Acknowledgment

The problem generalization and analysis in this paper was inspired by the work of Dr. V. Radeka, Institute "Ruder Bošković," Zagreb, Yugoslavia, who considered the speed limits of decimal counting schemes based on the binary-to-decimal conversion and described an "ideally fast" decimal logic.

References

[1] V. Radeka, "Fast decimal counting with binary-decimal logic," *IEEE Trans. on Nuclear Science*, vol. NS-11, pp. 296–301, January 1965.
[2] H. Stopper, "Ermittlung des Codes und der logischen Schaltung einer Zählerdekade," *Telefunken Ztg.*, vol. 33, pp. 13–19, March 1960.
[3] Z. Tarczy-Hornoch, "Five-binary counting technique makes faster decimal counting units," *Electronic Design*, vol. 9, pp. 34–37, January 1961.
[4] W. Hilberg, "Die Eccles-Jordan-Schaltung mit Transistoren für sehr schnelle Zählen," *Telefunken Ztg.*, vol. 33, pp. 98–108, June 1960.
[5] W. Hilberg, "Das Zählen sehr schnell aufeinander folgender Impulse," *NTZ*, vol. 17, pp. 24–34, January 1964.
[6] J. A. Goss, "ISABEL," *Electronic Engr.*, vol. 32, pp. 630–634, October 1960.

A Critical Comparison of Two Kinds of Adaptive Classification Networks

K. STEINBUCH AND B. WIDROW, MEMBER, IEEE

Adaptive pattern-recognition systems have been under study since 1959 at the Technische Hochschule Karlsruhe, Karlsruhe, West Germany; in the Department of Electrical Engineering at Stanford University; Calif., and elsewhere. This note results from a visit to Stanford by Dr. Steinbuch in May, 1964. The note summarizes the results of discussion in which the authors compared the "Learning Matrix" developed at Karlsruhe [1]–[4], with the "Madaline" developed at Stanford [5], [6]. In several cases, conclusions were drawn on experimental evidence only. These conjectured results have value, it is believed, because they give a certain insight based on experience, and at the same time they represent unsolved theoretical problems.

The basic structures of the Learning Matrix and the Madaline systems are shown in Fig. 1. The input signals (or patterns) may be binary or nonbinary. The output signals of both systems are binary (except in certain situations where the Madaline is used with multilevel quantizers producing ternary or other forms of output, or where no quantizers are used at all, producing analog outputs). For purposes of comparison, let the number of input-signal parameters to both systems be n, and let the number of pattern classes (output categories) be m. Both the Learning Matrix and the Madaline have n input lines; the Learning Matrix has m output lines, while the Madaline has k output lines, where k is the smallest integer equal to or greater than $\log_2 m \triangleq \mathrm{ld}\, m$.[1]

$$(1 + \mathrm{ld}\, m) > k \geq \mathrm{ld}\, m. \tag{1}$$

The Learning Matrix and the Madaline are structurally quite similar, forming arrays of output signals from linear combinations of input signals.[2] They differ primarily in their training methods and in their output logic. The training of the Learning Matrix is on a one-pass basis, while the training of the Madaline is an iterative process in which the individual training patterns may be repeated several times until adaptation[3] is completed.

The various training procedures for the Madaline are described

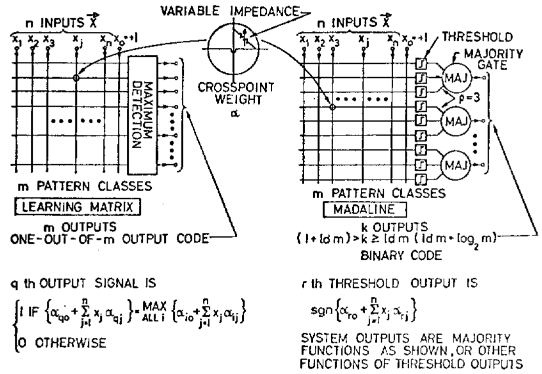

Fig. 1. Comparison of structures of Learning Matrix and Madaline.

in [5]–[7]. In the Madaline system each adaptive linear threshold logic element (called an "Adaline") may be adapted by a least-mean-square-error (LMS) procedure developed by Widrow and Hoff [5]; by a procedure due to Ridgway, described by Mays in [7]; or by still other procedures [7].

