1959 Kolsky File

PRODUCT PLANNING

Technical Report

DATE 4/1/59

NUMBER P 16

TITLE AN EXAMINATION OF THE NEED FOR ENLARGING PRESENTLY AVAILABLE CHARACTER SETS

AUTHOR E. G. Law

LOCATION Product Planning - Poughkeepsie

SUMMARY

The number of different characters which our equipment will read and print has been enlarged on more than one occasion in the past in response to customer need. The major changes have included alphabetic characters and a larger set of special symbols. A small but important segment of opinion now proposes a further enlargement to help customers make better use of our large-scale 700-7000 series systems. Although customer opinion at this time is far from unanimous, there is good reason to believe that the proposal accurately represents the direction customer demand will take. After discussing the nature of the problem and differences of opinion as to its solution, this report concludes by recommending that we begin a phased program of developing equipment to use an expanded character set.

THE NATURE OF THE PROBLEM

The largest generally available character set that we offer at the present time consists of forty-seven characters plus blank. It includes twenty-six alphabetic, ten numeric, and eleven special characters. In general, customers have found it reasonably easy to live within these limits. With only a few special-purpose exceptions, this set has proved quite satisfactory in handling the bulk of their work. But within the last two years, a fast growing use has brought into question the adequacy of the present character set.

This use is made by advanced semi-automatic programming techniques for our large-memory stored-program machines. The goal is very worthwhile: more efficient and profitable use of the machines and of the time and talents of valuable programmers. It calls for the means of expressing complex arithmetic and logical sequences in a concise and unambiguous manner. These problem statements must then be machineable without any loss of meaning. Extremely ingenious programs can and have been written to "compile" a correctly functioning applied program from these statements. The processors themselves, particularly those with binary internal organization, have no trouble at all with enlarged character sets. The weak link is in getting the statements into the machine undistorted by the language limitations of regular input preparation equipment.

It is also necessary to be able to print the results of the compiling run and later debugging runs. Experience has shown that it is essential that these printouts be in the same language as the original problem statements. Ideally it should be possible to print these either on-line or peripherally. There are very real obstacles to this; therefore, it may be necessary to compromise. The ability to list as well as keypunch input cards in the extended language is also important.

The problem stems from the increased abilities of the computer line and greater knowledge of how to put these abilities to best use. It is a problem of adequate support equipment, not of the computers. Proponents of additional characters say that without them, an unnecessary ceiling will be placed on effective computer use.

DIFFERENCES OF OPINION

Most customers agree readily on the desirability of a problem oriented compiler language which is direct, concise, and unambiguous. Unfortunately for an easy evaluation of the question which this report attempts to answer, the general agreement ends right about there. Some of the more pertinent differences of opinion follow:

1. Is an expansion of the presently available character set really a necessary condition for achieving a direct, concise, and unambiguous compiler language?

2. How much better a language will additional characters buy?

3. How much additional equipment cost can be justified by a richer language achieved with an expanded character set.

4. Just how much actual customer interest is there at the present time in an expanded character set and the means of implementing it?

5. How many additional characters are needed to yield the desired goal within economic limits?

ARE ADDITIONAL CHARACTERS A NECESSARY CONDITION ?

If the character set could be enlarged without requiring additional or more costly equipment, there would be no problem in finding large numbers of interested customers. But since most customers know that this is not likely to be the case, many of the cost conscious ones are asking very seriously if the same results can't be obtained without enlarging the character set. They suggest using combinations of the present set to represent the symbols which are being proposed as additions. This idea is not new, it is already being done to a lesser extent in many present systems. And regardless of how much the character set is expanded, it will still have to be done to express all of the relationships involved. The difference is one of degree, not of kind.

The International Algebraic Language distinguishes between what it calls the hardware language and its reference and publication languages. This recognizes the fact that there is bound to be some practical upper limit on the number of characters which can be provided. Multiple character combinations are required beyond that. This form of substitution, while undoubtedly not as convenient, is still a practical alternative to expanding the character set. Indeed, a SHARE committee has already proposed an IAL representation within the limits of the present character set.

HOW MUCH BETTER A LANGUAGE ?

The generally agreed upon goal is a compiler language which is direct, concise, unambiguous, and as easy to use as possible. Single characters instead of multiple character substitutes are certainly more concise. However, the degree to which this affects overall length of statement depends upon the frequency of use of these particular symbols. This is difficult to estimate. Mr. R. W. Bemer in Applied Programming has estimated that forty additional characters would cut scientific problem statement length to 60% of what it is with the present set. Other estimates are not quite as optimistic as this one.

