

## Chemical Combinatorics for Alkane-Isomer Enumeration and More

L. Bytautas<sup>†</sup> and D. J. Klein\*

Texas A&M University at Galveston, Galveston, Texas 77553-1675

Received May 18, 1998

Standard combinatorial enumeration techniques for alkanes are considered with a view to the extension to a widened range of chemically interesting features. As one brief point it is noted that these standard techniques naturally associate to generational schemes and thence have nomenclatural interpretations, which may be made to achieve some similarity to the standard IUPAC nomenclature. Our primary focus is the illustration that such combinatorial techniques are sufficient to enable computation of several graph-theoretic structural invariants averaged over (different types of) isomer classes. Such averages (and associated isomer counts) are tabulated for structural isomers for up to  $N = 40$  carbons, where there are  $\sim 10^{14}$  isomers (though the computational methodology should rather readily extend to at least  $N = 80$  where there should be  $\sim 10^{28}$  isomers). The averages for invariants are utilized to estimate several physicochemical properties averaged over these same isomer classes. The properties currently so considered are heat of formation, index of refraction, and magnetic susceptibility. Further, various asymptotic results for counts, mean invariants, and mean properties are noted, so that the exact graph-theoretic data are extrapolated with high accuracy to arbitrarily large alkanes.

### 1. INTRODUCTION

One of the classic areas of chemistry concerns isomerism, starting in the 1860s, when it was realized (as in ref 1) that there were different possible structural formulas for the same atomic composition. That such different structural formulas corresponded to different compounds provided crucial evidence for the validity of classical structural formulas and the existence of structured molecules. The problem of formal isomer enumeration was initiated by Cayley<sup>2</sup> in 1874 for the case of alkanes, and this continued as a topic of interest for some time,<sup>3</sup> particularly with the group of Henze<sup>4</sup> making several enumerations, for alkanes and homologous sequences of various derivatives. Then in 1935 motivated by the chemical isomer problem Pólya<sup>5,6</sup> developed a powerful combinatorial theory for the enumeration of symmetry-mediated equivalence classes of “colorings”, and this enumeration theory has now become standard fare in combinatorics texts. Beyond the formal mathematical theory, Pólya applied his theory to a few chemical problems, but in particular the problem of alkane isomer enumeration, and following Pólya there have been further (often very formal) refinements for this chemical problem in numerous papers, e.g., as refs 7–16. Work has been done emphasizing<sup>7,8,10,13</sup> the generality of the method in dealing with additional systems beyond alkanes, and there has been (a lesser amount of) work<sup>5,9–12</sup> dealing with the asymptotic behavior of the isomer counts, primarily for alkanes. The pre-1986 work on Pólya enumeration for chemical purposes is nicely reviewed by Read<sup>7</sup> who also gives a translation (made by D. Aepli) of Pólya’s foundational paper, and Fujita<sup>13</sup> considers in detail many recent theoretical extensions (primarily concerning chirality and symmetry questions not

only for isomers but also for reaction processes). Yeh<sup>17</sup> and Cyvin and co-workers<sup>18,19</sup> have recently repeatedly noted that much of Pólya’s general mathematical formalism can be foregone in many specific isomer enumerations—that is, whole chapters often occurring in standard combinatorics texts concerning formal consideration of permutation groups, general “cycle index” functions, and “pattern inventories” as well as general theory of implicit and holomorphic functions often need not be gone through for many specific cases.

Here we adopt this straightforward point of view for the classic case of alkanes, thereby refining some earlier results, and we develop possibilities for some extensions to new chemically oriented applications beyond simple isomer enumeration. Initially in section 2 the classical enumeration problem is reconsidered using such a straightforward approach. It is noted in passing that there is a potential relation of these mathematical enumeration procedures to structure generation and associated nomenclatural systems, and this is detailed in the Appendix. The primary aim here of developing a more widely useful combinatoric chemistry is pursued in sections 3 and 4 where we show how the standard combinatorial enumerative techniques can be readily extended to compute average values of certain graph-theoretic invariants, the averages being over the members of isomer classes. A few averages of some interest in “polymer statistics” are considered, including average numbers of conformations per isomer, the average through-bond diameter of a class of isomers, and the average through-bond (or shortest-path) distance between carbons in a class of isomers. This last quantity is obtained via a computation of the average so-called<sup>20,21</sup> “Wiener number” for isomeric alkanes. Also average counts of primary, secondary, tertiary, and quaternary carbons are made. Such graph-theoretic invari-

<sup>†</sup> On leave from Institute of Theoretical Physics & Astronomy, Gostauto 12, 2600 Vilnius, Lithuania.

ants counting different types of local structures have long been used in chemistry to linearly expand (and thereby compute) certain properties, e.g., as in refs 22–26. This approach is sometimes called<sup>24</sup> the “group-function” method (at least in application to heats of formation), though we prefer the phrase “cluster-expansion”<sup>25,26</sup> (which is a nomenclature more frequently used in expansions in other contexts). Here we use this to estimate values for heats of formation, indices of refraction, and magnetic susceptibilities averaged over structural isomer classes. Yet further techniques are utilized to obtain standard deviations over isomer classes for these properties. A variety of tabulated results for classes with up to a few trillions of members are reported. Such tabulations of average values for properties then provide a more thorough view of the overall class of compounds and indicate how various properties vary with isomer composition. Comparison of the standard deviations within isomer classes to the differences of averages between classes provides a measure of a degree of physicochemical distinction arising from the isomer classification. Sections 5 and 6 address the (many-atom) asymptotic behaviors of many of the computed properties, extending the applications earlier described by Pólya<sup>5</sup> and Otter.<sup>9</sup> Some polymer-statistical questions concerning conformational “sizes” are further discussed. Overall a useful theoretical “combinatoric” chemistry seems possible. With the present focus on properties one may perceive an overlap of this classical isomer-enumerative area of chemistry with

- \* modern “polymer statistics”, e.g., in ref 27,
- \* recent “dendrimer chemistry”,<sup>28</sup>
- \* even more recent “combinatorial chemistry”,<sup>29</sup> and
- \* discrete statistical-mechanical enumeration problems, e.g., in ref 30.

Thence though previously little explored such combinatoric chemistry seems of much potential use in several areas.

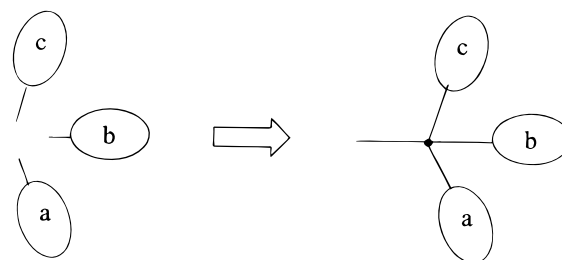
The key mathematical technique utilized here is that of “generating functions”. Generally in such a framework it is imagined that certain counts (as of alkane isomers) are collected together with each count appearing as a coefficient of a corresponding monomial powers of a “dummy” variable in the generating function. The idea is much more general than Pólya enumeration theory and is often used in combinatorics, where standard function manipulations often enable neatly encoded recursions for the coefficients of interest. Such generating functions also appear much like discrete-state partition functions in statistical mechanics, where the dummy variable appears as an “activity”. Some general ideas of use here are standard fare in several combinatorics texts, as ref 31. But the focus here is within the context of alkane-isomer examples, as we proceed with in the next section. And again the manipulations are presented with a minimum of mathematical overhead (i.e., without formal need for permutation or automorphism groups, group-action orbits, cycle-index functions for general permutation groups, double cosets, coset subduction, mark tables, etc.). As such the present work can be viewed as a direct extension of the classical pre-Pólya work, especially of Henze’s group<sup>4</sup>—the fuller merger with Pólya enumeration theory (now rather

extended as in refs 8, 10, 13) should be of use in the future in treating even more general problems.

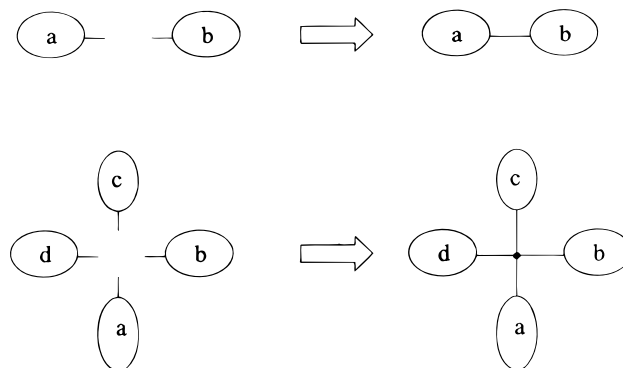
## 2. ENUMERATION METHODOLOGY

As already mentioned alkane enumeration has long been a standard theoretical problem, with several tabulations of isomer counts available.<sup>2–4,7,14,16</sup> The generating-function formalism too has been repeatedly developed, though less often<sup>12,15</sup> in the manner considered here relating to generation numbers and (through-bond) diameters. In all these treatments the enumeration is first made of alkyl radicals or graph-theoretically speaking in terms of *rooted* trees (each defined as a tree with a distinguished site called a *root*, which represents the radical site with a dangling bond). The graphs are “hydrogen-deleted” in the sense that only the carbons atoms and the C–C bonds are explicitly represented. Now for alkane-type case with the degree of a site restricted to no more than 4, an  $n$ -site rooted tree may be obtained from three smaller rooted  $a$ -,  $b$ -, and  $c$ -site rooted trees (where  $a + b + c = n - 1$  and  $a, b, c \geq 0$ ) by the process indicated in Figure 1. Once the various rooted trees are obtained they are then joined together to form alkanes as shown in Figure 2. For such rooted trees one may attend to *generation number*  $g$  defined as the number of carbon sites in a longest path from the root. Exactly what is done with this general idea depends on just what type of isomer one wishes to consider.

The first classical considerations were in terms of structural isomers, as we now also consider. The generating function



**Figure 1.** An illustration of the constructive manner considered for the iterative generation of alkyl radicals. Here three alkyl radicals (denoted  $a$ ,  $b$ , and  $c$ , with the possibility that some of them might be the zero-order alkyl radical consisting of a single H-atom with a dangling bond) are joined at a new radical site.



**Figure 2.** Two manners of joining generation- $g$  radicals so as to obtain an alkane: in (a) two generation- $g$  radicals are joined at a common bond; and in (b) four radicals are joined at a central site, with the constraint that two are of generation- $g$  while the two remaining ones have a generation number  $\leq g$ .

for degree-4-limited rooted trees of a given generation number  $g$  is defined as

$$\phi_g(t) \equiv \sum_{n \geq 0} \#_{g,n} t^n \quad (2.1)$$

where  $\#_{g,n}$  is the number of rooted trees with sites of degree  $\leq 4$  and  $t$  is a “dummy variable”. The construction of Figure 1 leads to a recursion for these generating functions

$$\phi_{g+1}(t) = t\{\phi_g(t)Z_2[\phi_{<g};t] + Z_2[\phi_g;t]\phi_{<g}(t) + Z_3[\phi_g;t]\} \quad (2.2)$$

where

$$\phi_{<g+1}(t) \equiv \phi_{<g}(t) + \phi_g(t)$$

$$Z_2[f;t] \equiv \{[f(t)]^2 + f(t^2)\}/2$$

$$Z_3[f;t] \equiv \{[f(t)]^3 + 3f(t)f(t^2) + 2f(t^3)\}/6$$

$$Z_4[f;t] \equiv \{[f(t)]^4 + 6[f(t)]^2f(t^2) + 8f(t)f(t^3) + 3[f(t^2)]^2 + 6f(t^4)\}/24 \quad (2.3)$$

with  $\phi_{<1}(t) \equiv 1$  and  $f$  is some function. The factor  $t$  in the recursion (2.2) arises because a new site (which is to be the radical site) is added not already occurring in the  $a$ -,  $b$ -, or  $c$ -groups of Figure 1. The rationale for the three different terms of (2.2) is best discussed separately for each. The first term concerns the case of Figure 1 where the  $a$ -group is of generation number  $g$ , while the  $b$ - and  $c$ -groups have generation numbers  $<g$ . Such  $b$ - and  $c$ -groups are accounted for in the  $[\phi_{<g}(t)]^2$  term of  $Z_2[\phi_{<g};t]$  with the division by 2 being appropriate if the  $b$ - and  $c$ -groups are distinct (since each of these groups may occur in connection with either of the  $\phi_{<g}(t)$  factors in  $[\phi_{<g}(t)]^2$ ). With identical  $b$ - and  $c$ -groups also arising in  $[\phi_{<g}(t)]^2$  one would not wish to divide by 2 (as we have done to correctly count the circumstance with them nonidentical), so to correct for this miscount by a factor of 1/2 for these identical groups, one adds on  $\phi_{<g}(t^2)/2$ , it being realized that  $\phi_{<g}(t^2)$  by itself makes correct counts for the case with both  $b$ - and  $c$ -groups identical (and of generation number  $<g$ ). The second term of (2.2) concerns the circumstance with the  $a$ - and  $b$ -groups having generation number  $g$ , while the  $c$ -group has generation number  $<g$ , and the rationale for  $Z_2[\phi_g;t]$  is much the same as for  $Z_2[\phi_{<g};t]$  in the first term. The third term of (2.2) concerns the case where the  $a$ -,  $b$ -, and  $c$ -groups are all of generation number  $g$ , and now up to three of these groups may be identical and the different terms of  $Z_3[\phi_g;t]$  may be somewhat similarly rationalized. The functional  $Z_4[f;t]$  has a similar utility in later manipulations. In any event it may also be noted that in fact these functionals  $Z_n[f;t]$  are permutation-group  $S_n$  “cycle-index functions” of Pólya.<sup>5-7</sup>

