/SOLID

• Yutaka Tanaka • Takahiro Ueda • Koichi Akatsuka

(Manuscript received September 1, 1987)

This paper describes the technical problems encountered in making a solid CAD system fit for practical use, and the solutions to these problems. The paper is based on the results of the development of a three-dimensional solid CAD system, ICAD/SOLID.

The following three techniques and functions were developed to make a solid CAD system practical: A man-machine interface that allows a designer to create and operate solid models using color shading images, functions that allow two-dimensional drawing data to be used directly for three-dimensional shape operations, and functions to develop and add specific functions using a solid model.

These functions do not require a knowledge of the internal data structure of a solid model or geometrical calculations.

1. Introduction

Solid modeling techniques first appeared more than ten years ago. These techniques have been widely known to be both effective and useful for general-purpose applications. But solid CAD systems, which are CAD systems that use the solid modeling technique, have not become popular[1),2)]. Solid CAD systems are presently regarded as systems to be used only for a specific business by a specific user. The aim of this paper is to illustrate the technical problems that have prevented solid CAD systems from becoming popular, and to discuss the solutions to these problems.

Research on solid CAD systems usually means research on solid modeling techniques. Examples are the interference detection between solid models, and control of the shape of a sculptured surface. These are very important areas of research, and are necessary for developing a powerful solid CAD system. But a powerful solid modeling technique does not always result in a practical solid CAD system. There are other important areas of research that tend to be overlooked, for example, research on the man-machine interface for interactive shape operations. It is important for a designer to create and operate on solid models easily and quickly. A powerful solid modeling technique and an easily used man-machine interface allow a practicable solid CAD system to be developed. This paper discusses the second problem, the design of a natural and easy-to-use man-machine interface for a practical solid CAD system.

Our research started with a trial and evaluation of solid CAD systems, which clearly exposed the problems, and are discussed in chapter 2. Three new techniques were developed from the analysis of the problems, which are described in chapter 3. The techniques were applied to the three-dimensional solid CAD system, ICAD/SOLID, whose features are described in chapter 4. The application of these techniques to the design of a telephone is described in chapter 5.

The conclusions are given in chapter 6.

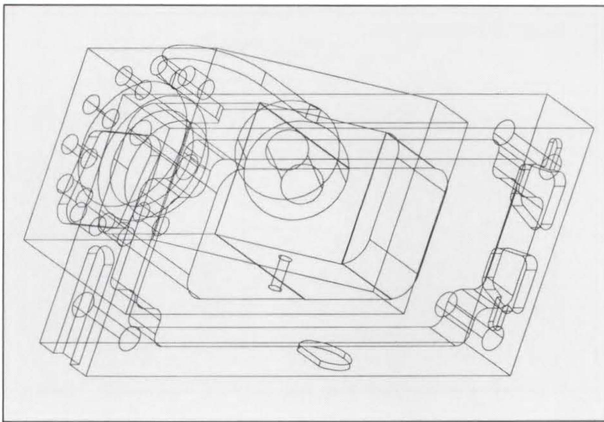


Fig. 1—A wireframe picture.

2. Technical problems with the solid CAD system

Tests and evaluations of solid CAD systems have exposed the three main problems that make solid CAD systems impractical.

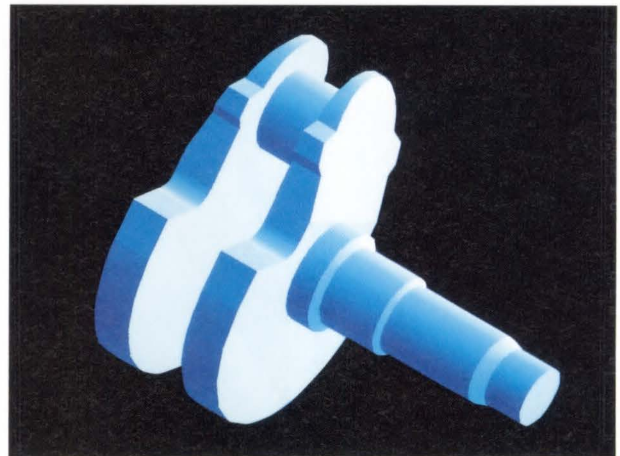
2.1 A method of interaction

To design products effectively, it is important for a designer to be able to create and operate solid models easily and quickly. Easy and quick operation of an interactive solid CAD system depends on the form of the display and the method of interaction. There are three main considerations for the man-machine interface: the display method, the element indication method and the delay in displaying the image.

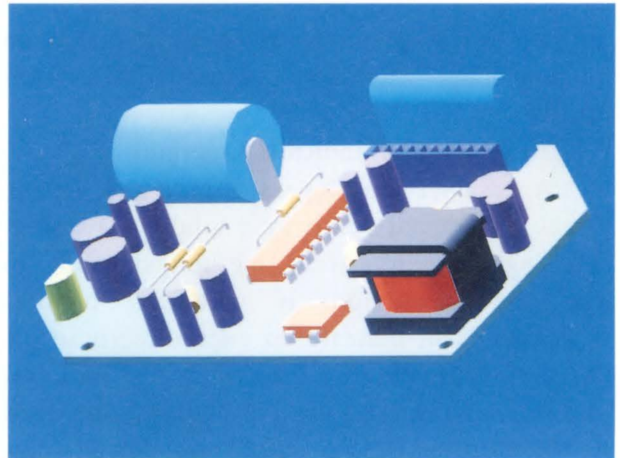
2.1.1 Display method

Usually, a man-machine interface based on wireframe pictures is used for an interactive shape operation, because such images require fewer calculations. The advantage of wireframe pictures is therefore little delay in displaying the image. An example of a wireframe picture is shown in Fig. 1. However, as a designer geneates and operates solids in a solid CAD system, and shape operation in a solid CAD system is based on edges, faces and solids, it is difficult for the designer to understand solid shapes from wireframe pictures and operate upon them.

Color shading makes it easy for a designer



a) Crankshaft



b) Printed circuit board

Fig. 2—Color shading images.

to understand solid shapes. Examples of color shading images are shown in Fig. 2. A man-machine interface based on images with color shading is suitable for solid shape operation. But as many calculations are needed to generate color shading images from solid models, the man-machine interfaces based on color shading images are seldom used for performing solid shape operations. The technique is normally used only for verifying finished shapes.

The above method, an interactive shape operation in wireframe pictures and a verification of the finished shapes using color shading

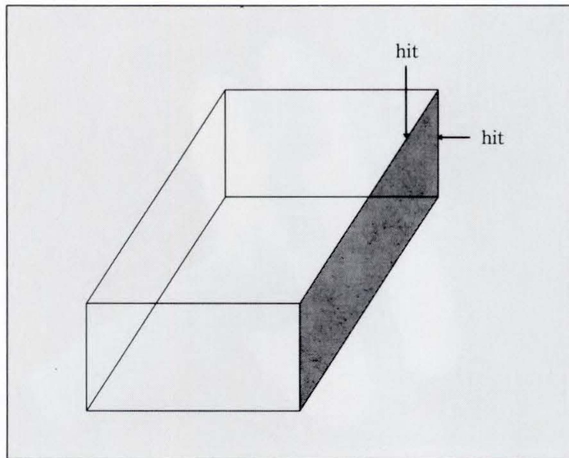


Fig. 3—A face indication by hits of two edges.

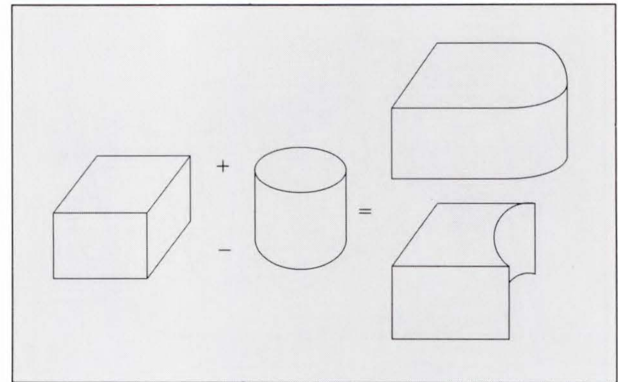


Fig. 4—A boolean operation.

images, is the usual one. However, it makes the man-machine interface of a solid CAD system difficult to use.

2.1.2 Element indication method

A designer must operate upon and indicate edges, faces, and solids in a solid CAD system. The element indication method depends on the display method. In wireframe pictures, a designer can indicate an edge directly in one hit action. However, a face cannot be indicated directly, as the faces are not actually displayed. It therefore takes two actions to indicate a face, for example, two of the edges on the face must be hit (see Fig. 3). In color-shaded images, a face can be indicated in one action, while an edge needs two actions.

For simple and quick operation, both an edge and a face should be indicated in one action. But neither wireframe pictures nor color-shaded images can satisfy this requirement.

2.1.3 Delay in displaying the image

Little delay in displaying the image is important in interactive solid CAD systems. When a shape operation has been performed, a designer must confirm the result of the operation in real time. The designer evaluates the shape operation by looking at the updated image, and decides upon subsequent action. In order not to interrupt this stream of thought and action, the updated images should be displayed in real time.

As described in subsection 2.1.1, wireframe pictures can nearly be displayed in real time. But it takes a lot of practice for the designer to understand solid shapes from wireframe pictures. And although it does not take long to understand solid shapes from color-shaded images, such images require many calculations and cannot be displayed in real time.

2.2 Shape operation method

The natural and efficient solid CAD system depends on the method used for the shape operation. The ideal operating method matches a designer's natural way of thinking while an unnatural shape operation method will make an interactive solid CAD system impractical.

The starting point for designs is still the drawing. Many designers still prepare their plans on drawing boards. They do not design three-dimensional products straight away. They put their ideas into drawings, and then they reconstruct solid images in their minds from these drawings.

The shape operation functions of a solid CAD system usually consist of a direct generation of solid primitives and a boolean operation between two primitives. An example of a boolean operation is shown in Fig. 4. Three-dimensional products are treated directly as solid primitives in a solid CAD system. But, this standard method does not always fit the designer's natural way of thinking. This is

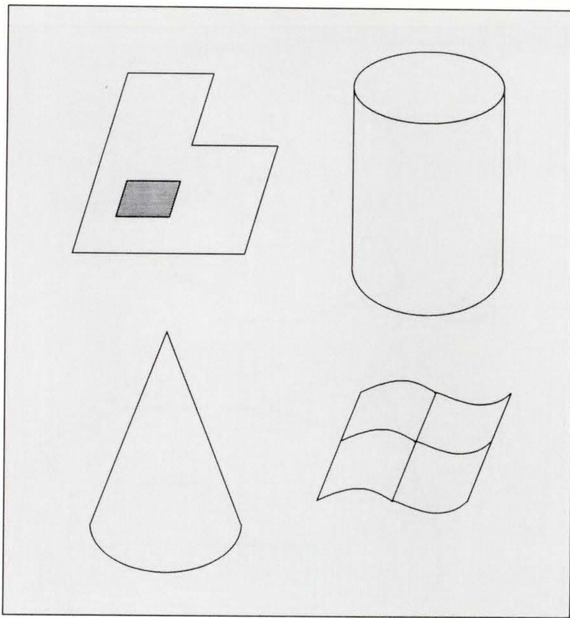


Fig. 5—Examples of input faces.

the second problem that often makes solid CAD systems impractical.

2.3 The method of adding specific functions

For a solid CAD system to be effective, the user must be able to easily add user-defined functions to the system. It is desirable that users develop these specific functions and add them to their systems themselves, as they understand their own business best.

However, as a lot of technical knowledge is needed to develop the programs, it is difficult for a user to develop specific function programs using a solid model. Knowledge of winged-edge data structures and boundary representations is needed, for example. This is the third problem that makes present solid CAD systems impractical.

3. Techniques for solving the problems

The following techniques were developed for solving the problems described in chapter 2.

3.1 Techniques for interaction

As described in section 2.1, man-machine interfaces based on color-shaded images are desirable for solid shape operation. However for easy and quick operation, edges and faces should be displayed in the same picture. It is highly desirable that the solid-model display used for the shape operation overlaps the wire-frame pictures that represent edges, with color-shaded images representing the faces. A three dimensional graphic display with a function for eliminating hidden surfaces is necessary for displaying overlapping pictures in real time. The functions described below have been developed. It is assumed that a three-dimensional graphic display with a hidden surface elimination function is used. Solid model data are converted into polygon data that the display can process. The polygon data are transmitted from the host computer to the display, which eliminates hidden surfaces, and displays the solid model.

3.1.1 Polygon division method

A polygon division method can transform a solid model given in boundary representations to three-dimensional polygon data form. It divides curved faces that include holes into planar polygons without holes, and curved edges on the boundary of the face into a chain of straight lines.

1) Algorithm

i) Input

Input data consists of the face, edges and vertices of the boundary representations. This method can divide a planar surface, a quadratic surface, or a sculptured surface, and faces can have holes. Examples of input faces are shown in Fig. 5.

ii) Division of edges

Curved edges on the boundary of a face can be divided into straight lines. For example, a circle edge can be divided into a number of short lines, as shown in Fig. 6. The width of each division can be controlled according to which the designer thinks is more important: a rapid display, or a precise display. This chain of straight lines is used for wireframe pictures and it becomes the input of the step "division of a face".

iii) Transformation into (u, v) space

Short straight lines in three-dimensional

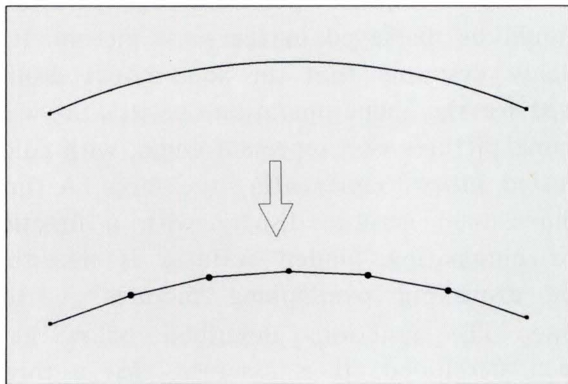


Fig. 6—Division of an edge.

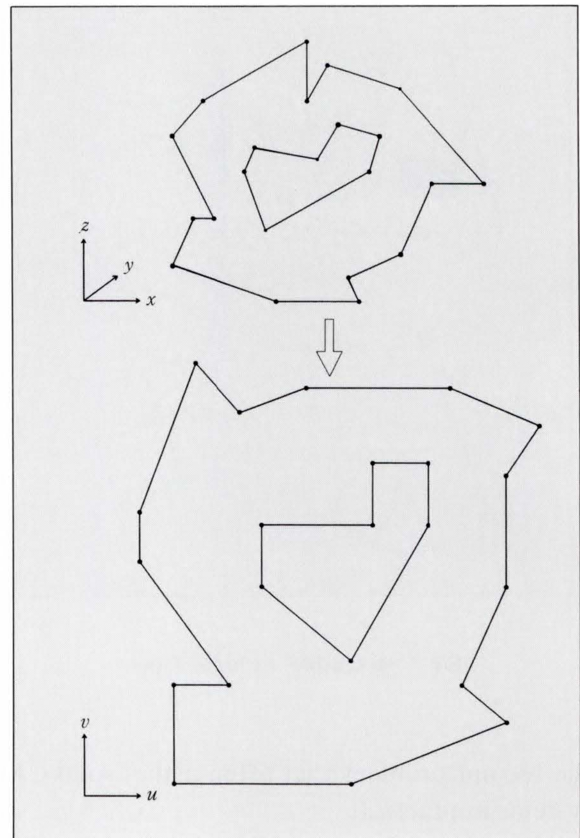


Fig. 7—Transformation from (x, y, z) to (u, v).

real space (x, y, z) can be converted into two-dimensional parametric space (u, v). This step transforms a curved face in (x, y, z) space into a polygon in (u, v) space, as shown in Fig. 7 and enables all the faces to be processed in the same way. In the next step, the polygon is divided into a number of triangles and rectangles that the display can process.

iv) Division of a face

It is important to fit wireframe pictures to the boundaries of the faces in overlapping pictures. If wireframe pictures do not fit exactly onto the boundary of the face, they may be covered by the face, and be put out of sight. Likewise, if the boundary of one face does not exactly fit the boundary of a neighboring face, there will be gaps between the two faces. Division of a face is usually done by a mesh method, which divides faces into grid squares by respective meshes of the faces, as shown in Fig. 8, repeats division until polygons in all meshes become triangles or rectangles. This means that the boundary of a face does not always fit the boundary of a neighboring face. To solve these problems, this algorithm generates color shading data for the neighboring faces from only the wireframe data of the common edge. Triangles and rectangles are generated in this step.

a) Generation of a triangle from an acute vertex

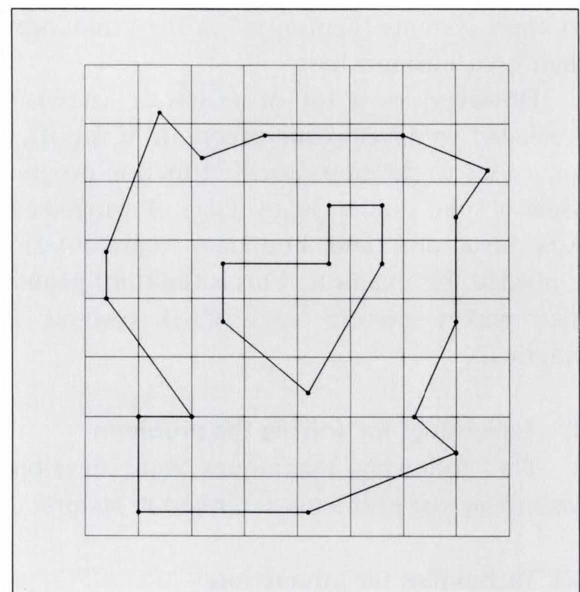


Fig. 8—Division by the mesh method.

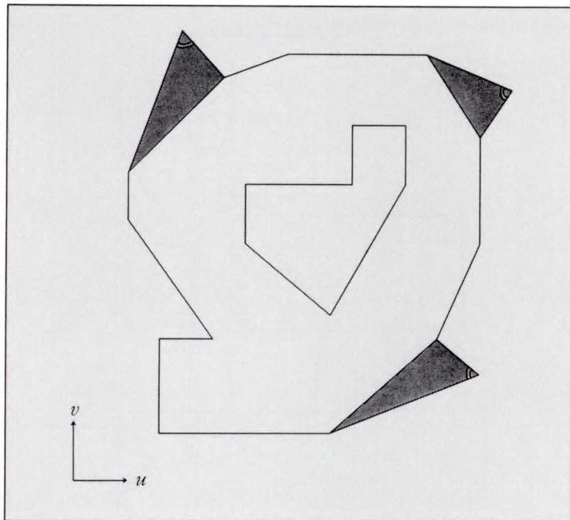


Fig. 9 – Generation of triangles.

Check the interior angle of every vertex of the input polygon. If the interior angle of the vertex is acute, generate a triangle polygon from the vertex and neighboring two vertices, as shown in Fig. 9. Delete the vertex from the chain and if there is no acute-angled vertex, go to the next step.

b) Generation of a right-angled triangle from a hypotenuse

Check the angle of inclination of every edge of the input polygon. Generate a right-angled triangle polygon from the edge. If the edge is not parallel to the u-axis and v-axis, a new vertiex is generated in the input polygon, as shown in Fig. 10. Add the vertex to the chain. The edge is the hypotenuse of the right-angled triangle, and the generated vertex is at the right-angled corner. If all edges are parallel to the u-axis or the v-axis, go to the next step.

c) Generation of a rectangle

Arrange the vertices in increasing order of v-coordinates. Arrange the vertices that have same v-coordinates in increasing order of u-coordinates.

Search this chain of vertices, cut at the same v-coordinates, generate rectangle polygons and delete the vertices from the chain, as shown in Fig. 11, if there is no vertex in the chain, go to the next step.

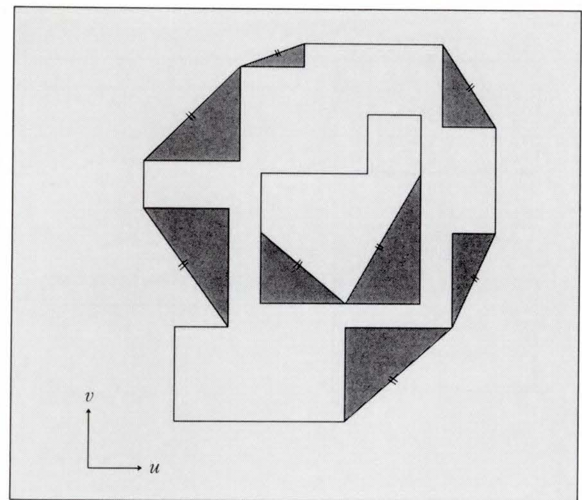


Fig. 10 – Generation of right-angled triangles.

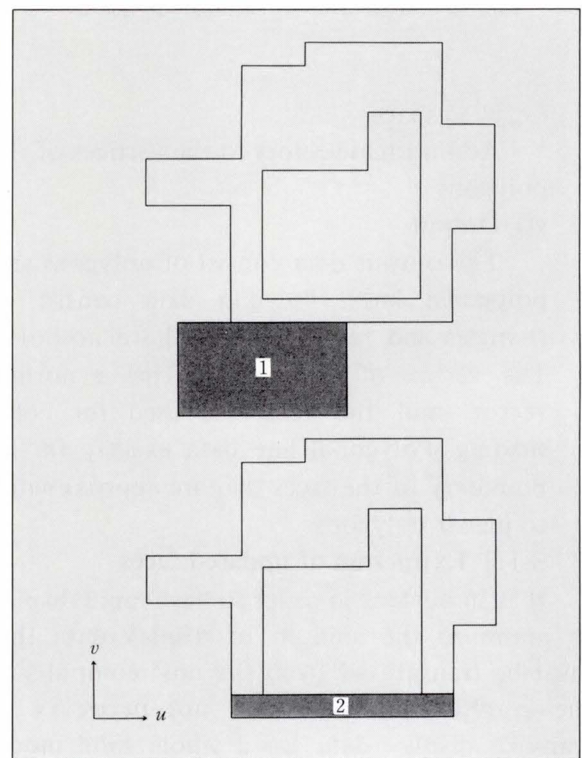


Fig. 11 – Generation of rectangles.

d) Subdivision of a rectangular polygon

Subdivide a rectangular polygon at even intervals of u-coordinates into smaller rectangular polygons, as shown in Fig. 12.

v) Transformation into (x, y, z) space

Transform polygon in (u, v) space into

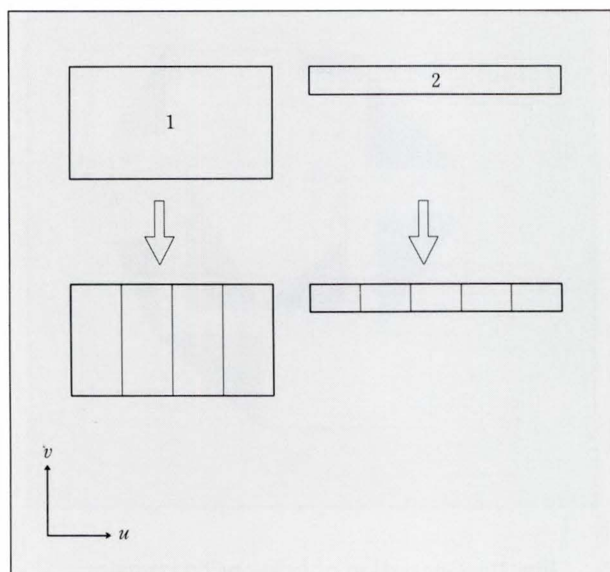


Fig. 12–Subdivision of rectangle polygons.

