

Mean Wiener Numbers and Other Mean Extensions for Alkane Trees

L. Bytautas[†] and D. J. Klein*

Texas A&M University at Galveston, Galveston, Texas 77551-1675

Received September 16, 1999

The “Wiener number” (the sum over intersite graph distances of a structure) as averaged over all alkane structural isomers of a fixed number N of carbon atoms is considered. This and several other measures of average graphical “extension” of N -site alkanes are computed for N up to 90 (where there are over 10^{35} such isomers). Fits are then made for several surmised or derived asymptotic forms, and a heuristic argument is made relating these results to geometric extensions of a random mix of (N -site) alkanes.

1. INTRODUCTION

There has long been interest in different graph-distance-based graph invariants for the graphs of the prototypical case of alkanes. Here the *graph* G associated to an alkane is just the network of C–C bonds (with H atoms and C–H bonds deleted). Also, the *graph distance* between two vertexes of G is the (unique) number of C–C bonds along the shortest path between the two vertexes. Then one graph invariant measuring “extension” (and introduced¹ over a half century ago) is the *Wiener number* $W(G)$, which is given as the sum over all distances between pairs of sites of G . This has often been used in structure–property or structure–activity correlations and is the subject of general articles² as well as a special issue³ of *MatCh*. A second invariant measuring graphical “extension” is the *diameter* $D(G)$, which is simply the maximal graph distance between two vertexes of G . The diameter is of central importance in the standard nomenclatural schemes for alkanes (and other structures). The mean such extents as averaged over isomeric classes of alkane structures is a natural point of interest and has been studied^{4–7} recently.

Lepović and Gutman⁴ studied the so-averaged Wiener number $\langle W(T) \rangle_N$ for structural isomers of alkanes with graphs T of up to $N = 20$ sites. Subsequently Dobrynin and Gutman⁶ used these data to suggest an asymptotic form for the mean

$$\langle W(T) \rangle_N \approx K \langle W(T) \rangle_{\text{lab},N} \quad (1.1)$$

where $K \approx 0.990$ and $\langle W(T) \rangle_{\text{lab},N}$ is the Wiener number as averaged over all site *labeled* trees with N sites. For these labeled trees, sites of degree greater than 4 are included in computing $\langle W(T) \rangle_{\text{lab},N}$ and also different weights are assigned to different structural isomers in accordance with the number of distinct ways of labeling the N sites with the integers from 1 to N . For example, for n -butane and isobutane the respective weights are 12 and 4. The utility of (1.1) is seen in that for the quantity on the right-hand side an analytic formula is available:

$$\langle W(T) \rangle_{\text{lab},N} = \sum_{k=1}^{N-1} \frac{k(k+1)N!}{2N^k(N-k-1)!} \quad (1.2)$$

as was earlier shown by Meir and Moon.⁸

In ref 5 a study was made of several mean graph invariants as averaged over N -site alkane isomers. A Pólya-counting-theoretic computational approach for these averages was developed, and for the average Wiener number $\langle W(T) \rangle_N$ a tabulation for T of up to $N = 40$ sites was reported. There an asymptotic form was suggested for the so-averaged Wiener number

$$\langle W(T) \rangle_N \approx \{A + BN^{-1/2} + CN^{-1}\} N^{3/2} (N-1)/2 \quad (1.3)$$

with $A \approx 1.2771$, $B \approx -1.437$, and $C \approx 0.86$. This was heuristically argued from analogy with some related earlier considered^{9–11} “extension measures” of randomly branched conformational structures. These earlier computed averages (which date back to about the same time as Wiener’s publications¹) are over all “idealized conformations” for all acyclic structures of a given number N of “monomers”, and the earlier “extension measure” is an average Euclidean intersite distance $\langle d_{E,ij}(G) \rangle_{\text{conf},N}$ for idealized geometric arrangements of the different conformations G . In any event the comparison might more naturally be to the mean intersite graph distance $\langle d_{ij}(G) \rangle_N$ rather than the Wiener number though these are simply related

$$\langle d_{ij}(G) \rangle_N = \langle W(G) \rangle_N / (2N(N-1)) \quad (1.4)$$

and indeed this accounts for a factor of $N(N-1)/2$ in (1.3).

There are several other plausible “extension measures” which might be compared to $\langle W(G) \rangle_N$. One such is the isomer-averaged diameter $\langle D(G) \rangle_N$, and another is the conformation-averaged diameter $\langle D(G) \rangle_{\text{conf},N}$. Also, these might plausibly be more naturally related to the mean intersite graph distance $\langle d_{ij}(G) \rangle_N$. Yet further there is a conformation-averaged Wiener number $\langle W(G) \rangle_{\text{conf},N}$.

Here a more comprehensive investigation is made of the interrelations between these different sorts of extension measures as applied for different sets of alkane structures of a fixed number N of sites. Since a computational scheme for $\langle W(T) \rangle_{\text{conf},N}$ does not seem to have been given previously,

[†] On leave from Institute of Theoretical Physics and Astronomy, Gostauto 12, Vilnius 2600, Lithuania.

we develop an analytic result for this in sections 2 and 3, where also exactly what is here meant by “conformation” is clarified. Section 4 develops some analytic asymptotic results. In section 5 some corrections to earlier computational formulas⁵ for the isomer-averaged Wiener numbers are made. These various somewhat technical sections 2–5 may be skipped over to the more numerical and descriptive results of sections 6–9. In section 6 the surmised asymptotic expression of the different extent measures in terms of N are detailed, and they are tested using exact results computed for up to $N = 90$ carbons. Section 7 then examines detailed asymptotic comparisons between the (five) different extent measures studied. Finally sections 8 and 9 present some discussion, with that of section 9 focusing on the effect of “volume exclusion” (e.g., as is relevant in “polymer statistics”).

2. ALKANE CONFORMATIONS AND THEIR ENUMERATION

Enumerations and mean Euclidean extents of idealized branched-polymer conformations have been considered previously.^{9,10} These idealized conformations embed an H-deleted graph G on a regular (e.g., diamond or cubic) lattice \angle without any attention to repeated embeddings of different sites of G onto the same site of \angle . That is, the sites of a graph G are mapped onto the sites of \angle such that, first, nearest neighbor sites of G are also nearest neighbor sites of \angle ; and, second, one of the sites of G is mapped onto the origin site of \angle . In ref 10 some improvement is made in avoiding embedment of next nearest neighbors of G on the same site of \angle . Notably this second improvement in the context of conformations simply for linear chain polymers has reached a wide popularity in the field of “polymer statistics” under the title of “rotational isomeric model”, e.g., as reviewed in Flory’s book.¹²

We may note some explanatory side points concerning different possible definitions of conformers and isomers. Basically different isomer classes correlate with different sets \mathcal{T} of transformations which relate different possible geometric structures in Euclidean 3-space as being equivalent. For structural isomers \mathcal{T} includes all transformations which when the associated formal graphical structure is considered in abstract space do not break bonds. Topological isomers replace transformations \mathcal{T} in abstract space with transformations in Euclidean 3-space, so that then, e.g., for a single cycle various knottings are distinguished—but for acyclic structures this refinement makes no difference. For stereoisomers the transformations (in Euclidean 3-space) are such that the geometric structures are not globally rigid but are rigid at each vertex (like a little tetrahedron at each site, for the case of alkanes) and all twistings, turnings, or stretchings of the bonds in Euclidean 3-space are allowed (again so long as they do not break). Diastereomers include in \mathcal{T} one additional transformation, namely, that of mirror imaging the whole structure; thence enantiomers are not distinguished, though different such pairs are distinguished as well as various “meso-structures”. Conformer structures differing via internal rotations or inversions are distinguished, whence \mathcal{T} includes only different degrees of overall translations and/or rotations. What exactly is chosen to be included in \mathcal{T} depends in part upon the intended use of the “conformer”

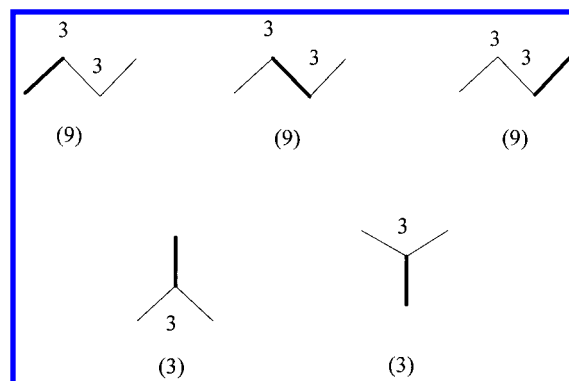


Figure 1. A presentation of the 33 “conformations” for different butanes (with $N = 4$). Here the given nonparenthetical numbers count the ways to make the embeddings of the next bond(s) at a site with the bonds fixed which are near to the origin bond (in boldface) of L . The total numbers of conformations with the center bond of L identified to the indicated boldface bond is then given by the parenthetical numbers.

concept. For polymer statistics purposes all degrees of freedom are to be summed over (in the system function), and rather than break the rotational sum up (with different factors for symmetric and nonsymmetric conformations) the rotations are imagined to be distinguished and only the translations are retained in \mathcal{T} . This leads to a definition we presently use (in the next paragraph). Organic or physical organic chemists, in considering conformers in correspondence with different local minima on a potential energy hypersurface, include all possible overall translations and rotations, and if attention were to be focused on symmetry inequivalent local minima, then mirror inversions may be included in \mathcal{T} . Such a definition is a little closer to the conformer definition we previously used in ref 5, though the auxiliary generating function for alkyl radicals is essentially the same as used here. In much of traditional polymer statistics, in Pólya enumeration theory, in refs 1–10 and 12, and in here the constraint of “volume exclusion” (requiring different monomer sites of a polymer to be assigned to distinct monomer cells in Euclidean space) is neglected to some degree, often because of mathematical convenience. Some consequences of this “volume exclusion” are indicated in ref 11, and the subject is returned to in the discussion section here.