The developers of the early automatic programming systems have placed the blame for most of the awkward notations which must be used on limitations in the present character set. It is questionable that all of the blame really belongs there; but it is true, for example, that FORTRAN would have had many more awkward notations than it has if the special set of symbols hadn't been adopted. And that was a substitution for mnemonic reasons rather than an enlargement.

Elimination of the need to make multiple character substitutions for symbols with well established meanings would certainly contribute to a better language. The people who have had the most experience with advanced programming languages feel that one of the greatest needs is for additional punctuation symbols. Multiple character substitutes seem particularly ineffective as an alternative in this case because they run more risk of introducing ambiguities of their own making. The purpose of extending the language in an IAL type implementation is to facilitate effective communication with the machine. Some feel that having to write multiple character substitutes will more than nullify any gain the system might produce in making the machine easier to use.

On the other hand, programmers at Systems Development Corporation state that they are feeling no inconvenience in using a multiple character substitute system for an extended language developed for multi-machine use. In fact, they suggest that an extended character set might actually be harder to use and less direct because of a loss in mnemonic value. This is the other side of the coin. It depends upon the character and its meaning. Where it is a specific well-known symbol with a meaning well established in mathematics or logic, it is certainly more direct to use the symbol itself. On the other hand, where it is a general purpose symbol whose meaning has been arbitrarily assigned, its use is not as direct as a less concise multiple character substitute chosen for its mnemonic value.

Directness in a system more often refers to internal processing considerations. There is little question that additional characters available for input convenience also result in a reduction in processing time by virtue of punctuation and ability to clearly distinguish between contextual ambiguities. What kind of an overall saving this results in is again difficult to say.

JUSTIFICATION FOR ADDITIONAL EQUIPMENT COST

One of the most often heard customer comments on enlarging the character set concerns the intangible nature of the resulting gain and the consequent difficulty in justifying any additional equipment expense to achieve it. And this comment comes from those who do acknowledge that there will be some benefit. Many of them are forced to justify additional equipment in terms of payoff over a relatively short period. Where there is an alternative, the justification must be stronger than ever.

At least part of this feeling of intangibility is due to lack of concrete experience with a larger character set. There are only a few individuals who have any feeling for the real part an expanded set will play in an advanced automatic programming system. Undoubtedly, the system will save both programmer time and machine time in arriving at a completely debugged program. But how much of that saving will be due to the use of additional characters? It isn't enough just to say some of it. A dollars and cents figure must be established in a defensible manner. Some solid experience is the only way to evaluate the actual effect additional characters will have.

EXTENT OF PRESENT SCIENTIFIC CUSTOMER INTEREST

At its most recent meeting, SHARE adopted a resolution proposed by its Committee on the International Algebraic Language which "deplored the inadequacy of the presently available limited character set and recognized a growing need for a more extensive set." The resolution stated that "an extended character set will eventually be required and that for an effective implementation of the IAL language an extended set of at least 100 characters is needed now." It concluded by recommending that "IBM consider providing across-the-board input/output equipment to meet this need."

On the face of it, this resolution would seem to indicate a high level of customer interest. What kind of interest is it? During floor debate which resulted in the weakening of the resolution's wording, Mr. W. Ramshaw of United Aircraft suggested that unless the members were ready to order equipment now, the resolution should merely ask that "IBM consider providing" the equipment. His suggestion resulted in an amendment to the resolution. Mr. B. Ferber of Convair, present chairman of SHARE, who presided over the discussion and adoption of the amended resolution, offered the opinion that the resolution was passed as a long range view of the direction in which users are headed rather than as an urgent request for action. He feels that it will be several months before any but a very few users would order even limited equipment, that it is a very indefinite and undefined need for most of the members who voted for the resolution.

There is interest, but it seems to be interest more in a general principle than in a specific need. There are a few advanced thinkers who are convinced that an expanded character set will play an essential part in any new automatic programming system. This is a small group. A slightly larger group is beginning to feel the problem, but hasn't yet developed any strong feelings as to what its solution might be. At this point in time, the second group is willing to be led by the first group. There is a far larger group, probably a majority, who haven't even felt the problem, but don't want to admit it for fear of being thought backward. The members of this third group have at one time or other been slightly inconvenienced by the present character set, so it isn't hard to agree to a resolution stating a general principle as long as it doesn't cost anything. The first group would order equipment if it were available. The second group will have to be convinced of the importance of the extra characters before they can be classed as equipment prospects. The third group will require a demonstration of the benefits of such use before they order anything.

