On the insertion time of random walk cuckoo hashing

Alan Frieze*and Tony Johansson†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
U.S.A.
alan@random.math.cmu.edu
tjohanss@andrew.cmu.edu

Abstract

Cuckoo Hashing is a hashing scheme invented by Pagh and Rodler [14]. It uses $d \geq 2$ distinct hash functions to insert n items into the hash table of size $m = (1+\varepsilon)n$. It has been an open question for some time as to the expected time for Random Walk Insertion to add items. We show that if the number of hash functions $d \geq d_\varepsilon = O(1)$ then the expected insertion time is $O(1)$ per item.

1 Introduction

Our motivation for this paper comes from Cuckoo Hashing (Pagh and Rodler [14]). Briefly each one of n items $x \in L$ has d possible locations $h_1(x), h_2(x), \ldots, h_d(x) \in R$, where d is typically a small constant and the h_i are hash functions, typically assumed to behave as independent fully random hash functions. (See [13] for some justification of this assumption.) We note that [14] only considered the case $d = 2$ and it was Fotakis, Pagh, Sanders and Spirakis [4] who first considered the generalisation to $d \geq 2$.

We assume each location can hold only one item. Items are inserted consecutively and when an item x is inserted into the table, it can be placed immediately if one of its d locations is currently empty. If not, one of the items in its d locations must be displaced and moved to another of its d choices to make room for x. This item in turn may need to displace another item out of one of its d locations. Inserting an item may require a sequence of moves, each maintaining the invariant that each item remains in one of its d potential locations, until no further evictions are needed.

We now give the formal description of the mathematical model that we use. We are given two disjoint sets $L = \{v_1, v_2, \ldots, v_n\}, R = \{w_1, w_2, \ldots, w_m\}$. Each $v \in L$ independently chooses a set $N(v)$ of $d \geq 2$ uniformly random neighbors in the set R of size $m = (1+\varepsilon)n$. We assume for simplicity that this selection is done with replacement. This provides us with the bipartite cuckoo graph Γ. Cuckoo Hashing can be thought of as a simple algorithm for finding a matching M of

*Research supported in part by NSF Grants DMS1362785, CCF1522984 and Grant 333329 from the Simons Foundation

†Research supported in part by NSF Grant DMS1362785

L into R in Γ. In the context of hashing, if $\{x, y\}$ is an edge of M then $y \in R$ is a hash value of $x \in L$.

Cuckoo Hashing constructs M by defining a sequence of matchings $M_1, M_2, \ldots, M_n$, where M_k is a matching of $L_k = \{v_1, v_2, \ldots, v_k\}$ into R. We let Γ_k denote the subgraph of Γ induced by $L_k \cup R$. We let R_k denote the vertices of R that are covered by M_k and define the function $\phi_k : L_k \to R_k$ by asserting that $M_k = \{\{v, \phi_k(v)\} : v \in L_k\}$. We obtain M_k from M_{k-1} by finding an augmenting path P_k in Γ_k from v_k to a vertex in $\bar{R}_{k-1} = R \setminus R_{k-1}$.

This augmenting path P_k is obtained by a random walk. To begin we obtain M_1 by letting $\phi_1(v_1)$ be a uniformly random member of $N(v_1)$, the neighbors of v_1. Having defined M_k we proceed as follows: Steps 1 – 4 constitute round k.

Algorithm INSERT:

Step 1 $x \leftarrow v_k$; $M \leftarrow M_{k-1}$;

Step 2 **If** $S_k(x) = N(x) \cap \bar{R}_{k-1} \neq \emptyset$ **then** choose y uniformly at random from $S_k(x)$ and let $M_k = M \cup \{\{x, y\}\}$, **else**

Step 3 Choose y uniformly at random from $N(x)$;

Step 4 $M \leftarrow M \cup \{\{x, y\}\} \setminus \left\{y, \phi_{k-1}^{-1}(y)\right\}$; $x \leftarrow \phi_{k-1}^{-1}(y)$; **goto** Step 2.

