Notes

Contribution from the Department of Chemistry,
The University of Texas at El Paso, El Paso, Texas 79968

Canonical Numbering, Stereochemical Descriptors, and Unique Linear Notations for Polyhedral Clusters

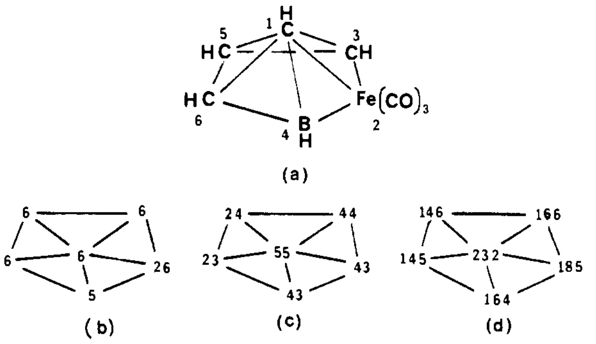
William C. Herndon* and Jack E. Leonard*[1]

Received April 19, 1982

The potential number of isomers in polynuclear cluster compounds is exceedingly large since isoelectronic–isolobal atom or group substitution[2-4] is in principle possible at any nuclear position in the cluster framework. Clusters may also have attached ligands that might themselves contain stereocenters or other stereochemically active moieties, and this increases the stereoisomeric possibilities. Finally, if two or more external substituents occur at a cluster atom, the orientation of the substituents may lead to stereoisomers. An ideal system of notation would allow one to obtain a unique name or other representation for every stereoisomer.

The problem of unique notations for cluster compounds is, of course, well recognized.[5,6] Graph-based nomenclature–notation systems, connection tables, or adjacency matrices generally distinguish only between constitutional isomers.[7-13] The stereochemical descriptors developed for organic systems[14] and coordination compounds[5,15,16] are not easily extended to polynuclear clusters. An appreciation of the difficulties that have been encountered in devising a general scheme can be obtained from two recent long papers by Casey et al.,[17,18] where rules are developed for naming the limited class of fully triangulated polyhedra (deltahedra)[19] containing at least one symmetry axis and/or plane. The suggested procedures[17,18] yield concise descriptive names for deltahedra with 4–24 vertices, but the rules for obtaining the names are lengthy and highly specific.

A compact linear form of a complete labeled adjacency matrix or connection table is suggested here as a useful general notation for polynuclear clusters. The explicit ordering of the listing will be shown to completely specify configuration at stereocenters within the cluster. The linear notation is written or printed with use of only numerical characters, hyphens, and parenthetical symbols, with no superscripts or subscripts, and with only a small number of formal rules. The procedure is easily carried out by hand, and the final listings can be used to provide unique registry numbers.



(e) (CH)0203040506(Fe(CO)$_3$)0304(CH)05(BH)06(CH)06(CH)

(f) (0601)0204060503(26(0608)-03)0304(0601)05(0501)06(0601)06(0601)

(g) 62465326683346556666

Figure 1. (a) Structure of $BC_4H_5Fe(CO)_3$ with canonical numbering according to (d). (b) Molecular graph labeled with atomic numbers. (c) First trial extended connectivity. (d) Final extended connectivity. (e) Constitutional linear notation with elemental symbols. (f) Linear notation with stereochemistry. (g) Unique identification number.

The application of our notation system must begin with a canonical numbering of the (three-dimensional) molecular graph of the cluster. Any one of the several algorithms[20-29] that have been devised to recognize constitutionally related structural features in molecular graphs may be used for this purpose. The procedure found most convenient uses a variant of the Chemical Abstracts Morgan algorithm,[20] which is similar to that suggested by Shelley and Munk.[22] In this variant the sum of an initial label and extended connectivity is iteratively calculated at each vertex of the molecular graph until the graph vertices have been partitioned into a sufficient number of equivalence classes. It can be shown that the Shelley–Munk algorithm and our variant are always convergent.[30] This is unlike the original Morgan rules, which in a few cases give oscillatory behavior or do not lead to a complete symmetry partitioning of molecular graph vertices.

