



Molecular graphs and the inverse Wiener index problem

Stephan G. Wagner^a, Hua Wang^b, Gang Yu^{b,*}

^a Department of Mathematics, Graz University of Technology, Steyrergasse 30, A-8010 Graz, Austria

^b Department of Mathematics, University of South Carolina, Columbia, SC 29208, United States

ARTICLE INFO

Article history:

Received 7 February 2005

Received in revised form 28 March 2008

Accepted 3 June 2008

Available online 29 July 2008

Keywords:

Wiener index

Sum of squares

Hardy–Littlewood method

ABSTRACT

In the drug design process, one wants to construct chemical compounds with certain properties. In order to establish the mathematical basis for connections between molecular structures and physicochemical properties of chemical compounds, some so-called structure-descriptors or “topological indices” have been put forward. Among them, the Wiener index is one of the most important. A long standing conjecture on the Wiener index [I. Gutman, Y. Yeh, The sum of all distances in bipartite graphs, *Math. Slovaca* 45 (1995) 327–334; M. Lepović, I. Gutman, A collective property of trees and chemical trees, *J. Chem. Inf. Comput. Sci.* 38 (1998) 823–826] states that for any positive integer n (except numbers from a given 49 element set), one can find a tree with Wiener index n . We proved this conjecture in [S. Wagner, A class of trees and its Wiener index, *Acta Appl. Math.* 91 (2) (2006) 119–132; H. Wang, G. Yu, All but 49 numbers are Wiener indices of trees, *Acta Appl. Math.* 92 (1) (2006) 15–20]. However, more realistic molecular graphs are trees with degree ≤ 3 and the so-called hexagon type graphs. In this paper, we prove that every sufficiently large integer n is the Wiener index of some caterpillar tree with degree ≤ 3 , and every sufficiently large even integer is the Wiener index of some hexagon type graph.

Published by Elsevier B.V.

1. Introduction

The structure of a chemical compound is usually modeled as a polygonal shape, which is often called the *molecular graph* of this compound. It has been found that many properties of a chemical compound are closely related to some topological indices of its molecular graph. Among these topological indices, the *Wiener index* is probably the most important one.

The *Wiener index* is a distance-based graph invariant, used as one of the structure descriptors for predicting physicochemical properties of organic compounds (often those significant for pharmacology, agriculture, environment-protection, etc.). The Wiener index was introduced by the chemist H. Wiener [13] about 60 years ago to demonstrate correlations between physicochemical properties of organic compounds and the topological structure of their molecular graphs. This concept has been one of the most widely used descriptors in relating a chemical compound's property to its molecular graph. Therefore, in order to construct a compound with a certain property, one may want to build some structure that has the corresponding Wiener index.

The biochemical community has been using the Wiener index to correlate a compound's molecular graph with experimentally gathered data regarding the compound's characteristics. In the drug design process, one wants to construct chemical compounds with certain properties. The basic idea is to construct chemical compounds from the most common molecules so that the resulting compound has the expected Wiener index. For example, larger aromatic compounds can be made from fused benzene rings as follows (Fig. 1):

* Corresponding address: Department of Mathematical Sciences, Kent State University, Kent, OH 44242, United States. Tel.: +1 512 471 3322x0587.

E-mail addresses: wagner@finanz.math.tu-graz.ac.at (S.G. Wagner), hwang0@math.sc.edu (H. Wang), yu@math.kent.edu (G. Yu).



Fig. 1. Larger aromatic compounds can be made from fused benzene rings.

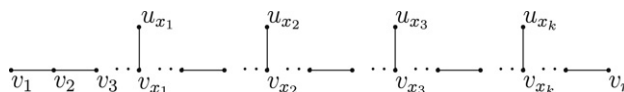


Fig. 2. Caterpillar tree with degree ≤ 3 .

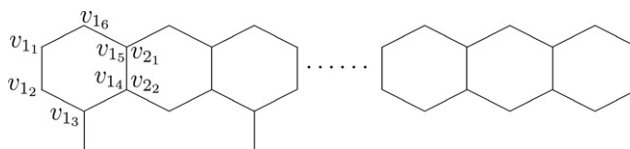


Fig. 3. The hexagon type graph.

Compounds with different structures (and different Wiener indices), even with the same chemical formula, can have different properties. For example, cocaine and scopolamine, both with chemical formula $C_{17}H_{21}NO_4$, have different properties and different Wiener indices. Hence it is indeed important to study the structure (especially the Wiener index) of the molecular graph besides the chemical formula.

From the close relation between the Wiener index and the chemical properties of a compound, it arises the important *inverse Wiener index problem* [5, 1]: Given a positive integer n , can we find a structure (graph) with Wiener index n ?

Goldman et al. [3] solved the inverse Wiener index problem for general graphs: they showed that for every positive integer n there exists a graph G such that the Wiener index of G is n .

Since the majority of the chemical applications of the Wiener index deal with chemical compounds that have acyclic organic molecules, whose molecular graphs are trees, the inverse Wiener index problem for trees attracts more attention and, actually, most of the prior work on Wiener indices deals with trees [2]. When the graph is restricted to trees, the problem is more complicated. Gutman and Yeh [5] conjectured that, for all but a finite set of integers n , one can find a tree with Wiener index n .

Lepović and Gutman [8] checked the integers up to 1206 and found that the following numbers are not Wiener indices of any trees:

2, 3, 5, 6, 7, 8, 11, 12, 13, 14, 15, 17, 19, 21, 22, 23, 24, 26, 27, 30, 33, 34, 37, 38, 39, 41, 43, 45, 47, 51, 53, 55, 60, 61, 69, 73, 77, 78, 83, 85, 87, 89, 91, 99, 101, 106, 113, 147, 159.

They claimed that those listed were the only “forbidden” integers and posed the following conjecture.