(x, y, z) space.

Add normal vectors to the vertices of the polygons.

vi) Output

The output data consist of polygons and polygonal lines. Polygon data consist of triangles and rectangles that have no holes. The vertex of the polygon has a normal vector, and the vector is used for color shading. Polygonal line data exactly fit the boundary of the faces that are approximated to planar polygons.

3.1.2 Extraction of updated faces

It is important in order to have rapid display to minimize the amount of display data that must be transmitted from the host computer to the graphic terminal. It is not necessary to transmit display data for a whole solid model when a shape operation is performed. Only the display data for the faces that are transformed A or generated B should be transmitted, as shown in Fig. 13.

A function for extracting the updated faces has been developed, which is built into the basic subroutines constituting the shape operation. During a shape operation, these subroutines write the records of generated, deleted and transformed elements into a table. The faces that must be transmitted are selected according to this table.

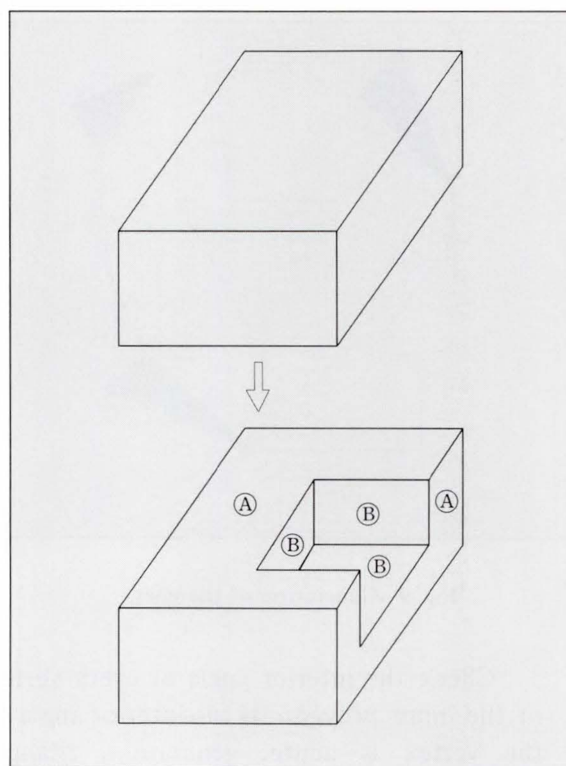
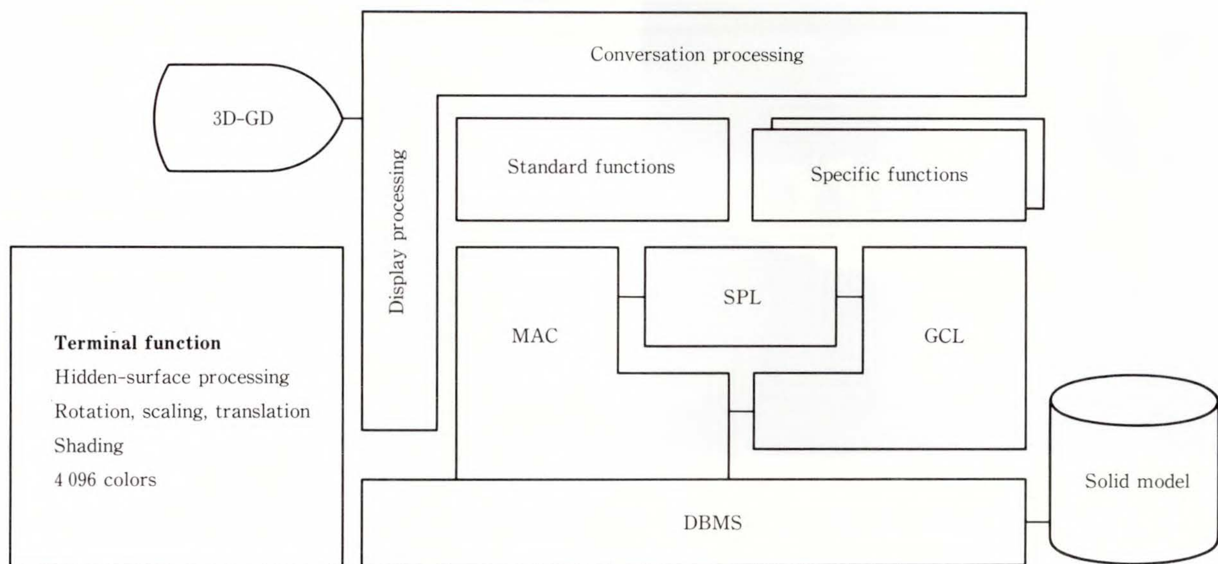


Fig. 13–Extraction of transformed A or generated B faces.

3.2 Technique for natural operation

According to the example of a boolean operation between simple primitives shown in Fig. 4, it seems that a boolean operation is easy to use. But some designers who have tried the solid CAD systems have pointed out that it is very difficult to generate cutter solid models for a boolean operation. A cutter solid model is necessary in a boolean operation to indicate the shape to be deleted. However it is unnatural and difficult for a designer to generate a complex cutter solid model directly without using drawings, although it is effective to generate a cutter model using drawing data. And it is more effective to use drawing data directly for a shape operation.

Addition of the data of the third coordinate is necessary in order to use two-dimensional drawing data for a three-dimensional shape operation.



MAC: Model ACcess Routine SPL: Shape Processing Library GCL: Geometrical Calculation Library

Fig. 14—System configuration of ICAD/SOLID.

Although it is difficult to add the data for the third dimension automatically, it is possible to add this data using the depth information of the viewport in which the drawing is displayed. Moreover, the drawing data is stored in the same format record as the three-dimensional edge for a quick transformation into a solid model.

3.3 Techniques for adding specific functions

It is difficult for a user to develop specific functions using a solid model. Therefore, the subroutines that are open to a user should not require detailed knowledge of the internal data structure of a solid model and geometrical calculations.

The subroutines were developed, based only on data structures that a user can understand through overlapping pictures. The data structure consists of solids, faces, edges, and vertices. The subroutines combine basic subroutines that use the internal data structure, and put the internal data structure out of sight.

4. Features of ICAD/SOLID

The three-dimensional solid CAD system ICAD/SOLID is a new part of Fujitsu's integrated computer aided design and manufacturing system (ICAD). Using the techniques described in chapter 3, the following functions have been realized in ICAD/SOLID. The system configuration of ICAD/SOLID is shown in Fig. 14.

4.1 An advanced man-machine interface

All the operations of ICAD/SOLID are carried out by combining color-shaded images and wireframe pictures. A solid, a face, and an edge can be indicated in a single action, consisting of one hit with the pointing device. If necessary, the display can be converted into color-shaded images or wireframe pictures, as shown in Fig. 15.

The updated image after a shape operation is displayed in a few seconds. Using the functions of a three-dimensional graphic terminal, cross sections and internal views of solid models, as shown in Fig. 16, can be obtained within a second.

ICAD/SOLID is an advanced man-machine interface for three-dimensional shape operations which is compatible with a designer's thought and work.

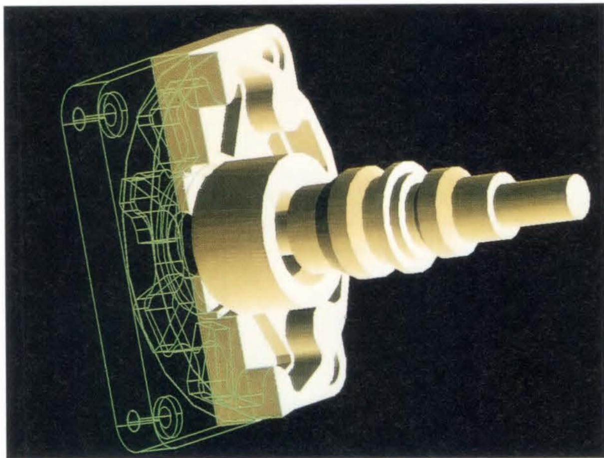
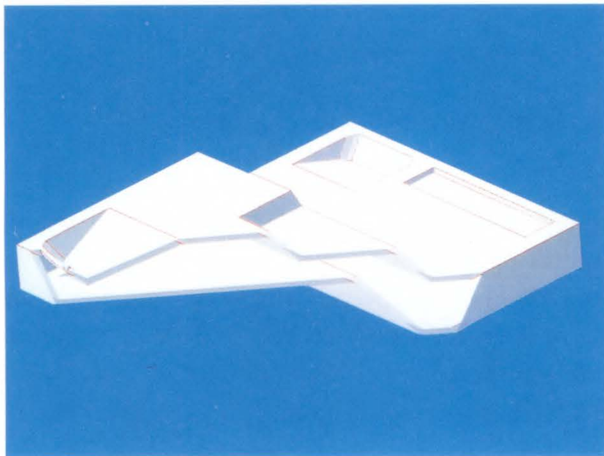
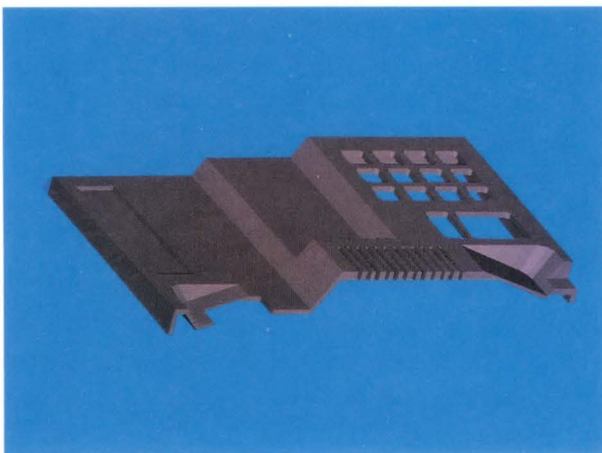


Fig. 15–A change of a display type.



a) Internal view of telephone



b) Cross section of telephone case

Fig. 16–Cross section and internal view.

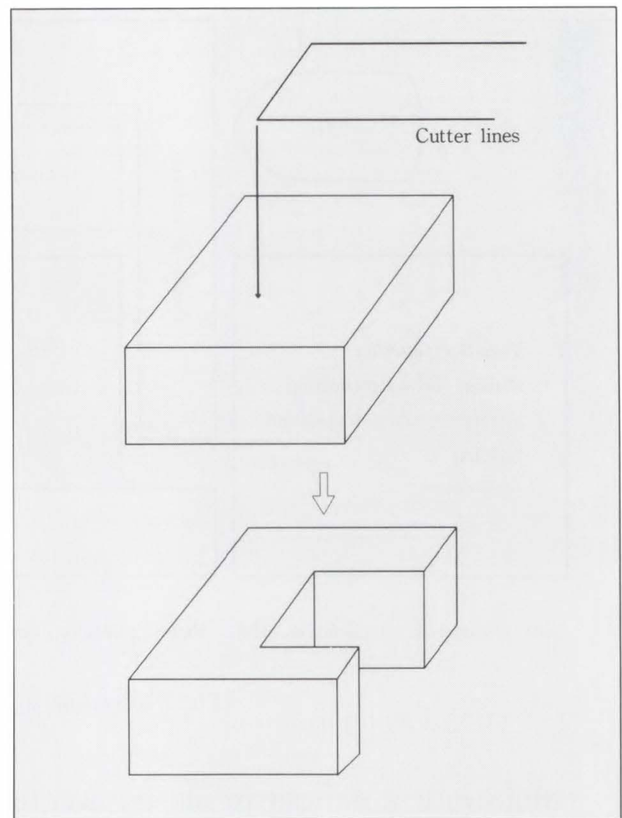


Fig. 17–MILL function.

4.2 Shape operation based on drawings

ICAD/SOLID can receive drawing data from ICAD/SDS[3] (integrated computer aided design and manufacturing system/standard design system). Drawing data is used for shape operations and generating solid models while the MILL function, shown in Fig. 17, is the typical function for a shape operation. This function adds the depth or height information to the drawing data and uses the drawing data as cutter lines which indicate the shape to be deleted. It does not use a cutter solid model, but uses drawings directly.

The method fits a designer's natural thought.

4.3 Structure open for adding user-defined functions

A user can incorporate programs written in FORTRAN77 into ICAD/SOLID. The user's additional programs can call up standard programs from ICAD/SOLID. ICAD/SOLID has the following standard subroutine libraries.

Fig. 18–A solid model of Fujitsu telephone GK.

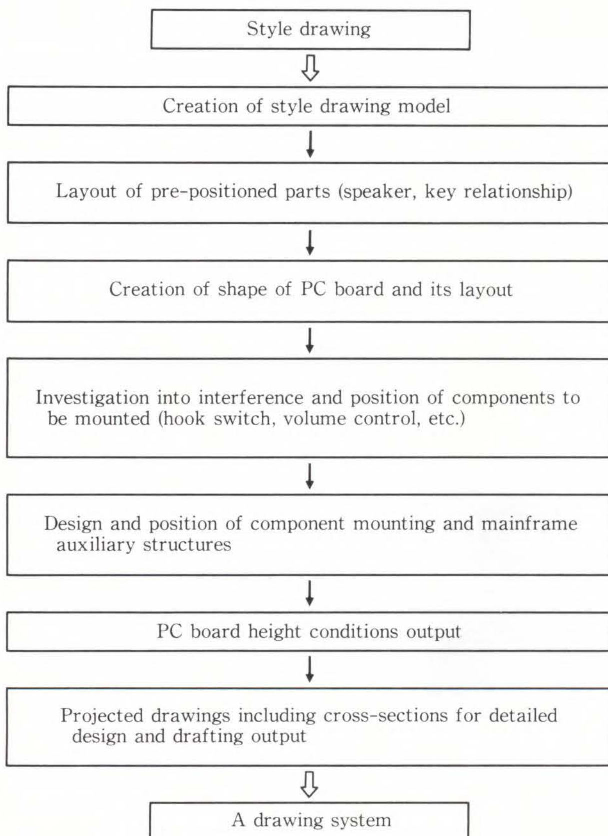


Fig. 19–Standard workflow.

1) Shape processing library
2) Geometrical calculation library
3) Model access library
4) Display and interaction library

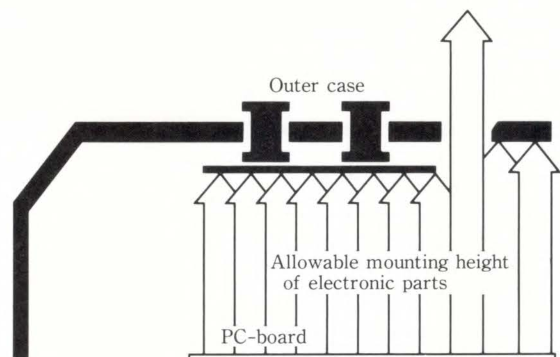


Fig. 20–Concept of packaging space calculation function.

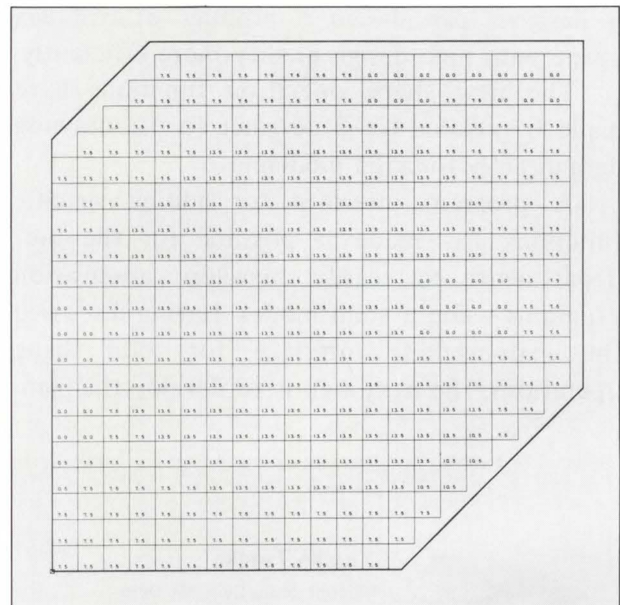


Fig. 21–Output of packaging space calculation function.

5. Application in layout design

ICAD/SOLID has been used for developing the layout and structural design of a telephone[4]. Solid models of the FUJITSU telephone GK are shown in Fig. 18. The standard workflow is shown in Fig. 19.

The user department developed the additional programs for the specific functions with the help of the libraries belonging to ICAD/SOLID, which is the parckaging space calculation function for printed circuit boards. The function divides the surface of a printed circuit board into square grids, and then

calculates and outputs the allowable mounting height of the electronic parts that are mounted on the printed circuit board. The concept of this function is shown in Fig. 20, and an example of the output is shown in Fig. 21.

Packaging space is normally calculated by hand and the information produced is approximate. However using ICAD/SOLID and this additional function, the packaging space can be calculated accurately.

6. Conclusion

With the development of a man-machine interface for three-dimensional shape operation, a designer can design a product of complex shape with high design quality more efficiently.

The new shape operation functions have made it possible for a designer to use drawing data directly for solid modeling.

A program interface for adding specific functions has made it possible for the user department to easily develop application programs using a solid model. Having discussed the man-machine interface for solid shape operations, the next step is to develop the man-machine interface that raises the efficiency of the application of solid models.

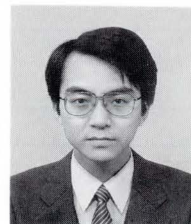
7. Acknowledgement

The authors wish to express their thanks to the members of the Fujitsu solid CAD project and to all the others who assisted in the development of this system.

References

1) Requicha, A.A.G., and Voelcker, H.B.: Solid Modeling: A Historical Summary and Contemporary Assessment. *IEEE Comput. Graph. Appl.*, **CG & A-2**, 2, pp. 9-24 (1982).
2) Requicha, A.A.G.: Representations for Rigid Solids: Theory, Methods, and Systems. *Comput. Surveys*, **12**, 4, pp. 437-464 (1980).
3) Hashimoto, E., and Yoshimura, M.: User-Friendly CAD/CAM System with a Japanese-Language Processing Capability for Computer Integrated Manufacture. *FUJITSU Sci. Tech. J.*, **20**, 4, pp. 585-604 (1984).
4) Hashimoto, E., Tanaka, Y., Ueda, T., and Itoh, K.: "Application of Solid CAD for Mold Design". Information Processing 86. Kugler, H.J., ed., North-Holland, Elsevier, 1986, pp. 433-439.

Yutaka Tanaka
Mechatronics Systems Dept.
FUJITSU LIMITED
Bachelor of Mechanical Eng.
University of Electro Communications
1970
Specializing in CAE/CAD/CAM System



Koichi Akatsuka
Mechatronics Systems Dept.
FUJITSU LIMITED
Bachelor of Science
Tohoku University 1980
Master of Science
Tohoku University 1982
Specializing in Solid CAD Systems

Takahiro Ueda
Mechatronics Systems Dept.
FUJITSU LIMITED
Bachelor of Mathematical Eng.
Kyoto University 1976
Specializing in CAD/CAM Systems

UDC 517.4:681.3.02

A Real-Time Hough Transform Processor

● Keishi Hanahara ● Tsugito Maruyama ● Takashi Uchiyama

(Manuscript received June 12, 1987)

One of the inherent disadvantages of the Hough transform, its long computation time, is preventing this transform from being used in real-time applications. This paper presents an architecture that was developed to address this problem and implemented in an experimental hardware model for real-time straight-line detection.

This paper also provides examples of applying this architecture to the automatic inspection and measurement of objects in factories and laboratories. In these examples, it took less than one second to extract straight-line parameters from captured pictures.

The highly parallel architecture proposed here for the Hough transform does not require any multiplication or trigonometric calculations at run time.

1. Introduction

The Hough transform for recognition can effectively process noisy backgrounds and boundary gaps. It was first applied to the detection of straight lines[1] and has since been generalized to successfully detect parametric curves such as circles, parabolas, and ellipses. Recent literature about the Hough transform includes many works related to applications such as recognizing nonparametric curves and 3-dimensional objects, determining the orientation of rigid bodies, and obtaining motion parameters for moving objects[2], [3].

In spite of the excellent recognition ability of the Hough transform, such disadvantages as long computation times and large memory requirements prevent its being used in real-time applications[4]. In general Hough transforms, computation time and memory requirements increase exponentially with the number of dimensions in the parameter space[5], [6].

Silberberg reports the execution time of a Hough transform program for straight-line detection on a VAX 11/785[7]. Only integer arithmetic was used with cosine and sine values stored in a table. The program he ran took about seven seconds to process a 521 x 512 image without peak detection.

One of the objectives of this paper is to present an architecture developed to address the time and memory load problems. This architecture has been implemented on an experimental hardware model for real-time straight-line detection.

The underlying premise of our study is that large memory requirements are not serious handicaps at current levels of memory integration technology for straight-line detection in two-dimensional arrays, which is the most fundamental and widely applied use of the Hough transform.

As can be seen from the literature, major applications of the Hough transform are still in the field of straight-line detection. Some of

This paper has been revised from the article "High-speed Hough transform processor and its applications to automatic inspection and measurement" as it appeared in the proceedings of the IEEE International Conference on Robotics and Automation, April 7-10, 1986.

these applications include gauge inspection[8], recognition of handprinted Hebrew characters[9], inspection of VLSI resist patterns[10], detection of seismic patterns[11], and vehicle guidance through road edge detection[12)-15].

