ilmedia

th
TECHNISCHE UNIVERSITÄT
ILMENAU

Mitzenmacher, Michael; Panagiotou, Konstantinos; Walzer, Stefan:

Load thresholds for cuckoo hashing with double hashing

TU Ilmenau | Universitätsbibliothek | ilmedia, 2020
http://www.tu-ilmenau.de/ilmedia

Load Thresholds for Cuckoo Hashing
with Double Hashing

Michael Mitzenmacher[1]

Harvard University, School of Engineering and Applied Sciences, Cambridge, USA
michaelm@eecs.harvard.edu
 https://orcid.org/0000-0001-5430-5457

Konstantinos Panagiotou[2]

University of Munich, Institute for Mathematics, Germany
kpanagio@math.lmu.de

Stefan Walzer

Technische Universität Ilmenau, Germany
stefan.walzer@tu-ilmenau.de
 https://orcid.org/0000-0002-6477-0106

—— **Abstract** ——

In *k-ary cuckoo hashing*, each of cn objects is associated with k random buckets in a hash table of size n. An ℓ-*orientation* is an assignment of objects to associated buckets such that each bucket receives at most ℓ objects. Several works have determined *load thresholds* $c^* = c^*(k, \ell)$ for k-ary cuckoo hashing; that is, for $c < c^*$ an ℓ-orientation exists with high probability, and for $c > c^*$ no ℓ-orientation exists with high probability.

A natural variant of k-ary cuckoo hashing utilizes *double hashing*, where, when the buckets are numbered $0, 1, \ldots, n - 1$, the k choices of random buckets form an arithmetic progression modulo n. Double hashing simplifies implementation and requires less randomness, and it has been shown that double hashing has the same behavior as fully random hashing in several other data structures that similarly use multiple hashes for each object. Interestingly, previous work has come close to but has not fully shown that the load threshold for k-ary cuckoo hashing is the same when using double hashing as when using fully random hashing. Specifically, previous work has shown that the thresholds for both settings coincide, except that for double hashing it was possible that $o(n)$ objects would have been left unplaced. Here we close this open question by showing the thresholds are indeed the same, by providing a combinatorial argument that reconciles this stubborn difference.

---

1 Introduction

1.1 The Threshold Question

Cuckoo hashing, introduced by Pagh and Rodler [13] and generalized in many subsequent works (see e.g. [1, 2], and [9] for additional background and references), has proven useful both as a theoretical building block and in practical systems. In *k-ary cuckoo hashing*, each of cn objects is associated with k random buckets in a hash table of size n. An *ℓ-orientation* is an assignment of objects to associated buckets such that each bucket receives at most ℓ objects. Several works have determined *load thresholds* $c^* = c^*(k, \ell)$ for k-ary cuckoo hashing; that is, for $c < c^*$ an ℓ-orientation exists with high probability, and for $c > c^*$ no ℓ-orientation exists with high probability. Beyond their theoretical interest, these load thresholds are important for designing systems that use cuckoo hashing, as they provide an accurate guide to what loads can be achieved in practical settings.

A natural variant of k-ary cuckoo hashing utilizes *double hashing*. Double hashing originally appeared in the context of open-address hash tables, where an object j would be placed by successively trying to find an open bucket at locations $h(i, j) = (h_1(j) + ih_2(j)) \bmod |T|$ for $i = 0, 1, \ldots$, where here $|T|$ represents the table size and h_1 and h_2 are two independently selected hash functions. In the context of cuckoo hashing when the buckets are numbered $0, 1, \ldots, n - 1$, the k choices of random buckets are of the form $h(i, j) = (h_1(j) + ih_2(j)) \bmod n$ for $i = 0, \ldots, k - 1$, so that the choices form an arithmetic progression modulo n.

