

A Precise Analysis of Cuckoo Hashing

MICHAEL DRMOTA and REINHARD KUTZELNIGG

Technische Universität Wien

Cuckoo hashing was introduced by Pagh and Rodler in 2001. Its main feature is that it provides constant worst case search time. The aim of this paper is to present a precise average case analysis of Cuckoo hashing. In particular, we determine the probability that Cuckoo hashing produces no conflicts and give an upper bound for the construction time, that is linear in the size of the table. The analysis rests on a generating function approach to the so called Cuckoo Graph, a random bipartite graph and an application of a double saddle point method to obtain asymptotic expansions. Furthermore, we provide some results concerning the structure of these kind of random graphs. Our results extend an analysis of Devroye and Morin in 2003. Additionally, we provide numerical results confirming the mathematical analysis.

Categories and Subject Descriptors: E.2 [Data]: Data Storage Representations—*hash table representations*; F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems—*Computations on discrete structures*; G.2.1 [Discrete Mathematics]: Combinatorics—*Counting problems; Generating functions*; G.2.2 [Discrete Mathematics]: Graph Theory

General Terms: Algorithms

Additional Key Words and Phrases: Hashing, Cuckoo hashing, random bipartite graphs, saddle point method, generating functions

1. INTRODUCTION

In computer science hash tables are dictionaries, where keys are mapped to the hash table with a so-called hash function. Basic operations on hash tables are insertion, look-up and deletion of data records. Standard hash algorithms such as open addressing or hashing with chaining (see *e.g.* [Knuth 1998],[Gonnet and Baeza-Yates 1991] for details) are widely spread and well analyzed algorithms, but a big drawback is their bad worst case behavior. In this paper we consider a relatively new hash algorithm, cuckoo hashing, that provides a constant worst case search time. The algorithm was introduced in [Pagh and Rodler 2004] and a further analysis was done by [Devroye and Morin 2003].

The authors were supported by the Austrian Science Foundation FWF, project S9604 and by the EU FP6-NEST-Adventure Programme, Contract number 028875 (NEMO).

Author's addresses: Michael Drmota, Institute of Discrete Mathematics and Geometry, Vienna University of Technology, Wiedner Hauptstr. 8–10/104, A-1040 Wien, Austria e-mail: michael.drmota@tuwien.ac.at. Reinhard Kutzelnigg, Institute of Discrete Mathematics and Geometry, Vienna University of Technology, Wiedner Hauptstr. 8–10/104, A-1040 Wien, Austria e-mail: reinhard.kutzelnigg@tuwien.ac.at

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 0000-0000/20YY/0000-0001 \$5.00

The critical point of every hash algorithm is the handling of colliding keys, that is, different keys which are mapped to the same location of the hash table by the hash function. The well known birthday paradox tells us, that such collisions are likely to appear, even if we only consider sparse tables. Usually, collisions are resolved by either allocating additional memory and linking this memory to the table (“chaining”) or by inspecting the other memory cells in a specified order (“open addressing”). Cuckoo hashing uses a different attempt. It restricts the number of possible storage locations of every key x to two, $h_1(x)$ and $h_2(x)$, each in a separate table V_1 and V_2 , respectively, and resolves collisions in rearranging keys. The data points are inserted sequentially, and each storage location can only hold a single data. A new key x_{new} is always inserted in the first table V_1 at location $v_1 = h_1(x_{new})$. If v_1 is already occupied by another key x_{old} , that is, $h_1(x_{old}) = h_1(x_{new}) = v_1$ then we “kick out” x_{old} and move it to its alternate position $v_2 = h_2(x_{old}) \in V_2$. If v_2 is occupied by another key then we proceed with this “kick out”-procedure until we access an empty cell. The algorithm is named after the cuckoo, because this ejection is similar to the birds nesting habits. The insertion procedure may of course end in an endless loop if the same keys are kicked out again and again. In the the latter case, we perform a rehash, that is, we rebuild the whole data structure using new hash functions and potentially also a larger table. Figure 1 depicts the evolution of a small Cuckoo hash table.

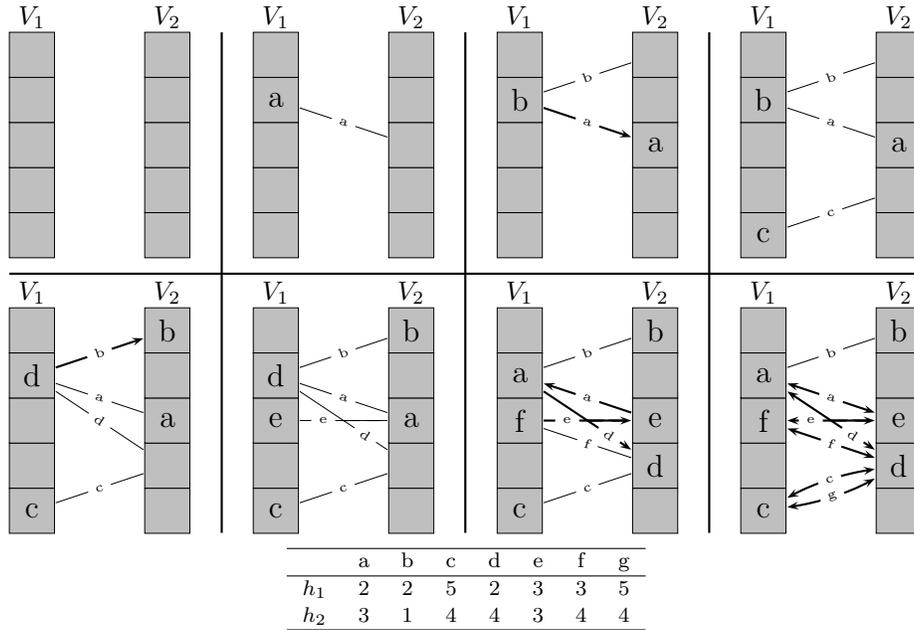


Fig. 1. An example of an evolving Cuckoo hash table. We start with an empty table and try to insert the keys a to g one after the other. Each of the further pictures depicts the data structure after the insertion of a single key. Thin lines connect the two possible storage locations $h_1(x)$ and $h_2(x)$ of a key x . Thick lines with arrows indicate the movement of the corresponding key, caused by the last insertion. The final picture displays the data structure after the attempt to insert g, what is impossible, and causes an endless loop.

There are also some generalizations of cuckoo hashing. One might use d instead of only two tables or storage places (d-ary Cuckoo Hashing) [Fotakis et al. 2005; Czyzowicz et al. 2006] or use cells with a storage capacity greater than one [Dietzfelbinger and Weidling 2007]. Implementation details are *e.g.* discussed in [Ross 2006; Tran and Kittitornkun 2007].

The paper is organized as follows. In Section 2 we introduce a bipartite graph (the Cuckoo graph) that is associated to the Cuckoo hash process and is fundamental for the analysis. The asymptotic results of the paper are then collected and discussed in Section 3. The proofs are given in Sections 4–7. Finally, in Section 8 we discuss a simplified version of Cuckoo hashing. Some technical proofs are put to an Appendix.

2. THE CUCKOO GRAPH

We model Cuckoo hashing with help of a labeled bipartite multigraph. The two (labeled) vertex sets V_1, V_2 represent the two hash tables. In the context of Cuckoo hashing we will assume that V_1 and V_2 have equal size $|V_1| = |V_2| = m$. An asymmetric variant is discussed in [Kutzelnigg 2008]. Further, let h_1, h_2 denote the two hash function with ranges V_1 and V_2 , respectively. The insertion of a key x is encoded by an edge $(h_1(x), h_2(x)) \in V_1 \times V_2$. It is also convenient to use labeled edges where the labels represent the evolution of the hash table, that is, the edge with label j corresponds to the j -th key that is inserted in the table. Note that repeated edges are possible.

This multigraph will be called Cuckoo graph. Interestingly the structure of this graph determines whether the insertion algorithm can resolve all conflicts or not. It is obviously necessary that every component of the Cuckoo graph has the less or equal edges than vertices. This means that all components are either trees or unicyclic (*i.e.* they contain exactly one cycle). On the other hand, it is easy to see that an endless loop in the insertion algorithms cannot occur in a tree or unicyclic component (see [Devroye and Morin 2003] for further details). It is common to call a component of a graph complex if it is neither a tree nor unicyclic. Thus, a Cuckoo graph is proper if and only if it contains no complex component.

Because of this close relation between the hash algorithm and the corresponding graph we can analyze Cuckoo hashing by considering bipartite multigraphs. In particular, if we are interested in the average case analysis of Cuckoo hashing, we can work with random bipartite multigraphs. For example, the probability that Cuckoo hashing works successfully with n keys and table size m equals the probability that a random bipartite multigraph with $2 \times m$ vertices and n edges has no complex component. Further, structural knowledge of tree and unicyclic components provides information about the running time. For instance, the insertion cost of a key x such that the edge $(h_1(x), h_2(x))$ is contained in a tree component is bounded by the diameter of this component.

3. RESULTS

Our analysis is based on the assumption, that the storage locations for the keys x form a sequence $(h_1(x), h_2(x))$ of independent uniform random pairs of integers drawn from $\{1, 2, \dots, m\} \times \{1, 2, \dots, m\}$. Further, if a rehash is necessary, we assume that all new hash values are independent from all previous attempts. One

might argue that this model is out of touch with reality. However, recall that uniform hashing (using a similar independence assumption), and double hashing (using very simple hash functions) behave practically identical [Gonnet and Baeza-Yates 1991]. Unfortunately, this simple hash functions do not work well for Cuckoo hashing. But it is sufficient to use polynomial hash functions with pseudo random behavior [Dietzfelbinger et al. 1992],[Dietzfelbinger and Woelfel 2003]. See also our experimental data.

THEOREM 1. *Suppose that $\varepsilon \in (0, 1)$ is fixed. Then the probability $p(n, m)$ that a Cuckoo hash of $n = \lfloor (1 - \varepsilon)m \rfloor$ data points into two tables of size m succeeds, (that is, the corresponding Cuckoo graph contains no complex component,) is equal to*

$$p(n, m) = 1 - \frac{(2\varepsilon^2 - 5\varepsilon + 5)(1 - \varepsilon)^3}{12(2 - \varepsilon)^2\varepsilon^3} \frac{1}{m} + \mathcal{O}\left(\frac{1}{m^2}\right).$$

This probability decreases to

$$p(n, n) = \sqrt{\frac{2}{3}} + o(1), \quad (1)$$

if n equals m .

There are several related results in the literature. For example, Lemma 2.1 of [Kalugin 1991] already states that the amount of graphs containing complex components tends to zero, but it does not provide an asymptotic approximation. Lemma 2 of [Devroye and Morin 2003] claims the bound $1 - \mathcal{O}(1/m)$, but does not provide a detailed expansion. Further, a formula corresponding to (1) also holds for usual random multigraphs [Janson et al. 1993], see also Section 8.

We provide experimental data in Table I to examine the practicality of the derived approximations. A pseudo random number generator is used to create hash values that satisfy the postulated conditions of randomness.

m	$\varepsilon = 0.4$		$\varepsilon = 0.2$		$\varepsilon = 0.1$		$\varepsilon = 0.06$		$\varepsilon = 0.04$	
	res.	exp.	res.	exp.	res.	exp.	res.	exp.	res.	exp.
$5 \cdot 10^3$	38	36.5	649	672	5070	7606	14467		25624	
10^4	16	18.2	288	336	3046	3803	9954		19857	
$5 \cdot 10^4$	1	3.65	62	67.2	676	761	3127	4008	8439	
10^5	1	1.82	31	33.6	362	380	1753	2004	5210	7202
$5 \cdot 10^5$	0	0.36	7	6.72	76	76.1	398	401	1284	1440

Table I. Comparison of simulation result (res.) and expected number of failures (exp.) during the construction of $5 \cdot 10^5$ cuckoo hash tables.

Our second result concerns the construction time $T_{m,n}$ of a hash table of size $2 \times m$ and n keys. Our measure is, of course, the number of insertions n plus the number of “kick out”-steps. Note that it is no loss of generality to assume that we can successfully fill the table. The probability to fail is $\mathcal{O}(1/m)$. In this case we just choose new hash functions and rebuild the table. The construction time of this table is again bounded by $\mathcal{O}(m)$ and, thus, the total expected error in the construction time will be $\mathcal{O}(1)$.

THEOREM 2. *Under the assumptions of Theorem 1, the expected construction time of a Cuckoo hash table is bounded above by*

$$\mathbb{E}T_{m,n} \leq \min\left(4, \frac{-\log \varepsilon}{1-\varepsilon}\right)n + \mathcal{O}(1),$$

where the constant implied by $\mathcal{O}(1)$ depends on ε .

We obtain these two bounds using different estimators for the insertion cost in a tree component, namely the component size and the diameter. The latter estimator provides of course the “better” bound $4n + \mathcal{O}(1)$, but as long as ε is not close to 0 even the constant 4 is a exaggerated estimate, as can be seen in Table II. See the proof of this theorem for more details.

[Devroye and Morin 2003] and [Pagh and Rodler 2004]¹ give the upper bound $\mathcal{O}(n)$ for the construction time, if ε greater zero is fixed.

m	$\varepsilon = 0.4$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.06$	$\varepsilon = 0.04$
$5 \cdot 10^3$	1.2721	1.3829	1.5108	1.6791	1.8663
10^4	1.2718	1.3782	1.4832	1.6151	1.7809
$5 \cdot 10^4$	1.2717	1.3753	1.4511	1.5196	1.6154
10^5	1.2717	1.3749	1.4469	1.5006	1.5697
$5 \cdot 10^5$	1.2717	1.3746	1.4430	1.4820	1.5147
$\frac{-\log \varepsilon}{1-\varepsilon}$	1.5272	2.0118	2.5584	2.9930	3.3530

sample size: $5 \cdot 10^5$

Table II. Insertion costs of cuckoo hashing.

Finally we present results on the structure of the Cuckoo graph.

THEOREM 3. *Suppose that $\varepsilon \in (0, 1)$ is fixed and that $n = \lfloor (1 - \varepsilon)m \rfloor$. Then a random labeled bipartite multigraph with $2 \times m$ vertices and n edges satisfies the following properties.*

- (1) *The number of unicyclic components $C_{mn,k}^{uc}$ with cycle length $2k$ has in limit a Poisson distribution $Po(\lambda_k)$ with parameter*

$$\lambda_k = \frac{1}{2k} (1 - \varepsilon)^{2k},$$

and the number of unicyclic components $C_{mn}^{uc} = \sum_{k \geq 1} C_{mn,k}^{uc}$ has in limit a Poisson distribution $Po(\lambda)$, too, with parameter

$$\lambda = -\frac{1}{2} \log(1 - (1 - \varepsilon)^2).$$

- (2) *The number of tree components $T_{mn,k}$ with k vertices satisfy a central limit theorem of the form*

$$\frac{T_{mn,k} - \mu m}{\sqrt{\sigma^2 m}} \rightarrow N(0, 1),$$

¹This analysis is even based on using universal hash functions and thus, weaker conditions on randomness.

where

$$\mu = 2 \frac{k^{k-2}(1-\varepsilon)^{k-1}e^{k(\varepsilon-1)}}{k!}$$

and

$$\sigma^2 = \mu - \frac{2e^{2k(\varepsilon-1)}k^{2k-4}(1-\varepsilon)^{2k-3}(k^2\varepsilon^2 + k^2\varepsilon - 4k\varepsilon + 2)}{(k!)^2}.$$

Furthermore, mean and variance are asymptotically given by $\mathbb{E}T_{mn,k} = \mu m + \mathcal{O}(1)$ and $\mathbb{V}\text{ar}T_{mn,k} = \sigma^2 m + \mathcal{O}(1)$ as $m \rightarrow \infty$, respectively.

- (3) The number of vertices V_{mn}^c contained in cycles has a limiting distribution V^c with characteristic function

$$\mathbb{E}e^{isV^c} = \sqrt{\frac{1 - (1-\varepsilon)^2}{1 - e^{2is}(1-\varepsilon)^2}},$$

mean and variance satisfy

$$\lim_{m \rightarrow \infty} \mathbb{E}V_{mn}^c = \mathbb{E}V^c = \frac{(1-\varepsilon)^2}{1 - (1-\varepsilon)^2},$$

and

$$\lim_{m \rightarrow \infty} \mathbb{V}\text{ar}V_{mn}^c = \mathbb{V}\text{ar}V^c = \frac{2(1-\varepsilon)^2}{(1 - (1-\varepsilon)^2)^2}.$$

- (4) Furthermore, the expected value of the number of vertices V_{mn}^{uc} in unicyclic components is asymptotically given by

$$\lim_{m \rightarrow \infty} \mathbb{E}V_{mn}^{uc} = \frac{(1-\varepsilon)^2}{\varepsilon(1 - (1-\varepsilon)^2)},$$

and its variance by

$$\lim_{m \rightarrow \infty} \mathbb{V}\text{ar}V_{mn}^{uc} = \frac{(1-\varepsilon)^2(\varepsilon^2 - 3\varepsilon + 4)}{\varepsilon^2(1 - (1-\varepsilon)^2)^2}.$$

Note that the second statement of Theorem 3.(1) is closely related to [Kalugin 1991, Theorem 2]. Further note that the parameters of the Poisson distributions are related by

$$\sum_{k \geq 1} \frac{1}{2k} (1-\varepsilon)^{2k} = -\frac{1}{2} \log(1 - (1-\varepsilon)^2).$$

Instead of considering the “symmetric” case where the graph possesses an equal number of vertices of both types, we can also analyze asymmetric graphs (see [Kutzelnigg 2008; 2009]) or a non-bipartite version (see Section 8) and obtain similar results.