The time required for our model to make a histogram in a 512 x 512 (ρ, θ) space for 1 024 feature points is 0.26 s. These results can be compared to those of the report[7] after correction by appropriate extrapolation. Execution time is linear for the number of the feature points and partitions in the θ coordinates. The report assumes 3 percent feature points in the 512 x 512 image and five degrees for θ resolution, which implies 12.9 s for our conditions.

Chapter 2 describes the processing flow of the Hough transform, and chapter 3 presents our design objectives for the hardware implementation of Hough transforms. Chapters 4 and 5 describe the optimum parameter space resolution and quantization errors. Chapter 6 shows the system configuration of our experimental hardware model, chapter 7 shows that the complementary Hough transforms are suitable for the developed processor architecture, and chapter 8 proposes a parallel architecture for Hough transforms. Chapter 9 presents examples of applying this architecture to automatic inspection and measurement tasks.

2. Processing flow

This chapter first defines the symbols used to set the coordinate systems of the image and parameter space. It then describes the parameterization of straight lines.

Raster-scanned TV signals are sampled to assemble an image space consisting of $N \times N$ pixels. Assuming an even aspect ratio, this space is a square.

A pair of horizontal and vertical addresses (H, V) is often used to indicate a location in a raster-scanned picture. This pair of addresses is referred to as the raster address. We use the following normal parameterization proposed by Duda and Hart[16] to specify the straight line:

$$\rho = x_i \cos\theta + y_i \sin\theta, \qquad \text{..........(1)}$$

where, x_i and y_i are the relative values from the center of the picture at the i-th feature point in the picture.

Assuming the raster address of this feature point to be (H_i, V_i), the address values are given by the relations

$$x_i = V_i - \frac{N}{2} \text{ and } y_i = H_i - \frac{N}{2},$$

where N is assumed to be an even number. In our system, $N = 256$. ρ is the distance of the line from the origin with raster address $(\frac{N}{2}, \frac{N}{2})$, and θ is the angle between the normal and the x axes.

The sinusoidal curve (1) in the parameter space represents all straight lines passing through the i-th feature point.

After all feature point curves (1) have been drawn in the parameter space (ρ, θ), the curves corresponding to collinear points in the picture should intersect at one point. Assuming the coordinate of this point to be (ρ_p, θ_p), the line that passes through the collinear points can be represented in the image space as follows:

$$\rho_p = x \cos\theta_p + y \sin\theta_p. \qquad \text{......(2)}$$

This reduces the problem of detecting the straight lines in the image space to a problem of searching for the intersections of sinusoidal curves in the parameter space.

The method used to search for these intersections is to quantize the (ρ, θ) space and treat it as a two-dimensional accumulator array. First, the array is cleared. Then, all cells of the array along the curves (1) are incremented by one unit. Once sinusoidal curves are accumulated for all feature points, the intersections appear as peaks.

3. Design approach

3.1 High resolution

Coarser quantization shortens processing time and reduces the demand for accumulator array memory space. Furthermore, the resulting histogram of the locus of the sinusoidal curve

will show legible peaks even for boundaries or edges of objects with poor linearity. The disadvantage of coarser quantization is that there are significant errors in parameters obtained from boundaries or edges with good linearity.

We have adopted a design for measuring objects with good linearity using high resolution by selecting the optimal step size.

Two-dimensional arrays are sufficient for detecting straight lines, and memory utilization requirements are not a serious problem with current memory integration technologies. The same histogram that results from a transform with a coarse step size can be obtained by merging multiple cells in postprocessing after histogram creation. This merge processing is a simple local operation. Our concern was that it is impossible to obtain fine histogram structures from histograms with coarse step sizes.

Over-sampling of the (ρ, θ) space, however, must be avoided. Research is now being done to find the optimum relationship between $\Delta\theta$ and $\Delta\rho$[17]. Chapter 4 introduces similar work resulting from a different point of view.

3.2 Parallel operations

We have separated sine curve generation and histogram generation so that they can be done in parallel.

Calculating the increment for a single feature point cannot begin until ρ_n has been calculated for that point. Normally, ρ calculation and increment processing cannot be done in parallel. They can be executed in parallel, however, if the feature points are input consecutively. ρ_n for the $i+1$-th point can then be calculated while the i-th point increment processing is being done because these two operations do not share hardware resources.

3.3 High-speed data transfer

In the edge extraction section, convolution, thinning, local maxima detection in nearby pixels and zero-crossing detection can be performed at video rates[18]. Because the Hough transform section cannot complete processing for each address within one sampling period, the addresses are temporarily buffered by a FIFO (first-in first-out) buffer. The histograms are transferred back to the space differential processing unit for edge extraction and local maxima are searched for at a video rate.

The sets of (ρ, θ) coordinates detected by hardware are also sent to the FIFO buffer in the order they are found along the raster scan. This means that the FIFO buffer is used twice during the complete process. The type of data passed through the buffer changes. It first carries addresses in the image space and then carries addresses in the parameter space.

4. Parameter space resolution

This chapter describes the optimum relationship between $\Delta\theta$ and $\Delta\rho$.

If the θ pitch is large in the discrete parameter space, ρ values must sometimes be interpolated to draw a sine curve. Interpolation is required when the difference in ρ values at adjacent θ coordinates exceeds $\Delta\rho$. This is likely to occur when the curve crosses the θ axis and x_i and y_i are both large. Hardware interpolation has the disadvantage of requiring either additional circuits or more complex processing, both of which lower processing speeds. It is possible, however, to find a $\Delta\rho$ value that makes interpolation unnecessary.

To achieve this, even when

$$\mathrm{dif}(\rho) \equiv \rho(\theta_{n+1}, x_i, y_i) - \rho(\theta_n, x_i, y_i),$$

is maximized requires that the following equation be satisfied.

$$\mathrm{dif}(\rho) \equiv \frac{\partial\rho}{\partial n} = \frac{\pi\sqrt{x_i^2 + y_i^2}}{N_\theta} \cdot$$

$$\cos\left(\frac{\pi_n}{N_\theta} + \theta_0\right) \leq \Delta\rho,$$

for all x_i, y_i and n, where we approximate the difference in ρ by differentiation and n is considered continuous.

Because the maximum value of $\sqrt{x_i^2 + y_i^2}$ is $\frac{N\Delta x}{\sqrt{2}}$, Equation is simplified as follows:

$$\frac{\pi N \Delta x}{N_\theta \sqrt{2}} \leq \Delta \rho. \qquad \qquad (4)$$

The optimum relationship between N_θ and $\Delta\rho$ is where equality holds.

So far we have assumed that $\cos\theta_n$ and $\sin\theta_n$ are of infinite precision. Now, we must check their accuracy. If the rounding error of $\cos\theta_n$ is δC and that of $\sin\theta_n$ is δS, the square of the ρ value rounding error satisfies the following Schwartz inequality:

$$(\delta\rho)^2 = (x_i\,\delta C + y_i\,\delta S)^2 \leq (x_i{}^2 + y_i{}^2)(\delta C^2 + \delta S^2) \leq N^2 (\Delta x)^2\, \delta^2,$$

where δ^2 is the maximum value of δC^2 or δS^2. For $\delta\rho$ to be less than $\Delta\rho$ along any curves, δ must be less than $\dfrac{\Delta\rho}{N\Delta x}$.

Finally, the maximum number for each cell of the accumulator array can be determined. The maximum count equals the maximum number of feature points that can be contained in a bar-shaped window[17] whose width is $\Delta\rho$. The window contains the maximum number of points when it is oriented diagonally in the image space. The area of the common portion of the image and the window is approximately $\sqrt{2}\, N\Delta x \cdot \Delta\rho$. There is one feature point in each square whose size is Δx, so the average number of points in the window is $\sqrt{2}\, N\dfrac{\Delta\rho}{\Delta x}$.

The bit length of the counter that represents one cell and must count up to

$\log_2 \sqrt{2}\, N\dfrac{\Delta\rho}{\Delta x}$ bits.

5. Hardware specifications

This chapter describes how the quantization parameters, whose optimum relationships were described in the preceding chapter, were reflected in the experimental hardware model.

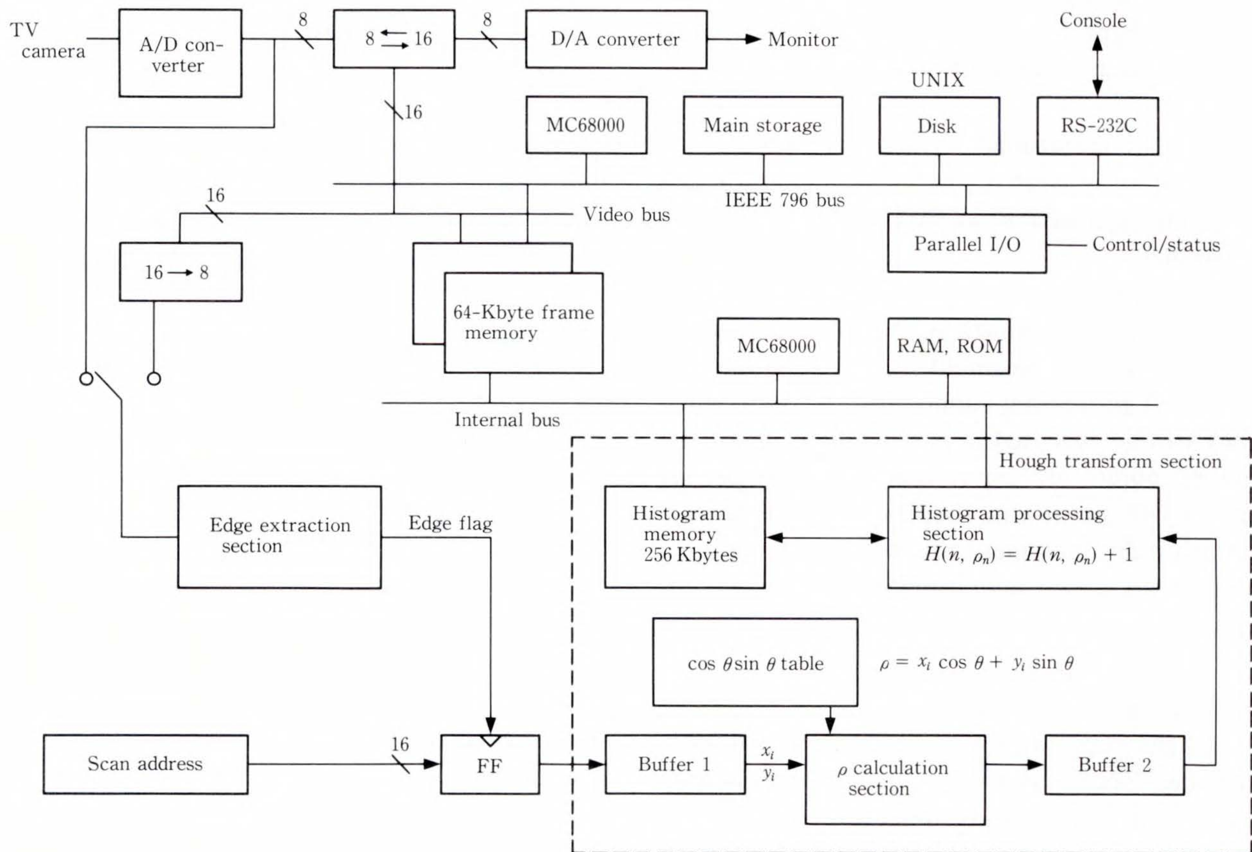


Fig. 1—System configuration.

The size of the image space is 256 x 256, so $N = 256$ and the ratio of $\Delta\rho$ to Δx is set to 1. In other words, adjacent horizontal or vertical lines can be differentiated. Under these conditions, the number of ρ partitions can reach $256 \cdot \sqrt{2}$. According to the optimum relationship between $\Delta\rho$ and $\Delta\theta$, the θ partition number is $\frac{256 \cdot \pi}{\sqrt{2}}$. To simplify hardware manufacturing requirements, the number of ρ and θ partitions were both set to 512. This results in some redundancy because $(512 - 256 \cdot \sqrt{2})\ N_\theta$ cells will never be used. The step size of θ is coarser than the optimum value, so interpolation is desirable, but we did not use interpolation because Equation (3) is satisfied for most of the feature points (i.e., those in a circular range within a radius of 163 picture elements).

Although $\log_2 \sqrt{2}\, N = 8.5$ bits are required for each cell of the accumulator array, only eight bits were taken. This was done for two reasons: first, to ensure compatibility with other parts of the image processors or microprocessors in the experimental model and second, because continuous straight diagonal lines are rare. The values $\cos\theta_n$ and $\sin\theta_n$ were quantized at eleven bits including signs.

6. System configuration

Figure 1 shows the system configuration, which comprises three main sections: the edge extraction section for extracting edge points from TV signals, the Hough transform section where sine curve histograms are created from extracted edge points according to Equation (1), and the microprocessor for general control of the total system and its peripheral devices.

6.1 Edge extraction section

TV signals are sampled at 6 MHz and quantized in 8-bit samples. The edge extraction section processes the quantized signals at the video rate. This section also has two convolver stages. By changing stage combinations, either Sobel operation + fine-line binary coding or smoothing + Laplacian + zero crossing detection can be selected (see Fig. 2). The mask is either 3 x 3 or 4 x 4 (see Table 1). The scan addresses of edge data obtained here are transferred to the Hough transform section at the video rate through a bipolar FIFO buffer with a capacity of 1 kW (1 W = 16 bits).

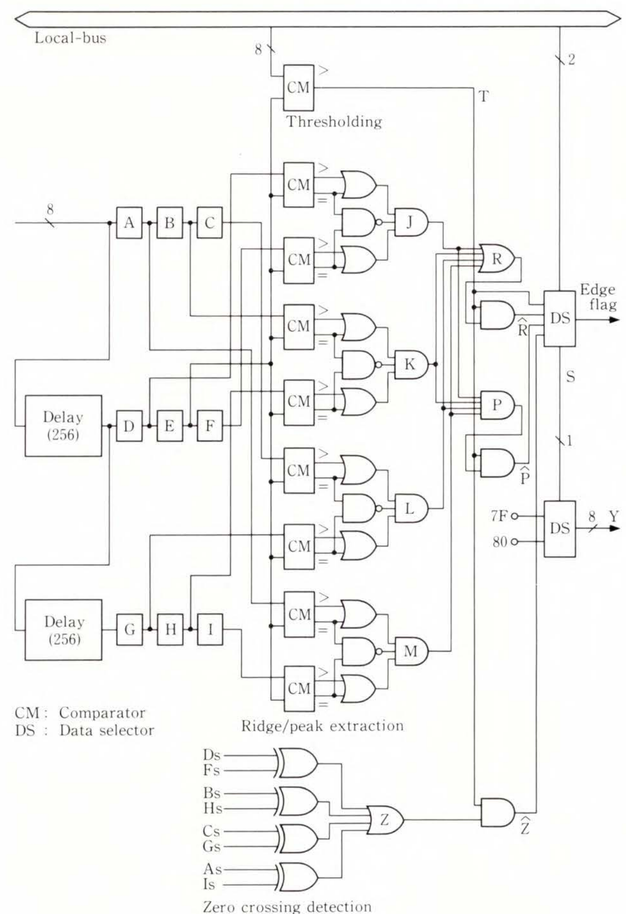


Fig. 2—Edge Extracting Structure.

6.2 Hough transform section

The Hough transform section consists of two subsections. The ρ calculation section uses Equation (1) to calculate ρ at each edge point and make a 512-word (n, ρ_n) table for each edge point. These tables are subsequently used by the histogram processing section. The contents of the histogram memory location addressed by (n, ρ_n) are incremented by one for $n = 0$ to 511. This procedure is repeated for all edge points. The ρ calculation section consists of bipolar ROMs, where sine and cosine values are stored to an accuracy of eleven bits, two 50 ns multipliers, and a 16-bit adder (see Fig. 3).

Table 1. Convolver coefficients

Filter	Coefficients		
	3 x 3	4 x 4 (1)	4 x 4 (2)
Smoothing	$\frac{1}{16}\begin{bmatrix}0&0&0&0\\0&1&2&1\\0&2&4&2\\0&1&2&1\end{bmatrix}$	$\frac{1}{64}\begin{bmatrix}1&2&2&1\\2&4&4&2\\2&4&4&2\\1&2&2&1\end{bmatrix}$	
Laplacian	$\begin{bmatrix}0&0&0&0\\0&-1&-1&-1\\0&-1&8&-1\\0&-1&-1&-1\end{bmatrix}$	$\begin{bmatrix}-1&-1&-1&-1\\-1&3&3&-1\\-1&3&3&-1\\-1&-1&-1&-1\end{bmatrix}$	$\begin{bmatrix}-1&-2&-2&-1\\-2&5&5&-2\\-2&5&5&-2\\-1&-2&-2&-1\end{bmatrix}$
Enhancement	$\begin{bmatrix}0&0&0&0\\0&-1&-1&-1\\0&-1&9&-1\\0&-1&-1&-1\end{bmatrix}$	$\frac{1}{4}\begin{bmatrix}-1&-1&-1&-1\\-1&4&4&-1\\-1&4&4&-1\\-1&-1&-1&-1\end{bmatrix}$	$\frac{1}{4}\begin{bmatrix}-1&-2&-2&-1\\-2&6&6&-2\\-2&6&6&-2\\-1&-2&-2&-1\end{bmatrix}$
Gradient (Sobel)	$\begin{bmatrix}0&0&0&0\\0&1&0&-1\\0&2&0&-2\\0&1&0&-1\end{bmatrix}$	$\begin{bmatrix}1&1&-1&-1\\2&2&-2&-2\\2&2&-2&-2\\1&1&-1&-1\end{bmatrix}$	$\begin{bmatrix}1&2&-2&-1\\2&4&-4&-2\\2&4&-4&-2\\1&2&-2&-1\end{bmatrix}$
	$\begin{bmatrix}0&0&0&0\\0&1&2&1\\0&0&0&0\\0&-1&-2&-1\end{bmatrix}$	$\begin{bmatrix}1&1&1&1\\2&2&2&2\\-2&-2&-2&-2\\-1&-1&-1&-1\end{bmatrix}$	$\begin{bmatrix}1&2&2&1\\2&4&4&2\\-2&-4&-4&-2\\-1&-2&-2&-1\end{bmatrix}$

The histogram memory has a capacity of $512 \times 512 \times 8$ bits. It consists of 64 K-bit CMOS RAMs with an access time of 150 ns.

7. The complementary Hough transform

Brown proposes a new scheme for Hough transforms which uses negative votes as well as the usual positive votes[19]. This complementary Hough transform, as he calls it, can eliminate off-peak positive votes in the parameter space, sharpen peaks, and reduce sidelobe bias and variance. For detecting circles, Brown uses feature point spread functions consisting of a ring of ls surrounded both inside and out by a ring of $-\frac{1}{2}$s.

The complementary Hough transform can be considered almost equivalent to performing the Laplacian on the final histogram based on positive votes. When there is only one feature point in the image space, the sinusoidal curves consisting of positive and negative votes are almost the same as the result obtained by applying the Laplacian to the curve.

Further, because the Laplacian is a linear operation and the superposition principle applies to the process of making histograms, the sequence of these two operations can be reversed freely.

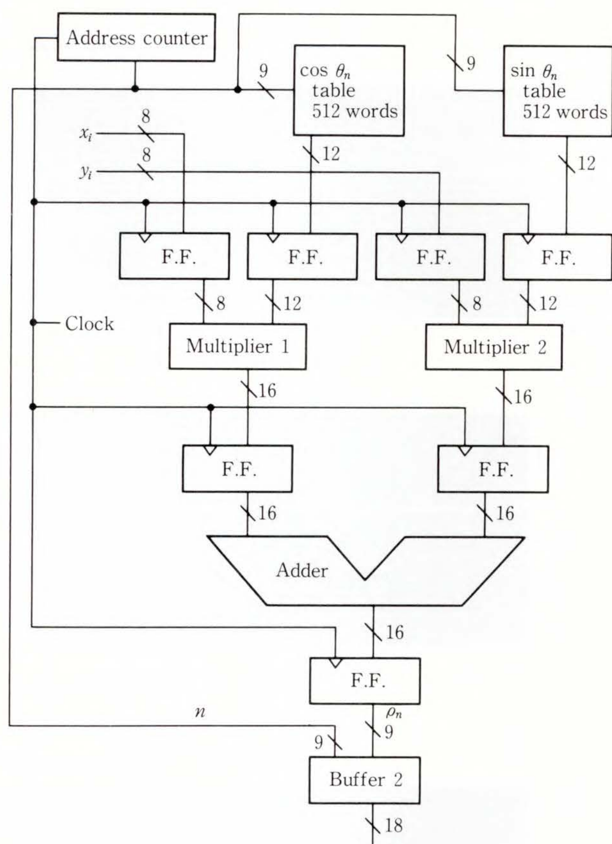


Fig. 3—ρ calculation section.

Modified complementary Hough transforms based on the Laplacian are suitable for the processor architecture. All that is required is to perform the Laplacian on the histogram in frame memory before performing peak detection. The extra processing time needed is 16.7 ms. Some experimental results of using this processing sequence are shown in Chapter 9.

8. Parallel architectures

Yalamanchili et al. proposes a parallel architecture in which the Hough space is divided into P partitions[5] and each processing element (PE) is required to increment $(\frac{N_\theta}{P})$ accumulators instead of N_θ for each feature point.

The parallel architectures in which each of the different θ coordinates is assigned a PE is an extreme case in which N_θ equals P. We refer to this as completely parallel architecture. The completely parallel architectures that fall within the framework proposed by Yalamanchili et al. are the fastest, but they also require the most hardware resources.