The various training procedures for the Learning Matrix are described in [1]–[4]. In the Learning Matrix, each linear weighting filter (represented by a horizontal line on Fig. 1) to which a pattern is matched is associated with a pattern class. Each pattern class has its own output line, so that the response is provided by a one-out-of-m code. The actual output selected is determined by a peak detector (minimum Hamming- or Euclidean-distance method).

In the Madaline system, the pattern class is indicated by an array of binary-coded output signals. Each output signal could come from a single threshold logic element connected to the inputs. Such an element could cause a space of input pattern points to be sectioned by a hyperplane [8]. Alternatively, each output signal could come from several such elements whose inputs are connected in parallel and whose outputs are combined in a *fixed-logic* element such as a majority element, an OR element, an AND element, a parity element, etc. The number of paralleled Adaline units used to provide a single output bit is represented as ρ. Thus each output bit is generated by sectioning the input space with ρ hyperplanes. In the system illustrated in Fig. 1, $\rho = 3$, and the outputs are based on majority logic. Training algorithms for Madaline systems using majority and OR output logic were devised by Widrow and Hoff [5]. The possibilities and advantages of paralleled rows, especially by OR combinations, have been considered for the Learning Matrix as well.

Both Learning Matrix and Madaline systems have been realized by adaptive hardware and by digital-computer simulation. It is felt that, although digital-computer circuitry and technology are vastly more advanced at present than adaptive-computer circuitry and technology, many future developments in data-processing systems for pattern discrimination will involve the use of physically realized adaptive circuits. The comparisons in this note between the Learning Matrix and the Madaline with regard to the amount and types of components, training time, and sensing time are based on realization by adaptive hardware rather than by simulation on a general-purpose digital computer.

The analog weights at the crosspoints of the Learning Matrix and the Madaline, as shown in Fig. 1, have been physically realized by means of a variety of circuit elements. At Karlsruhe, these analog storage functions have been performed by transfluxors and by square-loop toroids with beat-frequency RF readout [3]. At Stanford, analog weights have been realized by square-loop toroids with RF second-harmonic readout [9], and by electrochemical integrators called memistors (resistors with memory), which utilize the phenomenon of electroplating to vary resistance [6]. At the present time, an all-digital circuit realization of adaptive threshold elements is under development at Stanford University. The weight values are stored in a magnetic-core memory.

The physical Learning Matrix and Madaline systems are truly

Manuscript received March 10, 1965.

K. Steinbuch is with the Institute for Information Processing and Communication, Technische Hochschule Karlsruhe, Karlsruhe, West Germany.
B. Widrow is with the Department of Electrical Engineering, Stanford University, Stanford, Calif.

[1] Logarithmus dualis.

[2] It should be noted that the Learning Matrix can be used in a reversed mode called "BE Mode" [3], [4], where the inputs and outputs are exchanged. On exciting a row, a mean "input signal set" of the associated pattern class is generated. No comparable usage has been contemplated for the Madaline system.

[3] The adjustment of the weights by the processing of one input-signal set is called "adaptation."

parallel. Sensing times are essentially independent of the numbers of inputs and outputs. In the Learning Matrix, one output line is trained at a time. In the Madaline, as many output lines as necessary are trained in parallel (simultaneously). The times of one adaptation are assumed to be equal for the Learning Matrix and the Madaline.

A comparison of the Learning Matrix and the Madaline with regard to number of outputs, number of crosspoint weights, and number of adaptations required for training is presented in Table I. The application here is that of a one-to-one logic translator which will provide a precise and different output code word for each input code word (pattern). For each input pattern, the Learning Matrix has a separate output line. The capacity of the Learning Matrix (the number of input patterns that can be precisely classified) is certain and is equal to the number m of output lines. Without an analysis of the separability of the particular input patterns by a single hyperplane [10] or by ρ hyperplanes, the performance of the Madaline cannot be predicted with certainty. However, in the case of random input patterns in general position [11], a statistical capacity has been analytically established for the single Adaline [11]–[13], giving an average value for the number of input patterns that can be correctly classified. This average is $2(n+1)$, twice the number of weights per Adaline. Theoretical curves showing the probability of linear separability vs. the ratio of number of patterns N to number of weights $(n+1)$ are shown in Fig. 2 (taken from Brown [12]). No Madaline capacity formula for the case of $\rho>1$ has yet been derived, although Winder has established an upper bound [13]. Experimental measurements, made by Koford at Stanford, and privately communicated to Widrow, indicate that the statistical capacity tends to remain at twice the number of weights for both the majority and OR fixed-logic outputs. It has been conjectured, on the basis of extremely limited experimental evidence, that the average capacity will turn out to be $2(n+1)\rho$ for most nontrivial output logic functions, although choice of output logic affects the specific switching functions that can be realized.

