SIAM J. Discrete Math.
Vol. 29, No. 1, pp. 586–614

THE WIENER INDEX OF RANDOM DIGITAL TREES*

MICHAEL FUCHS† AND CHUNG-KUEI LEE‡

Abstract. The Wiener index has been studied for simply generated random trees, nonplane unlabeled random trees, and a huge subclass of random grid trees containing random binary search trees, random median-of-$(2k+1)$ search trees, random m-ary search trees, random quadtrees, random simplex trees, etc. An important class of random grid trees for which the Wiener index was not studied so far is random digital trees. In this work, we close this gap. More precisely, we derive asymptotic expansions of moments of the Wiener index and show that a central limit law for the Wiener index holds. These results are obtained for digital search trees and bucket versions as well as tries and PATRICIA tries. Our findings answer in the affirmative two questions posed by Neininger.

Key words. Wiener index, random trees, digital trees, moments, central limit theorem

AMS subject classifications. 05C05, 05C80, 60C05, 68W40, 68R10

DOI. 10.1137/140977989

1. Introduction and results. Topological indices of molecular graphs are of great importance in combinatorial chemistry, and many papers have been dedicated to them. One of the most well-known indices is the so-called *Wiener index*, which is defined as the sum of distances of all unordered pairs of nodes of a graph. This index was proposed by Wiener in [52] in order to investigate the boiling point of alkanes. It has been intensively studied, in particular for trees, since trees arise as molecular graphs of acyclic organic molecules; see the survey paper of Dobrynin, Entringer, and Gutman [8] for many results and references.

Here, we are interested in the Wiener index of random trees. The first class of random trees for which the Wiener index was studied was simply generated random trees. In [10], Entringer et al. showed that the mean of the Wiener index in a simply generated random tree of size n is of order $n^{5/2}$. The mean for families of random trees more relevant in chemistry has been investigated by Dobrynin and Gutman in [9] and Wagner in [49], [50].

As for deeper stochastic properties, Neininger in [38] was the first who considered variance and limit laws. More precisely, he showed for random binary search trees and random recursive trees that the mean of the Wiener index is of order $n^2 \log n$ and the variance is of order n^4. Moreover, he also proved a bivariate limit law of the Wiener index and the total path length (which is defined as the sum of distances of all nodes to the root). Janson in [25] then carried out a similar study for simply generated random trees whose Wiener index has variance of order n^5 and again satisfies a bivariate limit law with the total path length (however, the limiting distribution is quite different from the one found by Neininger for random binary search trees and random recursive trees). The same results were very recently also proved to hold for nonplane unlabeled trees by Wagner [51] (he considered both rooted and unrooted cases).

Finally, also very recently, Munsonius in [35] extended the above results of Nein-

*Received by the editors July 17, 2014; accepted for publication (in revised form) January 20, 2015; published electronically March 24, 2015. This work was partially supported by the NSC project under grant NSC-102-2115-M-009-002.

http://www.siam.org/journals/sidma/29-1/97798.html

†Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan (mfuchs@math.nctu.edu.tw).

‡Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan (enix45@gmail.com).

inger to the class of random split trees which was introduced by Devroye in [5]. The class of split trees is a very large class of random trees containing many important types of random trees as special cases, e.g., binary search trees, m-ary search trees, median-of-$(2k+1)$ search trees, quadtrees, simplex trees, and digital trees, among others. Munsonius proved in [35] that for a huge subclass of the class of random split trees, the variance of the Wiener index has order n^4 and a bivariate limit law with the total path length holds. The subclass he considered includes most of the classes of random trees mentioned above but not the important class of digital trees. It is the purpose of this work to fill this gap. Moreover, our work will answer in the affirmative two questions of Neininger from [38] who asked whether or not periodic oscillations are present in the moments of the Wiener index for digital trees and whether or not the Wiener index is asymptotically normally distributed, in contrast to all other classes of random trees studied before whose limit law was nonnormal.

Before recalling the definition of digital trees and discussing our results in more detail, we want to mention that apart from limit laws, results about tail probabilities of the distribution of the Wiener index have been proved as well; see Janson and Chassaing [26], Ali Khan and Neininger [3], Fill and Janson [11], and Munsonius [36]. Moreover, a quantity which is closely related to the Wiener index is the *distance* of two random nodes in a graph which was also extensively studied for many classes of random trees (including digital trees); see Meir and Moon [34], Dobrow [7], Mahmoud and Neininger [33], Devroye and Neininger [6], Panholzer [42], Panholzer and Prodinger [43], Christophi and Mahmoud [4], Aguech, Lasmar, and Mahmoud [1], [2], and Munsonius and Rüschendorf [37].

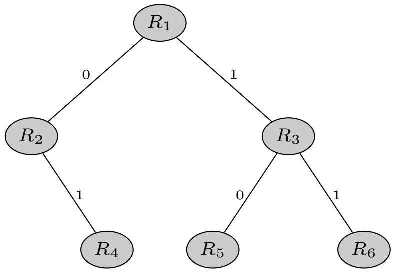
Now, we turn to digital trees, which are fundamental data structures in computer science; see, for instance, the textbooks of Mahmoud [32] or Szpankowski [48]. They are built from data whose keys are infinite 0-1 strings. We equip them with the so-called Bernoulli model, which assumes that every bit is independent and has a Bernoulli distribution with the probability of 0 equal to p. For the sake of simplicity, we consider in this paper only the unbiased Bernoulli model for which $p = 1/2$. The resulting random trees are called *symmetric random digital trees*.

One important subclass of digital trees is *digital search trees*. Here, the tree is constructed as follows. The first key is placed in the root. Then all other keys are distributed to the left or right subtree according to whether their first bit is 0 or 1, respectively. Finally, the first bit of every key is removed and the subtrees are constructed recursively using the same principle; see Figure 1. Digital search trees, although less important from a practical point of view, are the most mathematically challenging class of digital trees; see the paper of Hwang, Fuchs, and Zacharovas [20] and references therein. We will discuss results for the Wiener index and give detailed proofs for this class first. Then, in section 3, we will briefly discuss similar results for variants of digital search trees, namely, bucket digital search trees, tries, and PATRICIA tries (the definitions of these classes of digital trees will be postponed to this section).

Now, fix a random digital search tree of size n and denote by T_n its total path length and by W_n its Wiener index. Then we have the following result for first and second moments.

THEOREM 1. *We have for the mean of the total path length and the Wiener index of digital search trees*

$$\mathbb{E}(T_n) = n\log_2 n + nP_1(\log_2 n) + \mathcal{O}(\log n),$$
$$\mathbb{E}(W_n) = n^2\log_2 n + n^2P_1(\log_2 n) - n^2 + \mathcal{O}(n\log n),$$



$R_1 = 000001\cdots$
$R_2 = 000110\cdots$
$R_3 = 110111\cdots$
$R_4 = 011011\cdots$
$R_5 = 100001\cdots$
$R_6 = 111110\cdots$

FIG. 1. *A digital search tree built from 6 keys with total path length = 8 and Wiener index = 32.*

where $P_1(z)$ is a one-periodic function given in Remark 1 below. Moreover, variances and covariances of the total path length and the Wiener index of digital search trees are given by

$$\operatorname{Var}(T_n) = nP_2(\log_2 n) + \mathcal{O}(1),$$
$$\operatorname{Cov}(T_n, W_n) = n^2 P_2(\log_2 n) + \mathcal{O}(n \log n),$$
$$\operatorname{Var}(W_n) = n^3 P_2(\log_2 n) + \mathcal{O}(n^2 \log n),$$

where $P_2(z)$ is again a one-periodic function given in Remark 2 below.

Remark 1. The result for the mean of the total path length is not new and was obtained first by Knuth in [30]; see also Flajolet and Sedgewick [14]. The periodic function is given by

$$P_1(z) = \frac{\gamma - 1}{\log 2} + \frac{1}{2} - \sum_{k\geq 1} \frac{1}{2^k - 1} + \frac{1}{\log 2} \sum_{k\neq 0} \Gamma(-1 - \chi_k) e^{2k\pi i z},$$

where γ is Euler's constant and $\chi_k = 2k\pi i/\log 2$.

Note that the result for the mean of the Wiener index is also not new since it can be derived from the result in [1].

Finally, we want to remark that with our method of proof, it is straightforward to compute longer asymptotic expansions of the means.

Remark 2. Similar to the mean, the result about the variance of the total path length is also not new; see Kirschenhofer, Prodinger, and Szpankowski [29]. In [20], the following explicit expression was given for the periodic function:

$$P_2(z) = \frac{1}{\log 2} \sum_k \frac{G_2(2 + \chi_k)}{\Gamma(2 + \chi_k)} e^{2k\pi i z},$$

where

$$G_2(2 + \chi_k) = Q_\infty \sum_{j,h,l\geq 0} \frac{(-1)^j 2^{-\binom{j+1}{2}}}{Q_j Q_h Q_l 2^{h+l}} \varphi(2 + \chi_k; 2^{-j-h} + 2^{-j-l}).$$

Here, $Q_j = \prod_{1\leq l\leq j}(1 - 2^{-l}), Q_\infty = \lim_{j\to\infty} Q_j$, and

$$\varphi(\omega; x) = \begin{cases} \dfrac{\pi(1 + x^{\omega-2}((\omega - 2)x + 1 - \omega))}{(x-1)^2 \sin(\pi\omega)} & \text{if } x \neq 1; \\ \dfrac{\pi(\omega - 1)(\omega - 2)}{2\sin(\pi\omega)} & \text{if } x = 1. \end{cases}$$

Moreover, it was proved in [29] that $P_2(\log_2 n) > 0$ for all n; see also Schachinger [46] for a more elementary proof of this fact.

As for the covariance between total path length and Wiener index and the variance of the Wiener index, these results are new. In particular, note that the variance is of order n^3, which is different from the order obtained for other random split trees; see [35]. This smaller order is actually not too surprising since it has been observed many times that random digital search trees are "less random" than other random split trees (and this result gives further confirmation of this fact).

Again it is straightforward to obtain more terms in the asymptotic expansions of the variances and covariance.

As a corollary of Theorem 1, we obtain the following result.

COROLLARY 1. *For the correlation coefficient of the total path length and the Wiener index of digital search trees, denoted by* $\rho(T_n, W_n)$, *we obtain that* $\lim_{n\to\infty} \rho(T_n, W_n) = 1$.

This will allow us to prove the following result.

THEOREM 2. *We have*

$$\left(\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X, X),$$

where X *is a standard normal distributed random variable and* $\xrightarrow{d}$ *denotes weak convergence.*

Remark 3. Again the central limit theorem for the total path length is not new; see Jacquet and Szpankowski [23] and the discussion in section 5 in [20]. In fact, our result will follow from Jacquet and Szpankowski's result and Corollary 1.

Next, we give a brief description of the method we will use in order to prove our results. First, note that from the definition of the total path length and the Wiener index, we immediately get the following distributional recurrences: for $n \geq 0$, we have

(1.1)
$$T_{n+1} \stackrel{d}{=} T_{B_n} + T^*_{n-B_n} + n,$$
(1.2)
$$W_{n+1} \stackrel{d}{=} W_{B_n} + W^*_{n-B_n} + (B_n + 1)(T^*_{n-B_n} + n - B_n) + (n - B_n + 1)(T_{B_n} + B_n),$$

where $B_n = \mathrm{Binomial}(n, 1/2)$, (T^*_n, W^*_n) denotes an independent copy of (T_n, W_n), and (T_n, W_n) and (B_n) are independent. Also, note that initial conditions are given by $T_0 = W_0 = 0$.

This system of distributional recurrences will be the starting point of our analysis. In order to obtain the moments, we will use the Poisson–Laplace–Mellin method from [20], which was a refinement of a previous approach that used only two ingredients, namely, analytic depoissonization and the Mellin transform; see Jacquet and Szpankowski [24] for the former and Flajolet, Gourdon, and Dumas [13] for the latter. We will give a brief review of this method at the beginning of the next section. Another key ingredient of our proof is the use of poissonized variances and covariances which will also be explained in the next section (this was also one of the key contributions in [20]).

It is interesting to point out that Schachinger in [47] studied a general distributional recurrence which is very similar to the two recurrences above. More precisely, he investigated the distributional recurrence

$$X_n \stackrel{d}{=} X_{B_n} + X^*_{n-B_n} + T_n,$$

where the notation is as above and T_n is a general random variable called the *toll function* (this recurrence is actually the same as that encountered in the analysis of shape parameters in tries which behave similarly to digital search trees; see our results in section 3). For the case $T_n = n^\alpha$, $\alpha > 0$, he proved that the limit law is normal if and only if $\alpha \le 3/2$. In view of this result, it might come as a surprise that the Wiener index is asymptotically normally distributed because the toll sequence in (1.2) should be roughly of order $n^2 \log n$ (since the mean of T_n is of order $n \log n$ and the random variable B_n is highly concentrated at $n/2$). However, note that in Schachinger's result T_n is deterministic and hence independent of X_n, whereas in our situation we have strong dependence.

