DIGITAL HARDWARE: MASS DIGITIZATION

MR. DEAN EDSON: Hardware. This is where the action is. I am really looking forward to the next six sessions. Because of the immense amount of progress being made in the area of hardware development it was necessary to devote a major share of the program in AUTO CARTO III to the subject of hardware.

Ray Boyle from the University of Saskatchewan has come to moderate this very important panel, and to my knowledge there is no one who has a better grasp of the total hardware picture than Ray. In the way of formal introduction, Ray is a native of England, and has been in some form of automated cartography since 1960. When you consider that this is 1978, that is a fair time to be involved in a relatively new field. He is currently teaching electrical engineering at the University of Saskatchewan. Ray received his advanced degree, Doctorate in Chemistry, from the University of Birmingham in England, and has been in Canada 13 years. That really accounts for a marvelous disposition. Canada is a great place to live, and if you do not believe it, just ask Ray sometime. At any rate, it is a pleasure for me to have Ray here to have him take charge of this very important part of the program. So, without any further ado we will turn the program over to Ray and launch into our Hardware.

DR. RAY BOYLE: Thank you, Dean. It is, as usual, an honor to be here at these AUTO CARTO meetings, and I am very impressed with the very large number of people coming to this one, showing the enormous interest in this field at this time. Over the period of time that I have been involved with automated cartography, we have been hitting our heads against brick walls; nobody wanted to listen, nobody was a believer, but things have now broken through on many fronts -- the psychological front and very much in the last few years on the hardware front. I will talk more about that later. Dean described me as a moderator; one of the reasons he asked me to come along is that if it is necessary I am also an inciter, so do not expect me necessarily to damp anything down. The only thing I damp down is people who talk for too long in too long-winded a way; I am quite strict about this. The signal that I will give if in my opinion you are speaking for too long is a digital one (indicating).

The things that we have to talk about have been selected by me as the most important ones. If you do not like these areas and you think there should be others, then only blame me. I am an individual, so I have no department that you can blame and no government organization.

In the next hour and a half, or little longer, we are going to try and cover a six-month university course. It is a teaching, an edcational operation. The people on this panel are not here as representatives of their organization or of their company. They are here as representatives of a methodology. I have asked them to be completely fair in their descriptions of things that are wrong, as well as things that are good. We do not only want descriptions of things that work; we also want to know the things that are available or are being used at this time but which have serious problems. There are many of these and in fact, I think that there are probably a larger number of things that do not work than do work.

The panel members will each have about five to ten minutes as an introduction to say why they are on the panel and what is their area of expertise. I have tried to be representative in my selection of people. You will notice that this is quite a large panel; there are others that I would like to have had on this panel, and I hope they will speak from the audience.

As far as the organization of time is concerned, it is not a proper division to have one and a half hours for digitization and then three hours for mass storage. They are about equal in problem and in advancement at this time.

Do not forget that this is your meeting. It is you who have to keep the discussions going, you who have the questions to ask. I hope the people on this panel will be the resource people, the inciters to your questions. And I will generally try to keep the ball rolling; if everything else fails, Efraim Arazi, on my left here, can always tell one of his wonderful stories.

I think that for the first time in my life I wrote a pre-paper for this session indicating what we would not be talking about; I think that that is a unique departure. I hope you will find it useful. It was to give you the general background of the things that we have discussed in earlier AUTO CARTO meetings. In these AUTO CARTO meetings we are always trying to give you the next stage of advancement. You have the exhibitors to give you the present state of advancement. So, if they should feel left out and ask why things are not being covered, then that is the reason. We are talking about the next stage on this panel; the exhibitors are there to show you the present stage.

Company names will be mentioned. We are not trying to avoid this. They are of great assistance in our work. Without them we would not be in business in automated cartography. No one should hesitate about using a company name. If I hear one person using it more than five times I will suspect he is a salesman, but --

(Laughter). Your work here today, the sort of questions you ask, the sort of responses we give, will advance or retard new developments appreciably. However, I believe that nothing will stop them; these developments are bound to happen and the die is cast. Nevertheless, if you can gain the necessary background, if you can pass this onto your administrators, if you can appreciate the good from the bad and make your proper decisions, you will advance the state of the art very considerably from this meeting. The high number of participants makes me think that we will be getting good results.

During my present sabbatical year I have been visiting many different countries, and examining their assessments of their future requirements in automated cartography. Some of these have been by internal committees and others by external consultants. A few of these are good. In general -- and I am going to be quite honest about this -- I have never seen such a pile of garbage in my life. You must be very careful about such "assessments." Maybe we have to have assessments of assessments, and even assessments of assessments of assessments. I do not know how we deal with this, but it is clear you cannot avoid -- and I am presuming you are all here as cartographers or pseudocartographers -- doing your own thinking, your own cartographic thinking. Other people can help you in other areas, but cartographic thinking you have to do yourselves. You cannot farm that out to somebody and pay him to think for you.

The first discussion area -- digitization -- is one that has been a bottleneck for a considerable number of years. I have been personally involved over the last few months, particularly as I had the opportunity of working at the U.S. Geological Survey.

I am an electrical engineer; I like designing systems that have real applications. I also like trying to design for five years hence.

I really believe that the things that are being developed now, will meet your needs almost entirely in automated cartography. Over the next five years your tools will have become all that you need as we presently see automated cartography. Let me avoid saying anything about geographic information systems; that is slightly different. That is my opinion; you will be able to form your own opinion from our discussions. It even makes me wonder whether I should not get out of this business of automated cartography entirely and look for other challenging problem areas. Perhaps this is the last and most exciting hardware meeting we will have in the AUTO CARTO series. From now on it may be straightforward engineering. There will be problems; there will be troubles, there will be minor variations, but to me, it looks as if breakthroughs have occurred on every front.

DIGITIZATION. We have limited ourselves in the next discussion to scanning digitizing. We are not going to avoid talking a little about manual digitizing; we are not going to avoid talking a little about automatic line following. However, these are not the subject matters of this panel. We are talking about large throughput digitizing. At one time it was called mass automatic digitization, but we did not like the acronym, so we changed it to mass digitization.

The discussion is covering scan-to-line or scan-to-vector conversion software as well. Perhaps this is an infringement on the software panel, but it is an integral part of the hardware. We heard yesterday about the possibilities of working entirely in raster format, but I think that at present most people want the vector format for their data. So we will regard this as a necessary adjunct to the actual scanning operation.

The panel is representative of the general state of the art. We have generalists who have seen, by putting out contracts, what is happening in many areas. We have Joe Palermo and John Baumann, from RADC. One will talk more about hardware and the other will talk more about software. I have also asked Joe Palermo to give you a general introduction to the subject. Some of you will not know the problems of automatic scanning so, before we get into too much detail, we must have some background. Richard Clark from ETL will tell us about some of their problems and what they have been doing on scanning. Bill Switzer is from the Canada Geographic Information System. He has been doing digitization, scanning digitization for a very long time from about 1966; in fact, probably the longest period of time as far as production operations are concerned anywhere. In the experimental runs I have been doing this year, he has been most helpful and has proved to me that the costs of automatic digitization are very economic indeed. I am very grateful to him for the work he has done. Tom Kreifelts from Germany has been doing experimental work over there. He is basically a mathematician and has been working on the scan-to-line conversion. Efrain Arazi on my left will be talking about the hardware. One of the reasons he is here is to make it clear to you that such units for digitization are available "off the shelf."

Most of these people will be talking about drum scan digitizing, but we also have asked Leonard Laub from Xerox (who is really mainly concerned with the mass storage panel) because he knows the work that is being done with the laser scanning flatbed operations; these may well overtake the drum method, as they should be cheaper and faster. We are talking about complete sheets scanned in a few minutes, however much data there is. We are talking about costs of one hundred dollars per mapsheet when we say "good economics." I am now going to hand the microphone over to Joe and ask him to introduce the subject of mass digitization by scanning.

MR. JOE PALERMO: Basically what we are talking about, when we talk about raster scanning, is the need for high speed digitization of large volumes of data. We have already seen, there are a number of different ways of doing this. There are flatbed scanning systems. Some of them employ self-scan solid state arrays, and these go up to a thousand elements. You can adjust these to be anywhere from a thousandths of an inch up to four or five thousandths of an inch in resolution, and you would make a pass over the sheet by moving the head. You also have rotating mirror laser scanners. In this you would typically get to the order of about 500 lines per inch or about two thousandths of an inch, and you would move the scan head. The drum scanners are basically what we have now. We have a number of different types. We have the reflective black and white scanner, and this is typified by a raster plotter scanner, technology resides at companies like Image Graphics Inc. (IGI). This is on the order of, approximately, 26 minutes for a full format scan, which is 122 centimeters by 175 at .025 millimeter resolution. We also have the reflective color scanner. That technology base resides in a number of different companies, one being SCI-TEX. We have some work being done by IBM and some work being done by Hamilton Standard. In these areas we are basically mentioning scanners that go up to 12 or so colors, on the order of 12 to 16 colors, and approximately a thousand lines per inch. We look at the scanners, and we have some general attributes. For economic reasons in generating the scanners, whether it be flatbed or the drum types, about the highest resolution you will normally find is a thousand lines per inch. They have some other ones, you can vary most of them from approximately 250 lines per inch, or four thousandths of an inch spot size, up to a thousand lines per inch.

