SEISMOLOGICAL SERIES

of the

DOMINION OBSERVATORY

#50

1967-2

Computer Hardware and Programming Requirements for the Delay-Sum-and-Correlate Method of Processing Seismic Array Data

D. H. Weichert
1967

Seismological Service
of Canada

OTTAWA, CANADA

Department of Energy, Mines and Resources

OBSERVATORIES BRANCH

1967

ABSTRACT

The data from the Yellowknife seismic array have been processed digitally in Ottawa since early 1966. The method used was the sum-delay-and-correlate method and required approximately 16,000 words of computer memory when conducted at twice real-time speed, because of trade-offs between memory and processing speed. For the particular configuration of the Yellowknife array, and for a sufficiently uniform crust, it is demonstrated that, with an average third generation computer with about 4 μsec addition time and 24-bit word length, the data could be processed continuously at four times the recording speed. About 100 beams would be maintained, covering uniformly the third zone.

The routine formation of digital files for all detected events is not possible at four times the recording speed, unless more than one direct-memory-access channel is incorporated in the digital computer system. With two such channels, approximately one digital tape would be required per recording week. With only one channel as presently approved for purchase, digital tape formation for international data exchange is possible, but at twice real-time speed the tape utilization is only 18%.

INTRODUCTION

A medium aperture seismic array has been operated at Yellowknife, Canada, since 1963. The data from the array have been processed digitally on the CDC-3100 computer of the Department of Energy, Mines and Resources in Ottawa since early 1966. The construction of the array and its response characteristics, the processing method, the seismological results obtained to date, and the relevant computer program logic have been described in earlier papers (Manchee and Somers, 1966; Somers and Manchee, 1966; Weichert, Manchee and Whitham, 1967). The purely technical requirement of forming a relatively large number of delay-and-sum beams and a non-linear output for each beam has led to some interesting results concerning the relationship between the requirements of fast access computer memory, machine speed and programming methods. The purpose of this note is to describe the experience from one year of digital operations and the requirements which have governed the selection of a new processing facility, which is being installed at the Dominion Observatory in the late spring of 1967. In view of the current interest in the field of advanced seismological data exchange for detection purposes, technical considerations on the exchange of digital data are discussed.

COMPUTER HARDWARE AND PROGRAMMING REQUIREMENTS

The Yellowknife Array (YKA) consists of two lines of ten equispaced seismometers each, which run in the NS and EW directions (Figure 1) and cross at a common seismometer. The 19 seismometer signals, a coded time channel and two auxiliary seismic channels are recorded on FM tape in Yellowknife. In the Ottawa facility the tape is played back at twice the recording speed, the signals are filtered and then digitized on line at the rate of 20 samples per recorded second. Thus one sample of each channel is fed to the computer every 25 msec. The signals are added with different time delays to form up to about 150 beams, covering the third zone to Yellowknife in a systematic way. This results in a beam overlap at approximately the 3 db level, or slightly higher, for teleseismic frequencies. In addition to the summation, a non-linear output is formed for each beam by multiplying the phased sums of two lines and averaging the product over time with a digital RC filter.

The requirements of computer memory will first be considered. It will be shown later that an appreciable part of memory is needed for the actual summation program in order to achieve the necessary processing speed. Independently of this trade-off between memory and speed, there exists a basic storage requirement which depends only on the length of the necessary delay lines, and thus on the array configuration and the lowest velocities which are considered of interest. The simple-minded approach, of course, is to store signals for the necessary length of time and to form the sums by tapping the delay lines at the appropriate points. In another method, the incoming signals are added into the proper taps of the delay lines, which now hold the partially formed sums (Briscoe and Fleck, 1966). The optimum choice between the two methods depends completely on the relative number of seismic channels and beams which are to be formed. For the 19 YKA seismometers and the greater number of beams, the first arrangement is preferred. The minimum delay line length for each channel can be calculated from the distance of the seismometer