Conjecture 1.1. *There are exactly 49 positive integers that are not Wiener indices of trees, namely the numbers listed above.*

A recent computational experiment by Ban, Baspamyatnikh and Mustafa [1] shows that every integer $n \in [10^3, 10^8]$ is the Wiener index of some caterpillar tree. Thus, the conjecture is proved if one is able to show that every integer greater than 10^8 is the Wiener index of a tree.

In [11, 12], we proved that every integer $n > 10^8$ is the Wiener index of some tree. Combined with Ban, Baspamyatnikh and Mustafa’s results, we proved that **Conjecture 1.1** is indeed true.

However, the molecular graphs of most practical interest have natural restrictions on their degrees corresponding to the valences of the atoms and are typically trees or have hexagonal or pentagonal cycles [9, 4].

In this paper, we study the inverse Wiener index problem for the following two kinds of structures:

- (1) trees with degree ≤ 3 (Fig. 2);
- (2) hexagon type graphs (Fig. 3).

We define a family of trees $T = T(n, x_1, x_2, \dots, x_k)$, where

$$V = \{v_1, \dots, v_n\} \cup \{u_{x_1}, \dots, u_{x_k}\},$$

$$E = \{(v_i, v_{i+1}), 1 \leq i \leq n-1\} \cup \{(v_{x_i}, u_{x_i}), 1 \leq i \leq k\},$$

where n and x_i , $1 \leq i \leq k$, are integers such that $1 \leq x_1 \leq \dots \leq x_k \leq n$ (Fig. 2).

We also define a family of hexagon type graphs $G = G(n, x_1, x_2, \dots, x_k)$, where we have n adjacent hexagons $v_{i_1}v_{i_2} \dots v_{i_6}$, for $i = 1, 2, \dots, n$. The edges $v_{i_4}v_{i_5}$, $v_{(i+1)_2}v_{(i+1)_1}$ are identified for $i = 1, 2, \dots, n-1$. On the x_j th hexagon there is a pendant edge incident to v_{j_3} , for $j = 1, \dots, k$ (Fig. 3).

Another popular structure involves pentagons. We note that our proofs can be easily modified to solve the inverse Wiener index problem in that case. For the two kinds of graphs (Figs. 2 and 3) to be considered, we shall prove the following results:

Theorem 1.1. Every sufficiently large integer n is the Wiener index of a caterpillar tree with degree ≤ 3 .

Theorem 1.2. Every sufficiently large integer n is the Wiener index of a hexagon type graph.

Remark 1.3. Even though our proofs are not algorithmic, they can be turned into algorithms by merely checking all the possible cases. Unfortunately, their complexity is quite high; the running time for finding a graph from our graph classes with given Wiener index W is pseudo-polynomial in W .

Notation: In the proofs of Theorem 1.2 and Lemma A.1 (in the Appendix), we shall adopt the standard notation \ll and O . For a complex number A and a positive real number B , $A \ll B$ or $A = O(B)$ means that there is a positive absolute constant c such that $|A| \leq cB$.

2. Preliminaries

For a graph $T = (V, E)$, denote by $d(v_i, v_j)$ the length of the shortest path between two distinct vertices $v_i, v_j \in V$. Define $d_T(v) = \sum_{u \in V} d(v, u)$. The Wiener index $W(T)$ is then defined as

$$W(T) = \frac{1}{2} \sum_{v \in V} d_T(v).$$

For $T = T(n, x_1, x_2, \dots, x_k)$, as shown in Fig. 2, we have

$$\begin{aligned} W(T) &= \sum_{1 \leq i < j \leq n} d(v_i, v_j) + \sum_{i=1}^n \sum_{j=1}^k d(v_i, u_{x_j}) + \sum_{1 \leq i < j \leq k} d(u_{x_i}, u_{x_j}) \\ &= \frac{n^3 - n}{6} + \sum_{i=1}^n \sum_{j=1}^k (1 + |x_j - i|) + \sum_{1 \leq i < j \leq k} (2 + x_j - x_i), \end{aligned}$$

which can be rewritten as

$$\frac{n^3}{6} + \frac{kn^2}{4} + \frac{(6k-1)n}{6} - \frac{k^3 - 12k^2 + 14k}{12} + \sum_{j=1}^k \left(x_j + j - 1 - \frac{k+n}{2} \right)^2 \quad (1)$$

after some elementary simplification steps.

For $G = G(n, x_1, x_2, \dots, x_k)$ as shown in Fig. 3, we have

$$W(G) = \frac{16n^3 + 36n^2 + 26n + 3}{3} + \sum_{1 \leq i < j \leq k} (2 + 2(x_j - x_i)) + \sum_{i=1}^k (4n^2 + 8x_i^2 - 8nx_i + 12n - 8x_i + 7). \quad (2)$$

We note that, from (2), $W(G)$ and k have different parity. Due to this (somewhat annoying) phenomenon, the Wiener indices of our hexagon type graphs with a fixed number of “leaves” comprise at most half of positive integers. To show that every large integer is the Wiener index of such a graph, one should consider at least two different k , with different parities.

Expanding the last sum in (2) and collecting terms, we see that $W(G)$ is equal to

$$\frac{16n^3 + 36n^2 + 26n + 3}{3} + k(4n^2 + 12n + k + 6) + \sum_{i=1}^k (8x_i^2 - (8n + 2k - 4i + 10)x_i).$$

Completing squares is not necessary for our proof of Theorem 1.2, but it may make the expression look better. By doing so, we have

$$W(G) = \frac{16n^3 + 36n^2 + 26n + 3}{3} + k \left(2n^2 + 8n + k + 4 - \frac{k^2 - 1}{24} \right) + \frac{1}{8} \sum_{i=1}^k (8x_i - 4n - k - 5 + 2i)^2. \quad (3)$$