The alkyl-radical generating functions at each different generation number may be used to obtain generating functions for the nonradical alkanes. This is imagined from joining different alkyl radicals together at the “center” of an alkane. Here the center of a tree graph may be precisely defined (in a standard fashion) as the last nonzero graphical remnant remaining from successive *prunings*, each pruning entailing the deletion of all end (degree-1) vertexes. The

joining together of two generation- $g$  radicals at a central bond yields an alkane of diameter  $2g - 1$ , with a tree's diameter being precisely specifiable as the number of bond steps between two sites of the tree. If instead the center is a single site, then there being four places of attachment at a site, four radicals are joined to it with two of the radicals of  $g$  generations, while the other two have generation number  $\leq g$ , thereby yielding alkanes of diameter  $2g$ . These odd- and even-diameter constructions are respectively indicated in Figure 2a, b. With the introduction of a generating function  $P_D(t)$  for alkane counts, these center-based constructions then lead to the relations

$$P_{2g-1}(t) = Z_2[\phi_g;t]$$

$$P_{2g}(t) = t\{Z_2[\phi_g;t]Z_2[\phi_{<g};t] + Z_3[\phi_g;t]\phi_{<g}(t) + Z_4[\phi_g;t]\}$$

Here for  $P_{2g}(t)$  the first, second, and third summands respectively concern center-site constructions (of Figure 2(b)) with 2, 3, and 4 generation- $g$  radicals, while there are respectively 2, 1, and 0 radicals of generation number  $<g$ . The overall generating function for all alkanes then is

$$P(t) \equiv \sum_{D \geq 0} P_D(t) \quad (2.5)$$

The auxiliary functions  $\phi_g(t)$  and  $P_D(t)$  are also useful in developing various properties (in section 3). The final generating function when expanded out in powers of  $t$  as

$$P(t) \equiv \sum_{N \geq 0} \#'_N t^N \quad (2.5)$$

with coefficients  $\#'_N$  counting the number of alkane structural isomers.

The derivation closely parallels the usual ones following Pólya, but we build on these formulas and ideas in the following sections. Also most commonly generation number and diameter are sought (excepting in refs 12 and 15) to be avoided, though in fact they really are rather directly involved, and used in the following sections. Perhaps to it should be noted that the previous work on the enumeration of stereoisomers extends in much the same way as is here done for structural isomers. (It primarily involves new cycle index functionals  $Z'_n[f;t]$  which generally have fewer terms than the  $Z_n[f;t]$  of (2.3), but we do not presently consider this stereoisomeric case.)

The use of these relations is best accommodated via computer manipulation of polynomials. That is, the coefficients ( $\#_{g,N}$  and  $\#_N$ ) of the various powers of  $t$  are stored as vectors, and these are manipulated in manners simulating the various operations involving polynomials as involved in the recursions. This may be done quite explicitly using standard computer languages as FORTRAN or presumably even more readily by any one of several widely available mathematics manipulation programs such as “Mathematica” or “Maple”. The traditional enumerations are then just the coefficients ( $\#'_N$ ) of  $t^N$  in the final polynomial  $P(t)$ , which is not computed in totality but just up through some highest considered diameter  $D$ , known to be sufficiently great that no  $N$ -atom isomers have a longer diameter.

Beyond enumeration there are problems of generating all possible isomeric structures and of assigning them systematic

names. The alkanes are one of the first classes of molecules for which the nomenclatural problem was addressed, and typically this class is still the first so considered even in modern introductory organic chemistry courses. The IUPAC rules are set down,<sup>32</sup> but still there are some supposed problems.<sup>33</sup> At a more mathematical stage of “codes” (which we might simply identify as labels without a convenient verbal pronunciation) there are general mathematical suggestions, e.g., as in refs 34 and 35. The recursive enumeration method is directly associated to a generation scheme (as suggested in connection with Figure 1), which in turn links directly to a systematic conceivable nomenclature. This we note here in passing, and a more detailed consideration is presented in the appendix. Indeed we surmise that a systematic nomenclature implies and is implied by a systematic enumeration procedure.

### 3. RECURSIVE PROPERTY COMPUTATION

The generating function techniques of the preceding section may be extended to compute mean values for a variety of graph invariants, as the numbers of primary, secondary, tertiary, and quaternary carbons, and these in turn may be used to estimate mean values of various physico-chemical properties. For these four *atom types* one can simply imagine (much as in ref 16) four different *t*-variables  $t_1, t_2, t_3, t_4$  and an alkyl-radical generating function  $\phi_g(t_1, t_2, t_3, t_4) \equiv \phi_g(t)$  where the powers of  $t_1, t_2, t_3, t_4$ , respectively, identify the numbers of primary, secondary, tertiary, or quaternary carbons in the species being counted (with the radical site being assigned a type as though the dangling bond is attached to another carbon, as it will be when center joinings are made). The point is that the recursion for these  $\phi_g(t_1, t_2, t_3, t_4)$  is rather like that of (2.2) for the earlier  $\phi_g(t)$ —now in place of the initial *t* factor in (2.2) there now is a  $t_d$  in correspondence with the (H-deleted graphical) degree *d* of the root site—a mild complication occurs because one needs to separate from  $\phi_{<g}(t_1, t_2, t_3, t_4)$  the terms  $\phi_0(t_1, t_2, t_3, t_4) = 1$  to properly account for the root-site type. That is, in terms of  $\phi_{<g}^-(t) \equiv \phi_{<g}(t) - 1$

$$\begin{aligned} \phi_{g+1}(t) = & t_2 \phi_g(t) + t_3 \{ \phi_g(t) [\phi_{<g}^-(t)] + Z_2[\phi_g; t] \} + \\ & t_4 \{ \phi_g(t) Z_2[\phi_{<g}^-; t] + Z_2[\phi_g; t] [\phi_{<g}^-(t)] + Z_3[\phi_g; t] \} \end{aligned} \quad (3.2)$$

These alkyl-radical generating functions then are combined much as in (2.4) to give overall alkane enumerations, with the powers of the different  $t_d$  recording the numbers of different types of atoms. This recursion could be used to generate all the individual  $\#_{g,N}$ , but this leads to many different indexes, and a “mess”. Rather we proceed by way of derivative generating functions

$$t_d \cdot \partial_d \phi_g(t) = \sum_N N_d \#_{g,N} t^N \quad (3.3)$$

where  $\partial_d$  indicates a derivative with respect to  $t_d$ ,  $\mathbf{N} \equiv (N, N_1, N_2, N_3, N_4)$  with *N* the total number of atoms and  $N_d$  the number of degree-*d* atoms,  $\#_{g,N}$  the count of such alkyl radicals, and  $t^N$  an abbreviation for the product of the  $t_d$  to the powers  $N_d$ . Of course these derivative generating functions may be computed recursively via a relation

obtained by taking derivatives of (3.2) at any given value for the  $t_d$  variables. The explicit result is most easily expressed using some additional notation, first

$$\begin{aligned} F_2[\phi_g; t] & \equiv \phi_g(t) \\ F_3[\phi_g; t] & \equiv \phi_g(t) [\phi_{<g}^-(t)] + Z_2[\phi_g; t] \\ F_4[\phi_g; t] & \equiv \phi_g(t) Z_2[\phi_{<g}^-; t] + Z_2[\phi_g; t] [\phi_{<g}^-(t)] + \\ & Z_3[\phi_g; t] \end{aligned} \quad (3.4)$$

for the bracketed terms in (3.2) (so that now  $\phi_{g+1}(t)$  is a *d*-sum of  $t_d F_d[\phi_g; t]$ ) and next their derivatives

$$\begin{aligned} \partial_d F_2[\phi_g; t] & \equiv \partial_d \phi_g(t) \\ \partial_d F_3[\phi_g; t] & \equiv [\partial_d \phi_g(t)] [\phi_{<g}^-(t)] + \phi_g(t) [\partial_d \phi_{<g}^-(t)] + \\ & \partial_d Z_2[\phi_g; t] \\ \partial_d F_4[\phi_g; t] & \equiv [\partial_d \phi_g(t)] Z_2[\phi_{<g}^-; t] + \phi_g(t) (\partial_d Z_2[\phi_{<g}^-; t]) + \\ & (\partial_d Z_2[\phi_g; t]) [\phi_{<g}^-(t)] + Z_2[\phi_g; t] [\partial_d \phi_{<g}^-(t)] + (\partial_d Z_3[\phi_g; t]) \end{aligned} \quad (3.5)$$

The  $\partial_d Z_n[f; t]$  are similar sorts of derivatives (based on the formulas of (2.3)). Then the recursion relations for the desired derivative generating functions are

$$\partial_d \phi_{g+1}(t) = t \cdot F_d[\phi_g; t] + \sum_e t_e \cdot (\partial_d F_e[\phi_g; t]) \quad (3.6)$$

which we imagine is applied at  $t_1 = t_2 = t_3 = t_4 = 1$  (with the variables  $t_d$  having been introduced solely so that the formal derivatives could be taken, after which all are set to 1). That is, numerically one works with recursions for single-variable polynomials

$$\partial_d \phi_{g+1}(t) = t \cdot F_d[\phi_g; t] + \sum_e t_e \cdot (\partial_d F_e[\phi_g; t]) \quad (3.7)$$

and these derivative functions for alkyl radicals may be used to obtain corresponding derivative functions  $\partial_d P_D(t)$  and  $\partial_d P(t)$  for alkanes, by taking derivatives of (2.4) and (2.5). Defining the coefficient of  $t^N$  in  $\partial_d P(t)$  to be  $\partial_d \#_N$  one may obtain the mean number of type-*d* atoms in isomeric *N*-atom alkanes as

$$\langle N_d \rangle_N = (\partial_d \#'_N) / \#'_N \quad (3.8)$$

It may be noted that these different means are not “independent”, e.g., since a sum over *d* of these  $\langle N_d \rangle_N$  must evidently sum to *N*. But also there is another relation, since every time a branch is introduced another end is simultaneously introduced. That is

$$\begin{aligned} \langle N_1 \rangle_N + \langle N_2 \rangle_N + \langle N_3 \rangle_N + \langle N_4 \rangle_N & = N \\ \langle N_3 \rangle_N + 2 \langle N_4 \rangle_N + 2 & = \langle N_1 \rangle_N \end{aligned} \quad (3.9)$$

which can be either used to avoid two of the constructions or else to check the numerical results from the constructions.

The ideas of the preceding paragraph can be extended to obtain standard deviations for the numbers of atoms of each different type. This involves taking second derivatives, so



that in addition to (3.7) one has

$$\partial_d^2 \phi_{g+1}(\mathbf{t}) = 2t \cdot (\partial_d F_d[\phi_g; \mathbf{t}]) + \sum_e t e \cdot (\partial_d^2 F_e[\phi_g; \mathbf{t}]) \quad (3.10)$$

and finally also  $\partial_d^2 P_D(t)$  and  $\partial_d^2 P(t)$ , the latter of whose coefficients might be denoted  $\partial_d^2 \#'_N$ . Then (remembering that if the first derivative brings down a power  $N_d$  the second derivative overall brings down an overall factor of  $N_d(N_{d-1})$ ) the mean square values for numbers of different atom types are

$$\langle (N_d)^2 \rangle_N = \{(\partial_d^2 \#'_N) + (\partial_d \#'_N)\} / \#'_N \quad (3.11)$$

and the standard deviation for the count of type- $d$  carbon atoms is

$$\sigma(N_d)_N \equiv \{ \langle (N_d)^2 \rangle_N - (\langle N_d \rangle_N)^2 \}^{1/2} \quad (3.12)$$

In principle the same techniques should be applicable to computing cross-correlations, or higher moments, but this is not presently pursued. (And for the mean squares there seem to be no “sum rules” as in (3.9), without the cross-correlations, which if desired could presumably be obtained by similar methods.)

Numerical results for structural isomer counts and associated averages and standard deviations for the numbers of different types of atoms are given in Table 1 for alkanes of

up to  $N = 40$  atoms. The structural isomer counts are already known, e.g., in ref 14, with which we agree, and the average atom-type counts satisfy the relations (3.9) as they should. It seems that each average atom-type count goes up more-or-less linearly with the number of carbons, while the corresponding standard deviations go up more slowly—and all this is addressed more quantitatively in sections 5 and 6.

