

Computer Generation of Wiswesser Line Notation. II. Polyfused, Perifused, and Chained Ring Systems

STEPHEN R. HELLER and DEENA A. KONIVER*

Heuristics Laboratory, Division of Computer Research and Technology,
National Institutes of Health, Department of Health, Education and Welfare,
Bethesda, Md. 20014

Received November 12, 1971

The computer program for the generation of Wiswesser Line Notation (WLN) has been extended to include polyfused rings, methyl contraction rules, chain of two ring systems, some perifused rings, some chelates, and some metallocenes. Salts and ions are also handled, but in a different manner than what is normally found. Multipliers are not used by the program. The normal input for the WLN generation is an easy input program using a Rand Tablet; however, teletype and connection table input can also be used in most cases.

While the universal chemical sign language—the structural diagram—is the means by which the chemist prefers to communicate, other methods are both valuable and necessary. Chemical nomenclature is one such method, but suffers from the major deficiencies of being redundant and variable with time. The use of connection tables, giving the atom types and the connections between atoms, is a very powerful way to store precisely chemical structures.

Another method for the representation of a chemical structure is a one-dimensional linear notation. Examples of these include the Wiswesser Line Notation (WLN), the Dyson, and Hayward Notations.¹ These linear notations are more compact than the corresponding connection tables. However, searches for fragments or substructures using the WLN are limited in cases such as fragments which are both inside and outside a ring. While the Chemical Abstracts Service file of connection tables² (some 1.8 million to date) is probably the largest file of its kind, the WLN has a wide following, and WLN appears in many files as well as in the literature.^{3, 4} The WLN Generation Program discussed in this paper runs on a DEC PDP-10 time-sharing computer.

The WLN component of our experimental Chemical Information System⁵ is shown in Figure 1. Not indicated in Figure 1 is the option of taking a WLN string and generating a connection table.⁶

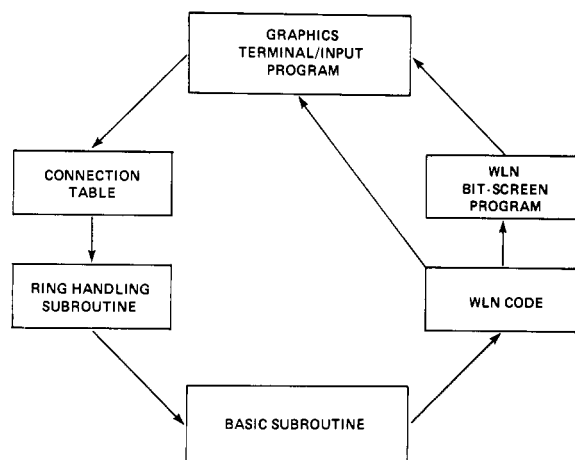


Figure 1. WLN component of the DCRT/CIS

INPUT PROGRAM

In an attempt to make the WLN Generation Program as psychologically appealing to the chemist as possible, the normal input into the system is via a Rand Tablet and Cathode Ray Tube (CRT). This allows the user to communicate directly with the computer on his own terms and has been found to be very valuable.

Figure 2 shows the menu of actions for the user to choose from. All actions are interactive and can be corrected easily if a mistake has been made. The boxes on the right-hand side from S down to C stand for eleven of the more common elements, and the X allows the user to input any one of the 103 elements. Below these atom boxes are the ZB box, which is the symbol for Zero Bond. This is used for ions. Next is the PI box, which is the symbol for a π bond. This is used for metallocene structures.