3. Proof of Theorem 1.1

We will use formula (1) with some special k and show that all sufficiently large integers can be written as $W(T(n, x_1, \dots, x_k))$. Due to the restriction $x_i \neq x_j$ (for $1 \leq i < j \leq k$), the well-known Four Square Theorem (with $k = 4$) does not directly yield what we want. We thus need to take some larger k , and we find that $k = 8$ is good enough for our purpose. Taking $k = 8$ and $n = 2s$, we can rewrite (1) as

$$W(T(n, x_1, \dots, x_8)) = \frac{4s^3}{3} + 8s^2 + \frac{47s}{3} + 12 + \sum_{j=1}^8 y_j^2 \quad (4)$$

where $y_j := x_j + j - s - 5$ with restrictions

$$-(s + 3) \leq y_1 < y_2 < \dots < y_8 \leq s + 3.$$

Of course, we note that $y_{j+1} - y_j \geq 2$ because $x_i \neq x_j$ if $i \neq j$. Now we need the following lemma, which is a slight modification of Lagrange's famous four-square theorem:

Lemma 3.1. *Let $N > 103$ and $4 \nmid N$. Then N can be written as $a_1^2 + a_2^2 + a_3^2 + a_4^2$ with nonnegative integers $a_1 < a_2 < a_3 < a_4$ and $a_2 \geq 2$.*

Proof. It is well known (see [6, Theorem 386]) that the number of representations of a positive integer N as the sum of 4 squares (representations which differ only in order or sign counting as different) is

$$r_4(N) = 8 \sum_{\substack{d|N \\ 4 \nmid d}} d,$$

while the number of representations of N as the sum of 2 squares is

$$r_2(N) = 4 \prod_{\substack{p^r || N \\ p \equiv 1 \pmod{4}}} (r + 1).$$

if every prime factor $q \equiv 3 \pmod{4}$ appears with an even power in the factorization of N (and 0 otherwise). The representations violating the first condition correspond to representations of the form $2a^2 + b^2 + c^2$. For each fixed $a \geq 0$ and each representation $b^2 + c^2$ of $N - 2a^2$, we have at most 24 representations of N as a sum of 4 squares (six possible choices for the positions of the two a 's, and two additional choices of sign).

The representations violating the second condition correspond to representations of the form $1 + a^2 + b^2$. For each representation $a^2 + b^2$ of $N - 1$, this gives us at most 24 representations of N as a sum of 4 squares (twelve possible choices for the positions of 0 and 1, and one additional choice of sign).

So the number of representations violating any of the conditions is at most

$$24 \sum_{a \leq \sqrt{N/2}} r_2(N - 2a^2) + 24r_2(N - 1).$$

Now, for a non-negative integer r , by induction we have

$$r + 1 \leq \frac{3}{\sqrt{5}} \cdot 5^{r/4}, \quad r + 1 \leq \frac{2}{\sqrt[4]{13}} \cdot 13^{r/4}, \quad \text{and} \quad r + 1 \leq (1 + 1)^r \leq p^{r/4} \quad \text{if } p \geq 17.$$

Thus

$$r_2(n) = 4 \prod_{\substack{p^r || n \\ p \equiv 1 \pmod{4}}} (r + 1) \leq 4 \cdot \frac{3}{\sqrt{5}} \cdot \frac{2}{\sqrt[4]{13}} \prod_{\substack{p^r || n \\ p \equiv 1 \pmod{4}}} p^{r/4} \leq 4 \cdot \frac{3}{\sqrt{5}} \cdot \frac{2}{\sqrt[4]{13}} \cdot n^{1/4}.$$

Therefore, if $4 \nmid N$ and $N \geq 28561 = 13^4$, the number of representations violating one of the conditions is at most

$$24 \left(\sqrt{N/2} + 2 \right) \cdot \frac{24}{\sqrt[4]{325}} N^{1/4} \leq 96\sqrt{2} \left(\sqrt{N/2} + 2 \right) N^{1/4} < 104N^{3/4} < 8N < r_4(N).$$

So there must be some representation not violating any of the conditions. This proves the lemma for $N > 28560$, but it turns out that it also holds true for $N \in [104, 28560]$ by explicit testing. \square

Remark 3.2. The condition $4 \nmid N$ may not be skipped – for example, 4^k cannot be represented as a sum of four squares without violating the conditions.

Corollary 3.3. *If $4 \nmid N$, $N > 103$, one can always find integers z_1, z_2, z_3, z_4 such that $N = z_1^2 + \dots + z_4^2$, $z_1 < \dots < z_4$ and no two of the z_i are consecutive.*

Let $a_1 < a_2 < a_3 < a_4$ satisfy the conditions of the lemma. Choose $z_1 = -a_3, z_2 = -a_1, z_3 = a_2$ and $z_4 = a_4$. Then,

$$z_1 < -a_2 < z_2 < 1 < z_3 < a_3 < z_4,$$

which already proves the claim. \square

Remark 3.4. Obviously, $z_4 \leq \lfloor \sqrt{n} \rfloor$ and $|z_1| \leq \lfloor \sqrt{N} \rfloor - 1$.

Proposition 3.5. *Let $K \geq 15$. Then any integer N in the interval $[4K^2 - 8K + 112, 5K^2 - 16K + 21]$ can be written as $y_1^2 + \dots + y_8^2$, where the y_i are integers satisfying*

$$-K \leq y_1 < y_2 < \dots < y_8 \leq K$$

and no two of them are consecutive.

Proof. Take $y_1 = -K, y_7 = K - 2, y_8 = K$ and either $y_2 = -K + 2$ or $y_2 = -K + 3$. By the corollary and the subsequent remark, any integer $M \in [104, (K - 3)^2 - 1], 4 \nmid M$, can be written as $y_3^2 + \dots + y_6^2$, where

$$-K = y_1 < y_2 < -K + 4 < y_3 < y_4 < y_5 < y_6 < K - 3 < y_7 < y_8 = K$$

(no two of them being consecutive). Now

$$(-K)^2 + (-K + 2)^2 + (K - 2)^2 + K^2 = 4K^2 - 8K + 8 \equiv 0 \pmod{4}$$

and

$$(-K)^2 + (-K + 3)^2 + (K - 2)^2 + K^2 = 4K^2 - 10K + 13 \equiv 2K + 1 \pmod{4}.$$