In a Madaline network having k output bits, it is shown [12] that when the outputs are assigned independently and when $n \gg k$, the statistical capacity of the entire network is less than, but very close to, that of a network having a single output bit. The reasons for this are 1) that the probability of separability for k output bits is the kth power of the probability of separability with $k=1$; and 2) that the curves showing probability of separability are initially flat and close to unity, and drop off sharply as shown in Fig. 2.

The Learning Matrix is compared in Table I with a Madaline whose designed average statistical capacity is at least twice the deterministic capacity of the Learning Matrix. The purpose of this difference is to make the probability of separation of the input patterns by the Madaline very close to unity, giving a fair basis of comparison. In the logic translator, the number of input patterns equals the number of pattern classes m. Let the integer ρ for the Madaline be chosen so that

$$\left(\frac{m}{n+1}\right) > \rho \geqq \frac{m}{n+1}. \tag{2}$$

Accordingly, the statistical capacity of the Madaline is conjectured to be

$$\text{statistical capacity} = (2n+1)\rho > 2m. \tag{3}$$

By comparison, the capacity of the Learning Matrix is m.

The number of adaptations for the Learning Matrix is certain. The training time for the Madaline is variable. The average number of adaptations is conservatively taken as $5m$ for the case where the number of training patterns is half capacity. Analytical work has been done by Mays [7] in estimating means and bounds on convergence time for the single Adaline ($\rho=1$) using the adaptation procedure of Ridgway and others; but so far, no simple closed-form expression for these parameters has been obtained. They are very sensitive to pattern characteristics. Experimental work has been necessary, especially where $\rho>1$.

A comparison of equipment requirements can be obtained by means of a specific example. Suppose that the number of input bits is $n=100$, and suppose that the number of classes m is $2^{10}=1024$. The Learning Matrix would have 1024 output lines and $n\cdot 1024=102\,400$ weights. The Madaline would have $k=10$ output lines and, in order to obtain the proper statistical capacity, ρ would be 10. The Madaline would have $(n+1)\rho k=10\,100$ weights. Since the two systems have the same number of input lines, the numbers of sense amplifiers that would need to be connected to the outputs of the two matrices of weights would be proportional to the numbers of weights. The equipment ratio will, therefore, be taken as the ratio of the number of weights, which for the preceding example is approximately 10:1—ten times as much equipment in the Learning Matrix as in the Madaline.

In general, the equipment ratio will be $m:\rho k$, as indicated in Table I. The quantities k and ρ must be integers, determined by

TABLE I

COMPARISON OF LEARNING MATRIX AND MADALINE AS ONE-TO-ONE LOGIC TRANSLATORS

Learning Matrix	Madaline
Number of input lines $=n$	Number of input lines $=n$
Number of pattern classes $=m$	Number of pattern classes $=m$
Number of output bits $=m$	Number of output bits $=k$ $(1+\text{ld } m) > k \geqq \text{ld } m$
Capacity $=m$ patterns, exactly (deterministic)	Statistical capacity $=2(n+1)\rho$ patterns
For comparison, half the statistical capacity of the Madaline is made equal to the capacity of the Learning Matrix: $\left(1+\frac{m}{n+1}\right) > \rho \geqq \frac{m}{n+1}$	
Number of matrix outputs $=m$	Number of threshold elements $=\rho k$
Equipment ratio $m:\rho k = \left(\frac{n+1}{\text{ld } m}\right):1$	
Number of weights[4] $=(n+1)m$	Number of weights $=(n+1)\rho k$
Training time, exactly m adaptations	Average training time with Ridgway adaptation, $5m$ adaptations (where convergence is achieved)
Training-time ratio = 1:5	
Convergence[5] is certain	Convergence[5] is not certain

[4] For binary input signals, only nm weights are required (no normalization).
[5] Convergence is defined as a condition in which all of the training patterns are classified correctly.

