PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Pages 000–000
S 0002-9939(XX)0000-0

ROTA'S BASIS CONJECTURE HOLDS ASYMPTOTICALLY

and ALEXEY POKROVSKIY

(Communicated by)

ABSTRACT. Rota's Basis Conjecture is a well known problem from matroid theory, that states that for any collection of n bases in a rank n matroid, it is possible to decompose all the elements into n disjoint rainbow bases. Here an asymptotic version of this is proved. We show that it is possible to find $n - o(n)$ disjoint rainbow independent sets of size $n - o(n)$.

1. INTRODUCTION

In 1989, Rota made the following conjecture "in any family $B_1, \ldots, B_n$ of n bases in a vector space V, it is possible to find n disjoint *rainbow bases*" (see [11], Conjecture 4). Here a *rainbow basis* means a basis of V consisting of precisely one vector from each of $B_1, \ldots, B_n$. In the context of this conjecture "disjoint" means that we do not have two rainbow bases using the same vector from the same basis B_i. Rota's conjecture has attracted attention due to its simplicity and connections to apparently unrelated areas. For example Huang and Rota [11] found connections between it and problems about Latin squares and supersymetric bracket algebra. Amongst other things, the recent collaborative Polymath project [3] studied an approach to Rota's conjecture using topological tools.

It was observed (by Rota as well), that this might hold in the much more general setting of matroids rather than vector spaces. Matroids are an abstraction of independent sets in vector spaces, which also generalize many other "independence structures". They are defined on a set V called the *ground set* of the matroid. A matroid M is a nonempty family of subsets of V (called independent sets) which is closed under taking subsets and satisfies the following additional property (called the "augmentation property"): that if $I, I' \in M$ are two independent sets with $|I| > |I'|$, then there is some element $x \in I \setminus I'$ such that $I' \cup \{x\}$ is also an independent set in M. A *basis* of M is a maximal independent sets. By the augmentation property all bases of M must have the same size, which is called the *rank* of M. Using this terminology, the general Rota's Basis Conjecture (see [11]) can be phrased as:

Conjecture 1.1 (Rota's Basis Conjecture)**.** Let $B_1, \ldots, B_n$ be disjoint bases in a rank n matroid M. Then it is possible to decompose $B_1 \cup \cdots \cup B_n$ into n disjoint rainbow bases.

Rota's conjecture attracted a lot of attention due to its simple formulation, and due to a large range of possible approaches towards it (coming from the many different settings in which matroids can naturally arise). One research direction

2020 *Mathematics Subject Classification.* Primary 05B35.

is to prove the conjecture for some particular naturally-arising class of matroids. For example for matroids arising from real vector spaces (called real-representable matroids), the conjecture is known to hold whenever $n-1$ or $n+1$ is prime. This was proved in a combination of papers. First, Huang and Rota [11] reduced the conjecture for real-representable matroids to the Alon-Tarsi Conjecture (which is a conjecture unrelated to matroids and states that for all n the number of even and odd Latin squares of order n is different). The Alon-Tarsi Conjecture was proved for $n-1$ prime by Drisko [5] and for $n+1$ prime by Glynn [10]. Rota's Conjecture is known to hold for some other classes of matroids too. It was proved for paving matroids by Geelen and Humphries [7], for strongly base orderable matroids by Wild [14], and for rank ≤ 4 matroids computationally by Cheung [2].

Another research direction is to try and establish a weaker conclusion which holds for *all matroids*. When $B_1, \dots, B_n$ are bases in a matroid, define a *rainbow independent set* as a (possibly empty) independent set containing at most one element from each B_i. There are a number of natural approaches here:

(1) **Find many disjoint rainbow bases in $B_1 \cup \dots \cup B_n$:** Finding one rainbow basis is easy using the augmentation property. Finding more is already challenging. Geelen and Webb [8] found $\Omega(\sqrt{n})$ disjoint rainbow bases. This was improved to $\Omega(n/\log n)$ by Dong and Geelen [4], and further to $n/2 - o(n)$ by Bucic, Kwan, Sudakov, and the author [1].

(2) **Decompose $B_1 \cup \dots \cup B_n$ into few rainbow independent sets:** The conjecture asks for a decomposition into n independent sets. Aharoni and Berger showed that you can decompose into $2n$ independent sets. This was investigated further during the Polymath 12 project [13] where it was improved to $2n-2$.

(3) **Find n disjoint rainbow independent sets of large total volume:** This means rainbow independent sets $I_1, \dots, I_n$ with $\sum_{i=1}^n |I_n|$ as large as possible. Rota's conjecture says that we can get $\sum_{i=1}^n |I_n| = n^2$. Both of the previous approaches give something here — having s rainbow bases clearly gives a family of independent sets of volume sn, whereas in a decomposition of $B_1 \cup \dots \cup B_n$ into t rainbow independent sets, the n largest of these must have total volume at least $(n/t)n^2$. Thus the best known results about (1) and (2) both give a family of independent sets of volume around $n^2/2$.

In each of the above three approaches it is desirable to obtain an asymptotic version of the conjecture. In other words: Can you find $(1-o(1))n$ disjoint rainbow bases? Can you decompose $B_1 \cup \dots \cup B_n$ into $(1+o(1))n$ disjoint rainbow independent sets? Can you find n disjoint rainbow independent sets of total volume $(1-o(1))n^2$? Previously such results were proved only for special classes of matroids — Friedman and McGuinness [6] proved an asymptotic version for large girth matroids. Combining the results of [11, 5, 10] with the Prime Number Theorem gives an asymptotic version for real-representable matroids [12]. In this paper, we prove the first asymptotic version of the conjecture which holds *for all matroids*.

Theorem 1.2. *Let $B_1, \dots, B_n$ be disjoint bases in a rank n matroid M. Then there are $n - o(n)$ disjoint rainbow independent sets in $B_1 \cup \dots \cup B_n$ of size $n - o(n)$.*

Notice that the union of these independent sets has size $(1-o(1))n^2$, and so this theorem gives an asymptotic version of the conjecture, when one takes approach (3) above. Going forward, it would be interesting to obtain stronger asymptotic versions of the conjecture as well as a proof for large rank matroids. Theorem 1.2 is likely to be a good starting point in proving such results — in recent years "absorption techniques" have been used in related problems to turn asymptotic solutions like Theorem 1.2 into exact ones.

2. Proof outline

Here, we explain the ideas of our proof by presenting a simplified version of it with some complications missing. Aside from some definitions, everything here is not used in the actual proof.

In this paper, we use the term *"coloured matroids"* to mean a matroid with a colour assigned to each element in the ground set such that the colour classes are independent. We will work with families $\mathcal{T} = (T_1, \ldots, T_m)$ of disjoint rainbow independent sets in a coloured matroid M. These families will always be ordered, with repetition allowed. We use $(\mathcal{T})_i$ to denote the ith independent set in $\mathcal{T}$, i.e., if $\mathcal{T} = (T_1, \ldots, T_m)$, then $(\mathcal{T})_i = T_i$. We use $E(\mathcal{T})$ for the subset $T_1 \cup \cdots \cup T_m$ of the ground set of M. For two families $\mathcal{T} = (T_1, \ldots, T_m)$, $\mathcal{S} = (S_1, \ldots, S_m)$, we write $\mathcal{T} \leq \mathcal{S}$, if $T_i \subseteq S_i$ for all i. For a colour, we use $E_{\mathcal{T}}(c)$ to denote the set of colour c elements of $\mathcal{T}$ and fix $e_{\mathcal{T}}(c) = |E_{\mathcal{T}}(c)|$. For a coloured matroid M, use $C(M)$ for the set of colours occurring on M. For a colour $c \in C(M)$, we use $E(c)$ to denote the set of colour c elements of M.A common consequence of the augmentation property that we use is that given any independents sets I, J with $|I| \geq |J|$, there is an independent set $B \subseteq I \cup J$ with $|B| = |I|$ and $J \subseteq B$ (which is proved by repeatedly adding elements from I to J using the usual augmentation property until the two sets have the same size). We'll also use the following consequence of the augmentation property.