There is a wide range of these products that you can buy off the shelf or one of a kind. Resolution alone does not really buy you very much. We have to be able to repeat what we scan, and repeatability on these instruments is normally on the order of plus or minus one least count. If we are talking about a thousand lines per inch device, you are normally talking about a one mil repeatability. Accuracy is distinct from repeatability and is the ability to locate in absolute manner any point on the surface of the chart, and typically, these, in an absolute fashion, are on the order of plus or minus two or so counts. So we would be talking about plus or minus two mils for the finer scanning systems. Some of the scanned results will be talked about and ways to handle them. If we are now looking at a large format chart, let us say 40 by 60 inches, we are taling on the order of about two and a half billion pixels in that chart. This is a very large amount of data. Most of the people today run-length code their data. We are not talking about photoimagery where each pixel is distinct, we are talking about charts, and we can run-length code to reduce data volume. It still ends up with a very large amount of data which has to be put somewhere -- on disk, on tape, and stored.

We are taling about a discreet system that would operate in approximately a one mil viewing plane, and it does not do many things that the human eye does. The human eye will look at something like a brown contour running over a green background area, and will see the same brown contour overprinted on a white area. The actual color scanning system, when it sees those, will be looking spectrally at the colors and will see, in many cases, varying colors. It is a difficult task to separate those and recognize brown under a number of varying different background conditions. Also, since we have a discreet system, we are going to have to quantize the output data, usually at the finest one mil steps. When we are viewing this type of data, you encounter problems of, say, an aperture half filled and half not filled, whether it be a black and white scanner or a color scanner. What does the scanner actually see? If we are looking at a four mil line, would our output be on the order of two, or so, mils, or would it be on the order of five mils? When you go to use this later on it is important, for the quality of the output product, that these types of scanning problems be taken care of by the system, either on the scanner or by some backup system. We also have a number of other situations which are encountered which have to be handled, and they are not basically yet being handled, I think, by most of the hardware you will hear about. That is processes like screened data. It is, basically using 120 line screen, the color scanner, using a one mil aperture, is going to see areas of color and areas of background, or white, or whatever happens to be screened on. This is going to greater generate a deal of data. Ways have to be arrived at to be able to detect these screens and outline these screens, because the massive amounts of data you get from picking up the individual dots basically burdens the system, and it would take a lot more time to process than you would actually like to use. Also, there is an increasing use of process color, which is a very difficult situation to handle. The one mil scanning aperture is going to see individual process color dots and will not tell you what the apparent color is. You have to take a macro view of that, and that, either, takes some sophisticated optical system, or it is going to take a large scale processing system to be able to get a macro view of colors.

Basically we are all scanning data for a purpose. Some of us would like to be able to manipulate data in a totally raster form. You heard something about that yesterday. The SCI-TEX system does that, it will work with that. We are also talking about, in many cases, if we want to actually change the data -- if we want to scan it, if we want to change scales, we are going to have to start generalizing data, things like this. You will hear about this. We may have to convert from raster like line to vector formats and things like that, you are also going to hear about that. Scanners alone will not solve your problem. We have to have systems which back them up. One scanning system which I was going to mention is the raster plotter scanner. I have two slides. This is a view of the drum subsystem of the raster plotter scanner. The black and white scanner is capable of 16 shades of gray. Resolution is 400 lines per centimeter over the drum. Spot size is .025 millimeters. Repeatability is plus or minus that element. The absolute accuracy over the system is plus or minus .05 millimeters. This is arrived at by having temperature control within the cabinet for the drum and the bearings, things of this nature. The lead screws tolerances are plus or minus one degree Fahrenheit to arrive at that mechanical accuracy. Also we have a substrate thickness compensation. We are looking at things which have to be in the system. If we are scanning for a four mil black and white film or a seven mil black and white film, the actual dimensions or the length dimensions you will get around the drum will vary or change by approximately 20 thousandths of an inch. This has to be compensated for somewhere. We have chosen to do it on the hardware on the fly. The format, as I say, is 122 centimeters by 175. At .025 millimeters we can scan that format in 26 mintues. We are going to end up generating about two and a half billion pixels, if you had a binary format, in about 26 minutes. Next slide, please. The associated processing system of this scanner consists of a PDP 11/40. It operates under CAPS -11. We decided on that system because we really did not need more than that. It has approximately 32K of core, 16 bit words. It outputs to two nine-track 1600 BPI, 125 inch per second mag tape units.

Now, these mag tape units, in order to handle that volume of data, even in run-length coded format, have to operate simultaneously. What is actually done is that the output from the system is, the even scan lines like two, four and six, and the odd ones, one, three, five and seven, are put on alternate tape units, so anybody who wants to decode this data would have to have this same type of setup. 125 IPS is necessary to keep up with the output data rate. The copy, it can scan now, is black and white, positive or negative. It can scan transparencies, translucent or opaque, black on white or white on black. This is simply handled by a flip of the switch. These systems are being used. They have been put in, on an initial basis, at the Defense Mapping Agency Aerospace Center and the Topo Center down in Washington. They are initially starting to use our backup by software, and they should be going into a production mode in the relatively near future. Thank you. (Applause.)

DR. BOYLE: I will have to ask Efraim to come to the microphone next. We are going essentially along the order of the table. I would like to stress the fact that there is a difference between people who want to digitize the separations of which there are a very, very large number; these are the straight separations which make up, say, for example, 1:24,000 areas, or to take a color printed map and scan it. These are very different jobs, and I would like to stress that difference when you are doing your thinking.

MR. EFRAIM ARAZI: My talk will be divided also into two panels, today and tomorrow, because the system I am familiar with, the Sci-Tex RESPONSE Design System, comprises a color separation or black and white scanner, a graphic interactive station with soft display, which is an inseparable part of the scanner, -- as you will see in a minute -- and also a laser plotter; a high precision laser hard copy plotter which outputs the 40 by 72 inch format separations which are generated in the system.

Our background, very briefly; Sci-Tex is an industrial company which has developed with its own funds for the graphic arts industry and for the textile industry which I will describe in a little while as a fairly high performance scanning and editing system for color work. We have installed to date 76 of these systems around the world. They are installed in the United States, all over Europe and in the Far East. The equipment used, the RESPONSE 200, includes the scanner which I will be describing this morning as well as the color television interactive display and the laser plotter.

This is our first visit, our first call on the cartographic community. We found to our great delight and surprise that our equipment -- which has been, by the way, already deployed in some industrial and some Government organizations for over three years -- is apparently capable of handling some of the finest cartographic products that are around. Let me start with one of the more difficult experiments in scanning, that of scanning a printed map. I will ask my associate to show progressive combinations, but first put all the layers together. This is a corner of a standard USGS 1:24,000 scale topographic map that was scanned on our scanner. I will explain in a few minutes how this scanner operates. As the sample is lifted layer by layer, let me say that in one pass the scanner separated the eight colors that were in use in this particular map. Ilan, you can flip through the separations now. These are by now perfectly clean separations. They are clean not because the scanner is perfect. I will be the first one to admit that while scanning color reflectance copy, there are some problems that have to do with either creased papers or with inks printing on top of other inks.
You will probably remember that cartographers do not pay great attention to brown lines printing over green areas and other such over prints. Let me ask Ilan to show us the original output -- take the brown color, the blue color, any one of them. If you look carefully, you will see under the word "Hyco River" a shadow of a line, a very thin line, which is really a line in another color. This particular sheet is the black printer. These are picked up by the scanner, except they are cleaned up later by the combination of the software that exists in the equipment. The equipment has software that will identify, locate by its continuity, by its location or proximity, all sorts of errors that were picked up.

Let's go back to the original drawing. Now, Ilan, if you can show us the other contour map that we obtained by scanning a black and white transparency. This is displayed with absolutely no retouching, no manual intervention, no editing whatsoever. This is typical of the quality that can be obtained. Now, let us turn off the projector, and I will very briefly describe the attributes of the scanner.

As I mentioned before, this scanner is part of a problem-solving computer system which is deployed all over the world. The scanner we now have accepts opaque or transparent material in sizes up to 36 by 36 inches, in up to 12 colors. The scanner operates at 200 scanning lines per minute at a resolution of about 1000 lines per inch. It recognises the color on the fly at the rate of about 250,000 picture elements per second, and codes them in run-length coding, and sends them to the disc memory. We use disc storage as an intermediate storage. The systems can be made available with 20 megabyte or a 300 megabyte disc store. The 300 megabyte disc store has so far taken care of any and all cartographic products we have run through at the sizes specified, although these may include up to 1.6 billion picture elements. Apparently the coding efficiency is such that we had no trouble with overflow.