We conclude the introduction with a sketch of the paper. In the next section, we will recall the Poisson–Laplace–Mellin method from [20] and use it to prove Theorems 1 and 2. In section 3, we will look at variants of digital search trees and state similar results for the Wiener index for these variants. Proofs are also similar to the digital search tree case, and consequently we will not give details. However, we will list necessary differential-functional equations (or functional equations in the cases of tries and PATRICIA tries) for the proofs in an appendix. Finally, in section 4, we will give some concluding remarks.

2. Wiener index for digital search trees. Here, we will prove Theorems 1 and 2 from the introduction. We will start with the result on the moments. As explained in the introduction, we will use the method from [20]. Note that the total path length was already analyzed in [20]. In fact, we will heavily rely on results from this analysis in our derivation below (all these results will be carefully reviewed below; for more details see sections 2.5 and 2.6 in [20]).

As promised in the introduction, we will first recall the Poisson–Laplace–Mellin method from [20]; see Figure 7 in [20] for a flowchart depicting the method and a comparison with a closely related approach of Flajolet and Richmond [12]. The method consists of the following steps.

– We first use Poisson-generating functions of means and second moments, where the Poisson-generating function of a sequence f_n is given by

$$\tilde{f}(z) := e^{-z} \sum_{n\ge 0} f_n \frac{z^n}{n!}.$$

All Poisson-generating functions satisfy a differential-functional equation of the form

$$\tilde{f}(z) + \tilde{f}'(z) = 2\tilde{f}(z/2) + \tilde{t}(z), \tag{2.1}$$

where $\tilde{t}(z)$ is a suitable function.

– Next, we carefully define "poissonized" variances and covariances. This was also one of the crucial steps in the analysis of [20] (see the explanation in the introduction of [20]). Poissonized variances and covariances also satisfy a differential-functional equation of type (2.1).

– The next task is to asymptotically solve (2.1). Therefore, we first apply the Laplace transform to (2.1) to get rid of the differential operator. This yields the following functional equation:

$$(1+s)\mathscr{L}[\tilde{f}(z);s] = 4\mathscr{L}[\tilde{f}(z);2s] + \mathscr{L}[\tilde{t}(z);s].$$

- Next, set

$$Q(s) := \prod_{j\geq 1}(1 - s/2^j)$$

and

$$\bar{\mathscr{L}}[\tilde{f}(z);s] := \frac{\mathscr{L}[\tilde{f}(z);s]}{Q(-s)}, \qquad \bar{\mathscr{L}}[\tilde{t}(z);s] = \frac{\mathscr{L}[\tilde{t}(z);s]}{Q(-2s)}.$$

Dividing the functional equation from the previous step by $Q(-2s)$ yields the slightly simplified functional equation

$$\bar{\mathscr{L}}[\tilde{f}(z);s] = 4\bar{\mathscr{L}}[\tilde{f}(z);2s] + \bar{\mathscr{L}}[\tilde{t}(z);s].$$

- An asymptotic expansion of $\bar{\mathscr{L}}[\tilde{f}(z);s]$ as $s \to 0$ is derived by a standard application of the Mellin transform; see [13].
- The inverse Laplace transform then yields an asymptotic expansion of $\tilde{f}(z)$ as $z \to \infty$.
- Finally, depoissonization is used in order to get an asymptotic expansion of f_n from that of $\tilde{f}(z)$; see [23] and section 2.3 in [20].

Now, we will start with our analysis. Therefore, set

$$\tilde{f}_{1,0}(z) = e^{-z}\sum_{n\geq 0}\mathbb{E}(T_n)\frac{z^n}{n!} \qquad \text{and} \qquad \tilde{f}_{0,1}(z) = e^{-z}\sum_{n\geq 0}\mathbb{E}(W_n)\frac{z^n}{n!}.$$

Then, from (1.1), (1.2), and a straightforward computation, one obtains

$$\begin{aligned}
&\tilde{f}_{1,0}(z) + \tilde{f}'_{1,0}(z) = 2\tilde{f}_{1,0}(z/2) + z,\\
&\tilde{f}_{0,1}(z) + \tilde{f}'_{0,1}(z) = 2\tilde{f}_{0,1}(z/2) + (z+2)\tilde{f}_{1,0}(z/2) + \frac{z^2}{2} + z
\end{aligned} \tag{2.2}$$

with $\tilde{f}_{1,0}(0) = \tilde{f}_{0,1}(0) = 0$. Similarly, set

$$\tilde{f}_{2,0}(z) = e^{-z}\sum_{n\geq 0}\mathbb{E}(T_n^2)\frac{z^n}{n!},\ \tilde{f}_{1,1}(z) = e^{-z}\sum_{n\geq 0}\mathbb{E}(T_nW_n)\frac{z^n}{n!},\ \tilde{f}_{0,2}(z) = e^{-z}\sum_{n\geq 0}\mathbb{E}(W_n^2)\frac{z^n}{n!}.$$

Then, again from (1.1), (1.2) with a slightly more involved computation,

$$\tilde{f}_{2,0}(z) + \tilde{f}'_{2,0}(z) = 2\tilde{f}_{2,0}(z/2) + 2\tilde{f}^2_{1,0}(z/2) + 4z\tilde{f}_{1,0}(z/2) + 2z\tilde{f}'_{1,0}(z/2) + z^2 + z,$$

$$\begin{aligned}
\tilde{f}_{1,1}(z) + \tilde{f}'_{1,1}(z) = {}& 2\tilde{f}_{1,1}(z/2) + 2\tilde{f}_{1,0}(z/2)\tilde{f}_{0,1}(z/2) + z\tilde{f}_{1,0}(z/2)\tilde{f}'_{1,0}(z/2)\\
&+ (z+2)\tilde{f}_{2,0}(z/2) + (z+2)\tilde{f}^2_{1,0}(z/2) + (2z^2+5z)\tilde{f}_{1,0}(z/2)\\
&+ \frac{3z^2+4z}{2}\tilde{f}'_{1,0}(z/2) + 2z\tilde{f}_{0,1}(z/2) + z\tilde{f}'_{0,1}(z/2) + \frac{z^3+4z^2+2z}{2},
\end{aligned}$$

$$\begin{aligned}
\tilde{f}_{0,2}(z) + \tilde{f}'_{0,2}(z) = {}& 2\tilde{f}_{0,2}(z/2) + \left(\frac{z^3}{2} + 3z + 2\right)\tilde{f}_{2,0}(z/2) + (2z+4)\tilde{f}_{1,1}(z/2)\\
&+ (2z+4)\tilde{f}_{1,0}(z/2)\,\tilde{f}_{0,1}(z/2) + 2z\tilde{f}_{1,0}(z/2)\,\tilde{f}'_{0,1}(z/2) + 2\tilde{f}_{0,1}(z/2)^2
\end{aligned}$$

$$+ (2z^2+4z)\tilde f_{0,1}(z/2) + (2z^2+2z)\tilde f'_{0,1}(z/2)$$
$$+ \left(\frac{z^2}{2}+2z+2\right)\tilde f_{1,0}(z/2)^2 + (z^2+2z)\tilde f_{1,0}(z/2)\,\tilde f'_{1,0}(z/2)$$
$$+ \frac{z^2}{2}\tilde f'_{1,0}(z/2)^2 + (z^3+6z^2+6z)\tilde f_{1,0}(z/2)$$
$$+ (z^3+5z^2+2z)\tilde f'_{1,0}(z/2) + \frac{z^4}{4} + 2z^3 + 4z^2 + z,$$

where $\tilde f_{2,0}(0) = \tilde f_{1,1}(0) = \tilde f_{0,2}(0) = 0$.

Next, we define poissonized variances and covariances. In our context, it turns out that a good choice is given by

$$\tilde V(z) = \tilde f_{2,0}(z) - \tilde f_{1,0}(z)^2 - z\tilde f'_{1,0}(z)^2,$$
$$\tilde C(z) = \tilde f_{1,1}(z) - \tilde f_{1,0}(z)\tilde f_{0,1}(z) - z\tilde f'_{1,0}(z)\tilde f'_{0,1}(z),$$
$$\tilde W(z) = \tilde f_{0,2}(z) - \tilde f_{0,1}(z)^2 - z\tilde f'_{0,1}(z)^2.$$

The reason we define them in this way will become clear in the depoissonization step; see also the detailed description in the introduction of [20]. Using the above differential-functional equations, a long computation (which can be done with Maple) gives the following differential-functional equation for $\tilde V(z)$, $\tilde C(z)$, and $\tilde W(z)$:

$$\tilde V(z) + \tilde V'(z) = 2\tilde V(z/2) + z\tilde f''_{1,0}(z)^2,$$

(2.3)
$$\tilde C(z) + \tilde C'(z) = 2\tilde C(z/2) + (z+2)\tilde V(z/2) + z\tilde f''_{1,0}(z)\tilde f''_{0,1}(z),$$

$$\tilde W(z) + \tilde W'(z) = 2\tilde W(z/2) + (2z+4)\tilde C(z/2) + \left(\frac{z^2}{2}+3z+2\right)\tilde V(z/2) + z^2\tilde f'_{1,0}(z/2)^2$$
(2.4)
$$+ 2z^2\tilde f'_{1,0}(z/2) + z\tilde f''_{0,1}(z)^2 + z^2$$

with $\tilde V(0) = \tilde C(0) = \tilde W(0) = 0$.

We will now apply the above approach to these differential-functional equations. We will start with the mean value.

Mean value of Wiener index. We will start from (2.2). According to the above method, we first apply the Laplace transform, which yields

$$(1+s)\mathscr{L}[\tilde f_{0,1}(z);s] = 4\mathscr{L}[\tilde f_{0,1}(z);2s] - 2\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde f_{1,0}(z);2s] + 4\mathscr{L}[\tilde f_{1,0}(z);2s] + \frac{1+s}{s^3}.$$

Next, dividing by $Q(-2s)$ and setting

$$\bar{\mathscr{L}}[\tilde f_{0,1}(z);s] = \frac{\mathscr{L}[\tilde f_{0,1}(z);s]}{Q(-s)}, \qquad \bar{\mathscr{L}}[\tilde f_{1,0}(z);s] = \frac{\mathscr{L}[\tilde f_{1,0}(z);s]}{Q(-2s)}$$

gives
(2.5)
$$\bar{\mathscr{L}}[\tilde f_{0,1}(z);s] = 4\bar{\mathscr{L}}[\tilde f_{0,1}(z);2s] - \frac{2}{Q(-2s)}\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde f_{1,0}(z);2s] + 4\bar{\mathscr{L}}[\tilde f_{1,0}(z);2s] + \frac{1+s}{s^3Q(-2s)}.$$

Observe that

$$\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);2s] = \mathscr{L}[\tilde{f}_{1,0}(z);2s]\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{Q(-2s)} + \frac{1}{Q(-2s)}\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde{f}_{1,0}(z);2s]. \tag{2.6}$$

Moreover, logarithmic differentiation yields

$$\frac{\mathrm{d}}{\mathrm{d}s}Q(-2s) = \frac{\mathrm{d}}{\mathrm{d}s}\exp\{\log(Q(-2s))\} = Q(-2s)\frac{\mathrm{d}}{\mathrm{d}s}\sum_{j\geq 0}\log\left(1+\frac{s}{2^j}\right) = Q(-2s)\sum_{j\geq 0}\frac{1}{2^j+s}.$$

Set $A(s) = \sum_{j\geq 0}\frac{1}{2^j+s}$, whose Maclaurin series is given by

$$A(s) = \sum_{j\geq 0}\sum_{k\geq 0}\frac{(-s)^k}{2^{(k+1)j}} = \sum_{k\geq 0}\frac{2^{k+1}}{2^{k+1}-1}(-s)^k.$$

Next,

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{Q(-2s)} = -\frac{1}{Q(-2s)^2}\frac{\mathrm{d}}{\mathrm{d}s}Q(-2s) = -\frac{A(s)}{Q(-2s)} = -\frac{2}{Q(-2s)} - \frac{\bar{A}(s)}{Q(-2s)}, \tag{2.7}$$

where $\bar{A}(s)$ is the meromorphic extension of $\sum_{k\geq 1}2^{k+1}(-s)^k/(2^{k+1}-1)$. Plugging (2.7) into (2.6), and (2.6) in turn into (2.5), gives

$$\bar{\mathscr{L}}[\tilde{f}_{0,1}(z);s] = 4\bar{\mathscr{L}}[\tilde{f}_{0,1}(z);2s] - 2\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);2s] - 2\bar{A}(s)\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);2s] + \frac{1+s}{s^3Q(-2s)}. \tag{2.8}$$

