(11) Harrison, J. M., and Lynch, M. F., "Computer Analysis of Chemical Reactions for Storage and Retrieval," *J. Chem. Soc.* **1970**, 2082–7.
(12) Corey, E. J., and Wipke, W. T., "Computer-assisted Design of Complex Organic Synthesis," *Science* **166**, 178–92 (1969).
(13) Corey, E. J., "General Methods for the Construction on Complex Molecules," *Pure Appl. Chem.* **14**, 19–37 (1967).
(14) Corey, E. J., "Computer-assisted Analysis of Complex Synthetic Problems" (Centenary lecture), *Quart. Rev.* **25**, 455–82 (1971).
(15) Corey, E. J., Wipke, W. T., Cramer, R. D., III, and Howe, W. J., "Computer-assisted Synthetic Analysis. Facile Man-machine Communication of Chemical Structure by Interactive Computer Graphics," *J. Amer. Chem. Soc.* **94**, 421–30 (1972).
(16) Corey, E. J., Wipke, W. T., Cramer, R. D., III, and Howe, W. J., "Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules," *J. Amer. Chem. Soc.* **94**, 431–9 (1972).
(17) Corey, E. J., Cramer, R. D., III, and Howe, W. J., "Computer-assisted Synthetic Analysis of Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates," *J. Amer. Chem. Soc.* **94**, 440–59 (1972).
(18) Corey, E. J., and Petersson, G. A., "An Algorithm for Machine Perception of Synthetically Significant Rings in Complex Cyclic Organic Structures," *J. Amer. Chem. Soc.* **94**, 460–5 (1972).
(19) Weygand, C., "Organic Preparations," VII-XIII, 1-3, Interscience, New York, 1945.
(20) Theilheimer, W., "Synthetische Methoden der Organischen Chemie," **1**, VI-VII S. Karger, Basel, New York, 1946.
(21) Sugasawa, S., and Nakai, S., "Reaction Index of Organic Synthesis," Wiley, New York, 1967.
(22) Vleduts, G. E., "Concerning One System of Classification and Codification of Organic Reactions," *Inf. Stor. Retr.* **1**, 117–46 (1963).
(23) Mishchenko, G. L., "Information Retrieval in the Field of Chemistry," 117–36, *Nat. Techn. Inf. Serv. JPRS–53523.* From *Zh. Vses. Khim. Obshch.* **16**, 55–63 (1971).
(24) Smith, E. G., "The Wiswesser Line-formula Chemical Notation," McGraw-Hill, New York, 1968.

# Error Checking Digit for Nonconventional Chemical Codes*

K. SUBRAMANIAM and P. V. SANKAR**
Department of Electrical Communication Engineering and Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Received September 22, 1972

**A method of constructing a check digit for nonconventional chemical codes is described.**

In handling large volumes of data such as chemical notations, serial numbers for books, etc., it is always advisable to provide checking methods which would indicate the presence of errors. The entire new discipline of coding theory is devoted to the study of the construction of codes which provide such error-detecting and correcting means.[1] Although these codes are very powerful, they are highly sophisticated from the point of view of practical implementation. With this in view, several inexpensive, but fairly effective means to guard against errors resulting from improper data transcription have been evolved. Some of these checks are already available for use in documentation and libraries—in the International Standard Serial Numbering (ISSN) for books and journals, and CODEN used for periodic Titles and in the Chemical Abstracts Service (CAS).

The ISSN system[2] uses a seven digit decimal code $\alpha\beta\gamma\delta\epsilon\phi\psi$ (where $\alpha,\beta,\gamma,\delta,\epsilon,\phi,\psi, = 0, 1, 2, \ldots 9$) and derives a check digit $\sigma$ from these seven digits by using the equation

$$\sigma \equiv [11 - (8\alpha + 7\beta + 6\gamma + 5\delta + 4\epsilon + 3\phi + 2\psi) \bmod 11]$$

and sets for $\sigma = \mathrm{X}$ for $\sigma \equiv 10 \bmod 11$

The Coden System,[3] on the other hand, uses only a five alphanumeric character code $\alpha\beta\gamma\delta\epsilon$ and derives the check character by the equation

$$\sigma \equiv (11\alpha + 7\beta + 5\gamma + 3\delta + \epsilon) \bmod 34$$

by assigning integral weights as follows:

CODEN: A, B, ...... Y, Z, 1, 2, .... 9, 0
Equivalent: 1, 2, ...... 25, 26, 27, 28, .... 35, 36.

The remainder $\sigma$ is converted into a check character by the following set of equivalents:

Remainder: 1, 2, ..... 25, 26, ..... 33, 34 (or zero)
Check character: A, B, ...... Y, Z, 2, 3, ..... 8, 9

The numeric check characters one (1) and zero (0) have been eliminated to avoid confusion with the alphabetic characters I and O.

Since the above methods of constructing check digits are applicable only for fixed length codes, it becomes necessary to consider a different scheme for constructing the check digits for a variable length code like WLN, etc.