Here we follow the polymer statistics idea, defining a *conformation* of G as a mapping of edges (or bonds) of G onto edges of \angle such that first adjoining edges of G are mapped to adjacent edges of \angle , and second some edge of G is mapped onto a fixed “origin” edge of \angle . Further, as appropriate for alkanes, we here choose \angle to be the diamond lattice. For example, then the different conformers at $N = 4$ are indicated in Figure 1.

The development proceeds to set up generating functions for counts of such idealized conformations first for (mono) radicals and second for nonradical alkanes. The radical structures may formally be viewed as *rooted alkyl trees* with the *root* (or radical) site of degree no greater than 3, and no other site of degree greater than 4. A *root* bond (representing an unsatisfied valence) is viewed to dangle from the root site. Let $\#_n$ be the number embeddings on \angle for such n -site rooted trees with the root bond of T embedded on the “origin bond” of \angle . Then the generating function for these acyclic radical conformations is defined as

$$\phi(t) \equiv \sum_n \#_n t^n \quad (2.1)$$

Here we take $\#_n = 0$ for all $n \leq 0$. Now if one considers the construction of Figure 2 to obtain a larger radical by appending three smaller ones to a new site to be the new root, then there corresponds a self-consistent self-recursion for the generating function

$$\phi(t) = t[1 + \phi(t)]^3 \quad (2.2)$$

Here the 1 added to the $\phi(t)$ on the right accounts for the fact that any of the three radical sites appended to the new root site may be a simple H-radical (with 0 carbons). With this recursion in hand one may neatly determine the $\#_n$ via Lagrange's (rather standard) expansion, whence

$$\#_n = \frac{d^{n-1} \{F(\phi)\}^n}{d\phi^{n-1}} \Big|_{\phi=0} \frac{1}{n!} \quad (2.3)$$

where $F(\phi)$ is the function of $\phi(t)$ appearing on the right-hand side of (2.2), excepting the factor of t . That is, $F(\phi) = (1 + \phi)^3$, and

$$\#_n = \frac{1}{2n+1} \binom{3n}{n} \quad (2.4)$$

where the parenthetic quantity here is a binomial coefficient, $(3n)!/(n!(2n)!)$. The nonradical alkane conformations can be generated by gluing two alkyl radicals together, making their two dangling root bonds one, at the location of the "origin" bond of \angle . The generating function for the alkane nonradicals (or $N \geq 2$ sites) then is $[\phi(t)]^2$. That is, the coefficient of t^N in $[\phi(t)]^2$ is just the number $\#_N^{(0)}$ of alkane conformations with some bond embedded on the "origin" bond of \angle . Thence

$$\#_N^{(0)} = \sum_{n=1}^{N-1} \#_n \#_{N-n} = \frac{2}{N} \binom{3N}{2N+2} \quad (2.5)$$

This really is a standard sort of result^{9,10} (and method of derivation), though more often it has been developed in the "polymer statistics" field imagining \angle to be the square-planar or cubic lattice. Here the manipulations have set the stage for the determination in the next section of the conformation-averaged Wiener number (such does not seem to have previously been considered in polymer statistics).

3. WIENER NUMBER AVERAGE OVER CONFORMATIONS

To obtain the mean graph distance averaged over conformers, a neat result dating back to Wiener¹ is utilized. This result gives the Wiener number of an acyclic structure T as a sum of bond contributions

$$W(T) = \sum_{i-j}^T m(i,j) m(j,i) \quad (3.1)$$

where the sum is over edges of T and $m(i,j)$ is the number of sites of T closer to i than to j . The bond contribution for the "origin" bond of L is particularly simple to deal with. For this bond the sum of such contributions over all conformations with n sites on one side of this bond and N

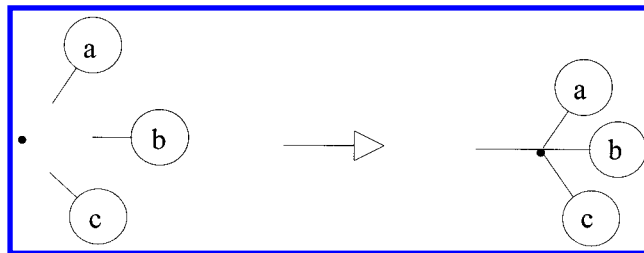


Figure 2. Imagined construction combining a-site, b-site, and c-site alkyl radicals together at a new (radical) site to form an $(a + b + c + 1)$ -site alkyl radical.

— n sites on the other side must be $n\#_n(N-n)\#_{N-n}$. Thence the contribution for the "origin" bond as averaged over all N -site alkane conformations must be

$$\sum_{n=1}^{N-1} n\#_n(N-n)\#_{N-n}/\#_N^{(0)} \quad (3.2)$$

Now the contributions for the bonds other than the "origin" bond of \angle are desired. Notably because of the regularity of \angle (and in particular the bond-transitivity of \angle), each alkane graph T has the same number of embeddings on \angle for any choice of bond of T to be identified to the "origin" bond of \angle . Thus the average total Wiener number is just a simple multiple of the contribution for the "origin" bond of \angle , and this multiple is $N - 1$, the number of choices of bond for any N -site T . That is, the Wiener number averaged over N -site conformations is

$$\langle W(T) \rangle_{\text{conf},N} = (N-1) \sum_{n=1}^{N-1} n\#_n(N-n)\#_{N-n}/\#_N^{(0)} \quad (3.3)$$

Then, since there are $N(N-1)/2$ intersite distances in an N -site graph, the average graph distance between two vertexes of N -site conformations is

$$\langle d_{ij} \rangle_{\text{conf},N} = \frac{2}{N(N-1)} \langle W(T) \rangle_{\text{conf},N} = \left(\binom{3N}{2N+2} \right)^{-1} \sum_{n=1}^{N-1} \frac{n}{2n+1} \binom{3n}{n} \frac{N-n}{2N-2n+1} \binom{3N-3n}{N-n} \quad (3.4)$$

This then is the final fairly concise result of this section (and is somewhat reminiscent of Meir and Moon's labeled-graph result,⁸ as given in (1.2)).

4. ANALYTIC ASYMPTOTICS

Analytic asymptotic behaviors for counts or averages may be derived in a few cases. All proceed by way of Stirling's factorial expansion

$$\ln(M!) \approx M \ln M - M + \frac{1}{2} \ln(2\pi M) + \frac{1}{12} \frac{1}{M} - \frac{1}{360} \frac{1}{M^3} + \dots \quad (4.1)$$

which is described in many standard texts.

First we consider the asymptotics of the conformer counts $\#_n$ and $\#_N^{(0)}$ in terms of the C-atom count N . These asymptotic forms result rather readily through substitution of the Stirling expansion for the various factorials appearing in the

binomial coefficients of (2.4) and (2.5), to yield

$$\begin{aligned} \#_n &\approx \frac{1}{4n} \left(\frac{3}{\pi n} \right)^{1/2} \left(\frac{27}{4} \right)^n \left\{ 1 - \frac{7}{6} \frac{1}{n} - \frac{5}{9} \frac{1}{n^2} \right\} \\ \#_n^{(0)} &\approx \frac{1}{4N} \left(\frac{3}{\pi N} \right)^{1/2} \left(\frac{27}{4} \right)^N \left\{ 1 - \frac{2}{3} \frac{1}{N} + \frac{25}{72} \frac{1}{N^2} \right\} \end{aligned} \quad (4.2)$$

Presumably even higher-order corrections might be generated by this approach.

As an aside, a brief comment might be made about the alternative definition of conformers used in ref 5. From some empirical evidence it seems that the asymptotic form for the count of these conformers differs by a factor of $3/2(N-1)$ from the asymptotic form for $\#_N^{(0)}$ in (4.2).