Double hashing both simplifies implementation and requires less randomness. Moreover, a classical result in the theory of open-address hash tables is that double hashing yields asymptotically the same cost for an unsuccessful search as using full randomness [8], showing that there is negligible performance cost in using double hashing. This type of result, that using double hashing does not change the performance, has since been shown for other hashing-based data structures using several choices, such as Bloom filters [4] and balanced allocation hash tables [10, 11]. We therefore expect that the load thresholds for cuckoo hashing would be the same using double hashing as when using full randomness. Indeed, as we describe in more detail below, previous work has *almost* shown that the thresholds are the same, but completing the argument has proven stubbornly elusive. Here we complete the proof through a suitable combinatorial argument.

1.2 Terminology

In the rest of the paper, we make use of the following terminology.

Fully random graph.   Let $H_{n,cn}^k$ be a k-uniform random hypergraph with vertex set $\mathbb{Z}_n$ and cn edges or something "morally equivalent". Specifically, for our purposes it is often convenient to have perfect independence of edges, each edge e being picked as $e = \{x_1, \ldots, x_k\}$ where $x_1, \ldots, x_k$ are chosen independently and uniformly from $\mathbb{Z}_n$. Note that this may result in some edges of size less than k, as well as duplicate edges. These deviations do not change the threshold, as is known via standard arguments.

Double-Hashing graph.   Similarly, $D_{n,cn}^k$ is also a random hypergraph with vertex set $\mathbb{Z}_n$, but the cn edges must be k-term arithmetic progressions. More precisely, each edge e is independently sampled as $e = \{a + ib \bmod n \mid 0 \le i < k\}$ for $a \in \mathbb{Z}_n$ and $1 \le b < n/2$ chosen independently and uniformly at random. If n is prime then each edge has size k and, for convenience, we assume this to be the case.
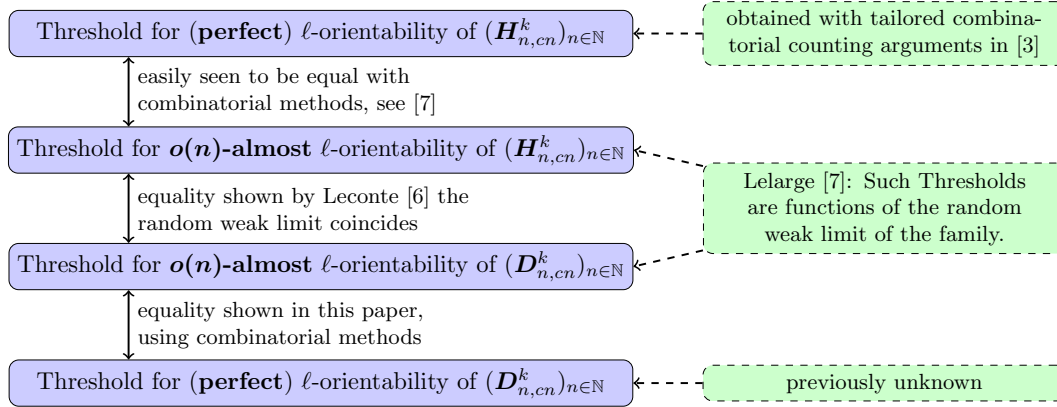
Figure 1 Two out of three steps for Theorem 1 are already known.

Orientability.   A hypergraph $H = (V, E)$ is (perfectly) ℓ-*orientable* if there is function $f : E \to V$ mapping each edge $e \in E$ to an incident vertex $f(e) \in e$ such that each vertex v is the image of at most ℓ edges. For $d \in \mathbb{N}$ we say $H = (V, E)$ is d-*almost* ℓ-*orientable* if there is $E' \subseteq E$ of size $|E'| = |E| - d$ such that $H' = (V, E')$ is ℓ-orientable.

Orientability threshold.   A family $(H_n)_{n \in \mathbb{N}}$ of random hypergraphs depending on a parameter c has an ℓ-*orientability threshold* $c^* \geq 0$ if for $c < c^*$, H_n is ℓ-orientable whp ("with high probability", i.e. with probability $1 - o_{n \to \infty}(1)$) and for $c > c^*$, H_n is *not* ℓ-orientable whp. Similarly we can define d-almost ℓ-orientability thresholds; we may even allow for d to be a function of n.