Comment: $N(x)$ is a multi-set and so in Step 2, neighbors are selected proportional to multiplicity.

This algorithm was first discussed in the conference version of [4]. Our interest here is in the expected time for INSERT to complete a round. Our results depend on d being a large enough constant. In this case we will improve on the bounds on insertion time given in Frieze, Melsted and Mitzenmacher [7], Fountoulakis, Panagiotou and Steger [5], Fotakis, Pagh, Sanders and Spirakis [4]. The paper [4] studied the efficiency of insertion via Breadth First Search and also reported on some experiments with the random walk approach. The paper by Khosla [10] describes a nice linear time algorithm for placing the n items. The papers [7] and [5] considered insertion by random walk and proved that the expected time to complete a round can be bounded by $\log^{2+o_d(1)} n$, where $o_d(1)$ tends to zero as $d \to \infty$. The paper [5] improved on the space requirements in [7]. They showed that given ε, their bounds hold for any d large enough to give the existence of a matching w.h.p. Mitzenmacher [12] gives a survey on Cuckoo Hashing and Problem 1 of the survey asks for the expected insertion time.

Frieze and Melsted [6], Fountoulakis and Panagiotou [3] give information on the relative sizes of L, R needed for there to exist a matching of L into R w.h.p. Our choice of d will be such that there is a matching of L into R w.h.p. Conditioning on there being a matching can only increase the probability of an event by a factor $1 + o(1)$ and we will not explicitly include this factor in the calculations below.

We will prove the following theorem. We show that the expected insertion time is $O(1)$, but only for a large value of d. When the load factor n/m is close to zero then the expected size of the component of Γ containing a fixed item can easily be shown to be bounded. Thus the theorem below focusses on the more interesting case where the load factor n/m is close to one.

Theorem 1 *Suppose that $n = (1-\varepsilon)m$ where ε is a fixed positive constant, assumed to be small. Let $0 < \theta < 1$ also be a fixed positive constant and let*

$$\gamma = 5(1-\varepsilon)^{d/2}. \tag{1}$$

If $d^2\gamma \leq (1-\theta)(d-1)$ then w.h.p. the structure of Γ is such that over the random choices in Steps 2,3,

$$\mathsf{E}(|P_k|) \leq 1 + \frac{2}{\theta} \text{ for } k = 1, 2, \ldots, n. \tag{2}$$

Here $|P_k|$ is the length (number of edges) of P_k.

When d is large the value of θ in (2) is close to $d\varepsilon/2$. It can be seen from the proof that as $\varepsilon \to 0$, the value of d needed is of the order $\left(\frac{\log 1/\varepsilon}{\varepsilon}\right)$. This is larger than the value $O(\log(1/\varepsilon))$ needed for there to be a perfect matching from L to R and finding an $O(1)$ bound on the expected insertion time for small d remains as an open problem. We note that Theorem 1 of [4] allows small values of $d \geq 5 + 3\log 1/\varepsilon$, but the BFS algorithm it relies on requires more space, shown to be $O(n^\delta)$ extra space for a constant $\delta > 0$, and is shown to have an expected insertion time of $d^{O(\log 1/\varepsilon)}$, which is actually worse than than that implied by our theorem.

The problem here bears some relation to the *On-line bipartite matching problem* discussed for example in Chaudhuri, Daskalakis, Kleinberg and Lin [2], Bosek, Leniowski, Sankowski and Zych [1] and Gupta, Kumar and Stein [8]. In these papers the bipartite graph is arbitrary and has a perfect matching and vertices on one side A of the bipartition arrive in some order, along with their choice of neighbors in the other side B. As each new member of A arrives, a current matching is updated via an augmenting path. The aim is to keep the sum of the lengths of the augmenting paths needed to be as small as possible. It is shown, among other things, in [2] that this sum can be bounded by $O(n \log n)$ in expectation and w.h.p. This requires finding a shortest augmenting path each time. Our result differs in that our graph is random and $|A| = (1-\varepsilon)|B|$ and we only require a matching of A into B. On the other hand we obtain a sum of lengths of augmenting paths of order $O(n)$ in expectation via a random choice of path.