Generally, nonequivalent vertices will have nonequivalent extended connectivities. The two-digit number 01 is assigned

---

(1) Present address: Department of Chemistry, Texas A&M University, College Station, TX 77843.
(2) Elian, M.; Chen, M. M. L.; Mingos, D. M. P.; Hoffmann, R. *Inorg. Chem.* **1976**, *15*, 1148.
(3) Mingos, D. M. P. *Adv. Organomet. Chem.* **1977**, *15*, 1.
(4) Brint, R. P.; Pelin, K.; Spalding, T. R. *Inorg. Nucl. Chem. Lett.* **1980**, *16*, 391.
(5) Sloan, T. E. *Top. Stereochem.* **1981**, *12*, 1.
(6) Fernelius, W. C. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 213.
(7) Lozach, N.; Goodson, A. L.; Powell, W. H. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 887.
(8) Goodson, A. L. *Chem. Inf. Comput. Sci.* **1980**, *20*, 167.
(9) Dittmar, P. G.; Stobaugh, R. E.; Watson, C. E. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111.
(10) Shelley, C. A.; Munk, M. E. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247.
(11) Tarjan, R. E. *ACS Symp. Ser.* **1977**, *No. 46*, 1.
(12) O'Korn, L. J. *ACS Symp. Ser. No. 46*, **1977**, 122.
(13) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E. "Computer Handling of Chemical Structure Information"; American Elsevier: New York, 1971.
(14) Cahn, R. S.; Ingold, C.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385.
(15) Thewalt, U.; Jenson, K. A.; Schaffer, C. E. *Inorg. Chem.* **1972**, *11*, 2129.
(16) Brown, U. F.; Cook, B. R.; Sloan, T. E. *Inorg. Chem.* **1975**, *14*, 1273; **1978**, *17*, 1563.
(17) Casey, J. B.; Evans, W. J.; Powell, W. H. *Inorg. Chem.* **1981**, *20*, 1333.
(18) Casey, J. B.; Evans, W. J.; Powell, W. H. *Inorg. Chem.* **1981**, *20*, 3556.
(19) Rudolph, R.; Pretzer, W. R. *Inorg. Chem.* **1972**, *11*, 1974.
(20) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107.
(21) Randić, M. *J. Chem. Inf. Comput. Sci.* **1976**, *15*, 105; **1977**, *17*, 171.
(22) Shelley, C. A.; Munk, M. E. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 110.
(23) Schubert, W.; Ugi, I. *J. Am. Chem. Soc.* **1978**, *100*, 37.
(24) Freeland, R. G.; Funk, S. A.; O'Korn, L. J.; Wilson, G. A. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 94.
(25) Jochum, C.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 113; **1979**, *19*, 49.
(26) Carhart, R. E. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 108.
(27) Uchino, M. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 116, 121.
(28) Randić, M.; Brissey, G. M.; Wilkins, C. L. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 52.
(29) Golender, V. E.; Drboglav, V. V.; Rosenblit, A. B. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 196.
(30) Herndon, W. C.; Davis, M. L.; Ellzey, M. L., Jr., to be submitted for publication. Also see ref 21.

0020-1669/83/1322-0554$01.50/0 © 1983 American Chemical Society

Chart I

group	no.	group	no.
(CH)	(0601)	(η^6-C_6H_6)	((0601)-06)
(CH_2)	(060101)	(η^5-C_5H_5)	((0601)-05)
(CH_3)	(0601-03)	(η^1-C_5H_5)	(06-0501-05)
(BH)	(0501)	(CO)	(0608)
(BH_2)	(050101)	($Fe(CO)_3$)	(26(0608)-03)
(CF_3)	(0609-03)	($P(C_6H_5)_3$)	(15(06-0601-05)-03)
(C_6H_5)	(06-0601-05)		

to the vertex with the largest value of extended connectivity, and the remaining vertices are ordered according to descending extended connectivity values. After numbering, a linear form of the labeled adjacency matrix is written by using the following rules.