4. COUNTING SPARSE BIPARTITE GRAPHS

In this section, we establish a framework of generating functions and give the proof of the first part of Theorem 1. The construction is similar to the multigraph process in [Janson et al. 1993], but we are dealing with bipartite graphs.

We start counting all bipartite graphs without restrictions to the type of their components. Let $G_{m_1, m_2, n}$ denote the set of all vertex and edge labeled bipartite multi graphs (V_1, V_2, E) with $|V_1| = m_1$, $|V_2| = m_2$, and $|E| = n$. By definition it is clear that the number of all graphs of the family $G_{m_1, m_2, n}$ equals

$$\#G_{m_1, m_2, n} = m_1^n m_2^n. \quad (2)$$

In particular, we are interested in the case $m_1 = m_2 = m$ and $n = \lfloor (1 - \varepsilon)m \rfloor$, where $\varepsilon \in (0, 1)$. This means that the graph is relatively sparse. However, for technical reasons it is necessary to take all possible bipartite graphs into account.

Next, let $G_{m_1, m_2, n}^\circ$ denote those graphs in $G_{m_1, m_2, n}$ without complex components, that is, all components are either trees or unicyclic. Further,

$$g^\circ(x, y, v) = \sum_{m_1, m_2, n} \#G_{m_1, m_2, n}^\circ \frac{x^{m_1} y^{m_2} v^n}{m_1! m_2! n!}$$

denotes the corresponding generating function. Our next goal is to describe this generating function. For this purpose we will first consider bipartite trees.

We call a tree bipartite if the vertices are partitioned into two classes V_1 (“black” vertices) and V_2 (“white” vertices) such that no vertex has a neighbor of the same class. They are called labeled if the vertices of type 1, that is vertices in V_1 , are labeled by $1, 2, \dots, |V_1|$ and the vertices of type 2 are labeled by $1, 2, \dots, |V_2|$.

Let T_1 denote the set of bipartite rooted trees, where the root is contained in V_1 , T_2 the set of bipartite rooted trees, where the root is contained in V_2 , and \tilde{T} the class of unrooted bipartite trees. Furthermore, let t_{1, m_1, m_2} and t_{2, m_1, m_2} denote the number of trees in T_1 and T_2 , respectively, with m_1 vertices of type of type 1 and m_2 of type 2. Similarly we define \tilde{t}_{m_1, m_2} . The corresponding generating functions are defined by

$$t_1(x, y) = \sum_{m_1, m_2 \geq 0} t_{1, m_1, m_2} \frac{x^{m_1} y^{m_2}}{m_1! m_2!},$$

$$t_2(x, y) = \sum_{m_1, m_2 \geq 0} t_{2, m_1, m_2} \frac{x^{m_1} y^{m_2}}{m_1! m_2!},$$

and by

$$\tilde{t}(x, y) = \sum_{m_1, m_2 \geq 0} \tilde{t}_{m_1, m_2} \frac{x^{m_1} y^{m_2}}{m_1! m_2!}.$$

LEMMA 1. *The generating functions $t_1(x, y)$, $t_2(x, y)$, and $\tilde{t}(x, y)$ are given by*

$$t_1(x, y) = x e^{t_2(x, y)}, \quad t_2(x, y) = y e^{t_1(x, y)} \quad (3)$$

and by

$$\tilde{t}(x, y) = t_1(x, y) + t_2(x, y) - t_1(x, y)t_2(x, y).$$

Furthermore we have

$$t_{1, m_1, m_2} = m_1^{m_2} m_2^{m_1 - 1}, \quad t_{2, m_1, m_2} = m_1^{m_2 - 1} m_2^{m_1}$$

and

$$\tilde{t}_{m_1, m_2} = m_1^{m_2 - 1} m_2^{m_1 - 1}.$$

The explicit formula for \tilde{t}_{m_1, m_2} is originally due to [Scoins 1962].

Note that $t_1(x, y) = t_2(y, x)$ and that $t_1(x, x)$ equals the usual tree function $t(x) = \sum_{n \geq 1} n^{n-1} x^n / n!$ that is given by $t(x) = x e^{t(x)}$. Thus, $t_1(x, y)$ and $t_2(x, y)$ are surely analytic functions for $|x| < e^{-1}$ and $|y| < e^{-1}$. This is due to the fact that the radius of convergence of $t(x)$ equals $1/e$.

The partial derivatives of the functions $\tilde{t}(x, y)$ and $t_1(x, y)$ are given by

$$\frac{\partial}{\partial x} \tilde{t}(x, y) = \frac{t_1(x, y)}{x}, \quad \frac{\partial}{\partial y} \tilde{t}(x, y) = \frac{t_2(x, y)}{y},$$

and

$$\frac{\partial}{\partial x} t_1(x, y) = \frac{t_1(x, y)}{x(1 - t_1(x, y)t_2(x, y))}, \quad \frac{\partial}{\partial y} t_1(x, y) = \frac{t_1(x, y)t_2(x, y)}{y(1 - t_1(x, y)t_2(x, y))}.$$

Further, the generating function of (usual) unrooted labeled trees is given by $t(x) - t(x)^2/2$. The relation (1) is a generalization of this result.

PROOF. The functional equations (3) are obvious by their recursive description. Next, note that the partial derivations of $\tilde{t}(x, y)$ and $t_1(x, y) + t_2(x, y) - t_1(x, y)t_2(x, y)$ are equal, thus (1) holds.

There is also a combinatorial interpretation of (1). Consider a rooted tree, possessing a black root labeled by 1, as an unrooted tree. Next, examine an unordered pair (t_1, t_2) of trees from $T_1 \times T_2$, and join the roots by an edge. If the black vertex labeled by 1 is contained in t_1 , consider the root of t_2 as new root, and we obtain a tree possessing a white root and at least one black vertex. Otherwise, consider the root of t_1 as new root, and we obtain a tree with a black vertex not labeled by 1.

Lagrange inversion applied to $t_1(x, y) = x \exp(y e^{t_1(x, y)})$ yields:

$$\begin{aligned} [x^{m_1} y^{m_2}] t_1(x, y) &= [y^{m_2}] \frac{1}{m_1} [u^{m_1-1}] \left(e^{y e^u} \right)^{m_1} \\ &= [y^{m_2}] \frac{1}{m_1} [u^{m_1-1}] \sum_{k \geq 0} \sum_{l \geq 0} \frac{m_1^k y^k}{k!} \frac{u^l l^l}{l!} \\ &= [y^{m_2}] \frac{1}{m_1} \sum_{k \geq 0} \frac{m_1^k y^k}{k!} \frac{k^{m_1-1}}{(m_1-1)!} = \frac{m_1^{m_2} m_2^{m_1-1}}{m_1! m_2!} \end{aligned}$$

Furthermore, $\tilde{t}_{m_1, m_2} = t_{1, m_1, m_2} / m_1 = t_{2, m_1, m_2} / m_2$ since there are exactly m_1 ways to choose the root of type 1 in an unrooted tree with m_1 vertices of type 1. \square

With help of these functions, we can describe the generating function $g^\circ(x, y, v)$.

LEMMA 2. *The generating function $g^\circ(x, y, v)$ is given by*

$$g^\circ(x, y, v) = \frac{e^{\frac{1}{v} \tilde{t}(xv, yv)}}{\sqrt{1 - t_1(xv, yv)t_2(xv, yv)}}.$$

PROOF. We have to count graphs where each component is either an unrooted tree (that is counted by $\tilde{t}(x, y)$) or a graph with exactly one cycle.

Of course, a cycle has to have an even number of vertices (say $2k$), where k vertices are black and the other k vertices are white. A cyclic vertex of black color can be considered as the root of a rooted tree of type 1 and similarly, a white cyclic

vertex can be considered as the root of a rooted tree of type 2. Note that we have to divide the product of the generating functions $t_1(x, y)^k t_2(x, y)^k$ by $2k$ to account for cyclic order and change of orientation. Hence, the corresponding generating functions of a unicyclic graph with $2k$ cyclic points is given by

$$\frac{1}{2k} t_1(x, y)^k t_2(x, y)^k.$$

Consequently the generating function of a connected graph with exactly one cycle is given by

$$c(x, y) = \sum_{k \geq 1} \frac{1}{2k} t_1(x, y)^k t_2(x, y)^k = \frac{1}{2} \log \frac{1}{1 - t_1(x, y) t_2(x, y)}.$$

Since a cyclic component of size $m_1 + m_2$ has exactly the same number of edges and since there are $(m_1 + m_2)!$ possible labels, the corresponding generating function that takes the number of edges into account is given by $c(xv, yv)$.

Similarly, a tree of size $m_1 + m_2$ has exactly $n = m_1 + m_2 - 1$ edges. Consequently the generating function $\tilde{t}(xv, yv)/v$ corresponds to a bipartite unrooted tree.

Finally the generating function $g^\circ(x, y, v)$ is given by

$$g^\circ(x, y, v) = e^{\frac{1}{v} \tilde{t}(xv, yv) + c(xv, yv)} = \frac{e^{\frac{1}{v} \tilde{t}(xv, yv)}}{\sqrt{1 - t_1(xv, yv) t_2(xv, yv)}},$$

which completes the proof of the lemma. \square

COROLLARY 1. *The number of graphs $\#G_{m_1, m_2, n}^\circ$ is given by*

$$\#G_{m_1, m_2, n}^\circ = \frac{m_1! m_2! n!}{(m_1 + m_2 - n)!} [x^{m_1} y^{m_2}] \frac{\tilde{t}(x, y)^{m_1 + m_2 - n}}{\sqrt{1 - t_1(x, y) t_2(x, y)}}. \quad (4)$$

Hence, by Cauchy's formula

$$\#G_{m, m, n}^\circ = \frac{-(m!)^2 n!}{4\pi^2 (2m - n)!} \oint \oint \frac{\tilde{t}(x, y)^{2m - n}}{\sqrt{1 - t_1(x, y) t_2(x, y)}} \frac{dx}{x^{m+1}} \frac{dy}{y^{m+1}}. \quad (5)$$

This is in fact an integral that can be asymptotically evaluated with help of a double saddle point method, see Lemma 3.

LEMMA 3. *Let $f(x, y)$ and $g(x, y)$ be analytic functions locally around $(x, y) = (0, 0)$ such that all coefficients $[x^{m_1} y^{m_2}] f(x, y)$ and $[x^{m_1} y^{m_2}] g(x, y)$ are non negative and that there exists M such that all indices (m_1, m_2) with $m_1, m_2 \geq M$ can be represented as a finite linear combination of the set $\{(m_1, m_2) | [x^{m_1} y^{m_2}] f(x, y) > 0\}$ with positive integers as coefficients.*

Let R_1 and R_2 be compact intervals of the positive real line such that $R = R_1 \times R_2$ is contained in the regions of convergence of $f(x, y)$ and $g(x, y)$. Furthermore set

$$S = \left\{ \left(\frac{x}{f(x, y)} \frac{\partial}{\partial x} f(x, y), \frac{y}{f(x, y)} \frac{\partial}{\partial y} f(x, y) \right) : (x, y) \in R \right\}.$$

Then we have

$$[x^{m_1} y^{m_2}] g(x, y) f(x, y)^k = \frac{g(x_0, y_0) f(x_0, y_0)^k}{2\pi x_0^{m_1} y_0^{m_2} k \sqrt{\Delta}} \left(1 + \frac{H}{24\Delta^3} \frac{1}{k} + \mathcal{O}\left(\frac{1}{k^2}\right) \right),$$

uniformly for $(m_1/k, m_2/k) \in S$, where x_0 and y_0 are uniquely determined by

$$\frac{m_1}{k} = \frac{x_0}{f(x_0, y_0)} \left[\frac{\partial}{\partial x} f(x, y) \right]_{(x_0, y_0)}, \quad \frac{m_2}{k} = \frac{y_0}{f(x_0, y_0)} \left[\frac{\partial}{\partial y} f(x, y) \right]_{(x_0, y_0)}$$

and the constants Δ and H are given in the following way: Let κ_{ij} and $\bar{\kappa}_{ij}$ be the cummulants

$$\kappa_{ij} = \left[\frac{\partial^i}{\partial u^i} \frac{\partial^j}{\partial v^j} \log f(x_0 e^u, y_0 e^v) \right]_{(0,0)}, \quad \bar{\kappa}_{ij} = \left[\frac{\partial^i}{\partial u^i} \frac{\partial^j}{\partial v^j} \log g(x_0 e^u, y_0 e^v) \right]_{(0,0)}.$$

Then $\Delta = \kappa_{20}\kappa_{02} - \kappa_{11}^2$ holds and H is given by

$$H = \alpha + \beta + \hat{\beta} + \gamma\bar{\kappa}_{10} + \hat{\gamma}\bar{\kappa}_{01} + \delta\bar{\kappa}_{10}\bar{\kappa}_{01} + \eta\bar{\kappa}_{10}^2 + \hat{\eta}\bar{\kappa}_{01}^2 + 4\eta\bar{\kappa}_{20} + 4\hat{\eta}\bar{\kappa}_{02} + 4\delta\bar{\kappa}_{11},$$

where

$$\begin{aligned} \alpha &= 54\kappa_{21}\kappa_{11}\kappa_{12}\kappa_{20}\kappa_{02} + 6\kappa_{22}\kappa_{20}\kappa_{02}\kappa_{11}^2 - 12\kappa_{22}\kappa_{11}^4 + 4\kappa_{03}\kappa_{11}^3\kappa_{30} \\ &\quad + 36\kappa_{21}\kappa_{11}^3\kappa_{12} + 6\kappa_{22}\kappa_{20}^2\kappa_{02}^2 + 6\kappa_{03}\kappa_{11}\kappa_{30}\kappa_{20}\kappa_{02}, \\ \beta &= -5\kappa_{02}^3\kappa_{30}^2 + 30\kappa_{02}^2\kappa_{30}\kappa_{11}\kappa_{21} - 24\kappa_{02}\kappa_{30}\kappa_{12}\kappa_{11}^2 - 6\kappa_{02}^2\kappa_{30}\kappa_{12}\kappa_{20} \\ &\quad - 12\kappa_{11}\kappa_{02}^2\kappa_{31}\kappa_{20} - 36\kappa_{02}\kappa_{21}^2\kappa_{11}^2 - 9\kappa_{02}^2\kappa_{21}^2\kappa_{20} + 3\kappa_{02}^3\kappa_{40}\kappa_{20} \\ &\quad - 3\kappa_{02}^2\kappa_{40}\kappa_{11}^2 + 12\kappa_{11}^3\kappa_{02}\kappa_{31}, \\ \gamma &= 12\Delta (\kappa_{02}^2\kappa_{30} - \kappa_{11}\kappa_{20}\kappa_{03} - 3\kappa_{21}\kappa_{11}\kappa_{02} + \kappa_{12}\kappa_{11}^2 + \kappa_{12}(\kappa_{02}\kappa_{20} + \kappa_{11}^2)), \\ \delta &= 24\Delta(\kappa_{11}\kappa_{20}\kappa_{02} - \kappa_{11}^3), \\ \eta &= 12\Delta(\kappa_{02}\kappa_{11}^2 - \kappa_{02}^2\kappa_{20}), \end{aligned}$$

and $\hat{\cdot}$ indicates to replace all functions of type κ_{ij} by κ_{ji} .

Lemma 3 is a generalization of a result of [Good 1957], where $g(x, y) = 1$. Its proof of is given in the Appendix.

We fix $\varepsilon > 0$ and consider the sequence of integer pairs $(m, n) = (m, \lfloor (1 - \varepsilon)m \rfloor)$. For technical reasons we also define the ratio

$$\varepsilon' = \varepsilon'_m = \frac{m - n}{m} = 1 - \frac{\lfloor (1 - \varepsilon)m \rfloor}{m} = \varepsilon + \mathcal{O}(m^{-1})$$

which is always very close to ε . By Stirling's formula

$$n! = \frac{n^n}{e^n} \sqrt{2\pi n} \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \mathcal{O}\left(\frac{1}{n^3}\right) \right)$$

we obtain the asymptotic expansion

$$\frac{(m!)^2 n!}{(2m - n)!} = \frac{2\pi m m^{2m} n^n}{e^{2n} (2m - n)^{2m - n}} \sqrt{\frac{n}{2m - n}} \left(1 + \frac{1 + \varepsilon' - \varepsilon'^2}{6(1 - \varepsilon')^2 m} + \mathcal{O}\left(\frac{1}{m^2}\right) \right). \quad (6)$$

For our problem, it turns out that the saddle point is given by

$$x_0 = y_0 = \frac{n}{m} e^{-\frac{n}{m}} = (1 - \varepsilon') e^{\varepsilon' - 1} < \frac{1}{e}.$$

This can be easily checked. By symmetry it is clear that $x_0 = y_0$. Further, $t_1(x, x) = t(x) = x e^{t(x)}$ equals the tree function. Hence we get

$$t_1(x_0, x_0) = 1 - \varepsilon' = \frac{n}{m}, \quad \tilde{t}(x_0, x_0) = 1 - \varepsilon'^2 = \frac{n}{m} \left(2 - \frac{n}{m} \right).$$

For instance, we further obtain

$$\begin{aligned}\kappa_{20} &= \frac{t_1(x_0, y_0)}{(1 - t_1(x_0, y_0)t_2(x_0, y_0))\tilde{t}(x_0, y_0)} - \frac{t_1(x_0, y_0)^2}{\tilde{t}(x_0, y_0)^2} = \frac{\varepsilon'^2 - \varepsilon' + 1}{(\varepsilon' + 1)^2 \varepsilon' (2 - \varepsilon')}, \\ \kappa_{11} &= \frac{t_1(x_0, y_0)t_2(x_0, y_0)}{(1 - t_1(x_0, y_0)t_2(x_0, y_0))\tilde{t}(x_0, y_0)} - \frac{t_1(x_0, y_0)t_2(x_0, y_0)}{\tilde{t}(x_0, y_0)^2} = \frac{1 - 2\varepsilon'}{(\varepsilon' + 1)^2 \varepsilon' (2 - \varepsilon')},\end{aligned}$$

and

$$\Delta = \frac{1 - \varepsilon'}{\varepsilon'(2 - \varepsilon')(\varepsilon' + 1)^3} = \frac{m^4 n}{(2m - n)^3 (m^2 - n^2)}.$$

The other cummulants can be calculated in the same way, but have been computed with help of a computer algebra system in a half-automatic way.