2 Proof of Theorem 1

2.1 Outline of the main ideas

Let

$$B_k = \left\{v \in L_k : N(v) \cap \bar{R}_{k-1} = \emptyset\right\}.$$

If $x \notin B_k$ in Step 2 of INSERT then we will have found P_k.

Let $P = (x_1, \xi_1, x_2, \xi_2, \ldots, x_\ell)$ be a path in Γ, where $x_1, x_2, \ldots, x_\ell \in L_{k-1}$ and $\xi_1, \xi_2, \ldots, \xi_{\ell-1} \in R_{k-1}$. We say that P is *interesting* if $x_1, x_2, \ldots, x_\ell \in B_k$. We note that if the path $P_k = (x_1 = v_k, \xi_1, x_2, \xi_2, \ldots, x_\ell, \xi_\ell, x_{\ell+1}, \xi_{\ell+1})$ then $Q_k = (x_1, \xi_1, x_2, \xi_2, \ldots, x_\ell)$ is interesting. Indeed, we must have $x_i \in B_k$, $1 \leq i \leq \ell$, else INSERT would have chosen $\xi_i \in \bar{R}_{k-1}$ and completed the round.

Our strategy is simple. We show that w.h.p. there are relatively few long interesting paths and because our algorithm (usually) chooses a path at random, it is unlikely to be long and interesting. One caveat to this approach is that while all augmenting paths yield interesting sub-paths, the reverse is not the case. That being so, it would be better to estimate the number of possible long augmenting paths. The problem with this approach is that we then need to control the distribution of the matching M_k. This has been the difficulty up to now and we have avoided the problem by focussing on interesting paths. Of course, there is a cost in that d is larger than one would like, but it is at least independent of n.

To bound the number of interesting paths, we bound $|B_k|$ and use this to bound the number of paths.

The probability space for the analysis is the product of two spaces. One generates Γ and the other generates the walks. One observes that except for (8) below, our probability statements are actually in respect of Γ and are high probablity statements. It is only in (8) that we explicitly consider the random walks and there we suffice with a bound on expectation.

2.2 Detailed proof

Let y_k be the element y chosen at Step 3 of round $k = 1, 2, \dots, n$. Then, fix $1 \leq k \leq n$. We observe that if $R_{k-1} = \{y_1, y_2, \dots, y_{k-1}\}$ then

$$y_k \text{ is chosen uniformly from } \bar{R}_{k-1} \tag{3}$$

and is independent of the graph Γ_{k-1} induced by $L_{k-1} \cup R_{k-1}$. This is because we can expose Γ along with the algorithm. When we start the construction of M_k we expose the neighbors of v_k one by one. In this way we either determine that $S_k(v_k) = \emptyset$ or we expose a uniformly random member of $S_k(v_k)$ without revealing any more of $N(v_k)$. In general, in Step 2, we have either exposed all the neighbors of x and these will necessarily be in R_{k-1}. Or, we can proceed to expose the unexposed neighbors of x until either (i) we determine that $S_k(x) = \emptyset$ and we choose a uniformly random member of $N(x)$ or (ii) we find a neighbor of x that is a uniformly random member of $\bar{R}_{k-1}$. Thus

$$R_{k-1} \text{ is a uniformly random subset of } R.$$

We need to show that B_k is small. It is clear that $v_1 \notin B_1$ i.e. $B_1 = \emptyset$ and so we deal next with $2 \leq k \leq d$. If $v_k \in B_k$ then $N(v_k) \subseteq R_{k-1}$ and when v_k is considered, R_{k-1} is fixed. But then

$$\mathsf{P}(\exists 2 \leq k \leq d : N(v_k) \subseteq R_{k-1}) \leq (d-1)\left(\frac{d-1}{m}\right)^d = O(n^{-d}) = o(1).$$