The leading asymptotic form for $\langle d_{ij} \rangle_{\text{conf},N}$ can also be deduced. One again simply uses Stirling's approximation for the various factorials appearing in the binomial coefficients of (3.4) to obtain

$$\langle d_{ij} \rangle_{\text{conf},N} \approx \frac{4(N+2)(2N+1)}{N(N-1)} \sum_{n=1}^{N-1} \frac{n}{2n+1} \frac{N-n}{2N-2n+1} \left(\frac{3N}{4\pi n(N-n)} \right)^{1/2} \quad (4.3)$$

Here the Stirling approximation is not extremely accurate when either n is near 1 or near $N-1$, and these terms in the summand are larger than any other single summand, but only marginally so. Thence it turns out the other terms overwhelm these near $n=1$ and $N-1$, so long as we seek only the leading term in the asymptotic expansion in (inverse) powers of $N^{1/2}$, whence

$$\langle d_{ij} \rangle_{\text{conf},N} \approx \left(\frac{3N}{4\pi} \right)^{1/2} \sum_{n=1}^{N-1} \{n(N-n)\}^{-1/2} \quad (4.4)$$

Conversion of the summation to an integral (à la Euler–Maclaurin) followed by analytic integration then leads to

$$\langle d_{ij} \rangle_{\text{conf},N} \approx \frac{(3\pi)^{1/2}}{2} N^{1/2} \quad (4.5)$$

which is the desired form.

The lead asymptotic dependence of the Wiener number averaged over labeled alkane trees of a given number of sites can also be similarly deduced. One again uses Stirling's factorial expansion now in Meir and Moon's⁸ result of (1.2) to obtain

$$\langle d_{ij} \rangle_{\text{lab},N} \approx \frac{2}{N} \sum_{k=1}^{N-1} k(k+1) \left(\frac{N}{N-k} \right)^{1/2} \exp \left\{ -\frac{(k+1)^2}{2N} \right\} \quad (4.6)$$

Then, noting that the contributions of the various summands here are strongly dominated by the relatively smaller k values (in fact by k up to $\sim N^{1/2}$), one can for the lead asymptotic behavior convert this to an integral over a variable $x \equiv k(2N)^{-1/2}$ to obtain

$$\langle d_{ij} \rangle_{\text{lab},N} \approx 2^{3/2} \int_0^\infty x^2 e^{-x^2} dx N^{1/2} \quad (4.7)$$

Thence

$$\langle d_{ij} \rangle_{\text{lab},N} \approx (\pi/2)^{1/2} N^{1/2} \quad (4.8)$$

which is a desired result.

5. WIENER-NUMBER AVERAGE FOR ALKANE STRUCTURAL ISOMERS

The Wiener number averaged over structural isomers was computed by a powerful generating-function technique in ref 5, but the development of the details of the technique is incomplete, and two key formulas given are incorrect. The approach in broad scope is similar in philosophy to that of section 2, but is more involved in what might be termed the elimination of repeated generation of the same structure. There are (in ref 5) generating functions $\phi_g(t)$ for g -generation alkyl radicals; the coefficient of t^n gives the number of n -site g -generation alkyl radicals. Suitably combining these alkyl-radical generating functions together (again somewhat as in Figure 2) then yields generating functions for enumeration of alkane structural isomers of a given diameter, and thereafter summation over diameters gives generating functions for enumeration of all alkane structural isomers. The development of generating functions for various properties involves inclusion of further variables attending to more graph invariants (beyond the site count), the whole trick being to identify constructive recursions for the associated generating functions. This overall approach seems to generally be neatly developed in two stages: first, developing multivariable generating functions with relevant invariants appearing as exponents of the different variables in the generating functions; second, taking derivatives with respect to the new auxiliary variables after which these auxiliary variables are set to 1, thereby yielding single-variable average-attendant generating functions (with a lesser proliferation of coefficients). The case of the Wiener number is somewhat more complicated than other graph invariants, and there is a third stage where the averages more conveniently so generated are then suitably combined (using the relation of (3.1)) to finally yield the Wiener-number averages. Anyway, in the first stage for the isomer case one introduces in parallel to the $\phi_g(t)$ of ref 5 two new two-variable generating functions

$$\phi_g^\zeta(s,t) \equiv \sum_n \#_{g,n,p}^\zeta s^p t^n \quad (5.1)$$

with s the second variable, with ζ taking two values (say α and β) for two properties, and with $\#_{g,n,p}^\zeta$ being the number of n -site generation- g alkyl radicals having a value p for our ζ th property. The first presently relevant property for a radical (or rooted) tree T is

$$m_\alpha(T) \equiv \sum_e^T m(T,e) \quad (5.2)$$

where the sum is over all edges e of T including the dangling (radicaloid) edge and $m(T,e)$ is the number of sites on the side of e away from the root site. The second presently relevant property for a radical (or rooted) tree T is

$$m_\beta(T) \equiv \sum_e^T [m(T,e)]^2 \quad (5.3)$$

where again the sum is over all edges e of T . Then in (5.1) $\#_{g,n,p}^\alpha$ is the number of n -site generation- g alkyl radicals T with $m_\alpha(T) = p$, and $\#_{g,n,p}^\beta$ is the number of n -site generation- g alkyl radicals T with $m_\beta(T) = p$. In ref 5 the function $\phi_g^\alpha(s,t)$ was introduced and called $\phi_g(s,t)$, and the rationale leading to the recursion ((4.4) in ref 5) for $\phi_g(s,t) = \phi_g^\alpha(s,t)$ may be used for $\phi_g^\beta(s,t)$ to obtain a formula which is the same except that the argument st in (4.4) is replaced by s^2t . The recursions of ref 5 also lead to generating functions $P_D^\zeta(s,t)$ for (unrooted nonradical) alkane trees of diameter D , merely with the addition of the superscript ζ . Thence the first stage is completed. Paralleling the ideas in ref 5 at the second stage, the Wiener-sum generating functions for even- and odd-diameter alkane trees are obtained as

$$\begin{aligned} W_{2g}(t) &= [\partial_s \partial_t P_{2g}^\alpha(s,t) - \partial_s P_{2g}^\beta(s,t)]_{s=1} \\ W_{2g-1}(t) &= [\partial_s \partial_t P_{2g-1}^\alpha(s,t) - \partial_s P_{2g-1}^\beta(s,t)]_{s=1} - \\ &\quad \{[t \partial_t \phi_g(t)]^2 + (t^2 \partial/\partial(t^2))^2 \phi_g(t^2)\}/2 \end{aligned} \quad (5.4)$$

where here ∂_x denotes the partial derivative with respect to x (for $x = s, t$). This then parallels the formulas (4.9) and (4.10) of ref 5 except that mistakenly in place of the present $\partial_s P_{2g}^\beta(s,t)$ there was $\partial_s(s \partial_s P_{2g}^\alpha(s,t))$; i.e., in ref 5 a first derivative of the second-power property was replaced by a second derivative of the first-power property. Though these two formulas given in ref 5 were incorrect, the computations reported there in fact used the correct formulas.

6. NUMERICAL RESULTS AND ASYMPTOTIC FITS

With several different measures of mean “extents” in hand, we have computed their values for alkane graphs with up to $N = 90$ carbons. Values for $\langle W(T) \rangle_N$ are obtained as in ref 5, with the formulas reported in ref 5 as corrected in section 4. The values for $\langle D \rangle_N$ and $\langle D \rangle_{\text{conf},N}$ are efficiently computed following the approach described in ref 5. The various numerical values which ultimately are used in asymptotic fits (i.e., for N from 40 to 90) are reported in Table 1. This then extends our previous tabulation⁵ (up to $N = 40$); the exact isomer counts up to $N = 100$ are given in ref 13. The values for $\langle W(T) \rangle_{\text{lab},N}$ are taken from (1.2) and those for $\langle W(T) \rangle_{\text{conf},N}$ from (3.4).

The comparisons are to be made in terms of intersite distances. That is, instead of now thinking in terms of some sort of mean Wiener number (say \bar{W}_N), we choose to think in terms of the corresponding mean intersite distance (say δ_N), it being understood that these two are simply related via

$$\bar{W}_N = N(N-1)\delta_N/2 \quad (6.1)$$

Naturally one has some intuition about δ_N , e.g., this clearly being between 1 and N . As (heuristically) argued previously⁵ (as well as in section 9 here), a “representative” distance average is expected to scale like $N^{1/2}$ (even if the “representative” distance is taken as the diameter rather than the mean intersite distance, all averaged over any one of our several

collections of N -site structures). This is further supported for $\langle d_{ij} \rangle_{\text{conf},N}$ in our analysis leading to (3.7) as well as our analysis for $\langle d_{ij} \rangle_{\text{lab},N}$ leading to (4.3). Indeed we imagine even higher order (asymptotic) corrections as given by

$$\delta_N \approx AN^{1/2} + B + C/N^{1/2} \quad (6.2)$$

(with A , B , and C constants, independent of N , but dependent on our choice of type of mean δ_N). Then a plot of

$$\delta_N/N^{1/2} \quad \text{versus} \quad 1/N^{1/2} \quad (6.3)$$

should approach linearity as the ordinate approaches 0. In Figure 3 such plots are shown for our five different choices of δ_N . It is seen that these plots appear to approach a definite limit in the $1/N^{1/2} \rightarrow 0$ limit, so that the correctness of the lead term is supported. That this limit seems to be approached with a (nonzero) slope supports the plausibility of the second term of our surmised asymptotic form of (6.2). That (at least some of) these plots display some curvature indicates the relevance of a higher order correction as the C -term of (6.2). Plots of δ_N/N^α vs $1/N^\alpha$ for α sufficiently close to $1/2$ (say with $|\alpha - 1/2| < 0.05$) look rather similar, but as already demonstrated (in section 4) for two of the extension measures $\langle d_{ij} \rangle_{\text{lab},N}$ and $\langle d_{ij} \rangle_{\text{conf},N}$ the exponents are exactly $\alpha = 1/2$, and we presume that this exponent is common for all five of our extension measures, as is supported by the similarity of the five different plots. A further argument as to the “universality” of this exponent is given in the first part of section 9.