2   Outline of the Argument

Our goal is to prove the following theorem.

▶ **Theorem 1.** *For any fixed constants $k \geq 3, \ell \geq 1$, the ℓ-orientability threshold for $(H^k_{n,cn})_{n \in \mathbb{N}}$ and the ℓ-orientability threshold for $(D^k_{n,cn})_{n \in \mathbb{N}}$ coincide.*

We review what is known about double hashing in the context of cuckoo hashing, to explain what remains left to show (see Figure 1). Leconte [6] showed that the families $(H^k_{n,cn})_{n \in \mathbb{N}}$ and $(D^k_{n,cn})_{n \in \mathbb{N}}$ have the same Galton-Watson Tree as random weak limit. Lelarge [7] showed that the threshold for $o(n)$-almost ℓ-orientability of a graph family only depends on the random weak limit of the family. It is fairly easy to reconcile the ℓ-orientability and the $o(n)$-almost ℓ-orientability for $(H^k_{n,cn})_{n \in \mathbb{N}}$ (see [7]), showing that the thresholds are the same for that family. In order to establish Theorem 1 all we need to prove is an analogous result for $D^k_{n,cn}$, which is done in the following proposition. Note that ℓ-orientability trivially implies $o(n)$-almost ℓ-orientability, so only the non-trivial direction is given.

▶ **Proposition 1.** *Let $k \geq 3$ and $\ell \geq 1$ be fixed constants. Let c^* be the $o(n)$-almost ℓ-orientability threshold for $(D^k_{n,cn})_{n \in \mathbb{N}}$. Then for any $c < c^*$, $D^k_{n,cn}$ is ℓ-orientable whp.*

The proof uses two lemmas that are proved in Sections 3 and 4. To understand them, we need another concept. In the context of discussing ℓ-orientability of a hypergraph $H = (V, E)$, we call $V' \subseteq V$ a *Hall-witness* if the set $E(V')$ of edges induced by V' has size $|E(V')| > \ell \cdot |V'|$. By Hall's Theorem (restated in Section 4), H is ℓ-orientable if and only if no Hall-witness exists.

The lemmas we utilize are as follows:

▶ **Lemma 2.** *Let $k \geq 3$, $\ell \geq 1$ and $c > 0$ be fixed constants. Then there exists a constant $\delta > 0$ such that, whp, no Hall-witness of size less than δn exists for $D_{n,cn}^k$.*

▶ **Lemma 3.** *If $H = (V, E)$ is d-almost ℓ-orientable and $e \in E$ is contained in some minimal Hall-witness, then $H^{(e)} = (V, E - \{e\})$ is $(d-1)$-almost ℓ-orientable.*

Given these lemmas, we prove Proposition 1, following [7].

Proof of Proposition 1. Let $c = c^* - \varepsilon$ for some $\varepsilon > 0$. We may sample $D_{n,cn}^k$ by first sampling $D_{n,c'n}^k$ for $c' = c^* - \varepsilon/2$ and then removing $\varepsilon n/2$ edges. More precisely, we set $D^{(0)} := D_{n,c'n}^k$ and obtain $D^{(i+1)}$ from $D^{(i)}$ by removing an edge uniformly at random for $0 \leq i < \varepsilon n/2$. Then $D^{(\varepsilon n/2)}$ is distributed as $D_{n,cn}^k$.

For $0 \leq i \leq \varepsilon n/2$, let d_i be the smallest d such that $D^{(i)}$ is d-almost ℓ-orientable. By choice of c^* we have $d_0 = o(n)$ whp. We take δ from Lemma 2 and condition on the high probability event that any Hall-witness of $D^{(0)}$ has size at least δn. Of course, the same bound applies to Hall-witnesses of the subgraphs $D^{(i)}$ with $i > 1$.