We set $f = \tilde{t}$, $g = 1/\sqrt{1 - t_1 t_2}$, $k = 2m - n$, $m_1 = m$, and $m_2 = m$, in Lemma 3 and obtain the leading coefficient

$$\frac{\tilde{t}(x_0, y_0)^{2m-n}}{2\pi(2m-n)x_0^m y_0^m \sqrt{1 - t_1(x_0, y_0)t_2(x_0, y_0)}\sqrt{\Delta}} = \frac{e^{2n} m^{2n+1} (2m-2)^{2m-n-1}}{2\pi n^n m^{2m} \sqrt{\Delta} \sqrt{m^2 - n^2}},$$

and

$$H = \frac{\varepsilon'^6 - 10\varepsilon'^5 + 21\varepsilon'^4 - 2\varepsilon'^3 - 27\varepsilon'^2 + 20\varepsilon' - 5}{12\varepsilon'^3(-2 + \varepsilon')^2(1 - \varepsilon')}.$$

Combining these results with (6) we obtain

$$\begin{aligned}\#G_{m,m,n}^\circ &= m^{2n} \left(1 + \frac{H}{(1 + \varepsilon')m} + \frac{1 + \varepsilon' - \varepsilon'^2}{6(1 - \varepsilon')^2 m} + \mathcal{O}\left(\frac{1}{m^2}\right) \right) \\ &= m^{2n} \left(1 - \frac{1}{m} \frac{(2\varepsilon'^2 - 5\varepsilon' + 5)(1 - \varepsilon')^3}{12(2 - \varepsilon')^2 \varepsilon'^3} + \mathcal{O}\left(\frac{1}{m^2}\right) \right)\end{aligned}$$

Finally we can safely replace ε' by $\varepsilon = \varepsilon' + \mathcal{O}(m^{-1})$ without changing the expansion. All changes go into the error term $\mathcal{O}(m^{-2})$.

Hence, if $p(n, m)$ denotes the probability, that every component of the cuckoo graph is either a tree or unicyclic, after the insertion of n edges then we have

$$p(n, m) = \frac{\#G_{m,m,[(1-\varepsilon)m]}^\circ}{\#G_{m,m,[(1-\varepsilon)m]}} = 1 - \frac{1}{m} \frac{(2\varepsilon^2 - 5\varepsilon + 5)(1 - \varepsilon)^3}{12(2 - \varepsilon)^2 \varepsilon^3} + \mathcal{O}\left(\frac{1}{m^2}\right),$$

which completes the proof of the first part of Theorem 1.

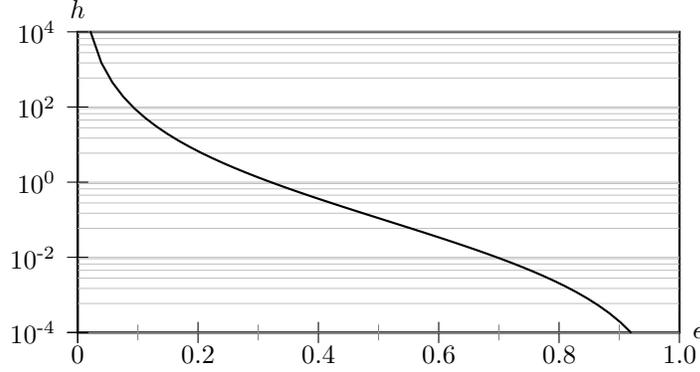
Figure 2 displays the graph of $h(\varepsilon) = (2\varepsilon^2 - 5\varepsilon + 5)(1 - \varepsilon)^3 / (12(2 - \varepsilon)^2 \varepsilon^3)$. Note that $h(\varepsilon)$ expands as

$$h(\varepsilon) = \frac{5}{48}\varepsilon^{-3} - \frac{5}{16}\varepsilon^{-2} + \frac{21}{64}\varepsilon^{-1} - \frac{13}{96} + \frac{3}{256}\varepsilon + \frac{1}{256}\varepsilon^2 + \frac{1}{1024}\varepsilon^3 + \mathcal{O}(\varepsilon^4) \quad (7)$$

if $\varepsilon \rightarrow 0$.

We want to note that it is also possible to obtain a little bit more precise asymptotic expansion for

$$p(n, m) = 1 - \frac{h(\varepsilon')}{m} - \frac{\hat{h}(\varepsilon')}{m^2} + \mathcal{O}\left(\frac{1}{m^3}\right),$$

Fig. 2. The graph of $h(\varepsilon)$.

where $\tilde{h}(\cdot)$ is again explicit. This can be done by refining the calculations related to Lemma 3.

For example, we can apply these expansions in order to obtain asymptotic representations for the probability $q(n+1, m)$ that the insertion of the $n+1$ -st edge creates a bicyclic component, conditioned on the property, that the first n insertions did not create such a component.

LEMMA 4. *The probability that the insertion of the $n+1$ -st inserted key forces a rehash is given by*

$$q(n+1, m) = -\frac{h'(1 - \frac{n}{m})}{m^2} + \mathcal{O}\left(\frac{1}{m^3}\right).$$

This is uniform for $n \leq (1 - \eta)m$, assuming $\eta > 0$.

PROOF. By definition we have $p(n+1, m) = (1 - q(n+1, m))p(n, m)$. Hence we get

$$\begin{aligned} q(n+1, m) &= \frac{p(n, m) - p(n+1, m)}{p(n, m)} \\ &= \left(\frac{h(\varepsilon')}{m} - \frac{h(\varepsilon' - \frac{1}{m})}{m} + \frac{\hat{h}(\varepsilon')}{m^2} - \frac{\hat{h}(\varepsilon' - \frac{1}{m})}{m^2} + \mathcal{O}\left(\frac{1}{m^3}\right) \right) \\ &\quad \times \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right) \\ &= -\frac{h'(\varepsilon')}{m^2} - \frac{\hat{h}'(\varepsilon')}{m^3} + \mathcal{O}\left(\frac{1}{m^3}\right) \\ &= \frac{(1 - \varepsilon')^2(-\varepsilon'^3 + 8\varepsilon'^2 - 15\varepsilon' + 10)}{4(2 - \varepsilon')^3\varepsilon'^4} \frac{1}{m^2} + \mathcal{O}\left(\frac{1}{m^3}\right). \quad \square \end{aligned}$$

5. THE “CRITICAL CASE”

It is also interesting to consider the case $\varepsilon \rightarrow 0$, that is, $n/m \rightarrow 1$. This case is more delicate since limiting the saddle point $x_0 = 1/e$ coalesces with the singularity of

the denominator. (Note that $t_1(1/e, 1/e) = 1$.) Hence we expect a phase transition where the singularity behavior of the denominator gets more and more important. This is definitely an involved analytic problem and not easy to handle. In particular, one has to handle the singularity structure of $t_1(x, y)t_2(x, y)$ around $x = 1/e$ and $y = 1/e$, which is surely feasible, but the choice of the double contour integral is not clear.

For the sake of shortness we will only work out the limiting case $\varepsilon = 0$, that is, $m = n$. Even in this case we do not work directly with the representation (4) but apply Lagrange's inversion formula first. In particular, we use the fact that $t_1(x, y)$ satisfies the equation $t_1 = x \exp(ye^{t_1})$. This leads (again) to a saddle point integral, where the denominator is explicit and does not contain implicitly defined functions as before.

LEMMA 5. *We have the identity*

$$\#G_{m,m,m}^\circ = (m!)^2 \sum_{k \geq 0} \binom{2k}{k} \frac{1}{4^k} [x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k.$$

PROOF. We use Corollary 1 and the series expansion

$$\frac{1}{\sqrt{1-z}} = \sum_{k \geq 0} \binom{2k}{k} \frac{1}{4^k} z^k. \quad \square$$

Further, we define the functions

$$\begin{aligned} f(u, y) &= (u + ye^u(1-u)) \exp(ye^u), \\ l(u, y) &= uye^u, \end{aligned}$$

and

$$h(u, y) = h_k(u, y) = u \frac{mu - mye^u u^2 + ku + kye^u + ku^2 - ku^2 ye^u}{u(u + ye^u(1-u))}.$$

LEMMA 6. *The following identity holds:*

$$[x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k = \frac{1}{m} [u^m y^m] f(u, y)^m l(u, y)^k h_k(u, y).$$

PROOF. Set $\phi(u, y) = \exp(ye^u)$. Then the generating function $t_1(x, y)$ satisfies the relation $t_1(x, y) = x\phi(t_1(x, y), y)$. Furthermore with

$$g(u, y) = (u + ye^u(1-u))^m u^k (ye^u)^k,$$

we have

$$\tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k = g(t_1(x, y), y).$$

Note that $\phi(0, y) \neq 0$, so that we can apply Lagrange's Inversion Theorem and obtain:

$$\begin{aligned} [x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k &= [x^m y^m] g(t_1(x, y), y) \\ &= [u^{m-1} y^m] \frac{1}{m} \phi(u, y)^m \frac{\partial}{\partial u} g(u, y) \\ &= [u^{m-1} y^m] \frac{1}{m} (\exp(ye^u))^m \frac{\partial}{\partial u} \left((u + ye^u(1-u))^m u^k (ye^u)^k \right) \end{aligned}$$

Differentiation and simplification completes the proof. \square

Obviously, the coefficient $[u^m y^m] f(u, y)^m l(u, y)^k h(u, y)$ equals zero if k is greater than m . Furthermore, for moderate k the saddle point method is applicable to obtain asymptotics for this coefficient. However, for relatively large k there is an easy upper bound.

LEMMA 7. *Assume that $k \geq m^{\frac{1}{3} + \xi}$ is satisfied for a positive ξ . Then, there exists a positive constant c such that*

$$[x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k = \mathcal{O}\left(e^{2m - cm^\xi}\right)$$

holds.

PROOF. Note that the (trivial) bound

$$[x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k \leq r^{-2m} \tilde{t}(r, r)^m t_1(r, r)^k t_2(r, r)^k$$

holds for all r satisfying $0 < r < 1/e$. We set $r = (1 - \eta)/e$. Recall that $t_1(x, x)$ equals the usual tree function $t(x)$ and that $\tilde{t}(x, x) = 2t(x) - t^2(x)$ holds. Further, the singular expansion of $t(x)$ around its singularity $1/e$ is well known to be (cf. [Flajolet and Sedgewick 2009])

$$t(x) = 1 - \sqrt{2}\sqrt{1 - ex} + \frac{2}{3}(1 - ex) - \frac{11}{18\sqrt{2}}(1 - ex)^{\frac{3}{2}} + \mathcal{O}((1 - ex)^2).$$

Thus, we obtain the inequality

$$[x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k \leq e^{2m + \frac{4}{3}\sqrt{2}m\eta^{\frac{3}{2}} - 2\sqrt{2}\eta^{\frac{1}{2}}k + \mathcal{O}(m\eta^2) + \mathcal{O}(\eta k)}.$$

Setting $\eta = m^{-2/3}$ allows us to compute the claimed bound. \square

LEMMA 8. *Assume that $k = \mathcal{O}(m^{1/3 + \xi})$ holds. Then, there exists a positive real constant c such that*

$$\begin{aligned} & \frac{1}{m} [u^m y^m] f(u, y)^m l(u, y)^k h_k(u, y) \\ &= \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \int_0^\infty s e^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} \sin\left(\sqrt{3}ksm^{-\frac{1}{3}} + \frac{\pi}{3}\right) ds \\ &+ \mathcal{O}\left(e^{2m - cm^{-\frac{1}{12}}}\right) + \mathcal{O}\left(m^{-\frac{7}{6} - \frac{1}{24}} e^{2m} \int_0^\infty s e^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} ds\right) \quad (8) \end{aligned}$$

holds.

The proof of Lemma 8 is given in the Appendix.

Interestingly the integral appearing in Lemma 8 is closely related to the Lommel function²

$$S_{1,2/3}(x) = 1 + \left(\frac{2}{x}\right)^{\frac{2}{3}} \int_0^\infty \exp\left(-u^3 - 3\left(\frac{x}{2}\right)^{\frac{2}{3}} u\right) u du.$$

²The Lommel functions $s_{\mu,\nu}(x)$ and $S_{\mu,\nu}(x)$ (of first and second kind) are solutions of the inhomogeneous Bessel differential equation $x^2 y'' + xy' + (x^2 - \nu^2)y = x^{\mu+1}$. (cf. [Zwillinger 1992])

In particular the main term in (8) can be rewritten to

$$\begin{aligned} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \int_0^\infty s e^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} \sin\left(\sqrt{3}ksm^{-\frac{1}{3}} + \frac{\pi}{3}\right) ds \\ = \frac{\sqrt{2\pi}}{2\pi^2} \frac{kie^{2m}}{m^{\frac{3}{2}}} \left(S_{1, \frac{2}{3}} \left(i \frac{4k^{\frac{3}{2}}}{3\sqrt{m}} \right) - S_{1, \frac{2}{3}} \left(-i \frac{4k^{\frac{3}{2}}}{3\sqrt{m}} \right) \right). \end{aligned}$$

With the help of the above results we are now able to calculate the number of graphs without complex components. By combining Lemma 5 and Lemma 6 we obtain

$$\#G_{m,m,m}^{\circ} = \frac{(m!)^2}{m} \sum_{k=0}^{m-1} \binom{2k}{k} \frac{1}{4^k} [u^m y^m] f(u, y)^m l(u, y)^k h(u, y) = (m!)^2 (S_1 + S_2),$$

where the sums S_1 and S_2 collect moderate $k \leq m^{\frac{1}{3}+\epsilon}$ and large k , respectively; see below. We start with an upper bound of S_2 , which follows from Lemma 7:

$$S_2 = \frac{1}{m} \sum_{k=\lceil m^{\frac{1}{3}+\epsilon} \rceil}^{m-1} \binom{2k}{k} \frac{1}{4^k} [u^m y^m] f(u, y)^m l(u, y)^k h(u, y) \leq m \mathcal{O}\left(e^{2m-cm^\epsilon}\right).$$

The sum S_1 is more difficult to handle. First, we use the result of Lemma 8 and split it up into four terms:

$$\begin{aligned} S_1 &= \frac{1}{m} \sum_{k=0}^{\lceil m^{\frac{1}{3}+\epsilon} \rceil} \binom{2k}{k} \frac{1}{4^k} [u^m y^m] f(u, y)^m l(u, y)^k h(u, y) \\ &= e^{-\frac{\pi i}{6}} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \sum_{k=0}^{\lceil m^{\frac{1}{3}+\epsilon} \rceil} \binom{2k}{k} \frac{1}{4^k} \int_0^\infty s e^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}} ks} ds \\ &\quad + e^{\frac{\pi i}{6}} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \sum_{k=0}^{\lceil m^{\frac{1}{3}+\epsilon} \rceil} \binom{2k}{k} \frac{1}{4^k} \int_0^\infty s e^{-\frac{2}{3}s^3 + \frac{2\bar{\zeta}}{\sqrt[3]{m}} ks} ds \\ &\quad + \sum_{k=0}^{\lceil m^{\frac{1}{3}+\epsilon} \rceil} \binom{2k}{k} \frac{1}{4^k} \mathcal{O}\left(e^{2m-cm^{-\frac{1}{12}}}\right) \\ &\quad + \sum_{k=0}^{\lceil m^{\frac{1}{3}+\epsilon} \rceil} \binom{2k}{k} \frac{1}{4^k} \mathcal{O}\left(m^{-\frac{7}{6}-\frac{1}{24}} e^{2m} \int_0^\infty s e^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} ds\right) \\ &= S_{11} + S_{12} + S_{13} + S_{14}. \end{aligned}$$

The integral term S_{11} is split up into three parts $[0, \infty) = [0, m^{-\delta}) \cup [m^{-\delta}, m^{\frac{1}{3}-\gamma}) \cup$