WLN GENERATION PROGRAM

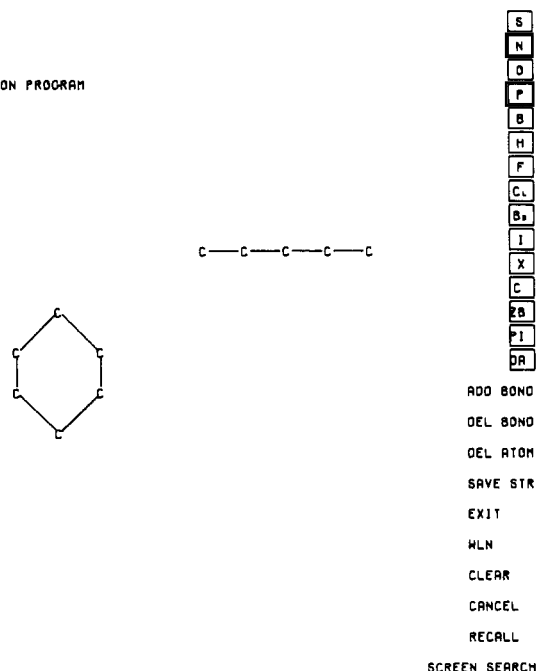


Figure 2. CRT screen and the menu

* To whom correspondence should be addressed.

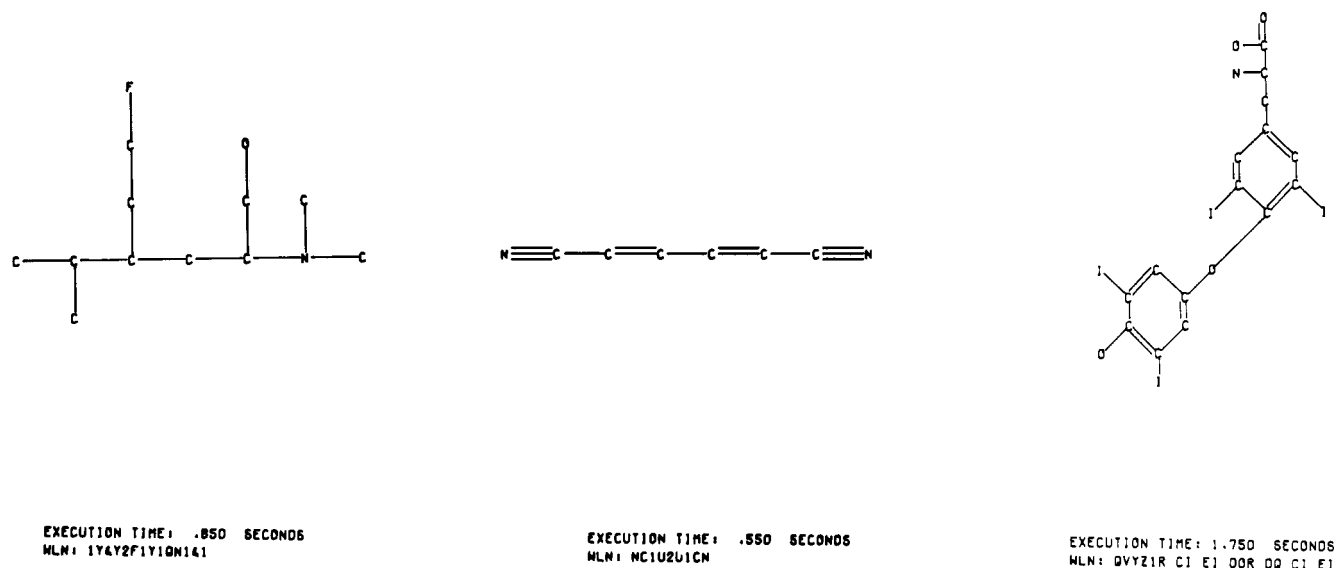


Figure 3. Examples of the WLN generated for aliphatic structures and benzene rings

In the last box is DA, which is the symbol for a dative or chelate bond. This is used for chelate structures.