With the presumption of this asymptotic form of (6.2), a quantitative least-squares fitting can be made for each of these five cases, using the numerical results for $N = 40-90$. In imagining the surmised asymptotic form the error due to the next higher term should be $\sim 1/N$, so that in making the fit we weight the lower orders less to compensate for this error; that is, the square error in the fit for δ_N at a given N is multiplied by a weight N^2 . The results of these consequent fits are displayed in Table 2, where we also report the computed least-squares errors (for this range of N values). The values obtained here for A , B , and C for the case of the structural-isomer mean $\langle d_{ij} \rangle_N$ compare modestly with the values ($A \approx 1.277$, $B \approx -1.44$, and $C \approx 0.86$) in ref 5, and for the case of the structural-isomer mean diameter $\langle D \rangle_N$, modest comparison is also obtained with the values ($A \approx 3.312$ and $B \approx -4.90$) in ref 7. In both cases the present values should be more accurate, since presently the fitting is to notably higher N . For the conformational mean $\langle d_{ij} \rangle_{\text{conf},N}$ only modest comparison is obtained with the exact value $A = (3\pi)^{1/2}/2 \approx 1.534\,990\,06$ of (4.5). If further $\langle d_{ij}(G) \rangle_{\text{conf},N}$ up through $N = 300$ are computed via (3.4) and a (weighted) fitting is made (for $N = 40-300$), one obtains $A \approx 1.532\,82$, which is in much closer correspondence with the exact result (and the agreement even improves further to $A \approx 1.534\,00$ if we restrict the fit to the range $N = 200-300$). Finally for $\langle d_{ij} \rangle_{\text{lab},N}$ modest comparison is obtained with the value $A = (\pi/2)^{1/2} \approx 1.253\,314\,137$ of (4.3), and again use of (1.2) with higher N values up to $N = 300$ leads to even closer agreement.

7. ASYMPTOTIC COMPARISONS

One may also imagine representing any one of the representative distance (or extension) measures in terms of

Table 1. Numerical Values for $\langle W \rangle_N$, $\langle d_{ij} \rangle_N$, $\langle D \rangle_N$ and $\langle D \rangle_{\text{conf},N}$

N	$\langle W \rangle_N$	$\langle d_{ij} \rangle_N$	$\langle D \rangle_N$	$\langle D \rangle_{\text{conf},N}$	N	$\langle W \rangle_N$	$\langle d_{ij} \rangle_N$	$\langle D \rangle_N$	$\langle D \rangle_{\text{conf},N}$
2	1.000 000	1.000 000	1.000 000	1.000 000	46	7610.147 596	7.352 800	17.444 548	18.662 468
3	4.000 000	1.333 333	2.000 000	2.000 000	47	8048.216249	7.445158	17.684 337	18.904 472
4	9.500 000	1.583 333	2.500 000	2.818 182	48	8501.253 577	7.536 572	17.921 698	19.143 821
5	18.000 000	1.800 000	3.000 000	3.582 418	49	8969.434 172	7.627 070	18.156 703	19.380 598
6	31.000 000	2.066 667	3.800 000	4.270 588	51	9951.914 570	7.805 423	18.619 924	19.846 747
7	48.000 000	2.285 714	4.333 333	4.913 313	52	10 466.554 742	7.893 329	18.848 269	20.076 266
8	69.555 556	2.484 127	4.833 333	5.515 260	53	10 997.019 025	7.980 420	19.074 518	20.303 505
9	96.685 714	2.685 714	5.342 857	6.082 852	54	11 543.473 498	8.066 718	19.298 730	20.528 530
10	129.520 000	2.878222	5.840000	6.621230	55	12 106.082 686	8.152 244	19.520 959	20.751 401
11	168.251 572	3.059 119	6.308 176	7.134551	56	12 685.009 593	8.237 019	19.741 257	20.972 179
12	213.622535	3.236 705	6.771 831	7.625 810	57	13 280.415 752	8.321 063	19.959 675	21.190 918
13	265.658 354	3.405 876	7.206 983	8.097 472	58	13 892.461 262	8.404 393	20.176 261	21.407 674
14	324.968 245	3.571 080	7.636 706	8.551 626	59	14 521.304 821	8.487 028	20.391 061	21.622 497
15	391.734 069	3.730 801	8.050 610	8.990 023	60	15 167.103 769	8.568 985	20.604 119	21.835 437
16	466.333 333	3.886 111	8.453 133	9.414 139	61	15 830.014 117	8.650 281	20.815 476	22.046 541
17	549.003 374	4.036 790	8.843 858	9.825 231	62	16 510.190 582	8.730 931	21.025 175	22.255 854
18	640.122 268	4.183 806	9.225 551	10.224 383	63	17 207.786 620	8.810 951	21.233 254	22.463 421
19	739.903 550	4.326 921	9.596 781	10.612 539	64	17 922.954 454	8.890 354	21.439 750	22.669 282
20	848.670 500	4.466 687	9.959 448	10.990 526	65	18 655.845 103	8.969 156	21.644 699	22.873 479
21	966.666 833	4.603 175	10.313 421	11.359 075	66	19 406.608 414	9.047 370	21.848 136	23.076 049
22	1094.176024	4.736693	10.659 654	11.718 835	67	20 175.393 085	9.125 008	22.050 095	23.277 031
23	1231.441 028	4.867 356	10.998 411	12.070 382	68	20 962.346 690	9.202 084	22.250 608	23.476 459
24	1378.726 337	4.995 385	11.330 334	12.414 234	69	21 767.615 711	9.278 609	22.449 705	23.674 368
25	1536.270 837	5.120 903	11.655 721	12.750 858	70	22 591.345 553	9.354 594	22.647 417	23.870 793
26	1704.322 923	5.244 071	11.975 026	13.080 675	71	23 433.680 573	9.430 053	22.843 773	24.065 763
27	1883.116 445	5.365 004	12.288 536	13.404 067	72	24 294.764 102	9.504 994	23.038 800	24.259 312
28	2072.887 446	5.483 829	12.596 587	13.721 382	73	25 174.738 463	9.579 429	23.232 525	24.451 467
29	2273.863 224	5.600 648	12.899 440	14.032 940	74	26 073.744 998	9.653 367	23.424 975	24.642 259
30	2486.270 273	5.715 564	13.197 366	14.339 032	75	26 991.924 081	9.726 819	23.616 174	24.831 716
31	2710.329 043	5.828 665	13.490 596	14.639 925	76	27 929.415 144	9.799 795	23.806 148	25.019 863
32	2946.257 892	5.940 036	13.779 357	14.935 868	77	28 886.356 689	9.872 302	23.994 918	25.206 728
33	3194.270 696	6.049 755	14.063 849	15.227 089	78	29 862.886 315	9.944 351	24.182 510	25.392 335
34	3454.578 714	6.157 894	14.344 264	15.513 799	79	30 859.140 726	10.015 950	24.368 943	25.576 709
35	3727.389 587	6.264 520	14.620 777	15.796 195	80	31 875.255 756	10.087 106	24.554 241	25.759 874
36	4012.908 247	6.369 696	14.893 550	16.074 459	81	32 911.366 380	10.157 829	24.738 423	25.941 852
37	4311.336 522	6.473 478	15.162 736	16.348 762	82	33 967.606 735	10.228 126	24.921 510	26.122 665
38	4622.873 601	6.575 923	15.428 474	16.619 263	83	35 044.110 130	10.298 005	25.103 521	26.302 336
39	4947.715 915	6.677 080	15.690 898	16.886 110	84	36 141.009 065	10.367 472	25.284 476	26.480 885
40	5286.057 394	6.776 997	15.950 130	17.149 444	85	37 258.435 244	10.436 536	25.464 392	26.658 332
41	5638.089 462	6.875 719	16.206 287	17.409 394	86	38 396.519 590	10.505 204	25.643 288	26.834 697
42	6004.001 194	6.973 288	16.459 478	17.666 085	87	39 555.392 259	10.573 481	25.821 181	27.009 998
43	6383.979 361	7.069 745	16.709 805	17.919 631	88	40 735.182 650	10.641 375	25.998 087	27.184 256
44	6778.208 535	7.165 125	16.957 366	18.170 143	89	41 936.019 425	10.708 892	26.174 024	27.357 487
45	7186.871 153	7.259 466	17.202 251	18.417 723	90	43 158.030 512	10.776 038	26.349 006	27.529 708

any other, as, e.g., was considered for $\langle d_{ij} \rangle_N$ in terms of $\langle d_{ij} \rangle_{\text{lab},N}$ by Dobrynin and Gutman.⁶ Generally suppose a representative distance ξ is to be expressed in terms of another ζ , where we now suppress the N -dependence, it being understood that both are taken at the same value of N . Then asymptotically we anticipate

$$\xi \approx A_{\xi\zeta}\zeta + B_{\xi\zeta} + C_{\xi\zeta}/\zeta \quad (7.1)$$