When the degree of parallelism increases, the percentage of circuitry dedicated to multiplica-

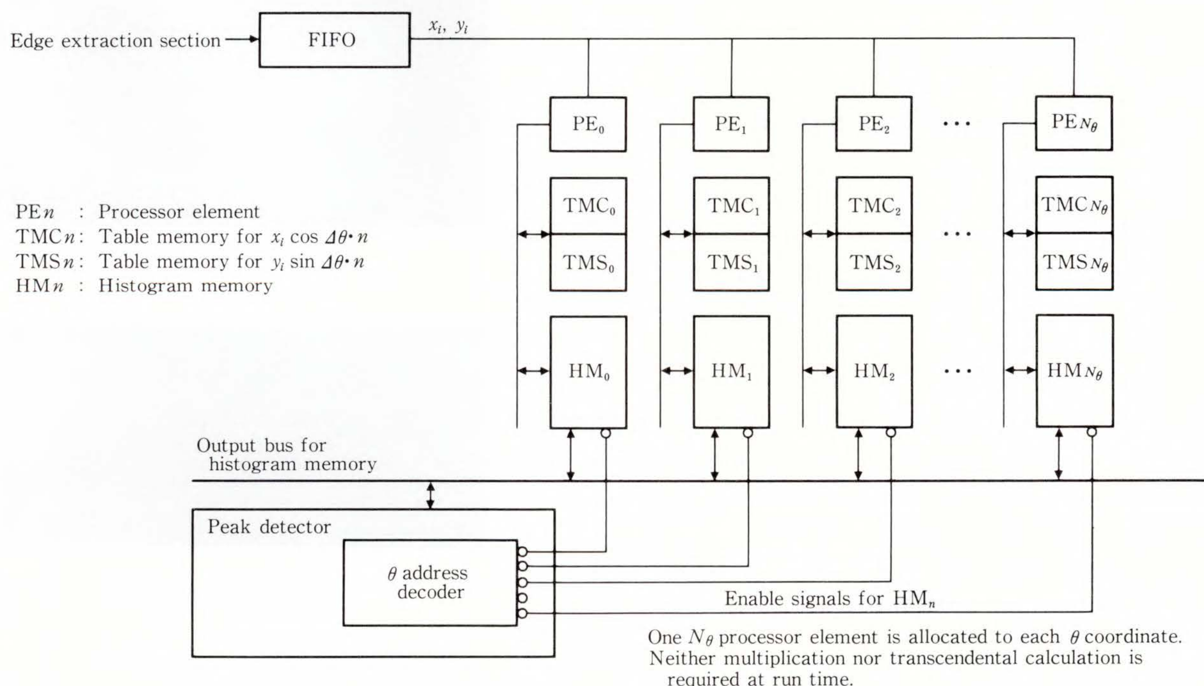


Fig. 4—Parallel architecture for Hough transform.

tion is the most important factor affecting economy. One set of parallel multipliers (12 bits x 12 bits) consists of 1 800 gates, so if N_θ is 512, a total of 920 K-gate are required for multiplication alone. Trigonometric functions, however, become less significant as the degree of parallel processing increases. This is because in completely parallel architectures, each PEn need hold only one set of cos $n\Delta_\theta$ and sin $n\Delta_\theta$ values.

Here, we propose a completely parallel architecture that does not use any multipliers. Figure 4 is a block of diagram of this architecture.

Each x_i and y_i input can take 256 different values, so each x_i cos $n\Delta_\theta$ and y_i sin $n\Delta_\theta$ result for one PE can also have one of a possible 256 values. This means that a 2 x 256 x 10 bit table can be used instead of the multipliers. Multiplication results were assumed to be 10-bit values after taking into account that the sum of both results would be quantized to nine bits.

This architecture is advantageous from the point of view both of scale and of speed. First, if we assume a completely parallel architecture implemented with VLSI, we can compare the proposed system with a multiplier configuration from the standpoint of chip area. On one hand, a recent Fujitsu gate array design contains 45 000 gates in an 11.5 x 11.8 mm^2 area[20], so if the above-mentioned multiplier of 12 x 12 bits is configured into it, it will occupy 5.43 mm^2. On the other hand, Fujitsu's 4 M-bit RAM with an access time of 70 ns measures 4.84 x 13.16 mm^2 in report[21]. If the above table is implemented with the same technology, it will occupy 0.078 mm^2. The area required by the lookup table is thus about 1/70 of that required by a multiplier with 1 800 gates. These values may contain some errors. To be completely accurate, the multiplier should also have a circuit configuration suitable for 8-bit x 12-bit multiplication. The purpose here, however, is to estimate relative and qualitative trends of both systems.

The operating speeds of systems based on table look-up architecture are faster because the table capacities are small and decoding time is much shorter than for 4 M-bit RAM chips.

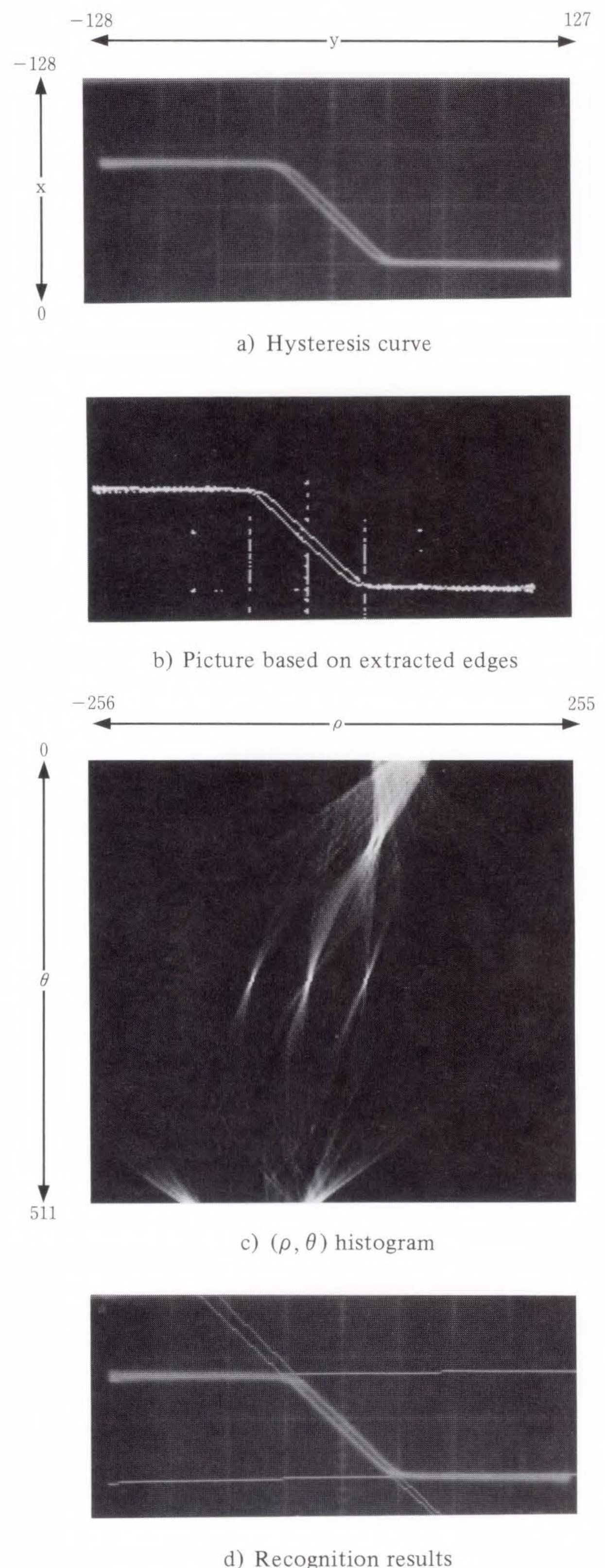
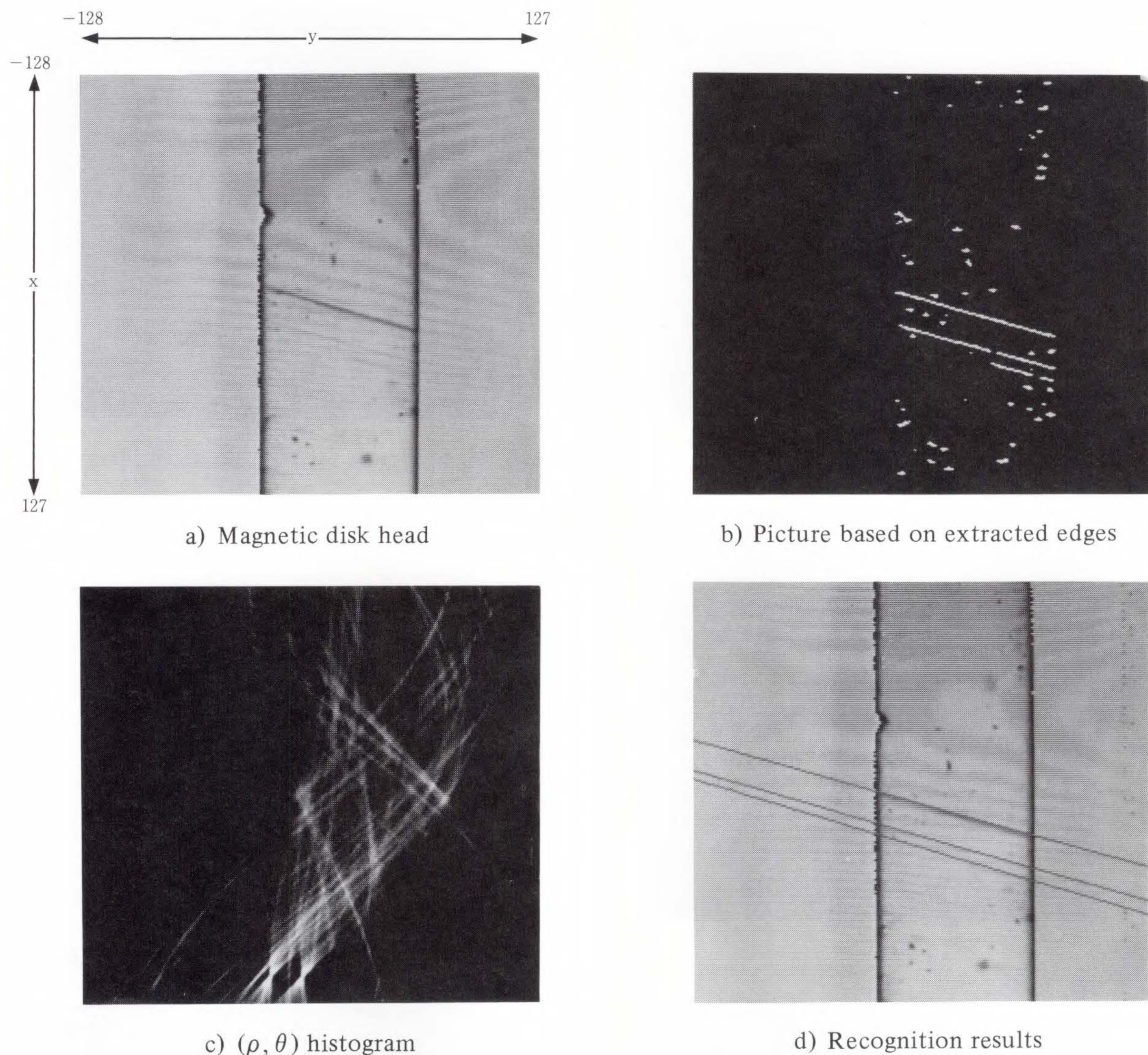


Fig. 5–Automatic measurements.



a) Magnetic disk head

b) Picture based on extracted edges

c) (ρ, θ) histogram

d) Recognition results

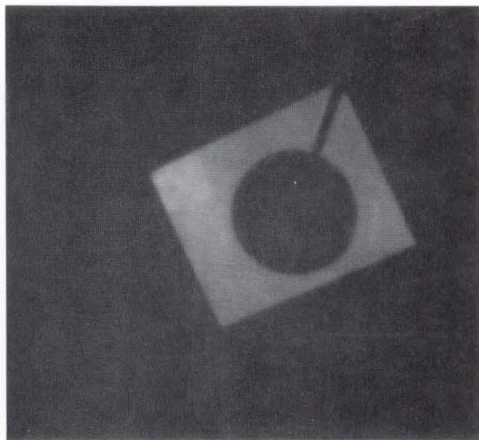
Fig. 6–Inspection applications.

The proposed system has another advantage. It can be assembled using the most primitive type of commercially available microprocessors, such as those without multipliers. Each microprocessor references the table twice per feature point, performs one addition operation and increments the contents of the indicated memory address by the added value plus the displacement. The time required for these operations can be in the order of a few tens of microseconds.

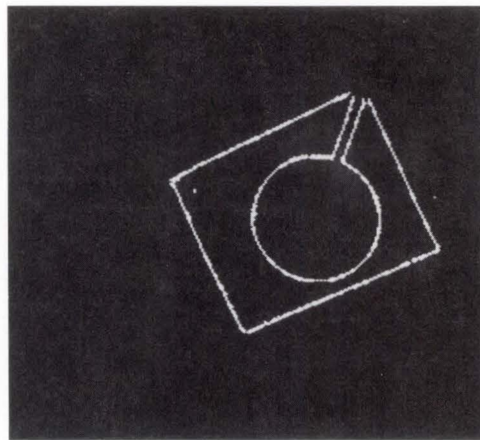
9. Processing examples

This chapter presents the experimental results obtained using the model processor on three different images. Figures 5-a) to -d) and Figs. 6-a) to -d) show the results of using this processor in automatic measurement and inspection applications. Figures 7-a) to -f) show the effects of using the modified complementary Hough transform.

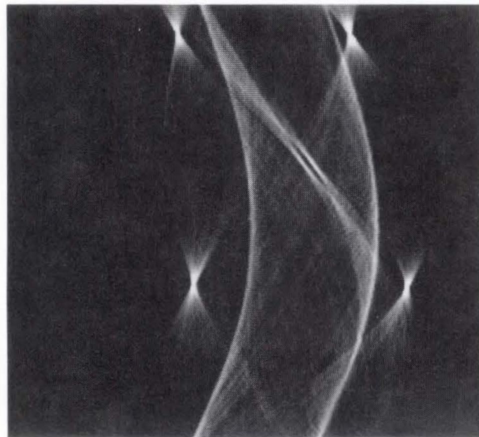
Figure 5-a) is a hysteresis curve of a thin-film head of a magnetic disk unit monitored by a curve tracer. Shown here is the upper half of the image plane. The main items to be measured are the upper and lower limits of the magnetic flux density (vertical axis) and the slope of the tangent line. The horizontal axis represents the



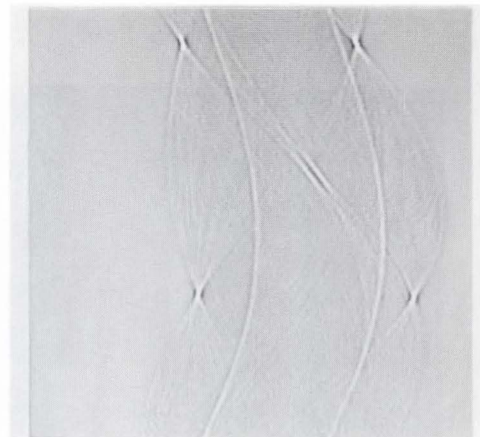
a) Mechanical part



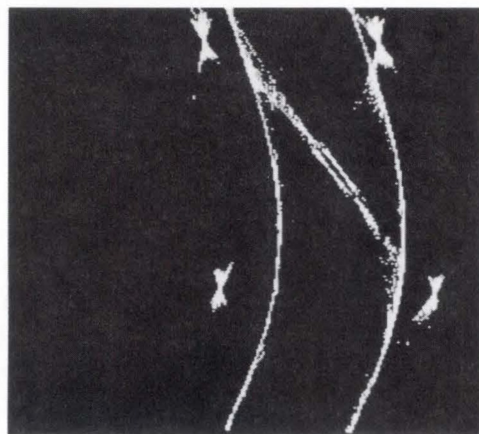
b) Picture based on extracted edges



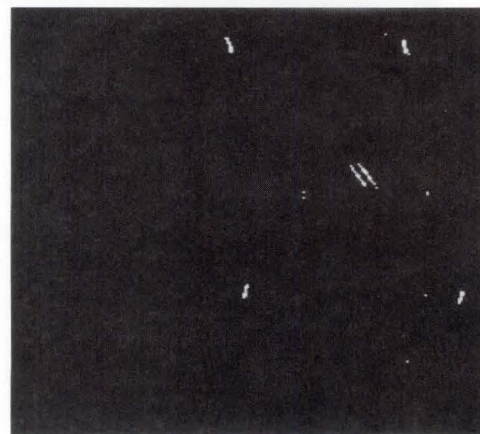
c) (ρ, θ) histogram



d) Modified complementary Hough transform



e) Cells in Fig. 7-c) whose counts exceed one half the peak height because of short line segments



f) Cells in Fig. 7-d) whose counts exceed one half the peak height because of short line segments

Fi.g 7—Modified complementary Hough transforms.

external magnetic field. Figure 6-a), taken through a microscope, shows the magnetic disk head. Our objective was to detect any oblique cracks or damage, which show up as the one

short and two long black lines in the figure.

Figure 7-a) is an image of a mechanical part. Figures 5-b), 6-b), and 7-b) are based on extracted edge points.

Figures 5-b), 6-b), and 7-b) show, respectively, the use of zero crossing after the Laplacian (4 x 4 Laplacian (2) in Table 1), enhanced vertical differentiation (4 x 4 enhancement (1), and the lower of 4 x 4 Sobel (1) in Table 1 cascaded) using Sobel operation (4 x 4 Sobel (1) in Table 1).

Figures 5-b), 6-b), and 7-b) respectively show the use of zero crossing after the Laplacian from the edge images.

The recognition results were converted back to their original picture space coordinates and overlaid on the original pictures, as shown in Fig. 5-d) and 6-d).

In these examples, thresholding and non-maximum suppression are used. The thresholds are 40 and 25, and the 5 x 5 and 7 x 7 windows are searched for maxima.

Figure 7-d) shows the (ρ, θ) space based on the modified complementary Hough transform. Applying the Laplacian after smoothing (4 x 4 Laplacian (2) and 4 x 4 smoothing in Table 1) to Fig. 7-c) results in Fig. 7-d).

Figures 7-c) to -f) provide evidence of how this method sharpens peaks and reduces sidelobe bias[19]. Figures 7-e) and -f) show the cells whose counts exceed one half of the peak height as a result of the two parallel short line segments for Figs. 7-c) and -d), respectively.

10. Conclusion

We have developed dedicated straight-line detection hardware that uses the Hough transform method, and which is only slightly affected by noise and gaps even in complicated backgrounds. This hardware takes less than one second to produce straight-line parameters after video image capture.

We have demonstrated the good recognition capabilities of the processor in automatic inspection and measurement applications in factories and laboratories.

We have proposed a highly parallel architecture that does not require any multiplication or trigonometric calculations at run time. The advantages of this architecture have been discussed from points of view of scale and speed.

Future work will be directed towards incorporating this processor into robot vision systems.

11. Acknowledgement

The authors wish to thank Mr. S. Kawakami for his encouragement during this study.

References

1) Hough, P.V.C.: Method and Means for Recognizing Complex Patterns. U.S. Patent 3 069 654, (1962).
2) Evans, F.: A Survey and Comparison of the Hough Transform. IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, 1985, pp. 378-380.
3) Iannino, A., and Shapiro, S.D.: A Survey of the Hough Transform and Its Extensions for Curve Detection. Proc. IEEE Computer Society Conf. on Pattern Recognition and Image Processing, 1978, pp. 32-38.
4) Merlin, P.M., and Farber, D.J.: A Parallel Mechanism for Detecting Curves in Pictures. *IEEE Trans. Comput.*, **C24**, 1, pp. 96-98 (1975).
5) Yalamanchili, S., and Aggarwal, J.K.: A System Organization for Parallel Image Processing. *Pattern Recognition*, **18**, 1, pp. 17-29 (1985).
6) Ballard, D.H.: Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, **13**, 2, pp. 111-122 (1981).
7) Silberberg, T.M.: The Hough Transform on the Geometric Arithmetic Parallel Processor. IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, pp. 387-393 (1985).
8) Dyer, C.R.: Gange Inspection Using Hough Transforms. *IEEE Trans.*, **PAMI-5**, 6, pp. 621-623 (1983).
9) Kushnir, M., Abe, K., and Matsumoto, K.: Recognition of Handprinted Hebrew Characters Using Features Selected in the Hough Transform space. *Pattern Recognition*, **18**, 2, pp. 103-113 (1985).
10) LI, C.C, Mancuso, J.F., Shu, D.B., Sun, Y.N., and Roth, L.D.: A Preliminary study of Automated Inspection of VLSI Resist Patterns. Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 474-480 (1985).
11) Huang, K.Y., Fu, K.S., Sheen, T.H., and Cheng, S.W.: Image Processing of Seismograms: (A) Hough Transformation for the Detection of Seismic

Patterns; (B) Thinning Processing in the Seismogram. *Pattern Recognition*, **18**, 6, pp. 429-440 (1985).

12) Inigo, R.M., McVey, E.S., Berger, B.J., and Wirtz, M.J.: Machine Vision Applied to Vehicle Guidance. *IEEE Trans*. **PAMI-6**, 6, pp. 820-826 (1984).

13) Wallace, R., Stentz, A., Thorpe, C., Moravec, H., Whittaker, W., and Kanade, T.: First Results in Robot Road-Following. Proc. of the 9th Int. Joint Con. on Artificial Intelligence, Aug. 1985, Los Angeles, Calif., 2, pp. 1089-1095 (1985).

14) Waxman, A.M., Moigne, J.L., and Srinivasan, B.: Visual Navigation of Roadways. Proc of the IEEE Int. Conf. on Robotics and Automation, pp. 862-867 (1985).

15) Tsuji, S., Zhen, J.Y., and Asada, M.: Stereo Vision of a Mobile Robot: World Constraints for Image Matching and Interpretation. Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 1594-1599 (1986).

16) Duda, R.O., and Hart, P.E.: Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Communs. Ass. Comput.*, **15**, 1, pp. 11-15 (1972).

17) Van Veen, T.M., and Groen, F.C.A.: Discretization Errors in the Hough Transform. *Pattern Recognition*, **14**, 1-6, pp. 137-145 (1981).

18) Maruyama, T., and Uchiyama, T.: Real-time Image Processor with Two Convolution Filter Modules and a Peak Extraction Module. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 546-549 (1983).

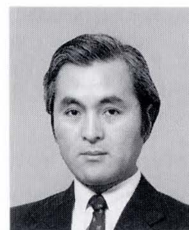
19) Brown, C.M.: Inherent Bias and Noise in the Hough Transform. *IEEE Trans*., **PAMI-5**, 5, pp. 493-505 (1983).

20) Takahashi, H., Fujii, S., Kawauchi, K., Inada, T., and Gambe, H.: A Sub-Micron CMOS Echo Canceller Using a DSP Cell. IEEE Int. Solid-State Circuits Conf., pp. 148-149 (1987).