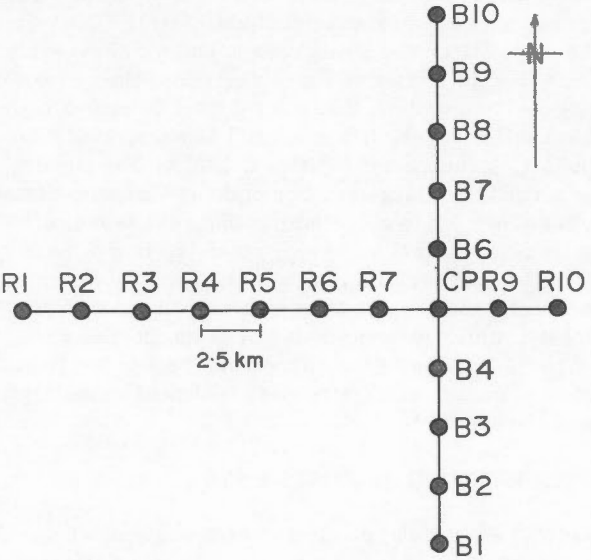

Figure 1.

The Yellowknife array, Northwest Territories, Canada. CP is at 62°29'34.3"N, 114°36'16.5"W, geographic. The mean bearing of the Blue and Red lines are 360° and 90° true respectively.

to the reference point and the parameters of the beam to be formed. However, for the general set of beams used to cover the third zone around the array, the proper time delay must be applied to signals from opposite directions, thus doubling the minimum line length for each seismometer. Although delay lines for the interior seismometers may in principle be made shorter than those for the extreme ones, it can easily be seen that the extra time needed for the handling of the incoming data and the different time pointers, which would be needed for these unequal delay lines, is prohibitive. All delay lines are therefore made equal to the longest one required by the particular set of velocities. Although the optimum reference point is about 1 km north of the R6 seismometer, the centre point has been chosen as reference. The longest delays are then necessary for R1, and the total memory requirement for a minimum velocity of 10 km/sec becomes 20 samples/sec x 21 channels x 2 x 7 x 2.5 km/10 km/sec = 1470 memory locations.

The centre point seismometer is included into both line sums, and the 21st channel holds the time track. The minimum velocities should, however, preferably be lower than 10 km/sec in order to undertake research on local events and to guard against triggering on their side lobes when scanning for third zone events. The calculated memory is thus a minimum.

In the simplest case the delay lines, which have been described, act as a form of ring buffer: the location of the zero time point moves from one end of the storage area to the other and is reset periodically. The taps for the different beams will be either above or below the zero point, although they are at predetermined and fixed locations relative to zero time. Their absolute address must be determined before each addition of a signal sample into the phased sums, using address arithmetic. To achieve this, the sum of relative address and time pointer must be formed and compared to the delay line length: if too large, it must be decremented. If only one arithmetic register is available, the partial signal sum must be stored and reloaded. Several times the basic addition time would be wasted in this way. If the absolute address of the signal sample to be added can always be kept on the same side of the zero time point, the address modification can be achieved by simple indexing. This requires usually only a fraction of a machine cycle and often is done without any increase in the addition time. The incoming data can easily be arranged to allow this simple indexing, if the reserved data area is doubled. Incoming data are at once duplicated into the lower half of the respective line and the zero-time pointer moves only over the inner two quarters of the lines. For the above illustration with the lowest velocity chosen to be 10 km, the signal storage area required is now 2940 locations. If local earthquakes and shear waves are to be studied, an 8000-word memory module would be filled with data alone.

Packing of two signals per computer word has been considered, but the recovery of either 12-bit signal word requires one free arithmetic register and takes about as long as the addition of the signal to the sum.

Up to this point in the argument, computer memory has been repeatedly sacrificed in order to increase processing speed, tacitly implying that speed is more important. The design has required for practical and economic purposes that the processing be done at least at twice the recording speed, and that between 100 and 200 beams must be formed to take advantage of the practical and demonstrated array resolution. This allows 25 msec per calculation cycle, or about 100 to 200 μsec per beam. Some of the time will be spent on housekeeping, on input/output, etc. Since one beam for the array uses 20 signals, one addition together with its share of load and store, multiplication of line sums and averaging must not take more than 5 to 10 μsec. With a computer having an addition time of the order of 4 μsec, it is therefore desirable to avoid address arithmetic, packing and unpacking, or additional unnecessary indexing.

All these tricks except the last possibility have already been considered. However, additional indexing can immediately be dismissed, since the predetermined relative addresses of the signal to be added-up are stored in tabular form and would have to be accessed by indirect addressing, which requires an additional machine cycle. The octal machine operation code is therefore inserted before each relative address in the table, and the necessary load, store and jump-to-correlate instructions are inserted into the table, which is now a self-contained summation subprogram, written in extensio. For 150 beams and 20 channels, this program would occupy about 3000 locations, if no use is made of symmetry.