The next step is to apply the Mellin transform. Therefore, note that from [20], we know that

$$\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);s] = \begin{cases}\mathcal{O}\left(|s|^{-2}|\log s|\right) & \text{as } s\to 0;\\ \mathcal{O}\left(|s|^{-b}\right) & \text{as } s\to\infty\end{cases}$$

uniformly for s with $|\arg(s)|\leq \pi-\epsilon$, where $b>0$ is an arbitrary large constant. Moreover, again from [20], for $Q(-2s)$ (and consequently also for $\bar{A}(s)$), we have the bounds

$$Q(-2s) = \begin{cases}1+\mathcal{O}(|s|) & \text{as } s\to 0;\\ \mathcal{O}(|s|^{-b}) & \text{as } s\to\infty,\end{cases} \qquad \bar{A}(s) = \begin{cases}\mathcal{O}(|s|) & \text{as } s\to 0;\\ \mathcal{O}(1) & \text{as } s\to\infty,\end{cases} \tag{2.9}$$

again uniformly for s with $|\arg(s)|\leq\pi-\epsilon$, where $b>0$ is an arbitrary large constant. As a consequence of this and Ritt's theorem (see Chapter 1, section 4.3 in Olver [41]), the Mellin transform of

$$\tilde{s}_{0,1}(s) = -2\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);2s] + \frac{1+s}{s^3Q(-2s)},$$

which we denote by $S_{0,1}(\omega)$, exists for $\Re(\omega)>3$, and the Mellin transform of

$$\tilde{t}_{0,1}(s) = -2\bar{A}(s)\bar{\mathscr{L}}[\tilde{f}_{1,0}(z);2s],$$

which we denote by $T_{0,1}(\omega)$, exists for $\Re(\omega)>1$. Moreover, by Proposition 5 in [13], we have, as $|t|\to\infty$,

$$S_{0,1}(c+it) = \mathcal{O}\left(e^{-(\pi-\epsilon)|t|}\right), \qquad T_{0,1}(c+it) = \mathcal{O}\left(e^{-(\pi-\epsilon)|t|}\right) \tag{2.10}$$

for all $c \in \mathbb{R}$ contained in the fundamental strip. In fact, using the expression for the Mellin transform for $\mathscr{L}[\tilde{f}_{1,0}(z);s]$ from [20], we obtain for $S_{0,1}(\omega)$ the expression

$$S_{0,1}(\omega) = \frac{Q(2^{\omega-3})\Gamma(\omega)\Gamma(2-\omega)}{2Q_\infty(2^{\omega-3}-1)} + \frac{Q(2^{\omega-3})\Gamma(\omega-1)\Gamma(2-\omega)}{Q_\infty} + \frac{Q(2^{\omega-2})\Gamma(\omega)\Gamma(1-\omega)}{Q_\infty}.$$

Note that from this it follows that (2.10) holds for all $c \in \mathbb{R}$. Finally, by applying the Mellin transform to (2.8),

$$\mathscr{M}[\bar{\mathscr{L}}[\tilde{f}_{1,0}];\omega] = \frac{S_{0,1}(\omega) + T_{0,1}(\omega)}{1 - 2^{2-\omega}}.$$

From this and the above explicit expression for $S_{0,1}(\omega)$, we obtain by the inverse Mellin transform

$$\begin{aligned} \bar{\mathscr{L}}[\tilde{f}_{1,0}(z);s] &= 2s^{-3}\log_2\frac{1}{s} + \left(\frac{1}{\log 2} - 1 - 2c\right)s^{-3} \\ &\quad + \frac{1}{\log 2}\sum_{k\neq 0}\Gamma(3+\chi_k)\Gamma(-1-\chi_k)s^{-3-\chi_k} + \mathcal{O}\left(|s|^{-2}|\log s|\right), \end{aligned} \tag{2.11}$$

where $c = \sum_{k\geq 1} 1/(2^k-1)$, χ_k was defined in Remark 1 in section 1, and the above asymptotic expansion holds uniformly as $s \to 0$ with $|\arg(s)| \leq \pi - \epsilon$. Moreover, due to (2.9), the same asymptotic expansion holds for $\mathscr{L}[\tilde{f}_{1,0}(z);s]$ as well.

Next, we apply the inverse Laplace transform. More precisely, we use Proposition 2.6 in [20], which we first recall since there is a small mistake in the statement of [20] ($|s+1|$ on the right-hand side of (29) in [20] should be replaced by $|s|$). We only state the result for the transfer of $\mathcal{O}$-bounds.

Proposition 1 (Hwang, Fuchs, and Zacharovas [20]). *Let $\tilde{f}(z)$ be a function whose Laplace transform exists and is analytic in $\mathbb{C} \setminus (-\infty, 0]$. Assume that*

$$\mathscr{L}[\tilde{f}(z);s] = \mathcal{O}\left(|s|^{-\alpha}|\log s|^m\right)$$

uniformly for $s \to \infty$ with $|\arg(s)| \leq \pi - \epsilon$, where $\alpha \in \mathbb{R}$ and $m \geq 0$ is an integer. Moreover, assume that

$$\mathscr{L}[\tilde{f}(z);s] = \mathcal{O}\left(|s|^{-1-\epsilon}\right)$$

uniformly for $s \to \infty$ with $|\arg(s)| \leq \pi - \epsilon$. Then

$$\tilde{f}(z) = \mathcal{O}\left(|z|^{\alpha-1}|\log z|^m\right)$$

uniformly as $z \to \infty$ and $|\arg(z)| \leq \pi/2 - \epsilon$.

Remark 4. A similar result holds for the transfer of terms of the form $cs^{-\beta}(\log(1/s))^m$, where $c, \beta \in \mathbb{C}$ and $m \geq 0$ is an integer. More precisely, under the same assumptions as above, we have

$$\mathscr{L}[\tilde{f}(z);s] = cs^{-\beta}\left(\log\frac{1}{s}\right)^m \implies \tilde{f}(z) = cz^{\beta-1}\sum_{0\leq j\leq m}\binom{m}{j}(\log z)^{m-j}\frac{\partial^j}{\partial\omega^j}\frac{1}{\Gamma(\omega)}\Big|_{\omega=\beta}.$$

(Note that another small mistake in Proposition 2.6 in [20] is that $|_{\omega=\beta}$ at the end of this formula is missing.)

Applying this result to (2.11) yields

$$\tilde{f}_{0,1}(z) = z^2 \log_2 z + z^2 P_1(\log_2 z) - z^2 + \mathcal{O}(|z \log z|) \tag{2.12}$$

uniformly as $z \to \infty$ with $|\arg(z)| \le \pi/2 - \epsilon$, where $P_1(z)$ was introduced in Remark 1.

The final step is depoissonization, which is done by using the tools from section 2.3 in [20] which rest on the notion of Jacquet–Szpankowski admissibility (JS-admissibility, for short; see Definition 1 in that section) and its closure properties. In particular, from Lemma 2.3 and Proposition 2.4 of that section, we obtain that $\tilde{f}_{0,1}(z)$ is JS-admissible. Hence,

$$\mathbb{E}(W_n) = \tilde{f}_{0,1}(n) - \frac{n}{2}\tilde{f}''_{0,1}(n) + \text{lower order terms.}$$

Note that from (2.12) and Ritt's theorem, we obtain that the second term on the right-hand side above is of order $\mathcal{O}(n \log n)$. Consequently, the above gives the claimed expansion for the mean.

Covariance of total path length and Wiener index. Here, we start from (2.3) and use the same method as for the mean. First, in [20], we have proved that

$$\tilde{f}_{1,0}(z) = z \log_2 z + z P_1(\log_2 z) + \mathcal{O}(|\log z|) \tag{2.13}$$

uniformly as $z \to \infty$ with $|\arg(z)| \le \pi/2 - \epsilon$. From this, (2.12), and Ritt's theorem, we obtain the bounds

$$z \tilde{f}''_{1,0}(z) \tilde{f}''_{0,1}(z) = \begin{cases} \mathcal{O}(|z|) & \text{as } z \to 0; \\ \mathcal{O}(|\log z|) & \text{as } z \to \infty \end{cases} \tag{2.14}$$

uniformly for z with $|\arg(z)| \le \pi/2 - \epsilon$.

Next, we apply Laplace transform to (2.3) and divide it by $Q(-2s)$. Then, by similar manipulations as for the mean, we obtain

$$\bar{\mathscr{L}}[\tilde{C}(z); s] = 4\bar{\mathscr{L}}[\tilde{C}(z); 2s] - 2\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{V}(z); 2s] - 2\bar{A}(s)\bar{\mathscr{L}}[\tilde{V}(z); 2s] + \bar{g}_{1,1}(s), \tag{2.15}$$

where

$$\bar{g}_{1,1}(s) = \frac{\mathscr{L}[z \tilde{f}''_{1,0}(z) \tilde{f}''_{0,1}(z); s]}{Q(-2s)}.$$

Before applying the Mellin transform, we note that from [20], we have

$$\bar{\mathscr{L}}[\tilde{V}(z); s] = \begin{cases} \mathcal{O}\left(|s|^{-2}\right) & \text{as } s \to 0; \\ \mathcal{O}\left(|s|^{-b}\right) & \text{as } s \to \infty \end{cases}$$

uniformly for s with $|\arg(s)| \le \pi - \epsilon$, where $b > 0$ is an arbitrary large constant. Moreover, from (2.14) and (2.9), we obtain

$$\bar{g}_{1,1}(s) = \begin{cases} \mathcal{O}\left(|s|^{-1} |\log s|\right) & \text{as } s \to 0; \\ \mathcal{O}\left(|s|^{-b}\right) & \text{as } s \to \infty \end{cases}$$

again uniformly for s with $|\arg(s)| \le \pi - \epsilon$, where $b > 0$ is an arbitrary large constant. Hence, the Mellin transform of

$$\tilde{s}_{1,1}(s) = -2\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{V}(z); 2s],$$

which we denote by $S_{1,1}(\omega)$, exists for $\Re(\omega) > 3$, and the Mellin transform of

$$\tilde{t}_{1,1}(s) = -2\bar{A}(s)\bar{\mathscr{L}}[\tilde{V}(z); 2s] + \bar{g}_{1,1}(s),$$

which we denote by $T_{1,1}(\omega)$, exists for $\Re(\omega) > 1$. Also, both Mellin transforms satisfy a bound of the form (2.10) inside their fundamental strips. Moreover, in [20], we showed that

$$\mathscr{M}[\bar{\mathscr{L}}[\tilde{V}]; \omega] = \frac{G_2(\omega)}{1 - 2^{2-\omega}},$$

where $G_2(\omega)$ is analytic for $\Re(\omega) > 0$ and satisfies a bound of the form (2.10) in this half-plane. Consequently, by applying the Mellin transform to (2.15),

$$\mathscr{M}[\bar{\mathscr{L}}[\tilde{C}]; \omega] = \frac{S_{1,1}(\omega) + T_{1,1}(\omega)}{1 - 2^{2-\omega}} = \frac{2^{2-\omega}(\omega - 1)G_2(\omega - 1)}{(1 - 2^{3-\omega})(1 - 2^{2-\omega})} + \frac{T_{1,1}(\omega)}{1 - 2^{2-\omega}}.$$

From this by the inverse Mellin transform

$$\bar{\mathscr{L}}[\tilde{C}(z); s] = \frac{1}{\log 2} \sum_k (2 + \chi_k) G_2(2 + \chi_k) s^{-3-\chi_k} + \mathcal{O}\left(|s|^{-2}\right)$$

uniformly as $s \to 0$ with $|\arg(s)| \leq \pi - \epsilon$. For $G_2(\omega)$, we showed in [20] the expressions given in Remark 2 in section 1. Moreover, from (2.9), we get the same asymptotics for $\mathscr{L}[\tilde{C}(z); s]$.

Applying Proposition 1 yields

$$\tilde{C}(z) = z^2 P_2(\log_2 z) + \mathcal{O}(|z|) \tag{2.16}$$

uniformly as $z \to \infty$ and $|\arg(z)| \leq \pi/2 - \epsilon$, where $P_2(z)$ is given in Remark 2 in section 1.

The final step is depoissonization. Therefore, observe that by the results in section 2.2 in [20], $\tilde{f}_{1,0}(z)$, $\tilde{f}_{0,1}(z)$, and $\tilde{f}_{1,1}(z)$ are all JS-admissible. Hence,

$$\mathrm{Cov}(T_n, W_n) = \tilde{C}(n) - \frac{n}{2}\tilde{C}''(n) - \frac{n^2}{2}\tilde{f}''_{1,0}(n)\tilde{f}''_{0,1}(n) + \text{lower order terms.}$$

Note that due to Ritt's theorem, the second term on the right-hand side is $\mathcal{O}(n)$ and the third term is $\mathcal{O}(n \log n)$. Hence, our claimed result for the covariance is proved.