Observation 2.1 (Rainbow exchange property)**.** Let S be a rainbow independent set in a coloured matroid M. Then for every element $e \notin S$, there is a set $Q \subseteq S$ of at most two elements with $S \setminus Q \cup \{e\}$ rainbow and independent.

Additionally, for any $T \subseteq S$ with $T + e$ rainbow and independent, we can choose $Q \subseteq S \setminus T$

Proof. Let S' be a maximum size independent subset of $S + e$ which contains $T + e$ (it exists because $T + e$ is independent). We must have $|S'| \geq |S|$, since otherwise the augmentation property would give some $x \in S$ with $S' + x$ independent. In particular, using that $e \in S' \subseteq S + e$ and $e \notin S$, we have $|S \setminus S'| = |(S \cup \{e\}) \setminus (S' \cup \{e\})| = |(S \cup \{e\}) \setminus S'| = |S| + 1 - |S'| \leq 1$. Note that since S' contains T we have that $S \setminus S'$ is disjoint from T. Let f be the colour $c(e)$ element of S (if it exists), noting that $f \notin T$ since $T + e$ is rainbow.

Set $Q = (S \setminus S') \cup \{f\}$ to get a set satisfying the observation. We have $S \setminus S'$, $\{f\}$ disjoint from T, which shows $Q \subseteq S \setminus T$. Also, $S \setminus Q \cup \{e\}$ is rainbow since $S \setminus Q$ was rainbow, missing colour $c(f) = c(e)$. Also $S \setminus Q \cup \{e\}$ is independent since it is contained in $S \setminus (S \setminus S') \cup \{e\} = S'$ which is independent. □

We call a family $\mathcal{T}$ *maximum* if $|E(\mathcal{T})|$ is maximum amongst families of disjoint rainbow independent sets in M. Rota's Conjecture is equivalent to saying that a maximum family has n^2 elements. Theorem 1.2 is equivalent to proving that a maximum family has $\geq (1 - \epsilon)n^2$ elements. We achieve this by studying how elements can be moved between the rainbow independent sets of the family. The key definition is that of a "reduced family of $\mathcal{T}$" which informally means deleting all elements from $\mathcal{T}$ which can be moved around robustly.

Definition 2.2 (Reduced family)**.** Let $\mathcal{T}$ be a family of disjoint rainbow independent sets in a coloured matroid. Define the *ℓ-reduced family* $\mathcal{T}'_\ell$ of $\mathcal{T}$ to be $\mathcal{T}$ minus all elements $e \in E(\mathcal{T})$ for which there are at least ℓ different choices of $T_j \in \mathcal{T}$ with $T_j + e$ a rainbow independent set. Define $\mathcal{T}^{(r)}_\ell = \mathcal{T}''''^{\cdots}_\ell$, where we repeat the operation r times. We fix $\mathcal{T}^{(0)}_\ell = \mathcal{T}$.

We will need the following monotonicity properties of reduction.

Observation 2.3. $\mathcal{T}'_\ell \leq \mathcal{T}_\ell$

Proof. If $\mathcal{T}_\ell = (T_1, \dots, T_m)$ and $\mathcal{T}'_\ell = (T'_1, \dots, T'_m)$, then each T'_i is formed by deleting elements from T_i and so $T'_i \subseteq T_i$, verifying the definition of $\mathcal{T}'_\ell \leq \mathcal{T}_\ell$. $\square$

Observation 2.4. If $\mathcal{T} \leq \hat{\mathcal{T}}$, then $\mathcal{T}'_\ell \leq \hat{\mathcal{T}}'_\ell$.

Proof. Consider some $\hat{T} \in \hat{\mathcal{T}}$, and let T, T', $\hat{T}'$ be the corresponding independent sets in $\mathcal{T}$, $\mathcal{T}'_\ell$, $\hat{\mathcal{T}}'_\ell$ respectively. We know that $T \subseteq \hat{T}$, and need to show that $T' \subseteq \hat{T}'$. Suppose for contradiction that there is some $e \in T' \setminus \hat{T}'$. Since Observation 2.3 gives $T' \subseteq T$, we know $e \in T \subseteq \hat{T}$. Since $e \notin \hat{T}'$, the definition of $\hat{\mathcal{T}}'_\ell$ tells us that there are at least ℓ choices of $\hat{T}_j$ in $\hat{\mathcal{T}}$ with $\hat{T}_j + e$ a rainbow independent set. Since $\mathcal{T} \leq \hat{\mathcal{T}}$, for each j there is some corresponding $T_j \in \mathcal{T}$ with $T_j \subseteq \hat{T}_j$. This implies that $T_j + e \subseteq \hat{T}_j + e$ is rainbow and independent for each j, i.e., we have at least ℓ choices of T_j with $T_j + e$ rainbow and independent. By the definition of "ℓ-reduced", this shows that $e \notin E(\mathcal{T}'_\ell)$, which is a contradiction. $\square$

Observation 2.5. Let $\ell \geq 2$ and let $\mathcal{T}$ be a family of disjoint rainbow independent sets. Suppose there is a colour c with $c \in C(T)$ for all $T \in \mathcal{T}$. Then $c \in c(T')$ for all $T' \in \mathcal{T}'_\ell$ also.

Proof. Consider some colour c element $e \in T \in \mathcal{T}$. Then for all $S \in \mathcal{T}$ with $S \neq T$, we have that $S + e$ is not a rainbow independent set (by assumption S contains some colour c element f. Since $S \neq T$ and S, T are disjoint rainbow independent sets we have $e \neq f$. Therefore, $S + e$ is not rainbow, since it contains both e and f). Thus there is only $1 < \ell$ choice of $S \in \mathcal{T}$ with $S + e$ rainbow and independent. $\square$

A maximum family of rainbow independent sets $\mathcal{T}$ has the property that no element e outside $\mathcal{T}$ can be added to any rainbow independent set $T \in \mathcal{T}$ without breaking either rainbowness or independence. Our proof rests on this property being preserved by reduction.

Lemma 2.6. *Fix $r = 0$ or 1. Let M be a coloured matroid and $\mathcal{T}$ a family of disjoint rainbow independent sets in M. Suppose we have $T \in \mathcal{T}^{(r)}_3$ and $e \notin E(\mathcal{T})$ with $T + e$ rainbow and independent. Then $\mathcal{T}$ is not maximum.*

Proof. Let $R \in \mathcal{T}$ with $T \subseteq R$. Since $T + e$ is rainbow and independent, by Observation 2.1, there is a set $Q \subseteq R \setminus T \subseteq E(\mathcal{T}) \setminus E(\mathcal{T}^{(r)}_3)$ of size ≤ 2 with $R \setminus Q \cup \{e\}$ rainbow independent. The proof proceeds differently based on the size of Q:

- If $Q = \emptyset$, then replacing R by $R + e$ gives a family larger than $\mathcal{T}$. In particular, this deals with the case $r = 0$ (since in that case $\mathcal{T}_3^{(r)} = \mathcal{T}_3^{(0)} = \mathcal{T}$ and $R = T$).
- If $Q = \{f_1\}$, then since $f_1 \notin E(\mathcal{T}_3')$, then, by the definition of "reduced", there is some $S_1 \in \mathcal{T}$ with $S_1 + f_1$ rainbow and independent. In fact we get 3 choices for S_1, so can choose it different from R. Now replace R by $R \setminus \{f_1\} \cup \{e\}$ and replace S_1 by $S_1 \cup \{f_1\}$ to get a family larger than $\mathcal{T}$.
- If $Q = \{f_1, f_2\}$, then since $f_1, f_2 \notin E(\mathcal{T}_3')$, then, by the definition of "reduced", there are some $S_1, S_2 \in \mathcal{T}$ with $S_1 + f_1, S_2 + f_2$ rainbow and independent. Since we have 3 choices for S_1, S_2, we can choose them different from each other and from R. Now replace R by $R \setminus \{f_1, f_2\} \cup \{e\}$, replace S_1 by $S_1 \cup \{f_1\}$, replace S_2 by $S_2 \cup \{f_2\}$ to get a family larger than $\mathcal{T}$.