Below these boxes are ten editing commands. The ADD BOND command allows the user continuously to add double (and triple) bonds throughout the structure he has drawn. The DEL BOND command allows the user to delete a bond which was added incorrectly or not wanted. The DEL ATOM command allows the user to delete an atom from the structure. The SAVE STR command initiates a subroutine to save on the computer disk the connection table of the structure that is presently on the screen. After the save is completed, the structure remains and the user can continue. The EXIT command allows the user to exit from the program. As a check, in case the exit command request was accidental, the program asks again if you wish to recycle—i.e., restart—or exit. The WLN command initiates the WLN Generation Program. The CLEAR command clears the screen of whatever structure is there, allowing the user a fresh start. The CANCEL command, probably one of the most useful, allows the user to go back to the state he was in before the previous command. For example, if the user put an N atom in a ring when he meant to put an O atom, the cancel command cancels the N atom and allows him to try again. If the user draws five bonds around a carbon atom, or two bonds around a chlorine atom, the valence subroutine reminds the user of his probable indiscretion, and the cancel command allows the error to be corrected. As a last example, after drawing a structure and obtaining the WLN, the user can use the cancel command to cancel the WLN generation request. This leaves the user with his structure, which can be altered and the WLN for the new structure requested. This is extremely convenient when obtaining the WLN for a series of related structures. The RECALL command allows the user to recall any structure that has been saved on the disk. Lastly, the SCREEN SEARCH command initiates the WLN Bit-Screen Generation and the WLN Substructure Search Program.⁷

The interactive input program is designed for the ease of structural diagram input. To "draw" a structure on the Rand Tablet, the user depresses the stylus and draws either an arbitrary circle—i.e., a closed loop—or a line. The former automatically generates a six-membered ring. The latter generates a chain of carbon atoms, the number of which is proportional to the length of the line drawn. Examples of the additions and modifications to the structure can be found elsewhere.⁵

After the user generates the base structure of the molecule of interest, the menu boxes are used to "refine" the structure by putting in the noncarbon atoms and the extra bonds. This "easy" input program⁵ requires $\frac{1}{2}$ to $\frac{1}{3}$ the number of pen actions and time compared with the input program described in a previous paper.⁸ The "easy" input is particularly facile in drawing multiple rings and in making hetero-atom substitutions in large molecules.

WLN GENERATION

A detailed description of the initial WLN Generation Program has been presented previously.⁸ Only the additions to the original program will be presented here.

Aliphatic Structures and Benzene Rings. For the generation of aliphatic structures and benzene rings, the subroutine has been completely rewritten and covers the contents of the first nine chapters of the WLN manual,⁴ with the exception of multipliers, which our program does not handle. Note that the rules of WLN treat the benzene ring in a different manner than other cyclic structures. The aliphatic subroutine is called CARL,⁹ for Chemical Algorithm for Reticulation Linearization. It contains three main sections: Basic, Linear, and Final. Basic takes the connection table from the input program and makes minor modifications, such as replacing a "V" for a carbonyl and an "R" for a benzene ring and a "U" for a double bond. Thus what is passed out of Basic is a modified connection table containing only single connections between entries. Most importantly, the symbols in the connection table are now exactly the ones that will appear in the WLN code (except for the locants and ampersands). To put in the locants and ampersands, each symbol has a flag associated with it, which indicates the number of ampersands and what locant (if any) follows the particular symbol.

The next section, Linear, calculates the flags for each symbol and finds the correct permutation of symbols. Final then takes the information derived from Linear and rewrites the WLN code with the ampersands and locants inserted in their proper place. Examples of the WLN generated by this subroutine are shown in Figure 3. This figure as well as all but Figure 1 used in this paper were obtained from Calcomp plots of the image on the CRT.