In the above paragraphs the processing of data from a seismic array with a general geometry is described and with the Yellowknife configuration used only as an illustration. For linear arrays the partial sums are symmetric about the lines, and if the set of beams to be formed is arranged properly, the summation program can be cut by almost a factor of two, both in memory occupied and in execution time.

Even greater savings can be obtained if the signals from either line are velocity-filtered independently, and the outputs then cross-correlated in all possible pairs. If the crust is uniform along the line of seismometers, and the seismometers of each group are precisely in line, the processing is optimum; if the geometry of a line is slightly 2-dimensional, the response will drop. If now, in addition, the seismometer separation in each line is constant, further simplifications can be introduced: velocity filtering for either line is performed by adding the channels with a constant delay between them. The tap positions, or add-addresses, no longer have to be precalculated and stored. Moreover, since one search condition is now obtained by cross correlation (one multiply and one usually exponential integrator) of two different velocity filter outputs, only $2\sqrt{N}$ filters are needed instead of 2N, as in the case of a general array configuration. Thus, for 169 conditions only 26 x 10 signals have to be added for YKA, instead of 169 x 10 x 2. The significant part of computing time now goes into multiplication and integrations.

Most of the arithmetic should naturally be done in fixed point format. Even if floating point hardware is incorporated into the central processor, floating point additions would be about 4 times slower than fixed point single precision additions (these figures for CDC-3100). This points out the importance of a thorough evaluation of the required dynamic range of the processing system, and a study of any necessity of using double precision integer arithmetic for some operations, with the associated decrease in processing speed. The basis for the evaluation is the analogue-to-digital input range and the seismometer sensitivity. The YKA FM tapes are nominally recorded with 10 mV per mμ/sec of ground velocity between about 0.9 and 6 cps, and the A/D converter codes ±5 volts into a 12-bit word. The 1 mμ amplitude of a signal of 1 cps would therefore appear in the machine as about a 6-bit magnitude. After summing 10 coherent signals this increases to 9 bits. Multiplication leads to 18-bit magnitudes and averaging over about 1.6 sec of record time to 23 bits. The maximum coherent input of 5 V results in a 34-bit magnitude. Thus, for sumall processing alone, a machine with a 17-bit word (including sign) would suffice if the dynamic input range is to be carried through the system regardless of its significance, while for the non-linear processing described earlier, about 35 bits are needed. If the 1 mμ signal is used to define the lower limit of dynamic range of the non-linear output, the line sums could be decreased 5 to 6 bits before multiplication and a 24-bit computer word could be used. It is, however, easier and faster to discard 11 bits of the product.

INTERNATIONAL EXCHANGE OF PRIMARY ARRAY DATA

While all standard processing of array data at the Dominion Observatory installation uses the original FM analogue tapes, an international exchange of data in analogue form is objectionable. The most convincing reason is the format of the

1-inch 24-channel FM tapes: very few installations only are set up to replay and digitize these tapes. For an exchange of data between these few organizations, FM tapes could, of course, easily be copied; but this results in a decrease in quality. A more attractive alternative is an exchange of digital magnetic tapes.

The problems connected with most digital data are usually those of "compatibility", often modified by three letters as "IBM compatibility", but the meaning of the expression is rarely well defined. It should not be interpreted to mean that a tape written on a CDC-3100 or DDP124 system can be read by any IBM user with a simple READ (u,i) or READ (u) Fortran Statement; nor should it mean that the data block length is necessarily one which is allowed by an IBM operating system or job supervising system. It does mean, however, that information on the tape can be read completely and without error by an IBM tapedeck and transmitted to any IBM computer, even though this computer may be 8-bit byte, rather than 6-bit-character oriented, and the transmitted code may be meaningless without some form of bit manipulation.