Note: This subject has been discussed with the following scientific customers:

Mr. W. Ramshaw	United Aircraft Corporation
Mr. R. Nutt	United Aircraft Corporation
Mr. B. Ferber	Convair Division, General Dynamics
Mr. I. Greenwald	RAND Corporation
Mr. F. Wagner	North American Aviation
Mr. J. Strong	North American Aviation
Mr. R. Porter	Boeing Airplane Company, Seattle
Mr. W. Bayless	Boeing Airplane Company, Wichita
Mr. F. Engel	Westinghouse Electric
Mr. H. Bright	Westinghouse Electric
Mr. R. Bosak	Systems Development Corporation
Mr. J. Schwartz	Systems Development Corporation
Mr. H. Bratman	Systems Development Corporation
Mr. W. Dobrusky	Douglas Aircraft Company

EXTENT OF PRESENT COMMERCIAL CUSTOMER INTEREST

Most commercial customers are unaware of any need to enlarge the present character set, at least in any general way. Not only that, they don't seem to mind admitting it. Many of them have felt a desire from time to time for a specific special purpose character which they may have RPQ's on a replacement basis. But as far as special characters for problem statement, very little interest can be evoked. It may be a tribute to the authors of the Autocoder system that there is such widespread satisfaction with its use and so little agitation for the development of a more elaborate business language compiler. It is probably also a measure of the somewhat lower level of sophistication among commercial programmers. The field is less rigidly defined; consequently the basis for an advanced compiler does not exist to the same degree that it does in the scientific area of use.

Logic would indicate, however, that the need for direct, concise, and unambiguous problem and data statement is just as great for the commercial as for the scientific user. As a matter of fact, there are factors which seem to indicate that it would be of even more importance. Commercial installations very often have to train personnel from within the company as programmers. This type less often achieves the status of the professional programmer capable of coding in any language equally well. A problem oriented language would aid immeasurably. There is a greater fear of having to reprogram large systems in order to take advantage of technological advances. A better language would make this type of change much less painful. Perhaps these same factors which indicate at least as great a need for a good problem oriented language in the commercial area are also delaying recognition of that need. In any case, and for whatever the reason, customer awareness of any need for enlarging the present character set in any general way is extremely weak among commercial users.

Note: This subject has been discussed with the following commercial customers.

Mr. M. Grosz	Standard Oil Company of New Jersey
Mr. J. Price	Standard Oil Company of New Jersey (formerly with National Security Agency)
Mr. W. Kraegel	Northwestern Mutual Life Insurance Co.
Mr. R. Judy	Boeing Airplane Company
Mr. E. Cooley	Prudential Insurance Company
Mr. J. Savage	Prudential Insurance Company
Mr. R. Prentice	General Telephone Company
Mr. M. Haigh	North American Aviation

APPLIED PROGRAMMING PLANS FOR INTERNATIONAL ALGEBRAIC LANGUAGE IMPLEMENTATION

Applied Programming has started work on an advanced system to implement IAL. There is a widespread feeling that the results of this effort will replace FORTRAN in general customer usage. The question has been asked, "Does IAL represent enough of a change to warrant the effort required for a major new system?" This question misses the point, perhaps because the effort itself is somewhat misnamed. The point is that enough has been learned from FORTRAN and other early systems to result in a significantly better system. IAL is simply a language choice which will help make the new system more powerful and easier to use. Properties of the language should make the result more direct, concise, and unambiguous if the proper input-output equipment is available for use. If adequate equipment is not available and too many arbitrary multiple character substitutes have to be used, there is a serious question of whether or not the gains made by extending the language won't be more than nullified by the awkwardness of its use. For this reason Applied Programming is one of the strongest proponents of support equipment to make additional characters available for use. There is a difference of opinion within Applied Programming, however, as to the extent of the gain which may be derived from additional characters.

INTERIM CONCLUSIONS

For the sake of clarity, it seems worthwhile drawing some conclusions at this point from the foregoing material.