The reasonability of these expansions for the asymptotic regime can be illustrated in terms of plots of ξ/ζ vs $1/\zeta$, which should turn out to be linear as ordinate 0 is approached. Such plots do so (as is more or less implied by the results of the preceding paragraph). Table 3 then gives these various A , B , and C coefficients for our suite of five extensions measures. All are estimated by a weighted least-squares fitting for $N = 40$ – 90 (with the weights chosen in the same manner as in the preceding paragraph). Thence we perform weighted least-squares fits for each pair of extension measures, using data for $N = 40$ – 90 . The results of the fits are displayed in Table 3, which then allows the asymptotic expression of any one of our representative distances in terms of any other, and of course Table 2 asymptotically expresses

each in terms of the simple atom count N . If an unweighted least-squares fit is made, the last two or three of the reported digits of A , B , and C in Tables 2 and 3 are different, but presumably the weighted-fit values reported are somewhat more accurate estimates of the true asymptotic limiting values. Incidentally we find the lead constant A for the expression of $\langle d_{ij} \rangle_N$ (averaging over structural isomers) in terms of $\langle d_{ij} \rangle_{\text{lab},N}$ (averaging over labeled alkane graphs) to be slightly different from that of Dobrynin and Gutman:⁶ our $A \approx 1.024\,43$ whereas that of ref 4 was estimated as 0.990. Moreover, examination of Figure 4 makes it difficult to imagine that our value for A is in error by even as much as 0.001, and in fact because of the empirical evidence for the correctness of the surmised asymptotic form to which the fit is made, we believe that our estimate is 10 times more accurate than this. Clearly, though, to obtain for a range of N the accuracy (of 0.1%) sought in ref 4, corrections to the leading term are needed (and are available now in our Table 3, likely to an even higher degree of accuracy). Incidentally this plot (of Figure 4) provides further evidence for the commonality of the lead exponent $\alpha = 1/2$ (at least between $\langle d_{ij} \rangle_N$ and $\langle d_{ij} \rangle_{\text{lab},N}$).

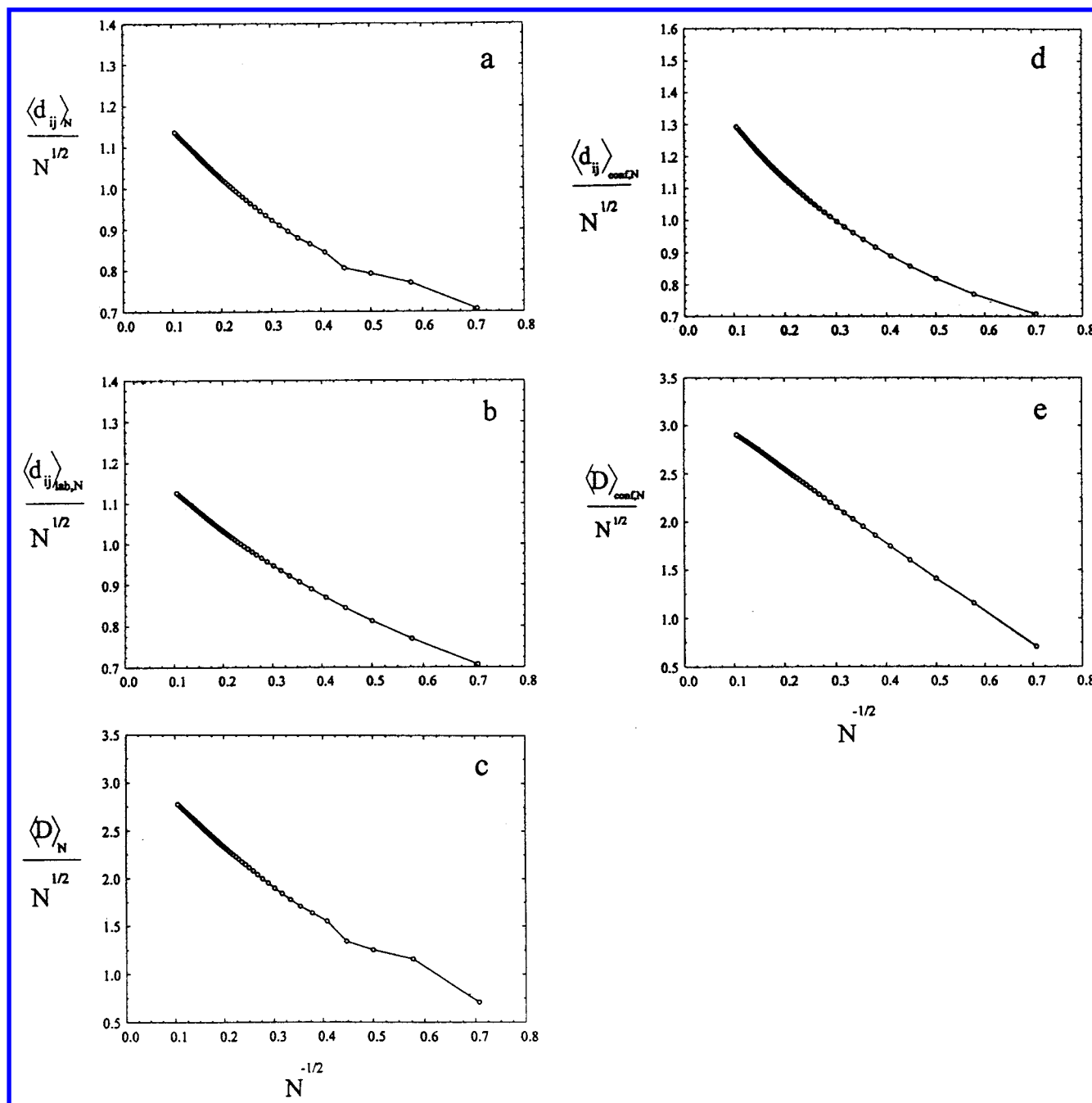


Figure 3. Different representative intersite distances δ_N for N -site alkanes scaled by $N^{1/2}$ and then plotted against $1/N^{1/2}$. For these plots (a) $\delta_N = \langle d_{ij} \rangle_N$; (b) $\delta_N = \langle d_{ij} \rangle_{\text{lab},N}$; (c) $\delta_N = \langle D \rangle_N$; (d) $\delta_N = \langle d_{ij} \rangle_{\text{conf},N}$; (e) $\delta_N = \langle D \rangle_{\text{conf},N}$.

Table 2. Asymptotic Results for Five Different Representative Intersite Distances δ_N for N -site Alkanes^a

	δ_N	A	B	C	err ₉₀ ^e	σ^d
1	$\langle d_{ij} \rangle_N$	1.28208	-1.4976	1.051	-2.83E-6	3.7E-6
2	$\langle d_{ij} \rangle_{\text{lab},N}$	1.25138	-1.2853	0.945	-2.29E-6	2.8E-6
3	$\langle D \rangle_N$	3.35295	-5.8711	3.898	-1.33E-5	3.9E-5
4	$\langle d_{ij} \rangle_{\text{conf},N}$	1.52851	-2.5049	2.590	-6.62E-6	9.3E-6
5	$\langle D \rangle_{\text{conf},N}$	3.21686	-2.5668	-3.990	2.55E-5	7.6E-5

^a Asymptotic form: $\delta_N \approx AN^{1/2} + B + C/N^{1/2}$. ^b The range of fitted data: $N = 40-90$. Weighted least-squares fit. ^c $\text{err}_N = (y_N^{\text{fit}} - y_N)/y_N$. ^d $\sigma = (\sum_i (y_i^{\text{fit}} - y_i)^2)^{1/2}/n$; where n is the number of fitted data points (=51).

General tests of the quality of the fits may be made. Such tests are reported in Table 4, where for each pair of extension measures there are given, first, the standard deviation σ for

the fit for $N = 40-90$; second, the relative error err_N at $N = 90$ between the fit and the exact expectation, and, third, another quality index $\Delta_{\xi\zeta}$ explained in the next paragraph, but essentially estimating an error in the A coefficients. Seemingly the expressions are fairly precise, to $<0.1\%$.