So all integers $\not\equiv 0 \pmod{4}$ in the interval $[4K^2 - 8K + 112, 5K^2 - 14K + 16]$ and all integers $\not\equiv 2K + 1 \pmod{4}$ in the interval $[4K^2 - 10K + 117, 5K^2 - 16K + 21]$ can be written in the required way. Since $0 \not\equiv 2K + 1 \pmod{4}$, this means that in fact all integers in the interval $[4K^2 - 8K + 112, 5K^2 - 16K + 21]$ can be written in the required way, which proves the claim. \square

Theorem 3.6. *All integers ≥ 3856 are Wiener indices of trees of the form $T(n, x_1, \dots, x_8)$ ($x_1 < x_2 < \dots < x_8$) and thus Wiener indices of chemical trees.*

Proof. By the preceding proposition, any integer in the interval $[4K^2 - 8K + 112, 5K^2 - 16K + 21]$ can be written as $y_1^2 + \dots + y_8^2$, where the y_i satisfy our requirements and $-K \leq y_1 < \dots < y_8 \leq K$. If we take the union of these intervals over $21 \leq K \leq s + 3$, we see that in fact any integer in the interval $[1708, 5s^2 + 14s + 18]$ can be written as $y_1^2 + \dots + y_8^2$, where the y_i satisfy our requirements and $-(s + 3) \leq y_1 < \dots < y_8 \leq s + 3$. Short computer calculations show that, for $s \geq 7$, even any integer in the interval $[224, 5s^2 + 14s + 18]$ can always be written that way. But this means that for any $s \geq 7$, all integers in the interval

$$\left[\frac{4s^3}{3} + 8s^2 + \frac{47s}{3} + 236, \frac{4s^3}{3} + 13s^2 + \frac{89s}{3} + 30 \right]$$

are Wiener indices of trees of the form $T(n, x_1, \dots, x_8)$. Taking the union over all these intervals, we see that all integers ≥ 12567 are contained in an interval of that type. By an additional computer search ($n \leq 40$ will do) in the remaining interval, one can get this number down to 3856. \square

Remark 3.7. By checking $k = 4, 5, 6, 7$ and finally all $n \leq 17$, one obtains a list of 250 integers (the largest being 927) that are not Wiener indices of trees of the form $T(n, x_1, \dots, x_k)$ with maximal degree ≤ 3 . Further computer search gives a list of 127 integers that are not Wiener indices of trees with maximal degree ≤ 3 – these are 16, 25, 28, 36, 40, 42, 44, 49, 54, 57, 58, 59, 62, 63, 64, 66, 80, 81, 82, 86, 88, 93, 95, 97, 103, 105, 107, 109, 111, 112, 115, 116, 118, 119, 126, 132, 139, 140, 144, 148, 152, 155, 157, 161, 163, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 199, 227, 239, 251, 255, 257, 259, 263, 267, 269, 271, 273, 275, 279, 281, 283, 287, 289, 291, 405 and the 49 values that cannot be represented as the Wiener index of any tree. This list reduces to the following values if one considers also trees with maximal degree four: 25, 36, 40, 49, 54, 57, 59, 80, 81, 93, 95, 97, 103, 105, 107, 109, 132, 155, 157, 161, 163, 167, 169, 171, 173, 177, 239, 251, 255 and 257.

4. Proof of Theorem 1.2

In order to prove Theorem 1.2 we use the following proposition for which we shall give a proof in the Appendix.

Proposition 4.1. *Suppose $g_i(y) = a_i y^2 + b_i y + c_i$ ($i = 1, \dots, 5$) are quadratic polynomials of integer coefficients, and $a_i > 0$ for $i = 1, \dots, 5$. d_i and D_i ($i = 1, \dots, 5$) are positive constants satisfying*

$$d_i < D_i, i = 1, \dots, 5, \quad \sum_{i=1}^5 a_i d_i^2 < 1 - \epsilon < 1 + \epsilon < \sum_{i=1}^5 a_i D_i^2 \tag{5}$$

for some constant $\epsilon > 0$. Suppose L is a sufficiently large integer. Let \mathcal{H} be the hypothesis that the congruence

$$g_1(y_1) + g_2(y_2) + \dots + g_5(y_5) \equiv L \pmod{p^{\gamma_p}}$$

is solvable for every prime power p^{γ_p} , where $\gamma_p = \max\{\gamma_{p,1}, \dots, \gamma_{p,5}\}$ with

$$\gamma_{p,j} := \begin{cases} \theta_{p,j} + 2 & \text{if } p = 2 \\ \theta_{p,j} + 1 & \text{if } p > 2, \end{cases}$$

and $\theta_{p,j}$ is the highest power of p such that

$$g_j'(x) \equiv 0 \pmod{p^{\theta_{p,j}}}$$

for all values of x .

If the hypothesis \mathcal{H} is satisfied, then the equation

$$g_1(y_1) + g_2(y_2) + \dots + g_5(y_5) = L \tag{6}$$

with $d_i \sqrt{L} < y_i \leq D_i \sqrt{L}$ has at least $cL^{\frac{3}{2}}$ integer solutions, where c is a certain positive constant depending only on a_i 's, d_i 's, and D_i 's, $i = 1, \dots, 5$.

As we have noticed that N and k have opposite parities, we have to prove the theorem separately in two cases subject to the parity of N . Nevertheless, since the proofs for odd N and even N are almost identical, we shall give a proof of the theorem for odd N only, and a proof for even large N follows the same way. Similar to the proof of Theorem 1.1, we need more variables than expected to guarantee that the side conditions $x_i \neq x_j$ (for $i \neq j \leq k$) are satisfied. For large odd N , we take $k = 10$ which is large enough for our purpose. (For even N , one can see that, with the same argument we shall carry out for odd N , it suffices to take $k = 9$ or any larger fixed odd integer.)