(1) Diagonal elements a_{ii} of the cluster matrix are taken to correspond to the atomic number of the atom represented by vertex i in the molecular graph. These atomic numbers and numbers for attached noncluster groups appear sequentially in the linear code as two-digit characters enclosed in parentheses.

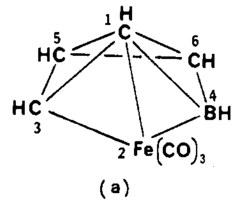
(2) Bonds of connectivities to adjacent cluster atoms are cited by a numeric string of two-digit vertex numbers following each enclosed set of atomic numbers. Only connections to higher numbered vertices are cited.

(3) For the specification of stereochemistry at a cluster atom, the order of citation of bonds is taken as clockwise when one views along an axis from the attached atom or group with the least priority in the numbering system. In the case of many polyhedral clusters this axis passes from outside the cluster through an *unnumbered* atom or group of atoms, through the specified cluster vertex, and into the body of the cluster.

(4) If there are two or more possible correct notations according to all preceding rules, then that numbering that is lowest at the first point of difference is chosen.

To avoid further abstract discussion and to illustrate some remaining notational conventions, we will give several examples. One of the postulated structures[31] for $BC_4H_5Fe(CO)_3$ is depicted in Figure 1. Numbering of the heteronuclear molecular graph commences with the atomic numbers of the atoms represented by each vertex in order to differentiate nonequivalence due to heterosubstitution (Figure 1b). The first trial extended connectivity of each vertex sets the connectivity value equal to the sum of the atomic numbers at adjacent vertexes plus the value at the vertex under consideration (Figure 1c). The process is continued till no further distinctions between vertices occur in two successive trials or until all vertices have been distinguished (Figure 1d). The resultant numbering is shown in Figure 1a. This canonical numbering is used to generate the alphanumeric linear notation without stereochemistry (Figure 1e). Substituting atomic numbers for element symbols and applying the clockwise stereochemical rule yields the notation with stereochemistry (Figure 1f). Either linear notation establishes the constitution, but the second representation is unique for the depicted stereoisomer.

The complete formula of the cluster is given by the atomic numbers enclosed in parentheses. Several types of polyatomic groups to be denoted within parentheses are given in Chart I. This list illustrates the convention that all atomic numbers are to be read as two-digit characters without intervening punctuation unless modified by hyphens. Hyphens that appear within parentheses indicate multiplicity of the preceding atom or group of atoms. Hyphens will also be used in a different way in large clusters with over 99 cluster vertices. In these cases the connectivity numbers (outside of parentheses) will be hyphenated if they require more than two digits. Thus a bond to cluster vertex 117 would be designated 01-17.

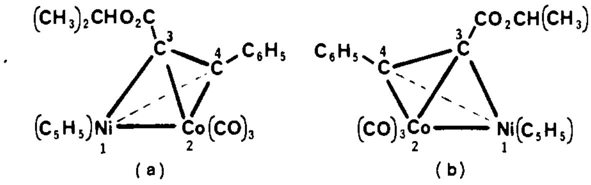


(b) (CH)0203040506(Fe(CO)$_3$)0304(CH)05(BH)06(CH)06(CH)

(c) (0601)0203050604(26(0608)-03)0403(0601)05(0501)06(0601)06(0601)

(d) 62356426683436556666

Figure 2. (a) $BC_4H_5Fe(CO)_3$ enantiomeric to compound in Figure 1a, canonical numbering. (b) Constitutional linear notation (identical with that of Figure 1e). (c) Linear notation with stereochemistry. (d) Unique identification number.