1. Customer opinion at this time is highly mixed on the need for enlarging presently available character sets. A large segment of it is unformed. The usual "leaders" of customer opinion are not in the lead on this question. Attitudes shift suddenly in this unsettled atmosphere. The general direction, however, is toward a realization of the value of an expanded character set. Commercial customer opinion is significantly less advanced than scientific customer opinion.

2. In the absence of actual experience, it is difficult to evaluate exactly how much additional characters will contribute to an advanced language system. Will it be enough to justify the necessary additional equipment expense for the customer? Those most able to answer feel that it will. Some solid affirmative experience is needed at the earliest opportunity to demonstrate this. Information on this experience will help clarify customer thinking as well as our own best course of action.

3. Because of the conditions stated above it is a temptation to adopt a wait and see attitude. However, a positive approach at this time will allow us to exercise a considerable influence in our favor over the form which the demand will eventually take. For the sake of more efficient and profitable use of 700-7000 series equipment, a little leadership on our part on this question seems called for.

IMPLEMENTATION POSSIBILITIES

Before making specific recommendations for implementing an expanded character set, I would like to briefly outline some of the possibilities. The three sections which follow deal with equipment, amount of expansion, and character code representation. This report than concludes with specific recommendations.

EQUIPMENT FOR IMPLEMENTING AN EXPANDED CHARACTER SET

A Reading and Punching Typewriter has been considered for some time as a possible first step toward implementing an expanded character set. This unit would consist of a transmitting and receiving typewriter connected through an adapter-translator to an 024 type keypunch. It would produce hard copy and suitably punched cards for material typed on the typewriter keyboard. It would also be able to list a deck of previously punched cards produced in a computer operation or in an earlier use of this same unit. Suitable control would be arranged in both modes of operation. Punched paper tape might also be used as an input-output form.

The Reading and Punching Typewriter would thus be able to prepare expanded character set coding and data for input to a computer and list expanded character set output from the computer. It could be made available at a reasonably low price as a modification to the Document Writing Attachment for the 024. This is important to the customer because justification for expanded character set equipment will be difficult until experience has shown the benefits to be derived. It is also important to us because the actual value of using expanded character sets can be measured and evaluated with only a comparatively modest investment of development funds. This requirement for as inexpensive a unit as possible on which to demonstrate the value of additional characters through actual use suggests the possibility that the RPT should be considered in two phases. The first phase would be to produce and field test a very modest unit. In addition to allowing a measurement of the unit's value, this phase would

show which features need expanding. The basic unit might or might not be marketed. If the first phase indicates that the value of extra characters is great and that additional features are needed, the second phase would be entered. The second phase would be to produce and market a more elaborate model which would incorporate abilities phase I had shown to be necessary. This two phase process would seem to fit all of the present and foreseeable conditions better than any other.

The Special Engineering Products Division along with at least two other manufacturers has bid on a version of the Reading and Punching Typewriter in which the Los Alamos Scientific Laboratory is interested. The purpose of this unit is to make available for use an expanded set of 123 characters. Los Alamos has been interested in such a unit for quite some time in order to print scientific notation along with tabled results. The latest specifications which they issued in their invitation to bid describe a very elaborate unit. It includes a triple case typewriter, half line spacing up or down, color control, large auxiliary keyboard, arbitrary and unique code representations, paper tape as well as card input-output, and a very large set of control functions. It is probably much too complex, elaborate, and thus expensive for the general customer, at least as a starter. Los Alamos' use of such a unit will certainly shed valuable light on the subject, but more general experience with a much simpler unit is needed at the same time.

The main objections to the Reading and Punching Typewriter center around its slow speed as an output unit. Experience with present compilers has convinced nearly everyone that debugging output must be in exactly the same language as the original problem statement input. There is also a need for more diagnostic output at compile time, again in the same language as input. These two requirements add up to a considerable volume of expanded character set output. This is further complicated by a requirement for minimum elapsed time between main frame run and hard copy in the coder's hands. Because the RPT would be a peripheral unit operating from cards, debugging output would have to be punched out for listing in a separate operation. Some customers say this additional step would be intolerable, that normal operating backlogs in a busy installation would tie any debugging schedule in knots. The fact that a great many very busy installations are using the monitor mode of operation in which debugging output is written on tape for peripheral printing would seem to partially contradict the charge. At least as regards the harm of an additional step. The objection to slower operating speed is certainly a valid one for installations with high and continuous debugging output.