Suppose N is a sufficiently large odd integer. Let $k = 10$, then from (3) we have

$$W(G) = \frac{16}{3}n^3 + 32n^2 + \frac{266}{3}n + \frac{399}{4} + \frac{1}{8} \sum_{i=1}^{10} (8x_i - 4n - 15 + 2i)^2.$$

We thus want to show that

$$N = \frac{16}{3}n^3 + 32n^2 + \frac{266}{3}n + \frac{399}{4} + \frac{1}{8} \sum_{i=1}^{10} (8x_i - 4n - 15 + 2i)^2$$

for certain integers x_i , $i = 1, 2, \dots, 10$ satisfying

$$1 \leq x_1 < x_2 < \dots < x_9 < x_{10} \leq n. \tag{7}$$

Let

$$f(x) = \frac{16}{3}x^3 + 32x^2 + \frac{266}{3}x + \frac{399}{4},$$

and $\alpha(N)$ be the positive real root of $f(x) - N + N^{\frac{1}{3}} = 0$. It is quite easy to see that $\alpha(N) = \left(\frac{3}{16}N\right)^{\frac{1}{3}} - 2 + O(N^{-\frac{1}{3}})$.

Let $n = \lfloor \alpha(N) \rfloor$. Then we have $n = \left(\frac{3}{16}N\right)^{\frac{1}{3}} + O(1)$, and thus $n < N^{\frac{1}{3}} < 2n$. Also, we have

$$0 \leq N - f(n) - N^{\frac{1}{3}} < f(n+1) - f(n) = 16n^2 + 80n + 126. \tag{8}$$

Since $8f(n) \equiv -2 \pmod{16}$, in order to obtain the statement for large odd N , it is sufficient to demonstrate that, for every integer M satisfying

$$8n \leq M \leq 8(16n^2 + 82n + 126) \quad \text{and} \quad M \equiv -6 \pmod{16}, \tag{9}$$

we have

$$M = \sum_{i=1}^{10} (8x_i - 4n - 15 + 2i)^2$$

for some x_i ($i = 1, 2, \dots, 10$) satisfying (7).

Let $K = \lfloor \sqrt{M}/24 \rfloor$, and

$$x_i = \lfloor n/2 \rfloor + K + i, \quad i = 6, \dots, 10. \quad (10)$$

Since $K \leq \sqrt{8(16n^2 + 82n + 126)}/24 < \frac{12}{25}n$, we have

$$n/2 + \sqrt{M}/24 < x_6 < x_7 < x_8 < x_9 < x_{10} \leq n. \quad (11)$$

It is very easy to check that

$$\sum_{i=6}^{10} (8x_i - 4n - 15 + 2i)^2 \equiv 8n - 3 \pmod{16} \quad (12)$$

and, noticing that M is sufficiently large,

$$\frac{5}{9}M < \sum_{i=6}^{10} (8x_i - 4n - 15 + 2i)^2 < \frac{3}{5}M. \quad (13)$$

From (12), (13) and (9), we see that it is sufficient to show that

$$\sum_{i=1}^5 (8x_i - 4n - 15 + 2i)^2 = L \quad (14)$$

for an integer L satisfying

$$\frac{2}{5}M \leq L \leq \frac{4}{9}M \quad \text{and} \quad L \equiv 8n - 3 \pmod{16} \quad (15)$$

with

$$1 \leq x_1 < x_2 < x_3 < x_4 < x_5 \leq n/2 + \sqrt{M}/24. \quad (16)$$

With the aid of Proposition 4.1, we shall show that there exists some integer solution to (14) subject to conditions (15), (16). Let

$$g_i(y) = (8y - 4(n - 2\lfloor n/2 \rfloor) - 15 + 2i)^2, \quad i = 1, \dots, 5.$$

It is clear that $y_p = 1$ for every prime $p \geq 3$. Note that each $\{g_j(y) \pmod{p} : y = 0, \dots, p-1\}$ contains $\frac{p+1}{2}$ residue classes modulo p . Thus, from the Davenport–Chowla Theorem (cf. [10], Lemma 2.14), $\{g_1(y_1) + g_2(y_2) \pmod{p} : y_1, y_2 = 0, \dots, p-1\}$ covers all residue classes modulo p . Thus, for every prime $p \geq 3$,

$$g_1(y_1) + g_2(y_2) + \dots + g_5(y_5) \equiv L \pmod{p}$$

is solvable.

For $p = 2$, we note that $\theta_{2,j} = 4$ is the largest integer such that

$$2^4 \mid g_i'(y) \quad \text{for all } y.$$

So, to show that congruence condition for $p = 2$ holds, it thus suffices to show that

$$g_1(y_1) + g_2(y_2) + \dots + g_5(y_5) \equiv L \pmod{2^6} \quad (17)$$

is solvable. Expanding the left-hand side of (17), we see that

$$\sum_{i=1}^5 g_i(y_i) \equiv 16 \left(\sum_{i=1}^5 (-1)^i y_i + (n - 2\lfloor n/2 \rfloor + 1)^2 \right) + 8(n - 2\lfloor n/2 \rfloor) + 45 \pmod{64}.$$

It is then easy to check that (17) has a non-trivial solution

$$y_1 = 0, \quad y_2 = y_3 = 1, \quad y_4 = \frac{L - 8(n - 2\lfloor n/2 \rfloor) + 19}{16}, \quad y_5 = (n - 2\lfloor n/2 \rfloor + 1)^2.$$

Therefore, the hypothesis \mathcal{H} is satisfied. Now, let

$$d_i = \frac{1}{18} + \frac{3^i}{4 \cdot 10^5}, \quad D_i = \frac{1}{18} + \frac{3^i}{2 \cdot 10^5}, \quad i = 1, \dots, 5. \tag{18}$$

Then we have

$$\sum_{i=1}^5 (8d_i)^2 = 0.994134 \dots < 1, \quad \sum_{i=1}^5 (8D_i)^2 = 1.000667 \dots > 1.$$

Now all conditions required by Proposition 4.1 are satisfied, thus, for the integer L satisfying (15), the Eq. (6) has solutions with $d_i\sqrt{L} < y_i \leq D_i\sqrt{L}, i = 1, \dots, 5$. Let $x_i = \lfloor n/2 \rfloor + y_i (i = 1, \dots, 5)$, and note that

$$d_i < D_i, \quad i = 1, \dots, 5, \quad \text{and} \quad D_i + 10^{-6} < d_{i+1} \quad i = 1, \dots, 4,$$

Proposition 4.1 guarantees a solution for (14) with

$$\lfloor n/2 \rfloor < x_1 < x_2 < x_3 < x_4 < x_5 \leq \lfloor n/2 \rfloor + D_5\sqrt{L} < n/2 + \sqrt{M}/24.$$

Theorem 1.2 thus follows. \square

Acknowledgement

The authors wish to thank Professor Ivan Gutman, who let the coauthoring groups know of each other and thus made this cooperation possible.