Variance of Wiener index. Next, we turn to the variance of the Wiener index. We start from (2.4), which we rewrite as

$$\tilde{W}(z) + \tilde{W}'(z) = 2\tilde{W}(z/2) + 2z\tilde{C}(z/2) + \frac{z^2}{2}\tilde{V}(z/2) + \tilde{g}_{0,2}(z)$$

with

$$\tilde{g}_{0,2}(z) = 4\tilde{C}(z/2) + (3z + 2)\tilde{V}(z/2) + z^2\tilde{f}'_{1,0}(z/2)^2 + 2z^2\tilde{f}'_{1,0}(z/2) + z\tilde{f}''_{0,1}(z)^2 + z^2.$$

In [20], we proved that

$$\tilde{V}(z) = zP_2(\log_2 z) + \mathcal{O}(1)$$

uniformly as $z \to \infty$ with $|\arg(z)| \le \pi/2 - \epsilon$. From this, (2.16), (2.13), (2.12), and Ritt's theorem it follows that

$$\tilde{g}_{0,2}(z) = \begin{cases} \mathcal{O}(|z|) & \text{as } z \to 0; \\ \mathcal{O}(|z|^2 |\log z|^2) & \text{as } z \to \infty \end{cases} \tag{2.17}$$

uniformly for z with $|\arg(z)| \le \pi/2 - \epsilon$.

Next, applying the Laplace transform to the above differential-functional equation and dividing by $Q(-2s)$ yields

$$\bar{\mathscr{L}}[\tilde{W}(z); s] = 4\bar{\mathscr{L}}[\tilde{W}(z); 2s] - \frac{4}{Q(-2s)}\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde{C}(z); 2s] + \frac{1}{Q(-2s)}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathscr{L}[\tilde{V}(z); 2s] + \bar{g}_{0,2}(s), \tag{2.18}$$

where

$$\bar{g}_{0,2}(s) = \frac{\mathscr{L}[\tilde{g}_{0,2}(z); s]}{Q(-2s)}.$$

Using the same manipulations as for mean and covariance, we have

$$-\frac{4}{Q(-2s)}\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde{C}(z); 2s] = -4\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{C}(z); 2s] - 4A(s)\bar{\mathscr{L}}[\tilde{C}(z); 2s]. \tag{2.19}$$

Moreover, observe that

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bar{\mathscr{L}}[\tilde{V}(z); 2s] = \frac{1}{Q(-2s)}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathscr{L}[\tilde{V}(z); 2s] - 2\frac{A(s)}{Q(-2s)}\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{L}[\tilde{V}(z); 2s] + \mathscr{L}[\tilde{V}(z); 2s]\frac{\mathrm{d}^2}{\mathrm{d}s^2}\frac{1}{Q(-2s)}$$

and note that

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\frac{1}{Q(-2s)} = -\frac{\mathrm{d}}{\mathrm{d}s}\frac{A(s)}{Q(-2s)} = \frac{A(s)^2}{Q(-2s)} - \frac{B(s)}{Q(-2s)},$$

where $B(s)$ is the meromorphic extension of

$$-\sum_{k\ge 0}\frac{k2^{k+2}}{2^{k+2}-1}(-s)^k.$$

This implies that

$$\frac{1}{Q(-2s)}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathscr{L}[\tilde{V}(z); 2s] = \frac{\mathrm{d}^2}{\mathrm{d}s^2}\bar{\mathscr{L}}[\tilde{V}(z); 2s] + 2A(s)\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{V}(z); 2s] + (A(s)^2 + B(s))\bar{\mathscr{L}}[\tilde{V}(z); 2s] \tag{2.20}$$

and plugging (2.20) and (2.19) into (2.18) yields

$$\bar{\mathscr{L}}[\tilde{W}(z); s] = 4\bar{\mathscr{L}}[\tilde{W}(z); 2s] - 4\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{C}(z); 2s] + \frac{\mathrm{d}^2}{\mathrm{d}s^2}\bar{\mathscr{L}}[\tilde{V}(z); 2s] + \tilde{t}_{0,2}(s)$$

with

$$\tilde{t}_{0,2}(s) = -4A(s)\bar{\mathscr{L}}[\tilde{C}(z); 2s] + 2A(s)\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{V}(z); 2s] + (A(s)^2 + B(s))\bar{\mathscr{L}}[\tilde{V}(z); 2s] + \bar{g}_{0,2}(s).$$

Before we apply the Mellin transform, note that from (2.17) and (2.9),

$$\bar{g}_{0,2}(s) = \begin{cases} \mathcal{O}(|s|^{-3}|\log s|^2) & \text{as } s \to 0; \\ \mathcal{O}(|s|^{-b}) & \text{as } s \to \infty \end{cases}$$

uniformly for s with $|\arg(s)| \leq \pi - \epsilon$, where $b > 0$ is an arbitrary large constant. Moreover, similar as for (2.9),

$$B(s) = \begin{cases} \mathcal{O}(1) & \text{as } s \to 0; \\ \mathcal{O}(|s|^{-2}|\log s|^2) & \text{as } s \to \infty, \end{cases}$$

again uniformly for s with $|\arg(s)| \leq \pi - \epsilon$. From this and corresponding bounds for $A(s), \bar{\mathscr{L}}[\tilde{C}(z);s]$, and $\bar{\mathscr{L}}[\tilde{V}(z);s]$ obtained in the analysis of the mean and covariance, we see that the Mellin transform of $\tilde{t}_{0,2}(s)$, which we denote by $T_{0,2}(\omega)$, exists for $\Re(\omega) > 3$. Similarly, the Mellin transform of

$$\tilde{s}_{0,2}(s) = -4\frac{\mathrm{d}}{\mathrm{d}s}\bar{\mathscr{L}}[\tilde{C}(z);2s] + \frac{\mathrm{d}^2}{\mathrm{d}s^2}\bar{\mathscr{L}}[\tilde{V}(z);2s],$$

which we denote by $S_{0,2}(\omega)$, exists for $\Re(\omega) > 4$. Both of these Mellin transforms satisfy a bound of the form (2.10) inside their fundamental strip. Moreover, observe that using the expressions from the analysis of the covariance, $S_{0,2}(\omega)$ is given by

$$S_{0,2}(\omega) = \frac{2^{2-\omega}(2^{3-\omega}+1)(\omega-1)(\omega-2)G_2(\omega-2)}{(1-2^{3-\omega})(1-2^{4-\omega})} + \frac{2^{3-\omega}(\omega-1)T_{1,1}(\omega-1)}{1-2^{3-\omega}},$$

where $G_2(\omega)$ is an analytic function for $\Re(\omega) > 0$, $T_{1,1}(\omega)$ is an analytic function for $\Re(\omega) > 1$, and both satisfy a bound of the form (2.10) in their half-plane of analyticity. Overall, we obtain for the Mellin transform of $\bar{\mathscr{L}}[\tilde{W}(z);s]$

$$\begin{aligned} \mathscr{M}[\bar{\mathscr{L}}[\tilde{W}];\omega] &= \frac{S_{0,2}(\omega)+T_{0,2}(\omega)}{1-2^{2-\omega}} \\ &= \frac{2^{2-\omega}(2^{3-\omega}+1)(\omega-1)(\omega-2)G_2(\omega-2)}{(1-2^{2-\omega})(1-2^{3-\omega})(1-2^{4-\omega})} \\ &\quad + \frac{2^{3-\omega}(\omega-1)T_{1,1}(\omega-1)}{1-2^{3-\omega}} + \frac{T_{0,2}(\omega)}{1-2^{2-\omega}}. \end{aligned}$$

From this, by applying the inverse Mellin transform

$$\bar{\mathscr{L}}[\tilde{W}(z);s] = \frac{1}{\log 2}\sum_k (3+\chi_k)(2+\chi_k)G_2(2+\chi_k)s^{-4-\chi_k} + \mathcal{O}(|s|^{-3-\epsilon})$$

uniformly as $s \to 0$ with $|\arg(s)| \leq \pi - \epsilon$. Moreover, due to (2.9), the same is also true for $\mathscr{L}[\tilde{W}(z);s]$.

Next, we apply Proposition 1 and obtain

$$\tilde{W}(z) = z^3 P_2(\log_2 z) + \mathcal{O}(|z|^{2+\epsilon})$$

uniformly as $z \to \infty$ with $|\arg(z)| \leq \pi/2 - \epsilon$.

The final step is the depoissonization step, where as above we use the results from section 2.2 in [20]. By these results, $\tilde{f}_{0,2}(z)$ and $\tilde{f}_{0,1}(z)$ are both JS-admissible. Consequently,

$$\mathrm{Var}(W_n) = \tilde{W}(n) - \frac{n}{2}\tilde{W}''(n) - \frac{n^2}{2}\tilde{f}''_{0,1}(n)^2 + \text{smaller order terms.}$$

By Ritt's theorem, the second term on the right-hand side is $\mathcal{O}(n^2)$ and the third term is $\mathcal{O}(n^2 \log^2 n)$. From this our result follows (the claimed error term in Theorem 1 is obtained by a slightly refined analysis, which we leave as an exercise to the reader).

This concludes our proof of Theorem 1 and consequently also Corollary 1. We will use now the latter to give a proof of Theorem 2. As a second ingredient, we need the following central limit theorem for the total path length.

Theorem 3 (Jacquet and Szpankowski [23]). *We have*

$$\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}} \xrightarrow{d} X,$$

where X has a standard normal distribution.

Proof of Theorem 2. First set

$$X_n = \frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}.$$

Then by the above result

$$X_n \xrightarrow{d} X,$$

where X has a standard normal distribution. Consequently,

$$(X_n, X_n) \xrightarrow{d} (X, X).$$

Next, define

$$Y_n = \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} - \frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}.$$

Note that

$$\begin{aligned}\mathbb{E}(Y_n^2) &= \frac{\mathbb{E}(W_n - \mathbb{E}(W_n))^2}{\mathrm{Var}(W_n)} + \frac{\mathbb{E}(T_n - \mathbb{E}(T_n))^2}{\mathrm{Var}(T_n)} - 2\frac{\mathbb{E}((W_n - \mathbb{E}(W_n))(T_n - \mathbb{E}(T_n)))}{\sqrt{\mathrm{Var}(W_n)\mathrm{Var}(T_n)}} \\ &= 2 - 2\rho(T_n, W_n).\end{aligned}$$

Hence, by Markov's inequality

$$P(|Y_n| \geq \epsilon) \leq \frac{\mathbb{E}(Y_n^2)}{\epsilon^2} \longrightarrow 0 \qquad \text{as } n \to \infty.$$

Thus, $Y_n \xrightarrow{P} 0$ and consequently $(0, Y_n) \xrightarrow{P} (0,0)$ (here, $\xrightarrow{P}$ denotes convergence in probability). Using Slutsky's theorem (also called Cramér's theorem; see Theorem 11.4 in Gut [17]) now implies

$$(X_n, X_n) + (0, Y_n) \xrightarrow{d} (X, X).$$

Since

$$(X_n, X_n) + (0, Y_n) = \left(\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}}(W_n)} \right),$$

this proves our claim. ∎

Remark 5. Alternatively, one could define the random variable

$$U_n = W_n - nT_n + n^2$$

whose mean by Theorem 1 is of order $\mathcal{O}(n \log n)$. By the same result, also the order of the variance is small:

$$\mathrm{Var}(U_n) = \mathrm{Var}(W_n) + n^2\mathrm{Var}(T_n) - 2n\mathrm{Cov}(T_n, W_n) = \mathcal{O}(n^2 \log n).$$

In particular, this is of a smaller order than the variance of W_n. Thus, again by Markov's inequality,

$$\frac{U_n - \mathbb{E}(U_n)}{\sqrt{\mathrm{Var}(W_n)}} = \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}} - \frac{n(T_n - \mathbb{E}(T_n))}{\sqrt{\mathrm{Var}(W_n)}} \xrightarrow{P} 0.$$

Consequently, Y_n from our above proof can be replaced by $(U_n - \mathbb{E}(U_n))/\sqrt{\mathrm{Var}(W_n)}$.

Overall, this suggests an alternative approach to Theorem 2 via a direct study of the moments of U_n. Such a study is possible with the same tools as above since U_n is easily seen to satisfy a distributional recurrence similar to those of the Wiener index.

3. Wiener index for variants of digital search trees. In this section, we are going to discuss results similar to those in section 1 for variants of digital search trees. Proofs of these results follow along the same lines (or are even easier, since in some cases the Laplace transform is not needed) and will not be given. For the reader's convenience, we will list the (differential-)functional equations for the poissonized mean, variances, and covariances which are crucial to the proofs in the appendix. Our results can be deduced from them with an approach similar to those used in section 2.

We start by defining the variants of digital search trees we want to investigate. The first variant is *bucket digital search trees* where every node can hold up to $b \geq 2$ keys with all internal nodes (nonleaf nodes) holding exactly b keys; for an example see Figure 2. Bucket digital search trees were discussed in many papers; see [20] and references therein. Note that there are two types of total path length in bucket digital trees: the sum of distances of all keys to the root and the sum of distances of all nodes to the root; the former is called *key-wise path length* and the latter *node-wise path length* (see [20] for more details). Accordingly, we also have a *key-wise Wiener index* and a *node-wise Wiener index*. Results for both Wiener indices in random bucket digital search trees will be presented below.