In general there are interrelations among the various comparison expressions of (7.1). For instance, the relation reverse to (7.1) may be obtained by, first, multiplying (7.1) by ζ to make it a quadratic in ζ , second, solving this quadratic to yield ζ in terms of ξ , and third, expanding this expression in inverse powers of ξ . Thence there results an interrelation

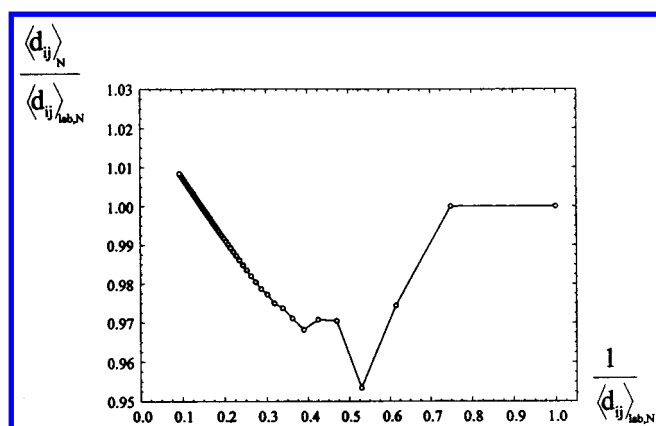
$$A_{\xi\xi} = 1/A_{\xi\zeta}, \quad B_{\xi\xi} = -B_{\xi\zeta}/A_{\xi\zeta}, \quad C_{\xi\xi} = -C_{\xi\zeta} \quad (7.2)$$

A similar approach gives the constants for relating ζ to η in terms of constants for relating ξ to ζ as well as those for

Table 3. Asymptotic Results for Interrelation of δ_N in Alkane Isomer Combinatorics^{a,b}

α	β				
	$\langle d_{ij} \rangle_N$	$\langle d_{ij} \rangle_{\text{lab},N}$	$\langle D \rangle_N$	$\langle d_{ij} \rangle_{\text{conf},N}$	$\langle D \rangle_{\text{conf},N}$
$\langle d_{ij} \rangle_N$		$A = 1.02443$ $B = -0.1779$ $C = 0.073$	$A = 0.38273$ $B = 0.7218$ $C = -0.792$	$A = 0.84037$ $B = 0.5491$ $C = -1.024$	$A = 0.39657$ $B = -0.3388$ $C = 5.421$
$\langle d_{ij} \rangle_{\text{lab},N}$	$A = 0.97617$ $B = 0.1733$ $C = -0.069$		$A = 0.37363$ $B = 0.8762$ $C = -0.917$	$A = 0.82037$ $B = 0.7086$ $C = -1.073$	$A = 0.38714$ $B = -0.1584$ $C = 5.127$
$\langle D \rangle_N$	$A = 2.61369$ $B = -1.9090$ $C = 0.996$	$A = 2.67776$ $B = -2.3800$ $C = 1.218$		$A = 2.19620$ $B = -0.4640$ $C = -1.622$	$A = 1.03629$ $B = -2.7821$ $C = 16.54$
$\langle d_{ij} \rangle_{\text{conf},N}$	$A = 1.19034$ $B = -0.6653$ $C = 1.153$	$A = 1.21969$ $B = -0.8843$ $C = 1.276$	$A = 0.45518$ $B = 0.2205$ $C = 1.448$		$A = 0.47179$ $B = -1.0542$ $C = 9.202$
$\langle D \rangle_{\text{conf},N}$	$A = 2.51825$ $B = 0.9252$ $C = -5.725$	$A = 2.57846$ $B = 0.5122$ $C = -5.719$	$A = 0.96573$ $B = 2.6117$ $C = -13.88$	$A = 2.11786$ $B = 2.2505$ $C = -8.632$	

^a Asymptotic form: $\alpha \approx A\beta + B + C/\beta$. ^b The range of fitted data: $N = 40-90$. Weighted least squares fit.

**Figure 4.** Plot of $\langle d_{ij} \rangle_N / \langle d_{ij} \rangle_{\text{lab},N}$ vs $1 / \langle d_{ij} \rangle_{\text{lab},N}$.**Table 4.** Statistics for the Fittings of Table 3^a

α		β				
		$\langle d_{ij} \rangle_N$	$\langle d_{ij} \rangle_{\text{lab},N}$	$\langle D \rangle_N$	$\langle d_{ij} \rangle_{\text{conf},N}$	$\langle D \rangle_{\text{conf},N}$
$\langle d_{ij} \rangle_N$	σ		1.0E-6	1.4E-5	8.5E-6	4.8E-5
	err ₉₀		-6.5E-7	1.1E-5	6.4E-6	-3.7E-5
	$\Delta_{\alpha\beta}$		1.6E-5	1.3E-4	2.8E-4	-5.3E-4
$\langle d_{ij} \rangle_{\text{lab},N}$	σ	1.0E-6		1.5E-5	9.4E-6	4.5E-5
	err ₉₀	6.6E-7		1.1E-5	6.9E-6	-3.6E-5
	$\Delta_{\alpha\beta}$	1.5E-5		1.8E-4	4.8E-4	-6.9E-4
$\langle D \rangle_N$	σ	3.3E-5	3.5E-5		1.2E-5	1.6E-4
	err ₉₀	-1.1E-5	-1.2E-5		-4.2E-6	-5.1E-5
	$\Delta_{\alpha\beta}$	8.6E-4	1.3E-3		-7.3E-4	8.1E-4
$\langle d_{ij} \rangle_{\text{conf},N}$	σ	9.3E-6	9.7E-6	4.0E-6		6.9E-5
	err ₉₀	-6.2E-6	-6.5E-6	3.1E-6		-4.7E-5
	$\Delta_{\alpha\beta}$	3.9E-4	7.2E-4	-1.5E-4		-3.8E-4
$\langle D \rangle_{\text{conf},N}$	σ	1.1E-4	1.0E-4	1.6E-4	1.3E-4	
	err ₉₀	3.3E-5	3.2E-5	4.7E-5	4.0E-5	
	$\Delta_{\alpha\beta}$	-3.4E-3	-4.6E-3	7.5E-4	-1.7E-3	

^a $\sigma = (\sum_i (y_i^{\text{fit}} - y_i)^2)^{1/2}/n$, where $n = 51$ is the number of fitted data points. $\text{err}_N = (y_N^{\text{fit}} - y_N)/y_N$. $\Delta_{\alpha\beta} = (A_{\alpha\beta} - 1/A_{\beta\alpha})$.

relating ξ to η , thus

$$A_{\eta\xi} = A_{\xi\xi}/A_{\xi\eta}, \quad B_{\eta\xi} = (B_{\xi\xi} - B_{\xi\eta})/A_{\xi\eta}, \\ A_{\xi\eta}, \quad C_{\eta\xi} = (C_{\xi\xi}/A_{\xi\eta}) - (C_{\xi\eta}/A_{\xi\xi}) \quad (7.3)$$

The actual fits do not exactly satisfy these relations, because of the “errors” due to neglected higher order terms. Thence the degree of satisfaction of these relations might be used as an estimate of the reliability of the fits. Indeed the third

quality index (for each pair ξ, ζ of extension measures) as reported in Table 4 is just the error in the lead term $A_{\xi\xi}$ as compared to $1/A_{\xi\xi}$, which as seen in (7.2) should (ultimately in the limit) be exactly the same. Thence this error is indicative of an error for the lead A coefficients $A_{\xi\xi}$.

8. DISCUSSION

The results of the preceding section may be utilized in different ways. A straightforward use is in developing simple asymptotic approximations for some of the mean properties. For example, following Dobrynin and Gutman,⁶ one can estimate the more difficult to compute mean $\langle d_{ij} \rangle_N$ in terms of the analytically available $\langle d_{ij} \rangle_{\text{lab},N}$, now with notably greater precision than previously. There are also other uses. The results of section 7 give the isomer averages $\langle d_{ij} \rangle_N$ and $\langle D \rangle_N$ in terms of N (which of course is even more easily computable, and the error estimates are nearly as small). Further, the manner of agreement between all these fits (of Tables 3 and 4) again suggests that all five of our extension measures δ_N share a common size-extension exponent $\alpha = 1/2$.