□

The above lemma generalizes to arbitrary r in Lemma 3.10 (by increasing "3" to something bigger). Our proof consists of showing that for any family $\mathcal{T}$ with $e(\mathcal{T}) \leq (1-\epsilon)n^2$, its reduction $\mathcal{T}_\ell^{(r)}$ becomes small for some r. We do this step by step by showing the inequality "$e(\mathcal{T}_\ell') \leq e(\mathcal{T}) - \epsilon^2 n^2$" for such families T. By iterating this inequality we get $e(\mathcal{T}_\ell^{(r)}) \leq e(\mathcal{T}) - r\epsilon^2 n^2$. The proof of this inequality rests on the following lemma which estimates how many edges every colour loses when reducing the family.

Lemma 2.7. *Fix $r = 0$ or 1. Let M be a coloured matroid with n colours of size n and let $\mathcal{T}$ be a maximum family of m disjoint rainbow independent sets in M. Let $T \in \mathcal{T}_3^{(r)}$ and let c be a colour missing from T. Then*

$$e_{\mathcal{T}_1^{(r+1)}}(c) \leq |T| - n + m$$

Proof. By the augmentation property there are $n - |T|$ colour c elements e with $T + e$ a rainbow independent set. Lemma 2.6 tells us that these all occur on $\mathcal{T}$. By definition of $\mathcal{T}_1'$, they are all absent from $\mathcal{T}_1^{(r+1)}$. Additionally there are at least $n - m$ colour c elements absent from $\mathcal{T}$, which remain absent in $\mathcal{T}_1^{(r+1)}$. □

In the actual proof we use a version of this $\mathcal{T}_\ell^{(r)}$ with larger r, ℓ (Lemma 3.8). The above lemma gives its best bound when T is as small as possible. This motivates us to define the *excess of a colour c in $\mathcal{T}$*:

$$\textsc{ex}(c, \mathcal{T}) := \max(0, e_{\mathcal{T}}(c) + n - m - \min_{T \in \mathcal{T}: c \notin T} |T|)$$

Lemma 2.7 now can be rephrased as saying that $e_{\mathcal{T}_1^{(r+1)}}(c) \leq e_{\mathcal{T}_3^{(r)}}(c) - \textsc{ex}(c, \mathcal{T}_3^{(r)})$ for every colour missing from some $T \in \mathcal{T}$. To get an inequality like "$e(\mathcal{T}^{(r+1)}) \leq e(\mathcal{T}^{(r)}) - \epsilon^2 n^2$" from this, we need to show that the average excess over all colours is $\epsilon^2 n$. We show the following:

Lemma 2.8. *Let M be a coloured matroid with n colours of size n and let $\mathcal{T}$ be a family of $(1-\epsilon)n$ disjoint rainbow independent sets in M. Then*

$$\frac{1}{n} \sum_{\text{colours } c} \textsc{ex}(c, \mathcal{T}) \geq \epsilon^2 n$$

The proof of this has nothing to do with matroids or colours. The essence of it turns out to be a very short lemma about bipartite graphs (see Lemma 3.4). We now have all the ingredients that go into the proof of Theorem 1.2. To summarize, the structure is:

- Start with M, a coloured matroid with n colours of size n.
- Consider a maximum family $\mathcal{T}$ of $(1-\epsilon)n$ disjoint rainbow independent sets in M.
- Suppose for contradiction that $e(\mathcal{T}) \leq (1-\epsilon)n^2$.
- By a variant of Lemma 2.8, we have $\sum_{\text{colours } c} \text{EX}(c, \mathcal{T}_\ell^{(r)}) \geq \epsilon^2 n^2$ for all r and large enough ℓ.
- By a variant of Lemma 2.7, and maximality, we have $e(\mathcal{T}_\ell^{(r)}) < (1 - r\epsilon^2)n^2$ for all r and large enough ℓ. At $r = 1/\epsilon^2$ this is a contradiction (meaning that the assumption "$\mathcal{T}$ is maximum" in one of the applications of Lemma 2.7 along the way was invalid).

In this sketch, there are a couple of things missing. Most of them are easy to fill in — namely Lemmas 2.6 and 2.7 can be proved for larger r and ℓ.

However there is one complication which appears to require significant changes to the above strategy — namely the requirement that "c is missing from some $T \in \mathcal{T}$" for the inequality "$e_{\mathcal{T}_1^{(r+1)}}(c) \leq e_{\mathcal{T}_3^{(r)}}(c) - \text{EX}(c, \mathcal{T}_3^{(r)})$'. When there are many colours that occur on all $T \in \mathcal{T}$, then it is possible that $\mathcal{T}_1' = \mathcal{T}$, which breaks the above strategy (as an example, consider a family $\mathcal{T}$ consisting of $(1-\epsilon)n$ rainbow independent sets of size $n/2$ all using the same $n/2$ colours and nothing else). The way we get around this issue is to change what "maximum family" means. Rather than asking them to have as many elements as possible, we instead ask them to be "lexicographically maximum" which means roughly that $\min_{\text{colours } c} e_{\mathcal{T}}(c)$ is as large as possible. The overall structure of the proof remains unchanged — it follows analogues of the above lemmas with suitable changes.

3. Proof of Theorem 1.2

Rather than working with the excess of a family as in the proof outline, we will associate an auxiliary bipartite graph to every family and study a parameter $k\text{-EX}_G(y)$ associated to the graph. All graphs we deal with will be simple. For a vertex v in a graph G, recall that the neighbourhood $N_G(v)$ is defined as the set of vertices connected to v, and the degree $d_G(v)$ is defined as $|N_G(v)|$. For a set of vertices A, we define $\delta_G(A) := \min_{v\in A} d_G(v)$. For two sets of vertices A, B, we let $e_G(A, B)$ be the number of edges ab with $a \in A, b \in B$. If G is bipartite with parts X, Y, and $A \subseteq X$, then we will often use that $e_G(A, Y) = \sum_{a\in A} d_G(A) \geq \delta_G(A)|A|$. Where there is no ambiguity, we abbreviate $N_G(v) = N(v), d_G(v) = d(v), \delta(G) = \delta_G(A), e(A, B) = e_G(A, B)$.

Definition 3.1. Let y be a vertex in a graph G. Let $N(y) = \{x_1, \dots, x_{d(y)}\}$ be ordered with $d(x_1) \geq d(x_2) \geq \dots \geq d(x_{d(y)})$. Define the k-excess of y in G

$$k\text{-EX}_G(y) = \max(0, d(x_k) - d(y)).$$

If a vertex y has less than k neighbours, then this definition says $k\text{-EX}_G(y) = 0$. It's immediate from the definition that $k\text{-EX}_G(y) \geq 0$ always.

Observation 3.2. Let G be a graph and $y \in V(G)$ with $d(y) \geq d(x)$ for all $x \in N(y)$. Then $k\text{-EX}_G(y) = 0$ for all k.

Proof. Let $N(y) = \{x_1, \ldots, x_{d(y)}\}$ be ordered with $d(x_1) \geq d(x_2) \geq \cdots \geq d(x_{d(y)})$. If $d(y) \leq k$, then we immediately have $k\text{-EX}_G(y) = 0$. If $d(y) > k$, then, by assumption, we have $d(x_k) \leq d(y)$, which gives $k\text{-EX}_G(y) = \max(0, d(x_k) - d(y)) = 0$. □

Observation 3.3. In a graph G, let M be a matching, which matches X to Y. Then

$$\sum_{y \in Y} 1\text{-EX}_G(y) \geq \sum_{x \in X} d(x) - \sum_{y \in Y} d(y).$$

Proof. For each $y \in Y$, let $a_y = \max_{x \in N(y)} d(x)$ and let x_y be the vertex matched to y by M, noting that $a_y \geq d(x_y)$. We have that $1\text{-EX}_G(y) = \max(0, a_y - d(y)) \geq a_y - d(y) \geq d(x_y) - d(y)$. Summing this over all $y \in Y$ gives the result. □

The following lemma will imply Lemma 2.8.