The equipment is self-teaching in color recognition. Basically, if the system operator places, say, a printed map or a transparency on the drum of the scanner, all he has to do is aim the scanning head onto the colors in use, and to press a button indicating "this is color no. 1, this is color no. 8, etc." The computer system recognises the particular reflectance characteristics of each color. Once the scanning is initiated, it recognizes and stores every picture element labelled with the particular code, also keeping an eye on the colors surrounding it, and sending a file of run-length encoded data, to the disc. So, this is a very brief explanation of the operation of the scanner. I will be very happy to go into greater detail on its attributes in response to questions which may come later. Tomorrow we will also discuss the interactive editing station which is used to clean up all the data and all the errors that creep into the scanning process.

Thank you (Applause).

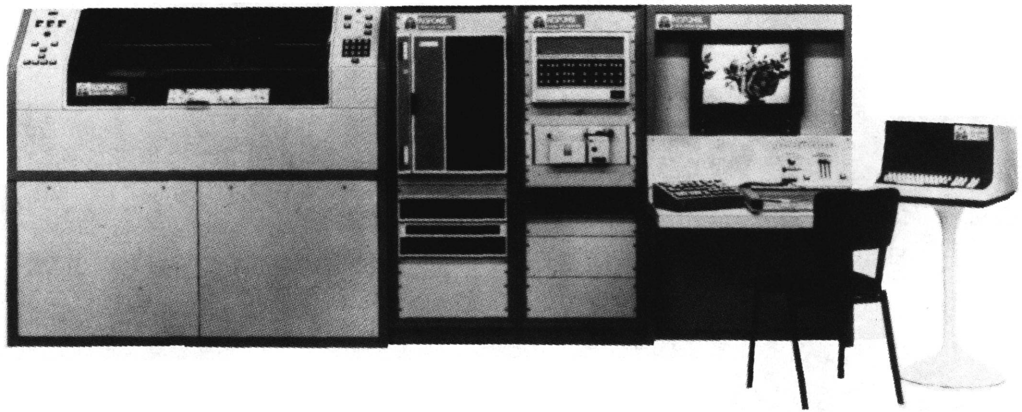

Fig. 1 The SCI-TEX RESPONSE - 200 Raster Design System.

DR. BOYLE: I think that adds an exciting new dimension. I mentioned that this scan editing will be added into the part that is called "small operational systems" tomorrow where Christian Hoinkes will be taling more about the vector. But I am getting Efrain Arrazi to talk in sort of a complement to this about the raster editing. So, for those of you interested in raster editing, I think that should be a good discussion tomorrow. The next one along the line is John Baumann.

MR. JOHN BAUMANN: I will try to keep my remarks fairly short here so we don't get the digital gong from Dr. Boyle. (Laughter.) In 1969 RADC exhibited a color raster scanner at the annual ASP-ACSM meeting in Washington. At that time that scanner generated quite a bit of interest. There were always a large number of people around the booth. We were hoarse at the end of the week. There was one question which came up very often, however, and that was, "How are you doing at scanning litho sheets?" Our answer to that at that time was, "Well, we really haven't done too much of that yet." What would be the possible application of it in the future? We saw it as a scanner for scanning hand-drafted manuscripts and for scanning black and white separation negatives.

Since that time there have been a number of experiments with scanning lithographic sheets. We can see that there are a large number of problems to be overcome in that area. We have been developing software mainly with the intent, again of scanning black and white color separation negatives or hand-drafted manuscripts while continuing to investigate the problems of litho scanning. As Dr. Boyle said previously, mass digitization, in this case raster scanning as we were mainly talking about this morning, means not just raster scanning and creating a tape in raster which would then be used by somebody, but also to process the data to put it into a lineal or a vector form which is more easily acceptable to the user.

Most users do not have either the hardware or the software processing capability at this time to handle raster data as raster data. It takes getting this raster data into a vector format so that the user can access the data, perform symbolization of the data, and other manipulations such as scaling.

There are a number of alternatives to raster scanning, as Dr. Boyle mentioned, and we do not want to dwell on them here. One is automatic line following. Semi-automatic line following is also being investigated. There has been a large investment in hardware in the government and also a large investment in software in the government. For private people to be using raster scanners which are inherently expensive, becomes a very difficult and never ending task to show that the raster scanning can be cost effective. There are some companies, however, some of them represented in the exibit area, who can take raster scan data and actually provide services for a particular user to raster scan the data and convert the data to a vector format.

Getting back to the lithographic scanning or actually scanning of color separation negatives and the problems involved, I see that probably one of the main problems now is the conversion of symbolized data and data which has broken lines in it which occurs by color crossing into the vector format. Over a number of years, and, as a matter of fact in 1969, the ability to convert from raster to lineal was really in its infancy. Most of the people who were doing that work were experiencing great difficulty either in processing time involved to do that conversion or in creating a large number of errors in the conversion. That was not acceptable, because that is the main reason for going to raster scanning, you can scan the data, pick up the information very cleanly and with very little degradation from the original. Since that time, a number of people have gone with a skeletonization type algorithm which thins down the raster data to a single element, and then the single elements can be strung together into a lineal format.

But the problem exists again, and I think this is the main point I want to get to, is that if we are going to scan lithocharts or color separation negatives with its symbolization such as railroad tracks with its tics, intermittent drains, and if you look at any map or chart of this area, almost every drain on there is an intermittent drain, that would at this time entail manual digitizing. So those problems I think still have to be attacked in the future in order for mass digitization to become a useful tool to most everyone.

Thank you. (Applause.)

DR. BOYLE: Thank you very much. I think Richard Clark is the next one down the line. Richard is from the Engineering Topographic Laboratory at Fort Belvoir.

MR. RICHARD CLARK: I am not sure whether I was selected as a representative of methodology that did not work or methodology that did work, since we have some of each. We at ETL became interested in raster processing along about 1967. About that era, XY digitizers were coming out, and everybody was trying them out and finding out that they did work but they were rather slow; a rather slow means of getting high density data into digital form. We thought there must be a faster way. We knew of some of the Canadian work through IBM with drum scanners, some work that I believe Bill Switzer is connected with. We awarded a contract to IBM to look into a drum scanner plotter for us. The plotter was tossed in because of commonality of hardware.

Our needs, as we saw them at the time, were more for digitizing planimetric type information than contour information, as we felt that the contouring problem was being taken care of more by other automatic means like the UNIMACE and automatic contouring programs. If I could have Vu-Graph No. 1. In 1972 IBM delivered a laboratory model scanner plotter which we used as a test bed to determine what would work, what wouldn't work, what we could do with raster data, and things of that nature. This device had a 24 by 30 inch format, and scanned in one mil, two mil, or four mil spot sizes. We determined later that four mil was a bit crude, but would pick up a goodly portion of the coarse line work, or I should say the low density line work. Two mils would pick up nearly everything. When you got into doing some actual contour scanning where the contours were very close to coalescing or where there was any other high density information, one mil spot size was necessary. Scan times typically were around 20 minutes for the four mil spot size, and double that, 40 minutes for the two mil, and 80 minutes for the one mil, which we considered was a fairly realistic time for getting that volume of information in.

Format-wise, we had two formats to select. One was a run-length coded format, which put out four data bytes per transition. The other was a pure binary form, which simply flooded data out, as you will see in Vu-Graph No. 2. The 19 by 22 inch format, which was typically our image area, and for a 250 spot per inch, which is the four mil resolution, you can see there are 26 million bits of information coming out. On the right are the number of reels of magnetic tape at 1600 BPI, with no compression--that is, one bit of information per bit on the tape. We have two-tenths of a reel of tape in this case. Further on down we hit the thousand lines or spots per inch; you can see we are starting to build up data in a hurry--418 million bits of information--and we are up to 1.6 reels of tape. Remember, this is 1600 BPI. If you jump to the bottom and look at a 50 by 70 inch format, which is the image size of the new DMA scanner, you can see the tremendous volume of information that is coming off that scanner. We reach over ten reels of tape if we have no compression. Well, of course, we will be using compression; run-length code will be employed since ten reels of tape becomes rather prohibitive to be carting around and storing.

The IBM scanner, just in passing, had several modes on it. As I said, it was also a plotter and plotted at the same resolutions, one mil, two mil and four mil. But, in addition, it had color scanning and gray shade capability. The color scanning capability used a two-path system, with a dichroic mirror to split the path essentially into the red and blue elements, and was found to be relatively unsatisfactory for discriminating colors. We also found that on the input end it was a bit difficult to actually encode in color. Here we are talking about inputs manuscripts that have hand-drawn center line information, with no symbolized information. It is one thing to do this with pen and ink, and it is another thing to do it with colors. We found it was a bit difficult to color encode, and especially to do it with colors that could be picked up reliably by the scanner. So, our use of the device has been almost exclusively for strictly black and white scanning.