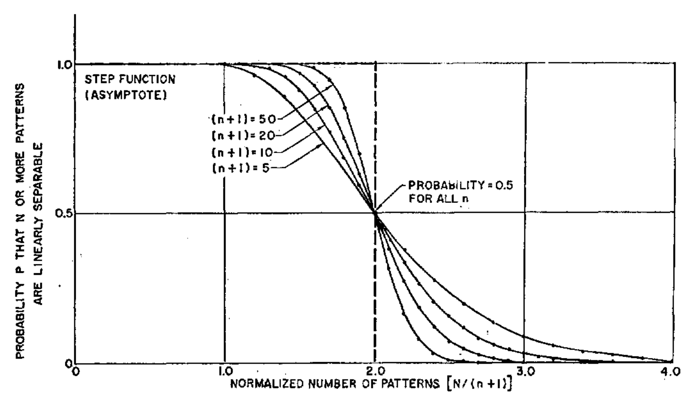

Fig. 2. Probability of linear separability of N or more random pattern vectors having a nonsingular probability-density function and equally probable desired output categories.

relations (1) and (2), respectively. The product ρk is related to the number of input lines n and to the number of pattern classes m. This product can be bounded by the following relation, derived from (1) and (2).

$$\left(\frac{m}{n+1}\,\mathrm{ld}\,m + \mathrm{ld}\,m + \frac{m}{n+1} + 1\right) > \rho k \geqq \frac{m}{n+1}\,\mathrm{ld}\,m. \qquad (4)$$

If the application of the Learning Matrix and the Madaline were statistical (rather than deterministic as in the translator application described previously), the comparison would be somewhat altered. Assume that the problem is to discriminate among noisy patterns with high reliability after training on a limited number of noisy sample patterns. Suppose that there are γ training patterns per class, and that there are m classes and n input lines as before. Comparison of amounts of equipment and training times is given in Table II. The parameter ρ of the Madaline is chosen so that the capacity of the Madaline closely matches the capacity C of the Learning Matrix.

The structure and amount of equipment required for the Learning Matrix are unchanged. The number of matrix output lines equals the number of classes m. Each line is trained γ times, corresponding to the number of training samples per class. The total number of adaptation cycles exactly equals the total number of training patterns, $m\gamma$ cycles. After the adaptation process terminates, it is possible that all of the training patterns will not be classified correctly. Such a condition is generally not as troublesome in this case as it would be when the Learning Matrix is used with noise-free patterns as in the logic-translator application. The important thing in statistical pattern classification is the achievement of a low error probability in response to patterns outside of, as well as including, the training set.

When the Madaline system is applied to the same noisy-pattern classification problem, adaptation need not necessarily proceed until the entire set of training patterns is correctly classified. Indeed, these patterns may not be separable by the particular structure.

It has been shown analytically [5] that, when a single Adaline element is trained to separate noisy patterns by the LMS procedure termination of the adaptation process after a number of adaptations, equal to 20 times the number of weights results in an error rate within approximately 5 percent of that which would be achieved if the adaptation process were continued indefinitely. Let this number of training cycles be considered adequate. A series of experiments performed at Stanford has indicated that the number of training cycles required for separation of noisy patterns increases by a factor of $\rho > 1$. Table II has been formulated on this basis.

The equipment ratios and the training-time ratios of the Learning Matrix and the Madaline may be compared in the applications of logic translator and classifier of noisy patterns by comparing Tables I and II. The equipment required by the Learning Matrix is the same for both of these applications. For the second application, the equipment required by the Madaline must be increased generally by the factor $f(m, n) > 1$. It is possible for the equipment required by the Madaline to exceed that required by the Learning Matrix, although this will not usually be so. The training time of the Learning Matrix increases by a factor of γ, while the training time of the Madaline increases by a factor of $4\rho[(n+1)/m]$ in the statistical classifier application as compared to the translator application. It is possible that, in certain statistical classification applications where γ is large, the average training time of the Madaline could be less than that of the Learning Matrix, although in most situations the training time of the Learning Matrix will be less than the average training time of the Madaline.