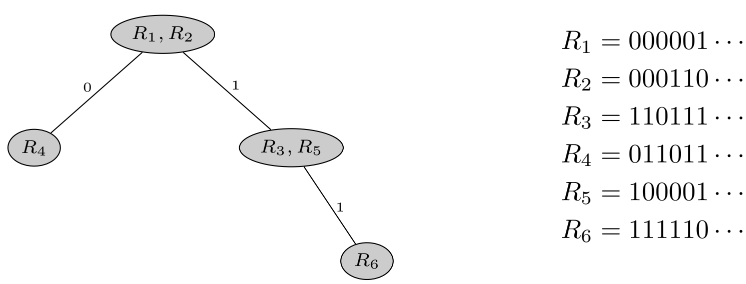


FIG. 2. *A bucket digital search tree with* $b = 2$ *built from* 6 *keys with key-wise path length* $= 5$, *key-wise Wiener index* $= 19$, *node-wise path length* $= 4$, *and node-wise Wiener index* $= 10$.

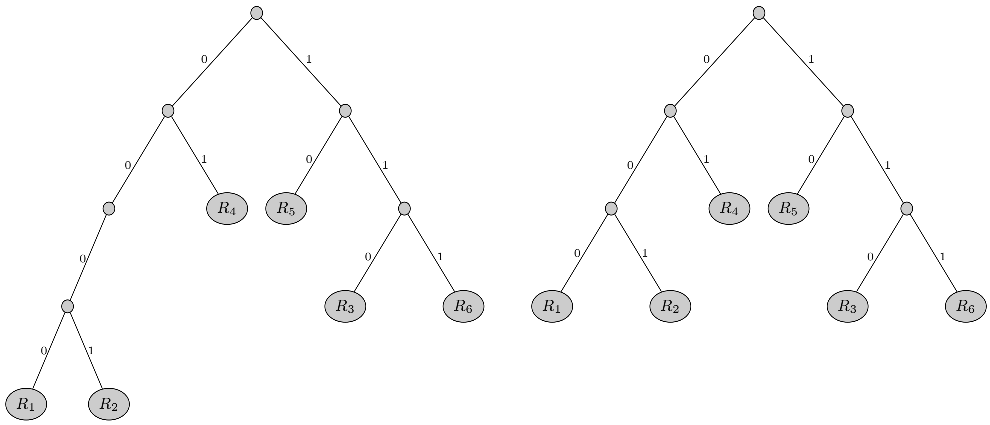


Fig. 3. *A trie built from the data from Figure* 2 *with external path length* = 18*, external Wiener index* = 72*, internal path length* = 9, *and internal Wiener index* = 35. *The corresponding PATRICIA trie is depicted on the right with total path length* = 16 *and Wiener index* = 64.

Another variant of digital search trees is *tries* (from the word data re*trie*val), which are one of the most important data structures on words with numerous applications; see [32], [48], and Park et al. [44] and references therein. For the reader's convenience, we recall the definition. As for digital search trees, start with a set of n data whose keys are infinite 0-1 strings. However, in contrast to digital search trees, a binary tree is built with keys only stored in the leaves. This is done as follows: Whenever a new key is stored, we use it to search in the already existing trie until we encounter a leaf (which already contains a key). Then the leaf is replaced by an internal node and the two keys are distributed to the two subtrees. If they go to the same subtree, then this procedure is repeated until both keys go to different subtrees where they are stored as leaves; see Figure 3 for an example. Again there are two types of total path length: the *external path length* (which uses the leaves) and the *internal path length* (which uses the internal nodes whose number is random); see [15]. Hence, there are also two different types of Wiener indices, namely, the *external Wiener index* and the *internal Wiener index*. Again both of these Wiener indices will be discussed below.

As a final variant of digital search trees, we consider *PATRICIA tries*; see [48]. The construction principle of PATRICIA tries is similar to that of tries, with the only difference being that one-way branching is suppressed (or, in other words, first a trie is built from the data and then all nodes with only one subtree are deleted); see Figure 3 for an example. Again there are two Wiener indices; however, since the number of internal nodes is now deterministic, they exhibit the same behavior. Therefore, we are going to give results only for the external Wiener index (which subsequently will be called the Wiener index for brevity).

As in section 1, we will denote by T_n the total path length and by W_n the Wiener index (both either key-wise or node-wise or external or internal depending on the context). Moreover, for the node-wise Wiener index and the internal Wiener index, we also need the number of nodes (internal in the case of internal Wiener index), which will be denoted by N_n.

Key-wise Wiener index of bucket digital search trees. Here, we have the following distributional recurrences for T_n and W_n: for $n \geq 0$,

$$\begin{aligned} T_{n+b} &\stackrel{d}{=} T_{B_n} + T^*_{n-B_n} + n, \\ W_{n+b} &\stackrel{d}{=} W_{B_n} + W^*_{n-B_n} + (B_n+1)(T^*_{n-B_n} + n - B_n) + (n - B_n + 1)(T_{B_n} + B_n), \end{aligned}$$

where the notation is as in section 1 and initial conditions are given by $T_0 = \cdots = T_{b-1} = W_0 = \cdots = W_{b-1} = 0$.

From these recurrences, we obtain the following results for the mean and variance.

THEOREM 4. *We have for the mean of the key-wise path length and key-wise Wiener index of bucket digital search trees*

$$\begin{aligned} \mathbb{E}(T_n) &= n\log_2 n + nP_1(\log_2 n) + \mathcal{O}(\log n), \\ \mathbb{E}(W_n) &= n^2\log_2 n + n^2 P_1(\log_2 n) - n^2 + \mathcal{O}(n\log n), \end{aligned}$$

where $P_1(z)$ is a one-periodic function given in the remark below. Moreover, variances and covariances of the key-wise path length and key-wise Wiener index of bucket digital search trees are given by

$$\begin{aligned} \mathrm{Var}(T_n) &= nP_2(\log_2 n) + \mathcal{O}(1), \\ \mathrm{Cov}(T_n, W_n) &= n^2 P_2(\log_2 n) + \mathcal{O}(n\log n), \\ \mathrm{Var}(W_n) &= n^3 P_2(\log_2 n) + \mathcal{O}(n^2\log n), \end{aligned}$$

where $P_2(z)$ is again a one-periodic function given in the remark below.

Remark 6. The results for the mean and variance of the key-wise path length were first obtained by Hubalek in [18]. In [20], the authors gave the following expressions for the periodic functions:

$$P_1(z) = \frac{\gamma - 1}{\log 2} + \frac{1}{2} + \frac{c}{\log 2} + \frac{1}{\log 2}\sum_{k\neq 0}\frac{G_1(2+\chi_k)}{\Gamma(2+\chi_k)}e^{2k\pi i z},$$

where

$$G_1(\omega) = \int_0^\infty \frac{s^{\omega-3}}{Q(-2s)^b}\mathrm{d}s, \qquad c = \lim_{\omega\to 2}(G_1(\omega) - 1/(\omega-2)),$$

and

$$P_2(z) = \frac{1}{\log 2}\sum_k \frac{G_2(2+\chi_k)}{\Gamma(2+\chi_k)}e^{2k\pi i z},$$

where

$$G_2(\omega) = \int_0^\infty \frac{s^{\omega-1}}{Q(-2s)^b}\int_0^\infty e^{-zs}\tilde{g}(z)\mathrm{d}z\mathrm{d}s$$

with

$$\begin{aligned} \tilde{g}(z) = &\left(\sum_{0\leq j\leq b}\binom{b}{j}\tilde{f}^{(j)}_{1,0}(z)\right)^2 + z\left(\sum_{0\leq j\leq b}\binom{b}{j}\tilde{f}^{(j+1)}_{1,0}(z)\right)^2 \\ &- \sum_{0\leq j\leq b}\binom{b}{j}\left(\tilde{f}^2_{1,0}(z) + z\tilde{f}'_{1,0}(z)^2\right)^{(j)} \end{aligned}$$

and $\tilde{f}_{1,0}(z)$ denotes the Poisson generating function of $\mathbb{E}(T_n)$.

Note that the result for the mean of the Wiener index also follows from [4].

Moreover, we have the following bivariate central limit theorem.

THEOREM 5. *We have*

$$\left(\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X, X),$$

where X is a standard normal distributed random variable and $\xrightarrow{d}$ denotes weak convergence.

Remark 7. The central limit theorem for the key-wise path length was first proved by Hubalek et al. [19].

Node-wise Wiener index of bucket digital search trees. Here, the distributional recurrences for N_n, T_n, and W_n are as follows: for $n \geq 0$,

$$N_{n+b} \stackrel{d}{=} N_{B_n} + N^*_{n-B_n} + 1,$$
$$T_{n+b} \stackrel{d}{=} T_{B_n} + T^*_{n-B_n} + N_{B_n} + N^*_{n-B_n},$$
$$W_{n+b} \stackrel{d}{=} W_{B_n} + W^*_{n-B_n} + (N_{B_n}+1)(T^*_{n-B_n} + N^*_{n-B_n}) + (N^*_{n-B_n}+1)(T_{B_n} + N_{B_n}),$$

where B_n is as in section 1, (N^*_n, T^*_n, W^*_n) denotes an independent copy of (N_n, T_n, W_n), and (N_n, T_n, W_n) is independent of (B_n). Initial conditions are given by $T_0 = \cdots = T_{b-1} = W_0 = \cdots = W_{b-1} = N_0 = 0$ and $N_1 = \cdots = N_{b-1} = 1$.

From this, we obtain the following result.

THEOREM 6. *We have for the mean of the number of nodes, node-wise path length, and node-wise Wiener index of bucket digital search trees*

$$\mathbb{E}(N_n) = nP_1(\log_2 n) + \mathcal{O}(1),$$
$$\mathbb{E}(T_n) = n(\log_2 n)P_1(\log_2 n) + \mathcal{O}(n),$$
$$\mathbb{E}(W_n) = n^2(\log_2 n)P_1(\log_2 n)^2 + \mathcal{O}(n^2),$$

where $P_1(z)$ is a one-periodic function given in the remark below. Moreover, variances and covariances of the number of nodes, node-wise path length, and node-wise Wiener index of bucket digital search trees are given by

$$\mathrm{Var}(N_n) = nP_2(\log_2 n) + \mathcal{O}(1),$$
$$\mathrm{Cov}(N_n, T_n) = n(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n),$$
$$\mathrm{Var}(T_n) = n(\log_2 n)^2 P_2(\log_2 n) + \mathcal{O}(n \log n),$$
$$\mathrm{Cov}(N_n, W_n) = 2n^2(\log_2 n)P_1(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n^2),$$
$$\mathrm{Cov}(T_n, W_n) = 2n^2(\log_2 n)^2 P_1(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n^2 \log n),$$
$$\mathrm{Var}(W_n) = 4n^3(\log_2 n)^2 P_1(\log_2 n)^2 P_2(\log_2 n) + \mathcal{O}(n^3 \log n),$$

where $P_2(z)$ is again a one-periodic function given in the remark below.

Remark 8. The results for the number of nodes were first proved in [19]. Moreover, the results were re-proved in [20], where the authors also proved the results for the node-wise path length and derived the following expressions for $P_1(z)$ and $P_2(z)$:

$$P_1(z) = \frac{1}{\log 2} \sum_k \frac{G_1(2+\chi_k)}{\Gamma(2+\chi_k)} e^{2k\pi i z},$$

where

$$G_1(\omega) = \int_0^\infty \frac{s^{\omega-2}}{Q(-2s)^b}(s+1)^{b-1}\mathrm{d}s$$

and

$$P_2(z) = \frac{1}{\log 2}\sum_k \frac{G_2(2+\chi_k)}{\Gamma(2+\chi_k)}e^{2k\pi iz},$$

where

$$G_2(\omega) = \int_0^\infty \frac{s^{\omega-1}}{Q(-2s)^b}\left(\int_0^\infty e^{-zs}\tilde{g}(z)\mathrm{d}z + \frac{(s+1)^{b-1}-(-1)^b(2b-3+(b-1)s)}{(s+2)^2}\right)\mathrm{d}s$$

with

$$\tilde{g}(z) = \left(\sum_{0\le j\le b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right)^2 + z\left(\sum_{0\le j\le b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right)^2 - \sum_{0\le j\le b}\binom{b}{j}\left(\tilde{f}_{1,0}^2(z)+z\tilde{f}'_{1,0}(z)^2\right)^{(j)}$$

and $\tilde{f}_{1,0}(z)$ denotes the Poisson generating function of $\mathbb{E}(T_n)$.

Theorem 6 yields the following trivariate central limit theorem.