I might point out that there is no magic cure-all in software that is going to make up for sloppy draftsmanship on the input document. If your hand happens to shake when you are drawing a road or stream or whatever, that is pretty much what is going to come out the other end, unless you have an editing device to go in and actually do some straightening at that point or putting in the proper curvature. If I could have Vu-Graph 3. This is just a quick overview of the process we are talking about. The raster scanner is on the left. Regardless of the computer doing the software work for you, most of the raster processing operations involve a thinning operation, (also called skeletonizing or cloud elimination), and, some kind of automatic editing to pick up line breaks, perhaps, and typically a vectorizing process. As Jay pointed out, it is not absolutely necessary to vectorize your data. You can go through the whole system and stay strictly in raster. It is quite fast to do it that way. But most users either for editing purposes or data base type work, want to have some form of vector output in there.

Could I have Vu-Graph 4 now. This represents data that was already scanned. You can see the staircase effect in some of the lines. This would be a proofing output. May I have the next Vu-Graph. This simply shows what the lines look like after the thinning process was accomplished by the software. We are down to one spot wide, whatever that spot happens to be. Next slide, please. Typically you run into places, such as the junction of this stream network, where there is a break in the data. Automatic editing can take care of small breaks like that. Next one, please. It has increased the information by one spot and caused continuity through that junction. Next slide. The vectorizing process, which is also called raster-to-vector conversion or linealization, creates some type of vector format from this data.

We are on the output end of things now, consisting of somewhere in the neighborhood of six or seven classes of data. Each one of these classes would have been a separate overlay or a separate data sheet going into the scanner, as it only handles one class of data at a time.

Next slide. We also have areas, in addition to straight line networks such as roads and streams. Now, areas, are very nicely handled in raster format. All you need to go in with is the outline of the area. It can be a relatively coarse scan such a four mil in this case. The draftsmanship of the lines does not have to be as critical as with roads and streams. The area is filled on an output process with whatever class or whatever symbol it is that you want. In this case I believe we have the orchard symbol in there. It does not really matter what the symbol is, it will fill these areas, and then, if desired, will suppress the border around the data.

May I have the next one. One passing comment on the raster output end. I will not dwell on that because another panel will be covering output, but with raster data you can also do some other tricks such as merging and suppressing of data. As I indicated, you can suppress the line, and you can merge all of your classes of data to make up one sheet, such as shown on one of the earlier Vu-Graphs with the various roads and railroads and so on. It is a very simple operation in raster to do that. If you want to suppress one class of data in favor of another, that is also very easily accomplished in raster. Software-wise, I will give you a quick sketch of the process we went through. We started out with some programs that ran on an IBM 1130, and we marked our progress into days of computing time per overlay. We came up the scale with some 360 programs, and we got our time down into hours of overlay. Later, very similar programs were run on the CDC 6400, and the time per overlay was down into minutes.

The handwriting on the wall all of the time seems to be that if you could afford a drum scanner, the big problem was what were you going to do with all that volume of data? We did not feel that sequential processors were going to cut the mustard, so to speak.

In 1974 we started some work through Goodyear with their associated array processor called the STARAN, as shown on this Vu-Graph. The STARAN works with 256 bit by 256 bit arrays. Because of its array architecture, the STARAN is ideally suited to raster processing. This was why we looked into processing on the STARAN. The STARAN has access to all data on a content addressable basis rather than strictly on a coordinate addressable basis, as you have with your sequential processors. You have operations going on in parallel, whether it be arithmetic or search or logical operations. This means you can acquire any data from memory in a single memory step without knowing its location. We processed quite a bit of data through the STARAN, enough to come up with some representative times and we also produced an output map called Lake Istokpoga. If I could have the next Vu-Graph. I show this because it represents the different classes of data that were processed. We started with 11 overlays, all of them hand drawn. As you will see, it is relatively sparse data. We processed the whole thing with four mil technology--four mil spot size, and four mil output spot size. The total STARAN time to process all of this information was 39 minutes. Of that, three and a half minutes was actual STARAN time. The rest was devoted to tape I.O. time, which, to me, was the first glimmer of light that raster processing really could be cost effective. We have another program that is being installed at ETL currently which will do essentially the same thing except it now works with two mil and one mil technology, which means now we can get into the finer data and can get rid of some of our coarse line work where the staircase effect from the raster processing shows up.

We have also found out some other interesting applications of the STARAN for processing raster data. One of them is line separation by line width. Here we are getting into a form of deciphering or decoding of symbolized data, where we have selected line work that was specifically drawn at a controlled line width. This shows the line drawing that we scanned in. This is one of the line weights--I have forgotten which one it is, I think perhaps the seven mil line--that was separated out from the rest of the data. Here the secret is to start with controlled line widths, not just plain hand drawn lines. I might mention here also that line width processing time is a function of the width of the line; since you are doing line thinning it will cost you so much for each iteration on the line thinning before you get down to unity width. So it behooves one to not have too coarse a line to thin. We can see various applications for this line width detection, because you can now go in and code various classes of data by line width. You can have perhaps, as they have here, road classes. Or, you could have three or four classes of contour lines which could be decoded in the software simply by the line width. We hope to be pursuing this further.

Even though we started out excluding contours from the raster process, we soon found out there was a good application--in fact, I think perhaps one of the biggest applications of raster processing--for digitizing some of these massive amounts of already existing contour plates that the various services and other organizations have. We are in the process of starting a contract with the STARAN people to write a program we are going to be calling Contragrid. This will essentially replace the operation going on at the DMA Topographic Center now, where they manually digitize contour sheets on a floating arm graphic recorder. Of course, you recognize that can be a massive number of lineal inches of data on a contour sheet. If I could have the last Vu-Graph. This would be a typical contour sheet that somebody has to sit down, (I believe shackled to the equipment), and manually digitize. We think this is a prime application for raster processing. We know that the data can be scanned in. You may have to use one mil spot size to pick up all of the fine line work or, more particularly, line spacing. We know that the numbers that appear in there can be recognized and then be deleted out.

The hard part with all of this is how do you associate the elevations? Because it is just a series of lines to the software until you put an elevation tag on it. The STARAN software will be doing a sizable amount of automatic elevation detection. Part of it will be done by going in with a list of the index contours at the borderlines. A good portion of the automatic processing can be done just based on that information. We are certainly going to need some interactive capability in there to pick up the elevations that cannot be done automatically. We may be able to use the actual numbers and do character recognition (some of that work was done earlier by IBM) to further augment the recognition process. The rest of the contouring program will be simply to break up this information into a uniform grid of elevations and store this for data base work.

I hear Ray's magic hook over here, so I will sum this up quickly. Feature tagging is another area that we want to get into, and this will be, as I see it now, predominantly based on the line width detection. One other area of a slightly different type scanning that we are getting into is with a laser plate maker. We have a contract with EOCOM Corporation to make a laser plate maker for the DMA Hydrographic Center. It will use a laser to scan along a curved platen. We have talked about flatbed digitizers, and we have talked about rotating drum digitizers. This application uses a stationary curved platen-type dgitizer.

The world around us is now becoming increasingly involved with raster processes of one kind or another. I still see the software as being the weak part of the link, but I think we are close to solving that part of it. As Professor Marble said yesterday, there are applications for raster and there are applications for XY digitizing, and it behooves one to pick the right one or else the operation may not be very cost effective. Thank you.

Dr. Boyle: Thank you very much. The next one is Bill Switzer.

Mr. WILLIAM SWITZER: Thank you, Ray. Good morning, ladies and gentlemen. First, let me thank you for the opportunity to come and talk to you about the application of an operational geographic information system. I would like you to note the name. It is a geographic information system. It is not a system for automated cartography. There has been some confusion in the last little while as to the name of the system as a result of a name change in our organization. It is the Canada Geographic Information System, and the name of the organization is the Canada Land Data Systems Division.

I am going to talk to you about an operational information system that is a solution, not necessarily the solution, to the problem of raster to vector or line segment data. This system makes use of an optical drum scanner developed by IBM, and delivered to us sometime in late 1967. The software has been fully operational since the beginning of the '70's. Currently we use two methods to get data into our data base. We make use of the drum scanner, about which I am going to talk further and blind hand digitizing a technique with which everyone is familiar.

As an aside, and partially in response to problems raised by an earlier speaker, I will briefly describe our recent acquisition of an interactive editing system from a commercial software house. The system is called AUTOMAP; we do have the source code. It is written in Fortran 5 and comes with a pre-processor to convert it to Fortran 4. It is written in a structured programming fashion and consequently it is very modular. The company that provided the package will train us to our heart's content and will support it, provided we have the dollars. But, it would have been naive of me to assume that they would let me see the software before I put my name on the dotted line.