The authors also want to thank Professor László Székely for his suggestions in writing this paper.

The first author’s research was supported in part by Austrian Science Fund project no. S-8307-MAT. The second author’s research was supported in part by the NSF contracts Nr. 0072187 and 0302307. This research of the third author was partially supported by NSF grant DMS0601033.

Appendix

The result of Proposition 4.1 is probably the most trivial case of representing large integers as a sum of polynomials of integer coefficients. Surprisingly, this result, though seemingly well known to the experts in Hardy–Littlewood method, is not found in literature. We thus give a proof here. The argument is very similar to [7]. While the main result in [7] is on sums of polynomials of higher degrees, most lemmas are given in a general setting. For a shorter proof, we shall make use of various results in [7], though some of them are not necessarily the best for quadratic polynomials.

Throughout, ϵ is any sufficiently small positive number, not necessarily the same at all places. As usual, $e(t) = \exp(2\pi it)$ for real number t , and (x, y) denotes the greatest common divisor of integers x, y .

For $j = 1, 2, \dots, 5$, let

$$f_j(\alpha) := \sum_{d_j\sqrt{L} < y \leq D_j\sqrt{L}} e(g_j(y)\alpha).$$

Let $R(L)$ be the number of solutions of Eq. (6) subject to the given conditions. For an integer n , and any real number δ , we have

$$\int_{\delta}^{1+\delta} e(n\alpha) d\alpha = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0, \end{cases} \tag{19}$$

which yields that

$$R(L) = \int_{\delta}^{1+\delta} f_1(\alpha)f_2(\alpha) \cdots f_5(\alpha)e(-L\alpha) d\alpha \tag{20}$$

for any $\delta \in \mathbb{R}$.

As in [7], we let $Q = L^{\frac{1}{3}}$, and $\delta = \frac{Q}{L}$. For integers a, q satisfying $1 \leq a \leq q \leq Q$ and $(a, q) = 1$, let $\mathfrak{M}(q, a)$ be the set of real numbers α satisfying $|q\alpha - a| < \delta$. Since $\mathfrak{M}(q, a) = \left(\frac{a-\delta}{q}, \frac{a+\delta}{q}\right)$, it follows that $\mathfrak{M}(q, a) \subset [\delta, 1 + \delta]$ and $\mathfrak{M}(q, x) \cap \mathfrak{M}(q, y) = \emptyset$ if $x \neq y$. We define the *major arcs* \mathfrak{M} as the union of such $\mathfrak{M}(q, a)$ (with $1 \leq a \leq q \leq Q$ and $(a, q) = 1$), and the *minor arcs* \mathfrak{m} as $\mathfrak{m} = [\delta, 1 + \delta] \setminus \mathfrak{M}$. Then we have

$$R(L) = R_{\mathfrak{M}}(L) + R_{\mathfrak{m}}(L) \tag{21}$$

with the two parts respectively corresponding to the integral of the integrand in (20) on \mathfrak{M} and \mathfrak{m} .

Upper bound of $R_{\mathfrak{m}}(L)$. We first state two lemmas that will be needed in the estimation of $R_{\mathfrak{m}}(L)$.

Lemma A.1 (Weyl's Inequality). Suppose $k \geq 2$ is an integer, a, q are integers satisfying $1 \leq a \leq q$, and $(a, q) = 1$. Let $\phi(x)$ be a polynomial with real coefficients given by

$$\phi(x) = \alpha x^k + \alpha_1 x^{k-1} + \cdots + \alpha_{k-1} x + \alpha_k, \quad \text{where } \left| \alpha - \frac{a}{q} \right| \leq q^{-2}.$$

Then for sufficiently large integer X ,

$$\sum_{x=1}^X e(\phi(x)) \ll X^{1+\epsilon} (q^{-1} + qX^{-k})^{2^{1-k}}. \quad (22)$$

Proof. This is Lemma 2.4 of [10]. \square

Lemma A.2. Suppose $g(x) = ax^2 + bx + c$ is a fixed quadratic polynomial with integer coefficients. For a positive integer X , let

$$f(\alpha) := \sum_{x=1}^X e(g(x)\alpha).$$

Then we have

$$\int_0^1 |f(\alpha)|^4 d\alpha \ll X^{2+\epsilon}.$$

Proof. This is a special case of the famous Hua's Lemma. Note that the integral equals the number of solutions of the equation

$$g(x_1) - g(x_2) = g(x_3) - g(x_4), \quad 1 \leq x_j \leq X \quad (j = 1, 2, 3, 4).$$

Multiplying $4a$ to the equation, we get $y_1^2 - y_2^2 = y_3^2 - y_4^2$, where $y_j = 2|a|x_j + \frac{a}{|a|} \cdot b \ll X$. The lemma then follows immediately by considering the number of solutions of the new homogeneous equation. \square