THEOREM 7. *We have*

$$\left(\frac{N_n-\mathbb{E}(N_n)}{\sqrt{\mathrm{Var}(N_n)}}, \frac{T_n-\mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n-\mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X,X,X),$$

where X is a standard normal distributed random variable and $\xrightarrow{d}$ denotes weak convergence.

Remark 9. The central limit theorem for the number of nodes was first proved in [19]. Also note that the problem of proving a bivariate central limit law of number of nodes and node-wise path length was posed as an open question in section 5 of [20].

External Wiener index of tries. Here, the distributional recurrences for T_n and W_n are as follows: for $n \ge 2$,

$$T_n \stackrel{d}{=} T_{B_n} + T^*_{n-B_n} + n,$$
$$W_n \stackrel{d}{=} W_{B_n} + W^*_{n-B_n} + B_n(T^*_{n-B_n}+n-B_n) + (n-B_n)(T_{B_n}+B_n),$$

where the notation is as in section 1 and initial conditions are given by $T_0 = T_1 = W_0 = W_1 = 0$.

From this, we obtain the following theorem.

THEOREM 8. *We have for the mean of external path length and external Wiener index of tries*

$$\mathbb{E}(T_n) = n\log_2 n + nP_1(\log_2 n) + \mathcal{O}(\log n),$$
$$\mathbb{E}(W_n) = n^2\log_2 n + n^2P_1(\log_2 n) - n^2 + \mathcal{O}(n\log n),$$

where $P_1(z)$ is a one-periodic function given in the remark below. Moreover, variances and covariances of the external path length and external Wiener index of tries are given by

$$\begin{aligned}\mathrm{Var}(T_n) &= nP_2(\log_2 n) + \mathcal{O}(1),\\ \mathrm{Cov}(T_n, W_n) &= n^2 P_2(\log_2 n) + \mathcal{O}(n\log n),\\ \mathrm{Var}(W_n) &= n^3 P_2(\log_2 n) + \mathcal{O}(n^2\log n),\end{aligned}$$

where $P_2(z)$ is again a one-periodic function given in the remark below.

Remark 10. The result about the mean of the total path length was first obtained in [30]. A detailed analysis of the variance of the total path length was first undertaken by Kirschenhofer, Prodinger, and Szpankowski [27] (see also Jacquet and Régnier [21] for preliminary results). In Fuchs, Hwang, and Zacharovas [15], the following expressions for the periodic functions were obtained:

$$P_1(z) = \frac{\gamma}{\log 2} + \frac{1}{2} - \frac{1}{\log 2}\sum_{k\neq 0}\Gamma(-\chi_k)e^{2k\pi iz}$$

and

$$P_2(z) = \frac{1}{\log 2}\sum_{k} G_2(-1-\chi_k)e^{2k\pi iz},$$

where

$$G_2(\omega) = \Gamma(\omega+1)\left(1 - \frac{\omega^2+\omega+4}{2^{\omega+3}}\right) + 2\sum_{l\geq 1}\frac{(-1)^l\Gamma(\omega+l+1)}{l!(2^l-1)}(l(\omega+l)-1).$$

Note that the result about the mean of the Wiener index also follows from [4].

From the previous result, we again obtain the following theorem.

THEOREM 9. *We have*

$$\left(\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X, X),$$

where X is a standard normal distributed random variable and $\xrightarrow{d}$ denotes weak convergence.

Remark 11. The central limit theorem for the external path length was first proved in [21].

Internal Wiener index of tries. Here, the distributional recurrences for N_n, T_n, and W_n are as follows: for $n \geq 2$,

$$\begin{aligned}N_n &\stackrel{d}{=} N_{B_n} + N^*_{n-B_n} + 1,\\ T_n &\stackrel{d}{=} T_{B_n} + T^*_{n-B_n} + N_{B_n} + N^*_{n-B_n},\\ W_n &\stackrel{d}{=} W_{B_n} + W^*_{n-B_n} + (N_{B_n}+1)(T^*_{n-B_n} + N^*_{n-B_n}) + (N^*_{n-B_n}+1)(T_{B_n} + N_{B_n}),\end{aligned}$$

where the notation is as for the node-wise Wiener index and initial conditions are given by $N_0 = N_1 = T_0 = T_1 = W_0 = W_1 = 0$.

Then we have the following result for mean values, variances, and covariances.

THEOREM 10. *We have for the mean of the number of internal nodes, internal path length, and internal Wiener index of tries*

$$\begin{aligned}\mathbb{E}(N_n) &= nP_1(\log_2 n) + \mathcal{O}(1),\\ \mathbb{E}(T_n) &= n(\log_2 n)P_1(\log_2 n) + \mathcal{O}(n),\\ \mathbb{E}(W_n) &= n^2(\log_2 n)P_1(\log_2 n)^2 + \mathcal{O}(n^2),\end{aligned}$$

where $P_1(z)$ is a one-periodic function given in the remark below. Moreover, variances and covariances of the number of internal nodes, internal path length, and internal Wiener index of tries are given by

$$\begin{aligned}\mathrm{Var}(N_n) &= nP_2(\log_2 n) + \mathcal{O}(1),\\ \mathrm{Cov}(N_n, T_n) &= n(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n),\\ \mathrm{Var}(T_n) &= n(\log_2 n)^2 P_2(\log_2 n) + \mathcal{O}(n \log n),\\ \mathrm{Cov}(N_n, W_n) &= 2n^2(\log_2 n)P_1(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n^2),\\ \mathrm{Cov}(T_n, W_n) &= 2n^2(\log_2 n)^2 P_1(\log_2 n)P_2(\log_2 n) + \mathcal{O}(n^2 \log n),\\ \mathrm{Var}(W_n) &= 4n^3(\log_2 n)^2 P_1(\log_2 n)^2 P_2(\log_2 n) + \mathcal{O}(n^3 \log n),\end{aligned}$$

where $P_2(z)$ is again a one-periodic function given in the remark below.

Remark 12. The result for the mean of the number of internal nodes was first proved in [30]. The variance of the number of internal nodes was first derived by Régnier and Jacquet [45] (see also [21], [22]). In [15], the authors derived the following expression for the periodic functions:

$$P_1(z) = \frac{1}{\log 2} + \frac{1}{\log 2}\sum_{k\neq 0} \chi_k \Gamma(-1-\chi_k)e^{2k\pi i z}$$

and

$$P_2(z) = \frac{1}{\log 2}\sum_{k} G_2(-1-\chi_k)e^{2k\pi i z},$$

where

$$G_2(\omega) = (\omega+1)\Gamma(\omega)\left(1 - \frac{\omega^2+4\omega+8}{2^{\omega+3}}\right) + 2\sum_{l\geq 1}\frac{(-1)^l l\Gamma(\omega+l+1)}{(l+1)!(2^l-1)}(l(\omega+l+1)-1).$$

The results for mean and variance of internal path length and covariance with the number of internal nodes are due to Nguyen-The [40].

As before, we have a central limit theorem which now reads as follows.

THEOREM 11. *We have*

$$\left(\frac{N_n - \mathbb{E}(N_n)}{\sqrt{\mathrm{Var}(N_n)}}, \frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X, X, X),$$

where X is a standard normal distributed random variable and $\xrightarrow{d}$ denotes weak convergence.

Remark 13. The central limit theorem for the number of internal nodes was first proved in [21] and [22]. The bivariate central limit theorem for the number of internal nodes and the internal path length was wrongly stated in [40] (the author of this work did not observe that the covariance matrix is singular, leading to a wrong proof).

Wiener index of PATRICIA tries. Here, we have the following for T_n and W_n: for $n \geq 2$,

$$T_n \stackrel{d}{=} \begin{cases} T_{B_n} + T^*_{n-B_n} + n & \text{if } B_n \neq 0 \text{ or } B_n \neq n; \\ T_n & \text{otherwise,} \end{cases}$$

$$W_n \stackrel{d}{=} \begin{cases} W_{B_n} + W^*_{n-B_n} + B_n(T^*_{n-B_n} + n - B_n) & \\ \quad +(n-B_n)(T_{B_n} + B_n) & \text{if } B_n \neq 0 \text{ or } B_n \neq n; \\ W_n & \text{otherwise,} \end{cases}$$

where the notation is as in section 1 and $T_0 = T_1 = W_0 = W_1 = 0$.

Then we have the following result.

THEOREM 12. *We have for the mean of the total path length and Wiener index of PATRICIA tries*

$$\mathbb{E}(T_n) = n \log_2 n + nP_1(\log_2 n) + \mathcal{O}(\log n),$$
$$\mathbb{E}(W_n) = n^2 \log_2 n + n^2 P_1(\log_2 n) - n^2 + \mathcal{O}(n \log n),$$

where $P_1(z)$ is a one-periodic function given in the remark below. Moreover, variances and covariances of the total path length and Wiener index of PATRICIA tries are given by

$$\mathrm{Var}(T_n) = nP_2(\log_2 n) + \mathcal{O}(1),$$
$$\mathrm{Cov}(T_n, W_n) = n^2 P_2(\log_2 n) + \mathcal{O}(n \log n),$$
$$\mathrm{Var}(W_n) = n^3 P_2(\log_2 n) + \mathcal{O}(n^2 \log n),$$

where $P_2(z)$ is again a one-periodic function given in the remark below.

Remark 14. The result for the mean of the external path length was first derived in [30]. The result for the variance of the total path length is due to Kirschenhofer, Prodinger, and Szpankowski [28]. In [15], the authors deduced the following expressions for the period functions:

$$P_1(z) = \frac{\gamma - 1}{\log 2} + \frac{1}{\log 2} \sum_{k \neq 0} \Gamma(-\chi_k) e^{2k\pi i z}$$

and

$$P_2(z) = \frac{1}{\log 2} \sum_k G_2(-1-\chi_k) e^{2k\pi i z},$$

where

$$G_2(\omega) = \Gamma(\omega+1)\left(2^{\omega+1}(\omega+2) - \frac{\omega^2 + 3\omega + 6}{4}\right) + 2^{\omega+2} \sum_{l \geq 1} \frac{(-1)^l \Gamma(\omega + l + 2)}{(l-1)!(2^l - 1)}.$$

The latter result again implies the following bivariate central limit theorem.

THEOREM 13. *We have*

$$\left(\frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}}, \frac{W_n - \mathbb{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right) \xrightarrow{d} (X, X),$$

where X is a standard normal distributed random variable and $\xrightarrow{d}$ denotes weak convergence.

Remark 15. To the best of our knowledge, the result for the total path length was first obtained by Neininger and Rüschendorf in [39].

4. Conclusion. In this paper we investigated the Wiener index, which was previously studied for simply generated families of random trees, nonplane unlabeled random trees, and a huge subclass of random grid trees. A notable family of random grid trees which was left open were random digital trees. It was the main purpose of this paper to fill this gap.

We studied the Wiener index for various types of random digital trees, namely, random digital search trees, random bucket digital search trees, tries, and PATRICIA tries, and proved (i) that moments exhibit periodic fluctuations (a phenomenon observed for many shape parameters of digital trees), and (ii) that the Wiener index (suitably centralized and normalized) is asymptotically normally distributed. In particular, the node-wise Wiener index was mentioned as an open problem in [35]. We solved this problem here for random digital trees. Also, we note that one further notion of the Wiener index was treated recently by Fuchs and Lee in [16] (see also the Ph.D. thesis of the second author [31]).

As for open problems, the most straightforward question is, what about the asymmetric case? In fact, similar results can be proved for this case as well. We content ourselves with briefly explaining the results and highlighting differences.

First, we consider random digital search trees. One difference for this class of digital trees is that the Poisson–Laplace–Mellin method from [20] cannot be applied since the method only works for symmetric digital search trees. However, one can still apply a combination of analytic depoissonization and Mellin transform with the disadvantage that periodic functions in the results become less explicit. The asymptotic expansions for the mean of total path length and Wiener index are then essentially the same as in the symmetric case (with a different period for the periodic functions which in addition become constant for $\log p/\log q$ irrational where $q := 1-p$). Asymptotic expansions for variances and covariances are slightly different since their order increases from n^k to $n^k \log n$. More precisely, we have

$$\mathrm{Var}(T_n) \sim \frac{pq\log^2(p/q)}{h^3} n\log n,$$
$$\mathrm{Cov}(T_n, W_n) \sim \frac{pq\log^2(p/q)}{h^3} n^2\log n,$$
$$\mathrm{Var}(W_n) \sim \frac{pq\log^2(p/q)}{h^3} n^3\log n,$$

where $h = -p\log p - q\log q$ is the entropy; for the result for the total path length, see, for instance, [23]. The periodic functions (in case $\log p/\log q$ is rational) are still present but now constitute the second order terms of these expansions (again they are constant when $\log p/\log q$ is irrational). From these expansions, we again have Corollary 1 and Theorem 2. Moreover, these results also hold for the key-wise Wiener index of bucket digital search trees, the external Wiener index of tries, and the Wiener index of PATRICIA tries, where in the latter two cases, the periodic functions in the second order term can be made explicit with the tools from [15].