Further these mean values for these graph invariants may be utilized to predict mean property values for a number of properties, representative ones of which are heat of formation, index of refraction, and magnetic susceptibility. That is, such properties for a particular molecule may be estimated in terms of an expansion in terms of local substructure counts, say of different types of atoms and of different types of bonds. For a structure represented by a graph  $G$ , such a property  $X$  is represented in terms of an additive *cluster expansion* as

$$X(G) = \sum_{\xi}^G \nu(\xi; G) X(\nu; \xi) \quad (3.13)$$

where the  $\xi$  label different types of substructures,  $\nu(\xi; G)$  is (at least in the current article just) the number of such substructures in  $G$ , and  $X(\nu; \xi)$  is a parameter dependent on the property  $X$  but not on the structure  $G$ . For example, if  $\xi$  labels a type of atom or bond, then  $X(\nu; \xi)$  is an atomic or bond contribution to the total property for the structure  $G$ . In principle the expansion if fully extended is<sup>25,26</sup> exact, but

**Table 1.** Some Characteristics of Alkane Structural Isomers

$N$	$\#_N$	$\langle \#_{\text{pri}} \rangle_N$	$\langle \#_{\text{sec}} \rangle_N$	$\langle \#_{\text{ter}} \rangle_N$	$\langle \#_{\text{qua}} \rangle_N$
1	1	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
2	1	2.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
3	1	2.00 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
4	2	2.50 ± 0.50	1.00 ± 1.00	0.50 ± 0.50	0.00 ± 0.00
5	3	3.00 ± 0.82	1.33 ± 1.25	0.33 ± 0.47	0.33 ± 0.47
6	5	3.20 ± 0.75	1.80 ± 1.33	0.80 ± 0.75	0.20 ± 0.40
7	9	3.55 ± 0.83	2.22 ± 1.40	0.88 ± 0.74	0.33 ± 0.47
8	18	4.00 ± 0.94	2.44 ± 1.54	1.11 ± 0.87	0.44 ± 0.60
9	35	4.34 ± 0.95	2.86 ± 1.57	1.26 ± 0.91	0.54 ± 0.60
10	75	4.72 ± 1.01	3.16 ± 1.67	1.52 ± 1.00	0.60 ± 0.65
11	159	5.10 ± 1.05	3.51 ± 1.73	1.68 ± 1.03	0.71 ± 0.69
12	355	5.48 ± 1.09	3.84 ± 1.79	1.90 ± 1.09	0.79 ± 0.72
13	802	5.85 ± 1.14	4.18 ± 1.86	2.09 ± 1.13	0.88 ± 0.75
14	1 858	6.23 ± 1.17	4.50 ± 1.92	2.29 ± 1.19	0.97 ± 0.79
15	4 347	6.61 ± 1.21	4.84 ± 1.98	2.49 ± 1.23	1.06 ± 0.82
16	10 359	6.99 ± 1.25	5.17 ± 2.04	2.69 ± 1.28	1.15 ± 0.85
17	24 894	7.37 ± 1.29	5.50 ± 2.09	2.89 ± 1.32	1.24 ± 0.88
18	60 523	7.75 ± 1.32	5.83 ± 2.15	3.09 ± 1.37	1.33 ± 0.91
19	148 284	8.13 ± 1.36	6.16 ± 2.21	3.28 ± 1.41	1.42 ± 0.93
20	366 319	8.51 ± 1.39	6.49 ± 2.26	3.48 ± 1.45	1.51 ± 0.96
21	910 726	8.89 ± 1.42	6.82 ± 2.32	3.68 ± 1.49	1.61 ± 0.99
22	2 278 658	9.27 ± 1.46	7.15 ± 2.37	3.88 ± 1.53	1.70 ± 1.01
23	5 731 580	9.65 ± 1.49	7.48 ± 2.42	4.08 ± 1.56	1.79 ± 1.04
24	14 490 245	10.03 ± 1.52	7.81 ± 2.47	4.28 ± 1.60	1.88 ± 1.06
25	36 797 588	10.42 ± 1.55	8.14 ± 2.52	4.47 ± 1.64	1.97 ± 1.09
26	93 839 412	10.80 ± 1.58	8.47 ± 2.57	4.67 ± 1.67	2.06 ± 1.11
27	240 215 803	11.18 ± 1.61	8.80 ± 2.62	4.87 ± 1.71	2.15 ± 1.14
28	617 105 614	11.56 ± 1.64	9.12 ± 2.66	5.07 ± 1.74	2.25 ± 1.16
29	1 590 507 121	11.94 ± 1.67	9.45 ± 2.71	5.27 ± 1.78	2.34 ± 1.18
30	4 111 846 763	12.32 ± 1.70	9.78 ± 2.76	5.46 ± 1.81	2.43 ± 1.20
31	10 660 307 791	12.71 ± 1.73	10.11 ± 2.80	5.66 ± 1.84	2.52 ± 1.22
32	27 711 253 769	13.09 ± 1.75	10.44 ± 2.84	5.86 ± 1.87	2.61 ± 1.24
33	72 214 088 660	13.47 ± 1.78	10.76 ± 2.89	6.06 ± 1.91	2.71 ± 1.27
34	188 626 236 139	13.85 ± 1.81	11.09 ± 2.93	6.26 ± 1.94	2.80 ± 1.29
35	493 782 952 902	14.24 ± 1.83	11.42 ± 2.97	6.45 ± 1.97	2.89 ± 1.31
36	1 295 297 588 128	14.62 ± 1.86	11.75 ± 3.01	6.65 ± 2.00	2.98 ± 1.33
37	3 404 490 780 161	15.00 ± 1.88	12.08 ± 3.05	6.85 ± 2.03	3.08 ± 1.35
38	8 964 747 474 595	15.38 ± 1.91	12.40 ± 3.10	7.05 ± 2.05	3.17 ± 1.36
39	23 647 478 933 969	15.76 ± 1.93	12.73 ± 3.13	7.24 ± 2.08	3.26 ± 1.38
40	62 481 801 147 341	16.15 ± 1.96	13.06 ± 3.17	7.44 ± 2.11	3.35 ± 1.40

it is virtually always used in a truncated (approximate) form where the substructures  $\xi$  are restricted to those of smaller size (e.g., perhaps just single sites and bonds). The substructures can be labeled to different extents (and there are different, yet complete, classes of substructures one may consider), so that different cluster expansions are possible.

The particular cluster expansions used next need to be specified. For molar refractivities a fairly accurate result is obtained<sup>24</sup> with subgraphs restricted to a single site labeled solely by the kind of atom (either C or H in our present circumstance). These contributions are

$$\text{Mn}_D(\nu;\text{C}) = +25.71 \text{ and } \text{Mn}_D(\nu;\text{H}) = -2.56 \quad (3.14)$$

For heats of formation  $\Delta H_f$  and magnetic susceptibilities  $\chi_{\text{mag}}$ , such an expansion limited to single sites does not work so well but is much improved<sup>22,26</sup> if the atoms are further labeled as to degree in the C-skeleton, i.e., as to whether they are primary, secondary, tertiary, or quaternary. The so-labeled atomic contributions for  $\Delta H_f$  are<sup>23</sup>

$$\begin{aligned} \Delta H_f(\nu;\text{pri}) &= -10.0, \quad \Delta H_f(\nu;\text{sec}) = -5.0, \\ \Delta H_f(\nu;\text{ter}) &= -2.4, \quad \Delta H_f(\nu;\text{qua}) = -0.1 \end{aligned} \quad (3.15)$$

with the units understood to be kcal/mol. For magnetic susceptibility  $\chi_{\text{mag}}$  the atomic contributions are<sup>26</sup>

$$\begin{aligned} \chi_{\text{mag}}(\nu;\text{pri}) &= 14.29, \quad \chi_{\text{mag}}(\nu;\text{sec}) = 11.48, \\ \chi_{\text{mag}}(\nu;\text{ter}) &= 9.19, \quad \chi_{\text{mag}}(\nu;\text{qua}) = 7.42 \end{aligned} \quad (3.16)$$

with the units understood to be m.u.  $\equiv -10^{-6}\text{erg/G}^2\cdot\text{mol}$ .

The computation of mean values and standard deviations is now straightforward in terms of the atom-type results and the expansion coefficients (of (3.14), (3.15), (3.16)). That is, one generally takes

$$\langle X(G) \rangle_N = \sum_{\xi}^G \langle \nu(\xi;G) \rangle_N X(\nu;\xi) \quad (3.17)$$

Consequent mean values and associated standard deviations for the three representative properties are shown in Figure 3. The standard deviations  $\sigma_N(X)$  are indicated by placing two points at  $+\sigma_N(X)$  and  $-\sigma_N(X)$  from our average estimate  $\langle X \rangle_N$ . The current estimate for index of refraction leads to standard deviations of 0 so that in this plot only a single point occurs at each  $N$  considered. It is seen that the heat of formation and magnetic susceptibility have at smaller  $N$  (say near  $N = 8$ ) a range within one isomer class such as to overlap a few neighboring isomer class averages on either side, and the number of such overlapping isomer classes goes up as  $N$  increases. That is, for these properties there are structural variations which become ever larger as  $N$  increases, exceeding by ever larger margins the variations between neighboring isomer classes. However, the range of overlap as measured in terms of the percentage range of an  $N$ -value seems to go down as  $N$  increases—that is, the fraction of isomer classes up to a given  $N$  overlapped by the  $N$ -atom

isomers seems to go down. More quantitative consideration of the large- $N$  behavior follows from the considerations of sections 5 and 6, whence the standard deviations for  $\Delta H_f$  and  $\chi_{\text{mag}}$  should increase  $\sim N^{1/2}$ .

#### 4. MEAN VALUE FOR THE WIENER NUMBER AND RELATED STRUCTURAL CHARACTERISTICS

The atom-type graph invariants of the preceding section are quite “classical”, and in some sense locally additive, but a graph invariant usually considered<sup>20</sup> more global is the so-called<sup>21</sup> “Wiener number”, and for this we now seek to compute isomer averages as well as some related structural/conformational averages. The usual definition<sup>21</sup> of the Wiener number depends on the *shortest-path distance* function on the vertexes of a connected graph—such a distance  $d(i,j)$  between two vertexes  $i$  and  $j$  of a graph  $G$  being the number of steps in a shortest path of  $G$  between  $i$  and  $j$ . Then the *Wiener number* for  $G$  is

$$W(G) \equiv \sum_{i < j} d(i,j) \quad (4.1)$$

This evidently seems quite “global”. Of course, for  $G$  a tree (as we currently deal with), the shortest path is unique, and it turns out that  $W(G)$  may, following Wiener,<sup>20</sup> be fairly readily re-expressed in terms of a sum over edges

$$W(G) \equiv \sum_{i \sim j} \nu(i,j) \nu(j,i) \quad (4.2)$$

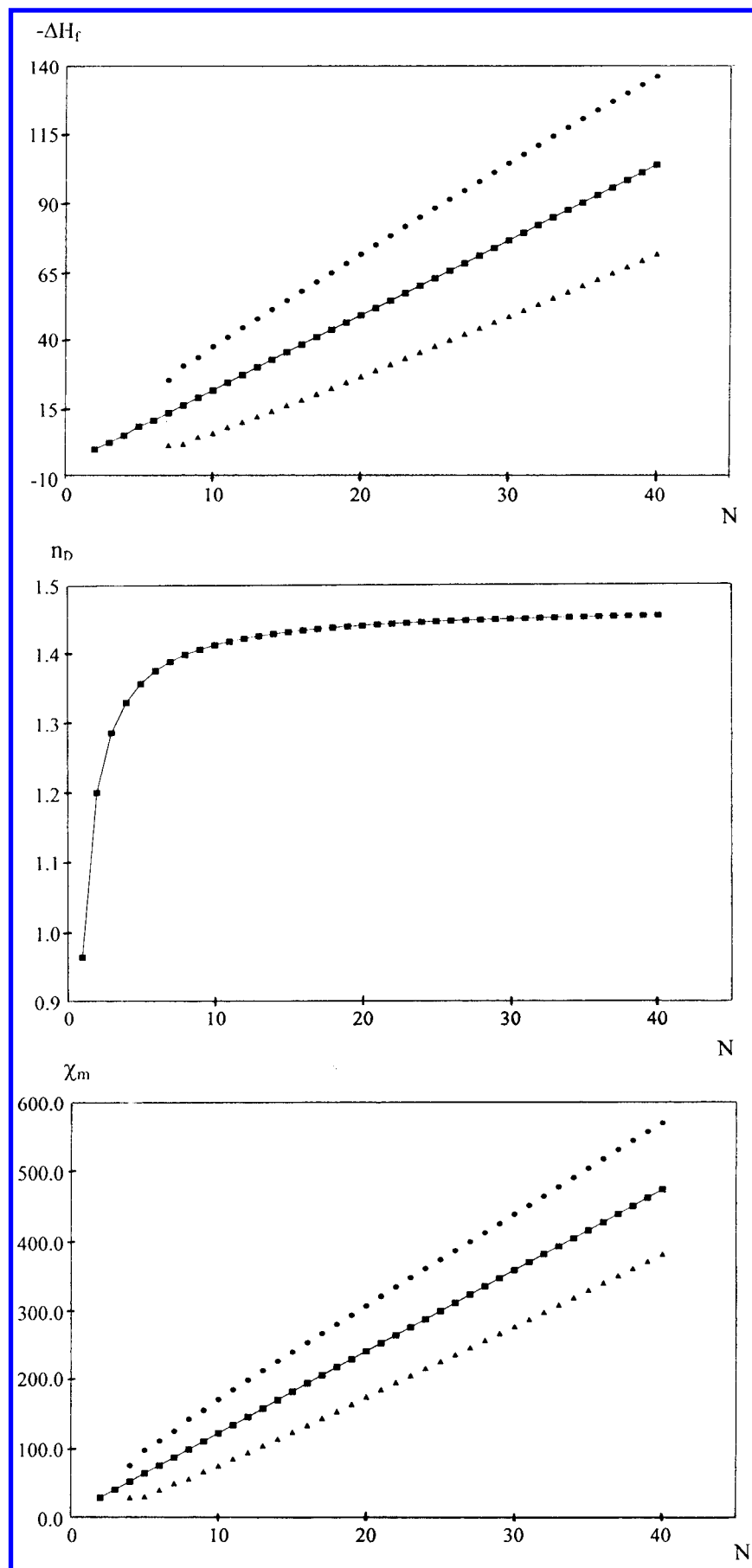
where  $\nu(i,j)$  is the number of sites of  $G$  closer to  $i$  than to  $j$ . And as it turns out after some preliminary development this is the formula used here.