We note that when $k = 2$, the right hand side of (22) is then replaced by $q^{\frac{1}{2}+\epsilon} + X^{1+\epsilon} q^{-1}$. Thus, when $\alpha \in \mathfrak{m}$, we have $f_j(\alpha) \ll (L/Q)^{\frac{1}{2}+\epsilon}$ from Lemma A.1. From this and Hölder's inequality, and Lemma A.2, we have

$$\begin{aligned} R_{\mathfrak{m}}(L) &\ll (L/Q)^{\frac{1}{2}+\epsilon} \int_{\mathfrak{m}} |f_2 f_3 f_4 f_5| d\alpha \ll (L/Q)^{\frac{1}{2}+\epsilon} \prod_{j=2}^5 \left(\int_0^1 |f_j(\alpha)|^4 d\alpha \right)^{\frac{1}{4}} \\ &\ll (L/Q)^{\frac{1}{2}+\epsilon} \cdot L^{1+\epsilon} \ll L^{\frac{3}{2}-\epsilon}. \end{aligned} \quad (23)$$

Treatment of $R_{\mathfrak{m}}(L)$. Next we shall approximate $f_j(\alpha)$ on major arcs with some "nicer" functions. Suppose $\alpha = \frac{a}{q} + \beta$, with $1 \leq a \leq q \leq Q$, $(a, q) = 1$, and $|\beta| < \frac{Q}{qL}$. For $j = 1, 2, \dots, 5$, let

$$v_j(\beta) := \int_{a_j \sqrt{L}}^{D_j \sqrt{L}} e(\beta a_j t^2) dt = \frac{1}{2} a_j^{-\frac{1}{2}} \int_{a_j D_j^2 L}^{a_j D_j^2 L} t^{-\frac{1}{2}} e(\beta t) dt,$$

and

$$u_j(\beta) := \frac{1}{2} a_j^{-\frac{1}{2}} \sum_{a_j D_j^2 L < m \leq a_j D_j^2 L} m^{-\frac{1}{2}} e(\beta m),$$

where a_j is the leading coefficient of $g_j(y)$. It should be noted that, by integration by parts,

$$\begin{aligned} u_j(\beta) &= \frac{1}{2} a_j^{-\frac{1}{2}} \int_{a_j D_j^2 L}^{a_j D_j^2 L} t^{-\frac{1}{2}} e(\beta t) d[t] \\ &= v_j(\beta) + O \left(\int_{a_j D_j^2 L}^{a_j D_j^2 L} \left| -\frac{1}{2} t^{-\frac{3}{2}} e(\beta t) + 2\pi i \beta t^{-\frac{1}{2}} e(\beta t) \right| dt \right) \\ &= v_j(\beta) + O \left(L^{-\frac{1}{2}} + L^{\frac{1}{2}} |\beta| \right). \end{aligned} \quad (24)$$

We also let

$$V_j(\alpha) = V_j(\alpha; q, a) = q^{-1} S_j(q, a) u_j(\beta),$$

where $S_j(q, a)$ is the Gaussian sum given by

$$S_j(q, a) = \sum_{x=1}^q e\left(\frac{ag_j(x)}{q}\right)$$

which is well-known to be $O(\sqrt{q})$.

Lemma A.3. Suppose $1 \leq a \leq q \leq Q$, and $(a, q) = 1$. For any $\alpha = \frac{a}{q} + \beta \in \mathfrak{M}(q, a)$, we have

$$f_j(\alpha) = q^{-1}S_j(q, a)v_j(\beta) + O\left(L^{\frac{1}{3}}\right).$$

Proof. This is the special case ($k = 2$) of Lemma 5.4 in [7]. \square

Lemma A.4. Suppose $n, k \geq 2$ are integers. For any real number $\beta \in [-\frac{1}{2}, \frac{1}{2}]$, we have

$$\sum_{m=1}^n \frac{1}{k} m^{k-1} e(\beta m) \ll \min\left(n^{\frac{1}{k}}, |\beta|^{-\frac{1}{k}}\right).$$

Proof. This is Lemma 2.8 in [10]. \square

From (24) and Lemma A.3, we know that for $\alpha = \frac{a}{q} + \beta \in \mathfrak{M}(q, a)$,

$$f_j(\alpha) = V_j(\alpha) + O\left(q^{-1}|S_j(q, a)|(L^{-\frac{1}{2}} + L^{\frac{1}{2}}|\beta|) + L^{\frac{1}{3}}\right) = V_j(\alpha) + O(L^{\frac{1}{3}}), \tag{25}$$

where we have used the fact that $|\beta| < \frac{Q}{qL}$.

This gives

$$R_{\mathfrak{M}}(L) = \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a, q) = 1}} \int_{\mathfrak{M}(q, a)} V_1(\alpha)f_2(\alpha) \cdots f_5(\alpha)e(-\alpha L)d\alpha + E,$$

where

$$\begin{aligned} E &= \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a, q) = 1}} \int_{\mathfrak{M}(q, a)} (f_1 - V_1)(f_2f_3f_4f_5)e(-\alpha L)d\alpha \\ &\ll L^{\frac{1}{3}} \int_0^1 |f_2f_3f_4f_5|d\alpha \ll L^{\frac{1}{3}}L^{1+\epsilon} \ll L^{\frac{3}{2}-\epsilon} \end{aligned}$$

by Hölder’s inequality and Lemma A.2. Continuing the same process, (noticing that $V_j(\alpha) \ll |f_j(\alpha)| + L^{\frac{1}{3}}$), we get

$$R_{\mathfrak{M}}(L) = \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a, q) = 1}} \int_{\mathfrak{M}(q, a)} V_1(\alpha)V_2(\alpha) \cdots V_5(\alpha)e(-\alpha L)d\alpha + O(L^{3/2-\epsilon}). \tag{26}$$

From Lemma A.4, we have, for any $\alpha = \frac{a}{q} + \beta$ with $|\beta| \leq \frac{1}{2}$,

$$V_j(\alpha) \ll q^{-1}|S_j(q, a)| \min\left(L^{\frac{1}{2}}, |\beta|^{-\frac{1}{2}}\right) \ll q^{-\frac{1}{2}} \cdot \frac{L^{\frac{1}{2}}}{(1 + L|\beta|)^{\frac{1}{2}}} \ll (L/q)^{\frac{1}{2}}(1 + L|\beta|)^{-\frac{1}{2}}. \tag{27}$$