This implies that w.h.p. $B_k = \emptyset$ for $2 \leq k \leq d$. We deal next with $d+1 \leq k \leq n^{9/10}$. Since $N(v_k)$ is uniformly random, we see that

$$\mathsf{P}(\exists k \leq n^{9/10} : v_k \in B_k) \leq n^{1-d/10} = o(1) \text{ for } d > 10.$$

Now fix $n^{9/10} \leq k \leq n-1$. Let $\nu_{k,\ell}$ denote the number of interesting paths with $2\ell - 1$ vertices. Let $\theta, \gamma > 0$ be as in the statement of Theorem 1.

Lemma 2 *Let A_0 be an arbitrary positive constant and assume that d is sufficiently large,*

$$\mathsf{P}\left(\exists 2 \leq \ell \leq A_0 \log\log n : \nu_{k,\ell} \geq (1+\theta)k\gamma(d^2\gamma)^{\ell-1}\right) = o(n^{-2}). \tag{4}$$

The bound $o(n^{-2})$ is sufficient to deal with the insertion of n items.

Before proving the lemma, we show how it can be used to prove Theorem 1. We will need the following claims:

Claim 3 *Let Δ denote the maximum degree in Γ. Then for any $t \geq \log n$ we have $\mathsf{P}(\Delta \geq t) \leq e^{-t}$.*

Proof of Claim: If $v \in L$ then its degree $\deg(v) = d$. Now consider $w \in R$. Then for $t \geq \log n$,

$$\mathsf{P}(\exists w \in R : \deg(w) \geq t) \leq m \binom{dn}{t} \frac{1}{m^t} \leq m \left(\frac{de}{t}\right)^t \leq e^{-t}.$$

End of proof of Claim

Claim 4 *With probability $1 - o(n^{-2})$, Γ contains at most $n^{1/2+o(1)}$ cycles of length at most $\Lambda = (\log \log n)^2$.*

Proof of Claim: Let C denote the number of cycles of length at most $2\ell = \Lambda$. Then

$$\mathsf{E}(C) \leq \sum_{s=2}^{\ell} \binom{n}{s} \binom{m}{s} (s!)^2 \left(\frac{d}{m}\right)^{2s} \leq \sum_{s=2}^{\ell} d^{2s} = n^{o(1)}.$$

Now let C_1 denote the number of cycles of length at most Λ where we can only use the first $\min\{\lambda, \deg(w)\}$, $\lambda = (\log n)^2$ edges incident with each vertex $w \in R$. Here "first" is defined in some canonical way. Then $C_1 \leq C$ and Claim 3 implies that

$$\mathsf{P}(C \neq C_1) \leq e^{-\lambda} = e^{-(\log n)^2}. \tag{5}$$

Now C_1 depends on the dn independent choices of edges in Γ and changing one choice of edge can only change C_1 by at most λ^Λ. Applying McDiarmid's inequality, and using $\mathsf{E}(C_1) \leq \mathsf{E}(C)$ we see that

$$\mathsf{P}(C_1 \geq \mathsf{E}(C) + n^{1/2}\lambda^{2\Lambda}) \leq \exp\left\{-\frac{2n\lambda^{4\Lambda}}{dn\lambda^{2\Lambda}}\right\} = o(n^{-2}).$$

Together with (5), this proves the claim.

McDiarmid's Inequality [11]: Let $Z = Z(Y_1, Y_2, \ldots, Y_N)$ be a function of N independent random variables $Y_1, Y_2, \ldots, Y_N$. Suppose that changing the value of a single Y_i can change the value of Z by at most c. Then

$$\mathsf{P}(Z - \mathsf{E}(Z) \geq t) \leq \exp\left\{-\frac{2t^2}{Nc^2}\right\}.$$

End of proof of Claim

These two claims imply the following property of Γ:

$$\text{With probability } 1 - o(n^{-2}) \text{ there are at most } n^{1/2+o(1)}(3\log n)^{2\ell} = n^{1/2+o(1)} \text{ vertices within distance at most } 2A_0 \log\log n \text{ of a cycle of length at most } \Lambda = (\log\log n)^2. \tag{6}$$

Now let $p_{k,\ell}$ denote the probability that INSERT requires at least ℓ steps to insert v_k.