The ratio of graph diameter to mean intersite graph distance could be used to characterize some aspect of the shape of a structure G . Both D and the mean d_{ij} for acyclic structures generally decrease with branching, so that maximum values occur for the linear chain (or path) and their minimum values occur for the maximally branched dendrimeric “near-Bethe” trees. The perfect (degree-4) *Bethe tree* has only vertexes of degrees 4 and 1, so that there is a condition on the total number N of sites. For other nearby N there are realizable *near-Bethe* trees which have a minimal number of degree 3 (defect) sites, and they have similar values for D and mean d_{ij} . This ratio

$$\rho \equiv DN(N-1)/2 \sum_{i < j} d_{ij} \quad (8.1)$$

may be computed to be $\approx 3(1 - 2/N)$ for the linear-chain (normal) alkane, and for maximally branched Bethe-tree-like structures $\rho \approx 1 + A/(\log N)$, as follows from the analytic results of Gutman et al.¹⁴ for these structures. Indeed, from their result for perfect Bethe trees we suspect that $A = 7/(8 \log 3)$. Presumably other acyclics generally have ratios

ρ intermediate between $3(1 - 2/N)$ and $1 + A/(\log N)$. Thence it seems that the farther ρ deviates above 1 (toward 3) the “skinnier” (more like a chain) is the structure, whereas the nearer ρ is to 1 the more highly branched is a structure. From Table 3 it is seen that the corresponding ratio of corresponding averages is (asymptotically) ≈ 2.61 for isomer averages, so that it seems that the class of isomers might not be dominated so strongly by extremely branched structures. The corresponding (asymptotic) ratio of ≈ 2.12 for conformations suggests that the more highly branched structures are a little more relevant in this case. Thence especially for the case of isomer counts (and related property averages), the “errors” due to the neglect of volume exclusion may be expected to be more minor (than for conformational averages). The general effect of volume exclusion is considered further in the next section.

9. SIZE-EXTENSION EXPONENTS AND VOLUME EXCLUSION

The various exponents which appear in our various asymptotic relations are of some interest. Such exponents in statistical mechanics and in particular are generally believed to exhibit some degree of “universality”, e.g., as discussed in the currently directly relevant context of “polymer statistics” in ref 15. One “classical” case of such an exponent is that of $\nu_0 = 1/2$ for the mean geometric extension $\sim N^{\nu_0}$ of an N -site random walk, as is of relevance in Brownian motion and other diffusion processes (as, e.g., discussed in ref 16 and also in ref 12). In the context of the present paper examples are found with the lead-term exponents of $\alpha = 1/2$ in the asymptotic relations of Table 2 (and of (4.5) and (4.8)). Evidently this value appears to be of some degree of “universality”, arising independently of the sort of average (over isomers or over conformers, presumably even if defined in other ways) and independently of whether the extension measure is a mean distance or a diameter. Presumably too it is largely independent of the “alkaneic” conditions on the degrees of the branch sites.

It may be argued that the random-walk geometric-extension exponent ν_0 and the branched-structure graphical-extension exponent α should be the same. To do so imagine a sequential labeling of a tree T of n edges to be obtained as follows:

- embed T (without crossing bonds) in the Euclidean plane;
- starting from a selected temporary “root” site of T , imagine tracing a walk around the outside edge of the embedding keeping the adjacent part of the tree always on one side of the walker;
- record a sequence of signs $\sigma_1, \sigma_2, \dots, \sigma_{2n}$ with $\sigma_i = +$ or $-$ as the i th step of the walk is directed away or toward the initial selected site.

For example, for the Y -tree of isobutane the sequence is $+, -, +, -, +, -$ if one starts from the central site, and it is $+, +, -, +, -, -$ if one starts from a terminal (degree 1) site. Evidently each edge is traversed twice (first with the walker stepping away from the root site and second with the walker stepping toward the root site), and it is seen that T may be reconstructed from the label $\sigma_1, \sigma_2, \dots, \sigma_{2n}$. Notably such a sequence also represents a one-dimensional random walk of $2n$ steps with σ_i indicating the direction of the i th step along the axis, and also notably the geometric extension

of this one-dimensional random walk corresponds precisely to graphical extension for the corresponding walk around T . Thus the two types of extension (including the exponents) are the same. The occurrence of a (temporary) root site is not important, in that to obtain the Wiener number one just needs to sum over results for all $(n + 1)$ choices of the root. Of course the one-dimensional walks are restricted, with the number of minus ($\sigma_i = -$) steps never at any stage of the walk exceeding the number of plus ($\sigma_i = +$) steps, but this is just a “decoration”, which we expect does not affect the “universality” of the $\nu_0 = 1/2$ exponent. In fact, this sort of random walk starting from 0 and forbidden to visit any negative site has been treated exactly sometime ago,¹⁷ and our expectation is correct. There is a further restriction on the trees T if they are to correspond to (degree-limited) alkanes, and this translates to a restriction on the corresponding one-dimensional walks, though again we expect that this does not affect the “universality” of the $\nu_0 = 1/2$ exponent. Thus it seems that our graphical-extension exponent for various ensembles of various N -site branched structures is

$$\alpha = \nu_0 = 1/2 \quad (9.1)$$

which again is supported by the analytic results of section 4 (for two special ensembles) as well our numerical data leading to the plots of Figure 3 (for all five of the considered ensembles). A question remains as to whether the exponent could be affected by the condition of “volume exclusion” (demanding that the implicit structures be embeddable on a regular lattice in Euclidean space) changes the universality class.

The present scaling result for extension may be utilized in a heuristic manner to consider the geometric extension of alkane structures satisfying the constraint of volume exclusion. That is, using our present result of graphical extensions $\sim N^\alpha$, with $\alpha = 1/2$, one may consider the scaling for mean geometric extension for conformations when self-avoidingly embedded on the lattice \mathcal{L} . It has been shown^{9,10} that the root-mean-square radii of gyration for the conformation models without volume exclusion scale as $\sim N^{1/4}$. Thus evidently a conformation’s “backbone” being identified as a graph diameter ($D \sim N^{1/2}$) seems to be embedded into the underlying lattice \mathcal{L} like a random walk of length $\sim N^{1/2}$. Also, one may surmise about what happens if instead one assumes that such typical same-length backbones are developed as self-avoiding chains, with some enlarged cross section due to the remnant sites branching off to the sides of the backbone. Then the characteristic number of sites involved in the cross-sectional area at a particular position along the backbone would be $\sim N/N^\alpha = N^{1/2}$, so that these could be imagined to spread out like a branched structure in the two-dimensional cross-sectional area at least $\sim N^{1/2}$, so that there is a characteristic linear (graphical) extent for a cross-sectional radius r of $\sim (N^{1/2})^\beta = N^{\beta/2}$, with β at least $1/2$ (as for a compacted cross section). That is, one might imagine the typical conformation behaves like a self-avoiding polymer chain with radius r (perhaps $\sim N^{1/4}$) and a length of $L \sim N^{1/2}$, or if we measure this length in units of the cross-sectional radius it is $\sim L/r \sim N^{1/4}$. Now use may be made of the standard¹⁵ presumed scaling behavior for ordinary self-avoiding polymer chains: a mean geometric extent in units of r for such a chain is $\sim (L/r)^\nu \sim N^{\nu/4}$ where ν is the so-

called “Flory exponent”, which has been simply estimated¹⁸ to take a value $\approx 3/5$. Thence the anticipated mean geometric extent in our basic units (where $r \sim N^{1/4}$) is

$$\sim L^\gamma \sim N^{\nu/4} r \sim N^{\nu/4} N^{1/4} \sim N^{2/5} \quad (9.2)$$

This consequent argued size-extension scaling exponent of $\gamma = 2/5$ for self-avoiding branched conformations turns out to be a little different than that of $\gamma = 1/2$ obtained via a different heuristic argument¹⁹ as well as some many-body work,²⁰ though it is a little closer to the value of $\gamma \approx 0.46$ estimated²¹ from numerical fitting to Monte Carlo samplings (only up through $N = 60$).

Because the present argument for the value of this exponent γ does not seem to agree very closely with other “accepted” values, a closer examination is warranted. First the heuristic argument for the mean cross-sectional radius r was really a lower bound, in that the mean number $\sim N^{1/2}$ of side branches should in general not be maximally compacted and expand to cover a geometric range greater than $\sim (N^{1/2})^{1/2}$. A better estimate would be to take the geometric range to be $\sim (N^{1/2})^\beta$ where β is the geometric size-extension exponent for self-avoiding structures in two dimensions, this exponent being believed²⁰ to have an exact value $\beta = 3/5$, in fairly close agreement ($\beta \approx 0.62$) with²¹ two-dimensional Monte Carlo results²¹ (now up through $N = 600$), and with a value ($\beta = 5/8$) from a heuristic argument,¹⁹ though a more recent Monte Carlo result²² estimates $\beta \approx 0.64$. Further, an improved empirically tested value²³ for the three-dimensional self-avoiding-chain exponent is quite accurately estimated as $\nu \approx 0.588$. Finally, one might imagine that the exponent for mean graph distance for self-avoiding branched conformations is different than that (of $1/2$) we have for non-self-avoiding branched conformations. Then introduction of these improved values into the argument of the preceding paragraph gives

$$\gamma = \alpha\nu + (1 - \alpha)\beta(1 - \nu) \quad (9.3)$$

If one again assumes that $\alpha = 1/2$ (but now with $\nu \approx 0.588$ and with β in the range from 0.60 to 0.64), then $\gamma \approx 0.418$ – 0.426 , which still differs from the “accepted” value²⁰ of $\gamma = 1/2$. Thus it seems that for self-avoiding structures $\alpha \neq 1/2$, and to bring agreement must be $\alpha \approx 0.732$ – 0.742 . That is, we have evidence that the graph-distance-extension exponent α takes different values depending on whether the ensemble of branched conformations averaged over is non-self-avoiding (when $\alpha = 1/2$) or self-avoiding (when $\alpha \approx 0.74$). Thus volume exclusion evidently (rather noticeably) affects this exponent.