$[m^{\frac{1}{3}-\gamma}, \infty)$:

$$\begin{aligned}
S_{11} &= e^{-\frac{\pi i}{6}} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} \int_0^\infty s e^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} ds \\
&= e^{-\frac{\pi i}{6}} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} \int_0^\infty s e^{-\frac{2}{3}s^3} \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} e^{\frac{2\zeta}{\sqrt[3]{m}}ks} ds \\
&= e^{-\frac{\pi i}{6}} \frac{\sqrt{2\pi}}{2\pi^2} \frac{e^{2m}}{m^{\frac{7}{6}}} (I_1 + I_2 + I_3).
\end{aligned}$$

For the first part we have the upper bound

$$\begin{aligned}
|I_1| &= \left| \int_0^{m^{-\delta}} s e^{-\frac{2}{3}s^3} \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} e^{\frac{2\zeta}{\sqrt[3]{m}}ks} ds \right| \\
&\leq \int_0^{m^{-\delta}} s e^{-\frac{2}{3}s^3} \sum_{k=0}^\infty \binom{2k}{k} \frac{1}{4^k} e^{-\frac{1}{\sqrt[3]{m}}ks} ds \\
&= \int_0^{m^{-\delta}} s e^{-\frac{2}{3}s^3} \left(1 - e^{-\frac{s}{\sqrt[3]{m}}}\right)^{-\frac{1}{2}} ds = \mathcal{O}\left(m^{\frac{1}{6}-\delta}\right),
\end{aligned}$$

where we have used the inequality $1/\sqrt{1-e^{-x}} \leq 1 + 1/\sqrt{x}$. This inequality is also useful for bounding I_3 :

$$\begin{aligned}
|I_3| &= \left| \int_{m^{\frac{1}{3}-\gamma}}^\infty s e^{-\frac{2}{3}s^3} \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} e^{\frac{2\zeta}{\sqrt[3]{m}}ks} ds \right| \\
&\leq \int_{m^{\frac{1}{3}-\gamma}}^\infty s e^{-\frac{2}{3}s^3} \left(1 - e^{-\frac{s}{\sqrt[3]{m}}}\right)^{-\frac{1}{2}} ds \\
&\leq m^{\frac{1}{6}} \int_{m^{\frac{1}{3}-\gamma}}^\infty \sqrt{s} e^{-\frac{2}{3}s^3} ds + \int_{m^{\frac{1}{3}-\gamma}}^\infty s^2 e^{-\frac{2}{3}s^3} ds = m^{\frac{1}{6}} \mathcal{O}\left(e^{-cm^{1-3\gamma}}\right).
\end{aligned}$$

The integral part I_2 provides the main contribution:

$$\begin{aligned}
I_2 &= \int_{m^{-\delta}}^{m^{\frac{1}{3}-\gamma}} s e^{-\frac{2}{3}s^3} \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} e^{\frac{2\zeta}{\sqrt[3]{m}}ks} ds \\
&= \int_{m^{-\delta}}^{m^{\frac{1}{3}-\gamma}} s e^{-\frac{2}{3}s^3} \left(\sum_{k=0}^\infty \binom{2k}{k} \frac{1}{4^k} e^{\frac{2\zeta}{\sqrt[3]{m}}ks} + \mathcal{O}\left(e^{-m^{\xi-\delta}}\right) \right) ds
\end{aligned}$$

$$\begin{aligned}
 &= \int_{m^{-\delta}}^{m^{\frac{1}{3}-\gamma}} se^{-\frac{2}{3}s^3} \left(1 - e^{-\frac{2c}{\sqrt[3]{m}}s}\right)^{-\frac{1}{2}} ds + \mathcal{O}\left(e^{-m^{\xi-\delta}}\right) \\
 &= \int_{m^{-\delta}}^{m^{\frac{1}{3}-\gamma}} se^{-\frac{2}{3}s^3} \left(\frac{m^{\frac{1}{6}}e^{\frac{\pi i}{6}}}{\sqrt{2s}} + \mathcal{O}\left(m^{-\frac{\gamma}{2}}\right)\right) ds + \mathcal{O}\left(e^{-m^{\xi-\delta}}\right) \\
 &= \frac{m^{\frac{1}{6}}e^{\frac{\pi i}{6}}}{\sqrt{2}} \int_0^{\infty} \sqrt{s}e^{-\frac{2}{3}s^3} ds + \mathcal{O}\left(m^{-\frac{\gamma}{2}}\right) + \mathcal{O}\left(e^{-m^{\xi-\delta}}\right) \\
 &\quad + \mathcal{O}\left(m^{\frac{1}{6}-\delta}\right) + m^{\frac{1}{6}}\mathcal{O}\left(e^{-cm^{1-3\gamma}}\right) \\
 &= e^{\frac{\pi i}{6}} \frac{m^{\frac{1}{6}}\sqrt{\pi}}{2\sqrt{3}} + \mathcal{O}\left(e^{-m^{\xi-\delta}}\right) + \mathcal{O}\left(m^{\frac{1}{6}-\delta}\right) + m^{\frac{1}{6}}\mathcal{O}\left(e^{-cm^{1-3\gamma}}\right).
 \end{aligned}$$

For instance, we may set $\xi = \frac{1}{12}$, $\delta = \frac{1}{24}$, and $\gamma = \frac{1}{12}$. Thus, we finally obtain:

$$S_{11} = \frac{\sqrt{2}e^{2m}}{4\pi\sqrt{3}m} (1 + \mathcal{O}(m^{-\delta})).$$

The second sum can be handled in the same way. In particular we obtain the same result:

$$S_{12} = \frac{\sqrt{2}e^{2m}}{4\pi\sqrt{3}m} (1 + \mathcal{O}(m^{-\delta})).$$

What is now still missing, are bounds for the remaining sums. These can be straightforwardly attained for S_{13} :

$$S_{13} = \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} \mathcal{O}\left(e^{2m-cm^{-\frac{1}{12}}}\right) = \mathcal{O}\left(m^{\frac{1}{3}+\xi} e^{2m-cm^{-\frac{1}{12}}}\right).$$

The sum S_{14} is a bit more complicated to handle, but we can proceed as in the calculation of S_{11} :

$$S_{14} = \sum_{k=0}^{\lfloor m^{\frac{1}{3}+\xi} \rfloor} \binom{2k}{k} \frac{1}{4^k} \mathcal{O}\left(m^{-\frac{7}{6}-\frac{1}{24}} e^{2m} \int_0^{\infty} se^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}}\right) ds = \mathcal{O}\left(e^{2m} m^{-1-\frac{1}{24}}\right).$$

Putting these results together, we finally obtain the equation

$$\#G_{m,m,m}^{\circ} = (m!)^2 \frac{\sqrt{2}e^{2m}}{2\pi\sqrt{3}m} \left(1 + \mathcal{O}\left(m^{-\frac{1}{24}}\right)\right) = m^{2m} \sqrt{\frac{2}{3}} \left(1 + \mathcal{O}\left(m^{-\frac{1}{24}}\right)\right).$$

Recall that $p(m, m) = \#G_{m,m,m}^{\circ} / \#G_{m,m,m}$. Hence, by using (2), this completes the proof of the second part of Theorem 1.

6. STRUCTURE OF THE CUCKOO GRAPH

In this section, we calculate limiting distributions of some parameters of random bipartite graphs that strongly influence the behavior of Cuckoo hashing. These are

for instance the size of the tree components and the number of cycles. Some parameters might also be of interest in other applications, see *e.g.* [Blasiak and Durrett 2005]. In particular, we give the proof of Theorem 3. This proof is divided into several parts, each of it proves separately one of the claimed properties. Again, we use a generating function approach. We recall the representation of the generating function

$$g^\circ(x, y, v) = \frac{e^{\frac{1}{v}\tilde{t}(xv, yv)}}{\sqrt{1 - t_1(xv, yv)t_2(xv, yv)}}. \quad (9)$$

that counts graphs without complex components (and was established in Lemma 2). Now, we introduce an additional variable to “mark” the parameter of interest, see for instance [Flajolet and Odlyzko 1990], [Flajolet and Sedgewick 2009], and [Drmota and Soria 1995; 1997] for further details of this method.

We fix $\varepsilon > 0$ and suppose that $n = \lfloor (1 - \varepsilon)m \rfloor$. We also note that it is sufficient to consider graphs of $G_{m, m, n}^\circ$ (the set of bipartite graphs without complex components), since all results for $G_{m, m, n}^\circ$ hold for unrestricted random bipartite graphs too. This can be easily seen in the following way. Consider a random variable ξ defined on the set $G_{m, m, n}$ (with $n = \lfloor (1 - \varepsilon)m \rfloor$ and $\varepsilon > 0$) and ξ' its restriction to $G_{m, m, n}^\circ$. Then the corresponding distribution functions by F_ξ and $F_{\xi'}$ satisfy the relation

$$|F_\xi - F_{\xi'}| \leq \mathbb{P}(G_{m, m, n} \setminus G_{m, m, n}^\circ) = \mathcal{O}(1/m).$$

6.1 Number of Cycles

LEMMA 9. *The moment generating function of the number of cycles C_{nm}^{uc} and the number of cycles of length $2k$ $C_{nm, k}^{uc}$ in a graph of $G_{m, m, n}^\circ$ (with $n = \lfloor (1 - \varepsilon)m \rfloor$ and $\varepsilon > 0$) is given by*

$$\mathbb{E} e^s C_{nm}^{uc} = \exp\left(\frac{\log(1 - (1 - \varepsilon)^2)}{2} (1 - e^s)\right) \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right),$$

and

$$\mathbb{E} e^s C_{nm, k}^{uc} = \exp\left(-\frac{(1 - \varepsilon)^{2k}}{2k} (1 - e^s)\right) \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right),$$

respectively, where s is any fixed real number.

Since the moment generating function of a Poisson distribution $Po(\lambda)$ is given by $e^{\lambda(e^s - 1)}$ we immediately deduce the first part of Theorem 3.

PROOF. We start with the calculation of the total number of cycles. For this purpose we introduce a new variable w , that marks each cyclic component, that is, the exponent of w counts the number of cycles. Equation (9) generalizes to

$$\begin{aligned} g_c^\circ(x, y, v, w) &= \exp\left(\frac{1}{v}\tilde{t}(xv, yv) + \frac{w}{2} \log \frac{1}{1 - t_1(x, y)t_2(x, y)}\right) \\ &= \frac{\exp\left(\frac{1}{v}\tilde{t}(xv, yv)\right)}{(1 - t_1(xv, yv)t_2(xv, yv))^{w/2}}. \end{aligned}$$

Of course, we have $g_c^\circ(x, y, v, 1) = g^\circ(x, y, v)$. Hence, the moment generating function is given by

$$\mathbb{E} e^{s C_{nm}^{uc}} = \frac{[x^m y^m v^n] g^\circ(x, y, v, e^s)}{[x^m y^m v^n] g^\circ(x, y, v, 1)}.$$

Again, the number of tree components equals $2m - n$, thus the generating function simplifies to

$$\left[\frac{x^m y^m v^n}{(m!)^2 n!} \right] g_c^\circ(x, y, v, e^s) = \frac{n!(m!)^2}{(2m - n)!} [x^m y^m] \frac{\tilde{t}(x, y)^{2m-n}}{(1 - t_1(x, y)t_2(x, y))^{e^s/2}}.$$

We continue using Cauchy's formula and the double saddle point method as described in Lemma 3. Note that we can use the same saddle point $x_0 = y_0 = (1 - \varepsilon')e^{\varepsilon' - 1}$. The calculation is even easier because it is sufficient to calculate the leading term. We make use of the inequality

$$\left| (1 - t_1(x, y)t_2(x, y))^{-e^s/2} \right| \leq (1 - t_1(x_0, y_0)t_2(x_0, y_0))^{-e^s/2},$$

that is satisfied on the lines $|x| = x_0$, $|y| = y_0$ of integration. Furthermore, since $e^s = \mathcal{O}(1)$, the performed Taylor expansion is still applicable, and thus we obtain a corresponding result:

$$[x^m y^m] \frac{\tilde{t}(x, y)^{2m-n}}{(1 - t_1(x, y)t_2(x, y))^{e^s/2}} \sim \frac{1}{2\pi(x_0 y_0)^m k \sqrt{\Delta}} \frac{\tilde{t}(x_0, y_0)^{2m-n}}{(1 - t_1(x_0, y_0)t_2(x_0, y_0))^{e^s/2}}$$

Thus we obtain the moment generating function

$$\begin{aligned} \mathbb{E} e^{s C_{nm}^{uc}} &= \frac{\sqrt{1 - t_1(x_0, y_0)t_2(x_0, y_0)}}{(1 - t_1(x_0, y_0)t_2(x_0, y_0))^{e^s/2}} \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right) \\ &= (1 - (1 - \varepsilon)^2)^{(1 - e^s)/2} \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right), \end{aligned}$$

which completes the proof of the first part of the lemma.

The proof of the second part is very similar, we just replace g_c° by the generating function

$$g_k^\circ(x, y, v, w) = \frac{\exp\left(\frac{1}{v}\tilde{t}(xv, yv) + (w - 1)\frac{1}{2k}t_1(xv, yv)^k t_2(xv, yv)^k\right)}{\sqrt{1 - t_1(xv, yv)t_2(xv, yv)}}.$$

Hereby, w is used to mark cycles of length $2k$. Recall that the generating function of a component containing a cycle of length $2k$ is given by $\frac{1}{2k}t_1(x, y)^k t_2(x, y)^k$. We proceed as usual and yield

$$\begin{aligned} \left[\frac{x^m y^m v^n}{(m!)^2 n!} \right] g_k^\circ(x, y, v, e^s) \\ = \frac{n!(m!)^2}{(2m - n)!} [x^m y^m] \frac{\exp\left((e^s - 1)\frac{1}{2k}t_1(x, y)^k t_2(x, y)^k\right)}{\sqrt{1 - t_1(x, y)t_2(x, y)}} \tilde{t}(x, y)^{2m-n}. \end{aligned}$$

Finally, the moment generating function of $C_{nm,k}^{uc}$ equals

$$\mathbb{E} e^{s C_{nm,k}^{uc}} = \frac{[x^m y^m v^n] g_k^\circ(x, y, v, e^s)}{[x^m y^m v^n] g_k^\circ(x, y, v, 1)}$$

$$\begin{aligned}
&= \exp\left((e^s - 1)\frac{1}{2k}t_1(x_0, y_0)^k t_2(x_0, y_0)^k\right) \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right) \\
&= \exp\left(-\frac{(1-\varepsilon)^{2k}}{2k}(1 - e^s)\right) \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right),
\end{aligned}$$

which completes the proof of the lemma. \square

6.2 Trees with fixed size

The proof of the second part of Theorem 3 is more complicated, since we also normalize depending on m . In what follows, we make use of the generating function of a bipartite tree component with $2k$ vertices. Because of Lemma 1, this function is given by

$$\tilde{t}_k(x, y) = \sum_{m_1+m_2=k} m_1^{m_2-1} m_2^{m_1-1} \frac{x^{m_1} y^{m_2}}{m_1! m_2!}.$$

The following lemmata provide more detailed information about this function for $x = y$.

LEMMA 10. *We have*

$$\tilde{t}_k(x, x) = \sum_{l=0}^k l^{k-l-1} (k-l)^{l-1} \frac{x^k}{l! (k-l)!} = 2k^{k-2} \frac{x^k}{k!}.$$

PROOF. We apply Lagrange's Inversion Formula to obtain the coefficient of x^k in $\tilde{t}(x, x) = 2t(x) - t(x)^2$, where $t(x)$ denotes the usual tree function that satisfies $t(x) = x \exp(t(x))$. Because of the previous relation, it is also clear that the number of unrooted bipartite trees possessing k vertices equals twice the number of unrooted (usual) trees of size k . \square

LEMMA 11. *We have*

$$\left[\frac{\partial}{\partial u} \tilde{t}_k(x_0 e^u, x_0 e^v) \right]_{(u,v)=(0,0)} = x_0^k \sum_{l=0}^k l^{k-l} (k-l)^{l-1} \frac{1}{l! (k-l)!} = k^{k-1} \frac{x_0^k}{k!}.$$

PROOF. The proof of this lemma is a simple application of Abel's generalization of the binomial theorem,

$$x^{-1}(x+y+ka)^k = \sum_{l=0}^k \binom{k}{l} (x+la)^{l-1} (y+(k-l)a)^{k-l},$$

see, e.g., [Riordan 1968]. We set $x = k$, $y = k$ and $a = -1$ and obtain the claimed result. \square

As in the formulation of Theorem 3 we use the following notation:

$$\mu = 2 \frac{k^{k-2} (1-\varepsilon)^{k-1} e^{k(\varepsilon-1)}}{k!},$$

and

$$\sigma^2 = \mu - \frac{2e^{2k(\varepsilon-1)} k^{2k-4} (1-\varepsilon)^{2k-3} (k^2 \varepsilon^2 + k^2 \varepsilon - 4k\varepsilon + 2)}{(k!)^2}$$

where $k \geq 1$ is an integer and $0 < \varepsilon < 1$.

We are now able to prove the following lemma.