We finish the proof of the theorem by showing that

$$\mathsf{E}(|P_k|) = 1 + 2\sum_{\ell=2}^{\infty} p_{k,\ell} \leq 1 + \frac{2}{\theta}. \tag{7}$$

Here we are using $\mathsf{E}(Z) = \sum_{k\geq 1} \mathsf{P}(Z \geq k)$ for non-negative integer valued Z. We observe that if v_k has no neighbor in $\bar{R}_{k-1}$ and has no neighbor in a cycle of length at most Λ then for some $\ell \leq A_0 \log\log n$, the first $2\ell-1$ vertices of P_n follow an interesting path. Hence, if $d^2\gamma \leq (1-\theta)(d-1)$ then

$$\sum_{\ell=2}^{A_0 \log\log n} p_{k,\ell} \leq O(n^{-1/2+o(1)}) + \sum_{\ell=2}^{A_0 \log\log n} \frac{\nu_{k,\ell}}{k(d-1)^{\ell}}. \tag{8}$$

Explanation of (8)**:** Following (6), we find that the probability v_k is within $2A_0 \log\log n$ of a cycle of length at most Λ is bounded by $n^{-1/2+o(1)}$. The $O(n^{-1/2+o(1)})$ term accounts for this and also absorbs the error probability in (4). Failing this, we have divided the number of interesting paths of length $2\ell - 1$ by the number of equally likely walks $k(d-1)^{\ell}$ that INSERT could take. To obtain $k(d-1)^{\ell}$ we argue as follows. We carry out the following thought experiment. We run our walk for ℓ steps regardless. If we manage to choose $y \in \bar{R}_{k-1}$ then instead of stopping, we move to v_k and continue. In this way there will in fact be $k(d-1)^{\ell}$ equally likely walks. In our thought experiment we choose one of these walks at random, whereas in the execution of the algorithm we only proceed as far the first time we reach $\bar{R}_{k-1}$.

Using (4) and (8) we now see that

$$\begin{aligned}\sum_{\ell=2}^{A_0 \log\log n} p_{k,\ell} &\leq O(n^{-1/2+o(1)}) + (1+\theta) \sum_{\ell=2}^{A_0 \log\log n} \frac{k\gamma(d^2\gamma)^{\ell-1}}{k(d-1)^{\ell}} \\ &\leq o(1) + (1+\theta)\sum_{\ell=2}^{\infty}(1-\theta)^{\ell-1} \\ &= o(1) + \frac{1-\theta^2}{\theta}.\end{aligned}$$

Finally, for the algorithm to take at least ℓ steps, it must choose an interesting path of length at least $2\ell - 1$.

Note next that

$$p_{k,A_0 \log\log n} \leq O(n^{-1/2+o(1)}) + 3^{-A_0 \log\log n}.$$

It follows that

$$\sum_{\ell=A_0 \log\log n}^{(\log n)^{A_0}} p_{k,\ell} \leq \sum_{\ell=A_0 \log\log n}^{(\log n)^{A_0}} p_{k,A_0 \log\log n} = o(1). \tag{9}$$

We will use the result of [7]: We phrase Claim 10 of that paper in our current terminology.

Claim 5 *There exists a constant $a > 0$ such that for any $v \in L_{k-1}$, the expected time for* INSERT *to reach $\bar{R}_{k-1}$ is $O((\log k)^a)$.*

It follows from Claim 5 that for any integer $\rho \geq 1$,

$$\mathsf{P}(|P_k| \geq \rho(\log k)^{2a}) \leq \frac{1}{(\log k)^{\rho a}}. \tag{10}$$

Indeed, we just apply the Markov inequality every $(\log k)^{2a}$ steps to bound $|P_k|$ by a geometric random variable.