Lemma 3.4 (1-excess sum)**.** *Let G be a bipartite graph with parts X, Y. Let $\delta(Y)$ denote the smallest degree in G out of vertices of Y. Then*

$$\sum_{y \in Y} 1\text{-EX}_G(y) \geq \delta(Y)(|Y| - |X|)$$

Proof. Let M be a maximum matching in G and C a minimum vertex cover. By König's Theorem we have $e(M) = |C|$ and so each edge of M contains precisely one vertex of C. In particular $V(M) \cap Y \setminus C$ is matched to $V(M) \cap X \cap C$. Since C is a vertex cover, we have $N(Y \setminus C) \subseteq V(M) \cap X \cap C$. This gives

$$\begin{aligned}
\sum_{y \in Y} 1\text{-EX}_G(y) &\geq \sum_{y \in V(M) \cap Y \setminus C} 1\text{-EX}_G(y) \geq \sum_{x \in V(M) \cap X \cap C} d(x) - \sum_{y \in V(M) \cap Y \setminus C} d(y) \\
&= e(V(M) \cap X \cap C, Y) - e(V(M) \cap Y \setminus C, X) \\
&\geq e(V(M) \cap X \cap C, Y \setminus C) - e(V(M) \cap Y \setminus C, X) \\
&= e(X, Y \setminus C) - e(V(M) \cap Y \setminus C, X) \\
&= e(Y \setminus V(M), X) \\
&= \sum_{y \in Y \setminus V(M)} d(y) \\
&\geq \delta(Y)|Y \setminus V(M)| \geq \delta(Y)(|Y| - |X|).
\end{aligned}$$

The first inequality is "$1\text{-EX}_G(y) \geq 0$ always". The second inequality is Observation 3.3 applied to $Y' = V(M) \cap Y \setminus C$ and $X' = V(M) \cap X \cap C$ (which we've already established are matched to each other). The third inequality is just $e(A, B) \geq e(A, B')$ for $B' \subseteq B$, which holds for all graphs. The second equation comes from "$N(Y \setminus C) \subseteq V(M) \cap X \cap C$". The third equation comes from $Y \setminus C$ being a disjoint union of $(Y \setminus V(M)) \setminus C = Y \setminus (V(M) \cup C) = Y \setminus V(M)$ and $(V(M) \cap Y) \setminus C$. The fourth inequality is $\sum_{a \in A} d(a) \geq \delta(A)|A|$ which holds for all graphs. The last inequality comes from $e(M) \leq |X|$ which happens because any matching in a bipartite graph G must be smaller than both of the parts of G. □

Lemma 3.5 (k-excess sum)**.** *Let G be a bipartite graph with parts X, Y. For a subset $Y' \subseteq Y$, let $\delta(Y')$ be the smallest degree in G out of vertices of Y'. Then*

$$\sum_{y \in Y'} k\text{-EX}_G(y) \geq \delta(Y')(|Y'| - |X|) - 2k|Y'|$$

Proof. If $\delta(Y') < k$, then the right hand side $\delta(Y')(|Y'| - |X|) - 2k|Y'| \le |Y'|(\delta(Y') - 2k)$ is negative and so the result trivially follows from "$k\text{-EX}_G(y) \ge 0$ always". So assume that $\delta(Y') \ge k$.

Construct a subgraph G' as follows: delete all vertices of $Y \setminus Y'$. For each $y \in Y'$, order $N(y) = \{x_1, \dots, x_{d(y)}\}$ with $d(x_1) \ge \dots \ge d(x_{d(y)})$ and delete the $k-1$ edges $yx_1, \dots, yx_{k-1}$.

Claim 3.6. $k\text{-EX}_G(y) \ge 1\text{-EX}_{G'}(y) - (k-1)$ for each $y \in Y'$.

Proof. Let $x \in N_{G'}(y)$ be the vertex with largest degree $d_{G'}(x)$, noting that $d_{G'}(x) \le d_G(x_k)$ (otherwise we'd have $d_G(x) \ge d_{G'}(x) > d_G(x_k)$, meaning that $x \in \{x_1, \dots, x_{k-1}\}$, contradicting that $x \in N_{G'}(y)$). From the definitions, $k\text{-EX}_G(y) = \max(0, d_G(x_k) - d_G(y))$ and

$$1\text{-EX}_{G'}(y) = \max(0, d_{G'}(x) - d_{G'}(y)) \tag{3.1}$$

$$\le \max(0, d_G(x_k) - d_{G'}(y)) = \max(0, d_G(x_k) - d_G(y) + k - 1). \tag{3.2}$$

When $1\text{-EX}_{G'}(y) = 0$, we have $d_G(x_k) - d_G(y) \le d_G(x_k) - d_G(y) + k - 1 \le 0$ and so $k\text{-EX}_G(y) = 0$, giving $k\text{-EX}_G(y) = 0 \ge -(k-1) \ge 1\text{-EX}_{G'}(y) - (k-1)$. When $1\text{-EX}_{G'}(y) > 0$, (3.2) gives $1\text{-EX}_{G'}(y) \le d_G(x_k) - d_G(y) + k - 1$, and so $k\text{-EX}_G(y) \ge d_G(x_k) - d_G(y) \ge 1\text{-EX}_{G'}(y) - (k-1)$. □

We also have $\delta_{G'}(Y') = \delta_G(Y') - (k-1)$. Applying Lemma 3.4 to G' gives us what we want

$$\begin{aligned}\sum_{y' \in Y'} k\text{-EX}_G(y) &\ge \sum_{y' \in Y'} (1\text{-EX}_{G'}(y) - (k-1)) \ge \delta_{G'}(Y')(|Y'| - |X|) - |Y'|(k-1) \\ &\ge (\delta_G(Y') - (k-1))(|Y'| - |X|) - |Y'|(k-1) \\ &\ge \delta_G(Y')(|Y'| - |X|) - 2k|Y'|.\end{aligned}$$

□

We associate a bipartite graph to every family $\mathcal{T}$.

Definition 3.7 (Availability graph)**.** Let $\mathcal{T} = (T_1, \dots, T_m)$ be a family of rainbow independent sets in a coloured matroid M. The *availability graph* of $\mathcal{T}$, denoted $A(\mathcal{T})$, is the bipartite graph with parts $\{T_1, \dots, T_m\}$ and $C(M)$, and with $T_i c_j$ an edge of $A(\mathcal{T})$ whenever $c_j \notin C(T_i)$.

Notice that the degree $d_{A(\mathcal{T})}(T_i)$ is the number of colours missing from T_i and the degree $d_{A(\mathcal{T})}(c)$ is the number of independent sets missing c. The two different definitions of excess that we introduced should now make sense because we have $\text{EX}(c, \mathcal{T}) = 1\text{-EX}_{A(\mathcal{T})}(c)$ for any colour missing from some $T \in \mathcal{T}$ (whereas for colours present on all $T \in \mathcal{T}$, the definitions disagree since we have $\text{EX}(c, \mathcal{T}) = n$ and $1\text{-EX}_{A(\mathcal{T})}(c) = 0$). Lemma 2.8 can now easily be deduced from Lemma 3.4 (although it is not used in the proof). The following is the analogue of Lemma 2.7 we use.

Lemma 3.8 (Increment lemma)**.** *Let $\mathcal{T}$ be a family of $\le n$ rainbow independent sets in a coloured matroid M with n colours of size $\ge n$, and c a colour. At least one of the following holds:*

(i) There is some $T_i \in \mathcal{T}$ for which there are at least $d_{A(\mathcal{T})}(c) + \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c)$ colour c elements $e \notin E(\mathcal{T})$ with $T_i + e$ a rainbow independent set.

(ii) The ℓ-reduced family has $e_{\mathcal{T}'_\ell}(c) \le e_{\mathcal{T}}(c) - \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c) + \ell n/k$.