LEMMA 12. *Mean value and variance of the number of tree components $T_{mn,k}$ with k vertices of a randomly chosen graph of $G_{m,m,n}^\circ$ (with $n = \lfloor (1 - \varepsilon)m \rfloor$ and $\varepsilon > 0$) are given by*

$$\mathbb{E} T_{mn,k} = m\mu + \mathcal{O}(1) \quad (10)$$

and by

$$\text{Var} T_{mn,k} = m\sigma^2 + \mathcal{O}(1).$$

PROOF. We start introducing the variable w to mark trees with of size k and obtain the generating function

$$g_{k,t}^\circ(x, y, v, w) = \frac{\exp\left(\frac{1}{v}\tilde{t}(xv, yv) + (w-1)\frac{1}{v}\tilde{t}_k(xv, yv)\right)}{\sqrt{1 - t_1(xv, yv)t_2(xv, yv)}}.$$

The l -th factorial moment is then given by

$$\mathbb{E} T_{mn,k}(T_{mn,k} - 1) \cdots (T_{mn,k} - l + 1) = \frac{[x^m y^m v^n] \left[\frac{\partial^l}{\partial w^l} g_t^\circ(x, y, v, w) \right]_{w=1}}{[x^m y^m v^n] g_t^\circ(x, y, v, 1)}.$$

The numerator of this expression simplifies to

$$\begin{aligned} & [x^m y^m v^n] \left[\frac{\partial^l}{\partial w^l} g_t^\circ(x, y, v, w) \right]_{w=1} \\ &= [x^m y^m] \left[\frac{\partial^l}{\partial w^l} \frac{(\tilde{t}(x, y) + (w-1)\tilde{t}_k(x, y))^{2m-n}}{(2m-n)! \sqrt{1-t(x, y)}} \right]_{w=1} \\ &= [x^{2m}] \frac{\tilde{t}(x, y)^{2m-n-l}}{(2m-n)! \sqrt{1-t(x, y)}} (2m-n)^l \tilde{t}_k(x, y)^l. \end{aligned}$$

Now, we apply Lemma 3 to calculate an asymptotic expansion. By using Lemma 10, we obtain that the leading term of $\mathbb{E} T_{mn,k}(T_{mn,k} - 1) \cdots (T_{mn,k} - l + 1)$ equals

$$\frac{(2m-n)^l}{\tilde{t}(x_0, y_0)^l} \tilde{t}_k(x_0, y_0)^l = \frac{m^l(1+\varepsilon)^l}{(1-\varepsilon^2)^l} \left(2 \frac{k^{k-2}}{k!} (1-\varepsilon')^k e^{(\varepsilon'-1)k} \right)^l \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right).$$

Hence, we have completed the proof of (10). Moreover, we conclude that the variance is of order $O(m)$ too, thus its calculation requires to determine the next term of the asymptotic expansion. We do this in a semi-automatic way using Maple and obtain the proposed result. \square

LEMMA 13. *Suppose that $n = \lfloor (1 - \varepsilon)m \rfloor$ for some and $\varepsilon > 0$. Then for every $k \geq 1$ and for every real number r we have, as $m \rightarrow \infty$,*

$$\mathbb{E} e^{ir(T_{mn,k} - \mu m)/\sqrt{\sigma^2 m}} = e^{-\frac{1}{2}r^2} \left(1 + \mathcal{O}\left(m^{-\frac{1}{2}+\delta}\right) \right),$$

where $0 < \delta < \frac{1}{6}$.

PROOF. The characteristic function $\mathbb{E} e^{irT_{mn,k}}$ is given by

$$\mathbb{E} e^{irT_{mn,k}} = \frac{[x^m y^m v^n] g_{k,t}^\circ(x, y, v, e^{ir})}{[x^m y^m v^n] g^\circ(x, y, v)},$$

where we can use the simplification

$$[x^m y^m v^n] g_{k,t}^\circ(x, y, v, e^{ir}) = [x^m y^m] \frac{(\tilde{t}(x, y) + (e^{ir} - 1)\tilde{t}_k(x, y))^{2m-n}}{\sqrt{1 - t_1(x, y)t_2(x, y)}}$$

Set $M = 2m - n = m(1 + \varepsilon')$. In order to normalize we substitute r by r/\sqrt{M} and use again an saddle point method (similarly to the methods of Lemma 3). More precisely we use the representation

$$\begin{aligned} & \left(\tilde{t}(x_0 e^{is}, y_0 e^{it}) + (e^{\frac{ir}{\sqrt{M}}} - 1)\tilde{t}_k(x_0 e^{is}, y_0 e^{it}) \right)^M \\ &= \tilde{t}(x_0 e^{is}, y_0 e^{it})^M \exp \left(M \log \left(1 + \frac{\tilde{t}_k(x_0 e^{is}, y_0 e^{it})}{\tilde{t}(x_0 e^{is}, y_0 e^{it})} \left(e^{\frac{ir}{\sqrt{M}}} - 1 \right) \right) \right) \end{aligned}$$

and (for s and t with $|s|, |t| \leq \alpha = M^{-\frac{1}{2} + \delta}$, where $0 < \delta < \frac{1}{6}$) the expansion

$$\begin{aligned} & M \log \left(1 + \frac{\tilde{t}_k(x_0 e^{is}, y_0 e^{it})}{\tilde{t}(x_0 e^{is}, y_0 e^{it})} \left(e^{\frac{ir}{\sqrt{M}}} - 1 \right) \right) \\ &= M \frac{\tilde{t}_k(x_0 e^{is}, y_0 e^{it})}{\tilde{t}(x_0 e^{is}, y_0 e^{it})} \left(e^{\frac{ir}{\sqrt{M}}} - 1 \right) - \frac{M}{2} \left(\frac{\tilde{t}_k(x_0 e^{is}, y_0 e^{it})}{\tilde{t}(x_0 e^{is}, y_0 e^{it})} \right)^2 \left(e^{\frac{ir}{\sqrt{M}}} - 1 \right)^2 + \mathcal{O} \left(M^{-\frac{1}{2}} \right) \\ &= c_{00} ir \sqrt{M} - c_{00} \frac{r^2}{2} - (c_{10}s + c_{01}t) r \sqrt{M} + c_{00}^2 \frac{r^2}{2} + \mathcal{O} \left(M^{-\frac{1}{2} + 2\delta} \right). \end{aligned}$$

Hereby, we used the abbreviations

$$c_{ij} = \left[\frac{\partial^i}{\partial u^i} \frac{\partial^j}{\partial v^j} \frac{\tilde{t}_k(x_0 e^u, y_0 e^v)}{\tilde{t}(x_0 e^u, y_0 e^v)} \right]_{(u,v)=(0,0)}.$$

In particular, we have

$$c_{00} = \frac{\mu}{1 + \varepsilon'} \quad \text{and} \quad c_{10} = c_{01} = \frac{\mu}{1 + \varepsilon'} \left(k - \frac{2}{1 + \varepsilon'} \right).$$

Now by proceeding as in the proof of Lemma 3 we eventually derive an asymptotic expansion for

$$\begin{aligned} \mathbb{E} e^{irT_{mn,k}/\sqrt{M}} &= \frac{[x^m y^m v^n] g_{k,t}^\circ(x, y, v, e^{ir/\sqrt{M}})}{[x^m y^m v^n] g^\circ(x, y, v)} \\ &= \exp \left(ir \mu \frac{m}{\sqrt{M}} - \frac{\sigma^2 m}{2M} r^2 \right) + \mathcal{O} \left(m^{-\frac{1}{2} + \delta} \right) \end{aligned}$$

which implies the lemma. \square

6.3 Vertices in cycles

For the next part of the proof of Theorem 3 we have to count the number of vertices V_{nm}^c contained in cycles. The corresponding result is rather easy to obtain. We

make use of the generating function

$$g_c^\circ(x, y, v, w) = \frac{\exp\left(\frac{1}{v}\tilde{t}(xv, yv)\right)}{\sqrt{1 - w^2 t_1(xv, yv) t_2(xv, yv)}},$$

where the exponent of w counts the number of cyclic points. Hence by using again the double saddle point methods we get the characteristic function

$$\begin{aligned} \mathbb{E} e^{isV_{nm}^c} &= \frac{[x^m y^m v^n] g_n^\circ(x, y, v, e^{is})}{[x^m y^m v^n] g_n^\circ(x, y, v, 1)} \\ &= \sqrt{\frac{1 - t_1(x_0, y_0) t_2(x_0, y_0)}{1 - e^{2is} t_1(x_0, y_0) t_2(x_0, y_0)}} \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right) \\ &= \sqrt{\frac{1 - (1 - \varepsilon')^2}{1 - e^{2is}(1 - \varepsilon')^2}} \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right) = \sqrt{\frac{1 - (1 - \varepsilon)^2}{1 - e^{2is}(1 - \varepsilon)^2}} \left(1 + \mathcal{O}\left(\frac{1}{m}\right)\right). \end{aligned}$$

Of course, we have to take care of the slightly modified conditions. In particular, we make use of the bound

$$\left| \frac{1}{\sqrt{1 - e^{2is} t_1(x, y) t_2(x, y)}} \right| \leq \frac{1}{\sqrt{|1| - |e^{2is} t_1(x, y) t_2(x, y)|}} = \frac{1}{\sqrt{\varepsilon'(2 - \varepsilon')}},$$

that is satisfied on the lines of integration. Hence we conclude that the contribution outside the arcs $(-\alpha, \alpha)$ is still negligible. It is further straightforward to calculate the limiting mean and variance.

Finally we use the series expansion

$$\begin{aligned} \sqrt{\frac{1 - (1 - \varepsilon)^2}{1 - e^{2is}(1 - \varepsilon)^2}} &= \sqrt{1 - (1 - \varepsilon)^2} \sum_{k \geq 0} \binom{-\frac{1}{2}}{k} (-1)^k (1 - \varepsilon)^{2k} e^{is2k} \\ &= \sqrt{1 - (1 - \varepsilon)^2} \sum_{k \geq 0} \frac{1 \cdot 3 \cdot 5 \cdots (2k - 1)}{2^k k!} (1 - \varepsilon)^{2k} e^{is2k}, \end{aligned}$$

to infer that the probability that exactly $2k$ vertices are contained in cycles equals

$$\frac{1 \cdot 3 \cdot 5 \cdots (2k - 1)}{2^k k!} \sqrt{1 - (1 - \varepsilon)^2} (1 - \varepsilon)^{2k},$$

in limit.

6.4 Vertices in cyclic components

If we count the number of all vertices contained in cyclic components, the generating function modifies to

$$g_v^\circ(x, y, v, w) = \frac{\exp\left(\frac{1}{v}\tilde{t}(xv, yv)\right)}{\sqrt{1 - t_1(xvw, yvw) t_2(xvw, yvw)}}.$$

Here we took care of all vertices of trees that are attached to cycles. It is straightforward to calculate asymptotic mean and variance. This completes the proof of Theorem 3.

7. RUNNING TIME

The aim of this section is to show that the average construction time of a Cuckoo hash table is linear. The cost of a single insertion is thereby measured by the number of moves during this procedure, hence it equals one plus the number of kick-out operations. We cannot give an exact result, but we obtain a suitable upper bound as given in Theorem 2.

Let p denote the failure probability of a simplified cuckoo hashing attempt. We already know that $p \sim c/m$ as $m \rightarrow \infty$. Clearly, the expected number of attempts to construct the data structure is hence given by $1/(1-p)$. This implies that the expected number $\mathbb{E}N$ of rehashes to build the hash table is $\mathcal{O}(1/m)$. Furthermore, the additional complexity of a failed attempt is $\mathcal{O}(n)$, because we detect an endless loop in the insertion procedure after at most $2n$ steps. We conclude that $\mathbb{E}C_i$, the expected number of steps required during the i -th unsuccessful construction is $\mathcal{O}(m)$. Hence

$$\mathbb{E} \sum_{i=1}^N C_i = \mathbb{E}N \mathbb{E}C_1 = \mathcal{O}(1)$$

holds, *cf.* [Devroye and Morin 2003].

Therefore, it remains to show that the proposed bound in Theorem 2 holds for the situation where cuckoo hashing succeeds, *i.e.* the cuckoo graph contains only trees and cyclic components.

Consider the Cuckoo graph just before the insertion of the l -th edge (or key) and denote the vertex of first type x_l and the other y_l . Recall that a new key is always inserted in the vertex of first type. The number of steps needed to perform this insertion is fully determined by the component containing x_l , and not affected by the component containing y_l , unless x_l belongs to a cyclic component. But this is a very rare event. We know from Theorem 3 that the expected number of vertices contained in cyclic components is finite.

LEMMA 14. *Suppose that the assumptions of Theorem 1 are fulfilled. Then the expected number of all steps performed in cyclic components is bounded by a constant.*

PROOF. Assume that exactly k vertices are contained in cycles. The insertion of each of the k corresponding keys takes at most $2k$ steps, because during an insertion, no vertex is visited more than twice. The total number of expected steps is therefore bounded by

$$\sum_k 2k^2 \mathbb{P}\{V_{nm}^c = k\},$$

which is finite because of the results from Theorem 3. \square

The cuckoo graph contains $2m - l + 1$ trees before the insertion of the l -th key. Given a subgraph S , we denote the number of vertices of first and second type by $m_1(S)$ and $m_2(S)$, respectively. Further, denote the maximum number of steps needed for the insertion in S by $\nu(S)$. Assume that x_l is contained in a tree component T . Note that each of the vertices of first type equals x_l with the same probability. Hence, choosing x_l corresponds to root the tree at a vertex of first

kind. Observe that there exists a unique node z of T that corresponds to an empty memory cell of the hash table. This node is somehow randomly (but not uniformly) selected among all the nodes of both types of T . The insertion itself corresponds to a walk in T , starting in x_l and ending in z .

The generating function $\tilde{t}(x, y)$ of unrooted bipartite trees can be rewritten as

$$\tilde{t}(x, y) = \sum_T \frac{x^{m_1(T)} y^{m_2(T)}}{m_1(T)! m_2(T)!}.$$

Consider a fixed bipartite tree T . Then, there exist $m_1(T)$ bipartite trees T^* possessing a root of first kind that correspond to T , what we denote by $T^* \sim T$. Each rooted tree T^* represents a unique choice of x_l among the nodes of T of first type and leads to an insertion cost of $\nu(T^*)$. Hence, we define the following function:

$$H_T(x, y) = \sum_T \sum_{T^* \sim T} \nu(T^*) \frac{x^{m_1(T)} y^{m_2(T)}}{m_1(T)! m_2(T)!}.$$

Similarly, assume that the insertion takes place in the unicyclic part U . There exist $m_1(U)$ selections of x_l among the nodes of U that lead to an insertion cost of $\nu(U)$. Thus we make use of the function

$$H_U(x, y) = \sum_U m_1(U) \nu(U) \frac{x^{m_1(U)} y^{m_2(U)}}{m_1(U)! m_2(U)!}.$$

Now, we put these things together. Recall that the Cuckoo graph contains $k = 2m - l + 1$ tree components T_1, \dots, T_k and an unicyclic part U :

$$\begin{aligned} H(x, y) &= \frac{1}{k!} \sum_{(T_1, \dots, T_k, U)} \left(\sum_{j=1}^k \sum_{T_j^* \sim T_j} \nu(T_j^*) + m_1(U) \nu(U) \right) \\ &\quad \times \binom{m}{m_1(T_1), \dots, m_1(T_k), m_1(U)} \binom{m}{m_2(T_1), \dots, m_2(T_k), m_2(U)} \frac{x^m y^m}{m! m!} \\ &= \frac{1}{(k-1)!} \frac{H_T(x, y) \tilde{t}(x, y)^{k-1}}{\sqrt{1 - t_1(x, y) t_2(x, y)}} + \frac{\tilde{t}(x, y)^k}{k!} H_U(x, y) \end{aligned}$$

By construction,

$$C(l) = \frac{1}{m \#G_{m, m, l}^\circ} \frac{(m!)^2 n!}{(2m - l)!} [x^m y^m] H(x, y)$$

bounds the average insertion cost of the l -th key.

First, consider

$$\frac{(m!)^2 n!}{(2m - l)!} \frac{1}{m \#G_{m, m, l}^\circ} [x^m y^m] \frac{\tilde{t}(x, y)^k}{k!} H_U(x, y).$$

This is $\mathcal{O}(1)$, because of Lemma 14. Hence, we obtain the upper bound

$$C(l) = \frac{1}{m \#G_{m, m, l}^\circ} \frac{(m!)^2 n!}{(2m - l)!} [x^m y^m] \frac{H_T(x, y) \tilde{t}(x, y)^{2m-l}}{\sqrt{1 - t_1(x, y) t_2(x, y)}} + \mathcal{O}(1)$$

for the average complexity of inserting the l -th key. Applying Lemma 3, we thus get

$$C(l) = \frac{H_T(x_0, x_0)(2m - l + 1)}{m\tilde{t}(x_0, x_0)} \left(1 + \mathcal{O}\left(\frac{1}{2m - l}\right) \right).$$

The analysis of this parameter is in fact similar to the previous calculations. A slight difference is the new occurring function $H_T(x, y)$, but it behaves like an additional constant factor. Thus, we only need to know $H_T(x_0, x_0)$, which we will consider next.