With regard to optimal schemes for the classification of noisy patterns, Table III is a limited list including 1) some of the conditions and types of problems in which it is known that the Learning Matrix is the better classifier; 2) some of the other conditions in which the Adaline is known to be the better classifier; and 3) still other conditions in which a certain form of the Learning Matrix using the conventional Adaline-type LMS adaptation procedure is optimal [14]. In the Learning Matrix, a separate linear filter is matched to the sample mean of each class of patterns. When the Madaline scheme is utilized, each Adaline element is "matched" to a variety of classes simultaneously.

In many situations, the application of the Learning Matrix is simpler because one can be very sure about its memory capacity, and one can be certain about its convergence in separating properly an MHD- or MED-separable set of training patterns. An example is the deterministic translator problem described above. Convergence of the Madaline in such situations would be uncertain without an extensive preliminary analysis to guarantee separability. There are a number of other important situations, however, in which the picture is reversed. One example is the realization of optimal switching surfaces by adaptive networks for a variety of contactor ("bang-bang") control processes.

TABLE II

COMPARISON OF LEARNING MATRIX AND MADALINE AS STATISTICAL CLASSIFIERS (Using γ Training Patterns per Class)

Learning Matrix	Madaline
Number of input lines $= n$	Number of input lines $= n$
Number of pattern classes $= m$	Number of pattern classes $= m$
Number of output bits $= m$	Number of output bits $= k$ $(1 + \mathrm{ld}\ m) > k \geqq \mathrm{ld}\ m$
Capacity[6] $= mf(m, n) \equiv C$ patterns	Statistical capacity $= 2(n+1)\rho > C$ patterns
For comparison, the statistical capacity of the Madaline is made equal to the average capacity of the Learning Matrix $\left(1 + \frac{C}{2(n+1)}\right) > \rho \geqq \left(\frac{C}{2(n+1)}\right)$	
Number of matrix outputs $= m$	Number of threshold elements $= \rho k$
Equipment ratio $= m : \rho k$	
Number of weights $= m(n+1)$	Number of weights $= (n+1)\rho k$
Number of training patterns available $= m\gamma$	Number of training patterns available $= m\gamma$
Training-time ratio $= m\gamma : 20(n+1)\rho$	
Training time, exactly $m\gamma$ adaptations	Average training time, *LMS* adaptation, $20(n+1)\rho$ adaptations
Convergence is not certain	Convergence is not certain

[6] The coefficient $f(m, n)$ of m proposed by H. Kazmierczak and relating to the capacity of the Learning Matrix may be greater than or less than γ, but up to now an exact estimate of $f(m, n)$ does not exist.

TABLE III

OPTIMAL CLASSIFIERS FOR A LIMITED LIST OF STATISTICAL SEPARATION PROBLEMS AND CONDITIONS

Classifier	Optimal Usage Conditions for Minimum Error Probability
Learning Matrix	Multiple-category separation. Noises in input signals mutually uncorrelated and uncorrelated with pattern class, and of equal variances among all input signals and among all classes. During adaptation, Learning Matrix weights made equal to sample means for each class. Weights along each matrix output line must be normalized. No influence on trained weight distribution if m' new additional pattern classes are admitted and trained in m' additional rows. Inverse operation ("BE mode") of Learning Matrices possible.
Madaline, single output with one Adaline ($\rho = 1$; $k = 1$)	Two-category separation. Gaussian noise in the two pattern classes may or may not be correlated, but should have identical covariance matrices in both classes. Least-mean-square (LMS) adaptation procedure should be used. Normalization of patterns or weights not required.
Learning Matrix with LMS adaptation (iterative training)	Multiple-category separation. Gaussian noises in each pattern class may be correlated but should have identical covariance matrices. All patterns in a given class should be trained into an individual line of weights that are adapted to minimize the mean-square error for a desired output of +1. It is important that the weights α_{q0} for each q be removed. Normalization of input patterns is required. Normalization of weights is not required.