Let i be an index with $d_i > 0$. Then $D^{(i)}$ is not ℓ-orientable and a minimal Hall-witness exists. Its size is at least δn, and it induces at least $\delta \ell n + 1$ edges. In particular, the probability that a random edge of $D^{(i)}$ is contained in this minimal Hall-witness is at least $\frac{\delta \ell n + 1}{c'n} \geq \delta \ell / c' = \Theta(1)$. If such an edge is chosen for removal, then by Lemma 3 we have $d_{i+1} = d_i - 1$. Until we reach $D^{(\varepsilon n/2)}$, there are $\varepsilon n/2 = \Theta(n)$ opportunities to reduce the d-value by 1, and each opportunity is realized with probability $\Theta(1)$. Since the initial gap is $o(n)$, the probability that we have $d_{\varepsilon n/2} > 0$ is $\Pr[X < o(n)]$ where $X \sim \mathrm{Bin}(\Theta(n), \Theta(1))$. Simple concentration bounds on binomial random variables prove that this is an $o(1)$-probability event, so we have $d_{\varepsilon n/2} = 0$ whp. Thus $D^{(\varepsilon n/2)} = D_{n,cn}^k$ is (perfectly) ℓ-orientable whp as desired.   ◀

3   No small Hall-witness exists

In this section, we prove Lemma 2. We argue first that it is enough to prove the statement in the case $k = 3$ and $\ell = 1$. Indeed, if $D_{n,cn}^k$ contains no $V' \subseteq V$ inducing more than $|V'|$ edges, then certainly no such V' induces more than $\ell|V'|$ edges. Moreover, let us write $e = \{a_e + ib_e : 0 \leq i < k\}$ for an edge e of $D_{n,cn}^k$. We project each edge e in $D_{n,cn}^k$ to $e' = \{a_e + ib_e : 0 \leq i < 3\}$; then the resulting 3-uniform hypergraph is distributed like $D_{n,cn}^3$, and each $V' \subseteq V$ induces at least as many edges as in $D_{n,cn}^k$. It therefore suffices to show the unlikeliness of certain Hall witnesses in the case of $k = 3$, $\ell = 1$, and fixed $c \in \mathbb{R}^+$.

We introduce the notion of an (s,t)-*set*, a set of size s that contains precisely t arithmetic triples for some $3 \leq s \leq n$ and $1 \leq t \leq \binom{s}{2}$. More specifically, a subset S of $[0, n-1]$ contains some number of arithmetic triples modulo n, which unfortunately does not depend solely on the size of the subset S, and we therefore parametrize the number of triples with an additional variable t. Our plan is to use first moment methods and bound the sum:

$$\sum_{(s,t)} Q_{s,t} p_{s,t}$$

where $Q_{s,t}$ is the number of (s,t)-sets that could be minimal Hall-witnesses, and $p_{s,t}$ is an upper bound on the probability that an (s,t)-set actually is a minimal Hall-witness in $D_{n,cn}^3$.

We separately deal with the following ranges of the parameters s and t.

Case 1: Small s. If $s = o(n^{1/2})$ we exploit that $\sum_t Q_{s,t}$ is sufficiently small by direct counting.

Case 2: Medium s, small-ish t. For $s = \omega(n^{2/5})$ and $t \leq \frac{s^2}{4ce^2}$, the probability $t/\binom{n}{2} \approx \frac{2t}{n^2}$ that random edges are contained in such an (s,t)-set is small enough to find a good bound on $p_{s,t}$.

Case 3: Medium s, large t. For $\delta n \geq s = \omega(n^{2/5})$ (for a small δ chosen later) and $t > \frac{s^2}{4ce^2}$ it turns out that t far exceeds the number of arithmetic triples that would be expected from a random set of size s. A concentration bound by Warnke [14] then gives a useful bound on $Q_{s,t}$.