The generating functions used parallel those of section 2, though now with an additional variable. We define

$$\phi_g(s,t) \equiv \sum_{m,n \geq 0} \#_{g,m,n} s^m t^n \quad (4.3)$$

where  $\#_{g,m,n}$  is the number of  $g$ -generation  $n$ -site rooted trees with a value of  $m$  for a particular characteristic—for a rooted tree  $T$  this characteristic value  $m$  is the sum of  $m(T,e)$  over all bonds  $e \in T$  including the root bond, with  $m(T,e)$  the numbers of sites on the side of  $e$  away from the root (including the end-site of  $e$  away from the root). The recursion for  $\phi_{g+1}(s,t)$  in terms of the  $\phi_g(\star, \star)$  with various arguments may be developed much as for  $\phi_{g+1}(t)$  in section 2, though special attention needs to be paid to the inclusion in  $m$  for the new  $g+1$ -generation trees the contribution for the root bond to the associated  $m$ . But this additional contribution  $\Delta m$  to  $m$  is just  $n$  (the number of sites in the radical structure) and can be counted along with  $n$  by replacing  $t$  in the recursion by  $st$ , so that in parallel to (2.2) we have

$$\begin{aligned} \phi_{g+1}(s,t) &= st \{ \phi_g(s,st) Z_2[\phi_{<g};s,st] + Z_2[\phi_g;s,st] \phi_{<g}(st) + \\ &\quad Z_3[\phi_g;s,st] \} \end{aligned} \quad (4.4)$$



**Figure 3.** Plots of average property values and standard deviations for alkanes of up to  $N = 40$  carbon atoms in (a) heats of formation (in kcal/mol); in (b) refractive indices; and in (c) magnetic susceptibilities (in  $m.u. \equiv -10^{-6} \text{erg/G}^2 \text{mol}$ ).

where

$$\phi_{<g+1}(s,t) \equiv \phi_{<g}(s,t) + \phi_g(s,t)$$

$$Z_2[f;s,t] \equiv \{[f(s,t)]^2 + f(s^2,t^2)\}/2$$

$$Z_3[f;s,t] \equiv \{[f(s,t)]^3 + 3f(s,t)f(s^2,t^2) + 2f(s^3,t^3)\}/6$$

$$Z_4[f;s,t] \equiv \{[f(s,t)]^4 + 6[f(s,t)]^2 f(s^2,t^2) + 8f(s,t)f(s^3,t^3) + 3[f(s^2,t^2)]^2 + 6f(s^4,t^4)\}/24 \quad (4.5)$$

with  $\phi_{<1}(s,t) \equiv 1$  and  $f$  is some two-variable function. Then still in analogy to the development in section 2, the new two-variable generating functions  $P_D(s,t)$  for unrooted alkanes of diameter  $D$  are written as

$$P_{2g-1}(s,t) = Z_2[\phi_g;s,t]$$

$$P_{2g}(s,t) = t\{Z_2[\phi_g;s,t]Z_2[\phi_{<g};s,t] + Z_3[\phi_g;s,t]\phi_{<g}(t) + Z_4[\phi_g;s,t]\} \quad (4.6)$$

though here it is important to note that in  $P_{2g-1}(s,t)$  the central bond is treated somewhat differently than the others. For the central bond  $e$  of an alkane tree  $T$  made of two  $g$ -generation rooted trees  $T_1$  and  $T_2$ , two contributions  $m(T_1,e)$  and  $m(T_2,e)$  are included in the associated  $m$  value (and this eventually entails some minor “correction” terms).

Now granted the basic generating functions of the preceding paragraph one may follow the ideas of section 3 to obtain recursions for their derivatives, with respect to  $s$  and  $t$ . Let  $\partial_s$  and  $\partial_t$ , respectively, indicate derivatives with respect to  $s$  and  $t$ . Then one sees that

$$st\partial_s\partial_t P_{2g}(s,t) = \sum_T^{2g} m(T) \cdot |T| \cdot s^{m(T)} t^{|T|} \quad (4.7)$$

where the sum is over all (unrooted) alkane trees of diameter  $2g$  and  $m(T) \equiv \sum_e m(T,e)$  with  $m(T,e)$  the number of sites on the side of the edge  $e$  away from the center of  $T$ . Of course if one lets  $w(T,e)$  be the number of sites on the side of  $e$  toward the center of  $T$ , then  $|T| = m(T) + w(T)$ , and

$$st\partial_s\partial_t P_{2g}(s,t) = \sum_T^{2g} m(T)w(T)s^{m(T)}t^{|T|} + \sum_T^{2g} \{m(T)\}^2 s^{m(T)} t^{|T|} \quad (4.8)$$

But this second summation is pretty close to the second derivative  $\partial_s^2 P_{2g}(s,t)$ , while the first sum is seen from (4.2) to yield a sum over Wiener numbers. Thus the coefficients  $S_{2g,N}$  of  $t^N$  in

$$W_{2g}(t) \equiv [t\partial_s\partial_t P_{2g}(s,t) - \partial_s(s\partial_s P_{2g}(s,t))]_{s=1} \quad (4.9)$$

are the sums over all Wiener numbers for all  $N$ -site diameter- $2g$  alkane trees. The odd-diameter alkane trees can be rather similarly handled, though special attention needs to be paid to the different manner in which the central bond in these species is treated in  $P_{2g-1}(s,t)$ . That is, this central bond is in essence counted twice since it has two sides away from the center, so that if we use an otherwise parallel development it remains to subtract off one of these duplicate contributions for this bond. Then noting that for this central

bond the  $m(T,e)$  count may be obtained from derivatives with respect to the single-variable generating functions  $\phi_g(t)$  for the radical groups being glued together, we find

$$W_{2g-1}(t) \equiv [t\partial_s\partial_t P_{2g-1}(s,t) - \partial_s^2 P_{2g-1}(s,t) - \partial_s P_{2g-1}(s,t)]_{s=1} - \{[t\partial_t\phi_g(t)]^2 + (t^2\partial/\partial t)^2\phi_g(t^2)\}/2 \quad (4.10)$$

with the coefficients  $S_{2g-1,N}$  of  $t^N$  in this polynomial being the sums over all Wiener numbers for all  $N$ -site alkane trees of diameter  $2g-1$ . The sums over all Wiener numbers for all  $N$ -site alkane structural isomers (independent of diameter) then are obtained as the coefficients  $S_N$  of  $t^N$  in the series

$$W(t) \equiv \sum_{D \geq 0} W_D(t) \quad (4.11)$$

Again as in the preceding section all the recursions may be done with the only value of the auxiliary variable  $s$  being at  $s = 1$ .

Thence we arrive at a mean value for the Wiener number for  $N$ -site alkane structural isomers, as

$$\langle W(G) \rangle_N = S_N / \#'_N \quad (4.13)$$

Further if one desires the mean shortest-path distances  $d(i,j)$  between all pairs  $\{i,j\}$  of vertexes as averaged over all  $N$ -site alkane isomers, this is immediately obtained as

$$\langle d(i,j) \rangle_N = \langle W(G) \rangle_N / 2N(N-1) \quad (4.14)$$

(since the number of pairs of such vertexes is just  $N(N-1)/2$ ).

The mean graph diameter  $\langle D \rangle_N$  for  $N$ -atom alkanes follows quite readily from the formulation set up in section 2. The number  $\#'_{D,N}$  of diameter- $D$  alkanes of  $N$  carbon-atoms is rather explicitly identified in the construction of the alkane-counting generating functions  $P_D(t)$ , whence one obtains the desired mean diameter as

$$\langle D \rangle_N = \{ \sum_g D \cdot \#'_{g,N} \} / \#'_N \quad (4.15)$$

Indeed one can even quite readily obtain standard deviations for the diameters.

Another property which we seek is the mean number of conformations per isomer. By conformation we refer to embeddings on a diamond lattice, without regard to multiple mappings of molecular graph sites onto the same lattice site, except that no two adjacent graph bonds are mapped onto the same lattice bond. That is, we do not assume “volume exclusion”, as is frequently done in treating linear-chain conformations, e.g., under the title of the “rotational isomeric model”, and is also done<sup>36,37</sup> for branched-chain isomers (though in this field of “polymer statistics” emphasis is often just on asymptotic forms). Also this neglect of volume exclusion is in the spirit of conventional Pólya counting theory, in that some alkane isomers counted sometimes have no volume-excluding conformations, as discussed elsewhere.<sup>12</sup> There have previously been some counts<sup>19,37</sup> for finite  $N$ , and it is in fact somewhat more straightforward than the usual Pólya-enumeration of alkane isomers because the (sometimes) redundant counting of what are in fact just different conformers of the same isomer is no longer to be



**Table 2.** Structural Isomer Properties of Alkanes

$N$	$\langle D \rangle_N$	$\langle W(T) \rangle_N$	$\langle d_{ij} \rangle_N$	$\langle \#_{\text{conf}} \rangle_N$
1	0.	0.	0.	1.0
2	1.0	1.0	1.0	1.0
3	2.0	4.0	1.333	6.0
4	2.5	9.5	1.583	6.5
5	3.0	18.0	1.800	24.3
6	3.8	31.0	2.067	62.4
7	4.333	48.0	2.286	160.7
8	4.833	69.56	2.484	396.1
9	5.343	96.69	2.686	1064.0
10	5.840	129.52	2.878	2618.7
11	6.308	168.25	3.059	6628.2
12	6.772	213.62	3.236	16 272.6
13	7.207	265.66	3.405	40 208.3
14	7.637	324.97	3.571	98 231.3
15	8.051	391.73	3.730	240 313.4
16	8.453	466.33	3.886	582 867.7
17	8.844	549.00	4.036	1 414 424.7
18	9.226	640.12	4.183	3 420 372.9
19	9.597	739.90	4.326	8 268 153.6
20	9.959	848.67	4.466	19 952 989.4
21	10.313	966.67	4.603	48 128 819.0
22	10.660	1094.18	4.736	115 972 406.0
23	10.998	1231.44	4.867	279 326 831.5
24	11.330	1378.73	4.995	672 364 624.3
25	11.656	1536.27	5.120	1 617 874 809.6
26	11.975	1704.32	5.244	3 891 513 588.9
27	12.289	1883.12	5.365	9 357 859 612.5
28	12.597	2072.89	5.483	22 496 671 901.0
29	12.899	2273.86	5.600	54 071 510 686.1
30	13.197	2486.27	5.715	129 936 591 379.8
31	13.491	2710.33	5.828	312 191 270 087.8
32	13.779	2946.26	5.940	749 966 801 707.7
33	14.064	3194.27	6.049	1 801 374 821 705.2
34	14.344	3454.58	6.157	4 326 254 387 940.6
35	14.621	3727.39	6.264	10 388 958 815 703.8
36	14.894	4012.91	6.369	24 945 274 334 025.8
37	15.163	4311.34	6.473	59 891 479 218 985.5
38	15.428	4622.87	6.575	143 782 413 460 353.7
39	15.691	4947.72	6.677	345 154 479 124 381.7
40	15.950	5286.06	6.776	828 497 234 580 482.0

eliminated. That is, the generating function for the enumeration of alkyl-radical conformers of generation number  $g$  being denoted  $\varphi_g(t)$  leads to a recursion

$$\varphi_{g+1} = 1 + t[\varphi_g(t) + \varphi_{<g}(t)]^3 \quad (4.16)$$

where

$$\varphi_{<g}(t) \equiv \sum_{f < g} \varphi_f(t) \quad (4.17)$$

with  $\varphi_0(t) \equiv 1$  and  $\varphi_{<0}(t) \equiv 0$ . In parallel to the development of section 2 one obtains generating functions  $P_D(t)$  for diameter- $D$  alkane conformations as

$$P_{2g-1}(t) = [\varphi_g(t)]^2$$

$$P_{2g}(t) = t\{[\varphi_g(t)]^4 + 4[\varphi_g(t)]^3\varphi_{<g}(t) + 6[\varphi_g(t)]^2[\varphi_{<g}(t)]^2\} \quad (4.18)$$

and the generating function for all alkane conformations as a sum over all  $D$  of these  $P_D(t)$ .