21) Mochizuki, H., Kodama, Y., Nakano, T., Ema, T., and Yabu, T.: A 70 ns 4Mb DRAM in a 300 mil DIP Using 4-Layer Poly. IEEE Int. Solid-State Circuits Conf., pp. 284-285 (1987).

Keishi Hanahara

Mechatronics-In-Space Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Physics
Tokyo Institute of Technology 1970
Master of Physics
Tokyo Institute of Technology 1972
Specializing in Computer Vision, Image and Signal Processing



Takashi Uchiyama

Mechatronics-In-Space Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Mechanical Eng.
Tokyo Institute of Technology 1970
Master of Production Mechanical Eng.
Tokyo Institute of Technology 1973
Specializing in Computer Vision, Intelligent Robotics

Tsugito Maruyama

Mechatronics-In-Space Laboratory
FUJITSU LABORATORIES, KAWASAKI
Bachelor of Electric Eng.
Niigata University 1974
Dr. of Electric & Communication Eng.
Tohoku University 1980
Specializing in Computer Vision

UDC 538.945:621.318

Reproducible Nb/AlO$_x$/Nb Josephson Junctions

• Shiro Ohara • Takeshi Imamura • Shinya Hasuo

(Manuscript received November 27, 1987)

Nb/AlO$_x$/Nb Josephson junctions were fabricated with different shapes and structures, and the influence of control lines, the size of the base electrode, contacts near the junctions, and junction size on junction characteristics was studies.
Junctions fabricated by dc sputtering were found to be of high quality, regardless of shapes or structures.

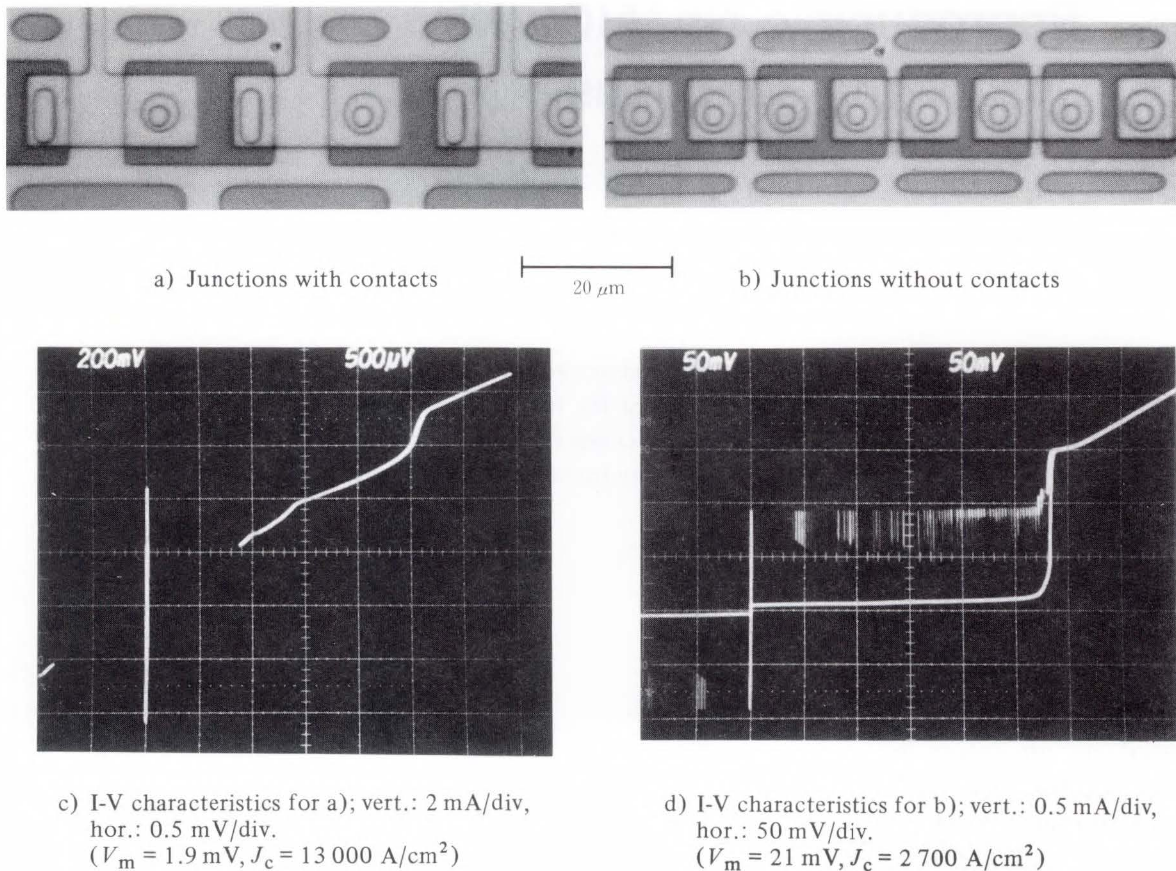
1. Introduction

The Nb/AlO$_x$/Nb Josephson junction has potential as an integrated Josephson circuit element because of its high quality and stability. The authors have developed the technology for fabricating Nb/AlO$_x$/Nb junctions, and applied the junctions to logic circuits. High-speed operation was confirmed for integrated Josephson circuits having several hundred gates.

To utilize the Nb/AlO$_x$/Nb junctions in circuits, the junctions must be of high quality. This requires small subgap currents and sharp increasing currents at gap voltages. Critical current uniformity in a chip is also important. In studying the fabrication techniques for Nb/AlO$_x$/Nb junctions the authors have clarified the ralationship between junction quality and deposition conditions[1]. In the course of this study, the authors encountered many difficulties in reproducing high-quality junctions. First, poor-quality junctions were frequently found on the same wafer with high-quality junctions. Second, the characteristics of high-quality junctions often deteriorated during subsequent processes, such as the formation of an insulation layer and Nb control lines. This phenomenon appears to be related to the size and shape of the junctions, base electrode, and control lines. This also seems to depend on the configuration of junctions and contacts.

Figure 1 compares the I-V characteristics observed in the same chip. Junction materials (Nb and Al) were deposited by *rf* magnetron sputtering. Figures 1-a) and -b) show microphotographs of junctions, and Figs. 1-c) and -d) are their respective current-voltage characteristics. Figure 1-a) is a pair connected in series consisting of a junction and a contact between the base electrode and the wiring layer. Figure 1-b) is 6-μm-diameter junctions connected in series. Despite the fact that they are on the same chip, their current-voltage characteristics are quite different. The quality parameter V_m is 1.9 mV for the junctions in Fig. 1-a) and 21 mV for those in Fig. 1-b). The critical current density J_c for Fig. 1-a) is five time that for Fig. 1-b). Here, V_m is defined as $I_c \times R_{sg}$ (I_c: critical current, R_{sg}: subgap resistance at 2mV). The only difference between these junctions is whether the contact is beside the measured junction. This suggests that junction characteristics are affected by the junction configuration. However, these deteriorated junctions could not be reproduced, and the causes of the deterioration are not clear.

To study this irregularity in the junction characteristics, the authors fabricated Nb/AlO$_x$/Nb junctions with different dimensions and configurations, and evaluated their characteristics. This time, junctions were deposited by



a) Junctions with contacts   20 μm   b) Junctions without contacts

c) I-V characteristics for a); vert.: 2 mA/div, hor.: 0.5 mV/div. (V_m = 1.9 mV, J_c = 13 000 A/cm^2)

d) I-V characteristics for b); vert.: 0.5 mA/div, hor.: 50 mV/div. (V_m = 21 mV, J_c = 2 700 A/cm^2)

Fig. 1—Microphotographs of junctions on the same chip and its I-V characteristics.

dc magnetron sputtering, as this technique yields junctions of higher quality than those fabricated by *rf* magnetron sputtering[1]. (Therefore, dc magnetron sputtering is now used for circuit fabrication[2].) This method is also used because it is important to study the properties of junctions fabricated by methods currently used for integrated circuits.

The authors examined:

1) the influence of the control line above the junctions,
2) the influence of contacts near the junctions, and
3) the influence of the dimensions of the base electrode.

2. Fabrication

Figure 2 shows cross sections of the test devices. The junction structure consists of a Nb base electrode (B), an Al-AlO$_x$ tunneling barrier,

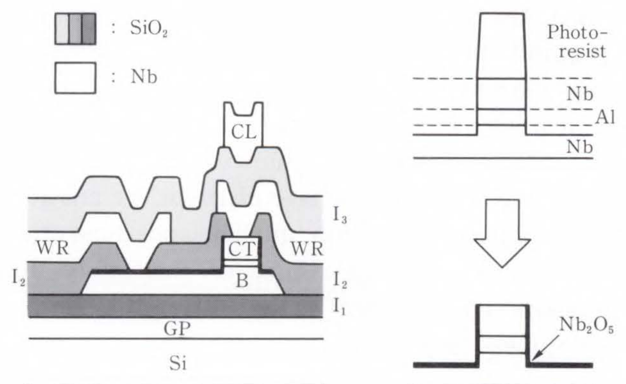


Fig. 2—Cross section of test deveices.

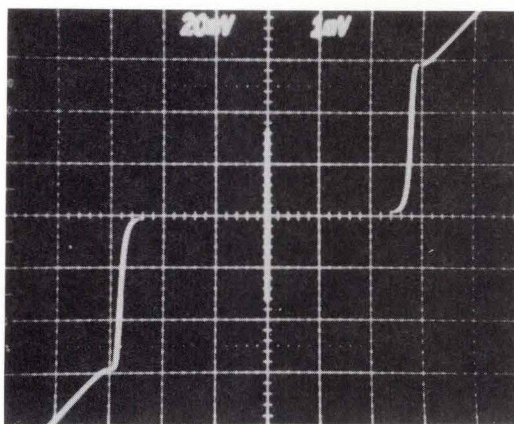
and a Nb counter electrode (CT). These components were deposited on a wafer two inches in diameter in the same vacuum run. In addition to the junction electrodes, three metal layers were deposited: a Nb ground plane (GP), a Nb wiring layer (WR), and Nb control line (CL).

They were insulated by SiO_2 except at the contact holes. All layers were deposited by sputtering. Nb and Al were deposited by dc magnetron sputtering, and SiO_2 was deposited by *rf* magnetron sputtering. All layers were patterned by reactive ion etching (RIE). The surface of the B layer and the edge of the junction were oxidized by liquid anodization. The thickness and etching gas for each layer are listed in Table 1.

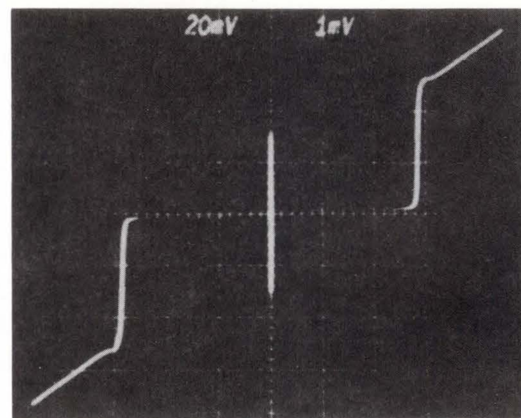
Table 1. Thickness and etching gas

Layer	Thickness (nm)	Etching gass
GP	400	CF_4 + 20% O_2
I_1	300	CHF_3 + 15% O_2
B	250	CF_4 + 10% O_2
CT	200	CF_4 + 5% O_2
I_2	400	CHF_3 + 15% O_2
WR	600	CF_4 + 10% O_2
I_3	800	CF_4 + 15% O_2
CL	1 000	CF_4 + 5% O_2

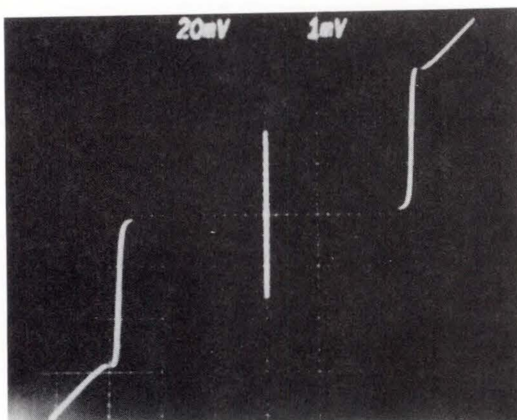
Junctions having different dimensions and configurations were made on the same wafers. To study the influence of the I_3 and CL layers, the wafers were cut into two parts after the wiring layer (WR) was patterned. For measurement, one of them was diced into chips, and the other was diced after the subsequent deposition and patterning of the I_3 and CL layers. Chips diced before the deposition of the I_3 layer are



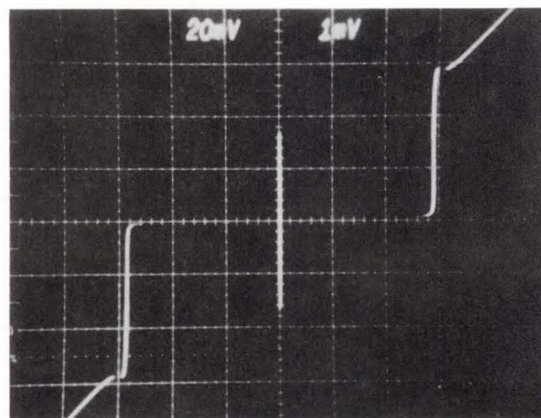
a) "Without CL" (V_m = 57 mV, J_c = 1 450 A/cm^2)



b) "With CL" (V_m = 41 mV, J_c = 1 400 A/cm^2)



c) "Without CL" (V_m = 41 mV, J_c = 1 400 A/cm^2)



d) "With CL" (V_m = 68 mV, J_c = 1 490 A/cm^2)

Fig. 3 – Junction characteristics of one wafer and the other wafer (vert.: 0.2 mA/div, hor.: 1 mV/div).

termed chips "without CL" and chips diced after the CL layer was patterned are termed chips "with CL".

3. Experimental results

The junctions were evaluated by the parameter V_m, and by critical current uniformity.

3.1 Influence of the CL layer

First, the influence of I_3 and CL layer deposition on junction characteristics was evaluated. Figure 3 compares the characteristics of junctions on chips "without CL" and those on chips "with CL". The junction is 6.5 μm in diameter, and the control line is 2.5 μm wide. Figures 3-a) and -b) show the change in the junction characteristics after the formation of the I_3 and CL layers in one wafer. Figures 3-c) and -d) show the change in the other wafer. Figures 3-a) and -c) represent the junction characteristics before deposition of I_3 layer. On one wafer, V_m decreases from 57 mV to 41 mV after the formation of the I_3 and CL layers, and on the other wafer, V_m increases from 40 mV to 68 mV. As a result, we could not observe any relationship between the quality of the junction and the deposition of I_3 and CL on the junctions. Variation in V_m shown in Fig. 3 should be considered due to V_m scattering on the wafer. For the critical current density, J_c, changes induced by the formation of I_3 and CL layers were also small. This was the same as the intrinsic J_c spread on a wafer.

The critical current density and V_m measured for different junction areas and control line widths are listed in Tables 2 and 3. Three different junction areas were prepared ranging from 3.5 μm to 6.5 μm in diameter. The control lines were 2.5 μm and 4.0 μm wide. A junction where the deposited CL was completely removed was also prepared on the chip. This is referred to as having a CL width of 0 μm. No distinct relationship between the junction characteristics and control line widths and/or junction diameters were observed.

The influence of CL layers on the I_c uniformity was also evaluated for 100 junctions connected in series. The results are shown in Table 4. The diameter of the junction area is 6.5 μm-diameter. The control line is 0 μm, 2.5 μm, and 22.5 μm wide for 25-junction chains. The I_c spread ranges from 5.6 percent to 7.4 percent. This does not indicate dependence on the width of the control line. It is almost the

Table 2. Current density

(A/cm^2)

Junction diameter (μm) \ Control line width (μm)	0	2.5	4.0
3.5	1 250	1 250	–
4.5	1 510	1 380	1 380
6.5	1 450	1 390	–

Table 3. Quality parameter V_m

(mV)

Junction diameter (μm) \ Control line width (μm)	0	2.5	4.0
3.5	63	52	–
4.5	44	42	42
6.5	62	53	–

Table 4. Control line width and critical current spread (junction dia.: 6 μm)

Control line width (μm)	I_c spread (%)	Number of junctions
0	6.0	25
2.5	7.4	25
22.5	5.6	25
–	6.4	100

Table 5. B layer sizea and junction characteristics

B layer size (μm^2)	Current density (A/cm^2)	Vm (mV)
3 250	650	50
1 750	680	51
1 350	740	49
1 000	700	49
700	680	52
450	690	51
250	730	52

same as the value of 6.4 percent measured for the 100-junction chain without the CL layer. It is reasonable, therefore, to conclude that the formation of the I_3 and CL layers has no significant influence on J_c spread.

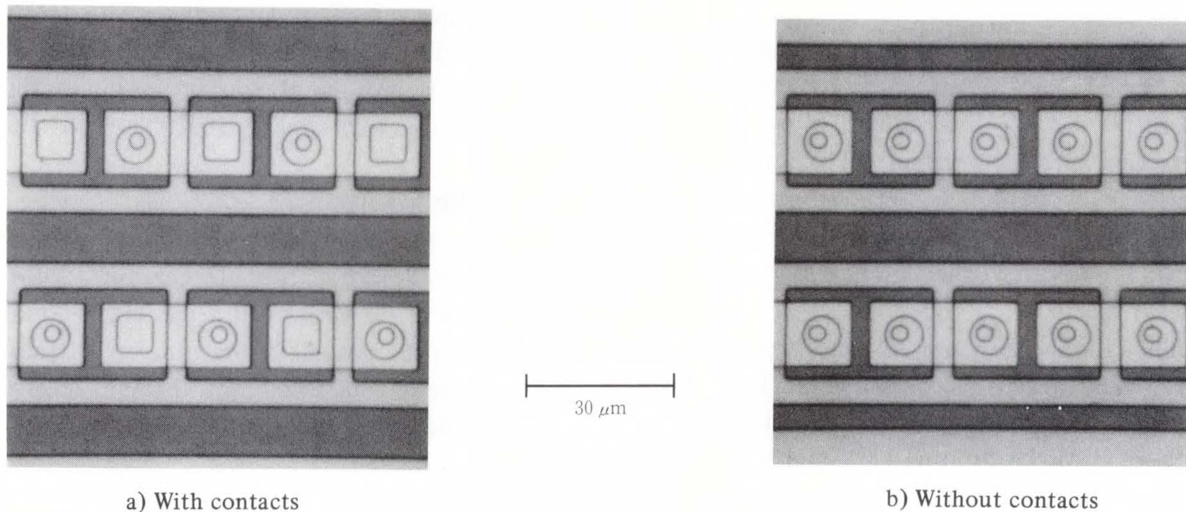
3.2 Influence of base electrode size

We evaluated the characteristics for junctions with different base electrode sizes. Other junction dimensions are the same for all samples. The critical current density and V_m are listed in Table 5. The area of the base electrode varies from 250 μm^2 to 3 250 μm^2, and the junction area is 4.5 μm in diameter. The current density and V_m are quite constant for all junctions, and no relationship with the base electrode size was observed.

3.3 Influence of contacts

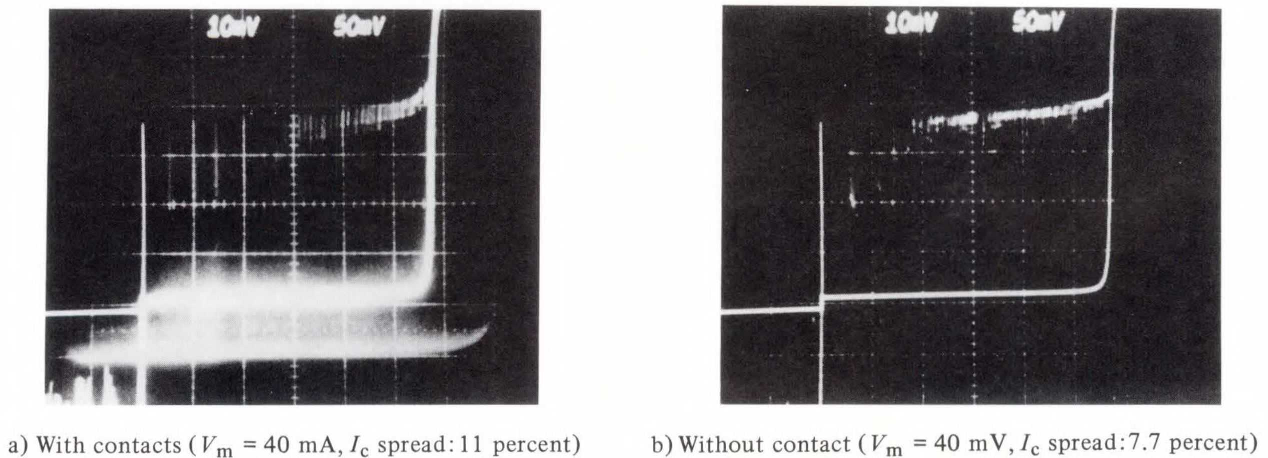
When the contact between the base electrode and the wiring is placed near a junction the characteristics of the junction often deteriorated in the same way as shown in Fig. 1. To study the influence of the contact, pairs of junctions and contacts were prepared in series, and compared with the simple junctions connected in series.

Figure 4 shows microphotographs of the test devices. The junction is 6.5 μm in diameter,



a) With contacts b) Without contacts

Fig. 4—Junction array for evaluation the influence of contacts.



a) With contacts (V_m = 40 mA, I_c spread: 11 percent) b) Without contact (V_m = 40 mV, I_c spread: 7.7 percent)

Fig. 5—I-V characteristics of 100 junctions connected in two series (vert.: 0.1 mA/div, hor.: 50 mV/div).

and the contact is 7 μm square. Figure 5 shows I-V characteristics for 100 junctions connected in series, Figs 5-a) with contacts and -b) without contacts. The value of V_m was evaluated at 40 mV for both samples. Here, an I_c distribution median was used to calculate V_m. The I_c spread was slightly different, 11 percent for Figs. 5-a) and 7.7 percent for -b). In Fig. 4-a), a few junctions with very large I_c were observed in the I-V characteristics, and they apparently enlarged the I_c spread. If they are omitted, I_c spread is thought to be almost constant. Thus, the authors concluded that the influence of the contact near the junction is not observed in the junction characteristics.

4. Conclusion

Junctions with different sizes and configurations were fabricated and evaluated to determine the influence of the control line, the base electrode, and the contact on junction characteristics. No distinct differences in the junction quality and the I_c spread were observed among junctions despite the different shapes and structures.