It should be noted that these objections all center around the adequacy of the Reading and Punching Typewriter as a unit for handling main frame output. They are objections to the lack of an expanded-character-set line-printer. (More of that in a moment). The RPT would still have an important function in an advanced expanded-character-set-using installation as an input preparation device, where the term input preparation includes the listing of decks in which changes have been made as well as the punching of original problem statement cards. In a smaller installation and in all installations during the period of experimentation with expanded character sets, the RPT can also serve as a regular output device. The price should be such that multiple units will be practical and desirable. This will offer backup in case of breakdown of like units or of an ECS line printer. At the same time, it will overcome some of the objections to the slower operating speed of the RPT. Another important and continuing use of the RPT would be in facilitating efficient deck maintenance during debugging. Many customers have programming groups located at a distance from the machine room. These programmers often maintain their own decks during debugging. Locating an RPT adjacent to a programming group would make possible more efficient use of the programmer's time in getting correction cards punched and new deck listings made.

To summarize, the Reading and Punching Typewriter represents a modest first step toward implementing the use of expanded character sets. It will enable the customer and IBM to measure and demonstrate the value of such use. It will at the same time help determine the forms such use should take and the features needed to facilitate improved use. It will create a demand for higher speed output equipment. And after it is replaced as an output device by higher speed units, it will still have an important function as an input preparation device.

Productive program life after debugging is typically very short in scientific installations. Some programs are written for a single production run. The majority see very few runs after debugging is completed and prior to revisions which are extensive enough to require major debugging again. As a result, it is not rare at all to find scientific installations in which the majority of output volume is debugging output. Clearly the Reading and Punching Typewriter would be inadequate as an output unit for this type of customer as soon as he had begun using an expanded character set system beyond the experimental state. A line printer capable of handling the same enlarged set of characters would be essential. In fact, these customers cannot

afford to entertain the idea of large scale use of an expanded character set system until assured that a suitable line printer will be available if needed. They'd get themselves in a box if they did. Quite naturally, they're less inclined to experiment with ideas for such a system if they feel that it will lead nowhere.

The chain printer offers the most immediate possibility for an expanded character-set line printer. The standard version divides the 240 character length chain into five identical sectors of forty-eight characters each. Without altering the basic principles of operations, it might be built with four sectors of sixty characters each, three of eighty each, two of 120 characters each, or one of 240 characters. Determination of when a given position prints is made by comparing the representation of the character to be printed with that of the known character on the chain at that point. The latter is determined by a counter which steps as the chain rotates and another which steps according to the print position being compared. The comparisions are made in one set of six bit counters on the standard model. Addition of a seventh bit would allow up to 128 combinations. Since the comparisons are made serially by printing position, the amount of extra circuitry is not as much as might be expected to accomplish a two-fold expansion of the character set. A larger portion of the chain must pass each print position before the line is complete; consequently, lines-per-minute speed is reduced in approximate proportion to the size of the character set. This reduction in speed is the main cost of the additional characters.

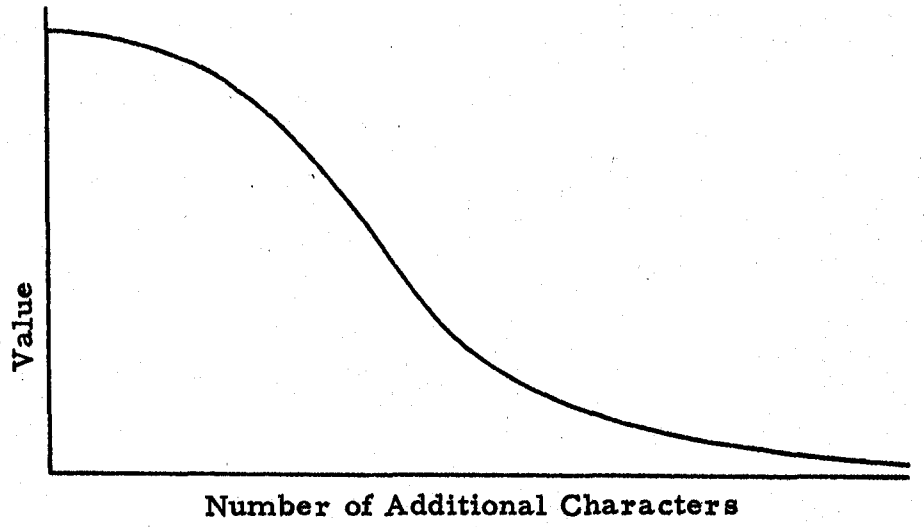
No further equipment modification is necessary in order to allow the binary scientific customer to make extensive use of an expanded character set. The present column binary attachment for the Type 714 Card to Tape and the Type 722 Tape to Card Units will allow peripheral reading and punching of expanded character set cards. Translation to and from whatever internal representation is desired is then relatively direct. The choice of external representation is the subject of a later section.