Proof. If $k\text{-EX}_{A(\mathcal{T})}(c) = 0$, then (ii) is true by Observation 2.3, so we can assume that $k\text{-EX}_{A(\mathcal{T})}(c) > 0$. By the definition of $k\text{-EX}_{A(\mathcal{T})}(c)$, there are k rainbow independent sets $T_1, \dots, T_k \in N_{A(\mathcal{T})}(c) \subseteq \mathcal{T}$ with $d_{A(\mathcal{T})}(T_1) \geq \cdots \geq d_{A(\mathcal{T})}(T_k)$ and $k\text{-EX}_{A(\mathcal{T})}(c) = d_{A(\mathcal{T})}(T_k) - d_{A(\mathcal{T})}(c)$. Using the definition of $A(\mathcal{T})$, we have that each T_i misses colour c and the number of colours each T_i misses is $d_{A(\mathcal{T})}(T_i) \geq d_{A(\mathcal{T})}(T_k) = d_{A(\mathcal{T})}(c) + k\text{-EX}_{A(\mathcal{T})}(c)$. Define a bipartite graph H whose parts are $X = \{T_1, \dots, T_k\}$ and $E(c)$ with T_ie an edge whenever $T_i + e$ is a (rainbow) independent set. The augmentation property tells us that $d_H(T_i) \geq |E(c)| - |T_i|$ (the augmentation property gives an independent B of size $|E(c)|$ with $B \subseteq E(C) \cup T_i$ and $T_i \subseteq B$. All $|B| - |T_i| = |E(C)| - |T_i|$ elements $e \in B \setminus T_i \subseteq E(C)$ have the property that $T_i + e$ is rainbow independent). Using this, and the fact that M consists of n colours of size $\geq n$ (each of which is an independent set), we have that for each $i = 1, \dots, k$

$$d_H(T_i) \geq |E(c)| - |T_i| \geq n - |T_i| = d_{A(\mathcal{T})}(T_i) \geq d_{A(\mathcal{T})}(c) + k\text{-EX}_{A(\mathcal{T})}(c).$$

Equivalently, there are at least $d_{A(\mathcal{T})}(c) + k\text{-EX}_{A(\mathcal{T})}(c)$ colour c elements e with $T_i + e$ a rainbow independent set. If, for some $i = 1, \dots, k$, at least $d_{A(\mathcal{T})}(c) + \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c)$ of these have $e \notin E(\mathcal{T})$, then case (i) of the lemma holds.

Thus we can assume that for all $i = 1, \dots, k$, there are $\geq \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c)$ colour c elements $e \in E(\mathcal{T})$ with $T_i + e$ a rainbow independent set. Let H' be the induced subgraph of H on $X = \{T_1, \dots, T_k\}$ and $E(c) \cap E(\mathcal{T})$ (so we have that the smallest degree in H' of vertices in X satisfies $\delta_{H'}(X) \geq \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c)$). Let $E_{\geq \ell} \subseteq E(c) \cap E(\mathcal{T})$ be the set of elements e with $d_{H'}(e) \geq \ell$. We have

$$\begin{aligned}|X||E_{\geq \ell}| + \ell|E(c) \cap E(\mathcal{T})| &\geq \sum_{e \in E_{\geq \ell}} |X| + \sum_{e \in E(c) \cap E(\mathcal{T}) \setminus E_{\geq \ell}} \ell \geq e(H') \\ &\geq |X|\delta_{H'}(X) \geq |X|\frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c).\end{aligned}$$

Using $|X| = k$ and rearranging gives

$$|E_{\geq \ell}| \geq \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c) - \ell|E(c) \cap E(\mathcal{T})|/k \geq \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c) - \ell n/k.$$

Here the second inequality used $|E(c) \cap E(\mathcal{T})| \leq n$. From the definition of the ℓ-reduced family $\mathcal{T}'_\ell$ and H, we have that every element $e \in E(c) \cap E(\mathcal{T})$ having $d_H(e) \geq \ell$ is absent from $E_{\mathcal{T}'_\ell}(c)$, i.e., that $E_{\mathcal{T}'_\ell}(c) \subseteq E_{\mathcal{T}}(c) \setminus E_{\geq \ell}$. This implies $e_{\mathcal{T}'_\ell}(c) \leq e_{\mathcal{T}}(c) - |E_{\geq \ell}| \leq e_{\mathcal{T}}(c) - \frac{1}{2}k\text{-EX}_{A(\mathcal{T})}(c) + \ell n/k$, demonstrating (ii). □

Next we prove the analogue of Lemma 2.6 we need. This analogue is Lemma 3.10. In order to facilitate its inductive proof, we first prove the following far more technical lemma, and then use it to deduce Lemma 3.10.

Lemma 3.9. *Let $\ell \geq 2$, $m \in \mathbb{N}$, and let M be a coloured matroid. Suppose that we have $r \leq \log_9 \ell - 1, \mathcal{S}, \mathcal{T}, R, F$ satisfying:*

(B1) $\mathcal{S} = (S_1, \dots, S_m), \mathcal{T} = (T_1, \dots, T_m)$ are families of disjoint rainbow independent sets with $\mathcal{T} \leq \mathcal{S}$.
(B2) $R \subseteq [m]$ with $|R| \leq \ell/4^r$.
(B3) $F \subseteq E(\mathcal{S})$ with $|F| \leq \ell/4^r$.
(B4) F is disjoint from $E(\mathcal{T}_\ell^{(r)})$.

Then there are $\mathcal{S}^*$, X, with $r, \mathcal{S}^*, X, \mathcal{S}, \mathcal{T}, R, F$ satisfying:

(C1) $\mathcal{S}^*$ *is a family of disjoint rainbow independent sets.*
(C2) $E(\mathcal{S}^*) = E(\mathcal{S}) \setminus X$.
(C3) For all $i \in R$, *we have* $(\mathcal{S}^*)_i = S_i \setminus F$.
(C4) $X \subseteq E(\mathcal{S}) \setminus E(\mathcal{T})$ *with* $|X| \le 2^r|F|$.

Proof. Fix ℓ, m, M. The proof is by induction on r. So we prove the cases $r = 0, 1, \dots, \lfloor \log_9 \ell \rfloor$ in order (for all valid choices of $\mathcal{S}, \mathcal{T}, R, F$). For the initial case "$r = 0$", consider some $\mathcal{S}, \mathcal{T}, R, F$ satisfying (B1) – (B4) with $r = 0$. We claim that $\mathcal{S}^* = (S_j \setminus F : j = 1, \dots, m)$, $X = F$ satisfy (C1) – (C4) (together with $r = 0, \mathcal{S}, \mathcal{T}, R, F$).

(C1) By construction we have $\mathcal{S}^* \le \mathcal{S}$. Since $\mathcal{S}$ is a family of disjoint rainbow independent sets, the same is true for $\mathcal{S}^*$.
(C2) We have $(\mathcal{S}^*)_j = S_j \setminus F$ for each j, which gives $E(\mathcal{S}^*) = \bigcup_{j=1}^m S_j \setminus F = E(\mathcal{S}) \setminus F = E(\mathcal{S}) \setminus X$.
(C3) We have $(\mathcal{S}^*)_i = (\mathcal{S})_i \setminus F$ for all $i \in [m]$ (and so in particular for all $i \in R$).
(C4) X satisfies $|X| = |F| = 2^0|F|$ and $X = F \subseteq E(\mathcal{S}) \setminus E(\mathcal{T}_\ell^{(0)}) = E(\mathcal{S}) \setminus E(\mathcal{T})$ (by (B3),(B4)).

Now let $r \ge 1$, and suppose that the lemma holds for all smaller r. Let $r, \mathcal{S}, \mathcal{T}, R, F$ satisfy (B1) – (B4). Partition F into $F_1 = F \cap E(\mathcal{T}_\ell^{(r-1)})$ and $F_2 = F \setminus E(\mathcal{T}_\ell^{(r-1)})$. For each $f \in F_1$, by the definition of ℓ-reduction, there are ℓ indices $i(f)$ for which $(\mathcal{T}_\ell^{(r-1)})_{i(f)} + f$ is rainbow and independent. For each $f \in F_1$, fix such an index with $i(f) \notin R \cup \{i : S_i \cap F \neq \emptyset\}$ and also $i(f), i(f')$ are distinct for different $f, f' \in F_1$ (in total there are $\le |R| + |F| + |F_1| \le |R| + 2|F| \le 3\ell/4^r < \ell$ indices to avoid, so there's always space to choose each $i(f)$). Note that since $i(f) \notin \{i : S_i \cap F \neq \emptyset\}$, we get that $f \notin S_{i(f)}$.