It follows from (10) that

$$\sum_{\ell \geq 3(\log k)^{2a}} p_{k,\ell} \leq \sum_{\rho=3}^{\infty} \sum_{\ell/(\log k)^{2a} \in [\rho,\rho+1]} p_{k,\ell} \leq \sum_{\rho=3}^{\infty} \frac{1}{(\log k)^{\rho a - 2a}} = o(1). \tag{11}$$

Theorem 1 now follows from (7), (8), (9) and (11), if we take $A_0 > 2a$.

2.3 Proof of Lemma 2

We will argue as in the proof of Claim 4 that

$$\mathsf{P}(\nu_{k,\ell} \geq \mathsf{E}(\nu_{k,\ell}) + n^{3/4}) \leq 2e^{-(\log n)^2}. \tag{12}$$

We let $\nu^*_{k,\ell}$ be the number of interesting paths that only use the first $\min\{\lambda, \deg(w)\}$, $\lambda = (\log n)^2$ edges incident with vertex $w \in R$. Then $\nu^*_{k,\ell} \leq \nu_{k,\ell}$ and Claim 3 implies that

$$\mathsf{P}(\nu_{k,\ell} \neq \nu^*_{k,\ell}) \leq e^{-\lambda} = e^{-(\log n)^2}. \tag{13}$$

Now $\nu^*_{k,\ell}$ depends on the dk independent choices of edges in Γ_k and changing one choice of edge can only change $\nu^*_{k,\ell}$ by at most $\lambda^{2\ell}$. Applying McDiarmid's inequality, and using $\mathsf{E}(\nu^*_{k,\ell}) \leq \mathsf{E}(\nu_{k,\ell})$ we see that

$$\mathsf{P}(\nu^*_{k,\ell} \geq \mathsf{E}(\nu_{k,\ell}) + n^{3/4}) \leq \exp\left\{-\frac{2n^{3/2}}{dk\lambda^{4\ell}}\right\} \leq e^{-(\log n)^2}.$$

Together with (13), this proves (12).

It follows from (12) that to finish the proof, all we need to show is that if $\theta > 0$ is an arbitrary positive constant then we have

$$\mathsf{E}(\nu_{k,\ell}) \leq (1+\theta)k\gamma(d^2\gamma)^{\ell-1}, \tag{14}$$

where γ is as in (1).

Claim 6 *Let*

$$\mathcal{B}_k = \{|B_k| \geq k\gamma\}.$$

Then

$$\mathsf{P}(\mathcal{B}_k) = O(e^{-\Omega(n^{1/2})}).$$

Proof of Claim:
Let $B_{k,1}$ denote the set of vertices $v_i \in L_k$ such that round i exposes at least $d/2$ edges incident with v_i. Then

$$\mathsf{P}(v_i \in B_{k,1}) \leq (1-\varepsilon)^{d/2}.$$

Now $|B_{k,1}|$ is the sum of independent $\{0,1\}$ random variables and so Hoeffding's theorem [9] implies that

$$\mathsf{P}\left(|B_{k,1}| \geq 2k(1-\varepsilon)^{d/2}\right) \leq e^{-\Omega(n^{1/2})}. \tag{15}$$

Next let

$$B_{k,2} = \{s \leq k : \text{ round } s \text{ does not end immediately in Step 2 with } x = v_s.\}$$

Then, $\mathsf{P}(s \in B_{k,2}) = \left(\frac{s-1}{m}\right)^d$ and this holds for each value of s independently and so

$$\mathsf{E}(|B_{k,2}|) \leq \sum_{s=1}^{k} \left(\frac{s-1}{m}\right)^d \leq \frac{k^{d+1}}{(d+1)m^d}.$$

Now $|B_{k,2}|$ is the sum of independent $\{0, 1\}$ random variables and so again we have that Hoeffding's theorem [9] implies that for a constant $\theta > 0$,

$$\mathsf{P}\left(|B_{k,2}| \geq (1+\theta)\frac{k^{d+1}}{(d+1)m^d}\right) = O(e^{-\varepsilon_1 k}) \text{ for some constant } \varepsilon_1 = \varepsilon_1(d, \varepsilon, \theta) > 0. \tag{16}$$