Thus, we have

$$\begin{aligned} &\int_{\mathfrak{M}(q, a)} V_1(\alpha) \cdots V_5(\alpha)e(-\alpha L)d\alpha - \int_{\frac{a}{q}-\frac{1}{2}}^{\frac{a}{q}+\frac{1}{2}} V_1(\alpha) \cdots V_5(\alpha)e(-\alpha L)d\alpha \\ &\ll \left(\int_{\frac{Q}{qL}}^{\frac{1}{2}} + \int_{-\frac{1}{2}}^{\frac{Q}{qL}}\right) |V_1(a/q + \beta) \cdots V_5(a/q + \beta)|d\beta \\ &\ll \left(\frac{L}{q}\right)^{\frac{5}{2}} \int_{\frac{Q}{qL}}^{\frac{1}{2}} (1 + L|\beta|)^{-\frac{5}{2}}d\beta \ll \frac{L^{\frac{3}{2}}}{q^{\frac{5}{2}}} \cdot \left(\frac{Q}{q}\right)^{-\frac{3}{2}} \ll \frac{(L/Q)^{\frac{3}{2}}}{q}. \end{aligned}$$

Taking this into (26), we get

$$\begin{aligned}
 R_{\mathfrak{M}}(L) &= \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \int_{\frac{a}{q}-\frac{1}{2}}^{\frac{a}{q}+\frac{1}{2}} V_1(\alpha)V_2(\alpha) \cdots V_5(\alpha)e(-\alpha L)d\alpha + O(L^{3/2-\epsilon}) \\
 &= \mathfrak{S}(L; Q)J(L) + O(L^{\frac{3}{2}-\epsilon}),
 \end{aligned}
 \tag{28}$$

where

$$\mathfrak{S}(L; Q) = \sum_{q \leq Q} q^{-5} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \left(\prod_{j=1}^5 S_j(q, a) \right) e\left(-\frac{aL}{q}\right)$$

is the partial singular series and

$$J(L) = \int_{-\frac{1}{2}}^{\frac{1}{2}} u_1(\beta) \cdots u_5(\beta)e(-\beta L)d\beta$$

is the singular integral. From (19), we see that

$$\begin{aligned}
 J(L) &= \frac{1}{32} (a_1 \cdots a_5)^{-\frac{1}{2}} \sum_{\substack{a_j d_j^2 L < m_j \leq a_j D_j^2 L \\ m_1 + \cdots + m_5 = L}} \frac{1}{\sqrt{m_1 \cdots m_5}} \int_{-\frac{1}{2}}^{\frac{1}{2}} e((m_1 + \cdots + m_5 - L)\beta)d\beta \\
 &= \frac{1}{32} (a_1 \cdots a_5)^{-\frac{1}{2}} \sum_{\substack{a_j d_j^2 L < m_j \leq a_j D_j^2 L \\ m_1 + \cdots + m_5 = L}} \frac{1}{\sqrt{m_1 \cdots m_5}} \\
 &\gg L^{-\frac{5}{2}} \sum_{\substack{a_j d_j^2 L < m_j \leq a_j D_j^2 L \\ m_1 + \cdots + m_5 = L}} 1 \gg L^{\frac{3}{2}},
 \end{aligned}
 \tag{29}$$

where I have used the fact that there are $\gg L^4$ solutions to the linear Diophantine equation $m_1 + \cdots + m_5 = L$ subject to the condition $a_j d_j^2 L < m_j \leq a_j D_j^2 L$ ($j = 1, \dots, 5$) because of (5).

As to the singular series, from the fact that $S_j(q, a) \ll \sqrt{q}$, we have

$$\mathfrak{S}(L; Q) = \mathfrak{S}(L) + O(Q^{-\frac{1}{2}}),$$

where

$$\mathfrak{S}(L) := \sum_{q=1}^{\infty} q^{-5} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \left(\prod_{j=1}^5 S_j(q, a) \right) e\left(-\frac{aL}{q}\right).$$

Lemma A.5. *If the condition \mathcal{H} given in (3) is satisfied, then*

$$\mathfrak{S}(L) > D > 0$$

for some constant D .

Proof. This is the special case ($k = 2$) of Lemma 7.9 in [7]. \square

This, along with (21), (23), (28) and (29), gives the desired result.

References

[1] Y.A. Ban, S. Bspamyatnikh, N.H. Mustafa, On a conjecture on Wiener indices in combinatorial chemistry, in: Proc. of the 9th International Computing and Combinatorics Conference '03, 2003, pp. 509–518 (The journal version will appear in *Algorithmica*, 2004).
 [2] A.A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: Theory and applications, *Acta Appl. Math.* 66 (2001) 211–249.
 [3] D. Goldman, S. Istrail, G. Lancia, A. Piccolboni, B. Walenz, Algorithmic strategies in combinatorial chemistry, in: Proc. 11th ACM-SIAM Sympos. Discrete Algorithms, 2000, pp. 275–284.
 [4] I. Gutman, O.E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986.
 [5] I. Gutman, Y. Yeh, The sum of all distances in bipartite graphs, *Math. Slovaca* 45 (1995) 327–334.
 [6] G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, fifth edition, The Clarendon Press Oxford University Press, New York, 1979.
 [7] L.K. Hua, On a generalized Waring problem, *Proc. London Math. Soc.* (2) 43 (1937) 161–182.
 [8] M. Lepovic, I. Gutman, A collective property of trees and chemical trees, *J. Chem. Inf. Comput. Sci.* 38 (1998) 823–826.
 [9] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1992.
 [10] R.C. Vaughan, *The Hardy–Littlewood Method*, 2nd edition, Cambridge University Press, 1997.
 [11] S. Wagner, A class of trees and its Wiener index, *Acta Appl. Math.* 91 (2) (2006) 119–132.
 [12] H. Wang, G. Yu, All but 49 numbers are Wiener indices of trees, *Acta Appl. Math.* 92 (1) (2006) 15–20.
 [13] H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* 69 (1947) 17–20.