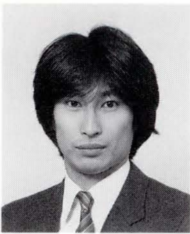
The inconsistencies in the production of high-quality junctions, as shown in Fig. 1, were observed frequently during the development of fabrication techniques. When junction electrodes were made by *rf* sputtering and the maximum V_m was no more than 25 mV, it was observed that the junction characteristics seemed to depend on the junction shapes. This time, junctions were fabricated by dc sputtering because this method yields superior junction characteristics. The quality parameter, V_m consistently exceeded 40 mV in all shapes and constructions tested. No influence of the shape or structure on the junction characteristics could be observed. Junctions fabricated by dc sputtering always have high-quality characteristics, regardless of their shapes and structures. The new fabrication technology is suitable for Josephson integrated circuits consisting of different shapes and structures. Deterioration in the junction characteristics seems to be prevented by increasing the value of V_m. However, the question of why the deterioration occurred and why dc sputtering suppresses this remain unanswered. Considering that the deterioration could be prevented by changing the method of film deposition, the internal film stresses or the conditions around the junction interface seem to cause the deterioration. Why this is the cause is not yet clear. The determination of the cause of the deterioration in characteristics is important to fabricating stable and reproducible high-quality Josephson junctions.

5. Acknowledgement

The present research effort is part of the National Research and Development Program on "Scientific Computing System", conducted under a program set by the Agency of Industrial Science and Technology, Ministry of Internatioal Trade and Industry.

References

1) Morohashi, S., and Hasuo, S.: Experimental Investigations and Analysis for High-Quality Nb/AlO$_x$/Nb Josephson Junctions. *J. Appl. Phys.*, **61**, 10, pp. 4835-4849 (1987).
2) Kotani, S., Fujimaki, N., Morohashi, S., Ohara, S., and Hasuo, S.,: Feasibility of an Ultra-High-Speed Josephson Multiplier. *IEEE J. Solid-State Circuits*, **SC-22**, 1, pp. 98-103 (1987).



Shiro Ohara
Device Development Dept.
FUJITSU LIMITED
Bachelor of Electrical Eng.
Waseda University 1982
Specializing in Josephson Integrated Circuits

Shinya Hasuo
Advanced Technology Division
FUJITSU LIMITED
Bachelor of Electronic Eng.
Tokyo Institute of Technology 1967
Dr. of Electronic Eng.
Tokyo Institude of Technology 1972
Specializing in Superconducting Electronics

Takeshi Imamura
Device Development Dept.
FUJITSU LIMITED
Bachelor of Applied Physics
Tokyo Institute of Technology 1973
Dr. of Electronic Eng.
The University of Tokyo 1987
Specializing in Josephson Integrated Circuit

UDC 621.382.3

Resonant-Tunneling Hot Electron Transistors (RHETs) Using an InGaAs/In(AlGa)As Heterostructure

● Kenichi Imamura ● Shun-ichi Muto ● Naoki Yokoyama

(Manuscript received September 4, 1987)

This paper describes the RHETs using an InGaAs/In (AlGa) As heterostructure. A common-emitter current gain of 10.0 was measured at 77 K for a 100 nm base RHET. The measured current gain of 25 nm base RHETs reaches as high as 25.0, which is about five times larger than that of GaAs/AlGaAs RHETs.

The collector current peak-to-valley ratio reached 15.0 for the same device.

1. Introduction

In 1985, we proposed and fabricated three-terminal resonant-tunneling devices, named the RHET (Resonant-tunneling Hot Electron Transistor), which used a GaAs/AlGaAs heterostructure[1]. We have demonstrated various attractive features of RHETs for memory and logic applications[2]. However, the common-emitter current gain in GaAs/AlGaAs RHETs was as low as 5 for a 25 nm base, which is not high enough for practical use.

The relatively low current gain in GaAs/AlGaAs RHETs is believed to be due to the narrow Γ-L valley separation energy of GaAs (0.30 eV). On the other hand, in InP and $In_{0.53}Ga_{0.47}As$, the Γ-L valley separation energy is considered to be about 0.55 eV.

In InAs, the Γ-L valley separation energy is considered to be about 0.78 eV. So, these materials should work well as the base layer for RHETs. Here, we tried to fabricate InGaAs base RHETs, which can be easily grown on InP substrate using MBE without lattice mismatch.

2. Experiment

Figure 1 is a schematic cross section of the InGaAs/In(AlGa)As RHET we fabricated. InGaAs and In(AlGa)As layers were grown on the InP semi-insulating substrate using pulsed molecular beam methods[3]. The device uses a quantum well which acts as the hot electron injector. This consists of a 3.81 nm $In_{0.53}Ga_{0.47}As$ layer sandwiched between two layers of 4.4 nm $In_{0.53}Al_{0.47}As$. We fabricated two types of RHETs with different base widths. The thicknesses of the base layers were 25 nm and 100 nm. The carrier concentration of the base layer was 1×10^{18} cm^{-3}. The thickness of the $In_{0.52}(Al_{0.5}Ga_{0.5})_{0.48}As$ collector barrier was 200 nm with a 6 nm compositional stepwise graded layer to reduce the quantum mechanical reflection at the collector barrier.

To investigate the effects of the graded layer, we also fabricated 25 nm base RHETs without the graded layer.

All the current-voltage characteristics were measured at 77 K.

3. Results

Figure 2 shows the common-emitter collector current-voltage characteristics measured at 77 K for a 100 nm base RHET. A common-emitter current gain of 10.0 was obtained at a collector-emitter voltage of 2 V and a base current of 1 mA. The current gain was about 15 times larger than that of the 100 nm base

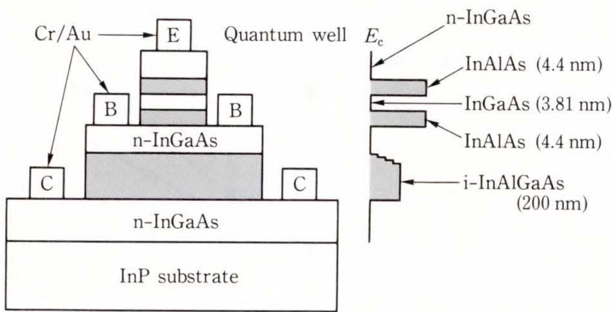


Fig. 1—Schematic cross section of InGaAs/In(AlGa)As RHET.

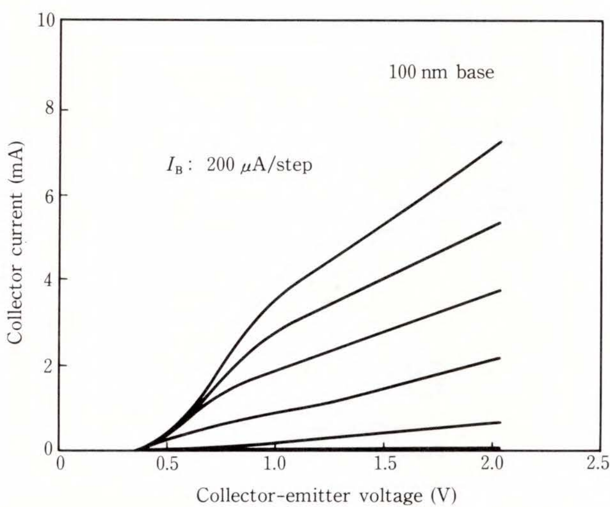


Fig. 2—Common-emitter collector current-voltage characteristics for the 100 nm-base RHET.

GaAs/AlGaAs RHETs[4]. Using thermionic emission current measurement, the conduction band discontinuity at the collector barrier was found to be 0.28 eV, which is about half that of the InAlAs ternary barrier (0.55 eV)[5]. The increased current gain of InGaAs/In(AlGa)As RHETs compared to that of GaAs/AlGaAs RHETs is believed to be due to the larger Γ-L valley separation energy of InGaAs compared to that of GaAs.

Figure 3 shows the common-emitter collector current-voltage characteristics for the 25 nm-base RHET with a graded layer. A common-emitter current gain of 18 was obtained at a collector-emitter voltage of 2 V, which is about twice that of the 100 nm-base RHET. Also a high current gain of 25 was obtained at a collector-emitter voltage of 3 V, which is about five times larger than that of GaAs/AlGaAs RHETs. Figure 4 shows the collector current as a function of the base-emitter voltage for the same device. The collector current exhibits a pronounced peak due to resonant-tunneling at a base-emitter voltage of 1 V. The peak collector current density is 2.8×10^4 A/cm^2 and the peak-to-valley ratio reached 15.0. The increased current

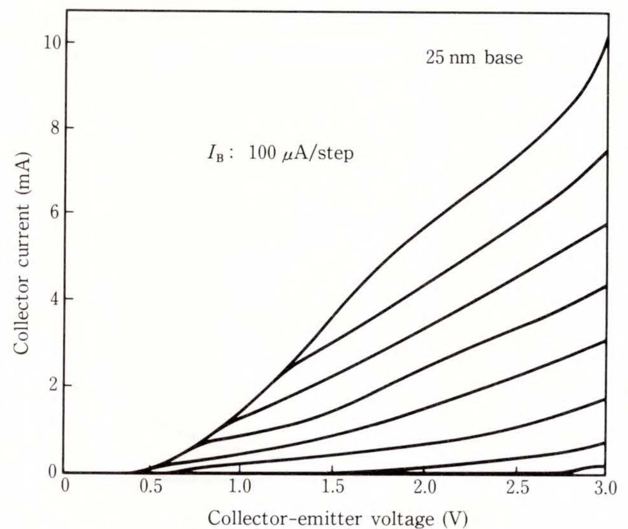


Fig. 3—Common-emitter collector current-voltage characteristics for the 25 nm-base RHET.

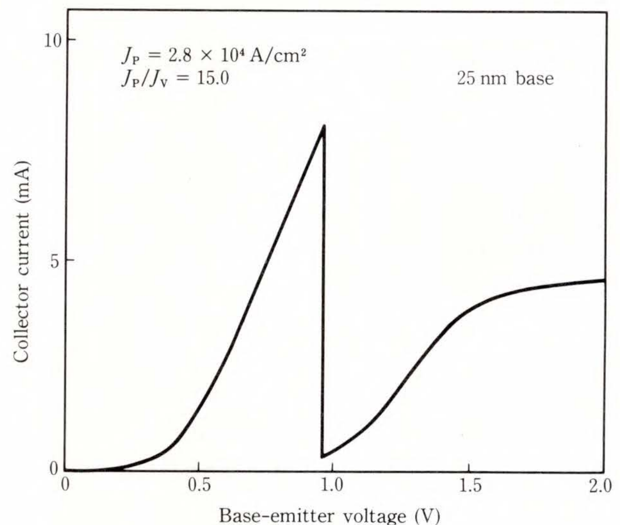


Fig. 4—Collector current as a function of base-emitter voltage for the 25 nm-base RHET.

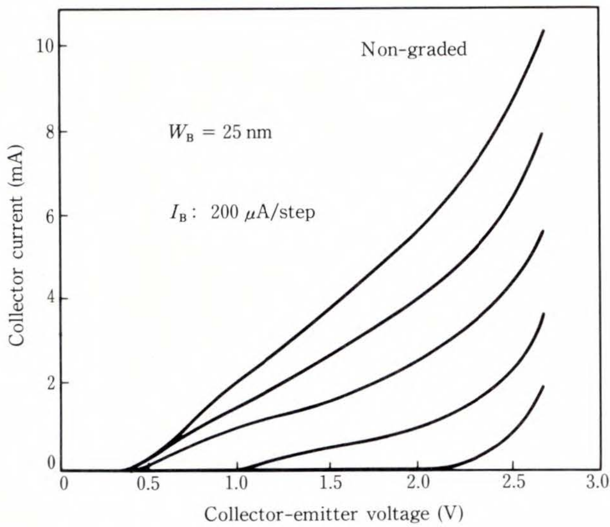


Fig. 5—Common-emitter collector current-voltage characteristics for the 25 nm-base RHET without graded layer.

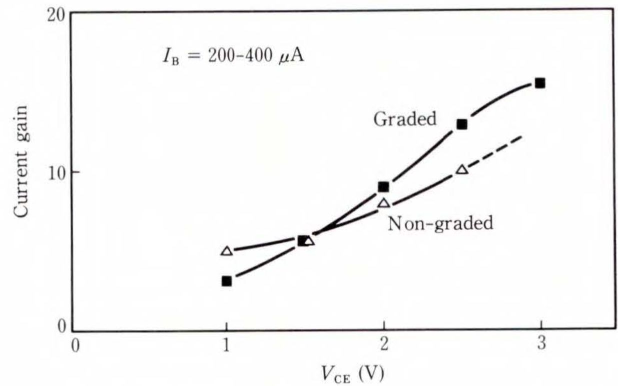


Fig. 6—Common-emitter current gain as a function of collector-emitter voltage for graded and non-graded RHET.

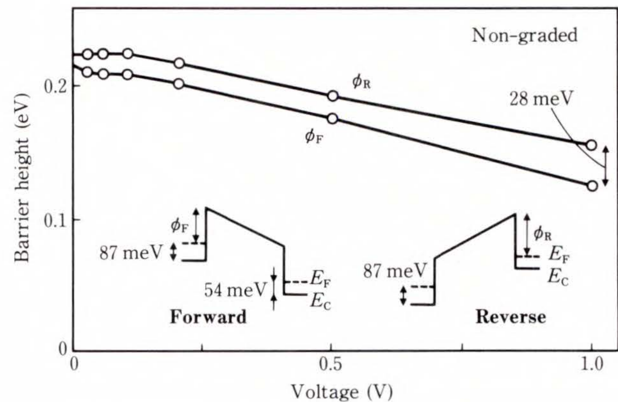


Fig. 7—Collector barrier height as a function of applied voltage for non-graded RHET.

density and peak-to-valley ratio is considered to be due to the large tunneling probability and high barrier height of the emitter InAlAs barrier compared to that of the AlGaAs barrier. The high current gain and peak-to-valley ratio should contribute to the stable operation of the circuits. The high peak current density should contribute to the high-speed operation of the RHET because it reduces the charging time for the emitter and collector capacitances.

Next, the effect on the current gain of the compositional graded layer in the collector barrier was investigated. We fabricated two kind of 25 nm-base RHETs. One has a 6 nm three stepwise compositional graded layer at the base and collector barrier interface (with a graded layer). The other has no graded layer (without a graded layer). Figure 5 shows the common-emitter collector current-voltage characteristics for the 25 nm-base RHETs without a graded layer. A common-emitter current gain of eight was obtained at a collector-emitter voltage of 2 V and a gain of ten was obtained at a collector-emitter voltage of 2.5 V, which is smaller than that of the devices with a graded layer.

Figure 6 shows the common-emitter current gain as a function of collector-emitter voltage for the graded and non-graded RHETs. Current gain was measured at the base current range from 200 μA to 400 μA. The current gain for the graded RHET is lower than that of the non-graded RHET at a lower V_{CE} than 1.5 V (V_{CB} of about 0.9 V).

Figure 7 shows the barrier height as a function of the applied voltage for non-graded collector barrier RHETs. The measured results are shown for the forward and reverse bias. The barrier height decreased as the applied voltage was increased. This figure clearly shows that there is a difference between the forward and reverse bias barrier heights and that the difference was almost constant between 0.01 V and 1 V. This difference can be explained as the difference of E_C-E_F (33 meV) due to the difference in the doping concentration for the

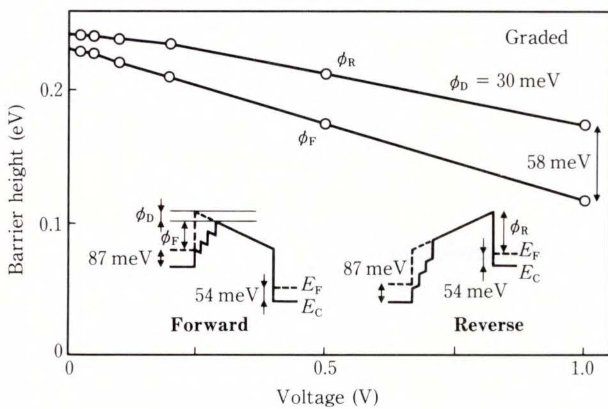


Fig. 8—Collector barrier height as a function of applied voltage for graded RHET.

base layer and collector layer ($N_D = 1 \times 10^{18}$ cm^{-3} for the base layer, $N_D = 5 \times 10^{17}$ cm^{-3} for the collector layer). Figure 8 shows the barrier height as a function of the applied voltage for the 25 nm base RHET with the graded layer. The difference in the barrier heights between the forward bias and the reverse bias is increased as the applied voltage is increased. This increase in the barrier height difference is believed to be caused by the lowering of the graded barrier with the applied voltage. Comparing the results for non-graded barrier RHETs, the barrier height lowering with an applied voltage of 1 V is calculated to be about 30 meV, which agrees well with the theoretically calculated barrier lowering. The initial barrier height lowering should contribute to the difference in the breakdown voltage in the common-emitter configuration. The characteristics shown in Fig. 6 can be well explained as the increase in current gain caused by the barrier height lowering at positive collector-base voltage for the graded RHETs. The effect of quantum mechanical reflections on the current gain was not clearly determined in this investigation.

Next, we estimated the device transit time for the InGaAs/In(AlGa)As RHETs. The total RHET delay time can be represented using Equation (1).

$$\tau_{EC} = \tau_E + \tau_B + \tau_C + \tau_{CC}. \qquad \ldots\ldots (1)$$

Here, τ_E is the emitter depletion-layer charging time, τ_B is the base transit time, τ_C is the collector depletion layer transit time, and τ_{CC} is the collector charging time. It is interesting to compare the total delay time of RHETs with that of GaAs/AlGaAs HBT (Heterojunction Bipolar Transistors). The same device size was assumed for RHETs and HBTs. The emitter/base junction was 2 μm x 5 μm and the base/collector junction was 4 μm x 7 μm. Cheng et al. reported a high cut-off frequency of 67 GHz. The individual delay time was 0.6 ps for τ_E, 0.6 ps for τ_B (80 nm base), 0.7 ps for τ_C, and 0.5 ps for τ_{CC}. The total delay time was 2.4 ps[6)].

For the InGaAs/In(AlGa)As RHETs, the base width was assumed to be 100 nm as was the thickness of the collector barrier.

The base transit time was calculated assuming ballistic transport in the base layer. The injection energy was supposed to be 0.6 eV, so the electron velocity was calculated to be about 1.24×10^8 cm/s. The base transit time was calculated to be about 0.08 ps[7)].

The collector depletion layer transit time was calculated to be about 0.01 ps, considering the velocity change at the base/collector barrier interface. The collector charging time was calculated to be about 0.06 ps. Then the total delay time, excluding the emitter depletion layer charging time, was calculated to be about 0.24 ps, which is about 1/7 less than that of the HBT. These results indicate the advantage of hot electron transistors for collecting only ballistic or near ballistic electrons.

The emitter delay time is related to the built in time, t_1, to achieve the resonant-tunneling state and the time, t_2, for electrons passing through the resonant-tunneling barrier. The time t_2 was calculated using $\hbar/\Delta E_0$. ΔE_0 was calculated to be 0.51 meV for the barrier used in this paper, so t_2 was calculated to be 1.3 ps. The time t_1 was considered to be the charging time for an ordinary parallel plate. The thickness of the capacitance was considered to be roughly the sum of the thicknesses of the barrier and the depletion layer. The conductance was calculated to be about the peak current divided by half of the peak voltage, which agrees well with the

measured results. The calculated value of t_1 was 2.4 ps. The simple sum of t_1 and t_2 was as large as 3.7 ps. Actually, the total emitter delay time was considered to be smaller than $t_1 + t_2$. If we assume that the total emitter delay time is the larger of t_1 and t_2, the emitter delay time can be calculated to be about 2.4 ps. In this case, the total delay time was calculated to be 2.64 ps which is almost equal to that of HBTs.

If the thickness of the emitter barrier was decreased, E_0 increased and t_2 decreased. A t_1 also decreased because of the increased current density and therefore increased conductance. The decreased barrier thickness (3.52 nm) and barrier width (3.22 nm) cause a peak current density of 4×10^5 A/cm^2 and the delay time t_1 will be about 0.62 ps (t_2 = 0.21 ps). In this case, the total delay time was calculated to be about 1 ps and the high cut-off frequency of 160 GHz will be achieved. These calculations indicate that InGaAs/In(AlGa)As RHETs have great potential as high-speed devices.

4. Conclusion

We have fabricated InGaAs/In(AlGa)As RHETs. A common-emitter current gain of 10 was obtained for the 100 nm base RHETs, which is about 15 times larger than that of GaAs/AlGaAs RHETs. A high current gain of 25 was obtained for the 25 nm base RHETs with a compositional stepwise graded layer. The collector current peak-to-valley ratio reached 15. The increased current gain compared to that of GaAs/AlGaAs RHETs is considered to be caused by the large Γ-L valley separation energy of InGaAs compared to that of GaAs. Compared to the current gain for the 25 nm base RHETs without a graded layer, the barrier height lowering at the applied voltage contributed to the increase in current gain.

The device delay time was estimated for RHETs compared to that of GaAs/AlGaAs HBTs. The estimated delay time was about 1 ps with a cut-off frequency of 160 GHz. We found that the RHETs have great potential as high-speed devices.

The obtained high current gain and high peak-to-valley ratio indicated that the InGaAs/In(AlGa)As RHETs are very promising for high speed and stable operation of future integrated circuits.

5. Acknowledgement

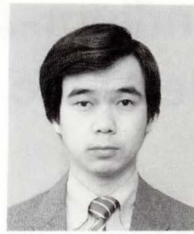
The authors thanks Messrs. A. Shibatomi, S. Hiyamizu, K. Kobayashi, and T. Misugi for their encouragement. They also thank Messers. T. Fujii, H. Ohnishi, T. Mori, Y. Nakata, Y. Sugiyama, K. Ishii for their support and discussion.

This work was performed under the management of the R and D Association for Future Electron Devices as a part of the Future Industries Project sponsored by the Agency of Industrial Science and Technology, MITI.