Fortunately, the code suggested by Black[4] for decimal systems can be easily extended to a general base. This code detects a single error or an adjacent transposition error.

## PRINCIPLE OF BLACK'S CODE

Given an N digit number, a check digit is constructed as a function of the N digits using a multiplication operation '*' defined below:

$$A * B = A \oplus (-1)^A B \qquad (1)$$

* Contribution No. 26 from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.
** To whom correspondence should be addressed.

where $0 \leq 9$ A, $B \leq 9$ and $\oplus$ is MOD-10 addition. The decimal digits 0 through 9 form a group with respect to this multiplication. A permutation which maps D into D, i.e., $T:D \rightarrow D$ is defined as

$$T(A) = \begin{cases} A \text{ if A is odd} \\ 8 - A \text{ if A is even} \end{cases} \quad (2)$$

For an N digit number $d_1 d_2 \ldots d_N$ there exists an unique check digit $d_{N+1}$ such that

$$\prod_{i=1}^{N+1} T^i(d_i) = 0 \quad (3)$$

since $T^i$ is a permutation and D satisfies the group properties.

## GENERALIZATION AND APPLICATION

This can be generalized to any given even base of representation provided the multiplication operation '*' is defined by

$$A * B = A \oplus (-1)^A B \quad (4)$$

where $0 \leq A, B \leq (\beta-1)$ and $\oplus$ is MOD-$\beta$ addition, and

$$T(A) = \begin{cases} A \text{ if A is odd} \\ \beta\text{-2-A if A is even} \end{cases} \quad (5)$$

Since the most probable errors which can occur in a chemical code like WLN[5] are just the alteration of a single character or a transposition of adjacent characters, the above code will guard against such errors in the computer handling of chemical codes.

The alphabet of the WLN consists of 40 characters, viz., (i) the letters of the English Alphabet A through Z, (ii) the decimal digits 0 to 9, (iii) the special characters &, -, / and (iv) the blank space. $\beta$ is chosen as 40 and each character in the alphabet is assigned a unique number from the set (0, 1, 2, . . . . . , 39) as follows

| | |
|---|---|
| WLN: | $\phi$, 1, . . . . . . . 9, A, B, . . . . Z, blank, &, -, / |
| Equivalents: | $\phi$, 1, . . . . . . . 9, 10, 11, . . . . 35, 36, 37, 38, 39. |

A Fortran program for constructing the check character as well as for checking whether the code is error-free is available in the microfilm edition of this journal.

**Example 1**

Let the given WLN be 3O2

This is coded as a number in base 40 with

$$d_1 = 3$$
$$d_2 = 24$$
$$d_3 = 2$$

The check digit $d_4$ is constructed such that

$$T^1(d_1)*T^2(d_2)*T^3(d_3)*T^4(d_4) = 0$$

Now using (5)

$$T^1(d_1) = 3$$
$$T^2(d_2) = 24$$
$$T^3(d_3) = 36$$

By (4)

$$((3 * 24) * 36) = (19 * 36) = 23$$

Hence the check digit is $T^4(d_4) = 23$.

$$\text{Since } (23 * 23) = 0$$
$$d_4 = 23$$

or the check character is N

The coded WLN is written as 3O2N

Let us assume that 2 and O are interchanged, i.e., 32ON where

$$d_1 = 3$$
$$d_2 = 2$$
$$d_3 = 24$$
$$d_4 = 23$$
$$T^1(d_1)*T^2(d_2)*T^3(d_3)*T^4(d_4)$$
$$= (((3 * 2) * 24) * 23)$$
$$= ((1 * 24) * 23) = (17 * 23)$$
$$= 34 \neq 0$$

Hence the error has been detected.

**Example 2**

WLN: OVVO & 3/ – FE– 2

Here, the check character is &

## ACKNOWLEDGMENT

The authors are indebted to E. V. Krishnamurthy for many valuable suggestions. Also thanks are due to the referees for the constructive remarks in revising the paper. Partial support from the National Institutes of Health (Grant No. AM-15964) is also acknowledged.

## LITERATURE CITED

(1) Peterson, W. W. and Weldon, E. J., Jr., "Error Correcting Codes," The MIT Press, Cambridge, Mass., and London, England, 1972.
(2) Bowker Annual of Library and Book Trade Information, R. R. Bowker, New York, London, 197–200, 1972.
(3) CODEN for Periodic Titles, Vol. I, American Society for Testing and Materials, Philadelphia, Pa., 1968.
(4) Black, W. L., "Error Detection in Decimal Numbers," *Proc. IEEE*, **60**, No. 3, 33 (1972).
(5) Smith, E. G., Ed., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill, 1968.

A Fortran program for error detection in WLN will appear following these pages in the microfilm edition of this volume of the Journal. Single copies may be obtained from the Business Operations Office, Books and Journals, Division, American Chemical Society, 1155 Sixteenth St., N. W. Washington, D. C. 20036. Refer to the following code number: JCHD-73-39. Remit by check or money order \$3.00 for photocopy or \$2.00 for microfiche.