We intend to use AUTOMAP's interactive editing capabilities to take data directly from the drum scanner have it edited by our technicians, and pass it on to the rest of the system.

AUTOMAP also provides intelligent digitizing, however, at this time we do not plan to use AUTOMAP in that fashion.

Currently we specially prepare our documents for scanning. We scribe them, not because of scanner requirements, but, because our software algorithms that analyze the raster data require a controlled line width. We hope to rewrite those algorithms in the near future so that we will be able to take data that has been prepared in the field, scan it, and using the AUTOMAP software, have our draftsmen edit the data. But these are future developments.

Today, we start with thematic data, maps of both the spaghetti and the classification or descriptive information, as our source document. The classification or descriptive data is converted to digitial format in a conventional way using key to cassette devices. The spaghetti or line information is scribed. The scribing tool produces a controlled line width. That document is put on the drum scanner, the output is on a nine-track magnetic tape. The scanner is run off-line, and the output goes to our CGIS software. Currently we generate not only our own data base, the image and descriptive data sets but in addition we can produce data in AUTOMAP format or in the INTERMAP format. INTERMAP, as some of you may know, is an interactive editing and digitizing system developed by Ray Boyle.

The drum scanner is several years old. It has a fixed hardware resolution of short one millimeter. The scanning time for a typical size document is about 15 minutes; the maximum scan size, 122 by 118 centimeters , black and white. The output is, off-line to a nine-track, 800 BPI, unlabeled magnetic tape. The volume of data for a typical size document is in excess of 40 million data points. Of course the problem is how to get rid of the excess data, at a reasonable cost and in a reasonable time frame. Averaged over 667 executions, the program that does the basic processing from raster to initial line segments costs slightly in excess of $20, and takes slightly less than 30 seconds of CPU time on a 370-168. I do not think that this is an excessive cost nor an excessive amount of time.

Since this is a mass digitization panel, the last foil will describe a project that we completed this summer on behalf of another federal government agency, Statistics Canada. We were asked to produce a data base for a significant area of Canada, approximately two and a half million square miles. We started with map sheets of varying scales. The system is capable of taking map sheets at various scales and processing them into a single coherent data base. But, for time and cost considerations we compiled the maps into slightly in excess of map sheets at the 1:250,000 and 1:500,000 scale.

The project took about six elapsed months and about 30 man months of effort. The costs were in the order of 60 to 70 cents per line inch. These cost figures are per true linear inch and are based on a sample of 12 maps from the 300. It is difficult to say if the sheets were typical . I believe the costs to be high

ie. the sample maps were typical. The average density of the 12 sheets was 841 inches.

That gives you a very, very brief introduction to a system or a solution to a problem. I will be pleased to answer any questions you may have on the project. Thank you.

DR. BOYLE: Thank you very much, Bill. He was also running the experimental work for my tests which I was doing with an experimental run on digitizing the 1:24,000 series for the Geological Survey, taking the separations which we pre-prepared by some of the photo mechanical work. Quite a bit of this was done at Menlo Park. Dick Zorker here was one of the people who was doing this work for me. We found that by doing a very small amount of image preparation we were able to make the whole process of digitizing and the conversion afterwards very much more efficient. Dean Edson has all the reports on this, and if anybody wants to follow up on that sort of experimental stage, no doubt they can. I was most impressed for my own purposes in the results that I did get.

If I could have Leonard Laub. His name is on the second one, for mass storage, but Xerox has also done quite a bit of work with the digitizing systems, so I have asked him to spend some time on this.

DR. LEONARD LAUB: Xerox, of course, is not immediately and visibly in the cartographic business, but Xerox clearly is in the business of handling documents or objects containing text, line drawings, graphs, the sorts of things which are used in business establishments and offices. That has proven to be motivation for an area of technology which seems to be quite germane to the immediate concerns, in that it is gradually being understood that however many trees you might want to plant, and however renewable a resource paper may be, even if you think you can get all the paper you might want -- storing it is another issue, getting to it is another issue, sending it from place to place is yet another issue -- so it is gradually becoming clear to Xerox and other companies in the document and business systems areas that some sort of electronic manipulation of the documents which are handled in business is appropriate. By manipulation I really mean everything that has to do with the document; its creation, whether it is typewritten text, line drawings, or even continuous tone images, half tones or colors. After the creation of the image you may want to store it locally; you may want to modify it; if it is words, you may want to do some word processing, for instance, reduce the average number of syllables in the words so it is more comprehensible to whomever, to your high executives.

After you have done all that you want to do locally with the document you may want to send it someplace else. In fact, you may want to send it many places simultaneously. You would like to have the people to whom you are sending it be able to access it today, tomorrow, next week, be able to get it either by its direct name or by some generic code, get it as part of a group. In other words, you want to be able to search over some data base using various types of keys in order to extract whatever document is of interest. Of course, in the midst of all of this you want to move vast quantities of stuff and you do not want to lose anything, nor to ignore the fact that people are very used to the good aspects of paper -- namely, that you can hold it and look at it and caress it, and also that it is capable of carrying rather a lot of visual information, at least of a subjective nature. It looks nice, typically. Good clean printing or nice reproduction of a drawing certainly looks pleasing, and the objective information may not be all that conveys the impression of value that a piece of paper has to its user. If it doesn't also look nice, if it is not crisp and sparkling looking, it may not please its user even though it conveys the same objective intent. Very few of us are totally objective. You will notice my ears are not pointed; I am not objective either. I have a great subjectivity, and I would hate to give up nice clean serif or sans-serif type faces for ASCII five-by-seven dot codes in reproduced text.

Now, this is not clearly agreed upon in the industry, to the extent that there is an industry of people doing electronic manipulation of documents, and, indeed, a lot of electronic document sending is now being done with a matrix type of printing. At Xerox the decision has been made for a variety of reasons that this is not good enough. Thus, when we consider scanners of the sort which you might want to take a document, the original, say, and send it into a system, or the scanners which you might use to take electronically transmitted or stored or retrieved or created document and get it out of the system onto paper in some hard or soft copy form -- soft copy being what you get when you just want to see, for editing or quick look purposes what is in that mass storage system and pull it up on a CRT -- in all cases we have taken our aesthetic inclination quite far. All of our scanners, hard and soft work at quite high resolution so that we can indeed reproduce all sorts of type faces. Suddenly all kinds of people in Xerox know what is Times Roman, what is Helvetica. They even know what a serif is. It is, of course, a good test to see whether the display is working well enough.

I come to you today not so much to give an advertisement for Xerox in the cartographic community, because this liaison is just really beginning, and we would like to improve it because we have the capability of responding to specific and non-corporate needs, as to tell you some of what we know about scanning of paper objects and transparencies. It so happens, for instance, that the need is incumbent upon us, if we want to reproduce various type faces of various sizes and do so almost as well as does good lithography, to scan with a spot no bigger than about two mils (about 50 micron). This, of course, is well understood by the people who make drum scanners for typographical purposes, the K. & S. Paul and Hell scanners and, of course, all the more recent ones. But it is also incumbent in the business world not to make users wait minutes and minutes. It is all very well to say that we prefer to have cleanly reproduced type faces, but it will not help us very much if it takes, say, 20 minutes to get a sheet of paper through. That is just not fair. It won't work. You can say it may take 20 minutes to find a piece of paper in the file cabinet, and it may sometimes take forever because you took it out last week and you did not put it back in or you put it in the wrong place and you'll never find it again. That is another problem.

It is much nicer if we can say that the new technology, which, of course, costs money, and, also of course, changes operating methods and requires acclimatization, at least in its performance aspects outdoes all the old technology. So, a very important aspect of work at Xerox on scanning has been speed combined with high·resolution. Our aim has been to produce scanners which can, for instance, move legal size documents covered with typescript logos, etc., resolved to two mils or better over their entire area, at a rate of two per second, into and out of a system. Now, we have to bear in mind that seconds and minutes are different things. The rates at which we scan, making scans of five or ten or fifteen thousand resolved spots across a page rapidly enough to scan that whole page in one half second, lead to data rates of several tens of megabits per second. Given the capability of digitizing at such rates, jobs which presently involve much tedium can be performed very quickly. From the base case of 500 milliseconds for an 8.5" x 11" page, time for scanning larger pages scales with area, leading to scan times of seconds for any likely cartographic sheet size. This is to be compared to the many minutes required by a typical drum scanner for digitization of standard size sheets, and shows about two orders of magnitude improvement in the rate at which images can be introduced to a computer system. The use of automatic sheet feeders permits this improvement to be practically realized when dealing with large numbers of images. Without such feeders, time consumed by manual insertion and removal of sheets can greatly exceed scan time.