Now if $B_{k,3} = \{s \in B_k : \exists \ell \leq k, \ell \neq s \text{ s.t. round } \ell \text{ ends with } x = v_s\}$ then $|B_{k,3}| \leq |B_{k,2}|$. Define $B_{k,4} = B_k \setminus (B_{k,1} \cup B_{k,2} \cup B_{k,3})$. Let $t > s$ be the first time that v_s is re-visited by INSERT or let $t = k$ if v_s is not re-visited. Then $s \in B_{k,4}$ only if in round t, at least $d/2$ unexposed edges incident to s are found to be in R_{t-1}. It follows that

$$E(|B_{k,4}|) \leq k(1-\varepsilon)^{d/2}.$$

Since membership of s in $B_{k,4}$ is determined by the random choices of v_s, $|B_{k,4}|$ is the sum of independent random variables and so

$$\mathsf{P}\left(|B_{k,4}| \geq 2k(1-\varepsilon)^{d/2}\right) = O(e^{-\Omega(n^{1/2})}). \tag{17}$$

The claim follows from (15), (16) and (17).
End of proof of Claim

Given Claim 6, we have

$$\begin{aligned}\mathsf{E}(\nu_{k,\ell}) &= \mathsf{E}(\nu_{k,\ell} \mid \neg\mathcal{B}_k)\mathsf{P}(\neg\mathcal{B}_k) + \mathsf{E}(\nu_{k,\ell} \mid \mathcal{B}_k)\mathsf{P}(\mathcal{B}_k) \\ &\leq k^\ell \gamma^\ell k^{\ell-1} \cdot \left((1+o(1))\frac{d}{k}\right)^{2\ell-2} + O(k^{2\ell-1} \cdot e^{-\Omega(n^{1/4})}), \\ &\leq (1+o(1))k\gamma(d^2\gamma)^{\ell-1} + o(1).\end{aligned} \tag{18}$$

This proves (14).

Explanation of (18)**:** We can choose the vertex sequence $\sigma = (x_1, \xi_1, \ldots, \xi_{\ell-1}, x_\ell)$ of an interesting path P in at most $|B_k|^\ell k^{\ell-1}$ ways, and we apply Claim 6. Having chosen σ we see that $((1 + o(1))d/k)^{2\ell-2}$ bounds the probability that the edges of P exist. To see this, condition on $\bar{R}_{k-1}$ and the random choices for vertices not on P. In particular, we can fix $R_{k-1} = \{y_1, y_2, \ldots, y_{k-1}\}$ from the beginning and this simply constrains the sequence of choices $y_1, y_2, \ldots, y_{k-1}$ to be a uniformly random permutation of R_{k-1}. Let $\mathcal{M}_k$ be the property that Γ has a matching from L_k to R. It is known that $\mathsf{P}(\mathcal{M}_k) = 1 - O(n^{4-d})$. This will also be true conditional on the value of $\bar{R}_{k-1}$. This follows by symmetry. The conditional spaces will be isomorphic to each other. So for large d, we can assume that our conditioning is such that with probability $1 - O(1/n^3)$ the edge choices by $x_1, x_2, \ldots, x_\ell$ are such that Γ_k has property $\mathcal{M}_k$ with probability $1 - O(n^{7-d})$. Recall from (3) that the disposition of the edges of Γ_{k-1} is independent of $\bar{R}_{k-1}$. Now each edge adjacent to a given

$x \in \sigma \cap L_k$ is a uniform choice over those edges consistent with x being in B_k. But there will be at least $k-1$ such choices for such an x viz. the vertices of R_{k-1}. Thus

$$\mathsf{P}(P \text{ exists} \mid \mathcal{M}_k) \leq \frac{\mathsf{P}(P \text{ exists})}{\mathsf{P}(\mathcal{M}_k)} \leq (1+o(1)) \left(\frac{d}{k}\right)^{2\ell-2}.$$

Note that $\mathsf{P}(\bar{\mathcal{M}}_k)$ is only inflated by at most $\frac{1}{(1-\varepsilon)^{d\ell}} = o(n^{o(1)})$ if we condition on $x_1, x_2, \ldots, x_\ell$ making their choices in $\bar{R}_{k-1}$. This has to be compared with the unconditional probability of $O(n^{7-d})$.