Numerical results for the averages of the shortest-path distances, the graph diameter, and the numbers of conformations per isomer are given in Table 2 for alkanes of up to  $N = 40$  carbon atoms. The through-bond-distance averages

$\langle\langle W(G) \rangle\rangle_N$  and  $\langle d(i,j) \rangle_N$  seem to grow somewhat more slowly than  $N$ , while the number of conformations per isomer seems to grow quite much more rapidly (perhaps exponentially fast), all as is discussed in more detail in the following two sections.

## 5. ASYMPTOTIC FORMS

The asymptotic behavior for alkane enumerations approaching the many-atom limit were in fact rather well discussed by Pólya,<sup>5,6</sup> and then further by Otter<sup>9</sup> and others.<sup>10,11</sup> But the more recent chemically oriented works seem meager in this regard, and asymptotic results for properties seems to be focused<sup>25,34,35</sup> on graph “sizes” when averaged over all conformations. Thence a more general consideration of asymptotic behaviors for isomer-averaged values of a variety of graph invariants or properties seems a reasonable goal, and the exact small- $N$  results obtained in the preceding sections can be of aid in making extrapolations to characterize these behaviors.

As a first step the earlier exact results<sup>5,6,9</sup> may be rederived in a simpler manner. For the structural isomers of the alkanes, an overall alkyl radical generating function

$$\phi(t) \equiv \sum_{g \geq 0} \phi_g(t) \equiv \sum_{N \geq 0} \#_N t^N \quad (5.1)$$

is of use (where  $\#_N$  is the number of alkyl radicals with  $N$  carbon atoms). This generating function exhibits a recursion

$$\phi(t) \equiv 1 + \frac{t}{6} \{[\phi(t)]^3 + 3\phi(t)\phi(t^2) + 2\phi(t^3)\} \quad (5.2)$$

which follows from that in section 2 for  $\phi_g(t)$ , such that the picturesque presentation there in essence applies here also. In fact the present recursion is that in the usual approach<sup>5,6,9</sup> to alkane enumeration. This recursion suggests the definition of a function

$$F(x) \equiv ax^3 + bx + c \quad (5.3)$$

whence the recursion of (5.2) can be rewritten as

$$\phi(t) = F(\phi(t)) \quad (5.4)$$

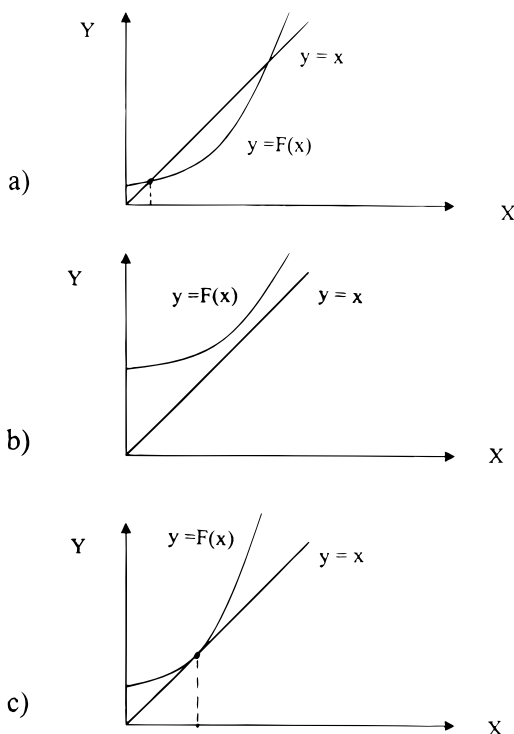
if it is somehow arranged that

$$a = t/6, \quad b = t\phi(t^2)/2, \quad c = 1 + t\phi(t^3)/3 \quad (5.5)$$

Now since  $\phi(t)$  is a power series in  $t$  with nonnegative coefficients  $\#_N$ , it must be a monotonic strictly increasing function of  $t$ . Since further the coefficients  $\#_N$  increase in such a way as to be bounded (for sufficiently great  $N$ ) by exponential forms

$$\kappa_{<}^N < \#_N < \kappa_{>}^N \quad (5.6)$$

one sees that  $\phi(t)$  is represented by a convergent series in  $t$  for sufficiently small  $t$  (say  $t < \kappa_{<}^{-1} > 0$ ), while the series diverges for sufficiently large  $t$  (say  $t > \kappa_{>}^{-1}$ ). Thence there must be some critical value  $t_c > 0$  (and  $t_c < 1$ ) of  $t$  such that for  $t < t_c$  there is convergence, while for  $t > t_c$  there is divergence. But now if one can understand some of this



**Figure 4.** Example plots of  $y = x$  and  $y = F(x)$ , in (a) for a circumstance where the two curves intersect, in (b) for a circumstance where the two curves miss, and (c) for the case where the two curves just touch. That is, in (a)  $t < t_c$ , in (b)  $t > t_c$ , and in (c)  $t = t_c$ .

critical behavior (as will govern the asymptotic form of the  $\#_n$ ) by considering the above-mentioned function  $F$  and the recursion (5.4)—basically one realizes that with  $t_c < 1$ , it follows that  $t_c^3 < t_c^2 < t_c$  so that with  $t$  near  $t_c$  one has  $\phi(t^2)$  and  $\phi(t^3)$  well behaved with changes in them proportional to changes of  $t$  away from  $t_c$ —thence it is feasible to incorporate the  $t$ -variation due to  $\phi(t^2)$  and  $\phi(t^3)$  into  $a$  and  $b$  as suggested in (5.5). Now from a consideration of (5.4) one sees that a simultaneous plot of  $y = F(x)$  and  $y = x$  as in Figure 4a should give a value for  $\phi(t)$  with  $t < t_c$  as the value of  $x$  at the (first) intersection between these two curves, so long as we have managed to choose  $a$  and  $b$  as in (5.5). Further one then sees that  $a$  and  $b$  increase monotonically with  $t$ , so that as  $t$  increases the  $y = F(x)$  curve will move upward till eventually it would no longer intersect the  $y = x$  curve, as in Figure 4b, whence there is no solution to the equation  $\phi = F(\phi)$  and  $t$  must exceed  $t_c$ . Thus  $t_c$  occurs where the  $y = F(x)$  and  $y = x$  curves just touch, as in Figure 4c, whence we have

$$1 = F'(\phi_c) \text{ or } 1 = 3a_c\phi_c^2 + b_c \quad (5.7)$$

where  $\phi_c \equiv \phi(t_c)$  and  $a_c$ ,  $b_c$ , and  $c_c$  are as in (5.5) with  $t = t_c$ . And as  $t$  approaches  $t_c$  (from below) the  $y = F(x)$  curve moves up proportionately to the variation in  $t$ , while the intersection moves ever more rapidly up the  $y = x$  curve. To quantify what is going on introduce a deviation  $\delta$  of  $t$  away from the critical value  $t_c$  and let

$$t = t_c - \delta, \quad a = a_c - a'\delta, \quad b = b_c - b'\delta + O(\delta^2), \\ c = c_c - c'\delta + O(\delta^2) \quad (5.8)$$

where

$$a' = 1/6, \quad b' = \{\phi(t_c^2) + 2t_c^2\phi'(t_c^2)\}/2, \\ c' = \{\phi(t_c^3) + 3t_c^3\phi'(t_c^3)\}/3 \quad (5.9)$$

Further introduce a deviation  $\epsilon \equiv \phi_c - \phi$  of  $\phi_c$  away from  $\phi$ . Then from (5.3) and (5.4)

$$\phi_c - \epsilon = (a_c - a'\delta)(\phi_c - \epsilon)^3 + (b_c - b'\delta)(\phi_c - \epsilon) + \\ (c_c - c'\delta) + O(\delta^2) \quad (5.10)$$

Recalling that  $\phi_c = a_c\phi_c^3 + b_c\phi_c + c_c$  one sees that here the terms of order  $\delta^0\epsilon^0 = 1$  on the left and right cancel, while from (5.7) one sees that the terms of order  $\delta^0\epsilon^1 = \epsilon$  on the right-hand side cancel, whence (5.10) reduces to

$$0 = \{3a_c\phi_c^2\}\epsilon^2 - \{a'\phi_c^3 + b'\phi_c\}\delta + O(\delta^2 \text{ and } \delta\epsilon) \quad (5.11)$$

Thus to leading order

$$\epsilon \approx C\delta^{1/2} \quad (5.12)$$

where  $C \equiv \{(a'\phi_c^3 + b'\phi_c)/3a_c\phi_c^2\}^{1/2}$ . Indeed one may see that the next higher term is of order  $\sim(\delta^{1/2})^2 = \delta$ .

Now the behavior of  $\phi(t)$  near  $t = t_c$  leads to asymptotic behaviors for the enumerative coefficients  $\#_N$  (and other related quantities). Indeed if we imagine that asymptotically  $\#_N$  is of the form

$$\#_N = N^{-\alpha}\kappa^N\{A + BN^{-1} + O(N^{-2})\} \quad (5.13)$$

then use of the leading part here in the power series of (4.1) leads to

$$\phi_c - \epsilon = 1 + \sum_{N \geq 1} \{\#_N - (A + B/N)N^{-\alpha}\kappa^N\}t^N + \\ A \sum_{N \geq 1} N^{-\alpha}\kappa^N t^N + B \sum_{N \geq 1} N^{-1-\alpha}\kappa^N t^N \quad (5.14)$$

Here one sees that in order for there to be convergence it is necessary and sufficient that  $t \leq 1/\kappa$ , so that  $t_c = 1/\kappa$ . Now the first summation leads to a finite value (so long as  $\alpha > 0$ ), while for the second and third sums the Euler–Maclaurin formula may be used to reexpress these sums as integrals over  $N$ . These integrals are (recalling that  $t = t_c - \delta = t_c(1 - \kappa\delta)$ ) of the general form

$$\int_0^\infty N^\beta (1 - \kappa\delta)^N dN \approx (\kappa\delta)^{-\beta-1} \int_0^\infty x^\beta e^{-x} dx = \\ (\beta!\kappa^{-\beta-1}) \cdot \delta^{-\beta-1} \quad (5.15)$$

(where  $x \equiv \kappa\delta N$ ). Then one obtains (for  $t$  approaching  $t_c$ , from below)

$$\phi \approx (\text{constant}) + A(-\alpha)!\kappa^{\alpha-1}\delta^{\alpha-1} + B(-1-\alpha)!\kappa^\alpha\delta^\alpha \quad (5.16)$$

Evidently the terms identified as “constant” here is just  $\phi_c$ , and to gain agreement with (5.12) the exponent  $\alpha$  is 3/2. One notes that not only does the asymptotic form of

(5.13) imply that of (5.16) or (5.12) but also (5.16) implies (5.13). That there is a one-to-one correspondence between the two asymptotic behaviors may be understood from the Cauchy relation

$$\#_N = \frac{1}{2\pi i} \oint \phi^{(N)}(t_c) \frac{N!}{(t - t_c)^N} dt \quad (5.17)$$

where  $\phi^{(N)}(t_c)$  denotes the  $N$ th derivative of  $\phi(t)$  with respect to  $t$ , as evaluated at  $t = t_c$ . Thus (5.13) gives the asymptotic form for the alkyl-radical counts, with  $\kappa = 1/t_c$  and  $\alpha = 3/2$ .

Granted the asymptotic forms of behavior for alkyl radicals, asymptotic results for (nonradical) alkanes readily follows. A power series  $\Phi(t) \equiv \sum_{N \geq 0} \#_N t^N$  for  $N$ -atom alkane-isomer counts  $\#_N$  may be defined and expressed in terms of  $\phi(t)$  following the ideas in section 2. The behavior of  $\Phi(t)$  follows from that of  $\phi(t)$ , and one finds

$$\#_N = N^{-1-\alpha} \kappa^N \{A' + B'N^{-1} + O(N^{-2})\} \quad (5.18)$$

with  $\alpha$  and  $\kappa$  the same as for the alkyl-radical case (but  $A'$  and  $B'$  being new constants). But actually there is a fairly straightforward way to understand this result. An  $N$ -atom alkane has  $2N + 2$  hydrogens which if removed yield a radical, though sometimes the removal of either of two different hydrogen may lead to the same radical. Indeed there are no more than  $N$  possible radicals, corresponding to the  $N$  different carbon atoms at which the dangling radical bond might conceivably be attached, and further this might be done only at nonquaternary carbons. Yet even then some carbons might be equivalent, so that attaching a radical bond at either of two such would give the same radical. Thus we anticipate that

$$f_N \#_N \approx \#_N' N s_N \quad (5.19)$$

where  $f_N$  is an average fraction of nonquaternary carbons and  $s_N$  is an average number of "equivalent" carbon sites per isomer for  $N$ -atom alkanes. But then granted (5.19) one sees that (5.13) implies (5.18).