A modification to the column binary attachment may be necessary for expanded character set use in the commercial machine area. The reason for this is discussed in the section on external representation. In brief, the modification would enable a special readout of the buffer to separate internally unallowable code combinations into a pair of allowable combinations.

AMOUNT OF EXPANSION TO BALANCE COST AGAINST GAIN

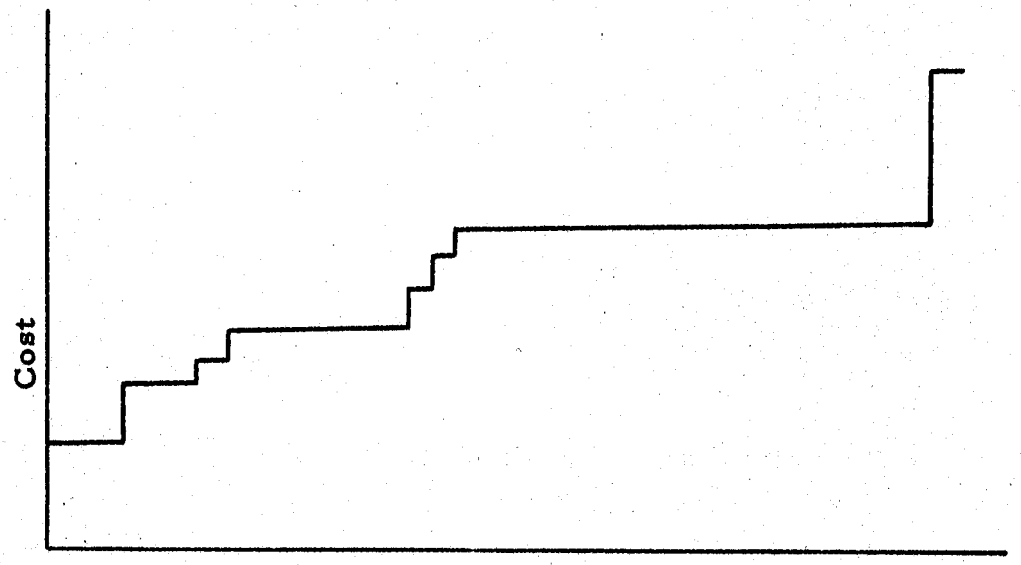
Stated another way, at what point does the cost of adding an additional character exceed the expected gain? It has already been stated that some experience is needed to determine the actual value of any additional characters. This is certainly even more true for determining the relative value of different amounts of set expansion. In spite of this, enough is known to make a certain amount of speculation worthwhile. For example, what do the curves representing the value of additional characters and the cost of additional characters look like and where might they be expected to cross? The value of a character is measured in terms of its effect in making the compiler language more direct, concise, and unambiguous, as well as easier to use. The difficulty of translating this effect into dollars for comparison with the cost curve is admitted. It is a function of the particular customer's policies among other things. Therefore, discussion of it here will be in general terms only.

The value of the first few additional characters is thought to be quite high. These would be well known and constantly used symbols. After ten or twenty of these, the relative value of additional ones begins to fall off. Beyond some point, it will be very small. The shape of this curve is probably something like the following:



The omission of any scale is deliberate. It would certainly vary from one customer to another. It will also be a function of the compiler itself.

Well before the tail of the value curve is reached, it has fallen below the curve which represents cost of the additional characters. Indeed, the possibility is admitted that the value curve is below the cost curve for the entire area beyond the present forty eight characters. The cost curve is probably of the following shape:



Vertical scale for this curve is unknown at this time. The points at which the steps occur are known along with some idea of the relative size of the step. The two curves probably cross at one of these steps. If they do not cross at a step, we are interested in the next higher step anyway, since the intervening characters can be provided at little or no additional cost.