We deal with these three cases below. We use the following simple bounds on $p_{s,t}$. As we are working in the setting where $\ell = 1$, for a set of size s to be a minimal Hall-witness, there must be at least $s+1$ edges whose elements are in the set. We therefore find:

$$p_{s,t} \leq \left(\frac{t}{\binom{n}{2}} \right)^{s+1} \binom{cn}{s+1} \leq \left(\frac{2t}{n^2} \right)^{s+1} \left(\frac{cne}{s} \right)^{s+1} = \left(\frac{2cet}{sn} \right)^{s+1} \tag{1}$$

$$\leq \left(\frac{ces}{n} \right)^{s+1}. \tag{2}$$

The bound is derived by taking the probability that for a set of $s+1$ edges, each edge turns out to be one of the t arithmetic triples contained in S. This is multiplied with the number of ways to choose $s+1$ out of the cn edges of $D^3_{n,cn}$. For the second line we used the trivial bound $t \leq \binom{s}{2} \leq \frac{s^2}{2}$.

Case 1: $s = o(\sqrt{n})$. Assume $S \subseteq \mathbb{Z}_n$ is a minimal Hall-witness for $D^3_{n,cn}$ inducing a set P of edges (with $|P| > 3|S|$). As a hypergraph, (S,P) is spanning, i.e. each vertex is contained in an edge, otherwise the isolated vertex can be removed for a smaller Hall-witness. Also, (S,P) is connected, i.e. for any $x,y \in S$ there is a sequence $e_1, \ldots, e_j \in P$ with $x \in e_1, y \in e_j$ and $e_i \cap e_{i+1} \neq \emptyset$ for $1 \leq i < j$). Otherwise, at least one connected component forms a smaller Hall-witness.

So for fixed s, we can count all (s,t)-sets (with arbitrary t) that might be minimal Hall-witnesses by counting vertex sets that can support connected spanning hypergraphs. We do this by counting *annotated depth-first-search-runs* (dfs-runs), associated with such (s,t)-sets, in the following way. A dfs-run through S starts at a root vertex $r \in S$ and puts it on the *stack*, whose topmost element is referred to as top. Then a sequence of steps follow, each of which either removes top from the stack (backtrack) or finds new vertices in S that are then put on the stack. More precisely, new vertices are found by specifying an arithmetic triple that is contained in S and involves top. The two vertices other than top may either both be new (find$_2$) or only one vertex is new, and a third vertex v was already found in a previous step (find$_1$). The following data about the dfs-run is needed to reconstruct S from it:

- The root vertex r. There are n possibilities.
- The type of each step, which can be backtrack, find$_1$ or find$_2$. Since there are at most $2s$ steps, there are at most 3^{2s} possibilities in total.
- For each step of type find$_1$, the vertex v that was previously found and that together with top and the new vertex forms an arithmetic triple. There are less than s possibilities. In addition we need the position of top and v in the arithmetic triple (essentially four possibilities). The newly discovered vertex can then be computed from top and v.

▬ For each step of type find_2, the difference between adjacent elements of the arithmetic triple – there are $n/2$ possibilities. Also, the position of top in that triple – there are 3 possibilities.

If f_1 and f_2 count the number of times the steps find_1 and find_2 are used in the dfs-run through S of size s, then we have $f_1 + 2f_2 = s - 1$. For $s = o(\sqrt{n})$, the find_2-steps yield a significantly higher number of possibilities per found vertex compared to find_1-steps, so we compute:

$$\sum_t Q_{s,t} \leq n \cdot 3^{2s} \cdot (4s)^{f_1} (3n/2)^{f_2} \leq n \cdot 3^{2s} \cdot (3n/2)^{(s-1)/2} \leq c_1^s n^{(s+1)/2}$$

where c_1 is a constant. Using Equation (2) we get:

$$\sum_{s=3}^{o(\sqrt{n})} \sum_t Q_{s,t} p_{s,t} \leq \sum_{s=3}^{o(\sqrt{n})} (\sum_t Q_{s,t})(\max_t p_{s,t})$$

$$\leq \sum_{s=3}^{o(\sqrt{n})} c_1^s n^{(s+1)/2} \left(\frac{ces}{n}\right)^{s+1} \leq \sum_{s=3}^{o(\sqrt{n})} \left(\frac{c_2 s}{\sqrt{n}}\right)^{s+1}$$

for a new constant c_2. Since each term in the sum is $O(n^{-2})$ and since there are $o(n^{1/2})$ terms, the sum is clearly $o(n^{-3/2}) = o(1)$, closing this case.