Cyclic Structures. The original program described pre-

COMPUTER GENERATION OF WISWESSER LINE NOTATION. II

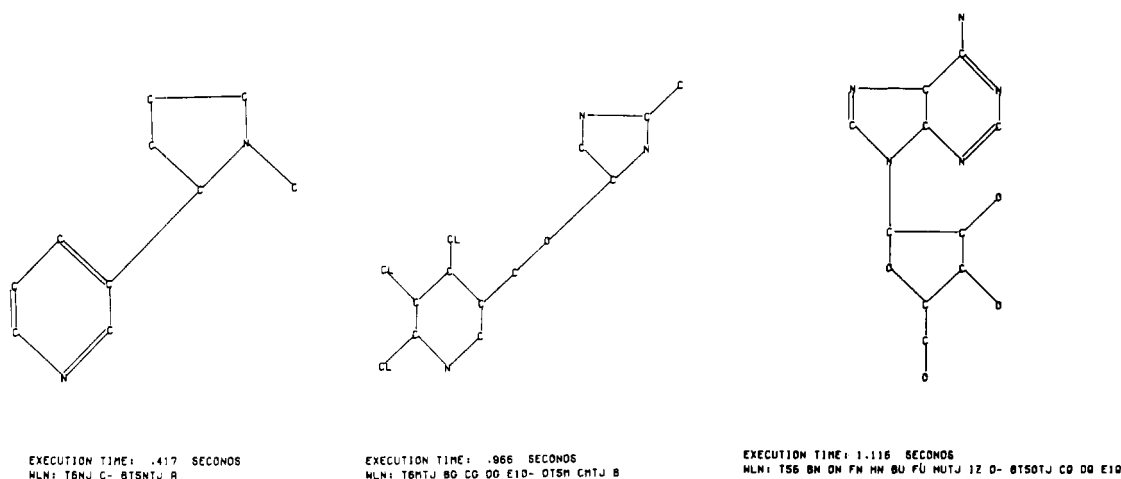


Figure 4. Examples of the WLN generated for a chain of two ring systems

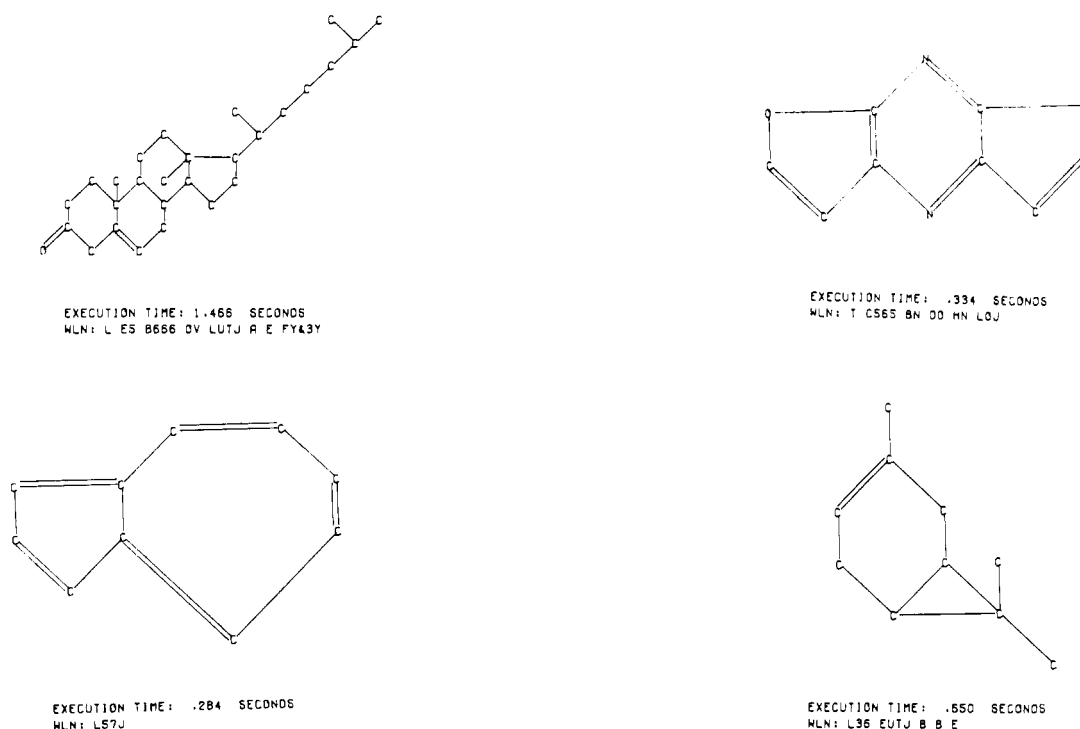


Figure 5. Examples of the WLN generated for polycyclic fused rings

viously handled only up to one cyclic (nonbenzene) ring structure. The program has been extended to cover a chain of two ring systems, polyfused ring systems, and perifused ring systems. The general limitations of the program are that a ring can contain a maximum of nine atoms, that there can be a maximum of 15 rings per structure, and lastly, there can be a maximum of one-hundred nonhydrogen atoms per structure. These limitations are all arbitrary and could be increased if it was deemed necessary to cover certain molecules. Examples of the subroutine that generates the WLN for a chain of two ring systems is given in Figure 4. The programs also handle arbitrarily large polyfused ring systems, examples of which are shown in Figure 5.

Lastly, the programs are able to handle a broad class of

perifused ring compounds. The function Peri is used to calculate the notation for the perifused ring system. The current limitations allow the WLN generation for both totally saturated—i.e., all T rings—and totally unsaturated—i.e., all & rings—as well as a totally unsaturated perifused ring system with one point of saturation in one of the rings. The program does not handle perifused ring systems requiring branch locants. Very briefly, the Peri function is given the connections and nodes directly involved with the ring system under consideration and returns the actual WLN for the perifused ring system, including the list of rings, their associated locants, the number and list of perifused atoms and the locant of the last cited node in the path. Further details and a discussion of the algorithm used can be found elsewhere.⁶ Examples of the WLN