References

1) Yokoyama, N., Imamura, K., Ohshima, T., Nishi, N., Muto, S., and Hiyamizu, S.: Characterization of double heterojunction GaAs/AlGaAs hot electron transistors. IEDM technical digest, p. 532 (1984).
2) Yokoyama, N., Imamura, K., Muto, S., Hiyamizu, S., and Nishi, H.: A new functional resonant-tunneling hot electron transistor (RHET). *Jpn. J. Appl. Phys.*, **24**, p. L254 (1986).
3) Fujii, T., Nakata, Y., Sugiyama, Y., and Hiyamizu, S.: MBE growth of InAlGaAs lattice-matched to InP by pulsed molecular beam method. *Jpn. J. Appl. Phys.*, **25**, p. 1254 (1986).
4) Sugiyama, Y., Inata, T., Fujii, T., Nakata, Y., Muto, S., and Hiyamizu, S.: Conduction band edge discontinuity of InGaAs/In(AlGa)As heterostructure. *Jpn. J. Appl. Phys.*, **25**, p. L648 (1986).
5) Mori, T., Ohnishi, H., Imamura, K., Muto, S., and Yokoyama, N.: Resonant tunneling hot-electron transistor with current gain of 5. *Appl. Phys. Lett.*, **49**, p. 1779 (1986).
6) Self-aligned AlGaAs/GaAs heterojunction bipolar transistors with improved high-speed performance. DRC abstract, (1978).
7) Characterization of GaAs/AlGaAs hot electron transistors using magnetic field effects on launched-electron transport. *Surface Sci.*, **174**, p. 481 (1986).

Kenichi Imamura
Exploratory Devices Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Applied Chemical Eng.
Keio University
Master of Physical Chemistry Eng.
Keio University
Specializing in Heterostructure Devices



Naoki Yokoyama
Exploratory Devices Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
Osaka University 1971
Dr. of Electrical Eng.
Osaka University 1984
Specializing in Heterostructure Devices

Shun-ichi Muto
Semiconductor Crystal Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
Tokyo Institute of Technology
Dr. of Physics
Tokyo Institute of Technology
Specializing in Superlattice Physics

UDC 669.7.018:669.872:681.327.634

Selenium-Indium-Antimony Alloy Film for Erasable Optical Disks

● Naga-aki Koshino ● Ken-ichi Utsumi ● Yasuyuki Goto

(Manuscript received April 10, 1987)

A new type of crystal phase change was found in selenium-indium-antimony alloy film, which is a unique material developed by Fujitsu. The phase change occurs reversibly between two different crystal states when the film is irradiated with laser pulses, resulting in different optical reflectivities. An erasable optical disk using this film has good carrier-to-noise-ratio of 45 dB, moderate recording sensitivity and long media life, but erasable cycles still vary from between 10^3 and 10^6 times at present. Though there remain some problems to be solved, the disk is nearly ready for practical use, and as such is very promising. A proposed model of the phase-change mechanism is also discussed.

1. Introduction

Optical disks have a high potential for large-volume information storage. Using a diffraction-limited focused laser beam (1 μm), it easily enables a storage density of 10^7-10^8 bit/cm^2, which is ten times larger than the storage density of the most advanced hard magnetic disks. Different from magnetic disks, optical disks store information on the thin film which is sandwiched between two relatively thick (1.2 mm) substrates. Because writing and reading are performed through the transparent substrate, dust on the substrate surface causes very few disk errors. Thus, optical disks can serve as easy-to-handle and high-performance file memories for a large computer system as well as for a small computer used in an office or at home.

Read-only type disks, from which only prerecorded information can be read, are already available in the form of video disks or CDs. As for write-once disks, on which information can be recorded by users, Fujitsu has already been shipping write-once disk drives (FACOM 6441 and 6442). Fujitsu has also been developing erasable optical disks, focusing on two types of erasable disks: magneto-optical disks and phase-change disks. This paper is about phase-change erasable disks.

The key to developing a phase-change erasable disk is to find a new material. The authors found out that a selenium-indium-antimony (SeInSb) alloy film shows reversible phase change when irradiated with laser light, and is thus promising as a recording material for erasable disks[1]. The authors also have found that an alloy film of indium and antimony (InSb) has the same erasability[2]. This paper describes the results of experiments evaluating the characteristics of the SeInSb disk and analyzing light-irradiated micro-bits of the alloy films. The paper also describes a proposed model of the phase-change mechanism of the alloy films[3)-4)].

2. Principle of writing/erasing method on phase-change media

The writing procedure on phase-change media changes the crystal state of the thin film by irradiating it with a laser beam. Phase transition is caused by the thermal effect of the irradiated laser light. No other external energy, such as an external magnetic field is used. Generally, two different irradiating conditions

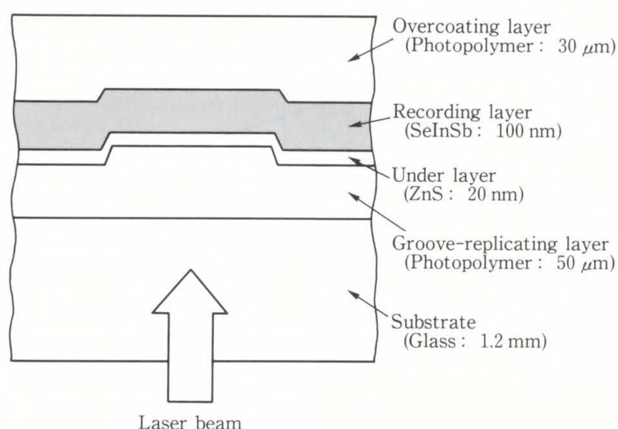


Fig. 1–Cross section of SeInSb medium.

Fig. 2–Photograph of SeInSb optical disk.

are used. One is to irradiate at high power for a short duration. The film changes to a liquid state in a short time and then solidifies quickly. The cooling rate is so high that the atoms cannot obtain energetically stable positions. The other condition is to irradiate at lower power for longer duration. In this case, cooling rate of melted film is so low that the atoms have time to recover energetically stable positions.

Tellurium alloy films have been developed as erasable phase-change materials[5)-8)]. In these films, information was recorded using a phase change between an amorphous state and a crystal state of the films. The newly developed alloy films, SeInSb and InSb films, have a different phase-change mechanism. They use a phase change between two different crystal states. Information is recorded via a transition between these two stable crystal states. Laser-irradiating conditions are nearly the same as with the tellurium alloy films.

3. Experiment

3.1 Fabrication of alloy films

SeInSb and InSb alloy films were deposited on a substrate by a vacuum evaporation technique. Vapor sources of selenium, indium and antimony were heated independently. The three vapors were mixed together on rotating substrate and formed a thin alloy film. The evaporation rate of each vapor was controlled independently, using a quartz-crystal oscillator to strictly control the film composition.

A cross section of the medium is shown in Fig. 1. The substrate is glass which is coated with UV polymerized photopolymer (2P)[9)]. Guiding grooves for the tracking servo-mechanism were formed on the surface of the 2P layer. Between the 2P layer and the recording layer (alloy film), zinc sulfide was deposited to reduce the reflectivity of the medium and this increased writing sensitivity. The composition of the alloy film was 20 at% selenium, 35 at% indium and 45 at% antimony for SeInSb, and 40 at% indium and 60 at% antimony for InSb. Finally, the top surface was coated with another 2P layer. The purpose of the overcoating was not to protect the alloy films from oxidation, but rather to prevent hole formation in the alloy film when it is irradiated with laser light. Figure 2 is a photograph of a 130-mm-diameter SeInSb optical disk.

3.2 Evaluation of write/erase characteristics

3.2.1 Static recording

SeInSb medium was irradiated with a focused light beam from a laser diode. In this case, the medium was not rotated. The diameter of the beam was 1 μm. The wavelength was 830 nm. The position of the objective lense which focused the beam was automatically

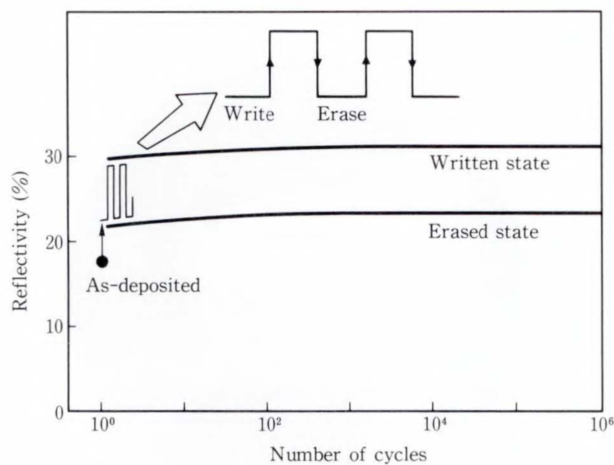


Fig. 3—Reflectivity change by cyclic irradiation.

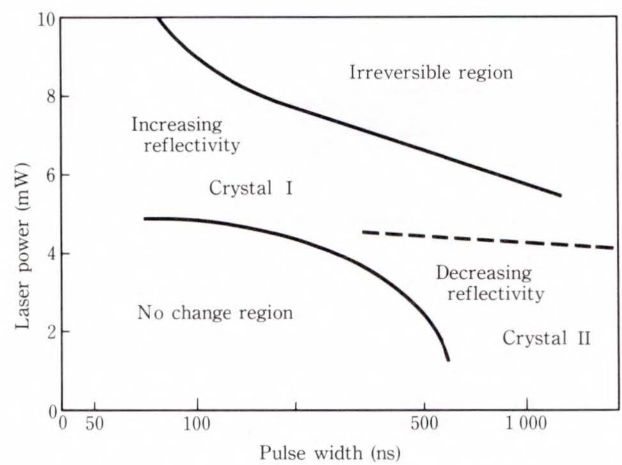


Fig. 4—Irradiating conditions for changing reflectivity.

controlled to focus the beam on the film in such a way that the beam diameter reached its minimum at the surface of the alloy film.

Between irradiations for writing and erasing, a weak continuous power was applied and reflected light from the small area of the film was detected, which gave the reflectivity of the small irradiated area of the medium. As the laser light was irradiated repeatedly on the medium, the reflectivity changed as shown in Fig. 3. At first, as-deposited alloy film had low reflectivity. Then, after an erasing beam was applied, the medium reached a new state of slightly higher reflectivity. The next irradiation with a writing beam moved the medium to a still-higher state of reflectivity. The next irradiation with an erasing beam returned the medium to the low-reflectivity state. After that, cyclic writing and erasing changed the reflectivity reversibly between the high and low states, respectively. The reversible change of reflectivity was confirmed beyond 10^6 cycles.

Figure 4 shows irradiating conditions necessary to make transitions between the two crystal states. An irradiation at high power for a short duration causes the transition from crystal state II to I. On the other hand, an irradiation at low power for a longer duration causes the reverse transition from I to II.

3.2.2 Dynamic recording

The irradiating method for a rotating disk is somewhat different from that for a stationary disk. Since each point of a rotating disk is irradiated only for the time during which the point traverses the laser beam, the rotation speed and the beam diameter determine the duration of irradiation. As was described in subsection 3.2.1, the duration of an erasing pulse is longer than that of a writing pulse. So, in this experiment, in order to obtain two different pulse durations at a fixed rotation speed, a two-beam optical head was used. One of the beams was a conventional circular-shaped beam for writing and reading. The other beam was an elliptical-shaped beam for erasing. The long axis of the erasing beam was aligned in the direction of the data tracks. The diameter of the circular-shaped beam was 0.9 μm. The elliptical-shaped beam was 3 μm long and 1 μm in width (i.e. perpendicular to the track direction).

The rotation speed was fixed at 900 rpm. The linear velocity was about 6 m/s. A signal of 3 MHz was recorded. The necessary peak power of the modulated laser pulse was 8 mW on the medium surface. The bit length was 1 μm. The same signal was erased by irradiating an elliptical-shaped DC laser beam on the recorded track. The necessary power was 12 mW. Overwriting was also performed. The prerecorded signal was erased by the elliptical beam and just after that a new signal was recorded on the same track during the same revolution of the disk.

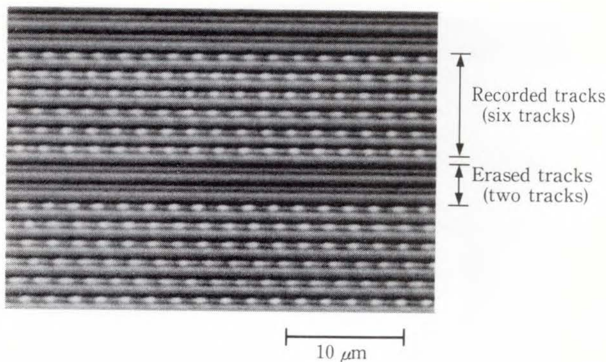
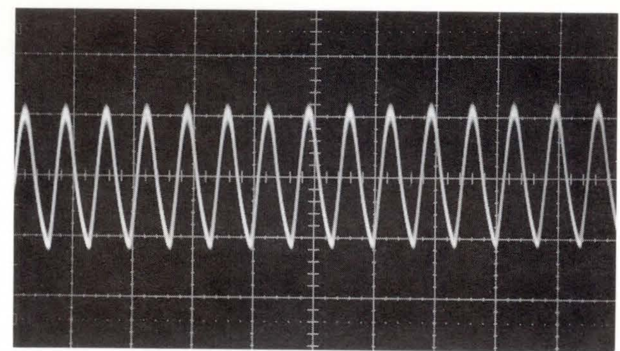
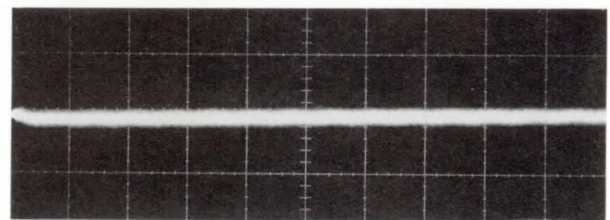


Fig. 5–Photomicrograph of recorded and erased tracks.



a) Recorded track



b) Erased track

Fig. 6–Wave forms of read-out signals.

Considering the linear velocity of 6 m/s, laser power and pulse duration are 8 mW, 170 ns for writing and 3 mW, 500 ns for erasing. These conditions coincide well with the data for static recording in Fig. 4.

A photomicrograph of the recorded tracks and erased tracks is shown in Fig. 5. Bits of 1 μm can be seen clearly. The carrier-to-noise ratio was measured to be 45 dB. The measuring bandwidth was 30 kHz. After erasure, the 3 MHz prerecorded signal was reduced to –40 dB of the initial level. Figure 6 shows the waveforms of the read-out signals of the recorded track and the erased track.

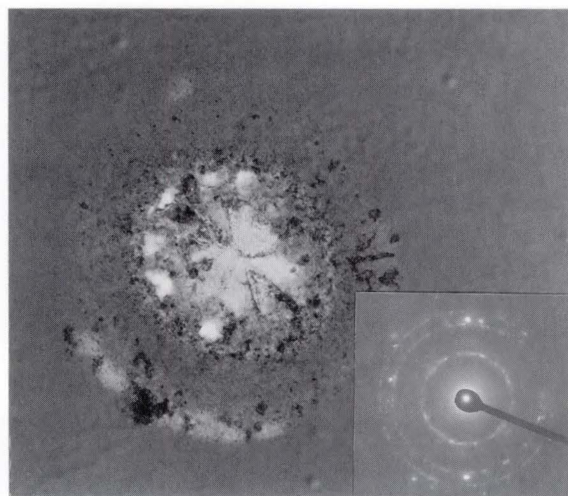
An experiment of cyclic writing and erasing was performed. It was observed that the carrier-to-noise ratio reduced gradually as number of cycles increased. The maximum number of erasable cycles varied between 10^3 and 10^6 cycles from disk to disk. As was determined by static recording, the reversible change of reflectivily exceeded 10^6 times. In this experiment, the amplitude of the read-out signal, which corresponds to the reflectivity change in the static experiment, did not change as the number of cycles increased. But the noise, which could not be observed in the static experiment, increased with the number of cycles.

3.3 Analysis of bits

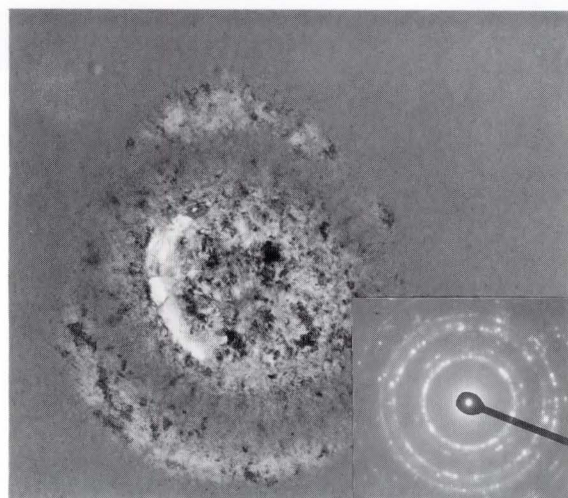
3.3.1 Static recording

The irradiated SeInSb alloy film was peeled off from the substrate and observed by transmission electron microscope. The bright field images of the irradiated area (bits) and diffraction patterns are shown in Fig. 7. The outside of the irradiated area, which remained in the as-deposited state, was amorphous. From the diffraction patterns it was clear that the irradiated bits were an assembly of small crystals. The difference between a written bit and an erased bit was not so definite, but the grain size of the central part of a written bit seemed to be bigger than that of an erased bit. Accordingly, the reflectivity change shown in Fig. 3 can be explained in terms of the crystalline states; the initial low-reflectivity state corresponds to the amorphous state. Laser light irradiation changed the film from the amorphous state to the crystal state. The film had two kinds of stable crystal states with different reflectivities. The cyclic irradiations with writing and erasing pulses caused reversible transitions between these crystal states, resulting in reversible reflectivity changes.

The InSb alloy medium showed the same reflectivity change and had same reversible cycles as the SeInSb medium shown in Fig. 3. Figures 8-a) and 9-a) are transmission electron

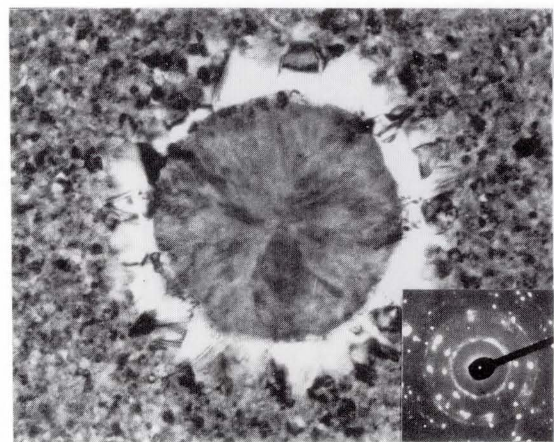


a) Written bit

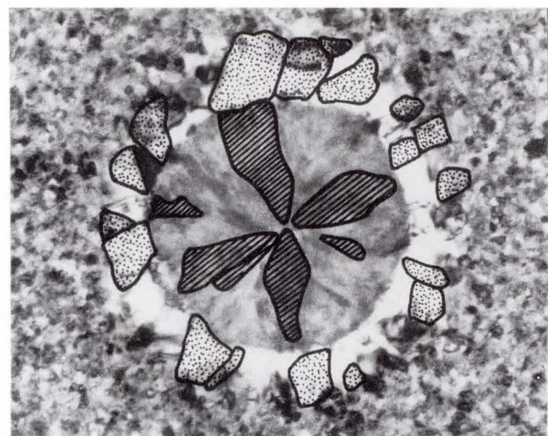


b) Erase bit

Fig. 7—Transmission electron micrographs of SeInSb film.



a) Bright field image

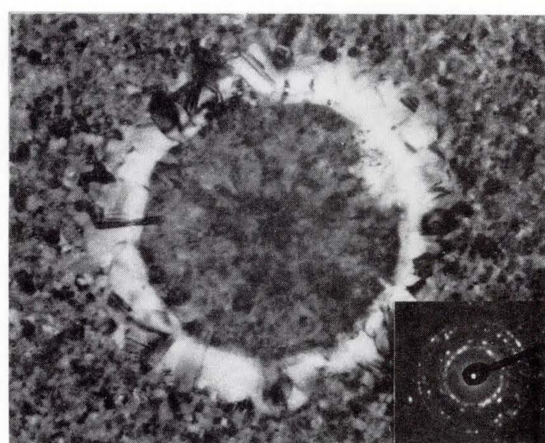


b) Results of dark field observation

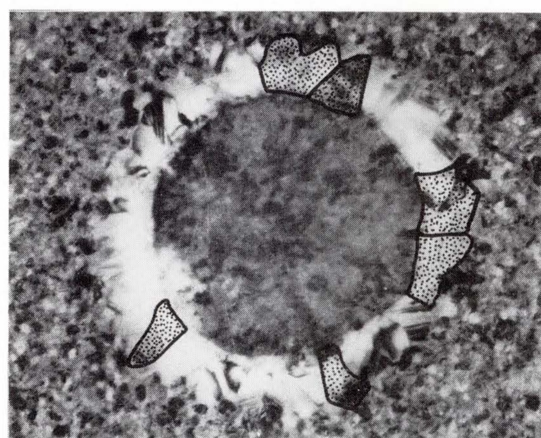
Fig. 8—Transmission electron micrographs of the written bit on InSb film. Identified crystals are marked b).

micrographs of InSb alloy film. In this case, the alloy film was crystallized during vacuum evaporation, so the as-deposited initial state is crystalline. It must be noted that the diameters of irradiated areas in Figs. 8-a) and 9-a) are smaller than in Fig. 7. The reason is thought to be that in an amorphous film, heat diffusion from the irradiated area to the outer amorphous area results in an excess crystallized area around the bit. The crystallization during deposition can be also applied to SeInSb film, which is effective to obtain small bits.

Compared with SeInSb, InSb is simpler. For this reason, the authors conducted a thorough investigation of InSb film in order to clarify the mechanism of the crystal phase change. From the analysis of the diffraction patterns, the outside of the irradiated bit of the InSb film was shown to be a mixture of very small Sb crystals and small InSb compound crystals. In this film a clear difference exists between written bits and erased bits. In the written bit {Fig. 8-a)}, the center part is composed of large crystals. In the erased bit {Fig. 9-a)}, however, the center part is an assembly of small crystals. At the outer part of the both bits, crystals of intermediate size



a) Bright field image



b) Results of dark field observation

0.5 μm

Fig. 9–Transmission electron micrographs of the erased bit on InSb film. Identified crystals are marked b).
Small crystals of Sb and InSb in the central part are not marked deliberately because of complications.

are observed. From the analysis of the diffraction patterns, it was concluded that all the crystals in both bits were composed of cubic $In_{50}Sb_{50}$ intermetallic compound (zinc blend structure) and rhombohedral Sb. No other crystal structure was found.