Case 2: $s = \omega(n^{2/5})$ and $t \leq \frac{s^2}{4ce^2}$. Combining the trivial bound of $Q_{s,t} \leq \binom{n}{s}$, Equation (1), and our assumption on t we obtain:

$$Q_{s,t} \cdot p_{s,t} \leq \binom{n}{s} \cdot \left(\frac{2cet}{sn}\right)^{s+1} \leq \left(\frac{ne}{s}\right)^s \cdot \left(\frac{s}{2ne}\right)^{s+1} \leq \left(\frac{1}{2}\right)^s.$$

This is clearly $o(1)$, even after summing over all $O(n)$ admissible choices for s and all $O(n^2)$ choices for t.

Case 3: $\omega(n^{2/5}) \leq s \leq \delta n$ and $t > \frac{s^2}{4ce^2}$. A random set $S \subseteq \mathbb{Z}_n$ of size s in this range behaves very much like a random set T that is obtained by picking each element of $\mathbb{Z}_n$ independently with probability $p = \frac{s}{n}$. Let X be the number of arithmetic triples in T. We have $\mu := \mathbb{E}[X] = \binom{n}{2}p^3 \leq \frac{s^3}{2n}$. In particular, the case $X > \frac{s^2}{4ce^2}$ is very rare if $s < \delta n$ for sufficiently small δ. We can therefore expect the number $Q_{s,t}$ to be significantly less than $\binom{n}{s}$. Formally we write:

$$Q_{s,t} \leq \binom{n}{s} \Pr[S \text{ contains } t \text{ a.p.}] = \binom{n}{s} \Pr[X = t \mid |T| = s] \leq \binom{n}{s} \Pr[X = t] O(\sqrt{n})$$

where $O(\sqrt{n})$ is the inverse of the probability of the event $|T| = s$. Using Theorem 1 from [14] with $k = 3$, $p = s/n$, we get positive constants $b, B > 0$ such that for sufficiently large n

$$\Pr[X = t] \leq \Pr[X \geq (1 + \tfrac{t-\mu}{\mu})\mu] \overset{[14]}{\leq} e^{-b\sqrt{\frac{t-\mu}{\mu}}\sqrt{\mu}\log(1/p)}$$

Using our bound on t and assuming $\delta \leq \frac{1}{4ce^2}$ we can bound the negated exponent by:

$$b\sqrt{t-\mu}\log(\tfrac{1}{p}) \geq b\sqrt{\tfrac{s^2}{4ce^2} - \tfrac{s^3}{2n}}\log(\tfrac{1}{\delta}) \geq bs\sqrt{\tfrac{1}{4ce^2} - \tfrac{\delta}{2}}\log(\tfrac{1}{\delta}) \geq bs\sqrt{\tfrac{1}{8ce^2}}\log(\tfrac{1}{\delta}) = sc_3\log(\tfrac{1}{\delta})$$

for some constant $c_3 > 0$ which yields $\Pr[X = t] \leq (\delta^{c_3})^s$. Combining this with Equation (2), this time assuming $\delta^{c_3} \leq \frac{1}{2ce^2}$, we can write

$$Q_{s,t} \cdot p_{s,t} \leq \binom{n}{s} \Pr[X = t] O(\sqrt{n}) \cdot p_{s,t} \leq \left(\frac{ne}{s}\right)^s (\delta^{c_3})^s O(\sqrt{n}) \left(\frac{ces}{n}\right)^{s+1}$$

$$\leq O(\sqrt{n}) \left(ce^2 \delta^{c_3}\right)^s \leq O(\sqrt{n}) \cdot 2^{-s}$$

which is $o(1)$, even when summing over all $\Theta(n)$ admissible values s and all $\Theta(n^2)$ admissible values for t.