First we use the trivial upper bound $\nu(T^*) \leq m_1(T^*) + m_2(T^*)$ and obtain for positive x and y

$$H_T(x, y) \leq \sum_T m_1(T) (m_1(T) + m_2(T)) \frac{x^{m_1(T)} y^{m_2(T)}}{m_1(T)! m_2(T)!}.$$

$$\sum_T m_1(T) \frac{x^{m_1(T)} y^{m_2(T)}}{m_1(T)! m_2(T)!} = x \frac{\partial}{\partial x} \tilde{t}(x, y) = t_1(x, y) \quad (11)$$

Recall that $t_1(x, x)$ equals $t(x)$, so we establish

$$H(x, x) \leq x \frac{\partial}{\partial x} t(x) = \frac{t(x)}{1 - t(x)} = \frac{t_1(x, x)}{1 - t_1(x, x)}.$$

Setting $x = x_0 = \frac{l}{m} e^{-l/m}$ and using the local expansions for $t(x)$ and $\tilde{t}(x)$ we obtain

$$C(l) \leq \frac{m}{m - l} \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right) \quad (12)$$

which leads to

$$\frac{1}{n} \sum_{l=1}^n C(l) \leq \frac{1}{n} \sum_{l=1}^n \frac{1}{1 - \frac{l}{m}} \left(1 + \mathcal{O}\left(\frac{1}{m}\right) \right) \rightarrow \frac{1}{1 - \varepsilon} \int_1^{\varepsilon} \frac{da}{a} = \frac{\log \frac{1}{1 - \varepsilon}}{1 - \varepsilon}, \quad (13)$$

where $n = \lfloor (1 - \varepsilon)m \rfloor$ and m goes to infinity. This completes the proof of the first bound of Theorem 2.

Next, we try to obtain a better bound using a more suitable estimate for $\nu(T)$. Recall that the selection of the vertex x_l in a tree component, transforms this component into a rooted bipartite tree. The insertion procedure starts at the root and the number of required steps is bounded by the height of this tree. Further, note that in the asymptotic analysis we are only interested in the special case $x = y = x_0$. Because of this, we can consider usual (non bipartite) rooted trees instead.

We introduce the notations

- $t_n^{[k]}$ for the number of rooted trees with n vertices and height less or equal k ,
- and h_n for the sum of the heights of all rooted trees with n vertices.

Moreover, we introduce the corresponding generating functions:

$$t^{[k]}(x) = \sum_{n \geq 0} t_n^{[k]} x^n, \quad h(x) = \sum_{n \geq 0} h_n x^n.$$

Due to [Flajolet and Odlyzko 1982] we know

$$t(x) - t^{[k]}(x) \sim 2 \frac{\delta(x) (1 - \delta(x))^n}{1 - (1 - \delta(x))^n},$$

where $\delta(x) = \sqrt{2(1 - ex)} + \mathcal{O}(1 - ex)$ and further,

$$h(x) = \sum_{k \geq 0} \left(t(x) - t^{[k]}(x) \right) \sim -2 \log \delta(x)$$

in a Δ -domain around its singularity e^{-1} .

Let T^* denote a bipartite tree possessing a root node of first kind. Clearly, the upper bound $\nu(T^*) \leq \text{height}(T^*)$ holds. Furthermore, $t_1(x, y)$ is given by

$$t_1(x, y) = \sum_T \sum_{T^* \sim T} \frac{x^{m_1(T)} y^{m_2(T)}}{m_1(T)! m_2(T)!}.$$

As in (11), we make use of $t_1(x, x) = t(x)$ and we derive the inequality

$$H_T(x, x) \leq h(x).$$

Thus, we use the asymptotic approximation of $h(x)$ as upper bound of $H(x, x)$ and obtain similarly to (12) the upper bound

$$C(l) \leq m \frac{-\log 2 \left(1 - e^{1 - l/m} \frac{l}{m} \right)}{l} \left(1 + \mathcal{O} \left(\frac{1}{m} \right) \right),$$

for the construction time. This is of course only valid near the singularity, that is for $1 - l/m$ close to zero. Nevertheless, this result is suitable to prove the second bound stated in Theorem 2. This is due the fact that the integral

$$\int_{1/2}^{\varepsilon} \frac{-\log 2 (1 - e^a (1 - a))}{1 - a} da$$

is obviously bounded for $\varepsilon \rightarrow 0$, in contrast to the corresponding integral of (13).

A numerical computation using both bounds leads to an estimated value of about 4 steps per insertion.

8. A SIMPLIFIED VERSION OF CUCKOO HASHING

8.1 Introduction

In this section, we discuss a modified, and in some sense simplified version of the Cuckoo hash algorithm. Instead of two tables of size m , we use just one table of size $2m$ and grant both hash functions access to the whole memory. This simplifies the analysis, because the bipartite Cuckoo graph is replaced by a usual random graph (a version without different types of vertices, but directed edges), and we do not need bivariate generating functions any longer. Despite this change, the modified Cuckoo hash algorithm succeeds if and only if this new Cuckoo Graph does not contain a complex component (as in the original case).

This approach was already suggested by [Pagh and Rodler 2004] and further by [Fotakis et al. 2005] for generalized d -ary Cuckoo hashing, which uses d tables (or hash functions), because it is easier to analyze, see also [Kutzelnigg 2008; 2009].

m	$\varepsilon = 0.4$		$\varepsilon = 0.2$		$\varepsilon = 0.1$		$\varepsilon = 0.06$		$\varepsilon = 0.04$	
	res.	exp.	res.	exp.	res.	exp.	res.	exp.	res.	exp.
$5 \cdot 10^3$	44	49.2	710	767	5272	8100	15276		26802	
10^4	27	24.6	386	383	3122	4050	10451		20536	
$5 \cdot 10^4$	5	4.92	87	76.7	737	810	3414	4159	8666	
10^5	3	2.46	32	38.3	417	405	1831	2079	5323	7380
$5 \cdot 10^5$	0	0.49	7	7.67	85	81	417	416	1358	1476

Table III. Comparison of simulation result (res.) and expected number of failures (exp.) during the construction of $5 \cdot 10^5$ simplified cuckoo hash tables.

Our analysis shows, that there are also advantages of this version of the algorithm, which recommend it for practical application.

8.2 Results

Almost the same results as stated in Theorems 1, 2, and 3 hold for this simplified algorithm too:

THEOREM 4. *Suppose that $\varepsilon \in (0, 1)$ is fixed. Then the probability that a simplified Cuckoo hash of $n = \lfloor (1 - \varepsilon)m \rfloor$ data points into a table of size $2m$ succeeds, (that is, the corresponding Cuckoo graph contains no complex component,) is equal to*

$$\bar{p}(n, m) = 1 - \frac{(5 - 2\varepsilon)(1 - \varepsilon)^2}{48\varepsilon^3} \frac{1}{m} + \mathcal{O}\left(\frac{1}{m^2}\right).$$

This probability decreases to

$$\bar{p}(n, n) = \sqrt{\frac{2}{3}} + o(1),$$

if n equals m .

As mentioned earlier, the second statement can be found in [Janson et al. 1993].

The only difference to the result of Theorem 1 is the modified asymptotic expansion.

$$\bar{h}(\varepsilon) = \frac{5}{48}\varepsilon^{-3} - \frac{1}{4}\varepsilon^{-2} + \frac{3}{16}\varepsilon^{-1} - \frac{1}{24}.$$

But the leading term is still the same as in (7), so the behavior is almost the same for ε close to zero. Figure 3 compares these two functions $\bar{h}(\varepsilon)$ and $h(\varepsilon) = (2\varepsilon^2 - 5\varepsilon + 5)(1 - \varepsilon)^3 / (12(2 - \varepsilon)^2\varepsilon^3)$. We conclude that the success probability of simplified cuckoo hashing is slightly smaller compared to the standard algorithm. This is also justified by our simulation results given in Tables I and III. For all tested combinations of m and ε , the entry in the first table is less or equal than the corresponding entry in the latter one. However, note that there is no serious difference, because the failure probability is still very small for all practical relevant instances.

We obtain the same bound on the construction cost for the simplified algorithm as for the original one.

THEOREM 5. *Under the assumptions of Theorem 4, the expected construction*

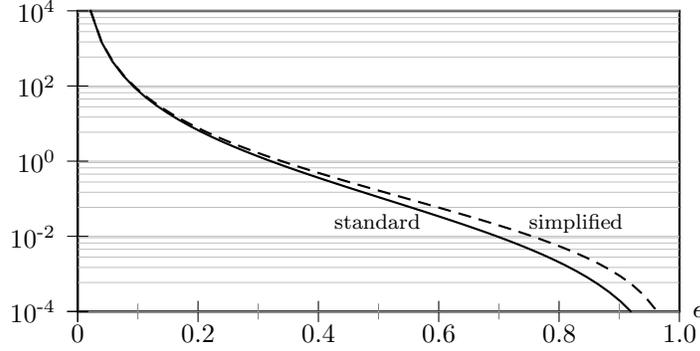


Fig. 3. The functions h (standard c.h.) and \bar{h} (simplified c.h.).

m	$\varepsilon = 0.4$	$\varepsilon = 0.2$	$\varepsilon = 0.1$	$\varepsilon = 0.06$	$\varepsilon = 0.04$
$5 \cdot 10^3$	1.1972	1.3149	1.4551	1.6376	1.8336
10^4	1.1970	1.3108	1.4263	1.5699	1.7415
$5 \cdot 10^4$	1.1967	1.3070	1.3945	1.4728	1.5715
10^5	1.1967	1.3063	1.3903	1.4512	1.5248
$5 \cdot 10^5$	1.1967	1.3060	1.3859	1.4321	1.4700
$\frac{-\log \varepsilon}{1-\varepsilon}$	1.5272	2.0118	2.5584	2.9930	3.3530

sample size: $5 \cdot 10^5$

Table IV. Insertion costs of simplified cuckoo hashing.

time of a simplified Cuckoo hash table is bounded above by

$$\min\left(C, \frac{-\log \varepsilon}{1-\varepsilon}\right)n + \mathcal{O}(1),$$

where the constant implied by $\mathcal{O}(1)$ depend on ε .

Although we give the same upper bound for both variations, we note that the actual behavior is different. The comparison of the results provided in Tables II and IV shows that the expected number of steps per insertion is smaller for the modified version of the algorithm, see [Kutzelnigg 2009] for further details.

The different behavior of the two versions of the algorithm is also influenced by some differences in the structure of the underlying Cuckoo graphs, which we investigate next.

THEOREM 6. *Suppose that $\varepsilon \in (0, 1)$ is fixed and that $n = \lfloor (1 - \varepsilon)m \rfloor$. Then a random labeled multigraph with $2m$ vertices and n edges satisfies the following properties.*

- (1) *The number of unicyclic components $\bar{C}_{mn,k}^{uc}$ with cycle length k has in limit a Poisson distribution $Po(\lambda_k)$ with parameter*

$$\lambda_k = \frac{1}{2k} (1 - \varepsilon)^k,$$

and the number of unicyclic components $\bar{C}_{mn}^{uc} = \sum_{k \geq 1} C_{mn,k}^{uc}$ has in limit a

Poisson distribution $Po(\lambda)$, too, with parameter

$$\lambda = -\frac{1}{2} \log \varepsilon.$$

- (2) The number of tree components $\bar{T}_{mn,k}$ with k vertices satisfy a central limit theorem of the form

$$\frac{\bar{T}_{mn,k} - \mu m}{\sqrt{\sigma^2 m}} \rightarrow N(0, 1),$$

where

$$\mu = 2 \frac{k^{k-2} (1-\varepsilon)^{k-1} e^{k(\varepsilon-1)}}{k!}$$

and

$$\sigma^2 = \mu - \frac{2e^{2k(\varepsilon-1)} k^{2k-4} (1-\varepsilon)^{2k-3} (k^2 \varepsilon^2 + k^2 \varepsilon - 4k\varepsilon + 2)}{(k!)^2}.$$

Furthermore, mean and variance are asymptotically given by $\mathbb{E} \bar{T}_{mn,k} = \mu m + \mathcal{O}(1)$ and $\text{Var} \bar{T}_{mn,k} = \sigma^2 m + \mathcal{O}(1)$ as $m \rightarrow \infty$, respectively.

- (3) The number of vertices \bar{V}_{mn}^c contained in cycles has a limiting distribution \bar{V}^c with characteristic function

$$\mathbb{E} e^{is\bar{V}^c} = \sqrt{\frac{\varepsilon}{1 - e^{is}(1-\varepsilon)}},$$

mean and variance satisfy

$$\lim_{m \rightarrow \infty} \mathbb{E} \bar{V}_{mn}^c = \mathbb{E} \bar{V}^c = \frac{1-\varepsilon}{2\varepsilon},$$

and

$$\lim_{m \rightarrow \infty} \text{Var} \bar{V}_{mn}^c = \text{Var} \bar{V}^c = \frac{(1-\varepsilon)}{2\varepsilon^2}.$$

- (4) Furthermore, the expected value of the number of vertices \bar{V}_{mn}^{uc} in unicyclic components is asymptotically given by

$$\lim_{m \rightarrow \infty} \mathbb{E} \bar{V}_{mn}^{uc} = \frac{(1-\varepsilon)}{2\varepsilon^2},$$

and its variance by

$$\lim_{m \rightarrow \infty} \text{Var} \bar{V}_{mn}^{uc} = \frac{(1-\varepsilon)(2-\varepsilon)}{2\varepsilon^4}.$$

The number of tree components of fixed size have the same Gaussian limiting distributions in both cases. As the tree components have the widest influence on the complexity of the insertion, we expect a similar behavior of both variants.

However, there are some differences concerning the cyclic components. The parameters of the limiting Poisson distributions differ in one way from the results of usual cuckoo hashing: The term $(1-\varepsilon)^2$ is replaced by $(1-\varepsilon)$. This explains that the expected number of cycles in a non-bipartite cuckoo graph is larger compared

to the bipartite counterpart, which is also confirmed by our experiments. Nevertheless, this does not seem to have big influence on the behavior of the hash algorithm, since the difference is bounded by $\log(2)/2$, and hence the expected number of cycles is still very small.

This result suggests that the expected number of vertices in cycles and cyclic components is larger, too, which is indeed verified by part (4) of Theorem 6. For the vertices in cycles themselves, the increase is bounded by a small additive constant, namely $1/4$. If we consider all vertices in cyclic components, they are surprisingly no longer bounded by a constant for small ε . But again, this is not of great influence as the ratio of the expectations tends to 1 as ε tends to 0.

8.3 Sketch of the Proofs of Theorems 4-6

We already mentioned above that the simplified algorithm is connected with a modified “non-bipartite” Cuckoo graph, and that the restrictions concerning the structure of the graph remain unchanged. The evolution of the graph is described by the multigraph process of [Janson et al. 1993], with the only difference that we consider labeled and directed edges.

Similarly to Section 4, let $G_{2m,n}$ denote the set of all vertex and edge labeled and directed multigraphs (V, E) with $|V| = 2m$ and $|E| = n$. Obviously we have

$$\#G_{2m,n} = (4m^2)^n.$$

Furthermore, let $G_{2m,n}^\circ$ denote those graphs of this family without complex components and set

$$g^\circ(x, v) = \sum_{m,n} \#G_{2m,n}^\circ \frac{x^{2m}}{(2m)!} \frac{v^n}{n!}.$$

Again we make use of the generating functions $t(x)$ and $\tilde{t}(x)$ of rooted and unrooted trees which satisfy the equations

$$t(x) = xe^{t(x)}, \quad \tilde{t}(x) = t(x) - \frac{1}{2}t(x)^2.$$

Similarly to Lemma 2, we obtain the following explicit representation (see also [Janson et al. 1993]):

LEMMA 15. *The generating function $g^\circ(x, v)$ is given by*

$$g^\circ(x, v) = \frac{e^{\frac{1}{2v}\tilde{t}(2xv)}}{\sqrt{1 - t(2xv)}}.$$

We use Cauchy’s formula and obtain a integral representation analog to (5)

$$\#G_{2m,n}^\circ = \frac{2^n(2m)!n!}{2\pi i(2m-n)!} \oint \frac{\tilde{t}(x)^{2m-n}}{\sqrt{1-t(x)}} \frac{dx}{x^{m+1}}. \quad (14)$$

Again, we proceed using the saddle point method. Hence we need a one dimensional version of Lemma 3.

LEMMA 16. *Let $f(x)$ and $g(x)$ be analytic functions locally around 0 such that all coefficients $[x^m]f(x)$ and $[x^m]g(x)$ are non negative and such that the “aperiodicity condition” $\gcd\{m|[x^m]f(x) > 0\} = 1$ holds.*

Let R be a compact interval of the positive real line that is contained in the radius of convergence of $f(x)$ and $g(x)$. Furthermore set

$$S = \left\{ \frac{x}{f(x)} \frac{\partial}{\partial x} f(x) : x \in R \right\}.$$

Then we have

$$[x^m]g(x)f(x)^k = \frac{g(x_0)f(x_0)^k}{x_0^m \sqrt{2\pi k \kappa_2}} \left(1 + \frac{H}{24\kappa_2^3} \frac{1}{k} + \mathcal{O}\left(\frac{1}{k^2}\right) \right),$$

uniformly for $m/k \in S$, where x_0 is uniquely determined by

$$\frac{m}{k} = \frac{x_0 f'(x_0)}{f(x_0)}.$$

and the constants κ_2 and H are given in the following way. Let κ_i and $\bar{\kappa}_i$ be the cumulants

$$\kappa_i = \left[\frac{\partial^i}{\partial u^i} \log f(x_0 e^u) \right]_{u=0}, \quad \bar{\kappa}_i = \left[\frac{\partial^i}{\partial u^i} \log g(x_0 e^u) \right]_{u=0}.$$

Then H is given by

$$12\kappa_2\kappa_3\bar{\kappa}_1 + 3\kappa_2\kappa_4 - 12\kappa_2^2\bar{\kappa}_1^2 - 12\kappa_2^2\bar{\kappa}_2 - 5\kappa_3^2.$$

The leading term of this asymptotic expansion is already given in [Gardy 1995], and the calculation of further terms is suggested as possible extension by the author. A detailed asymptotic expansion for the special case $g = 1$ can be found in [Drmota 1994].