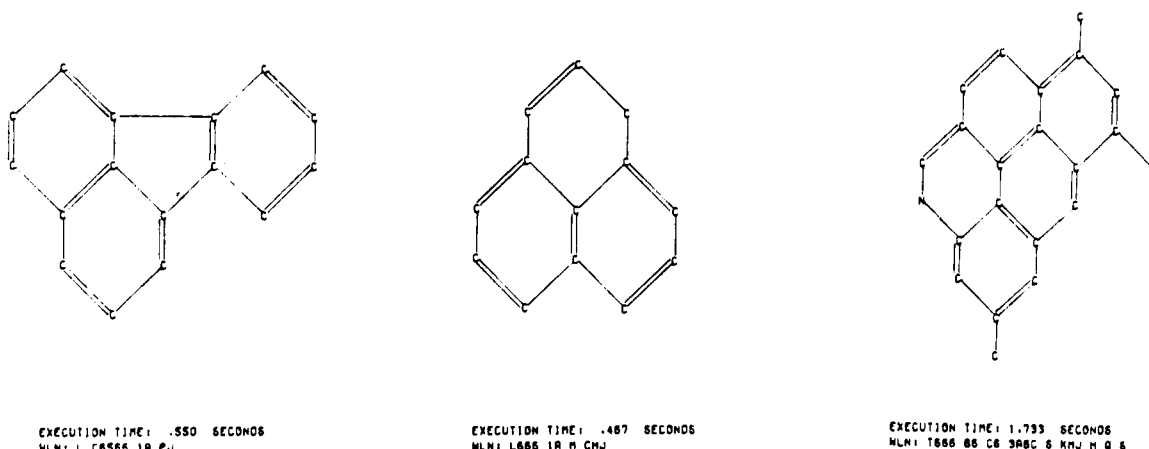


Figure 6. Examples of the WLN generated for perfluorinated rings

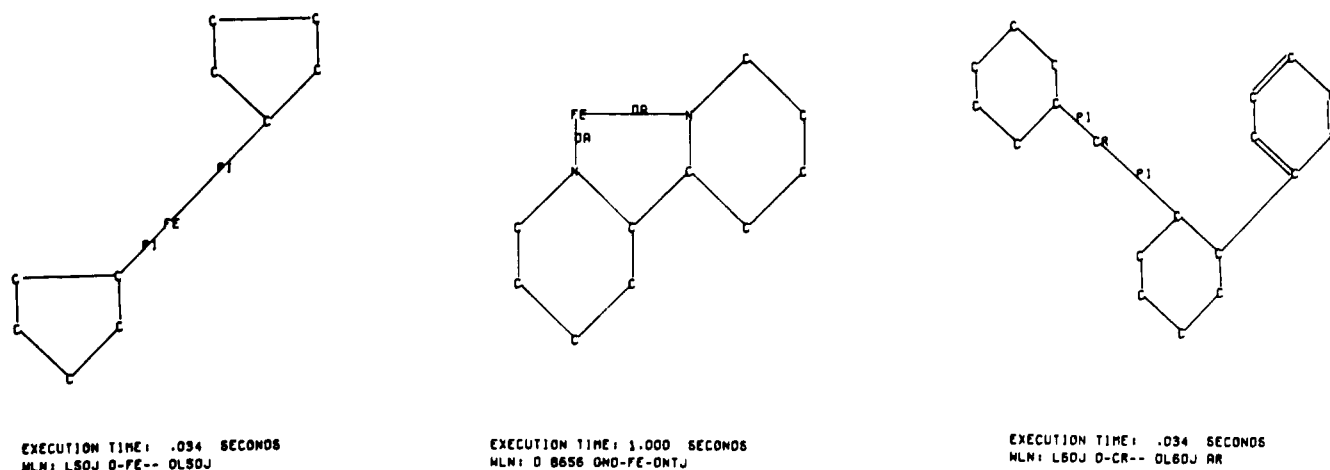


Figure 7. Examples of the WLN generated for metallocenes and chelates