4   The significance of Hall-witnesses

To understand how Hall's Theorem relates to our situation, we view a hypergraph $H = (V, E)$ as a bipartite graph with E on the "left", V on the "right" and a connection between $e \in E$ and $v \in V$ iff $v \in e$. We care about generalized $(1, \ell)$-matchings in this *incidence graph* of H, i.e. sets $M \subseteq E \times V$ such that any $e \in E$ has degree at most 1 in M and any $v \in V$ has degree at most ℓ in M. An ℓ-orientation f of H, viewed as a set of pairs $f \subseteq E \times V$, is then precisely an *edge-perfect* $(1, \ell)$-matching (each $e \in E$ has degree precisely 1). We call the corresponding notion of a vertex-perfect $(1, \ell)$-matching (each $v \in V$ has degree ℓ in M) an *ℓ-saturation*.

In this setting, Hall's Theorem is easily generalized to the following, where we use $N(X)$ to denote the direct neighbors of X in the incidence graph (note that $X \subseteq V$ and $X \subseteq E$ are both allowed) and $E(V')$ to denote the set of edges contained in $V' \subseteq V$.

▶ **Theorem 4** (Hall's Theorem).
(i) H *has an ℓ-orientation* $\Leftrightarrow \nexists E' \subseteq E$ *with* $\ell |N(E')| < |E'|$
     $\Leftrightarrow \nexists V' \subseteq V$ *with* $\ell |V'| < |E(V')| \Leftrightarrow$: *No Hall-witness exists.*
(ii) H *has an ℓ-saturation* $\Leftrightarrow \nexists V' \subseteq V$ *with* $|N(V')| < \ell |V'|$.
We are now ready to prove Lemma 3.

Proof of Lemma 3. Let $H = (V, E)$ be a non-ℓ-orientable hypergraph and $S \subseteq V$ be a minimal Hall-witness to this fact. Consider $H_S = (S, E(S))$, the sub-hypergraph of H induced by S. Within H_S we have $|N_{H_S}(S')| > \ell |S'|$ for any $\emptyset \neq S' \subseteq S$, as otherwise, i.e. assuming $|N_{H_S}(S')| \leq \ell |S'|$, we have

$$|E(S - S')| = |E(S) - N_{H_S}(S')| = |E(S)| - |N_{H_S}(S')| > \ell |S| - \ell |S'| = \ell |S - S'|$$

which would make $S - S'$ a smaller Hall-witness than S, contradicting minimality.

This means for $H_S^{(e)} := (S, E(S) - \{e\})$ we have (replacing ">" with "≥") $|N_{H_S^{(e)}}(S')| \geq \ell |S'|$ for any $S' \subseteq S$ (the claim is trivial for $S' = \emptyset$). By Theorem 4(ii), $H_S^{(e)}$ has an ℓ-saturation $M_S^{(e)}$.

Now if H is d-almost ℓ-orientable and $M \subseteq E \times V$ is a corresponding $(1, \ell)$-matching of size $|E| - d$, our task is to obtain a $(1, \ell)$-matching M' with $|M| = |M'|$ in $H^{(e)} = (V, E - \{e\})$ where an edge $e \in E(S)$ was removed. This will imply that $H^{(e)}$ is $(d - 1)$-almost ℓ-orientable as desired.

Constructing M' is easy, as we just remove all edges from $E(S)$ from M (this certainly gets rid of e if it was used) and re-saturate the vertices from S by adding an appropriate subset $Y \subseteq M_S^{(e)}$. Then $M' := (M \setminus E(S)) \cup Y$ has the same size as M. ◀

5   Conclusion

We have shown that for ℓ-orientations in k-ary cuckoo hashing (for constant k and ℓ), double hashing yields the same load thresholds as fully random hashing. This provides yet another example of a hashing structure with the same behavior when using only double hashing in place of random hashing. It seems somewhat unfortunate and perhaps a little mysterious that there does not yet appear to be a unifying argument for multiple such hashing structures; each structure, thus far, has required its own specialized argument. We optimistically suggest that a more unified approach may exist, that would shed more light on this phenomenon.