Even for this definition of compatibility, a few questions arise. Partly because of a preference for 8-bit bytes, and probably also for other reasons, the 9-level digital magnetic tape having 8 bits of information and 1 lateral parity bit per line has been introduced. Although the history of electronic data processing suggests that we will eventually write on 9-track tape, there are presently more 7-track tape drives in use than 9-track. In installations like the Canadian Government's Central Data Processing Bureau with its IBM 360/65 system, only 2 drives out of 8 are 7-track, but one of the two tape control units contains a data conversion feature, permitting the conversion of three 8-bit bytes to four 6-bit characters and vice versa. For the present, it seems desirable to use 7-track tapes for international exchange: conversion difficulties are not foreseen for some years.

Information on digital magnetic tapes is usually written and read in either of two modes, binary or BCD. In the binary mode, information is written onto the tape as it comes over the data channel from core storage. There are computers of various word lengths and the width of the data channels is usually closely related to the word length. Whatever the length of the transmitted byte, it must be stored by the tape controller, and then be sequentially written on tape in groups of 6 bits. The 7th track is filled with a controller-generated parity bit which makes the total of bits per line an odd number, in this mode. At the end of each block of data the tape controller automatically writes an additional parity line, spaced a little farther than the data (e.g., 0.0075 inch at 556 bpi). This line generates even longitudinal parity for the whole of the data block. There are no hardware restrictions to the length of a block and everything stated above for the "WRITE" operation is also true for the "READ". Thus the only true compatibility for digital seismic data tapes concerns the physical tape dimensions, spacing of tracks, magnetization level on the tapes and similar features. To the best of our knowledge, all major manufacturers of tape units are compatible with regard to these basic hardware details.

Once the binary data have been read by a tapedeck and transferred over a data channel into a computer core, compatibility is a completely different matter. In the system which we have used to the end of 1966 for processing Yellowknife array data, the A/D converter produced a 12-bit word, and coded the range of ±5V in ones-complement notation. Since the central processing unit (CDC-3100) also uses ones-complement notation, it was only necessary to extend the uppermost bit before performing arithmetic. Had the same A/D converter been attached to a digital processor such as the DDP124, the notation would have had to be changed to sign-and-magnitude. Many A/D converters produce twos-complement numbers, and at least one manufacturer even includes both ones-complement and twos-complement add-instructions in his repertoire. To change from one to another of these representations of integer numbers is trivial. If, however, a tape that was written in ones-complement notation in our installation was read into an IBM 1620, the data would have to be completely recoded for use on that decimal machine.

Very little will be said about BCD coding. In this mode the lateral as well as longitudinal parity, which is generated and checked by the controller, is even, but in addition the binary numbers coming from core memory are converted according to some code. Since the same 7-track BCD magnetic tape code is used by different manufacturers, alphameric information written by one company's system will be interpreted correctly by other systems, except for a few special characters. However, an example of what would happen to some numeric data if written in non-formatted BCD is the following: the octal sequence 12345670 would be coded on tape by a DDP124 as 12342070, representing the 4 alphameric characters 0(H . If read back by a CDC-3100 system, these 4 alphameric characters will be interpreted correctly, but numerically they will now appear in core as 00746030. Thus, tapes written in BCD have their main use in recording alphameric information and small amounts of numeric material written according to a certain FORMAT statement. If, for instance, file or tape description labels are found to be desirable, these could well be written in BCD. On the other hand, because of the difficulty of updating such labels without copying the whole tape, a simple numeric identifier record with a minimum length of 12 characters should preferably precede the seismic data file and a detailed description of the data could more easily be kept elsewhere.

Another unanswered question is the best block length. Apart from the short identifier, blocks should ideally be long enough to utilize the tape to a high degree, thus cutting distribution costs. A standard block length of 136 (6-bit) characters as used in some operating systems gives only about 25% tape utilization (at 556 bpi density), because each 0.24 inch of data is followed by a 0.75 inch inter-record gap. For a data exchange service, blocks of about 4000 to 8000 characters should be chosen. This gives about 90% to 95% tape packing, while the blocks are still small enough to be handled by most systems. The length should in any case be a multiple of 8 characters, since not all machines incorporate character-oriented input/output instructions, or flexible options to control the numbers of characters to be transmitted per word.