Finally, for the node-wise Wiener index of bucket digital search trees and the internal Wiener index of tries, there is no increase in the order and the asymptotic expansions from the symmetric case also hold in the asymmetric case (again with a different period for the periodic functions which become constant if $\log p/\log q$ is irrational; for the result for the number of nodes and the internal path length of tries, see [15]). Again, the result is less explicit for bucket digital search trees, but periodic functions can be made explicit in the trie case with tools from [15].

Computations in all the above cases are long and cumbersome, but still doable with the help of Maple. However, the resulting expressions would fill many pages. This is why, in this paper, we decided to concentrate entirely on the symmetric case.

Appendix. We use the same notation for poissonized means, variances, and covariances as in section 2. In addition, for the node-wise Wiener index of bucket digital search trees and the internal Wiener index for tries, we denote by $\tilde{h}_1(z)$ the Poisson generating function of $\mathbb{E}(N_n)$ and

$$\begin{aligned}\tilde{H}_N(z) &= \tilde{g}_N(z) - \tilde{h}_1(z)^2 - z\tilde{h}_1'(z)^2,\\ \tilde{H}_T(z) &= \tilde{g}_T(z) - \tilde{h}_1(z)\tilde{f}_{1,0}(z) - z\tilde{h}_1'(z)\tilde{f}_{1,0}'(z),\\ \tilde{H}_W(z) &= \tilde{g}_W(z) - \tilde{h}_1(z)\tilde{f}_{0,1}(z) - z\tilde{h}_1'(z)\tilde{f}_{0,1}'(z),\end{aligned}$$

where $\tilde{g}_N(z)$, $\tilde{g}_T(z)$, and $\tilde{g}_W(z)$ denote the Poisson generating function of $\mathbb{E}(N_n^2)$, $\mathbb{E}(N_nT_n)$, and $\mathbb{E}(N_nW_n)$, respectively.

Key-wise Wiener index of bucket digital search trees. We have

$$\begin{aligned}\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z) &= 2\tilde{f}_{1,0}(z/2) + z,\\ \sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z) &= 2\tilde{f}_{0,1}(z/2) + (z+2)\tilde{f}_{1,0}(z/2) + \frac{z^2}{2} + z\end{aligned}$$

and

$$\begin{aligned}\sum_{j=0}^{b}\binom{b}{j}\tilde{V}^{(j)}(z) &= 2\tilde{V}(z/2) + \left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right)^2 + z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right)^2\\ &\quad - \sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{1,0}(z)^2 + z\tilde{f}_{1,0}'(z)^2\right)^{(j)},\\ \sum_{j=0}^{b}\binom{b}{j}\tilde{C}^{(j)}(z) &= 2\tilde{C}(z/2) + (z+2)\tilde{V}(z/2) + \left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z)\right)\\ &\quad + z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j+1)}(z)\right)\\ &\quad - \sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{1,0}(z)\tilde{f}_{0,1}(z) + z\tilde{f}_{1,0}'(z)\tilde{f}_{0,1}'(z)\right)^{(j)},\\ \sum_{j=0}^{b}\binom{b}{j}\tilde{W}^{(j)}(z) &= 2\tilde{W}(z/2) + (2z+4)\tilde{C}(z/2) + \left(\frac{z^2}{2} + 3z + 2\right)\tilde{V}(z/2) + z^2\tilde{f}_{1,0}'(z/2)^2\\ &\quad + 2z^2\tilde{f}_{1,0}'(z/2) + z^2 + \left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z)\right)^2 + z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j+1)}(z)\right)^2\\ &\quad - \sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{0,1}(z)^2 + z\tilde{f}_{0,1}'(z)^2\right)^{(j)}.\end{aligned}$$

Node-wise Wiener index of bucket digital search trees. We have

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j)}(z)=2\tilde{h}_1(z/2)+1,$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(b)}(z)=2\tilde{f}_{1,0}(z/2)+2\tilde{h}_1(z/2),$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(b)}(z)=2\tilde{f}_{0,1}(z/2)+2\tilde{f}_{1,0}(z/2)\tilde{h}_1(z/2)+2\tilde{h}_1(z/2)^2+2\tilde{f}_{1,0}(z/2)+2\tilde{h}_1(z/2)$$

and

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{H}_N^{(j)}(z)=2\tilde{H}_N(z/2)+\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j)}(z)\right)^2+z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j+1)}(z)\right)^2 -\sum_{j=0}^{b}\binom{b}{j}\left(\tilde{h}_1(z)^2+z\tilde{h}_1'(z)^2\right)^{(j)},$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{H}_T^{(j)}(z)=2\tilde{H}_T(z/2)+2\tilde{H}_N(z/2)+\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right) +z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j+1)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right) -\sum_{j=0}^{b}\binom{b}{j}\left(\tilde{h}_1(z)\tilde{f}_{1,0}(z)+z\tilde{h}_1'(z)\tilde{f}_{1,0}'(z)\right)^{(j)},$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{V}^{(j)}(z)=2\tilde{V}(z/2)+4\tilde{H}_T(z/2)+2\tilde{H}_N(z/2)+\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right)^2 +z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right)^2-\sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{1,0}(z)^2+z\tilde{f}_{1,0}'(z)^2\right)^{(j)},$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{H}_W^{(j)}(z)=2\tilde{H}_W(z/2)+2\tilde{H}_T(z/2)(\tilde{h}_1(z/2)+1)+2\tilde{H}_N(z/2)(2\tilde{h}_1(z/2) +\tilde{f}_{1,0}(z/2)+1)+\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z)\right) +z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{h}_1^{(j+1)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j+1)}(z)\right) -\sum_{j=0}^{b}\binom{b}{j}\left(\tilde{h}_1(z)\tilde{f}_{0,1}(z)+z\tilde{h}_1'(z)\tilde{f}_{0,1}'(z)\right)^{(j)},$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{C}^{(j)}(z) = 2\tilde{C}(z/2) + 2\tilde{H}_W(z/2) + 2\tilde{V}(z/2)(\tilde{h}_1(z/2)+1) + 2\tilde{H}_T(z/2)(3\tilde{h}_1(z/2) + \tilde{f}_{1,0}(z/2) + 2) + 2\tilde{H}_N(z/2)(2\tilde{h}_1(z/2) + \tilde{f}_{1,0}(z/2) + 1) + \left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z)\right) + z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{1,0}^{(j+1)}(z)\right)\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j+1)}(z)\right) - \sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{1,0}(z)\tilde{f}_{0,1}(z) + z\tilde{f}'_{1,0}(z)\tilde{f}'_{0,1}(z)\right)^{(j)},$$

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{W}^{(j)}(z) = 2\tilde{W}(z/2) + 4\tilde{C}(z/2)(\tilde{h}_1(z/2)+1) + 4\tilde{H}_W(z/2)(2\tilde{h}_1(z/2) + \tilde{f}_{1,0}(z/2)+1) + 2\tilde{V}(z/2)\tilde{H}_N(z/2) + \tilde{V}(z/2)((2+z)\tilde{h}_1(z/2)^2 + 4\tilde{h}_1(z/2) + 2) + 2\tilde{H}_T(z/2)^2 + \tilde{H}_T(z/2)(8\tilde{h}_1(z/2)^2 + 16\tilde{h}_1(z/2) + 4z\tilde{h}'_1(z/2)^2 + 4\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2) + 2z\tilde{h}'_1(z/2)\tilde{f}'_{1,0}(z/2) + 4) + 4\tilde{H}_N(z/2)^2 + 8\tilde{H}_N(z/2)^2\tilde{h}_1(z/2)^2 + 8\tilde{H}_N(z/2)\tilde{H}_T(z/2) + \tilde{H}_N(z/2)(8\tilde{h}_1(z/2) + 4z\tilde{h}'_1(z/2)^2 + 8\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2) + 4z\tilde{h}'_1(z/2)\tilde{f}'_{1,0}(z/2) + 2\tilde{f}_{1,0}(z/2)^2 + 4\tilde{f}_{1,0}(z/2) + z\tilde{f}'_{1,0}(z/2)^2 + 2) + z^2\tilde{h}_1(z/2)^4 + 2z^2\tilde{h}_1(z/2)^3\tilde{f}_{1,0}(z/2) + z^2\tilde{h}'_1(z/2)^2\tilde{f}'_{1,0}(z/2)^2 + \left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j)}(z)\right)^2 + z\left(\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{0,1}^{(j+1)}(z)\right)^2 - \sum_{j=0}^{b}\binom{b}{j}\left(\tilde{f}_{0,1}(z)^2 + z\tilde{f}'_{0,1}(z)^2\right)^{(j)}.$$

External Wiener index of tries. We have

$$\tilde{f}_{1,0}(z) = 2\tilde{f}_{1,0}(z/2) + z - ze^{-z},$$
$$\tilde{f}_{0,1}(z) = 2\tilde{f}_{0,1}(z/2) + z\tilde{f}_{1,0}(z/2) + z^2/2$$

and

$$\tilde{V}(z) = 2\tilde{V}(z/2) + e^{-z}(4z\tilde{f}_{1,0}(z/2) + 2z\tilde{f}'_{1,0}(z/2) - 2z^2\tilde{f}'_{1,0}(z/2)) + e^{-z}(z - ze^{-z} + z^2e^{-z} - z^3e^{-z}),$$

$$\tilde{C}(z) = 2\tilde{C}(z/2) + z\tilde{V}(z/2) + e^{-z}\left(z\tilde{f}_{1,0}(z/2) + \frac{z^2}{2}\tilde{f}'_{1,0}(z/2) - \frac{z^3}{2}\tilde{f}'_{1,0}(z/2) + 2z\tilde{f}_{0,1}(z/2) + z\tilde{f}'_{0,1}(z/2) - z^2\tilde{f}'_{0,1}(z/2)\right) + e^{-z}\left(z^2 - \frac{z^3}{2}\right),$$

$$\tilde{W}(z) = 2\tilde{W}(z/2) + 2z\tilde{C}(z/2) + \left(\frac{z^2}{2} + z\right)\tilde{V}(z/2) + z^2\tilde{f}'_{1,0}(z/2)^2 + 2z^2\tilde{f}'_{1,0}(z/2) + z^2.$$

Internal Wiener index of tries. We have

$$\begin{aligned}
\tilde{h}_1(z) &= 2\tilde{h}_1(z/2) + 1 - e^{-z}(1+z),\\
\tilde{f}_{1,0}(z) &= 2\tilde{f}_{1,0}(z/2) + 2\tilde{h}_1(z/2),\\
\tilde{f}_{0,1}(z) &= 2\tilde{f}_{0,1}(z/2) + 2\tilde{f}_{1,0}(z/2)\tilde{h}_1(z/2) + 2\tilde{h}_1(z/2)^2 + 2\tilde{f}_{1,0}(z/2) + 2\tilde{h}_1(z/2)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{H}_N(z) &= 2\tilde{H}_N(z/2) + e^{-z}(4\tilde{h}_1(z/2) + 4z\tilde{h}_1(z/2) - 2z^2\tilde{h}_1'(z/2))\\
&\quad + e^{-z}(1 + z - e^{-z} - 2ze^{-z} - z^2e^{-z} - z^3e^{-z}),\\
\tilde{H}_T(z) &= 2\tilde{H}_T(z/2) + 2\tilde{H}_N(z/2) + e^{-z}(2\tilde{h}_1(z/2) + 2z\tilde{h}_1(z/2) - z^2\tilde{h}_1'(z/2)\\
&\quad + 2\tilde{f}_{1,0}(z/2) + 2z\tilde{f}_{1,0}(z/2) - z^2\tilde{f}_{1,0}'(z/2)),\\
\tilde{V}(z) &= 2\tilde{V}(z/2) + 4\tilde{H}_T(z/2) + 2\tilde{H}_N(z/2),\\
\tilde{H}_W(z) &= 2\tilde{H}_W(z/2) + 2\tilde{H}_T(z/2)(\tilde{h}_1(z/2) + 1) + 2\tilde{H}_N(z/2)(2\tilde{h}_1(z/2)\\
&\quad + \tilde{f}_{1,0}(z/2) + 1) + e^{-z}(2\tilde{h}_1(z/2)^2 + 2z\tilde{h}_1(z/2) + 2\tilde{h}_1(z/2) + 2z\tilde{h}_1(z/2)\\
&\quad - z^2\tilde{h}_1(z/2)\tilde{h}_1'(z/2) - z^2\tilde{h}_1'(z/2) + 2\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2) + 2z\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2)\\
&\quad - z^2\tilde{h}_1(z/2)\tilde{f}_{1,0}'(z/2) - z^2\tilde{h}_1'(z/2)\tilde{f}_{1,0}(z/2) + 2\tilde{f}_{1,0}(z/2) + 2z\tilde{f}_{1,0}(z/2)\\
&\quad - z^2\tilde{f}_{1,0}'(z/2) + 2\tilde{f}_{0,1}(z/2) + 2z\tilde{f}_{0,1}(z/2) - z^2\tilde{f}_{0,1}'(z/2)),\\
\tilde{C}(z) &= 2\tilde{C}(z/2) + 2\tilde{H}_W(z/2) + 2\tilde{V}(z/2)(\tilde{h}_1(z/2) + 1)\\
&\quad + 2\tilde{H}_T(z/2)(3\tilde{h}_1(z/2) + \tilde{f}_{1,0}(z/2) + 2) + 2\tilde{H}_N(z/2)(2\tilde{h}_1(z/2)\\
&\quad + \tilde{f}_{1,0}(z/2) + 1),\\
\tilde{W}(z) &= 2\tilde{W}(z/2) + 4\tilde{C}(z/2)(\tilde{h}_1(z/2) + 1) + 4\tilde{H}_W(z/2)(2\tilde{h}_1(z/2) + \tilde{f}_{1,0}(z/2) + 1)\\
&\quad + 2\tilde{V}(z/2)\tilde{H}_N(z/2) + \tilde{V}(z/2)((2+z)\tilde{h}_1(z/2)^2 + 4\tilde{h}_1(z/2) + 2) + 2\tilde{H}_T(z/2)^2\\
&\quad + \tilde{H}_T(z/2)(8\tilde{h}_1(z/2)^2 + 16\tilde{h}_1(z/2) + 4z\tilde{h}_1'(z/2)^2 + 4\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2)\\
&\quad + 2z\tilde{h}_1'(z/2)\tilde{f}_{1,0}'(z/2) + 4) + 4\tilde{H}_N(z/2)^2 + 8\tilde{H}_N(z/2)^2\tilde{h}_1(z/2)^2\\
&\quad + 8\tilde{H}_N(z/2)\tilde{H}_T(z/2) + \tilde{H}_N(z/2)(8\tilde{h}_1(z/2) + 4z\tilde{h}_1'(z/2)^2 + 8\tilde{h}_1(z/2)\tilde{f}_{1,0}(z/2)\\
&\quad + 4z\tilde{h}_1'(z/2)\tilde{f}_{1,0}'(z/2) + 2\tilde{f}_{1,0}(z/2)^2 + 4\tilde{f}_{1,0}(z/2) + z\tilde{f}_{1,0}'(z/2)^2 + 2)\\
&\quad + z^2\tilde{h}_1(z/2)^4 + 2z^2\tilde{h}_1(z/2)^3\tilde{f}_{1,0}(z/2) + z^2\tilde{h}_1'(z/2)^2\tilde{f}_{1,0}'(z/2)^2.
\end{aligned}$$