Contactor controllers produce real-time binary-actuation signals to "plants" being controlled. These instantaneous binary decisions depend on the state of the system. In many cases, it has been determined that, when the analog state variables are quantized and encoded with the linearly independent codes proposed by Smith [15], the resulting patterns and associated responses are linearly separable when discrete approximations to optimal switching surfaces (minimum settling time, minimum fuel, etc.) are realized. At Stanford University, a machine (the "broom-balancer") has been constructed to illustrate these principles. It is used to control a fourth-order process. An inverted pendulum on wheels is stabilized by encoding four state variables (angle, angle rate, position, and velocity of pendulum mount) and feeding the resulting binary patterns to a single adaptive threshold element. The threshold-element output controls a relay, which actuates a motor coupled to the rolling pendulum mount.

It is possible to realize the same kind of control surface with a Learning Matrix. The output signals could be decoded to produce a single binary output. It is likely, however, that a considerable amount of preliminary study would be required in order to determine the proper number of matrix outputs, how these outputs should be decoded to a single binary output to insure separability in the realization of an optimal switching surface, and how many adaptations would be required to achieve convergence.

The comparisons made previously allow few unequivocal conclusions. It can safely be said, however, that in most cases, the Learning Matrix will require fewer adaptation cycles than the Madaline, when applied to a given problem; while on the other hand, the Madaline will usually require less equipment than the Learning Matrix. However, in making a choice of techniques for a given situation, the overriding consideration will usually be the question of optimality. The method selected will be the one that can be shown to provide optimal responses or to approach this goal as closely as possible. Both the Learning Matrix and the Madaline, together with variations and combinations of the techniques described, have their own natural areas of application.

An increasingly important field for research is the development of a network theory (both statistical and deterministic) for adaptive systems, which will relate the nature of the problem to the optimal configuration of adaptive elements that would produce a solution.

Acknowledgment

The authors wish to thank Dr. H. Kazmierczak of the Technische Hochschule Karlsruhe; Prof. T. E. Cover of Stanford University; and Dr. R. J. Brown of the Autonetics Division, North American Aviation Corp., Anaheim, Calif., for their assistance in preparation of this note. Also Profs. J. B. Angell and J. G. Linvill of Stanford, for making possible the visit of Dr. K. Steinbuch under the sponsorship of the Stanford University Solid-State Affiliates Program.

References

[1] K. Steinbuch, "Die Lernmatrix," *Kybernetik*, vol. 1, pp. 36–45, January 1961.
[2] K. Steinbuch and H. Frank, "Nichtdigitale Lernmatrizen als Perzeptoren," *Kybernetik*, vol. 1, pp. 117–124, December 1961.
[3] K. Steinbuch and U. A. W. Piske, "Learning matrices and their applications," *IEEE Trans. on Electronic Computers*, vol. EC-12, pp. 846–862, December 1963.
[4] H. Kazmierczak and K. Steinbuch, "Adaptive systems in pattern recognition," *IEEE Trans. on Electronic Computers*, vol. EC-12, pp. 822–835, December 1963.
[5] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *1960 WESCON Conv. Rec.* pt. 4, pp. 96–104, August 1960.
[6] B. Widrow, "Generalization and information storage in networks of Adaline 'neurons' " in *Self Organizing Systems 1962*, M. C. Yovits, G. T. Jacobi, and G. D. Goldstein, Eds. Washington, D. C.: Spartan Books, 1962.
[7] C. H. Mays, "Adaptive threshold logic," Stanford Electronics Labs., Stanford, Calif., Rept. SEL-63-027, April 1963.
[8] J. K. Hawkins, "Self-organizing systems—A review and commentary," *Proc. IRE*, vol. 49, pp. 31–48, January 1961.
[9] H. S. Crafts, "Components that learn, and how to use them," *Electronics*, vol. 36, March 1963.
[10] F. C. Singleton, "A test for linear separability as applied to self-organizing machines" in *Self-Organizing Systems 1962*. Washington, D. C.: Spartan Books, 1962.
[11] T. M. Cover, "Geometrical and statistical properties of linear threshold devices," Stanford Electronics Labs., Stanford, Calif., Rept. SEL-64-052, May 1964.
[12] R. J. Brown, "Adaptive multiple-output threshold systems and their storage capacities," Stanford Electronics Labs., Stanford, Calif., Rept. SEL-64-018, June 1964.
[13] R. O. Winder, "Bounds on threshold gate realizability," *IEEE Trans. on Electronic Computers (Correspondence)*, vol. EC-12, pp. 561–564, October 1963.
[14] G. F. Groner, "Statistical analysis of adaptive linear classifiers," Stanford Electronics, Labs., Stanford, Calif., Rept. SEL-64-062, April 1964.
[15] B. Widrow and F. W. Smith, "Pattern recognizing control systems," *1963 Computer and Information Sciences (COINS) Symp. Proc.*