Dark field observation was also performed. Images of Sb crystals were formed by electrons that were diffracted from one of Sb crystal planes. In the same manner, images of InSb compound crystals were formed. Figure 8-b) is the combined result of the dark field observations. It suggests that the large crystals at the center part were Sb crystals, and the crystals of intermediate size at the outer part of the bit were InSb compound crystals.

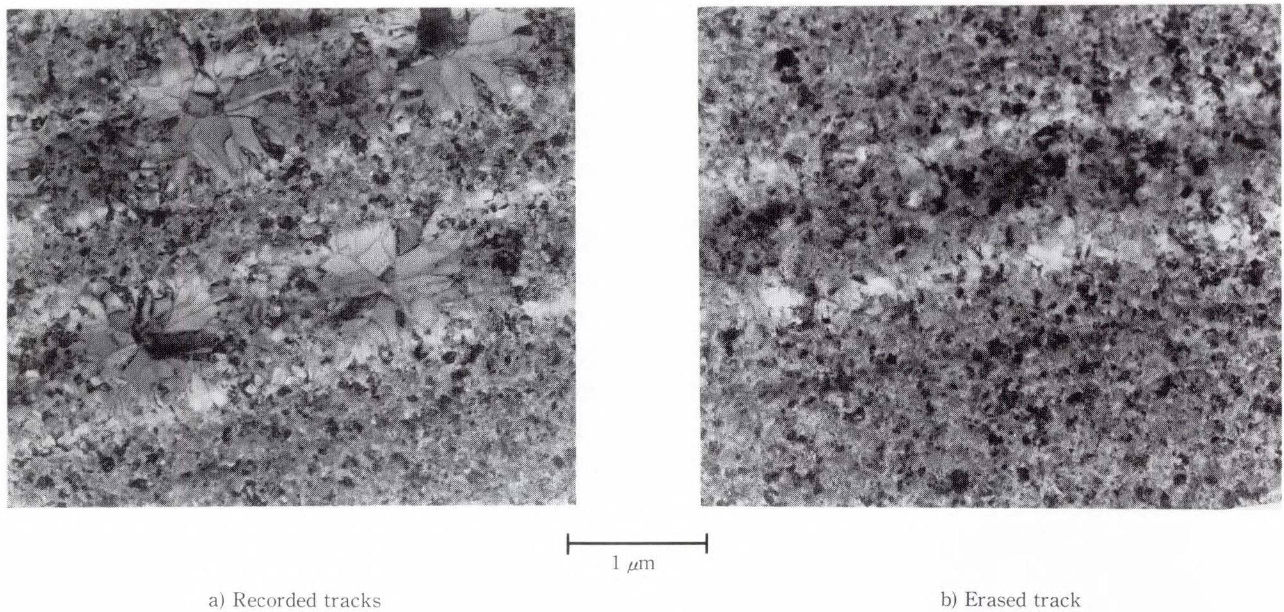
As for the erased bit, there were no large Sb crystals, but Sb and InSb compound crystals were distributed uniformly in the center part of the bit {Fig. 9-a)}, which was determined from the analysis of the diffraction patterns. Result of the dark field observations is shown in Fig. 9-b), in which only InSb crystals at the outer part of the bit are marked. (Small crystals of Sb and InSb compound in the central part are not illustrated because of complications.)

From the Figs. 8 and 9, it was concluded that the reversible change in InSb alloy film is related to the distribution of Sb crystals and InSb compound crystals. In SeInSb alloy film, too, the origin of the reversible change is thought to be change in distribution of Sb crystals and InSb compound crystals. The reflectivity change can also be explained in terms of the distribution change. It is known that reflectivity of bulk Sb and InSb compound is 70 percent and 40 percent, respectively. So, the written bit, which has large Sb crystals at the central part of the bit, shows high reflectivity. On the other hand, the erased bit, in which the central part is relatively poor in Sb crystals, has low reflectivity.

3.3.2 Dynamic recording

The recording layer was peeled off from the substrate and observed by transmission electron microscope. Figure 10-a) is the transmission electron micrograph of a bit array recorded on a rotating disk. This alloy film was also crystallized during vacuum deposition, because in the crystallized film smaller bits can be recorded than in amorphous film, as described in Subsec. 3.3.1. The recorded bits were located along the groove. At the center of each bit, large crystals appeared, which are assumed to be Sb crystals.

Figure 10-b) is the transmission electron micrograph of an erased track. The bit patterns in Fig. 10-a) have disappeared completely, but along both sides of the track erased by



a) Recorded tracks b) Erased track

Fig. 10—Transmission electron micrograph of recorded tracks and erased track on a rotating disk.

beam-irradiation, crystals of intermediate size were present, and are assumed to be InSb compound crystals. These photographs are basically the same as the transmission electron micrographs of static recording bits shown in Figs. 8 and 9.

3.4 Estimation of media life

SeInSb and InSb alloy films are chemically stable. The oxidization of both films has been monitored by measuring the optical transmission. It has been confirmed that optical transmittances of both films do not change for more than four months under the atomospheric condition of 60 °C and 90 percent relative humidity[1)]

In this experiment, stability of the crystal phases was examined. That is, the storage time of recorded information was examined by an accelerated aging test. If the written state (crystal I) is energetically unstable, information will be lost during an aging test. The media life was evaluated in terms of carrier-to-noise ratio. A signal of 3 MHz was recorded on both SeInSb and InSb disks and their carrier-to-noise ratio were measured after certain intervals under the testing atmosphere. The humidity was fixed at 90 percent relative humidity,

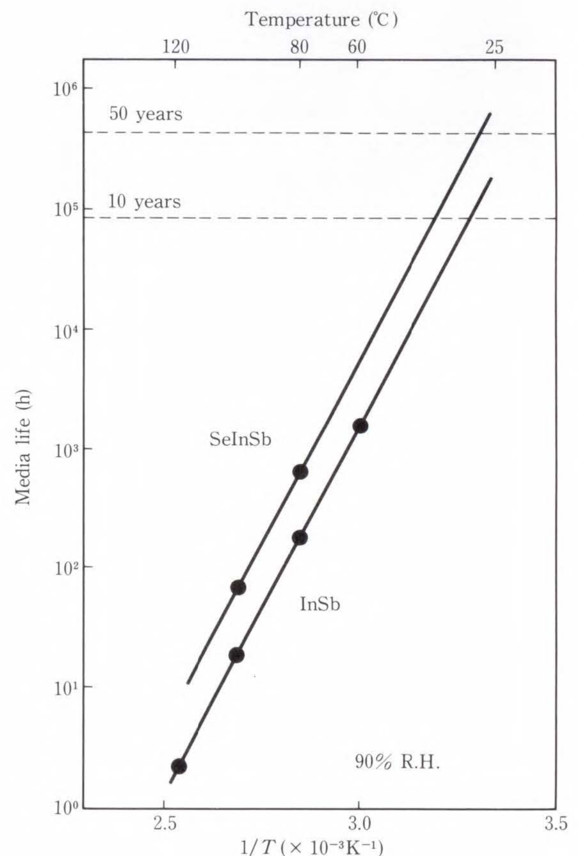


Fig. 11—Arrhenius plot of media life vs. aging temperature. Life was defined end at the time that C/N was reduced by 3 dB from the initial value.

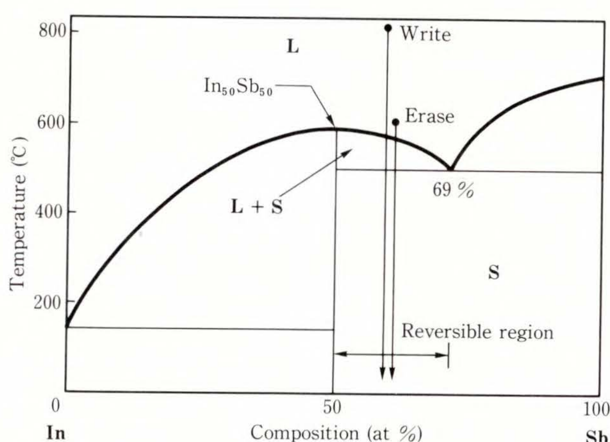


Fig. 12—Equilibrium phase diagram of InSb alloy system. Arrows show cooling processes of the written bit and erased bit.

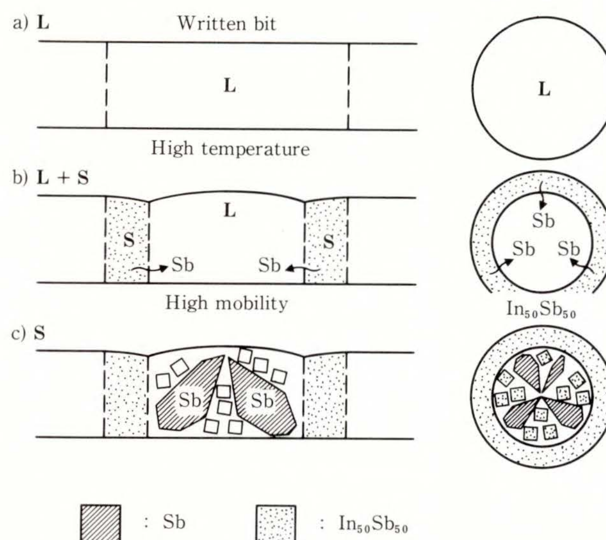


Fig. 13—Model of the process as a written bit cools.

and only temperature was varied (60-120 °C).

Life of the disk under the measured conditions was defined to be the time interval between the initial time and the specific time at which the carrier-to-noise ratio dropped to −3 dB from the initial value. The results are shown in Fig. 11. Extrapolation from the high-temperature region to 25 °C suggests that the life of an InSb disk is about 20 years at 25 °C and 90 percent relative humidity. The life of an SeInSb disk is estimated to be more than 50 years. Selenium itself does not have a direct influence on the mechanism of the reversible phase change, but it has an important role in increasing chemical stability and data storage time.

4. Discussion

4.1 Model of phase-change mechanism

Based on the observations described in section 3.3 and using equilibrium phase diagram, the authors have proposed a model of the writing and erasing mechanisms in SeInSb and InSb alloy films[4]. Figure 12 is the phase diagram of the In-Sb binary alloy system. The authors have found that the alloy film in which Sb concentration is between 50 at% and 69 at% shows a reversible reflectivity change when irradiated with a laser pulse. This is shown in Fig. 12.

The model is described in the following subsections.

4.1.1 Writing

While the film is irradiated at high power for a short duration, the temperature of the film increases rapidly, and when its temperature reaches the melting point, the film becomes liquid. After the irradiation of laser light, the temperature goes down. When it reaches the liquidus line in Fig. 12, InSb-compound crystals solidify at the outer part of the irradiated area as initial crystals. At this time, excess Sb atoms move from the outer part to the inner part of the irradiated area, as shown in Fig. 13. The central part still remains liquid, and the average composition goes towards the eutectic point along the liquidus line.

During this cooling process, Sb atoms preserve high mobility and move quickly towards the center of the bit. So, after the temperature comes down to the melting point, Sb atoms form large crystals.

4.1.2 Erasing

Because the laser power used in the erasing process is low, the maximum temperature of the liquid film is lower than for the writing process. During cooling, InSb compound crystals first solidify at the outer part of the bits. The mobility of the Sb atoms, however, is low because of the low temperature of the bits' center. Thus, Sb and InSb compound crystals

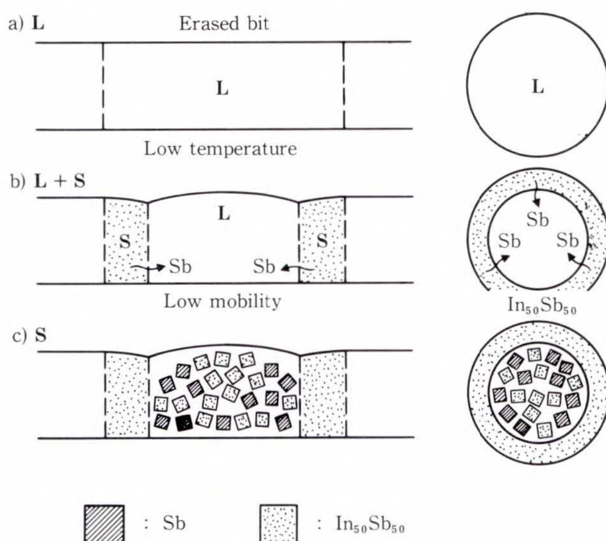


Fig. 14—Model of the process as an erased bit cools.

solidify as small crystals uniformly distributed in the bit.

This process is illustrated in Fig. 14.

The authors have examined the atomic distribution in the bits by EPMA (electron probe micro analysis). The results were in good accordance with this model[4].

The same model is applicable to the writing and erasing process on a rotating disk. A detailed analysis is now under way.

4.2 Media characteristics

The recording sensitivity of SeInSb and InSb disks is fairly good. Considering that currently available laser power is about 10 mW at the medium surface, maximum recordable velocity of the medium is 9 m/s, which closely corresponds to the linear velocity of a 130-mm-diameter disk rotating at 1 200 rpm. This is almost satisfactory for application to low-transfer-rate (less than 1 Mbyte/s) devices. But it is not sufficient for a high-transfer-rate application of optical disks. For example, recent magnetic disks have attained data transfer rates of 3 Mbyte/s, with a linear velocity of more than 20 m/s. Thus, an approximately two-fold increase in recording sensitivity is necessary. This is a problem that must be solved in the future. Development of a high-power laser would be another way to resolve the problem.

Carrier-to-noise ratio of 45 dB is sufficient for digital recording. Stable write/erase cycle is on the order of 10^3. The limitation is determined by the noise. The noise is thought to originate from grain growth of InSb-compound crystals located on both sides of the recorded tracks. Addition of other elements to the SeInSb alloy, or improvement of film fabrication method, will resolve the problem.

Long archaival life is an indispensable condition of optical memory media. The good chemical stability is a merit of SeInSb and InSb alloy films. Medium life is over ten years in normal environments, even if conservatively estimated. This is sufficient for practical use. It also proves that the mechanism of crystal-to-crystal phase change is stable. But this estimation was in terms of degradation of carrier-to-noise ratio. The authors are now studying stricter estimation by bit-error tests.

5. Conclusion

A new types of crystal phase change in SeInSb alloy film was found, and an erasable optical disk has been developed using this film. Investigations by transmission electron microscope strongly suggest that the phase change originates from a change in the distribution of crystals, and that this change is caused by irradiated laser light.

The medium characteristics can be summarized as follows.

1) Carrier-to-noise ratio is 45 dB.
2) Recording sensitivity is moderate; recordable medium velocity is 9 m/s using 10 mW laser power at the medium surface.
3) Disk life is over ten years.
4) Erasable cycles vary between 10^3 and 10^6, at present.

Though there remain some problems to be resolved, the disk has attained almost-practical characteristics. The SeInSb disk is very promising. The authors hope to contribute the progress of technology through the development of the SeInSb disk.

6. Acknowledgement

The authors would like to thank Dr. Kuro-

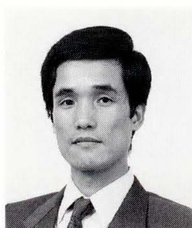
kawa, Dr. Henmi and Dr. Ogawa for their discussions and encouragement during the course of this work.

References

1) Koshino, N., Maeda, M., Goto, Y., Itoh, K., and Ogawa, S.: Selenium alloy film for new erasable optical disk media. *Proc. SPIE*, **529**, pp. 40-44 (1985).
2) Maeda, M., Utsumi, K., Goto, Y., Koshino, N., and Itoh, K.: Characteristics of InSb erasable optical disk medium. (in Japanese), Proc. 32th Spring Meet. of Jpn. Soc. Appl. Phys. and Related Socs. 1p-p-2 (1985).
3) Goto, Y., Maeda, M., Utsumi, K., Koshino, N., and Itoh, K.: Writing and erasing mechanism of InSb erasable optical disk medium. (in Japanese), Proc. 32th Spring Meet. of Jpn. Soc. Appl. Phys. and Related Soc., 1p-p-3 (1985).
4) Goto, Y., Utsumi, K., Ushioda, A., Tsugawa, I., and Koshino, N.: Write-erase mechanism of indium-antimony optical disk medium. *Proc. MRS*, **74**, pp. 251-256 (1986).
5) Feinleib, J., Neuville, J. de, Moss, S.C., and Ovshinsky, S.: Rapid reversible light-inducted-induced crystallization of amorphous semi-conductors. *Appl. Phys. Lett.*, **18**, 6, pp. 254-257 (1971).
6) von Gutfeld, F.J., and Chandrai, P.: Laser writing and erasing on chalcogenide film. *J. Appl. Phys.*, **43**, 11, pp. 4688-4694 (1972).
7) Bell, A.E., and Spong, F.W.: Reversible optical recording in trilayer structures. *Appl. Phys. Lett.*, **38**, 11, pp. 920-921 (1983).
8) Takenage, M., Yamada, N., Ohhara, S., Nishiuchi, K., Nagashima, M., Shirata, T., and Yamashita, T.: New optical erasable medium using suboxide thin film. *Proc. SPIE*, **420**, pp. 173-177 (1984).
9) Huijser, A., Jacobs, B., and Vriens, L.: Aging characteristics of DOR media. Tech. Digest Optical Data String Storage. pp. THB1-1-THB1-4 (1983).

Naga-aki Koshino
Functional Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
The University of Tokyo 1970
Master of Physics
The University of Tokyo 1972
Specializing in Optical Recording Materials

Yasuyuki Goto
Functional Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
Chiba University 1977
Master of Physics
Tokyo Institute of Technology 1979
Specializing in Optical Recording Materials
and Materials Analysis



Ken-ichi Utsumi
Functional Materials Laboratory
FUJITSU LABORATORIES, ATSUGI
Bachelor of Physics
Tokyo Institute of Technology 1973
Dr. of Physics
The University of Tokyo 1978
Specializing in Optical Recording Materials

Erratum

Experimental Modal Analysis of Vibrating System: A Method for Modal Parameter Identification and Its Software Package. Hiroyuki Kano, Fumio Matsumoto, and Junji Kaneko, FUJITSU Sci. Tech. J., **23**, 1, (1987).

Erros found in Table 1 on page 62 of the paper should be corrected as follows:

i	$a(i)$	$b(i-1)$		i	$a(i)$	$b(i-1)$
...	...	...		...	...	...
5	−5.52677E-01	...		5	5.52677E-01	...
...	...	...	to	...	...	...
17	8.05400E-01	...		17	8.05400E-02	...
...	...	...		...	...	...

The authors wish to thank Professor S. Sugimoto of Osaka University for bringing these errors to their attention.

Overseas Offices

Algiers Office
6, Boulevard Colonel, Amirouche,
Alger, Algeria
Telephone: (213)-63-42-55

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, Rubao 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. Bogota, D.E. Colombia
Telephone: (57-1)-296-8076
Telex: 35-45408

Hawaii Branch
6660 Hawaii Kai Drive, Honolulu,
Hawaii 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: 73-45944

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 V4 Siemens AG, Otto-Hahn-Ring 6,
D-8000, München 83,
F.R. Germany
Telephone: (49-89)-636-3244
Telex: 41-5210935

New Dehli Office
Flat No. 517-B, 5th Floor,
World Trade Centre, Barakhamba Lane,
New Dehli, 110001, India
Telephone: (91-11)-331-1311
Telex: 31-2964

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: FUJITSU 262661F

Shanghai Office
Room 1504 Ruijin Bldg., 205 Maoming,
Road South, Shanghai,
People's Republic of China
Telephone: (86-21) 336-462

Taipei Office
Sunglow Bldg. 66, Sung Chiang Road,
Taipei, Republic of China
Telephone: (886-2)-561-7715
Telex: 785-26773

Overseas Subsidiaries

		Telephone	Telex
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 PTAFJ CN
Fujitsu America, Inc.	3055 Orchard Drive, San Jose, CA 95134-2017, USA	(1-408)-432-1300	230-176207
Fujitsu Australia Ltd.	475 Victria 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 92130-2103, USA	(1-619)-481-4004	25-910-337-1127
Fujitsu do Brasil Communicação Eletrônica Máquinas e Serviços Limitada	Rua Manoel da Nóbrega, 1280-2º Andar, C.E.P. 04001, São Paulo, SP, Brazil	(55-11)-885-7099	38-1121777
Fujitsu Deutschland GmbH	Rosenheimer Strasse 145, D-8000, München 80, F.R. Germany	(49-89)-413010	41-17-897106
Fujitsu España, S.A.	Paseo de la Castellana, 95, Edioficio Torre Europa 5, 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 Italia S.r.l.	Centro Derezionale Milanofiori, Palazzo A2, I-20090 Assago-Milano, Italy	(39-2)-824-6170	–
Fujitsu Microelectronics Pacific Asia Ltd.	Room 805, Tsim Shatsui Centre, Mody Road, Kowloon, Hong Kong	(852-3)-7320100	802-31959
Fujitsu Microelectronics Asia Pte. Ltd.	No. 2 Second Chin Bee Road, Jurong Town, Singapore 2261, Singapore	(65)-265-6511	87-22194
Fujitsu Microelectronics, Inc.	3320 Scott Boulevard, Santa Clara, CA 95054-3197, USA	(1-408)-562-1000	25-910-338-0190
Fujitsu Microelectronics Ireland Ltd.	Unit D Greenhills Centre, Greenhills Rd., Tallaght, Dublin 24, Ireland	(353-1)-520-744	500-30152
Fujitsu Microelectronics Ltd.	Vaughan St., West Gorton, Manchester M12 5DU, England	(44-61)-232-6262	51-848955
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 New Zealand Ltd.	Cnr Khyber Pass Road & Nugent St., Grafton 3, Auckland, New Zealand	(64-9)-34-869	74-60157
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.	No. 2 Second Chin Bee Rd., Jurong Town, Singapore 2261, Singapore	(65)-265-6511	87-22194
Fujitsu Systems of America, Inc.	12670 High Bluff Drive, San Diego, CA 92130-2103, USA	(1-619)-481-4004	25-910-337-1127
Tatung-Fujitsu Co., Ltd.	5 Floor Tatung Bldg., 225, Nanking East Road 3rd Section, Taipei, Republic of China	(886-2)-713-5396	785-26209
Fujitsu Vitória Computadores e Serviçôs 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, Colorado 80501, USA	(1-303)-678-0697	–

(The information here as of March 22, 1988)

FUJITSU