This completes the proof of Theorem 1. □

Remark 7 *Along with an upper bound, we can prove a simple lower bound:*

$$\mathsf{E}(|P_k|) \geq \frac{2}{1-(1-\varepsilon)^d}.$$

This follows from the fact that Step 2 of INSERT *ends the procedure with probability* $1-(1-\varepsilon)^{|S_k(x)|}$ *and* $|S_k(x)| \leq d$.

3 Final Remarks

There is plenty of room for improvement in the bounds on d in Theorem 1. It would be most interesting to prove an $O(1)$ bound on the expected insertion time for small d, e.g. $d = 3, 4, 5$. This no doubt requires an understanding of the evolution of the matching M. It would also be of interest to consider the effect of deleting (random) elements. It is tempting to say that the analysis is robust enough but we won't make that claim.

Acknowledgement: We thank Wesley Pegden for his comments. We also thank Lutz Warnke for pointing out a minor error and one of the reviewers for poining out another minor error.

References

[1] B. Bosek, D. Leniowski, P. Sankowski and A. Zych, Online bipartite matching in offline time, *55th Annual IEEE Symposium on Foundations of Computer Science* 2014.

[2] K. Chaudhuri, C. Daskalakis, R. Kleinberg and H. Lin, Online Bipartite Matching With Augmentations, *28th IEEE INFOCOM* (2009) 1044-1052.

[3] N. Fountoulakis and K. Panagiotou, Sharp Load Thresholds for Cuckoo Hashing, *Random Structures and Algorithms* 41 (2012) 306-333.

[4] D. Fotakis, R. Pagh, P. Sanders, and P. Spirakis, Space Efficient Hash Tables With Worst Case Constant Access Time, *Theory of Computing Systems* 8 (2005) 229-248.

[5] N. Fountoulakis, K. Panagiotou and A. Steger, On the Insertion Time of Cuckoo Hashing, *SIAM Journal on Computing* 42 (2013) 2156-2181.

[6] A.M. Frieze and P. Melsted, Maximum Matchings in Random Bipartite Graphs and the Space Utilization of Cuckoo Hashtables, *Random Structures and Algorithms* 41 (2012) 334-364.

[7] A.M. Frieze, P. Melsted and M. Mitzenmacher, An Analysis of Random-Walk Cuckoo Hashing, *SIAM Journal on Computing* 40 (2011) 291-308.

[8] A. Gupta, A. Kumar and C. Stein, Maintaining Assignments Online: Matching, Scheduling, and Flows, *Proceedings of the 25th Annual ACM-SIAM Symposium on Disrete Algorithms* (2014)

[9] W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* 58 (1963) 13-30.

[10] M. Khosla, Balls into bins made faster, Proceedings of ESA 2013, 601-612.

[11] C. McDiarmid, On the method of bounded differences, *in Surveys in Combinatorics, ed. J. Siemons, London Mathematical Society Lecture Notes Series* 141, Cambridge University Press, 1989.

[12] M. Mitzenmacher, Some open questions related to cuckoo hashing. In *Proceedings of the 17th Annual European Symposium on Algorithms (ESA)* (2009) 1-10.

[13] M. Mitzenmacher and S. Vadhan. Why Simple Hash Functions Work: Exploiting the Entropy in a Data Stream. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA) (2008) 746-755.

[14] R. Pagh and F. Rodler. Cuckoo Hashing, *Journal of Algorithms* 51 (2004) 122-144.