Using Observation 2.1 (applied with $e = f, S = S_{i(f)}, T = (\mathcal{T}_\ell^{(r-1)})_{i(f)}$), for each $f \in F_1$, there is a set $Q(f) \subseteq S_{i(f)} \setminus (\mathcal{T}_\ell^{(r-1)})_{i(f)} \subseteq E(\mathcal{S}) \setminus E(\mathcal{T}_\ell^{(r-1)})$ of size at most 2 with $S_{i(f)} \setminus Q(f) \cup \{f\}$ rainbow and independent. Let $\hat{F} = F_2 \cup \bigcup_{f \in F_1} Q(f)$. Note that since indices $i(f), i(f') \notin \{i : S_i \cap F \neq \emptyset\}$ are distinct, we have $S_{i(f)} \cap F = \emptyset$ and $S_{i(f)} \cap \bigcup_{f' \in F} Q(f') = Q(f)$. This gives that $S_{i(f)} \cap \hat{F} = Q(f)$ and that $Q(f) \subseteq S_{i(f)}$ is disjoint from F.

Let $\hat{\mathcal{S}} = (S_j \setminus F_1 : j = 1, \dots, m)$, $\hat{\mathcal{T}} = (T_j \setminus F_1 : j = 1, \dots, m)$, $\hat{R} = R \cup \{i(f) : f \in F_1\}$. We claim that $r - 1, \hat{\mathcal{S}}, \hat{\mathcal{T}}, \hat{R}, \hat{F}$ satisfy (B1) – (B4).

(B1) $\hat{\mathcal{S}}, \hat{\mathcal{T}}$ are families of disjoint rainbow independent sets, since they're formed from such families $\mathcal{S}, \mathcal{T}$ by deleting elements. We have $\mathcal{T} \le \mathcal{S}$ which means that for all i, we have $T_i \subseteq S_i$ and hence $T_i \setminus F_1 \subseteq S_i \setminus F_1$, giving $\hat{\mathcal{T}} \le \hat{\mathcal{S}}$.
(B2) Note that $|\hat{R}| \le |R| + |F| \le 2\ell/4^r \le \ell/4^{r-1}$.
(B3) For "$\hat{F} \subseteq E(\hat{\mathcal{S}})$", note that by construction $E(\hat{\mathcal{S}}) = E(\mathcal{S}) \setminus F_1$. Next note that $\hat{F} \subseteq E(\mathcal{S})$. Indeed, $F_2 \subseteq F \subseteq E(\mathcal{S})$ (by (B3) holding for $F, \mathcal{S}$) and $Q(f) \subseteq E(\mathcal{S})$ for each f by their definition. Finally, note that $\hat{F} \cap F_1 = \emptyset$. Indeed, $F_1 \cap F_2 = \emptyset$ and "$Q(f)$ is disjoint from F" show this. Combining $\hat{F} \subseteq E(\mathcal{S})$, $\hat{F} \cap F_1 = \emptyset$, $E(\hat{\mathcal{S}}) = E(\mathcal{S}) \setminus F_1$ gives $\hat{F} \subseteq E(\hat{\mathcal{S}})$.
For the second part, note that $|\hat{F}| \le |F_2| + 2|F_1| \le 2|F| \le 2 \cdot \ell/4^r \le \ell/4^{r-1}$.
(B4) Recall that $Q(f) \subseteq E(\mathcal{S}) \setminus E(\mathcal{T}_\ell^{(r-1)})$ and $F_2 = F \setminus E(\mathcal{T}_\ell^{(r-1)})$, which shows that $\hat{F}$ is disjoint from $E(\mathcal{T}_\ell^{(r-1)})$. We also have $\hat{\mathcal{T}} \le \mathcal{T}$. Together with

Observation 2.4, this shows that $\hat{\mathcal{T}}_\ell^{(r-1)} \le \mathcal{T}_\ell^{(r-1)}$, and so $\hat{F}$ is disjoint from $E(\hat{\mathcal{T}}_\ell^{(r-1)})$.

By induction, there are $\hat{\mathcal{S}}^*, \hat{X}$, with $\hat{\mathcal{S}}^*, \hat{X}, r-1, \hat{\mathcal{S}}, \hat{\mathcal{T}}, \hat{R}, \hat{F}$ satisfying (C1) – (C4) (with "r" replaced by "$r-1$" in (C4)). Note that for each $f \in F_1$ we have $(\hat{\mathcal{S}}^*)_{i(f)} = (\hat{\mathcal{S}})_{i(f)} \setminus \hat{F} = S_{i(f)} \setminus Q(f)$ (the first equation uses (C3) and $i(f) \in \hat{R}$, while the second uses $(\hat{\mathcal{S}})_{i(f)} = S_{i(f)} \setminus F_1 = S_{i(f)}$, $\hat{F} \cap S_{i(f)} = Q(f)$). Let $\mathcal{S}^*$ be $\hat{\mathcal{S}}^*$ with $(\hat{\mathcal{S}}^*)_{i(f)} = S_{i(f)} \setminus Q(f)$ replaced by $S_{i(f)} \setminus Q(f) \cup \{f\}$ for each $f \in F_1$. We show that $\mathcal{S}^*, \hat{X}$ satisfy (C1) – (C4) (together with $r, \mathcal{S}, \mathcal{T}, R, F$), proving the lemma.

(C1) First note that for all i, $(\mathcal{S}^*)_i$ is rainbow and independent. Indeed for $i \notin \{i(f) : f \in F_1\}$, we have $(\mathcal{S}^*)_i = (\hat{\mathcal{S}}^*)_i$, which is rainbow and independent due to $\hat{\mathcal{S}}^*$ satisfying (C1). For $i = i(f)$ for $f \in F_1$, we have $(\mathcal{S}^*)_i = S_{i(f)} \setminus Q(f) \cup \{f\}$ which is rainbow and independent from Observation 2.1.

Next we show that for $i \neq j$ we have $(\mathcal{S}^*)_i, (\mathcal{S}^*)_j$ disjoint. To see this note that, by construction of $\mathcal{S}^*$, for all i we have $(\mathcal{S}^*)_i \setminus F_1 = (\hat{\mathcal{S}}^*)_i$. Together with $\hat{\mathcal{S}}^*$ satisfying (C1), this shows that for $i \neq j$, we have $(\mathcal{S}^*)_i \cap (\mathcal{S}^*)_j \subseteq F_1$. But, by construction of $\mathcal{S}^*$, we have that each $f \in F_1$ is in precisely one set of $\mathcal{S}^*$, which shows that $(\mathcal{S}^*)_i \cap (\mathcal{S}^*)_j = \emptyset$ always.

(C2) We have $E(\mathcal{S}^*) = E(\hat{\mathcal{S}}^*) \cup F_1$ by construction. Also $E(\hat{\mathcal{S}}^*) = E(\hat{\mathcal{S}}) \setminus \hat{X}$ (from $\hat{\mathcal{S}}^*, \hat{\mathcal{S}}, \hat{X}$ satisfying (C2)). From construction of $\hat{\mathcal{S}}$, we have $E(\hat{\mathcal{S}}) = E(\mathcal{S}) \setminus F_1$. From (C4), we have $\hat{X} \subseteq E(\hat{\mathcal{S}})$ which shows $\hat{X} \cap F_1 = \emptyset$. Combining all these gives

$$E(\mathcal{S}^*) = E(\hat{\mathcal{S}}^*) \cup F_1 = (E(\hat{\mathcal{S}}) \setminus \hat{X}) \cup F_1 = ((E(\mathcal{S}) \setminus F_1) \setminus \hat{X}) \cup F_1 = E(\mathcal{S}) \setminus \hat{X}$$

as required by (C2).

(C3) For all $i \in R \subseteq \hat{R}$, we have $(\hat{\mathcal{S}}^*)_i = (\hat{\mathcal{S}})_i \setminus \hat{F}$ by $\hat{\mathcal{S}}, \hat{\mathcal{S}}^*, \hat{R}, \hat{F}$ satisfying (C3). By construction of $\hat{\mathcal{S}}$, we have $(\hat{\mathcal{S}})_i = S_i \setminus F_1$, giving $(\hat{\mathcal{S}}^*)_i = (S_i \setminus F_1) \setminus \hat{F} = S_i \setminus (F_1 \cup F_2 \cup \{Q(f) : f \in F_1\})$. Since $Q(f) \subseteq S_{i(f)}$ for $i(f) \notin R$, we have that for $i \in R$ we have

$$S_i \setminus (F_1 \cup F_2 \cup \{Q(f) : f \in F_1\}) = S_i \setminus (F_1 \cup F_2) = S_i \setminus F.$$

Thus we have shown that for $i \in R$ we have $(\hat{\mathcal{S}}^*)_i = S_i \setminus F$, as required.