Concerning (14) we obtain the relation $t(x_0) = 1 - \varepsilon$ and hence the saddle point

$$x_0 = (1 - \varepsilon)e^{-(1-\varepsilon)}.$$

A direct application of Lemma 16 completes the proof of the first part of Theorem 4. As mentioned above the limit for $p(n, n)$ can be found in [Janson et al. 1993].

The proofs given in sections 6 and 7 can be easily adapted to proof the Theorems 5 and 6. We just replace the bivariate generating functions by their simplified counterparts and use a simple instead of a double saddle point method, see [Kutzelnigg 2009] for details.

9. CONCLUSION

The main contribution of this paper was a precise analysis of standard Cuckoo hashing introduced by [Pagh and Rodler 2004]. Unlike usual hash algorithms, this data structure offers constant worst case access time. This is achieved by rearranging keys to resolve conflicts. However, there is a non-zero probability that the creation of the hash table fails. We showed that the error probability is of order $1/m$ with an explicit constant depending on ε when the load factor is restricted below 0.5. Further, we proved that the failure rate increases asymptotically to approximately 18.4% for half-full tables. All these results were obtained using the Cuckoo graph, a bipartite random graph that is closely related to the data structure.

As a further important result, we analyzed the average running time required to build-up a Cuckoo hash table. In particular, we gave an upper bound that is linear

in the size of the data structure, conditioned that the load is restricted below 50%. This result was derived via a detailed investigation of the structure of the Cuckoo graph. Among other things, we provide asymptotic results covering the number of cycles and trees of given size, including limiting distributions.

Finally, we briefly discussed a simplified version of Cuckoo hashing. In contrast to the standard algorithm, both hash functions address the whole memory. Nevertheless, we obtain almost the same results as for the standard algorithm using a non-bipartite random graph model.

APPENDIX

We present the proofs of Lemma 3 and 8. We start with the proof of Lemma 3, in which we will use the formula

$$\int_{-\infty}^{\infty} e^{-z^2/2} z^k dz = \begin{cases} 1 \cdot 3 \cdot 5 \dots (k-1) \sqrt{2\pi} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

PROOF OF LEMMA 3. The technical conditions on the coefficients of $f(x, y)$ ensure that the function $f(x_0 e^{is}, y_0 e^{it})$, $(s, t) \in [-\pi/2, \pi/2]^2$, has its maximal modulus for $s = t = 0$. Furthermore, it can be seen that the saddle point (x_0, y_0) is unique, because the cummulants of second order are strictly positive (compare with [Good 1957]).

We start by applying Cauchy's Formula and substitute $x = x_0 e^{is}$ and $y = y_0 e^{it}$:

$$\begin{aligned} [x^{m_1} y^{m_2}] g(x, y) f(x, y)^k &= -\frac{1}{4\pi^2} \int_{|x|=x_0} \int_{|y|=y_0} \frac{g(x, y) f(x, y)^k}{x^{m_1+1} y^{m_2+1}} dy dx \\ &= \frac{1}{4\pi^2 x_0^{m_1} y_0^{m_2}} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} g(x_0 e^{is}, y_0 e^{it}) f(x_0 e^{is}, y_0 e^{it})^k e^{-m_1 is - m_2 it} dt ds. \end{aligned}$$

The contribution of the integral taken over the range $I = ([-\pi, \pi] \times [-\pi, \pi]) \setminus ([-\alpha, \alpha] \times [-\alpha, \alpha])$ is very small compared to the remaining integral, where $\alpha = k^{-1/2+\xi}$ and ξ denotes a real number satisfying $0 < \xi < 1/6$. By continuity we surely have $|f(x_0 e^{is}, y_0 e^{it})| \leq f(x_0, y_0) - \delta$ if $|s| \geq \eta$ or $|t| \geq \eta$, where $\delta > 0$ and $\eta > 0$ are chosen appropriately. Furthermore, for $|s| < \eta$ and $|t| < \eta$ we can use a local expansion of the form

$$e^{k \log f(x_0 e^{is}, y_0 e^{it}) - m_1 is - m_2 it} = f(x_0, y_0)^k e^{-\frac{k}{2}(\kappa_{20}s^2 + 2\kappa_{11}st + \kappa_{02}t^2) + \mathcal{O}(k^{-\frac{1}{2}+3\xi})}.$$

to deduce that (for some $c > 0$)

$$\begin{aligned} \left| \iint_I g(x_0 e^{is}, y_0 e^{it}) f(x_0 e^{is}, y_0 e^{it})^k e^{-m_1 is - m_2 it} dt ds \right| \\ \leq 4\pi^2 g(x_0, y_0) f(x_0, y_0)^k e^{-ck^{2\xi}}, \end{aligned}$$

Hence this part of the integral is negligible (as proposed).

Next, we substitute $u = \sqrt{k}s$ and $v = \sqrt{k}t$ and calculate Taylor expansions of the functions $\log f$ and $\log g$. More precisely, we obtain the expansions

$$\begin{aligned} k \log f \left(x_0 e^{i \frac{u}{\sqrt{k}}}, y_0 e^{i \frac{v}{\sqrt{k}}} \right) - m_1 i \frac{u}{\sqrt{k}} - m_2 i \frac{v}{\sqrt{k}} &= k \log f(x_0, y_0) \\ - \frac{1}{2} (\kappa_{20} u^2 + 2\kappa_{11} uv + \kappa_{02} v^2) - \frac{i}{6\sqrt{k}} (\kappa_{30} u^3 + 3\kappa_{21} u^2 v + 3\kappa_{12} uv^2 + \kappa_{03} v^3) \\ &+ \frac{1}{24k} (\kappa_{40} u^4 + 4\kappa_{31} u^3 v + 6\kappa_{22} u^2 v^2 + 4\kappa_{13} uv^3 + \kappa_{04} v^4) + \mathcal{O}(k\alpha^5), \end{aligned}$$

and

$$\begin{aligned} \log g(x_0 e^{i \frac{u}{\sqrt{k}}}, y_0 e^{i \frac{v}{\sqrt{k}}}) \\ = \log g(x_0, y_0) + \frac{i}{\sqrt{k}} (\bar{\kappa}_{10} u + \bar{\kappa}_{01} v) - \frac{1}{2k} (\bar{\kappa}_{20} u^2 + 2\bar{\kappa}_{11} uv + \bar{\kappa}_{02} v^2) + \mathcal{O}(\alpha^3) \end{aligned}$$

in the neighborhood of (x_0, y_0) . The linear terms vanish due to the choice of the saddle point. By using the expansions (in k)

$$\begin{aligned} \exp \left(-i \frac{1}{6\sqrt{k}} (\kappa_{30} u^3 + 3\kappa_{21} u^2 v + 3\kappa_{12} uv^2 + \kappa_{03} v^3) \right) \\ = 1 - \frac{i}{6\sqrt{k}} (\kappa_{30} u^3 + 3\kappa_{21} u^2 v + 3\kappa_{12} uv^2 + \kappa_{03} v^3) \\ - \frac{1}{72k} (\kappa_{30} u^3 + 3\kappa_{21} u^2 v + 3\kappa_{12} uv^2 + \kappa_{03} v^3)^2 + \mathcal{O}(k^3 \alpha^9), \end{aligned}$$

$$\begin{aligned} \exp \left(\frac{1}{24k} (\kappa_{40} u^4 + 4\kappa_{31} u^3 v + 6\kappa_{22} u^2 v^2 + 4\kappa_{13} uv^3 + \kappa_{04} v^4) \right) \\ = 1 + \frac{1}{24k} (\kappa_{40} u^4 + 4\kappa_{31} u^3 v + 6\kappa_{22} u^2 v^2 + 4\kappa_{13} uv^3 + \kappa_{04} v^4) + \mathcal{O}(k^2 \alpha^8), \end{aligned}$$

$$\exp \left(\frac{i}{\sqrt{k}} (\bar{\kappa}_{10} u + \bar{\kappa}_{01} v) \right) = 1 + \frac{i}{\sqrt{k}} (\bar{\kappa}_{10} u + \bar{\kappa}_{01} v) - \frac{1}{2k} (\bar{\kappa}_{10} u + \bar{\kappa}_{01} v)^2 + \mathcal{O}(\alpha^3),$$

$$\begin{aligned} \exp \left(-\frac{1}{2k} (\bar{\kappa}_{20} u^2 + 2\bar{\kappa}_{11} uv + \bar{\kappa}_{02} v^2) \right) = \\ 1 - \frac{1}{2k} (\bar{\kappa}_{20} u^2 + 2\bar{\kappa}_{11} uv + \bar{\kappa}_{02} v^2) + \mathcal{O}(\alpha^4) \end{aligned}$$

we can rewrite the remaining integral in the following way:

$$\begin{aligned} \frac{1}{4\pi^2 x_0^{m_1} y_0^{m_2}} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} g(x_0 e^{is}, y_0 e^{it}) e^{k \log f(x_0 e^{is}, y_0 e^{it}) - m_1 is - m_2 it} dt ds \\ = \frac{g(x_0, y_0) f(x_0, y_0)^k}{4k\pi^2 x_0^{m_1} y_0^{m_2}} \int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} \int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} e^{-\frac{1}{2}(\kappa_{20} u^2 + 2\kappa_{11} uv + \kappa_{02} v^2)} \end{aligned}$$

$$\begin{aligned}
 & \times \left(1 - \frac{i}{6\sqrt{k}} (\kappa_{30}u^3 + 3\kappa_{21}u^2v + 3\kappa_{12}uv^2 + \kappa_{03}v^3) \right. \\
 & \quad + \frac{i}{\sqrt{k}} (\bar{\kappa}_{10}u + \bar{\kappa}_{01}v) - \frac{1}{72k} (\kappa_{30}u^3 + 3\kappa_{21}u^2v + 3\kappa_{12}uv^2 + \kappa_{03}v^3)^2 \\
 & \quad + \frac{1}{24k} (\kappa_{40}u^4 + 4\kappa_{31}u^3v + 6\kappa_{22}u^2v^2 + 4\kappa_{13}uv^3 + \kappa_{04}v^4) \\
 & \quad + \frac{1}{6k} (\kappa_{30}u^3 + 3\kappa_{21}u^2v + 3\kappa_{12}uv^2 + \kappa_{03}v^3) (\bar{\kappa}_{10}u + \bar{\kappa}_{01}v) \\
 & \quad \left. - \frac{1}{2k} (\bar{\kappa}_{10}u + \bar{\kappa}_{01}v)^2 - \frac{1}{2k} (\bar{\kappa}_{20}u^2 + 2\bar{\kappa}_{11}uv + \bar{\kappa}_{02}v^2) + \mathcal{O}(\alpha^9 k^3) \right) dv du.
 \end{aligned} \tag{15}$$

Without loss of generality we can assume that $\kappa_{20} \geq \kappa_{02}$. We substitute $u = \sqrt{\kappa_{02}/\Delta} a$ and $v = -\kappa_{11}/\sqrt{\kappa_{02}\Delta} a + b/\sqrt{\kappa_{02}}$, where $\Delta = \kappa_{20}\kappa_{02} - \kappa_{11}^2$. Hence

$$\int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} \int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} e^{-\frac{1}{2}(\kappa_{20}u^2 + 2\kappa_{11}uv + \kappa_{02}v^2)} dv du = \frac{1}{\sqrt{\Delta}} \int_{-\mu}^{\mu} \int_{-\nu(a)}^{\nu(a)} e^{-\frac{a^2+b^2}{2}} db da,$$

where $\mu = \alpha\sqrt{k\Delta/\kappa_{02}}$ and $\nu(a) = \alpha\sqrt{k\kappa_{02}} + a\kappa_{11}/\sqrt{\Delta}$. Note that for all a satisfying $-\mu \leq a \leq \mu$, the inequality $\nu(a) \geq \nu_{\min} = \alpha\sqrt{k}(\sqrt{\kappa_{02}} - \kappa_{11}/\sqrt{\kappa_{02}})$ is valid. Furthermore the relation $\kappa_{20} \geq \kappa_{02}$ implies that $\sqrt{\kappa_{02}} - \kappa_{11}/\sqrt{\kappa_{02}} > 0$. Consequently we get (for some constant $0 < c < \frac{1}{2}$)

$$\int_{\nu(a)}^{\infty} e^{-\frac{b^2}{2}} b^l db \leq \int_{\nu_{\min}}^{\infty} e^{-\frac{b^2}{2}} b^l db = \mathcal{O}(e^{-ck^{2\xi}}). \tag{16}$$

The last part of (15) can be estimated by

$$\begin{aligned}
 & \left| \int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} \int_{-\alpha\sqrt{k}}^{\alpha\sqrt{k}} e^{-\frac{1}{2}(\kappa_{20}u^2 + 2\kappa_{11}uv - \kappa_{02}v^2)} \mathcal{O}(k^3 \alpha^9) dv du \right| \\
 & \leq \frac{\mathcal{O}(k^3 \alpha^9)}{\sqrt{\Delta}} \int_{-\mu}^{\mu} \int_{-\nu(a)}^{\nu(a)} e^{-\frac{a^2+b^2}{2}} db da \leq \mathcal{O}(k^3 \alpha^9) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{a^2+b^2}{2}} db da \\
 & = \mathcal{O}(k^{-\frac{3}{2}+9\xi}).
 \end{aligned}$$

Finally we introduce the notation

$$\begin{aligned}
 I(p, q) &= \int_{-\mu}^{\mu} \int_{-\nu(a)}^{\nu(a)} e^{-\frac{a^2+b^2}{2}} a^p b^q db da \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{a^2+b^2}{2}} a^p b^q db da + \mathcal{O}(e^{-ck^{2\xi}})
 \end{aligned} \tag{17}$$

Obviously, $I(p, q) = 0$ if either p or q is odd. Hence the main term of the integral in (15) rewrites to

$$\begin{aligned}
I(0, 0) &- \kappa_{03}^2 \frac{I(6, 0)}{72\kappa_{02}^3} + (\kappa_{04} + 4\kappa_{03}\bar{\kappa}_{01}) \frac{I(4, 0)}{24\kappa_{02}} - (\bar{\kappa}_{02} + \bar{\kappa}_{01}^2) \frac{I(2, 0)}{2\kappa_{02}} \\
&+ (\kappa_{11}^6 \kappa_{03}^2 + 6\kappa_{11}^4 \kappa_{03} \kappa_{02}^2 \kappa_{21} - 2\kappa_{11}^3 \kappa_{03} \kappa_{02}^3 \kappa_{30} - 6\kappa_{11}^5 \kappa_{03} \kappa_{02} \kappa_{12} \\
&\quad - 18\kappa_{11}^3 \kappa_{12} \kappa_{02}^3 \kappa_{21} + 9\kappa_{12}^2 \kappa_{11}^4 \kappa_{02}^2 + 6\kappa_{11}^2 \kappa_{12} \kappa_{02}^4 \kappa_{30} + 9\kappa_{21}^2 \kappa_{11}^2 \kappa_{02}^4 \\
&\quad - 6\kappa_{02}^5 \kappa_{11} \kappa_{21} \kappa_{30} + \kappa_{02}^6 \kappa_{30}^2) \frac{I(0, 6)}{72\kappa_{02}^3 \Delta^3} \\
&- (5\kappa_{11}^4 \kappa_{03}^2 - 20\kappa_{11}^3 \kappa_{03} \kappa_{02} \kappa_{12} - 2\kappa_{11} \kappa_{03} \kappa_{02}^3 \kappa_{30} + 12\kappa_{11}^2 \kappa_{03} \kappa_{02}^2 \kappa_{21} \\
&\quad + 18\kappa_{12}^2 \kappa_{11}^2 \kappa_{02}^2 + 2\kappa_{12} \kappa_{02}^4 \kappa_{30} - 18\kappa_{12} \kappa_{02}^3 \kappa_{21} \kappa_{11} + 3\kappa_{21}^2 \kappa_{02}^4) \frac{I(2, 4)}{24\kappa_{02}^3 \Delta^2} \\
&- (4\kappa_{11}^3 \kappa_{02} \kappa_{03} \kappa_{10} g + 12\kappa_{11}^3 \kappa_{02} \kappa_{12} \bar{\kappa}_{01} + 4\kappa_{11}^3 \kappa_{02} \kappa_{13} - 12\kappa_{11}^2 \kappa_{02}^2 \kappa_{21} \bar{\kappa}_{01} \\
&\quad - 6\kappa_{11}^2 \kappa_{02}^2 \kappa_{22} - 12\kappa_{11}^2 \kappa_{02}^2 \kappa_{12} \bar{\kappa}_{10} - \kappa_{02}^4 \kappa_{10} - 4\kappa_{02}^4 \kappa_{30} \bar{\kappa}_{10} \\
&\quad + 12\kappa_{11} \kappa_{02}^3 \kappa_{21} \bar{\kappa}_{10} + 4\kappa_{11} \kappa_{02}^3 \kappa_{30} \bar{\kappa}_{01} + 4\kappa_{11} \kappa_{02}^3 \kappa_{31} - \kappa_{11}^4 \kappa_{04} \\
&\quad - 4\kappa_{11}^4 \kappa_{03} \bar{\kappa}_{01}) \frac{I(0, 4)}{24\kappa_{02}^3 \Delta^2} \\
&+ (3\kappa_{02}^2 \kappa_{12}^2 + 2\kappa_{02}^2 \kappa_{03} \kappa_{21} - 10\kappa_{11} \kappa_{03} \kappa_{12} \kappa_{02} + 5\kappa_{11}^2 \kappa_{03}^2) \frac{I(4, 2)}{24\Delta} \\
&+ (2\kappa_{02}^2 \kappa_{21} \bar{\kappa}_{01} + \kappa_{02}^2 \kappa_{22} + 2\kappa_{02}^2 \kappa_{12} \bar{\kappa}_{10} - 2\kappa_{11} \kappa_{02} \kappa_{03} \bar{\kappa}_{10} - 6\kappa_{11} \kappa_{02} \kappa_{12} \bar{\kappa}_{11} \\
&\quad - 2\kappa_{11} \kappa_{02} \kappa_{13} + \kappa_{11}^2 \kappa_{04} + 4\kappa_{11}^2 \kappa_{03} \bar{\kappa}_{01}) \frac{I(2, 2)}{4\kappa_{02}^2 \Delta} \\
&- (\kappa_{02}^2 \bar{\kappa}_{20} + \kappa_{02}^2 \bar{\kappa}_{10}^2 + \kappa_{11}^2 \bar{\kappa}_{02} + \kappa_{11}^2 \bar{\kappa}_{01}^2 - 2\kappa_{11} \kappa_{02} \bar{\kappa}_{01} \bar{\kappa}_{10} - 2\kappa_{11} \kappa_{02} \bar{\kappa}_{11}) \frac{I(0, 2)}{2\kappa_{02} \Delta}.
\end{aligned}$$