(c) (28(0601)-05)020403(27(0608)-03)0304(060608-020601(0601-03)-02)04 (06(06-0601-05))

(d) 286524327683346682663246665

(e) (28(0601)-05)020304(27(0608)-03)0403(060608-020601(0601-03)-02)04 (06(06-0601-05))

(f) 286523427683436682663246665

Figure 3. (c) Linear notation with stereochemistry for (a). (d) Unique identification number for (a). (e) Linear notation with stereochemistry for (b). (f) Unique identification number for (b).

The concatenated numerals in the linear notation with hyphens and parentheses deleted form a unique registry or identification number for the cluster compound. This definition is analogous to that used by Wipke and co-workers[32,33] in defining the SEMA numerical name[34] for organic compounds, and a proof for uniqueness can be obtained as in this earlier work.[32] Deletion of initial 0's and 1's in the linear number gives a still shorter number that is also unique (Figure 1g). This identification number could be further shortened by standard hash techniques[35] if this is considered to be desirable.

The ferrocarborane shown in Figure 1 is chiral; the linear notation for its enantiomer is given in Figure 2. An odd number of permutations of the bond connectivity numbers is required to interconvert the notations for the two enantiomers. However, this is not a general requirement as can be demonstrated with use of the small tetranuclear chiral compounds[36] depicted in Figure 3. With the assumption of the Cahn-Ingold-Prelog rules[14] for assigning priority in the organic moieties, the cluster vertices are capable of being numbered canonically without being examined for extended connectivity. One then sees that the notations for the two enantiomers differ by only two permutations of the bond connectivity numbers. Therefore, although the procedures outlined will generate a unique notation for each stereoisomer, at present no simple

---

(31) Fehlner, T. P. *J. Am. Chem. Soc.* **1978**, *100*, 3250.

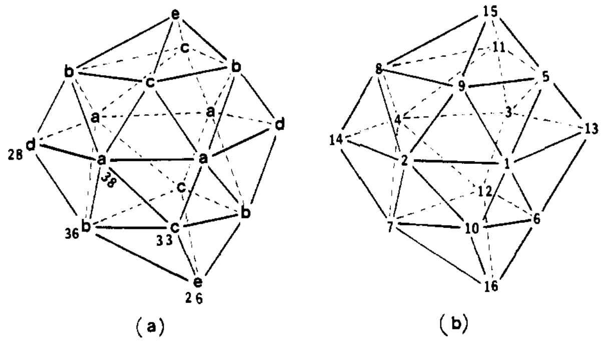
(32) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4834.

(33) Choplin, F.; Marc, R.; Kaufman, G.; Wipke, W. T. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 110.

(34) *S*tereochemical *E*xtension of the *M*organ *A*lgorithm.[32]

(35) Wipke, W. T.; Krishnan, S.; Ouchi, G. I. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 32.

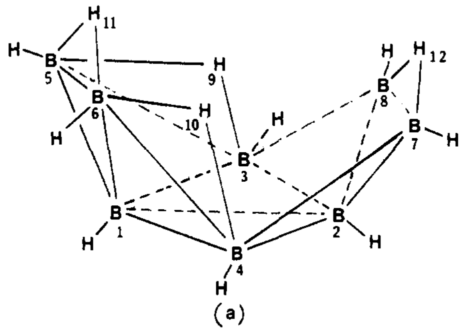
(36) Gérard, J.; Marinetti, A.; Saillard, J.-Y.; Sayer, B. G.; McGlinchey, M. J. *Organometallics* **1982**, *1*, 225.



(c) (0501)0209051306l0(0501)1007140809(0501)0412130511
(0501)1108140712(0501)09151113(0501)13121610
(0501)10161214(0501)141115(0501)15(0501)16(0501)15
(0501)16(0501)(0501)(0501)(0501)

(d) 5295136105107148954121351151181471259151113513121610
51016121451411155155165155165555

Figure 4. (a) Molecular graph, equivalent vertices, and extended connectivities for $B_{16}H_{16}^{2-}$. (b) Canonical numbering. (c) Linear notation with stereochemistry. (d) Unique identification number.