Logical Design of Analog-to-Digital Converters

PETER FELLGETT

Verster[1] has described a serial-parallel A-D converter which has improved accuracy, and he has embedded his description in an interesting discussion of A-D logic.

The speed of an A-D converter can be increased by processing several digits in parallel. The penalty is that subtraction of the equivalent analog feedback quantity cannot then take place until after the decision is made. This effectively opens the feedback loop of the A-D with respect to certain types of drift etc., and errors may result. Verster overcomes this objection by making the decision revocable within calculated limits, and is thus able to increase the *speed-accuracy* product.

There are many A-D applications which make only modest demands on the speed of modern transistors. At Royal Observatory Edinburgh (ROE) it has been found valuable to design an A-D in which speed is sacrificed for extreme simplicity of logic.[2] The object of the present note is to show that this design, although initially based on different objectives, is compatible with Verster's concepts, and that a combination of both sets of ideas enables a good *speed-accuracy-simplicity* product to be attained over a wide range of requirements.

Figure 1 outlines the ROE arrangement, which consists essentially of a simple counting register, feeding a weighting and summing network, which is fed back subtractively to the analog input in the ordinary way. A discriminator detects whenever the signal fed back differs from the input by more than a prescribed threshold value (in principle, half the least significant digit) and gates an oscillator into the counter which then cycles until balance is restored. There is provision (not shown) for either continuous or one-shot operation. A fuller description will be given in Matley et al.[2]

Using cheap silicon transistors capable of counting at 1.024 Mc/s, this circuit needs a maximum of 1 ms to make a 10-bit reading. This performance is fully sufficient for many applications. If it is desired to increase the speed, a first step is to use a reversible counter, and to gate the oscillator either to add or to subtract according to the sign of the error. The maximum reading time is thus halved, and the mean reading time for a reasonably smooth input waveform is reduced considerably.

The speed may be further increased by using multiple-level discrimination to gate the oscillator to more significant stages of the counter when the error is large. For example, two discriminators of appropriate sensitivity can control the zeroth and the Nth stages of a $2N$-bit register, as shown in Fig. 2. This effectively divides the register into two halves, which are set successively, and the maximum operation time is reduced by a factor of approximately 2^{N-1}.

By continuing this process, a configuration is eventually found in which every stage of the counter is controlled by a discriminator of appropriate threshold (see Fig. 3). For this configuration, it is not very appropriate to regard the pulses reaching the counting stages as "add" or "subtract" pulses but rather they are better regarded as "set" and "reset" pulses. It is simplest (but not necessarily optimum) to postulate a circuit in which carries are propagated from one stage to the next of the counter only in the absence of set or reset pulses affecting the two stages. The present treatment will follow that of Verster in omitting discussion of carry propagation times.

It may be judged that the arrangement shown in Fig. 3 is rather easier to implement than the more usual type of serial A-D which operates under the control of explict sequencing logic. In Fig. 3, the sequencing is obtained implicitly from the gradation in threshold between the various discriminators. This yields a number of apparent advantages. At each stage of taking a reading (including the first) the circuit changes the most significant bit, which is currently in error. It will, therefore, have a considerable speed advantage over the

Manuscript received May 1, 1965.
The author is with the Department of Applied Physics, University of Reading, England. He was formerly with Royal Observatory, Edinburgh, Scotland.
[1] T. C. Verster, "A method to increase the accuracy of fast-serial-parallel analog-to-digital converters," *IEEE Trans. on Electronic Computers*, vol. EC-13, pp. 471–473, August 1964.
[2] W. Matley, et al., *Publ. Royal Observatory, Edinburgh*, to be published.