Moreover, the argument leading to (5.18) by way of (5.19) extends to the understanding of asymptotic forms for other properties averaged over isomer classes. That is, the mean numbers of atom types ( $\Xi$  = primary, secondary, tertiary, or quaternary) are anticipated to be nonzero fractions of the total number of carbons. Thus we expect

$$\langle \#_{\Xi} \rangle_N \approx A_{\Xi} N + B_{\Xi} \quad (5.20)$$

Indeed a similar expression applies for bond types. And beyond the mean values one can look at standard deviations  $\sigma_N(\#_{\Xi})$  (as in (3.12)). For these we surmise that the asymptotic form is

$$\sigma_N(\#_{\Xi}) \approx A_{\sigma\Xi} N^{1/2} + B_{\sigma\Xi} \quad (5.21)$$

much as applies for the anticipated deviations of numbers of heads for a large number  $N$  of tossings of a random coin from the expected mean (of  $N/2$ ).

The asymptotic form for the mean (through-bond) distances between carbons or mean (through-bond) diameters is another matter. But related means have already been repeatedly investigated<sup>33,34</sup> when the averages are made over all conformations of all isomers. The relevant recursions for the associated generating functions are similar to what we have already been dealing with; e.g., in place of (5.2) the generating function for all alkyl radical conformations embeddable on the diamond lattice turns out to be of the form

$$\varphi(t) = 1 + t[\varphi(t)]^3 \quad (5.22)$$

And the modifications to obtain the mean distance covered on the diamond lattice have been developed.<sup>36,37</sup> (These treatments are without "volume exclusion" in the sense that the embeddings considered allow more than one site of the molecular graph to be mapped to the same lattice site—but in fact Pólya theory suffers from the same problem, as emphasized elsewhere.<sup>12</sup>) In these approaches the mean distance spanned by  $N$ -atom polymer graphs on a general regular lattice scales  $\sim N^{1/4}$ , evidently because<sup>37</sup> the "backbone length" (or diameter)  $\sim N^{1/2}$  performs a random walk, and the standard result that a random walk of  $L$  steps spans a distance  $\sim L^{1/2}$ , then leads for our  $N$ -atom branched polymers to a span  $\sim (N^{1/2})^{1/2} = N^{1/4}$ . Because of the similarity of the recursions (as (5.2) vs (5.21)) and because of the conceptual similarity, we anticipate that for the current case of interest the mean (through-bond) diameter  $D$  and the mean (through-bond) distances  $d(i,j)$  (between sites  $i$  and  $j$ ) both scale  $\sim N^{1/2}$ . That is, we anticipate asymptote forms

$$\begin{aligned} \langle D \rangle_N &\approx A_D N^{1/2} + B_D \\ \langle d(i,j) \rangle_N &\approx A_d N^{1/2} + B_d \end{aligned} \quad (5.23)$$

with  $A_D$ ,  $B_D$ ,  $A_d$ , and  $B_d$  different constants.

## 6. ASYMPTOTIC NUMERICS

Now the asymptotic forms obtained or surmised may be utilized along with the numerical data of Tables 1 and 2 to estimate the various constants appearing in the various asymptotic forms (5.18), (5.20), (5.21), and (5.23). First from the formal considerations connected with (5.2) there follows a powerful method for determining the growth factor  $\kappa$  very accurately, say to within a given tolerance  $\epsilon$ . The basic idea involves the use (5.2) as a recursion relation, which is easier to understand for a simplified circumstance when there are no quaternary carbons. For this case of aquaternary alkyl radicals this recursion relation simplifies to

$$\phi(t) \equiv 1 + \frac{t}{2} \{[\phi(t)]^2 + \phi(t^2)\} \quad (6.1)$$

Then one imagines, starting from high powers  $p > 2^M$  of  $t$ , where it turns out that  $\phi(t^p)$  is accurately known, that one recurs downward to give at the  $m$ th step  $\phi(t^{2^{M-m}})$  until finally at the  $M$ th step one obtains  $\phi(t^{2^0}) = \phi(t)$ . To start the recursion at high powers note that since  $\phi(t) = 1 + O(t)$ , it

follows that  $\phi(t^p) = 1 + O(\epsilon) \approx 1$  so long as  $t^p \ll \epsilon$ . That is, one takes  $\phi(t^p) = 1$  for  $p > 2^M$  with  $M > \{(\log \epsilon)/(\log t)\}/(\log 2)$ . The recursion at each step involves a solution of a polynomial (indeed quadratic) equation for  $\phi(t^{2^{M-m}})$  in terms of  $t$  and the higher order  $\phi(t^{2^{1+M-m}})$  already obtained, as follows from (6.1). Moreover, the recursion leads to a real polynomial root at the last step iff  $t \leq t_c$  (i.e., if we have a circumstance as in one of the plots as in Figure 4(a) or 4(c), though because of the way we initiated the recursion, an error  $\lesssim t^{2^M} \sim \epsilon$  arises. (Further if nonreal roots result at  $t^{2^{M-m}}$  before  $m$  reaches  $M$ , this means that for all larger  $m$  they will be nonreal too.) Thus up to a chosen tolerance  $\sim \epsilon$  one can rather readily decide whether a given  $t$  is  $> t_c$  or  $< t_c$ . And one may proceed cyclically as follows:

- (0) Choose initial bounds  $t_- = 0$  and  $t_+ = 1$ .
- (1) Guess  $t = (t_- + t_+)/2$  and decide how this compares to  $t_c$ ; if  $t < t_c$  replace  $t_-$  by  $t$ , whereas if  $t > t_c$  replace  $t_+$  by  $t$ .
- (2) If  $|t_- - t_+| < 2\epsilon$ , then take  $t_c = (t_- + t_+)/2$ , and otherwise return to (1).

Each cycle between (1) and (2) here cuts in half the interval in which  $t_c$  is localized, so that after 33 iterations one has  $t_c$  to a precision of  $\sim \epsilon \sim 2^{-33} \approx 10^{-10}$ , which is what we will be content with here.

The case for alkyl radicals without exclusion of quaternary carbons proceeds by much the same scheme, except that there are more different powers of  $t$  involved. Indeed since the recursion of (4.2) involves  $\phi(t)$ ,  $\phi(t^2)$ , and  $\phi(t^3)$ , one ends up considering a doubly indexed set of powers  $t_{m,n} \equiv t^{2^m 3^n}$  and of corresponding  $\phi(t_{m,n}) \equiv \varphi_{m,n}$ . The recursion of (5.2) now appears as

$$\varphi_{m,n} = 1 + t_{m,n} \{ \varphi_{m,n}^3 + 3\varphi_{m,n}\varphi_{m+1,n} + 2\varphi_{m,n+1} \} / 6 \quad (6.2)$$

and for sufficiently large  $m, n$  one takes  $\varphi_{m,n} = 1$ . Here "sufficiently large" means that the powers in  $2^m 3^n$  must be such that  $t_{m,n} \ll \epsilon$ , i.e., it means that

$$(\log 2) \cdot m + (\log 3) \cdot n > (\log \epsilon) / (\log t) \quad (6.3)$$

The recursion then may be viewed to start with a two-dimensional  $m, n$ -grid on which the values of  $\varphi_{m,n}$  are to be entered: all the values with  $m, n$  lying beyond the straight line identified by (6.3) are set to 1; then (6.2) is used as a 2-fold downward recursion to obtain real  $\varphi_{m,n}$  roots to a polynomial (now a cubic); the reality or nonreality of the roots for  $\varphi_{0,0}$  again identifies whether  $t$  is  $< t_c$  or  $> t_c$ . Finally the whole process is repeated as in the scheme (0), (1), (2) already indicated for the case of aquaternary alkanes. Implementing this scheme we obtain

$$\kappa = 1/t_c \cong 2.815\,460\,033\,29 \quad (6.4)$$

which exceeds in accuracy Otter's<sup>9</sup> earlier seven-digit number (and it should not be difficult to add a few more decimals of precision, if desired).

The case of conformer counts can be treated to a certain extent analytically. In analogy to the overall structural-

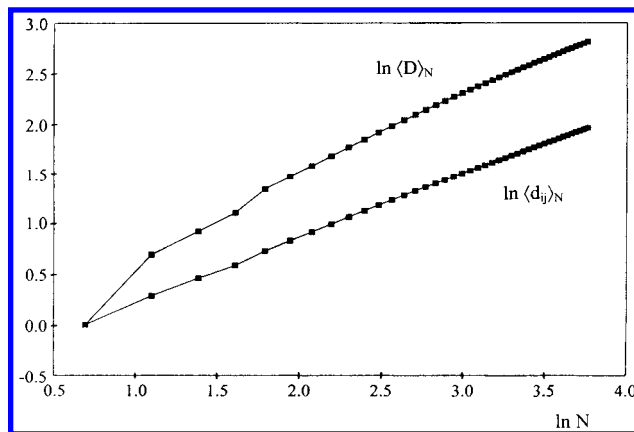


Figure 5. Plots of  $\log \langle d(i,j) \rangle_N$  and  $\log \langle D \rangle_N$  versus  $\log N$ , to test exponent values.

isomer generating function  $\phi(t)$  as in (5.1) we now have an overall conformer generating function  $\varphi(t)$ , which turns out to satisfy an especially simple recursion

$$\varphi(t) = 1 + t[\varphi(t)]^3 \quad (6.5)$$

This is analogous to the recursion of (5.2) but then leads to a result in analogy with (5.7) of  $1 = 3t_c\varphi(t_c)$ , whence one obtains  $t_c = 4/27$ , and

$$\kappa_{\text{conf}} = 27/4 \quad (6.6)$$

an exact growth factor for the case of alkane conformations.

For the properties one may test the surmises for exponents (as in (5.21) and (5.23)). Thus for the mean shortest-path distance between vertexes of an  $N$ -atom alkane a plot of

$$\log \{ \langle d(i,j) \rangle_N \} \text{ vs } \log N \quad (6.7)$$

should approach a straight line with a slope of  $1/2$ . A test of this and a corresponding plot involving  $\log \langle D \rangle_N$  is indicated in Figure 5, whence indeed our expectation of asymptotic straight lines seems plausible. Fitting the data up to  $N = 43$  yields slope estimates close to our expectations. In particular, if slopes from pairs of successive points on this plot are plotted vs  $N^{-1/2}$ , then the numerically extrapolated values for this slope as  $N \rightarrow \infty$  are 0.50 and 0.49 for  $\langle d(i,j) \rangle_N$  and  $\langle D \rangle_N$ , respectively. The exponents for the standard deviations of (5.21) were similarly tested as well as several of the others with a more rigorously founded reasons for the expected values, and all were found to fit close to expectations. Thus in our further considerations we assume that all the exponents are exactly as argued in the preceding section.

Finally numerical estimates for the various asymptotic  $A$ -,  $B$ -, and  $C$ -proportionality factors may be made. For the enumerations  $A_N$  one divides out  $N^{-5/2}\kappa^N$  to deal with the weaker remnant factor. In this case and for most of the properties then one seeks a fit of a sequence of values  $X_N$  for a range  $\mathbf{R}$  of  $N$  values to a form

$$N^\alpha (A + BN^{-\beta} + CN^{-2\beta}) \quad (6.8)$$

(with  $\alpha, \beta, A, B$ , and  $C$  dependent on  $X$ , but not  $N$  – and  $C$  being the coefficient higher term than the  $B$ -terms explicitly



**Table 3.** Asymptotic Numerics in Alkanes<sup>a</sup>

	$\Xi$	$\alpha$	$\beta$	A	B	C	err <sub>40</sub> , %
Structural Isomers							
1. $\#_N/\kappa^N$		-5/2	1	0.6570	-0.068	6.2	0.005
Conformers							
2. $\#'_N/\kappa^N$		-5/2	1	0.3664	-0.574	0.6	0.0000
Mean Properties of Structural Isomers							
3. $\langle D \rangle_N$		1/2	1/2	3.3339	-5.584	2.83	0.007
4. $\langle d_{ij} \rangle_N$		1/2	1/2	1.2771	-1.437	0.86	0.003
5. $\langle \#_p \rangle_N$		1	1	0.3831	0.800	0.96	0.002
6. $\langle \#_s \rangle_N$		1	1	0.3267	0.020	-1.20	-0.002
7. $\langle \#_i \rangle_N$		1	1	0.1973	-0.440	-0.49	-0.004
8. $\langle \#_a \rangle_N$		1	1	0.0929	-0.380	0.72	-0.003
9. $\langle \sigma_{\Delta H^\circ} \rangle_N$		1/2	1/2	4.9713	2.548	-10.3	0.088
10. $\langle \sigma_\chi \rangle_N$		1/2	1/2	14.221	7.462	-21.0	0.096

<sup>a</sup> Asymptotic form:  $\Xi = N^\alpha(A + BN^{-\beta} + CN^{-2\beta})$ .

displayed in section 5). And this we have done via a regression analysis, wherein we minimize a weighted least-squares error

$$\sum_{N \in \mathbf{R}} \{X_N - N^\alpha \{A + BN^{-\beta} + CN^{-2\beta}\}\}^2 \cdot N^{2(3\beta-\alpha)} \quad (6.9)$$