(b) (0501)0203050604(0501)04070803(0501)08095
(0501)061007(0501)091106(0501)1110(0501)1208(0501)12
(01)(01)(01)(01)

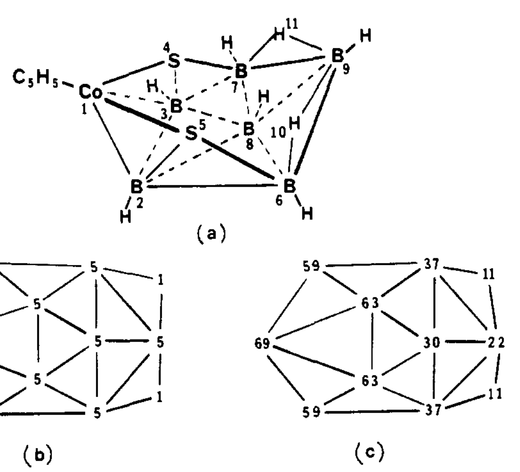
(c) 523564547835895561075911651110512085l2

Figure 5. (a) Molecular structure and canonical numbering for B_8H_{12}. (b) Linear notation with stereochemistry. (c) Unique identification number.

algorithm presents itself for the pairing of enantiomeric structures.

The 16-vertex deltahedron[17] $(B_{16}H_{16})^{2-}$ shown in Figure 4 has a large set of alternative numberings. Extended connectivity divides the skeleton into five equivalence classes of vertexes labeled a-e in Figure 4a. The membership in each of these classes is 4, 4, 4, 2, and 2, respectively. Thus, there are in principle 55 296 (=4:4:4:2:2:) possible numberings for the molecular graph. However, once we designate any one of the four possible vertices in group a as vertex 01, application of the right-hand rule (and the requirement that the numbering for class b vertices begins with 05, for class c with 09, etc.) leads to a single unique numbering for this skeleton.

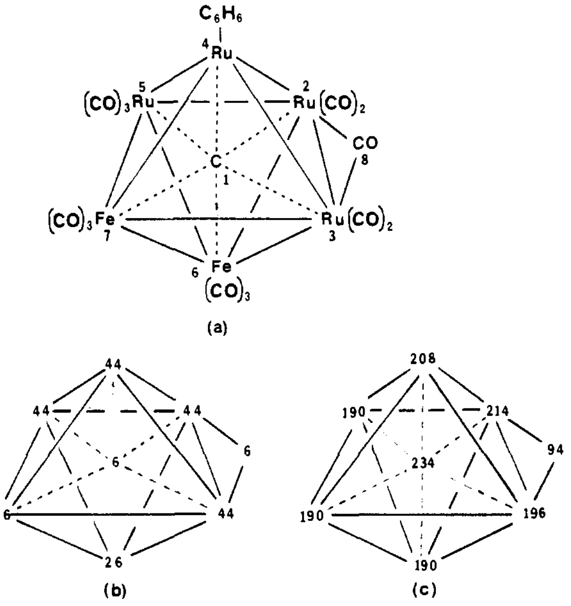
The borane structure[37] for B_8H_{12} given in Figure 5 demonstrates some of the general features of the canonical numbering system. In agreement with the results of other systems[20-22] that use extended connectivity as the basic numbering tool, high connectivity and a central position in the molecular graph, along with high atomic number, are found to be the main factors leading to high priority. As a result, ensuing numbers tend to cluster close to preceding numbers, and



(d) (27(0601)-05)02030405(0501)05060803(0501)080704(16)07(16)06(0501)
100908(0501)080911(0501)09(0501)1011(01)(01)

(e) 27652345556835874167166510985891159510ll

Figure 6. (a) Molecular structure and canonical numbering for $Co(C_5H_5)S_2B_6H_8$. (b) Molecular graph labeled with atomic numbers. (c) Extended connectivity. (d) Linear notation with stereochemistry. (e) Unique identification number.