One might also consider the influence of “volume exclusion” on branched conformations embedded instead in two-dimensions. Then the argument is similar, with (9.3) still applying, but now with α the geometric-size exponent for self-avoiding branched structures in two-dimensions²⁰ being 0.60–0.64, with ν the accepted geometric-size exponent for self-avoiding chains in two-dimensions^{15,24} being $3/4$, and with β the one-dimensional geometric-size exponent for self-avoiding chains being 1. Then for two-dimensional self-avoiding branched conformations we estimate a graph-extension exponent $\gamma \approx 0.70$ – 0.78 . Perhaps the general result that graph-extension exponents are different for non-self-avoiding and self-avoiding branched structures should

be different is not too surprising—it is well accepted (as, e.g., in refs 15 and 18) that different geometrical-extension exponents apply for self-avoiding and non-self-avoiding chain conformations.

10. CONCLUSIONS

Evidence has been provided that several different graph-theoretic-related “extent measures” are asymptotically inter-related in a systematic fashion. There are some new analytic results: for the mean intersite graph distance as averaged over (idealized) conformations of all N -carbon alkanes; and for asymptotic results for this mean as well as that for the mean graph distance as averaged over N -site labeled alkane trees. Beyond this, numerical fits for the asymptotic forms (with presumably exact exponents) have been systematically explored and tested. Thence for the five different extent measures here studied the asymptotic expansions are more firmly and accurately established. The graph-size-extension exponent α is argued to display a degree of universality, taking the value $\alpha = 1/2$ for each of our size-extension measures. This leads to average Wiener numbers scaling as $\sim N^{5/2}$ which (as noted by a referee) is a behavior intermediate between “area-like” ($\sim N^2$) and “volume-like” ($\sim N^3$) behaviors (each of which is intimated to have been previously surmised). Notably also the comparison (studied by Dobrynin and Gutman⁶) between $\langle d_{ij} \rangle_N$ and $\langle d_{ij} \rangle_{\text{lab},N}$ seems from Table 4 to be the most precise of all the comparisons made (i.e., the error estimates for this comparison are smaller than for others). The general results are utilized (in section 9) in an illuminating argument as regards behavior when “volume exclusion” is taken into account, and estimates for the graph-theoretic size-extension exponents for self-avoiding branched conformations in two- and three-dimensional Euclidean space are obtained. Presumably a similar difference also applies for exponents of size-extension exponents for self-avoiding branched structural isomers.

Also the present results again (following ref 5, 7, and 25) indicate a powerful alternative to the computation of isomer-class averages by way of brute-force generation of individual structures.

Indeed the techniques can be viewed as natural extensions of methods often used in statistical mechanics.

ACKNOWLEDGMENT

The authors acknowledge support from the Welch Foundation of Houston, TX.

REFERENCES AND NOTES

- (1) 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.
- (2) Polansky, O. E.; Bonchev, D. The Wiener Number of Graphs. I. General Theory Changes and Changes due to some Graph Operations. *Commun. Math. Chem. (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. Gutman, I.; Yeh, Y.; Lee, Y.-N.; Lau, Y.-L. Some recent results in the theory of the Wiener number. *Indian J. Chem.* **1993**, *32A*, 651–661. Nikolić, S.; Trinajstić, N.; Mihalić, Z. The Wiener Index: Development and Applications. *Croat. Chem. Acta* **1995**, *68*, 105–129.
- (3) *Commun. Math. Chem. (MatCh)* **35**, **1997**, Gutman, I., Klavzar, S., Mohar, B., Guest Eds.

- (4) Lepović, M.; Gutman, I. A Collective Property of Trees and Chemical Trees. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 823–826.
- (5) Bytautas, L.; Klein, D. J. Chemical Combinatorics for Alkane-Isomer Enumeration and More. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1063–1078.
- (6) Dobrynin, A. A.; Gutman, I. The Average Wiener Index of Trees and Chemical Trees. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 679–683.
- (7) L. Bytautas, D. J. Klein, Alkane Isomer Combinatorics: Stereo-Structure Enumeration, Graph-Invariant, Molecular-Property Distributions. *J. Chem. Inf. Comput. Sci.* **39**, **1999**, 803–818.
- (8) Meir A.; Moon, J. W. The Distance Between Points in Random Trees. *J. Comb. Theory* **1970**, *8*, 99–103.
- (9) Zimm, H. B.; 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. London, A* **1962**, *268*, 240–259. Dobson, G. R.; Gordon, M. Theory of Branching Processes and Statistics of Rubber Elasticity. *J. Chem. Phys.* **1965**, *43*, 705–713. 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.
- (10) Klein, D. J.; Seitz, W. A.; Kilpatrick, J. E. Branched polymer models. *J. Appl. Phys.* **1982**, *53*, 6599–6603.
- (11) 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.
- (12) Flory, P. J. *Statistics of Chain Molecules*; Cornell University Press: Ithaca, New York, 1969.
- (13) Trinajstić, N.; Nikolić, S.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Computational Chemical Graph Theory*; Horwood, New York, 1991.
- (14) Gutman, I.; Yeh, Y.-N.; Lee, S.-L.; Chen, J.-C. Wiener Numbers of Dendrimers. *Commun. Math. Chem. (MatCh)* **1994**, *30*, 103–115.
- (15) See, e.g.: deGennes, P. G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
- (16) Barber, M.; Ninham, B. *Random and Restricted Walks*; Gordon Breach, New York, 1970.
- (17) Weiss, G. H.; West, B. J. Random Walks: Theory and Selected Applications. *Adv. Chem. Phys.* **1983**, *52*, 363–505.
- (18) Chandrasekhar, S. Stochastic Problems in Physics and Astronomy. *Rev. Mod. Phys.* **1942**, *15*, 1–89.
- (19) Flory, P. J. The Configuration of Real Polymer Chains. *J. Chem. Phys.* **1949**, *17*, 303–310. Fisher, M. E. *J. Phys. Soc. Jpn.* **1969**, *26S*, 44–45.
- (20) Isaacson, J.; Lubensky, T. C. Flory exponents for generalized polymer problems. *J. Phys.* **1980**, *41*, L469–L471.
- (21) Lubensky, T. C.; Isaacson, J. Field theory and polymer size distribution for branched polymers. *J. Phys.* **1981**, *42*, 175–188. Parisi, G.; Sourlas, N. Critical behavior of branched polymers and the Lee-Yang edge singularity. *Phys. Rev. Lett.* **1981**, *46*, 871–874.
- (22) Seitz, W. A.; Klein, D. J. Excluded volume effects for branched polymers: Monte Carlo results. *J. Chem. Phys.* **1981**, *75*, 5190–5193.
- (23) Meirovitch, H. A new simulation of branched polymers. *J. Phys. A* **1987**, *20*, 6059–6073.
- (24) Hara, T.; Slade, G.; Sokal, A. D. New Lower Bounds on the Self-Avoiding-Walk Connective Constant. *J. Stat. Phys.* **1993**, *72*, 47–517. Conway, A. R.; Enting, I. G.; Guttmann, A. J. Algebraic techniques for enumerating self-avoiding walks on the square lattice. *J. Phys. A* **1993**, *26*, 1519–1534. Conway, A. R.; Guttmann, A. J. Square Lattice Self-Avoiding Walks and Corrections to Scaling. *Phys. Rev. Lett.* **1996**, *77*, 5284–5287.
- (25) Neinhuis, B. Exact critical point and critical exponents of $O(n)$ models in two dimensions. *Phys. Rev. Lett.* **1982**, *49*, 1062. de Forcrand, Ph.; Koukiou, F.; Petritis, D. Self-Avoiding Random Walks on the Hexagonal Lattice. *J. Stat. Phys.* **1986**, *45*, 459–470.
- (26) Bytautas, L.; Klein, D. J. Isomer combinatorics for acyclic conjugated polyenes: enumeration and beyond. *Theor. Chem. Acc.* **1999**, *101*, 371–387. Bytautas, L.; Klein, D. J. Formula Periodic Table for the Isomer Classes of Acyclic Hydrocarbons—Enumerative and Asymptotic Characteristics. *Croat. Chem. Acta* **2000**, in press. Bytautas, L.; Klein, D. J. Formula Periodic Table for Acyclic Hydrocarbon Isomer Classes: Combinatorially Averaged Graph Invariants. *Phys. Chem.—Chem. Phys.* **1999**, *1*, 5565–5572. Bytautas, L.; Klein, D. J.; Schmalz, T. G. All Acyclic Hydrocarbons: Formula Periodic Table and Average Properties via Chemical Combinatorics. *New J. Chem.* **2000**, in press.

CI990127B