So, speed that high may be more than many people need. Someone will ask about the enormous data rate; I will say two things with respect to that. One of them is that obviously we have to deal with it at Xerox. Our concerns have been met in part by dealing with a lot of special compression algorithms. Run-length encoding was a good point of departure, but we have gone past that into some more specialized, and, of course, some more proprietary techniques. Typically, with text or line drawings, we can compress somewhere between ten and twenty to one. The other aspect is that we felt forced, even with all the compression that we might be able to do, to investigate mass storage so as to accommodate the gigantic volumes of data that result when you start taking tens or hundreds of thousands of documents and storing them up. But mass storage is another issue, and I personally will be back to talk to you later in the morning about that. Right now I would like to describe some hardware, give you an idea how all this "gee whiz" stuff is built and a little bit of an idea of the practical or production environment in which it lives.

One device is intended to produce a repetative line scan with a red beam from a helium-neon laser. In order to deflect the beam we have to blow it up to a larger diameter and then bounce it against, as Ray Boyle pointed out that I should call it, a "multifaceted rotating mirror". We typically call this a polygon, but in this audience I know I'd better not. (Laughter). That scanning beam is then sent either against a sheet of paper or, if the device is an output scanner, against the xerographic belt or drum, the selenium photoreceptor, which is used to pick up images. That drum is then given its electrostatic treatment, covered with toner and the toner is transferred to plain paper. So, this sort of device is used as a printer. In fact, this type of device is very similar to the heart of the new Xerox 9700 printer which does indeed output legal size pages of two mil resolution, two per second.

It is appropriate to mention the modular characteristics of construction that we exploit in order to make these devices. Typically we accept the fact that our business machines are complex and that, no matter how well we might train the service personnel, we would rather that they did not have to come out too frequently and that, when they did, their jobs could be condensed. So, the difficult parts of the equipment are typically built into modules, and they are intended to be interchangeable with one another. The laser and beam shaping optics reside on a plate which can be removed from the main chassis and replaced in the field if necessary. In all these scanners, the main optical system produces a high-speed scan. The scanned line is being traversed every two or three ten-thousandths of a second. Then, a cross scanner is used to move that line across the page. So, at the rates of which I have been talking, the rates at which these machines typically work, that cross scan will take about half a second to cover a normal page. In this prototypical form, there is no document feed. In the more elaborate devices which are already coming into production, we have exploited our expertise in rapid paper motion. The basic scanner is thus built into a bigger machine which will feed paper quite rapidly and automatically.

These are samples of laser scanners, both for input and for output, which are beginning to become products not for cartography, but they are beginning to become products nonetheless. I guess the most directly useful thing I can say here is that the corporate development program, the usual multi-million dollar multi-hundred man year sort of thing which is aimed at satisfying what is perceived as a large scale corporate need, just because it happens to be very similar to the uses you have here may be indeed very beneficial to you. Further questions are, of course, welcome. Thank you. (Applause.)

DR.BOYLE: We are going to continue our panel discussion on mass digitizing. I am going to ask Dr. Kreifelts to give us a few minutes talk on their experiences in Germany on this matter.

DR. KREIFELTS: For the last three years we have run an experiment in Germany (ARLIP project at GMD) concerning automatic digitizing with raster scanners. I will give you a brief summary of the procedure we applied and some of the results. The starting point is a black and white map which is scanned on a large format drum scanner. We use an Optronics drum scanner which can take documents up to the size of 50 by 52 centimeters. Let me stress two points. One has already been mentioned this morning. The document has to be of good quality and high contrast, because what is lost during the scanning process is later on missing in the line file. With a low quality original one has to spend more time in manual error correction than one would otherwise have to do. The second point is that with our procedure, all lines on the map are treated alike, regardless of thickness and significance.

The result of the scanning is a binary image of the map, the 1-pixels representing line work and the 0-pixels, background. This binary image is then transformed by the

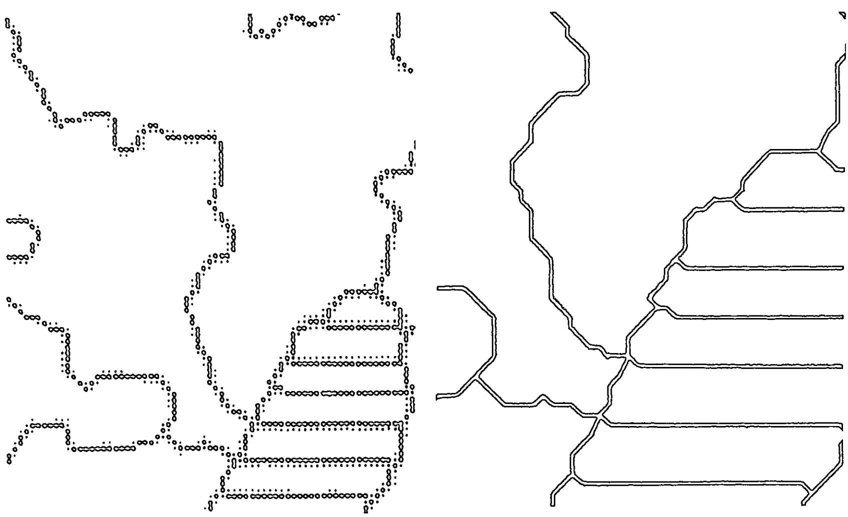

Fig.1 - Binary image, skeleton, and extracted lines.

raster-to-vector-conversion program to a line file consisting of strings of coordinates and a node file consisting of the coordinates of the nodes and references to the lines which meet in the respective nodes. This node file is very useful for constructing a polygon file from the line and node files, which is used in connection with, for instance, land use maps. The node file is also useful for detecting errors when digitizing contour maps. The technique we applied in the raster-to-vector-conversion program is the line thinning technique, which has also been mentioned this morning.

May I have the first Vu-Graph, please (fig.1). This shows the binary image of a map, in this case the municipality boundaries of Bavaria. The fat dots represent the skeleton. The skeletonization is the first step of the raster-to-vector-conversion. Hereafter the nodes are extracted, and then the lines are followed. May I have the next Vu-Graph (fig.1). This gives an impression of what the lines look like without any data compression.

I come now to the results. May I have the next slide, please (fig.2). This shows on which types of maps we have tried this program out, the size of the documents, the scanner step size, which is usually four mils, but occasionally can be two mils or one mil, the overall line length, which varies between 13 and 65 meters, and the number of lines and nodes encountered in the various maps, which is between roughly 1,000 and 16,000 lines.

May I have the next one (fig.3). This gives you some information about the program performance. It shows how much CPU time and how much core storage is used, and what the computer cost is for each linealization job. The test were run on a Siemens 4004/151 computer, which if compared to an IBM/360-50, has roughly twice the speed and the same instruction set. The core storage used is well below 100 kilobyte, and we think we can push it further down, below 64 kilobyte. We will try to implement the algorithm on our PDP 11/70. If you divide the dollars by the meters it boils down to between four and six cents per inch of digitized line. Of course this is only the running cost for the raster-to-vector-conversion. Thank you.

type of map	size [cm]	scanner step size [μ]	overall line length [m]	lines (nodes)
contour	28x48	50	62.60	5096
county boundary	40x43	100	13.00	984 (169)
community boundary	50x50	100	65.20	16414 (10500)
land use	50x46	100	28.60	1930 (1225)

Fig. 2 - Results of raster-to-vector-conversion .

type of map	overall line length [m]	CPU time [sec]	core storage [kbyte]	computer costs [US $]
contour	62.60	1017	93	110.20
county boundary	13.00	275	56	29.80
community boundary	65.20	1115	67	120.80
land use	28.60	393	59	42.60

Fig. 3 - Performance of raster-to-vector-conversion program .

DISCUSSION

DIGITAL HARDWARE, MASS DIGITIZATION

DR. BOYLE: Thank you very much. That is the end of our panel. I wish to mention the people that I would like to hear from in the audience. I would particularly mention Eocom. They are laser scanning and reproducing digital data directly onto printing plates. If anyone is here I would be pleased to hear a few words from them.

I would also ask if there is anyone from DEST. Unfortunately on this panel we have not had a representative of a different method of scanning using a "brush," "comb" or "broom" method.

Broomall is an example of one of the only groups of people doing service digitization by scanning. I think they should be commended for this; there is a great shortage of this in North America and I think there is a great need. I think that maybe some government departments are to blame for the fact that they keep this work much too closely to their chests and want it all in-house instead of setting up outside facilities which people at universities and others can use. I would like to see a much greater growth of these service bureau type of operations for scan digitizing and also for precision drafting, another area where there is great shortage and I feel a great need in this country.

The floor is now open to questions.

MR. MITCH MODELESKI: Mitch Modeleski, ESRI. Ray, this is directed to you. Do you know any operations that have used rasterization to produce a clean DIME file.