A problem closely related to the cuckoo hashing problem we have studied here is the question of the ℓ-core threshold of a k-uniform random hypergraph. The ℓ-core of a hypergraph is obtained by repeatedly removing any vertex of degree less than ℓ, and all adjacent edges. One can think of the ℓ-core as what is left after a "greedy" first stage in an offline algorithm for finding an $(\ell-1)$-orientation; each bucket with at most $\ell-1$ objects simply accepts those objects, and the remaining objects would then have to be more carefully placed to obtain an an $(\ell-1)$-orientation, if possible. For random hypergraphs on n vertices with cn edges, there are similar thresholds $c^* = c^*(k, \ell)$ for the existence of a non-empty ℓ-core; that is for $c < c^*$ the ℓ-core is empty with high probability, and for $c > c^*$ the ℓ-core consists of $\Omega(n)$ edges with high probability. Empirically, the double-hashing graph appears to have the same thresholds as random hypergraphs for the ℓ-core, and it is known the thresholds are the same when $\ell > k$ [12]. It might seem our approach would be useful for settling this question as well, but thus far we cannot currently rule out small $o(n)$-sized ℓ-cores under double hashing using these ideas. This question remains tantalizingly open.

References

1   Martin Dietzfelbinger and Christoph Weidling. Balanced allocation and dictionaries with tightly packed constant size bins. *Theoretical Computer Science*, 380(1-2):47–68, 2007.

2   Dimitris Fotakis, Rasmus Pagh, Peter Sanders, and Paul Spirakis. Space efficient hash tables with worst case constant access time. *Theory of Computing Systems*, 38(2):229–248, 2005.

3   Nikolaos Fountoulakis, Megha Khosla, and Konstantinos Panagiotou. The Multiple-Orientability Thresholds for Random Hypergraphs. In *Proc. 22nd SODA*, pages 1222–1236, 2011. URL: `http://www.siam.org/proceedings/soda/2011/SODA11_092_fountoulakisn.pdf`.

4   Adam Kirsch and Michael Mitzenmacher. Less hashing, same performance: Building a better bloom filter. *Random Structures & Algorithms*, 33(2):187–218, 2008.

5   M. Leconte, M. Lelarge, and L. Massoulié. Convergence of multivariate belief propagation, with applications to cuckoo hashing and load balancing. In *Proc. 24th SODA*, pages 35–46, 2013. URL: `http://dl.acm.org/citation.cfm?id=2627817.2627820`.

6   Mathieu Leconte. Double hashing thresholds via local weak convergence. In *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*, pages 131–137. IEEE, 2013. `doi:10.1109/Allerton.2013.6736515`.

7   Marc Lelarge. A New Approach to the Orientation of Random Hypergraphs. In *Proc. 23rd SODA*, pages 251–264, 2012. URL: `http://dl.acm.org/citation.cfm?id=2095139`.

8   George S Lueker and Mariko Molodowitch. More analysis of double hashing. *Combinatorica*, 13(1):83–96, 1993.

9   Michael Mitzenmacher. Some open questions related to cuckoo hashing. In *ESA*, pages 1–10. Springer, 2009.

10   Michael Mitzenmacher. Balanced allocations and double hashing. In *Proceedings of the 26th ACM symposium on Parallelism in algorithms and architectures*, pages 331–342. ACM, 2014.

11   Michael Mitzenmacher. More analysis of double hashing for balanced allocations. *CoRR*, abs/1503.00658, 2015. `arXiv:1503.00658`.

12   Michael Mitzenmacher and Justin Thaler. Peeling arguments and double hashing. In *50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012, Allerton Park & Retreat Center, Monticello, IL, USA, October 1-5, 2012*, pages 1118–1125. IEEE, 2012. `doi:10.1109/Allerton.2012.6483344`.

13   Rasmus Pagh and Flemming Friche Rodler. Cuckoo hashing. *J. Algorithms*, 51(2):122–144, 2004. `doi:10.1016/j.jalgor.2003.12.002`.

14   Lutz Warnke. Upper tails for arithmetic progressions in random subsets. *Israel Journal of Mathematics*, pages 1–49, 7 2017. `doi:10.1007/s11856-017-1546-3`.