(C4) Since (C4) applies to $\hat{X}, \hat{\mathcal{S}}, \hat{\mathcal{T}}, \hat{F}$ we have $|\hat{X}| \le 2^{r-1}|\hat{F}| \le 2^{r-1}(|F_2| + 2|F_1|) \le 2^r|F|$, and also $\hat{X} \subseteq E(\hat{\mathcal{S}}) \setminus E(\hat{\mathcal{T}}) = (E(\mathcal{S}) \setminus F_1) \setminus (E(\mathcal{T}) \setminus F_1) = E(\mathcal{S}) \setminus E(\mathcal{T})$. □

We now prove the analogue of Lemma 2.6 that we need. For technical reasons there are two families $\mathcal{S}, \mathcal{T}$ in this lemma, but the most important case is when $\mathcal{S} = \mathcal{T}$. In that case the lemma is an extension of Lemma 2.6 to larger r, ℓ.

Lemma 3.10 (Switching lemma)**.** *Let $\ell \ge 9^{r+1}$. Let $\mathcal{T}$, $\mathcal{S}$ be two families of disjoint rainbow independent sets with $\mathcal{T} \le \mathcal{S}$. Suppose we have an $e \notin E(\mathcal{S})$ with $T + e$ rainbow and independent for some $T \in \mathcal{T}_\ell^{(r)}$. Then there is a family of disjoint rainbow independent sets $\mathcal{S}^*$ with $E(\mathcal{S}^*) = \{e\} \cup (E(\mathcal{S}) \setminus X)$, for some $X \subseteq E(\mathcal{S}) \setminus E(\mathcal{T})$ with $|X| \le 2^{r+1}$.*

Proof. Let $\mathcal{S} = \{S_1, \dots, S_m\}$ and $S_i \in \mathcal{S}$ with $T \subseteq S_i$. By Observation 2.1, there is some $F \subseteq S_i \setminus T$ of size $\le 2 \le \ell/4^r$ with $S_i \setminus F \cup \{e\}$ a rainbow independent set. Notice that $r, \mathcal{S}, \mathcal{T}, R = \{i\}, F$ satisfy (B1) – (B4). Indeed (B1) comes from the lemma's assumptions, (B2) is true since $|R| = 1 \le \ell/4^r$, while (B3) and (B4) come from Observation 2.1. Applying that lemma gives us $\hat{\mathcal{S}}^*, X$ satisfying (C1) – (C4). Note that by (C3) and $i \in R$, we have $S_i \setminus F \in \hat{\mathcal{S}}^*$. Let $\mathcal{S}^*$ be $\hat{\mathcal{S}}^*$ with $S_i \setminus F$ replaced by $S_i \setminus F \cup \{e\}$. Using $e \notin E(\mathcal{S})$ and (C1), this is a family of disjoint rainbow independent sets. Using (C2), we have $E(\mathcal{S}^*) = \{e\} \cup E(\hat{\mathcal{S}}^*) = \{e\} \cup (E(\mathcal{S}) \setminus X)$ as required. Finally "$X \subseteq E(\mathcal{S}) \setminus E(\mathcal{T})$ with $|X| \le 2^{r+1}$" comes from (C4) and $|F| \le 2$. □

We are now ready to show that Rota's conjecture holds asymptotically.

Proof of Theorem 1.2. Let $\epsilon > 0$ with $\epsilon^{-1} \in \mathbb{N}$. Fix $r_0 = 100/\epsilon^2$, $\ell = 9^{r_0+1}$, $k = 8\epsilon^{-2}\ell$, $h = 3^{r_0}/\epsilon$, noting that these are all integers. Let n be sufficiently large compared to r_0, k, h, ϵ — formally, picking any $n \ge 100^{r_0}$ works. Let M be a coloured matroid with n colours of size n. All rainbow independent sets throughout the proof will use these colours (and so have size $\le n$). For a family of disjoint rainbow independent sets $\mathcal{S}$ we say that an ordering $\sigma = (c_1, \dots, c_n)$ of the colours $C(M)$ is *increasing for* $\mathcal{S}$ if it satisfies $e_{\mathcal{S}}(c_1) \le \dots \le e_{\mathcal{S}}(c_n)$.

Define a partial order $\succ_{\text{LEX}}$ on families of disjoint rainbow independent sets $\mathcal{S}, \mathcal{T}$ by defining $\mathcal{S} \succ_{\text{LEX}} \mathcal{T}$ if there exists an ordering $\sigma_S = (c_1, \dots, c_n)$ which is increasing for $\mathcal{S}$ and an ordering $\sigma_T = (d_1, \dots, d_n)$ which is increasing for $\mathcal{T}$ such that the smallest index q with $e_{\mathcal{S}}(c_q) \ne e_{\mathcal{T}}(d_q)$ has $e_{\mathcal{S}}(c_q) > e_{\mathcal{T}}(d_q)$.

Claim 3.11. $\succ_{\text{LEX}}$ is a partial order.

Proof. Define a vector $v(\mathcal{S}) \in \mathbb{Z}^n$ whose ith coordinate is $e_{\mathcal{S}}(c_i)$ where c_i is the ith colour in an increasing ordering for $\mathcal{S}$. Note that the vector $v(\mathcal{S})$ depends only on $\mathcal{S}$, rather than which increasing ordering σ_S we are considering for $\mathcal{S}$. Also, note that $\mathcal{S} \succ_{\text{LEX}} \mathcal{T} \iff v(\mathcal{S})$ is bigger than $v(\mathcal{T})$ in the usual lexicographic ordering on $\mathbb{Z}^n$. Now "$\succ_{\text{LEX}}$" being a partial order comes from the same being true in the usual lexicographic ordering — it is a general phenomenon that for a function $f : S \to P$ from a set S to a poset P, defining "$s > s' \iff f(s) > f(s')$" gives a partial order on S. □

Let $\mathcal{R} = (R_1, \dots, R_{(1-\epsilon)n})$ be a family of $(1-\epsilon)n$ disjoint rainbow independent sets in M which is maximal with respect to $\succ_{\text{LEX}}$. Suppose for the sake of contradiction that there are at more than $3\epsilon n$ colours c with $e_{\mathcal{R}}(c) \le (1-4\epsilon)n$.

Consider an arbitrary ordering $\sigma_0 = (c_1, \dots, c_n)$ of the colours which is increasing for $\mathcal{R}$.

Claim 3.12. There is some $m \ge 2\epsilon n$ with $e_{\mathcal{R}}(c_m) \le (1-2\epsilon)n$ and for all i with $m + h \le i \le n$, we have $e_{\mathcal{R}}(c_i) \ge e_{\mathcal{R}}(c_m) + 3^{r_0}$.