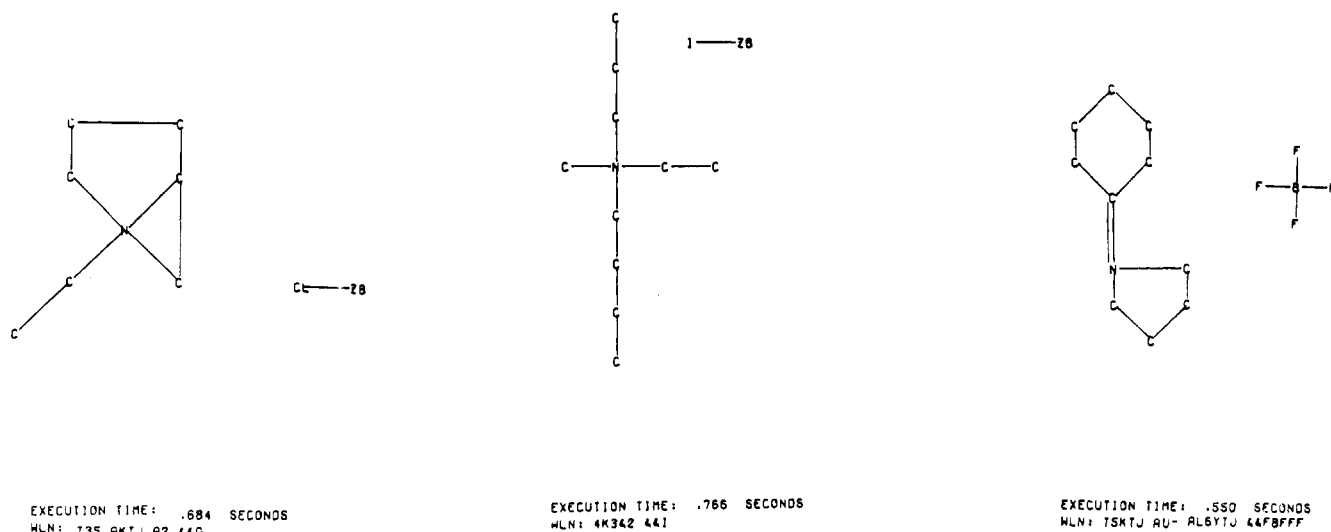


Figure 8. Examples of the WLN generated for salt and ions

Generation Program for perfused structures are shown in Figure 6.