The weight  $N^{2(3\beta-\alpha)}$  reflects the manner in which the square-residual errors are anticipated to scale, with this choice then making the remnant so-weighted errors the “same” for each data point. The ranges  $\mathbf{R}$  were chosen differently for different properties:  $\mathbf{R} = \{11 \rightarrow 80\}$  for  $\#_N$ ;  $\mathbf{R} = \{1 \rightarrow 44\}$  for  $\#_N(\text{conf})$ ;  $\mathbf{R} = \{11 \rightarrow 65\}$  for atom types; and  $\mathbf{R} = \{1 \rightarrow 43\}$  for  $\langle d(i,j) \rangle_N$  and the various  $\sigma_N(X)$ . The consequent numerical results for  $A$ ,  $B$ , and  $C$  are indicated in Table 3 for each of several different characteristics  $X$ . Also indicated there are the exponents  $\alpha$  and  $\beta$  used as well as the relative % error

$$\text{err}_N \% \equiv 100 \cdot \{X_N - N^\alpha \{A + BN^{-\beta} + CN^{-2\beta}\}\} / X_N \quad (6.10)$$

at  $N = 40$ . Also as another such check for the largest count given as an integer we have

$$\#_{45} = 8\ 227\ 162\ 372\ 221\ 203$$

$$\kappa^{45} (45)^{-5/2} \{A + B(45)^{-1} + C(45)^{-2}\} \cong 8.22366 \times 10^{15} \quad (6.11)$$

so that the discrepancy for our approximate formula is here  $\approx 0.043\%$ . Evidently the asymptotic forms appear to be quite accurate, and in agreement with the surmised forms of the preceding section. The behavior of the properties of Figure 3 follows from the behavior of the atom-type asymptotic results of Table 3, with the average heat of formation and magnetic susceptibility increasing linearly with  $N$ , while their standard deviations over a class increases as  $N^{1/2}$ . Notably the asymptotic forms for the mean shortest-path distance and mean shortest-path diameter exhibit  $N$ -dependences similar to those for the “polymer statistical” results,<sup>35,36</sup> despite the statistical-mechanical averages being over conformations and our present ones being over (structural) isomers.

## 7. CONCLUSIONS

Classical enumeration schemes have been reconsidered, with straightforward mathematical extensions enabling chemically new types of considerations. In particular, several different graph-theoretic characteristics averaged over the different isomer classes have been computed and then used to consider some interesting chemical properties. Tabulations (often exact) for up to about  $N = 40$  are made. But further asymptotic behaviors for isomer counts and various of these properties are derived or surmised and then numerically tested to yield quite accurate asymptotic results. Some polymer statistical properties averaged over isomers have been found to exhibit asymptotic forms as for the same properties averaged over conformations (as is done in statistical mechanics<sup>27,30,36</sup>). Our isomer-averaged properties (particularly with standard deviations) reveal something about the degree to which an identification to an isomer class also identifies molecular properties. For heats of formation and magnetic susceptibilities the  $N$ -atom isomer class properties overlap a couple of classes on each side at modest  $N$  (of  $10 \rightarrow 20$ ) and the degree of overlap grows as  $N^{1/2}$  (which however becomes ever smaller in comparison to  $N$ , such as the properties scale).

An interesting and hopefully useful formal “combinatoric” chemistry seems possible, yielding not only enumerations but also some properties. The computation of standard deviations achieved for some properties here holds promise that in further extensions higher moments or distributions may be obtained, with the possibility of picking out molecules with extreme (e.g., optimal) properties, and further extensions should be considered. Such may be especially desirable with other combinatoric collections of molecules other than the alkanes—and indeed much of what has been done should extend to other collections. That much of the formal “luggage” of Pólya enumeration theory<sup>5-8,10,13-16</sup> has been avoided here is not to say is irrelevant but rather that even more powerful tools for further questions are available for application to more general classes of molecules. That is, properties for many other classes of molecules should be similarly treatable (even far beyond some other classes which have already been enumerated<sup>8,18</sup>). The point made in the appendix that Pólya enumeration is closely related to structure generation and thence to nomenclature can be imagined to be reversible, so that systematic nomenclature (of which we have a great deal) could lead to enumeration and thence (average) properties. Of course properties beyond those which are often described as “locally additive” could be especially interesting, in that optimization of a property that is “locally additive” really just involves including an extreme number of the local groups with an extreme local value for this property. Thus the ability to compute the mean “Wiener” number is perhaps of interest in that it is often viewed as not locally additive but rather a global characteristic. It may be that one can ultimately mathematically simulate some aspects of current experimental dendrimer chemistry<sup>28</sup> or of what is currently called<sup>29</sup> “combinatorial chemistry”.

## ACKNOWLEDGMENT

The authors acknowledge the support of the Welch Foundation of Houston, Texas.

**Table 4.** Various Conceivable Associated Alkane Nomenclatures

methane	methane	methane	methane
ethane	dimethyl	dimethyl	dimethyl
propane	dimethylmethane	dimethylmethane	dimethylmethane
butane	di(methylmethyl)	diethyl	di(ethyl)
2-methyl-propane	trimethylmethane	trimethylmethane	trimethylmethane
pentane	di(methylmethyl)methane	diethylmethane	di(ethyl)methane
2-methyl-butane	(dimethylmethyl)(methylmethyl)	(trimethylmethyl)ethyl	(isopropyl)(ethyl)
2,2-dimethyl-propane	tetramethylmethane	tetramethylmethane	tetramethylmethane
hexane	di(methylmethylmethyl)	dipropyl	dipropyl
2-methylpentane	(dimethylmethyl)(methylmethyl)methane	(dimethylmethyl)ethylmethane	isopropylethylmethane
3-methylpentane	di(methylmethyl)methylmethane	(diethyl)(methyl)methane	diethylmethylmethane
2,2-dimethylbutane	(trimethylmethyl)methylmethyl	(trimethylmethyl)ethyl	(tert-butyl)ethyl
2,3-dimethylbutane	di(dimethylmethyl)	di(dimethylmethyl)	diisopropyl

## APPENDIX

The recursive method utilized for enumeration evidently is based upon a generation scheme as suggested with Figure 1, and the generation scheme in turn links directly to a systematic conceivable nomenclature. Actually with such close attention to generation number and diameter, the suggested nomenclatural scheme has some relation to the standard IUPAC scheme which is based on (graphical) diameters. But the current generational scheme pays much attention to the graph center, with the groups attached thereto naturally separately named. Further as the current recursive relations all build (as indicated in Figure 1) from a single carbon atom, the generation and naming can be all based on single-carbon-atom groups, i.e., on methyl groups, possibly with various substitutions. In this "minimal-vocabulary" scheme the resulting correspondence to the standard name is indicated for the alkanes up through  $C_6H_{12}$  in the second column of Table 4, the standard names being given in column 1. With a modification to allow for (usual) individual names for straight-chain (normal) alkyl groups, the names become as indicated in the third column of the Table 4. With a further modification to allow for (usual) individual names for alkyl groups with up to one branch point (either secondary or tertiary) at the chain end, the names become as indicated in the fourth column of the Table 4. Interestingly without much effort, the molecules are quite recognizable from these names. It seems a general rule that powerful enumeration schemes may aid in generation and in nomenclature.

Conversely a systematic nomenclature should associate to an enumerative and computational procedure. The usual IUPAC nomenclature then would relate to a generating-function approach to enumeration when one builds upon a straight-chain backbone to be identified as the diameter. That is, may build our generating functions  $P_D(t)$  based on the IUPAC nomenclature, e.g., for all alkanes with a diameter  $D = 2g - 1$

$$P_{2g-1}(t) = t^{2g} \prod_{n=2}^g \{(Z_2[\phi_{<n};t]^2 - Z_2[\phi_{<n}^2;t^2])/2\}$$

$$P_{2g}(t) =$$

$$t^{2g+1} Z_2[\phi_{<g};t] \prod_{n=2}^{g-1} \{(Z_2[\phi_{<n};t]^2 - Z_2[\phi_{<n}^2;t^2])/2\}$$

That is, we surmise that generally systematization in nomenclature and in enumeration are equivalent.

## REFERENCES AND NOTES

- (1) Brown, C. On the Theory of isomeric Compounds. *Trans. R. Soc. Edinburgh* **1864**, 23, 707–719.
- (2) Cayley, A. On the Mathematical Theory of Isomers. *Philos. Mag.* **1874**, 47, 444. Cayley, A. On the Analytical Forms called Trees with Applications to the Theory of Chemical Compounds. *Rep. Brit. Ass. Adv. Sci.* **1887**, 257–305.
- (3) Schiff, H. Zur Statistik chemischer Verbindungen. *Ber. Deutsch. Chem. Ges.* **1876**, 8, 1542–1547. Flavitzky, F. *Ber. Deutsch. Chem. Ges.* **1876**, 9, 267. Delannoy, H. A. Sur les arbres geometriques et leur emploi dans la theorie des combinaisons chimiques. *Compt. Rend. Ass. Franc. Avance. Sci.* **1894**, 23, part 2, 102–116. Herrmann, F. Über das Problem die Anzahl der isomeren Paraffine der Formel  $C_nH_{2n+2}$  zu bestimmen. *Ber. Deutsch. Chem. Ges.* **1897**, 9, 792 and 2423–2426. Losanitsch, S. M. Die isomerie-Arten bei den Homologender Paraffin-Riehe. *Ber. Deutsch. Chem. Ges.* **1897**, 30, 1917–1926.
- (4) Henze, H. R.; and Blair, C. The number of structurally isomeric alcohols of the methanol series. *J. Am. Chem. Soc.* **1931**, 53, 3042–3046; The Number of Isomeric Hydrocarbons of the Methane Series. 3077–3085. Blair, C. M.; Henze, H. R. The Number of Stereoisomeric and Non-Stereoisomeric Monosubstitution Products of the Paraffins. *J. Am. Chem. Soc.* **1932**, 54, 1098–1106; The Number of Stereoisomeric and Non-Stereoisomeric Paraffin Hydrocarbons. 1538–1545. Perry, D. The Number of Structural Isomers of Certain Homologues of Methane and Methanol. *J. Am. Chem. Soc.* **1932**, 54, 2918–2920. Coffman, D. D.; Blair, C. M.; Henze, H. R. *J. Am. Chem. Soc.* **1933**, 55, 252–253. Henze, H.; Blair, C. M. The number of structurally isomeric hydrocarbons of the ethylene series. *J. Am. Chem. Soc.* **1933**, 55, 680–686. Henze, H.; Blair, C. M. The Number of Structural Isomers of the More Important Types of Aliphatic Compounds. *J. Am. Chem. Soc.* **1934**, 54, 593.
- (5) Pólya, G. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.* **68**, 1937, 145–254.
- (6) Pólya, G. Algebraische Berechnung der Isomeren einiger organischer Verbindungen. *Zeit. Kryst. A* **1936**, 93, 414–443. Pólya, G. Über das Anwachsen der Isomerenzahlen in den homologen Reihe der organischen Chemie. *Vierteljschr. Naturforsch. Ges. (Zürich)* **1936**, 81, 243–258.
- (7) Pólya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*; Springer-Verlag: Berlin, 1987.
- (8) Ford, G. W.; Uhlenbeck, G. E. Combinatorial Problems in the Theory of Graphs. I. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, 42, 122–128, ...II. 203–208. Balaban, A. T.; Harary, F. Chemical graphs. Part 4. Dihedral groups and monocyclic aromatic compounds. *Rev. Roumaine Chim.* **1967**, 12, 1511–1515. deBruijn, N. G. Enumeration of Tree-Shaped Molecules. In *Recent Progress in Combinatorics*; Tutte, W. T., Ed.; Academic Press: New York, 1969; pp 59–68. White, D. E. Classifying Patterns by Automorphism Group: An Operator Theoretic Approach. *Discuss. Math.* **1975**, 13, 277–295. Balaban, A. T. Enumeration of Cyclic Graphs. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 63–105. Kerber, A. On Graphs and their Enumeration, Part I. *Math. Chem. Commun.* **1975**, 1, 5–10; "Part II" **1976**, 2, 17–34. Robinson, R. W.; Harary, F.; Balaban, A. T. The Numbers of Chiral and Achiral Alkanes and Monosubstituted Alkanes. *Tetrahedron* **1976**, 32, 355–361. Kerber, A. Counting Isomers and Such. In *The Permutation Group in Physics and Chemistry*; Hinze, J., Ed.; Springer-Verlag: Berlin, 1979; pp 1–18. Balasubramanian, K. A Generalized Wreath Product Method for the Enumeration of Stereo and Position Isomers of Polysubstituted Organic compounds. *Theor. Chim. Acta* **1979**, 51, 37–54. Hanlon, P. A Cycle Index Sum Inversion Theorem. *J. Comb. Theory A* **1981**, 30, 248–269. Davidson, R. A. Isomers and Isomerization: Redfield's Combinatorial Theory. *J. Am. Chem. Soc.* **1981**, 103, 312–314.