(d) [(06)0203040506}(44(0608)-02)0304050608(44(0608)-02)06070408
(44(0601)-06)0705(44(0608)-03)0706(26(0608)-03)07(26(0608)-03)(0608)

(e) 6234564468234568446826748446675446837626837268368

Figure 7. (a) Molecular structure and canonical numbering for $Ru_4Fe_2C(CO)_{14}C_6H_6$. (b) Molecular graph with atomic numbers. (c) Extended connectivity. (d) Linear notation with stereochemistry. (e) Unique identification number.

bridging hydrogen atoms or other divalent bridging groups ordinarily will appear last in the order of citation. In many cases this could allow the assignment of the correct canonical labels without actually proceeding through the numbering algorithm. The octaborane numbering was assigned in this manner, but the correct numbering is more difficult to perceive if several heteroatoms are present as illustrated for the cobaltadithiaborane[38] in Figure 6.

Clusters containing internal atoms can be treated by including the bonding to the internal atom or atoms as part of the molecular graph and within the notation system. The hypothetical $Ru_4Fe_2C(CO)_{14}C_6H_6$ carbide cluster given in

---

[37] Lipscomb, W. N. "Boron Hydrides"; W. A. Benjamin: New York, 1963.

[38] Zimmerman, G. J.; Sneddon, L. G. *J. Am. Chem. Soc.* **1981**, *103*, 1102.

Figure 7 is an example of such a compound with a structure analogous to that of a known[39] $Ru_6C(CO)_{14}$(mesitylene) complex. The stereochemistry is uniquely defined by the order of listed bonds at each of the external cluster atoms as in the previous examples, so it is not necessary to establish a stereochemical descriptor to define stereochemistry at the carbide carbon. Consequently, we simply list vertices attached to carbon in ascending order, realizing that the remaining notational ordering will uniquely define the enantiomer being considered. Bonding information where stereochemistry is not defined will be designated by enclosure in brackets.

These examples should suffice to show that this linear notation system provides a workable method to denote structures and gives unique registry numbers for any cluster residue. Some advantages of the system are that it requires less learning effort than standard nomenclature systems, the number of arbitrary rules and conventions is small, and the system is not limited to particular types of polyhedra with particular types of symmetry. It is also obvious that the numeric form of the notation system makes it suitable for computer-oriented search operations, and we think that the ability of the method to allow the laboratory chemist to obtain a unique registry or identification number could prove to be an additional useful feature.

One important disadvantage is that a good drawing or model of the compound to be denoted is required before the correct stereochemical notation can be obtained. If the absolute configuration at a particular chiral center in unknown, or if one is dealing with a racemic mixture, the notation is too specific; one name is required for each possible isomer. We suggest that brackets be used both where stereochemical information is redundant (as above) and also where such information is missing. A compound whose constitution is known but whose stereochemistry is yet to be determined would then be indicated by bracketing the entire linear notation. In the case of racemic mixtures we suggest that one simply list the notations for both enantiomers.

The question as to whether or not the proposed scheme is completely general and applicable to all compounds must obviously be answered no. It turns out that the stereochemistry of a complex cluster is actually rather easy to denote because the extensive three-dimensional cluster bonding restricts the available molecular shapes and conformations. These restrictions would not hold, for example, for inorganic or organic moieties attached to the cluster where a conformational description might be required. An illustration is provided by the restricted staggered X-ray structure of the metallodicarborane $(8\text{-}CH_3OC_2B_9H_{10})_2Ni$ found by Janousek et al.[40] They have proposed a system of locants and descriptors to treat the stereochemistry of cases of this type, but clearly a more general method is desirable. We recognize this problem, and we have devised simple extensions of the present rules to provide linear notations including conformational isomerism, which will be reported in a full paper. Suggestion for the application of the linear notation system to coordination compounds and bridged dinuclear or polynuclear compounds will also be advanced.