Metallocenes and Chelates. The WLN Generation Program now covers a limited class of structures with "nonclassical" connectivity, that is, nonclassical in the chemical sense. In the case of chelates, the limitation is only cyclic chelates with complete coordination. Work to cover cyclic structures with some classical and some nonclassical bonds is in progress. The basic connection table generated for the structure from the input program has been modified to contain a second connection table for "nonclassical" bonds, such as the coordinate or dative bonds found in chelate structures. The user designates a dative bond in the input by touching the DA box and then touching the two atoms at either end of the DA bond.

The coverage of metallocenes or π -complex structures is currently limited to totally unsaturated ring metallocenes. The problems in the connection table here are similar to those in the chelate structures. For example, chemically speaking, in ferrocene the iron atom is $\frac{1}{2}$ bonded to each atom of each cyclopentadienyl anion. However, the regular connection table allows for either zero bonds or one bond, with no intermediary values. Thus a second connection table is generated in the input program by touching the PI box and then touching the metal atom and the ring atom to which it happens to be attached. This causes that bond to be destroyed and a new metal- π bond to be generated in the new second connection table. Examples of the subroutine for chelates and metallocenes is shown in Figure 7.

Salts and Ions. The program now covers salts and ions; however, the WLN generated is not the WLN suggested in the manual. The DCRT WLN program indicates this by separating ions with a blank space and two ampersands. The main difference is that the program does not choose on a chemical basis where, for example, a hydrogen atom is attached in a salt. Wherever the user draws the bond, it remains. Since the program automatically fills in hydrogen atoms, if necessary, to satisfy the regular valence of an atom, the input program uses the ZB from the menu to generate salts and ions. The ZB, standing for Zero Bond, replaces what would usually be filled in with a hydrogen atom in the salt or ion with what amounts to a nonatom or imaginary atom, so that the WLN Generation Program will consider it a salt or ion structure. Each ion is encoded using the organic rules in the manual, rather than sometimes using the inorganic rules. See Figure 8 for examples.

Methyl Contractions. The program now handles methyl contractions. Examples of this can be seen in most of the figures in this paper.

USES OF THE PROGRAM

The program for the generation of WLN has been used to generate WLN for the Common Data Base, a collection of compounds compiled by the FDA and NLM. The program handled 81% of this file. In addition, the program has been used to generate WLN for most of the drugs in the second edition of the book "Psychotropic Drugs and Related Compounds."¹⁰ The program is being used to generate WLN for a two-volume catalog of NMR spectra.

CONCLUSIONS

The automatic encoding, via graphic input, of chemical structures into WLN appears to be a practical objective. The main advantages of the system described are the graphical input and the certainty that the same (and correct) WLN is always produced. Plans are currently underway to extend the program to cover bridged and spiro ring systems.

LITERATURE CITED

- (1) "Chemical Structure Information Handling, A Review of the Literature," 1962-1968, National Academy of Sciences, Washington, D.C., 1969.
- (2) Tate, F. A., *Chem. Eng. News* **45**, 79-90 (Jan. 23, 1967).
- (3) Wiswesser, W. J., "The Empty Column' Revisited," *Computers and Automation* **19**, 4, 2-6 (1970).
- (4) Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill, New York, 1968.
- (5) Feldmann, R. J., Heller, S. R., Shapiro, K. P., and Heller, R. S., "An Application of Interactive Computing: A Chemical Information System," *J. Chem. Doc.* **12**, 41-7 (1972).
- (6) Miller, G. A., "Encoding and Decoding WLN," *Ibid.*, **12**, 60-7 (1972).
- (7) Feldmann, R. J., and Koniver, D. A., "Interactive Searching of Chemical Files and Structural Diagram Generation from Wiswesser Line Notation," *Ibid.*, **11**, 154-9 (1971).
- (8) Farrell, C. D., Chauvenet, A. R., and Koniver, D. A., "Computer Generation of Wiswesser Line Notation," *Ibid.*, **11**, 52-9 (1971).
- (9) Miller, G. A., "CARL—Chemical Algorithm for Reticulation Linearization," DCRT internal publication, August 1970.
- (10) Usdin, E., and Efron, D. H., "Psychotropic Drugs and Related Compounds," Second ed., U.S. Printing Office, Washington, D.C., in press, 1972.