Some interesting numbers emerge when one considers a small computing system in which only one data input/output channel is available, such as the special purpose configuration now approved for the Dominion Observatory. Since relay responses are such that the FM tape cannot be stopped while a block of digital data is written out and then restarted, the block length must either be short enough to be written within the 25 msec (twice real-time) or 50 msec (real-time) digitization interval, or long enough to hold a complete seismic event. However, for the Yellowknife array, a computer memory of 16K (24-bit) words holds only a little more than 1 minute of data, and in the practical situation it will thus be necessary to write the digital tape in smaller blocks. In the extreme case, which has been used up to now, one record contains one sample of the 21 channels, leading to a tape utilization of less than 10%. If an allowance of 30% is made for tape start and stop times, the maximum achievable tape density is obtained with the described system running at twice the recording speed when the write time is nominally 25 msec - (0.75 inch inter-record gap/ 45 inch/sec) x 1.3 = 3.35 msec, corresponding to two samples of all channels per record, and a tape utilization of about 18%. It should be pointed out that the A/D input can be handled during digital tape deceleration. If the FM playback speed is decreased to real time, the tape utilization increases considerably to over 60%. Finally, on the basis of these figures, an estimate of the tape consumption for a routine digitization of data for events detected by the Yellowknife array can be made. If ten events per day are recorded for 5 minutes each, at real-time speed, between 600 and 700 feet of digital tape will be written each day.

Thus 50 minutes of data only out of 24 hours can be made available in digital library form with a length of tape approximately 30% of the length of the primary analogue FM tape. The economics of the process requires, therefore, considerable thought for permanent storage, as distinct from compatible exchange. In addition, unless the process can be automated at the time of detection, during the high speed search, an additional two hours or so of computer time each day will be required to relocate events and form library digital files.

CONCLUSIONS

It has been found that, using the simple sumall and correlator method, about 100 to 200 different beams are considered useful for continuous monitoring and automatic epicentre determination of seismic events in the third zone to the Yellowknife seismic array. In order to guard as much as possible against triggering on sidelobes of low velocity events, the lowest velocities which should be included in the beams are at most 8 km/sec. To implement such a search program at twice real-time speed on a digital computer with about 4 μ sec addition time, the necessary speed is achieved by a wasteful use of computer memory for signal storage and in-extensio programming, and requires about a 16K word memory.

If, instead of a general array geometry, the sensors are grouped in lines, the symmetry of seismic energy arrivals with respect to these lines leads to savings in both computer time and memory.

If the sensors are equispaced along the lines, and the crust in the vicinity of the array is sufficiently homogeneous, sumall and correlator processing can be further simplified. The addition time of the computer then loses its importance and the time for the multiplication and averaging dominates. If the computations can be done in single precision integer arithmetic, which requires about a 24-bit word in order to exploit the dynamic range of the Yellowknife recording system, or if fast double precision instructions are part of the computer's repertoire, a continuous search of the FM tapes can be made at four times the recording speed, with only a small reduction in the number of beams: about 120 beams are estimated to be possible. This still corresponds to a coverage of the third zone at approximately the 3 db response level, depending on frequency.

On the other hand, with the single direct-memory-access system presently approved for purchase by the Dominion Observatory, it is not possible to prepare digital tapes of selected events on a routine basis directly after triggering during this fast search. In order to form digital files of all events, it is necessary to relocate the events on the FM tape at a later time and digitize them, using a slower playback speed. Based on a 45 ips digital tape speed and making a 30% allowance for acceleration and deceleration while spacing over inter-record gaps, the digital tape can be 18% utilized when digitizing at twice the recording speed; in real time a utilization of about 63% becomes possible.

Based on this percentage and estimating an average of ten events per day and a record length of 5 minutes per event, a digital tape usage between one third and one quarter of the length of the initial FM tape can be estimated. Quite apart from the question of international distribution of seismic data in digital form the economy of preserving data in digital or in analogue libraries may therefore favour digital files, unless the present estimate of the average number of events turns out to be grossly underestimated.

REFERENCES

Briscoe, H.W. and Fleck, P.L. 1966. Proc. Spring Joint Computer Conference, Amer. Comp. Soc., Washington, D.C.

Manchee, E.B. and Somers, H. 1966. Publ. Dom. Obs., Canada, 32, 2.

Somers, H. and Manchee, E.B. 1966. Geophys. J., 10, 401.

Weichert, D.H., Manchee, E.B. and Whitham, K. 1967. Geophys. J. 14.

ROGER DUHAMEL, F.R.S.C.
QUEEN'S PRINTER AND CONTROLLER OF STATIONERY
OTTAWA, 1967