DR. BOYLE: Use raster data to produce a clean . . ?

MR. MODELESKI: DIME file.

DR. BOYLE: DIME file. No, I am sorry, I do not. If Duane Marble is here he might be able to tell you, or if a representative of his group is here they could answer that. I am sorry, I cannot.

MR. TOM WAUGH: Tom Waugh, University of Edinburgh. I believe it was Tom Peucker at AUTO CARTO I, when faced with presentations like this, said, "Well, we're not really talking about hardware, we're talking about big ware." (Laughter).

That prompted a university colleague near me, when he saw the ETL description, to say, "If you gave one percent of that to the universities," I should imagine it gave Jon Leverenz a corporate heart attack to look at some of those costs. Two of the speakers actually gave costs, but it seems to me that those costs in fact were running costs. They were not set up costs of buying the equipment, setting up the software. When one looks at the ETL setup, the STARAN and all the rest of it, one wonders how many maps would have to be generated to cover that initial cost. Could someone from ETL or any of the other speakers respond to that question?

MR. CLARK: That is a very good question. I do not have the figures at my disposal on how many maps it takes to amortize the costs of all the equipment we are talking about. We are in research and development -- not using that as an excuse, but in research and development we are more interested in developing the methodology and then determining whether it is cost effective enough to go on into production or semi-production mode. But there is certainly a sizable outlay of cash involved in any computer system. The STARAN computer system probably, a million and a half dollars is a conservative figure for the cost of that. It is not being used strictly for automated cartography but also on other projects, and that is an important difference in the STARAN over some of your other sequential processors where you suddenly find yourself totally dedicated to processing just raster data because of the time involved. The software projects are developmental. The costs vary all over the place. The drum scanners; the cost of developing a drum scanner I do not think is too important right now.

We had about half a million dollars into the IBM drum scanner, but the cost to duplicate that certainly is much lower. I have not had any recent estimates, but at one time after the development contract, about $150,000 was a going figure to reproduce the piece of equipment we had.

MR. ARAZI: I want to drive a very strong point home. As system engineers with experience in graphics, the SCI-TEX Corporation regards the cartography problem to be just one more graphic problem, to the extent you are dealing only with graphic features of those maps. If your purpose in life is to take a collection of all separations or producibles that you have, put them into digital data bank in order to update them in the future or to go from one scale to another or to start meshing together LANDSAT imagery with topographic imagery, for those kinds of work, those are mostly graphic enterprises. I am not talking about rectifying pictures or photogrammetry or all those other activities. To the extent that you want to stay in the graphic realm, we happen to have already installed, as I told you, '76 equipment that have a fixed firm price and do a very simple act: They will scan a piece of opaque or transparent imagery which can be a colored thing such as a printed map, or it will take less effort if it is a separated sheet, permit you to view it and edit it and correct and manipulate it on an interactive display, which is a color television type display, later output it into a 72 by 40 inch film, all for a fixed price of $400,000. Then you have to pay $28,000 a year for maintenance and software support. Those are very firm prices all over the world. So this gives you a number on how much it costs to have a digital in and out system for handling cartography.

DR. BOYLE: I think I could add to that, that if you are only interested in digitizing units, then about half of that.

MR. ARAZI: Yes. If you do not need our laser plotter, if you do not need the output device, and you just have need for the digital data, the scanner alone is in the $200,000 bracket. The scanner with the editing station and with all the computers that go with it is $325,000.

DR. BOYLE: That figure was given, and also the figure of the run costs, say, on a 360 and 370 computer. These costs were given in the talks. I think they are fairly realistic. They proved out to me this summer, anyway. Does anyone else wish to comment?

MR. DICK STOVER: Dick Stover with Broomall Industries. I would like to say a word about our system, which has been developed by private industry. It involves about eight years of development. Our approach has been to work with existing interactive systems, since there has been quite a growth of those. To add to a system of this type is about a $200,000 cost for our kind of system. Now, we do raster scanning, software, vectorization, and we then interactively edit the data. The cost of three to four hundred thousand will cover an entire system with all capabilities. As I say, $200,000 would be a system to add to an interactive graphic system. Now, we are doing scanning with a paint brush arrangement such as was mentioned here earlier, and we will have a presentation in the workshop across the way at 11:30 and one again at 4:00, which includes some work on an interactive tube with this kind of system. Thank you.

DR. BOYLE: Thank you. I think to go back to Tom Waugh's question. When people ask me about this I never recommend that they should buy these systems themselves unless they can keep them in very full operation. This is why I want to see a growth of service bureaus.

MR. KEN SMITH: Ken Smith, Bell Northern Research, Ottawa, Canada.

I have two questions I will pose to anyone on the panel. The first question is: Has anyone done much work on the automatic recognition of characters on the items that they are scanning? The other question is: What percentage of the time does the cleanup and the editing phase of the work have to the scanning process.

MR. CLARK: I can address the first part. We had some work done by IBM about six or seven years ago on character recognition. The work there indicated that the potential for doing character recognition, and here we were shooting for strictly hand drawn characters, looking for, first of all, the recognition rates on totally unconstrained handwriting, and then adding constraints until we got the recognition rate up to some reasonable level. Constraints in this case were simply to box the character in as far as the total height and the width, and then starting special features on the character such as slashing an "0," a serif on a "7," that type of thing. We reached recognition rates with about three sample sets, somewhere in the neighborhood of 92 to 96 percent recognition.

The biggest problem here was in getting the draftsmen to be consistent in the data set. As long as the same draftsman was drawing the characters on the same day, why, you could get fairly high recognition rates. But to do it over a long-term basis, the recognition would drop. We have not pursued that any further, except that some of that work is being utilized in our Goodyear program on the STARAN. We perhaps will be using those techniques in the Contragrid program that I referred to in picking off the numbers that exist on contour sheets where that should be a much easier problem, since most of them were very highly constrained characters when they were drawn.

DR. BOYLE: Is there anyone else on the second question?

MR. PALERMO: The second question about the amount of editing that has to be done. I think it was pointed out by a couple of people that the amount of editing that you have to do is directly dependent on the quality that goes into your system. If you have a high quality black and white positive and you just want an image of your data, then the amount of editing is very little. If you have a multicolored graphic and you have to deal with things such as mixing of colors, and/or color shading, then you have to start doing a great deal more editing. The old adage "garbage in, garbage out" is going to happen everywhere if you have a very poor manuscript. Any editing that would basically have to be done would have to be for a special case, your case, and what type of graphic you are doing and what input system you have. As a general question, I do not think we can really answer that, because we do not know. Depending on what inputs you have and what type of editing you are doing -- the amount of editing would have to be looked at from that point of view.

MR. BAUMANN: To go one step further there. The amount of editing which is required on some of the newer methods of linealization -- namely, those such as Broomall is doing, and the skeletonization routines being performed at other places in industry or in government, produce very good outputs which require very little editing. Again, as Joe mentioned, though, that input is very critical.

MR. ARAZI: From our recent experiments, if you start with a manuscript of good quality and you want it digitized -- say, for example, a topographic map, and topographic maps are, as far as we are concerned, the most difficult maps to handle, if you work with manuscripts you only have to spend about an hour per sheet scanning and verifying it and filing it. An hour per color sheet.

If you work with printed maps, if you work with a map that somebody brought you from Japan, a topographic map, and you would like to digitize it, and you do not have the reproducibles, the scanning is again about an hour, but then you have to spend anywhere from five minutes to one hour per square inch fixing it up, correcting it. It is a learning curve process. It has to do with what they call process color printing or multicolor tint printing and how many colors are in use.

The map that we projected in the early part of my presentation took in the vicinity of half an hour per square inch the first time around for an operator, someone who never dealt with cartography before. This can be further reduced.

MR. CLARK: One more point, Ken. If you drop me a line at ETL I would be glad to send you a report on that character recognition study.

DR. BOYLE: I would like to make a point about editing. One of the reasons I so like automatic scan digitizing is that I have very, very little correction to do. The data is almost perfect and anything that is not perfect could generally be dealt with entirely automatically. However, you have, of course, pure "spaghetti" coming from your device and you have to do labelling. This is sometimes called "editing" but is different. We are now looking at methods of reducing the amount of this labelling work appreciably -- for example, on the contours, only interactively label the index contours and then automatically label the intermediates. On the estimates I was doing on the standard one to

24,000 topographic sheets this summer, it looked as if we had about two hours of interactive labelling to do for all separations for one map.

MR. SMITH: I am interested in capturing not so much topographical maps as large scale urban maps with utility information, and I am looking towards being able to produce a complete data base from which I can do networking concepts. As I see the automatic scanning so far, we have a long way to go yet before you can tell me the size of pipe and what fits together and characteristics of wire and what-not.