Finally, following suggestions of reviewers, we would like to comment more specifically on potential uses of this notation scheme. We do not suggest that it provides descriptive names to be used in research papers or textbooks. Primarily the linear notation uniquely specifies connectivity and stereochemistry of a three-dimensional cluster of atoms, and the numerical notation itself or a shortened version is a unique registration number for the compound. As such it can be used to compare structures to establish whether or not they are identical, or it can be used in computer searches of files of cluster compounds, with the assumption that clusters have been catalogued with this method. The advantages of structure-generated registry numbers that can be determined by the user rather than sequentially assigned registry numbers (Chemical Abstracts) are factors to be considered in assessing utility.

In large cluster compounds, the notation also does not easily allow by-hand reconstruction of the structure from the linear sequence of atomic symbols and connectivity numbers. However, the linear notation is tantamount to a connection table augmented with stereochemical descriptors, and a computer program capable of producing a three-dimensional representation from this type of input data has recently been described by Wenger and Smith.[41] In principle then, the visualization of structures using the linear notation is a solved problem. Practical procedures associated with implementation of the visualization process and with the problems of searching for classes of clusters or particular substructures are under investigation.

Acknowledgment. Many useful detailed conversations with colleagues Michael Davis, M. Lawrence Ellzey, Jr., Keith H. Pannell, and Richard L. Armstrong helped in developing the concepts presented in this paper. We also gratefully acknowledge the financial assistance of the Robert A. Welch Foundation of Houston, TX.

---

(39) Mason, R.; Robinson, W. R. *Chem. Commun.* **1968**, 468.

(40) Janousek, Z.; Plesek, J.; Hermanek, S.; Base, K.; Todd, L. J.; Wright, W. R. *Collect. Czech. Chem. Commun.* **1981**, *46*, 2818.

(41) Wenger, J. C.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 29.

Contribution from the Laboratoire des Acides Minéraux,
LA 79, CNRS, Université des Sciences et Techniques
du Languedoc, 34060 Montpellier Cédex, France, and
Institute of Chemical Engineering, University of Palermo,
90128 Palermo, Italy

A Bent, Staggered $(\mu\text{-}H)[Cr_2(CO)_5]_2^-$ Monoanion. Crystal Structure of $KHCr_2(CO)_{10}$

J. Roziere,*[1a] P Teulon,[1a] and M. D. Grillone[1b]

Received March 12, 1982

In recent years the nature of the metal–hydrogen–metal bond in the binuclear transition-metal complexes $(\mu\text{-}H)[M(CO)_5]_2^-$ (M = group 6B metal) and related compounds has attracted much attention. Particular interest in the chemical reactivity of such systems stems in part from their use as reagents for the synthesis of organometallic derivatives as well as from their potential activity in catalytic systems.[2,3] For instance, these species have recently been observed in solutions of $M(CO)_6$ and potassium hydroxide, which serve as homogeneous catalysts for the water gas shift reaction.[4] From the structural viewpoint, previous investigations have placed particular emphasis on the MHM bridge geometry in relation to the extent of direct metal–metal overlap in the three-center two-electron-bridged hydride bond.

As discussed in a review of structural studies on transition-metal hydride complexes,[5] the geometry of group 6B

---

(1) (a) Université des Sciences et Techniques du Languedoc. (b) University of Palermo.

(2) Grillone, M. D.; Kedzia, B. B. *J. Organomet. Chem.* **1977**, *140*, 170.

(3) Darensbourg, M. Y.; Atwood, J. L.; Burch, R. R.; Hunter, W. E.; Walker, N. *J. Am. Chem. Soc.* **1979**, *101*, 2631.

(4) Darensbourg, D. J.; Darensbourg, M. Y.; Burch, R. R., Jr.; Froelich, J. A.; Incorvia, M. J. *Adv. Chem. Ser.*, **1979**, *No. 173*, 106.

(5) Bau, R.; Teller, R. G.; Kirtley, S. W.; Koetzle, T. F. *Acc. Chem. Res.* **1979**, *12*, 176.

0020-1669/83/1322-0557$01.50/0 © 1983 American Chemical Society