It will, therefore, be instructive to examine the location and approximate size of these steps. They are points at which the hardware must take a jump in order to include even a single addition character. Sixty-four is a very important one of these. Six binary bits will represent uniquely at most sixty-four characters. All of our present magnetic tape equipment is six channel in width. Certain internal operations are based on six bit groups. Fortunately, neither of these considerations is at all disabling. With some inconvenience and possibly some loss of efficiency, larger bit groupings can be manipulated correctly. This is especially true for pure binary machines, which have fewer arbitrary internal groupings. On tape a larger group of bits occupies more than one "horizontal" row. Sets of these can still be packed on a binary machine at least. Succeeding powers of two cause steps at 128 and 256 characters. These are not as serious or costly as the step at sixty-four. There are very few proposals to go beyond 256 characters. The STRETCH computer has a basic internal grouping of eight bits, which allows for 256 unique combinations.

The Reading and Punching Typewriter is responsible for two more steps. Use of any other than a standard typewriter would obviously be more expensive. The standard two-case forty-four key machine allows for eighty-eight characters. A three-case forty-two key machine which the ET Division is considering for SEPD would make available 126 characters. This machine will cost more per character than the two-case machine, and will itself represent another step. Versions of a four-case machine exist, but problems with its double basket probably rule it out for use as an RPT.

The possible subdivisions of the chain on the chain printer also cause steps in the cost curve. When the number of characters in each subdivision is enlarged, the speed of printing is slowed and the cost per line of printing rises. This is based on an assumption of full prime shift printer use which isn't necessarily the case. In general, however, slowing the lines-per-minute speed of the printer will tend to increase the cost per line. The steps caused by the chain printer are the factors of 240: 48, 60, 80, 120, and 240 itself. Beyond each of these points the cost of printing can be expected to rise.

Grouping the three causes, the successive steps beyond the present forty-eight characters occurs as follows: 60, 64, 80, 88, 120, 126, 128, 240, and 256. Because of the obvious desirability of providing the same set of characters on an RPT and a line printer, attention is drawn to the 80 and 120 character regions where there are groupings of the steps. The steps at 80 and 120 characters are caused by the chain printer. It has been admitted that these are not steps at all under certain circumstances. In spite of this, they are chosen for major consideration because they offer compatibility at the two levels. Further study will probably show that the value and cost curves cross at one of these two points. In discussions with knowledgeable customers, "around 100" is a frequently heard desirable size for an extended character set. Actual experience may show that the value of additional characters is not high enough to pay for a set of 120 characters. In this case, multiple character substitutes could be used for everything beyond eight characters. This would still provide a healthy increase over the present forty-eight.

There is a reasonably high degree of unanimity among the scientific users on the question of which additional characters are needed most. The IAL reference language provides a guide. If it appears advisable later to limit the size of the set to eighty characters. a controversy can be expected over the relative importance of lower case alphabet on the one hand and additional special symbols on the other. Information on type and relative frequency of use can be gathered on this question in advance for use if it needs to be settled.

CHOICE OF CHARACTER CODE REPRESENTATION FOR AN EXPANDED SET

There are a number of conflicting criteria which must be considered and given appropriate weight in making a choice of code representation for an extended character set. Those which take precedent into account are considered by many to be the most important. They are also the most troublesome in some respects. But as attractive as it might seem to make a fresh start, precedent simply cannot be ignored. Doing so very often results in inconvenient use in a total sense. It is an example of sub-optimization. Therefore, the following criteria are advanced as being of prime importance.

1. The present Hollerith characters must be retained as a useable subset.

2. The present Hollerith codes must be retained as a useable subset.

3. The present commercial collating sequence must be retained.

Also of considerable importance are the criteria which relate to convenience and efficiency of use. With the state of present hardware technology what it is, there is a strong customer feeling that convenience of use should never be subordinated to convenience of hardware fabrication unless significant savings can be made thereby. The following criteria are in order of importance as to their effect on convenience of use.

4. Degree of computer acceptance. Complexity of conversion to internal code system. Practical useability on as wide a range of machines as possible without compromising the efficiency of the most powerful.

5. Degree of input acceptance. Peripheral readability. Need for any special attachments.

6. Conciseness. Number of card columns used.

7. Fixed field length convenience. Same number of columns used for Hollerith and non-Hollerith. Ease of making changes to cards.

8. Degree to which the set can be further expanded in a direct and consistent manner.

In general, customers are quite willing to let us decide which choice will best satisfy these criteria. The 704-709 users do feel very strongly that the choice should not compromise binary machine use in order to include non-binary machines. This is the second part of the fourth point above.