MR. BRUCE OPITZ: I am from the Hydrographic Center of the Defense Mapping Agency. We will be getting a scanner that RADC is developing, the color scanner. To our best estimate, total editing time including editing that is not involved in scanner problems but which is involved in changes we have to make to the product for a final output, totally in raster is about five to one, five hours of edit time to each hour of scanning is about our best estimate, five to six to one. I have a question involved in one of the costs I heard. It was from the Canadian experiment where it was 60 cents per lineal inch. Is that what I understood for center line digital data?

MR. SWITZER: For what?

MR. OPITZ: For center line digital data. That seems to me from my experience to be kind of a high cost. I envisaged that the cost would be quite a bit less than that. That is on the order of what we are achieving with our manual lineal systems.

MR. SWITZER: That is the total cost of taking the data from its map format into its digital data base -- total cost, not just the cost for resolving the raster to scan. That includes a 25 percent overhead to cover managerial costs that are not directly recorded.

The data that I have indicates that the cost for the 12 maps that I picked out at random is 55.46 cents per inch. Again, I added a factor to that to cover the additional overhead.

FROM THE FLOOR: Yes. As I say, that seems to be on the order of what the lineal digitizing is currently costing, manual total cost, about that much per inch.

DR. BOYLE: I find it difficult to understand this high figure. It is entirely different from the cost of the work you did for me. We are talking about a hundred to one difference.

MR. SWITZER: If you take the numbers I gave you and you multiply them by the number of lineal inches I think there are in the total maps processed, then the cost for the entire project works out to be about three times the figure I have in front of me, so I suspect that the 12 maps chosen were not necessarily representative of the total, but it does at least give a feel for what the costs per inch are.

DR. BOYLE: I just do not understand it. I agree with Bruce Opitz, and I do not understand that figure.

MR. SWITZER: As I say, the costs come from our internal accounting system that records the computer costs and the manpower effort that goes into it.

DR. BOYLE: Are there any other questions?

MR. MIKE GREEN: Mike Green, University of Wisconsin. It is my impression from the talks that I have heard today that mostly what you are talking about is an efficient means of digitizing line drawing type of things. I wonder what equipment the panel is familiar with with digitizing formats that have gray tones associated with them, like, for instance, aerial photographs.

DR. BOYLE: This is really not part of the discussion of this meeting, but perhaps somebody would like to answer that.

MR. PALERMO: There are some raster systems that are around basically made for imagery. One that comes to mind immediately is known as the laser image processing system. That was produced by CBS Labs before they sold out. That system is basically capable of imagery, digitization in 256 levels of gray and down to on the order of two or so micrometer spots. There is also a transmissive system which works with a laser, single channel, and transmits through the imagery to acquire the total gray shade range of the imagery. With reflective systems you have the problem of going through the emulsion twice and basically doubling the density level. So you are limited on most of those systems to densities that are on the order of one. Transmissive systems are working on scanning up to two and better. That is the only real system that I have seen that has been in the field for a long time that will output a digital tape in 256 shades of gray from an image.

MR. GREEN: Could I ask the gentleman from Xerox if they have done any work along these lines?

DR. LAUB: Thank you for prompting me. I was about to tell you anyway. (Laughter). The output process that Xerox likes to use is Xerography, and that is not a very linear process, so output of continuous tone images almost has to be half tones. Based on that we have to have very, very good schemas, even somewhat better than the ones we have been talking about so that we can make proper variation in size of half tone dots from a screen that might be as fine as 250 per inch. We are doing that sort of thing. We are also scanning in half tone documents and are scanning in true continuous tone devices and electronically generating half tone dots for the output. So all of this is part of the work. This includes color work. Branching off from the color work there has been some multispectral scanning. Actually, it is sort of active scanning which is directed out into the world, and has even been employed in Southeast Asian countries. So, yes, such work is going on. It is also being done at Kodak, and I suspect elsewhere.

MR. JON LEVERENZ: Someone mentioned four to six cents per lineal inch. What does that refer to? Define one, please.

DR. KREIFELTS: That is the computer cost for the raster to vector conversion. You put the scanner tape in and get a disk file out with the lines, and the cost of the job comes to about four to six cents per inch line work.

MR. LEVERENZ: Mr. Switzer, you said 60 to 70. How does this four to six --

MR. SWITZER: The 60 to 70 is the total cost. It is not just the raster to vector --

MR. LEVERENZ: Yes. But how does the four to six cents compare with the --

MR. SWITZER: Costs are very, very difficult to pin down as well. In our case we buy time from a commercial service bureau. So we pay the commercial service rates. If we had in-house systems conceivably the costs that I quoted would be lower, possibly they would be higher as well. So that has an influence on it. I cannot tell you, as I do not have the figures in front of me, as to the actual cost per lineal inch from the raster to the vector. All I can give you is the average cost of running a particular program that does that over the past three or four months.

DR. BOYLE: I think that the 50 to 60 includes Bill Switzer's salary -- (Laughter).

MR. HARRISON: My name is Harrison, Australian Army Survey. I would like to ask Dr. Laub if he could indicate whether he is working in the analog or digital domain, and if he is working in the analog domain, what he sees as the real problems in converting that to digital format.

DR. LAUB: Luckily I can avoid the second question, because the answer to the first is that we are in the digital domain. Later on in the mass storage session I am going to be talking about some analog storage which winds up looking quasi-digital -- that is certainly analog-made-discrete. The half tones I was just mentioning in reference to another question are also analog-made-discrete. But we have decided as a corporate direction to do everything digitally. This lets us format any type of scan together with any type of computer data combined with voice, combined with telemetry, combined with God knows what, and also, to permit expedient access to it, search over it, whatever you might like. We have accepted the ideas that computers will be managing all of these data bases and transactions, so everything will be digital. (Amen).

MR. ARAZI: I think a remark well worth noting is that every color magazine you look at, every commercial catalog you look at, about 40 percent of the color photographs that are reproduced in processed colors are these days being separated by digital scanners. There are in the United States some 600 or 800 color separation scanners made by three or four companies which have the data converted into digital form for a variety of internal purposes, and are reproducing extremely high quality graphics in processed color. So, the digits are clicking along. They may not go in and out in the computers -- the data may be kept internally. But the business of taking pictorial data of the highest quality, digitizing it, and outputting it again in a color separated form is here, is now. Those people are working sometimes at 5,000 lines per inch scanning, and scanning color transparencies with any density range which is obtained by photographers. So, this state of the art has been around for six, seven years now.

DR. BOYLE: Are there any other questions?

DR. ROBERT AANGEENBRUG: I am a little bit concerned about your figure of five hours editing per one hour, say, acquisition. Did I hear that correctly, that there is a great deal of editing time or whatever you call it "annotation," I suppose, Ray . . .

DR. BOYLE: That is the same figure that I said. I said two hours when the digitizing time is about 20 minutes.

DR. AANGEENBRUG: I wonder if some of the panelists could comment on what quality personnel or what kind of training or what kind of preparation could go on, say, either in your institution or in ours to providing you with employees that would know how to do this facet. Because you alluded to the learning curve in terms of edit process. What kind of skills are you looking for? It is fairly significant to us in higher education.

MR. ARAZI: As far as our SCI-TEX experience, most of the people that are employed in operating our graphic design systems are not people with higher education. It is basically a technician job. If their editing has to make up for faults in the incoming document -- I will take the extreme case: You are requested to scan a printed map. Your problem in life is to find faults in the graphics which emanate from overprinting of colors or from an array of facts that have to do with screening in the original, and to prepare the data, to go through the data to straighten it up prior to digitizing and/or conversion to vector, it is not a job for an academically trained personnel, it is a job for the graphic technician. The same kind of people who are doing scribing, and it is actually a more interesting occupation for the very same kind of people.

If the job is to label, to identify features, that is more of a cartographic job, and that is more in the two hours per scan type category. But it is a straight cartographic job except that we give them a very, very fast graphic unit. You have your 12 layers together on a color television screen. If you annotate on one layer, you immediately update on all the other layers. You have some ground survey data that you want to update, you want to change the nomenclature on something, you can do it on all layers together. So, basically, you have a very high power pantograph with a color television screen display, and you do your usual cartographic work. But, here again, you know the kind of person that does that.

MR. SWITZER: Maybe I can comment on the one example that I used for volume processing. Looking at the information I have in front of me, it says that what we term manual error correction, which is the correction of the cartographic data, if I can call it that, just correction of the spaghetti where we join lines together, et cetera, took approximately one half hour of five and a half hours manual effort. So that the average effort was five and a half hours for each of the maps, and approximately one half hour of it was devoted to MEC. Now, in addition, of course, there is some other editing in the sense of capturing the classification data, so there is some editing that goes on at that stage, but this suggests it is considerably less than 50 percent.

DR. BOYLE: Are there any other questions? We will now close down this panel with very great thanks to the members of this panel, and we will start on the mass storage panel, if they will come to the rostrum, please. (Applause).