- Balasubramanian, K. The Combinatorics of symmetry Adaptation. *Theor. Chim. Acta* **1981**, *51*, 47–54. McClarnen, T. J. The numbers of polytypes in sheet silicates. *Zeit. Kristallogr.* **1981**, *155*, 247–268. Quintas, L. V.; Yarmish, J. Valence isomers for chemical trees. *Math. Chem. Commun.* **1981**, *12*, 65–73. Merris, R. Generalized Matrix Functions and Pattern Inventory. *Lin. Multilin. Alg.* **1983**, *12*, 315–327. Hässelbarth, W. A note on Pólya enumeration theory. *Theor. Chim. Acta* **1984**, *66*, 91–110. Balasubramanian, K. Applications of Combinatorics and Graph Theory to Spectroscopy and Quantum Chemistry. *Chem. Rev.* **1985**, *85*, 599–618. Mead, C. A. Table of Marks and Double Cosets in Isomer Counting. *J. Am. Chem. Soc.* **1987**, *109*, 2130–2137. Fujita, S. The USCI approach and elementary superposition for combinatorial enumeration. *Theor. Chim. Acta* **1992**, *82*, 473–498. Balasubramanian, K. Enumeration of Chiral and Positional Isomers of Substituted Fullerene Cages ( $C_{20}$ – $C_{70}$ ). *J. Phys. Chem.* **1993**, *97*, 6990–6998. Fujita, S. Subduction of dominant representations for combinatorial enumeration. *Theor. Chim. Acta* **1995**, *91*, 315–332.
- (9) Otter, R. The Number of Trees. *Ann. Math.* **1948**, *49*, 583–599.
  - (10) Ford, G. W.; Uhlenbeck, G. E. Combinatorial Problems in the Theory of Graphs. III. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, *42*, 529–535. Harary, F.; Robinson, R. W.; Schwenk, A. J. Twenty-Step Algorithm for determining the asymptotic Number of Trees of various Species. *Austral. Math. Soc.* **1975**, *20A*, 483–503.
  - (11) Lederberg, J.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. V.; Duffield, A. M.; Djerassi, C. Applications of Artificial Intelligence for Chemical Inference. The Number of possible Organic Compounds. I. *J. Am. Chem. Soc.* **1969**, *91*, 2973–2976. Cioslowski, J. Series analysis methods in enumeration of chemical isomers. *Theor. Chim. Acta* **1989**, *76*, 47–51.
  - (12) Klein, D. J.; Seitz, W. A. Graphs, Polymer Models, Excluded Volume, and Chemical Reality. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 430–445.
  - (13) Fujita, S. *Symmetry and Combinatorial Enumeration in Chemistry*; Springer-Verlag: Berlin, 1992.
  - (14) Read, R. C. The Enumeration of Acyclic Chemical Compounds. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 25–61. Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; SKTH/Kemija u industriji, Zagreb, 1985.
  - (15) Gordon, M.; Kennedy, J. W. The Counting and Coding of Trees of Fixed Diameter. *SIAM J. Appl. Math.* **1975**, *28*, 376–398. Balaban, A. T.; Kennedy, J. W.; Quintas, L. The Number of Alkanes having  $n$  Carbons and a longest Chain of Length  $d$ . *J. Chem. Educ.* **1988**, *65*, 304–313.
  - (16) Bailey, C.; Palmer, E.; Kennedy, J. Points by Degree and Orbit Size in Chemical Trees. I. In *The Theory and Applications of Graphs*; John Wiley and Sons: New York, 1981; pp 27–43. Quintas, L. V.; Yarmish, J. Degree Distributions for Random 4-Trees and Skeletons of Symmetry Weighted (1,4)-Trees. *Congressus Num.* **1982**, *36*, 115–125.
  - (17) Yeh, C.-Y. Isomer enumeration of alkanes, labeled alkanes, and monosubstituted alkanes. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 912–913. Yeh, C.-Y. Theory of Acyclic Chemical Networks and Enumeration of Polyeneoids via Two-Dimensional Chirality. *J. Phys. Chem.* **1996**, *100*, 15800–15804. Yeh, C.-Y. Isomer Enumeration of Alkenes and Aliphatic Cyclopropane Derivatives. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 854–856. Yeh, C.-Y. Counting linear polyenes by excluding structures with steric strain. *J. Chem. Phys.* **1996**, *105*, 9706–9707. Yeh, C.-Y. Isomerism of asymmetric dendrimers and stereoisomerism of alkanes. *J. Mol. Struct.* **1998**, *432*, 153–156.
  - (18) Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Fuji, Z.; Xiaofeng, G.; Tosić, R. Graph Theoretical Studies on Fluorantenoids and Fluorenoids: Enumeration of Some Catacondensed Systems. *J. Mol. Struct.* **1993**, *285*, 179–185. Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Brendsdal, E.; Enumeration and Classification of Certain Polygonal Systems Representing Polycyclic Conjugated Hydrocarbons: Annulated Catafusenes. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1174–1180. Cyvin, B. N.; Brendsdal, E.; Brunvoll, J.; Cyvin, S. J. A Class of Polygonal Systems Representing Polycyclic Conjugated Hydrocarbons: Catacondensed Monohiptafusenes. *Monat. Chem.* **1994**, *125*, 1327–1337. Cyvin, S. J.; Brunvoll, J.; Brendsdal, E.; Cyvin, B. N.; Lloyd, E. K. Enumeration of Polyene Hydrocarbons: A Complete Mathematical Solution. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 743–751. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N. Enumeration of constitutional isomers of polyenes. *J. Mol. Struct.* **1995**, *357*, 255–261. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N.; and Brendsdal, E. Enumeration of Isomers and Conformers: A complete Mathematical solution for Conjugated Polyene Hydrocarbons. *Adv. Mol. Struct. Res.* **1996**, *2*, 213–245.
  - (19) Brunvoll, J.; Cyvin, B. N.; Brendsdal, E.; Cyvin, S. J. Computerized Enumeration of Staggered Alkane Conformers. *Comput. Chem.* **1995**, *19*, 379–388. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N.; Lüttke, W. Enumeration of the Staggered Conformers of Alkanes. *Zeit. Naturforsch.* **1996**, *50a*, 857–863.
  - (20) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20. Wiener, H. Correlation of Heats of Isomerization, and Differences in heats of Vaporization of Isomers, Among the Paraffin hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636–2638.
  - (21) Polansky, O. E.; Bonchev, D. The Wiener Number of Graphs. I. General Theory and Changes due to some Graph Operations. *Match* **1986**, *21*, 133–186. Rouvray, D. H. The Role of the Topological Distance Matrix In Chemistry. In *Mathematical and Computational Concepts in Chemistry*; Trinajstić, N. Ed.; Ellis Horwood: Chichester, 1986; pp 295–306. Special issue of *Match* **1997**, *35*, Gutman, I., Klavzar, S., Mohar, B., Eds.
  - (22) Tatevskii, V. M.; Benderskii, V. A.; Yarovoi, S. S. *Rules and Methods for Calculating the Physico-Chemical Properties of Paraffinic Hydrocarbons* English translation from Pergamon Press: Oxford, 1961.
  - (23) Benson, S. W. *Thermochemical Kinetics*; John Wiley and Sons: New York, 1976. Cohen, N.; Benson, S. W. Estimation of Heats of Formation of Organic Compounds by Additivity Methods. *Chem. Rev.* **1993**, *93*, 2419.
  - (24) Vogel, A. I. Physical Properties and Chemical Constitution. Part XXIII. Miscellaneous Compounds. *J. Chem. Soc.* **1948**, 1833–1854. Vogel, A. I.; Cresswell, W. T.; Jeffery, G. H.; Leicester, J. Physical Properties and Chemical Constitution. Part XXIII. Aliphatic Aldoximes, Ketoximes, and Ketoxime O-Alkyl Ethers, NN-Dialkylhydrazines, Aliphatic Ketazines, Mono- and Dialkylaminopropionitriles, Dialkyl Azodiformates, and Dialkyl Carbonates. *J. Chem. Soc.* **1952**, 514–549.
  - (25) Smolenskii, E. A. [Engl. trans.] Application of the Theory of Graphs to Calculations of the Additive Structural Properties of Hydrocarbons. *Russ. J. Phys. Chem.* **1964**, *35*, 700–702. Gordon, M.; Kennedy, J. W. The Graph-like state of matter. part 2. LCGI Schemes for the thermodynamics of alkanes and the theory of inductive inference. *J. Chem. Soc., Faraday Trans. 2* **1973**, *69*, 484–504. Essam, J. W.; Gordon, M.; Kennedy, J. W.; Whittle, P. The Graph-like state of matter. part 8. LCGI schemes and the statistical analysis of experimental data. *J. Chem. Soc., Faraday Trans. 2* **1977**, *73*, 1289–1307. Klein, D. J. Chemical Graph-Theoretic Cluster Expansions. *Intl. J. Quantum Chem. S* **1986**, *20*, 153–173.
  - (26) Schmalz, T. G.; Klein, D. J.; Sandleback, B. L. Chemical Graph-Theoretic Cluster Expansion and Diamagnetic Susceptibility. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 54–57.
  - (27) de Gennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
  - (28) Tomalia, D. A.; Naylor, A. M.; Goddard, W. A., III Starburst Dendrimers: Control of Size, Shape, Surface Chemistry, Topology, and Flexibility in the Conversion of Atoms to Macroscopic Materials. *Angew. Chem.* **1990**, *102*, 119–157. Frechet, J. M. J. Functional Polymers and Dendrimers – Reactivity, Molecular Architecture, and Interfacial Energy. *Science* **1994**, *263*, 1710–1714. Newkome, G. R.; Moorefield, C. N.; Vögtle, F. *Dendritic Macromolecules: Concepts, Syntheses, Perspectives*; VCH Pub.: New York, 1996. Zeng, F.; Zimmerman, S. C. Dendrimers in Supramolecular Chemistry: from Molecular Recognition to Self-Assembly. *Chem. Rev.* **1997**, *97*, 1681–1712.
  - (29) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. Applications of Combinatorial Technologies to Drug Discovery. I. Background and Peptide Combinatorial Libraries. *J. Med. Chem.* **1994**, *37*, 1233–1251. Zuckermann, R. N.; Martin, E. J.; Spellmeyer, D. C.; Stauber, G. B.; Shoemaker, K. R.; Kerr, J. M.; Figliozzi, B. M.; Goff, D. A.; Siani, M. A.; Simon, R. J.; Banville, S. C.; Brown, E. G.; Wang, L.; Richter, L. S.; Moos, W. H. Discovery of Nanomolecular Ligands for 7-trans- membrane G-protein-coupled Receptors from a diverse  $n$ -(substituted)-glycine Peptoid Library. *J. Med. Chem.* **1994**, *37*, 2678–2685. *Combinatorial Chemistry*; Czarnik, A. W., DeWitt, S. H., Eds.; American Chemical Society: Washington, DC, 1997.
  - (30) Thompson, C. J. *Mathematical Statistical Mechanics*; Macmillan Co.: New York, 1972.
  - (31) Brualdi, R. A. *Introductory Combinatorics*; North-Holland: Amsterdam, 1977; Chapter 7. Wilf, H. S. *Generating Functionology*; Academic Press: Boston, 1994.
  - (32) IUPAC. In *Nomenclature of Organic Compounds*; Pergamon Press: New York, 1979; pp 10–11.
  - (33) Davidson, S. An Improved IUPAC–Based Method for Identifying Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 151–155.
  - (34) Read, R. C. A new System for the Designation of Chemical Compounds. I. Theoretical Preliminaries and the Coding of Acyclic Compounds. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 135–149. Goodson, A.; Lozac'h, N. Advantages of Nodal Numbering for Uniquely Identifying Atoms in Chemical Nomenclature. *Croat. Chem. Acta* **1986**, *59*, 547–563.



- (35) Bonchev, D.; Balaban, A. T. Topological Centric Coding and Nomenclature of Polycyclic Hydrocarbons. 1. Condensed Benzenoid Systems (Polyhex Fusedes). *J. Chem. Inf. and Comput. Sci.* **1981**, *21*, 223–229.
- (36) Zimm, B. H.; Stockmayer, W. The Dimensions of Chain Molecules Containing Branches and Rings. *J. Chem. Phys.* **1949**, *17*, 1301–1314. Gordon, M. Good's theory of cascade processes applied to the statistics of polymer distributions. *Proc. R. Soc. A* **1962**, *268*, 240–259. Dobson, G. R.; Gordon, M. Theory of Branching Processes and Statistics of Rubber Elasticity. *J. Chem. Phys.* **1965**, *43*, 705. deGennes, P. G. Statistics of Branching and Hairpin Helices for the dAT Copolymer. *Biopolymers* **1968**, *6*, 715–729. Kajiwar, K. Statistics of randomly branched polycondensates: Part 2. The application of Lagrange's expansion method to homodisperse fractions. *Polymer* **1971**, *12*, 57–66.
- (37) Klein, D. J.; Seitz, W. A.; Kilpatrick, J. E. Branched polymer models. *J. Appl. Phys.* **1982**, *53*, 6599–6603.

CI980095C