It remains to complete the integrals $I(p, q)$ to the range \mathbb{R}^2 (see (17)) and to calculate them according to (16). \square

PROOF OF LEMMA 8. In this proof, c_i denotes real positive constants. We use Lemma 6 and Cauchy's formula and obtain

$$[x^m y^m] \tilde{t}(x, y)^m t_1(x, y)^k t_2(x, y)^k = -\frac{1}{m} \frac{1}{4\pi^2} \oint \oint \frac{f(u, y)^m l(u, y)^k h(u, y)}{u^{m+1} y^{m+1}} dy du. \quad (18)$$

The corresponding saddle point is obtained using the equation system

$$\frac{\partial}{\partial u} (\log f(u, y) - \log u - \log y) = 0, \quad \frac{\partial}{\partial y} (\log f(u, y) - \log u - \log y) = 0,$$

obtaining the solution $u_0 = 1, y_0 = 1/e$. As in Lemma 3, the cummulants are

$$\kappa_{ij} = \left[\frac{\partial^i}{\partial \sigma^i} \frac{\partial^j}{\partial \tau^j} \log f(u_0 e^\sigma, y_0 e^\tau) \right]_{(\sigma, \tau) = (0, 0)},$$

particularly we obtain $\kappa_{20} = 0, \kappa_{11} = 0, \kappa_{02} = 1, \kappa_{30} = -4, \kappa_{21} = -1, \kappa_{12} = 0$, and $\kappa_{03} = 1$. In fact, we cannot proceed in the same way as in Lemma 3, because the

determinant $\Delta = \kappa_{20}\kappa_{02} - \kappa_{11}^2$ equals zero. Instead, we perform the substitution

$$y = e^{i\tau-1}, u = \begin{cases} e^{\zeta\sigma} & \text{if } \Im(u) \geq 0 \\ e^{\bar{\zeta}\sigma} & \text{if } \Im(u) < 0 \end{cases} \text{ where } \zeta = e^{i\frac{2\pi}{3}}.$$

We concentrate on the path in the upper half plane (the second part is similar) and obtain the integral

$$I = -\frac{i\zeta}{4\pi^2} \frac{1}{m} \int_0^{\frac{2\pi}{\sqrt{3}}} \int_{-\pi}^{\pi} \frac{f(e^{\zeta\sigma}, e^{i\tau-1})^m l(e^{\zeta\sigma}, e^{i\tau-1})^k h(e^{\zeta\sigma}, e^{i\tau-1})}{e^{\zeta\sigma m} e^{(i\tau-1)m}} d\tau d\sigma.$$

Let $\alpha \in (0, 2\pi/\sqrt{3})$ and $\beta \in (0, \pi)$. This choice will be discussed later. We divide the integral into two parts

$$I_1 = -\frac{i\zeta}{4\pi^2} \frac{1}{m} \int_0^{\alpha} \int_{-\beta}^{\beta} \frac{f(e^{\zeta\sigma}, e^{i\tau-1})^m l(e^{\zeta\sigma}, e^{i\tau-1})^k h(e^{\zeta\sigma}, e^{i\tau-1})}{e^{\zeta\sigma m} e^{(i\tau-1)m}} d\tau d\sigma,$$

and $I_2 = I - I_1$. We continue calculating an approximation of I_1 but before, we give an upper bound on I_2 .

Next, we consider the modulus of the function $f(e^{\zeta\sigma}, e^{i\tau-1}) e^{-\zeta\sigma} e^{1-i\tau}$ if $(\sigma, \tau) \in [0, 2\pi/\sqrt{3}] \times [-\pi, \pi]$. A plot can be found in Figure 4. The function is unimodal and attains its maximum at $(\sigma, \tau) = (0, 0)$. This can be verified using the mean value theorem, a bound on the derivative, and the evaluation of the function on points of a sufficient small grid. Further, a local expansion of this function around

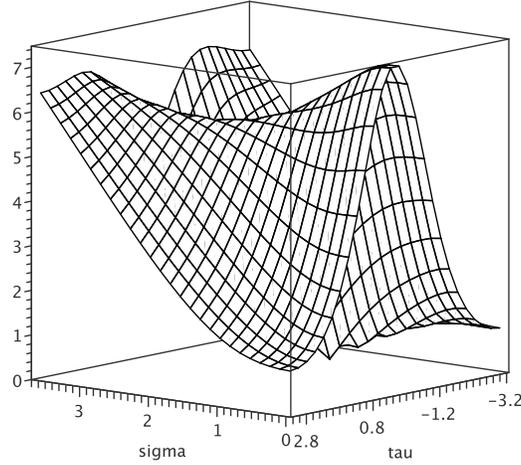


Fig. 4. The modulus of $f(e^{\zeta\sigma}, e^{i\tau-1}) e^{-\zeta\sigma} e^{1-i\tau}$ in the relevant area.

the origin is given by $\exp\left(2 - \frac{1}{2}\tau^2 - \frac{2}{3}\sigma^3 + O(\sigma^2\tau)\right)$. We choose

$$\alpha = m^{-\frac{7}{24}} \quad \text{and} \quad \beta = m^{-\frac{11}{24}},$$

such that $\alpha^3 m$ and $\beta^2 m$ tend to infinity as m goes to infinity, but on the other hand the terms of higher order like $\alpha^2 \beta m$ tend to zero. Hence, we conclude that the bound

$$\left| \frac{f(e^{\zeta\sigma}, e^{i\tau-1})}{e^{\zeta\sigma} e^{i\tau-1}} \right|^m \leq e^{2m - c_0 m^{-\frac{1}{8}}}$$

holds, if $\sigma \in [\alpha, 2\pi/\sqrt{3}]$ and $|\tau| \in [\beta, \pi]$ are satisfied. Thus, we further obtain the bound

$$|I_2| = \mathcal{O}\left(e^{2m - c_0 m^{-\frac{1}{8}}}\right).$$

Next, we calculate an approximation of I_1 with help of the following local expansions: (replace ζ by $\bar{\zeta}$ for the second case)

$$\begin{aligned} f\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right)^m &= \exp\left(m + \zeta s m^{\frac{2}{3}} + it\sqrt{m} - \frac{1}{2}t^2 - \frac{2}{3}s^3 + \mathcal{O}\left(m^{-\frac{1}{24}}\right)\right), \\ l\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right)^k &= \exp\left(\frac{2\zeta}{\sqrt[3]{m}}ks + \mathcal{O}\left(m^{-\frac{3}{24}+\xi}\right)\right), \\ h_k\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right) &= -2\zeta s m^{\frac{2}{3}} \left(1 + \mathcal{O}\left(m^{-\frac{3}{24}}\right)\right). \end{aligned}$$

Using these expansions, and assuming that $\xi \leq \frac{1}{12}$ holds, we infer:

$$\begin{aligned} I_1 &= \frac{-i\zeta}{4\pi^2 m^{\frac{11}{6}}} \int_0^{\alpha \sqrt[3]{m}} \int_{-\beta\sqrt{m}}^{\beta\sqrt{m}} \frac{f\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right)^m l\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right)^k h\left(e^{\frac{\zeta s}{\sqrt[3]{m}}}, e^{\frac{it}{\sqrt{m}}-1}\right)}{e^{\frac{\zeta s}{\sqrt[3]{m}}m} e^{\left(\frac{it}{\sqrt{m}}-1\right)m}} dt ds \\ &= \frac{i\zeta^2 e^{2m}}{2\pi^2 m^{\frac{7}{6}}} \int_0^{\alpha \sqrt[3]{m}} \int_{-\beta\sqrt{m}}^{\beta\sqrt{m}} s e^{-\frac{1}{2}t^2 - \frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} \left(1 + \mathcal{O}\left(m^{-\frac{1}{24}}\right)\right) dt ds \\ &= \frac{i\zeta^2 e^{2m}}{2\pi^2 m^{\frac{7}{6}}} \int_0^{\alpha \sqrt[3]{m}} \int_{-\beta\sqrt{m}}^{\beta\sqrt{m}} s e^{-\frac{1}{2}t^2 - \frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} dt ds + \mathcal{O}\left(m^{-\frac{29}{24}} e^{2m} \int_0^{\infty} s e^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} ds\right). \end{aligned}$$

We denote the remaining integral by I'_1 :

$$I'_1 = \int_0^{\alpha \sqrt[3]{m}} \int_{-\beta\sqrt{m}}^{\beta\sqrt{m}} s e^{-\frac{1}{2}t^2 - \frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} dt ds.$$

Our next step is to “complete the tails”:

$$\int_{\beta\sqrt{m}}^{\infty} e^{-\frac{1}{2}t^2} dt = \int_0^{\infty} e^{-\frac{1}{2}(h+\beta\sqrt{m})^2} dh = e^{-\frac{1}{2}\beta^2 m} \int_0^{\infty} e^{-\frac{1}{2}h^2 - h\beta\sqrt{m}} dh \leq \sqrt{\frac{\pi}{2}} e^{-\frac{1}{2}\beta^2 m},$$

$$\begin{aligned} \left| \int_{\alpha \sqrt[3]{m}}^{\infty} se^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} ds \right| &\leq \int_{\alpha \sqrt[3]{m}}^{\infty} se^{-\frac{2}{3}s^3 - \frac{ks}{\sqrt[3]{m}}} ds \\ &\leq e^{-\frac{2}{3}\alpha^3 m} \left(\int_0^{\infty} he^{-\frac{2}{3}h^3} dh + \alpha \sqrt[3]{m} \int_0^{\infty} e^{-\frac{2}{3}h^3} dh \right) \leq (c_1 + c_2 \alpha \sqrt[3]{m}) e^{-\frac{2}{3}\alpha^3 m}. \end{aligned}$$

We apply this bounds and get the representation

$$\begin{aligned} I'_1 &= \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt + \mathcal{O}(e^{-c_3\beta^2 m}) \right) \left(\int_0^{\infty} se^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} ds + \mathcal{O}(e^{-c_4\alpha^3 m}) \right) \\ &= \sqrt{2\pi} \int_0^{\infty} se^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} ds + \mathcal{O}(e^{-c_4\alpha^3 m}) + \mathcal{O}(e^{-c_3\beta^2 m}). \end{aligned}$$

The second part of the integral representing (18) has the main contribution

$$-\frac{\sqrt{2\pi}i\zeta^2 e^{2m}}{2\pi^2 m^{\frac{7}{6}}} \int_0^{\infty} se^{-\frac{2}{3}s^3 + \frac{2\zeta}{\sqrt[3]{m}}ks} ds$$

and the error terms are of the same order as before. \square

REFERENCES

- BLASIAK, J. AND DURRETT, R. 2005. Random oxford graphs. *Stochastic Process. Appl.* 115, 8, 1257–1278.
- BOLLOBÁS, B. 2001. *Random Graphs*, Second ed. Cambridge University Press, Cambridge, UK.
- CONWAY, J. B. 1978. *Functions of One Complex Variable*, Second ed. Springer, New York.
- CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L., AND STEIN, C. 2001. *Introduction to Algorithms*, Second ed. MIT Press, Cambridge, Mass., London, England.
- CZYZOWICZ, J., FRACZAK, W., AND WELFED, F. 2006. Notes about d-cuckoo hashing. Tech. Rep. RR 06/05-1, Université du Québec en Outaouais.
- DEVROYE, L. AND MORIN, P. 2003. Cuckoo hashing: Further analysis. *Information Processing Letters* 86, 4, 215–219.
- DIESTEL, R. 2005. *Graph Theory*. Springer, Berlin.
- DIETZFELBINGER, M., GIL, J., MATIAS, Y., AND PIPPENGER, N. 1992. Polynomial hash functions are reliable (extended abstract). In *ICALP '92*. LNCS, vol. 623. Springer, 235–246.
- DIETZFELBINGER, M. AND WEIDLING, C. 2007. Balanced allocation and dictionaries with tightly packed constant size bins. *Theoretical Computer Science* 380, 1-2, 47–68.
- DIETZFELBINGER, M. AND WOELFEL, P. 2003. Almost random graphs with simple hash functions. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. ACM, 629–638.
- DRMOTA, M. 1994. A bivariate asymptotic expansion of coefficients of powers of generating functions. *European Journal of Combinatorics* 15, 2, 139–152.
- DRMOTA, M. AND SORIA, M. 1995. Marking in combinatorial constructions: Generating functions and limiting distributions. *Theoretical Computer Science* 144, 1&2, 67–99.
- DRMOTA, M. AND SORIA, M. 1997. Images and preimages in random mappings. *SIAM Journal on Discrete Mathematics* 10, 2, 246–269.
- FLAJOLET, P. AND ODLYZKO, A. M. 1982. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences* 25, 2, 171–213.
- FLAJOLET, P. AND ODLYZKO, A. M. 1990. Random mapping statistics. *LNCS* 434, 329–354.

- FLAJOLET, P. AND SEDGEWICK, R. 2009. *Analytic Combinatorics*. Cambridge University Press, Cambridge, UK.
- FOTAKIS, D., PAGH, R., SANDERS, P., AND SPIRAKIS, P. G. 2005. Space efficient hash tables with worst case constant access time. *Theory Comput. Syst.* 38, 2, 229–248.
- GARDY, D. 1995. Some results on the asymptotic behaviour of coefficients of large powers of functions. *Discrete Mathematics* 139, 1-3, 189–217.
- GIMENEZ, O., DE MIER, A., AND NOY, M. 2005. On the number of bases of bicircular matroids. *Annals of Combinatorics* 9, 1, 35–45.
- GONNET, G. H. AND BAEZA-YATES, R. 1991. *Handbook of algorithms and data structures: in Pascal and C*, Second ed. Addison-Wesley, Boston, MA, USA.
- GOOD, I. J. 1957. Saddle-point methods for the multinomial distribution. *Ann. Math. Stat.* 28, 4, 861–881.
- GREENE, D. H. AND KNUTH, D. E. 1982. *Mathematics for the Analysis of Algorithms*, Second ed. Birkhäuser, Boston.
- JANSON, S., KNUTH, D. E., LUCZAK, T., AND PITTEL, B. 1993. The birth of the giant component. *Random Structures and Algorithms* 4, 3, 233–359.
- KALUGIN, I. B. 1991. The number of components of a random bipartite graph. *Discrete Mathematics and Applications* 1, 3, 289–299.
- KNUTH, D. E. 1998. *The Art of Computer Programming, Volume III: Sorting and Searching*, Second ed. Addison-Wesley, Boston.
- KUTZELNIGG, R. 2006. Bipartite random graphs and cuckoo hashing. In *Proceedings of the 4th Colloquium on Mathematics and Computer Science*. Discrete Mathematics and Theoretical Computer Science. 403–406.
- KUTZELNIGG, R. 2008. An improved version of cuckoo hashing: Average case analysis of construction cost and search operations. In *Proceedings of the 19th international workshop on combinatorial algorithms*. 253–266.
- KUTZELNIGG, R. 2009. *Random Graphs and Cuckoo Hashing*. Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken.
- PAGH, R. AND RODLER, F. F. 2004. Cuckoo hashing. *Journal of Algorithms* 51, 2, 122–144.
- RIORDAN, J. 1968. *Combinatorial Identities*. Wiley, New York.
- ROSS, K. A. 2006. Efficient hash probes on modern processors. IBM Research Report RC24100, IBM.
- SCOINS, H. I. 1962. The number of trees with nodes of alternate parity. *Proc. Cambridge Philos. Soc.* 58, 12–16.
- TRAN, T. N. AND KITTITORNKUN, S. 2007. Fpga-based cuckoo hashing for pattern matching in nids/nips. In *APNOMS*. 334–343.
- ZWILLINGER, D. 1992. *Handbook of differential equations*, Second ed. Academic Press, Boston.

Received March 2009;