Wiener index of PATRICIA tries. We have

$$\begin{aligned}
\tilde{f}_{1,0}(z) &= 2\tilde{f}_{1,0}(z/2) + z - ze^{-z/2},\\
\tilde{f}_{0,1}(z) &= 2\tilde{f}_{0,1}(z/2) + z\tilde{f}_{1,0}(z/2) + \frac{z^2}{2}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{V}(z) &= 2\tilde{V}(z/2) + e^{-z/2}(2z\tilde{f}_{1,0}(z/2) - z^2\tilde{f}_{1,0}'(z/2)) + e^{-z/2}\left(z + \frac{z^2}{2}\right)\\
&\quad - e^{-z}\left(z + \frac{z^3}{4}\right),
\end{aligned}$$

$$\tilde{C}(z) = 2\tilde{C}(z/2) + z\tilde{V}(z/2) + e^{-z/2}\bigg(z\tilde{f}_{1,0}(z/2) + \frac{z^2}{2}\tilde{f}_{1,0}(z/2) + \frac{z^2}{2}\tilde{f}'_{1,0}(z/2) - \frac{z^3}{4}\tilde{f}'_{1,0}(z/2) + z\tilde{f}_{0,1}(z/2) - \frac{z^2}{2}\tilde{f}'_{0,1}(z/2)\bigg) + z^2e^{-z},$$

$$\tilde{W}(z) = 2\tilde{W}(z/2) + 2z\tilde{C}(z/2) + \left(\frac{z^2}{2} + z\right)\tilde{V}(z/2) + z^2\tilde{f}'_{1,0}(z/2)^2 + 2z^2\tilde{f}'_{1,0}(z/2) + z^2.$$

Acknowledgments. We thank Svante Janson for pointing out Slutsky's theorem to us, which was the key tool for proving our central limit theorems. Moreover, we thank one anonymous referee for Remark 5 and both referees for many helpful suggestions. Parts of this work were done when the first author was visiting the Institute of Statistical Sciences, Academia Sinica. He thanks the institute for its hospitality.

REFERENCES

[1] R. Aguech, N. Lasmar, and H. Mahmoud, *Distances in random digital search trees*, Acta Inform., 43 (2006), pp. 243–264.

[2] R. Aguech, N. Lasmar, and H. Mahmoud, *Limit distribution of distances in biased random tries*, J. Appl. Probab., 43 (2006), pp. 1–14.

[3] T. Ali Khan and R. Neininger, *Tail bounds for the Wiener index of random trees*, in Proceedings of the 2007 Conference on the Analysis of Algorithms, Discrete Math. Theor. Comput. Sci. Proc., 2007, pp. 279–289.

[4] C. Christophi and H. Mahmoud, *The oscillatory distribution of distances in random tries*, Ann. Appl. Probab., 15 (2005), pp. 1536–1564.

[5] L. Devroye, *Universal limit laws for depths in random trees*, SIAM J. Comput., 28 (1998), pp. 409–432.

[6] L. Devroye and R. Neininger, *Distances and finger search in random binary search trees*, SIAM J. Comput., 33 (2004), pp. 647–658.

[7] R. Dobrow, *On the distribution of distances in recursive trees*, J. Appl. Probab., 33 (1996), pp. 749–757.

[8] A. A. Dobrynin, R. Entringer, and I. Gutman, *Wiener index of trees: Theory and applications*, Acta Appl. Math., 66 (2001), pp. 211–249.

[9] A. A. Dobrynin and I. Gutman, *The average Wiener index of trees and chemical trees*, J. Chem. Inf. Comput. Sci., 39 (1999), pp. 679–683.

[10] R. C. Entringer, A. Meir, J. W. Moon, and L. A. Székely, *The Wiener index of trees from certain families*, Australas. J. Combin., 10 (1994), pp. 211–224.

[11] J. A. Fill and S. Janson, *Precise logarithmic asymptotics of the right tails of some limit random variables for random trees*, Ann. Comb., 12 (2009), pp. 403–416.

[12] P. Flajolet and B. Richmond, *Generalized digital trees and their difference-differential equations*, Random Structures Algorithms, 3 (1992), pp. 305–320.

[13] P. Flajolet, X. Gourdon, and P. Dumas, *Mellin transforms and asymptotics: Harmonic sums*, Theoret. Comput. Sci., 144 (1995), pp. 3–58.

[14] P. Flajolet and R. Sedgewick, *Digital search trees revisited*, SIAM J. Comput., 15 (1986), pp. 748–767.

[15] M. Fuchs, H.-K. Hwang, and V. Zacharovas, *An analytic approach to the asymptotic variance of trie statistics and related structures*, Theoret. Comput. Sci., 527 (2014), pp. 1–36.

[16] M. Fuchs and C.-K. Lee, *A general central limit theorem for shape parameters of m-ary tries and PATRICIA tries*, Electron. J. Combin., 21 (2014), 1.68.

[17] A. Gut, *Probability: A Graduate Course*, Springer Texts in Statistics, Springer-Verlag, New York, 2005.

[18] F. Hubalek, *On the variance of the internal path length of generalized digital trees: The Mellin convolution approach*, Theoret. Comput. Sci., 242 (2000), pp. 143–168.

[19] F. Hubalek, H.-K. Hwang, W. Lew, H. Mahmoud, and H. Prodinger, *A multivariate view of random bucket digital search trees*, J. Algorithms, 44 (2002), pp. 121–158.

[20] H.-K. Hwang, M. Fuchs, and V. Zacharovas, *Asymptotic variance of random symmetric digital search trees*, Discrete Math. Theor. Comput. Sci., 12 (2010), pp. 103–166.

[21] P. Jacquet and M. Régnier, *Normal Limiting Distribution of the Size and the External Path Length of Tries*, Technical Report RR-0827, INRIA-Rocquencourt, 1988.
[22] P. Jacquet and M. Régnier, *Normal limiting distribution of the size of tries*, in Performance '87, North-Holland, Amsterdam, 1988, pp. 209–223.
[23] P. Jacquet and W. Szpankowski, *Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees*, Theoret. Comput. Sci., 144 (1995), pp. 161–197.
[24] P. Jacquet and W. Szpankowski, *Analytical de-Poissonization and its applications*, Theoret. Comput. Sci., 201 (1998), pp. 1–62.
[25] S. Janson, *The Wiener index of simply generated random trees*, Random Structures Algorithms, 22 (2003), pp. 337–358.
[26] S. Janson and P. Chassaing, *The center of mass of the ISE and the Wiener index of trees*, Electron. Comm. Probab., 9 (2004), pp. 178–187.
[27] P. Kirschenhofer, H. Prodinger, and W. Szpankowski, *On the variance of the external path length in a symmetric digital trie*, Discrete Appl. Math., 25 (1989), pp. 129–143.
[28] P. Kirschenhofer, H. Prodinger, and W. Szpankowski, *On the balance properties of Patricia tries: External path length viewpoint*, Theoret. Comput. Sci., 68 (1989), pp. 1–17.
[29] P. Kirschenhofer, H. Prodinger, and W. Szpankowski, *Digital search trees again revisited: The internal path length perspective*, SIAM J. Comput., 23 (1994), pp. 598–616.
[30] D. E. Knuth, *The Art of Computer Programming, Volume* 3: *Searching and Sorting*, Addison-Wesley, Reading, MA, 1973.
[31] C.-K. Lee, *Probabilistic Analysis of Additive Shape Parameters in Random Digital Trees*, Ph.D. thesis, National Chiao Tung University, 2014.
[32] M. Mahmoud, *Evolution of Random Search Trees*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, 1992.
[33] M. Mahmoud and R. Neininger, *Distribution of distances in random binary search trees*, Ann. Appl. Probab., 13 (2002), pp. 253–276.
[34] A. Meir and J. W. Moon, *The distance between points in random trees*, J. Combinatorial Theory, 8 (1970), pp. 99–103.
[35] G. O. Munsonius, *On the asymptotic internal path length and the asymptotic Wiener index of random split trees*, Electron. J. Probab., 16 (2011), pp. 1020–1047.
[36] G. O. Munsonius, *On tail bounds for random recursive trees*, J. Appl. Probab., 49 (2012), pp. 566–581.
[37] G. O. Munsonius and L. Rüschendorf, *Limit theorems for depths and distances in weighted random b-ary recursive trees*, J. Appl. Probab., 4 (2011)8, pp. 1060–1080.
[38] R. Neininger, *The Wiener index of random trees*, Combin. Probab. Comput., 11 (2002), pp. 587–597.
[39] R. Neininger and L. Rüschendorf, *A general limit theorem for recursive algorithms and combinatorial structures*, Ann. Appl. Probab., 14 (2004), pp. 378–418.
[40] M. Nguyen-The, *Distribution de valuations sur les arbres*, Ph.D. Thesis, LIX, Ecole polytechnique, 2003.
[41] F. W. J. Olver *Asymptotics and Special Functions*, Academic Press, New York, 1974.
[42] A. Panholzer, *The distribution of the size of the ancestor-tree and of the induced spanning subtree for random trees*, Random Structures Algorithms, 25 (2004), pp. 179–207.
[43] A. Panholzer and H. Prodinger, *Spanning tree size in random binary search trees*, Ann. Appl. Probab., 14 (2004), pp. 718–733.
[44] G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski, *Profiles of tries*, SIAM J. Comput., 38 (2009), pp. 1821–1880.
[45] M. Régnier and P. Jacquet, *New results on the size of tries*, IEEE Trans. Inform. Theory, 35 (1989), pp. 203–205.
[46] W. Schachinger, *On the variance of a class of inductive valuations of data structures for digital search*, Theoret. Comput. Sci., 144 (1995), pp. 251–275.
[47] W. Schachinger, *Asymptotic normality of recursive algorithms via martingale difference arrays*, Discrete Math. Theor. Comput. Sci., 4 (2001), pp. 363–397.
[48] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2001.
[49] S. G. Wagner, *A class of trees and its Wiener index*, Acta Appl. Math., 91 (2006), pp. 119–132.
[50] S. G. Wagner, *On the average Wiener index of degree-restricted trees*, Australas. J. Combin., 37 (2007), pp. 187–203.
[51] S. Wagner, *On the Wiener index of random trees*, Discrete Math., 312 (2012), pp. 1502–1511.
[52] H. Wiener, *Structural determination of paraffin boiling points*, J. Amer. Chem. Soc., 69 (1947), pp. 17–20.