One of the first proposals for extending the code was to use non-Hollerith punch combinations for non-Hollerith characters. This satisfies 1 thru 3 and 6 thru 8 above, but falls short of 4 and 5 in that, without equipment modification, it would be useable on binary machines only and would require a special attachment to the 714 unit. A second proposal suggests using two Hollerith codes to represent non-Hollerith additions to the set. This retains the present set as a subset and so satisfies 1 thru 3 above. It also satisfies 4 and 5, but falls short on 6 thru 8. It is not as concise in that it uses two card columns for non-Hollerith characters. Its use of a variable number of columns would make corrections very difficult. Finally, its further expandability is limited without taking out of use additional Hollerith characters for identification of non-Hollerith. The use of a variable number of columns is the most serious shortcoming. This would be disabling under some circumstances.

To overcome this disadvantage, a proposal was put forth to use two columns to represent both Hollerith and non-Hollerith characters. This violates 2 above but in a minimum way if a blank is used for the second code with Hollerith characters. It satisfies 1, 3, 4, and 5 reasonably well. It meets criteria 7 and 8 very well. But it fails badly in satisfying 6. All characters take two columns. A problem statement using all Hollerith characters would take twice the card columns it should take to express. A related proposal is to use a two column numeric code for Hollerith and non-Hollerith. This violates 2, 6, and 8 very badly. It compromises binary machine efficiency of use badly in order to ease problems of use by numeric machines.

A fifth proposal comes very close to satisfying all eight criteria. It is actually a combination of the first and second proposals. Hollerith characters would be represented by the standard Hollerith code. Non-Hollerith characters would be represented by two superimposed Hollerith codes chosen for uniqueness. Again, 1 thru 3 and 6 thru 8 are satisfied very well. Numbers 4 and 5 are satisfied in a compromise fashion. Cards of this type can be read and punched with column-binary equipped 714's and 722's. On magnetic tape from this source, or read on-line into a 704 or 709, they can be converted into whatever internal code is desired. For the non-binary main frame, a modification of the column binary attachment will enable the 714 to separate the superimposed double Hollerith into two legitimate Hollerith characters on readout. The separation will be defined and the superimposed Hollerith chosen so that the results are unique.

This choice avoids penalizing the 704-709 user, and does it in a fashion which does not exclude the 705-7070 customer. It will require a later piece of equipment modification for eventual use by commercial customers. But it appears to be a solution which takes into account the realities of the present situation. The only interest now is in the 704-709 area. An inexpensive first approach is needed to test the value of additional characters and build interest. It needs to satisfy binary machine use as much as possible to do this. At the same time, possible later use in the commercial area must be kept in mind.

RECOMMENDATIONS

With the preceding discussion and conclusions as background, I submit the following recommendations on the course of action which we should take.

1. Build a pair of Reading and Punching Typewriters based on a reasonably minimum modification to the Document Writing Attachment of the 024.

2. Field test the first of these in the Applied Programming Department's IAL Implementation Group. Field test the second with the aid and cooperation of selected customers who are interested in helping measure the value of additional characters and who have some plans for experimenting with them.

3. Evaluate these field tests very carefully in order to measure the value of expanded character set use, adequacy of the RPT concept, additional features needed on the RPT, and other related questions.

4. Collect and disseminate at the earliest opportunity results and conclusions drawn from field testing the basic RPT's. A greater flow of reliable information will help clear up much of the confusion that now exists on the question of expanded character sets.

5. Prepare specifications for a more advanced model of the Reading and Punching Typewriter if the field tests show the value of such use and the soundness of the basic concept, but indicate that additional features are needed.

6. Market either the first or second model of the RPT, depending on the results of the field tests, the success of Applied Programming in implementing IAL, and the development of a favorable customer attitude toward the use of such a device.

7. Intensify investigation of the feasibility of higher speed line or page printing of an expanded character set. Determine at the earliest opportunity answers to such questions as the following:

   a. What modification is necessary to the chain printer itself to allow it to print a set of 80 or 120 characters?

   b. Which of the various control units proposed for the chain printer can be modified to accept an expanded character set?

   c. How extensive would such modifications have to be? Will a special control unit be needed?

   d. Can on-line printing of the enlarged set be achieved for the 704-709?

   e. Can line printing of expanded-character-set card-input be achieved with a modification to the SPACE chain printer configuration?

   f. What are the possibilities of expanded-character-set printing with the magnetic printer? The shaped-beam CRT printers? Others in development? What are the limits?