Proof. Let $t \le n$ be the largest index with $e_{\mathcal{R}}(c_t) \le (1-2\epsilon)n$. Note that t must exist and satisfy $t \ge 3\epsilon n$, since otherwise we'd have a contradiction to "there are at more than $3\epsilon n$ colours c with $e_{\mathcal{R}}(c) \le (1-4\epsilon)n$". If $t > n - h$, then picking $m = t$ works (since then there are no indices i with $m + h \le i \le n$). So suppose $t \le n - h$. Note that to prove the lemma it is enough to find some m with $2\epsilon n \le m \le t$ with $e_{\mathcal{R}}(c_{m+h}) \ge e_{\mathcal{R}}(c_m) + 3^{r_0}$. So suppose for contradiction that

$$e_{\mathcal{R}}(c_{m+h}) - 3^{r_0} < e_{\mathcal{R}}(c_m) \text{ for all } 2\epsilon n \le m \le t \tag{3.3}$$

For each integer i, set $m_i := \lceil 2\epsilon n \rceil + ih$. Noting that $m_0 < 3\epsilon n \le t$ and the m_is are strictly increasing integers, there must exist some index $k \ge 0$, for which $m_k \le t$ and $m_{k+1} > t$. Note that we have $k \le t/h$ (since for $k' > t/h$, we have $m_{k'} = \lceil 2\epsilon n \rceil + k'h > \lceil 2\epsilon n \rceil + h(t/h) > t$). For $i \in \{0, \dots, k\}$ we have $2\epsilon n \le m_i \le t$, and so (3.3) gives $e_{\mathcal{R}}(c_{m_{i+1}}) - 3^{r_0} = e_{\mathcal{R}}(c_{m_i+h}) - 3^{r_0} < e_{\mathcal{R}}(c_{m_i})$. Add these inequalities for $i = 0, \dots, k$ to get $\sum_{j=0}^{k} e_{\mathcal{R}}(c_{m_{i+1}}) - \sum_{j=0}^{k} 3^{r_0} < \sum_{j=0}^{k} e_{\mathcal{R}}(c_{m_i})$ which is equivalent to $e_{\mathcal{R}}(c_{m_{k+1}}) - (k+1)3^{r_0} < e_{\mathcal{R}}(c_{m_0}) = e_{\mathcal{R}}(c_{\lceil 2\epsilon n \rceil})$. Using $m_{k+1} > t$ and the choice of t we have $e_{\mathcal{R}}(c_{m_{k+1}}) > (1-2\epsilon)n$. Thus

$$\begin{aligned} e_{\mathcal{R}}(c_{\lceil 2\epsilon n \rceil}) &> (1-2\epsilon)n - (k+1)3^{r_0} \\ &\ge (1-2\epsilon)n - (t/h+1)3^{r_0} = (1-2\epsilon)n - \epsilon t - 3^{r_0} > (1-4\epsilon)n \end{aligned}$$

The first inequality is "$e_{\mathcal{R}}(c_{m_{k+1}}) > (1-2\epsilon)n$" combined with "$e_{\mathcal{R}}(c_{m_{k+1}}) - (k+1)3^{r_0} < e_{\mathcal{R}}(c_1)$". The second inequality is $k \le t/h$. The equality is $h = 3^{r_0}/\epsilon$. The last inequality uses $3^{r_0} \le \epsilon n$, which is true because n is sufficiently large (and it also follows from $r_0 = 100/\epsilon^2, n \ge 100^{100/\epsilon^2}$). This is again a contradiction to "there are at more than $3\epsilon n$ colours c with $e_{\mathcal{R}}(c) \le (1-4\epsilon)n$". □

We call the colours $c_1, \dots, c_m$ *small*, the colours $c_{m+1}, \dots, c_{m+h}$ *medium*, and the colours $c_{m+h+1}, \dots, c_n$ *large*. Note that there are $m \ge 2\epsilon n > \epsilon n$ small colours. For a family $\mathcal{F}$, use $E_{\text{small}}(\mathcal{F})/E_{\text{medium}}(\mathcal{F})/E_{\text{large}}(\mathcal{F})$ to denote the sets of elements of corresponding colours in $\mathcal{F}$.

Claim 3.13. There is a family of disjoint rainbow independent sets $\mathcal{R}^*$ with $E_{\text{small}}(\mathcal{R}^*) = E_{\text{small}}(\mathcal{R}) + e$ for some element e outside $\mathcal{R}$, and also $E_{\text{medium}}(\mathcal{R}^*) = E_{\text{medium}}(\mathcal{R})$, and $|E_{\text{large}}(\mathcal{R}) \setminus E_{\text{large}}(\mathcal{R}^*)| \le 2^{r_0+1}$.

This claim implies the theorem since it implies $\mathcal{R}^* \succ_{\text{LEX}} \mathcal{R}$ (contradicting maximality of $\mathcal{R}$). To check the definition of $\mathcal{R}^* \succ_{\text{LEX}} \mathcal{R}$: we need two orderings σ/σ^* of the colours $C(M)$ which are increasing for $\mathcal{R}/\mathcal{R}^*$ respectively. Let c_t be the unique small colour with $e_{\mathcal{R}^*}(c_t) = e_{\mathcal{R}}(c_t) + 1$ which is guaranteed to exist by the claim. The claim also guarantees that for all the other small/medium colours c_i we have $e_{\mathcal{R}^*}(c_i) = e_{\mathcal{R}}(c_i)$. Let k be the largest index for which $e_{\mathcal{R}}(c_k) = e_{\mathcal{R}}(c_t)$, noting that $t \le k \le m+h$ (for the last inequality first observe $e_{\mathcal{R}}(c_t) \le e_{\mathcal{R}}(c_m)$, since t is small. Then note that, from Claim 3.12, if we had $k \ge m+h$, then we'd have $e_{\mathcal{R}}(c_k) \ge e_{\mathcal{R}}(c_m) + 3^{r_0} > e_{\mathcal{R}}(c_m) \ge e_{\mathcal{R}}(c_t)$. This would contradict that $e_{\mathcal{R}}(c_k) = e_{\mathcal{R}}(c_t)$). Thus the colours $c_1, \dots, c_k$ are all small/medium.

Now we construct the two orderings σ, σ^*. The ordering σ is constructed from σ_0 by moving the colour c_t to be between c_k and c_{k+1}. Let $d_{k+1}, d_{k+2}, \dots, d_n$ be the colours $c_{k+1}, c_{k+2}, \dots, c_n$, but arranged in order of how many elements they have in $\mathcal{R}^*$. The ordering σ^* is constructed from σ by rearranging $c_{k+1}, c_{k+1}, \dots, c_n$ in the order $d_{k+1}, d_{k+2}, \dots, d_n$. Formally, define:

$$\begin{aligned} \sigma &= (c_1, c_2, \dots, c_{t-1}, c_{t+1}, \dots, c_{k-1}, c_k, c_t, c_{k+1}, \dots, c_n) \\ \sigma^* &= (c_1, c_2, \dots, c_{t-1}, c_{t+1}, \dots, c_{k-1}, c_k, c_t, d_{k+1}, \dots, d_n) \end{aligned}$$

Note that σ/σ^* is increasing for $\mathcal{R}/\mathcal{R}^*$ respectively. Indeed σ is increasing for $\mathcal{R}$, since the original ordering σ_0 was increasing for $\mathcal{R}$ and since we have $e_{\mathcal{R}}(c_t) = e_{\mathcal{R}}(c_{t+1}) = \dots = e_{\mathcal{R}}(c_k)$ by definition of c_k. This also shows that $\mathcal{R}^*$ is increasing on the segment $(c_1, c_2, \dots, c_{t-1}, c_{t+1}, \dots, c_{k-1}, c_k)$ of σ^* (since we've established that these colours are small/medium and have $e_{\mathcal{R}^*}(c_i) = e_{\mathcal{R}}(c_i)$). Also the segment of $\mathcal{R}^*$ on $(d_{k+1}, \dots, d_n)$ is increasing for $\mathcal{R}^*$ by construction. Thus it remains to show that in $\mathcal{R}^*$ the colour c_t has at least as many elements as colours that precede it in σ^* and that c_t has at most as many elements as colours that succeed it in σ^*. This amounts to:

- For all $i > k$ we have $e_{\mathcal{R}^*}(c_i) \geq e_{\mathcal{R}^*}(c_t)$: If c_i is small/medium, then this happens because $e_{\mathcal{R}^*}(c_i) = e_{\mathcal{R}}(c_i) \geq e_{\mathcal{R}}(c_{k+1}) \geq e_{\mathcal{R}}(c_k)+1 = e_{\mathcal{R}}(c_t)+1 = e_{\mathcal{R}^*}(c_t)$. If c_i is large, then, using both claims, we have

$$\begin{aligned} e_{\mathcal{R}^*}(c_i) &\geq e_{\mathcal{R}}(c_i) - 2^{r_0} \geq e_{\mathcal{R}}(c_{m+h}) - 2^{r_0} \geq e_{\mathcal{R}}(c_m) + 3^{r_0} - 2^{r_0} \\ &\geq e_{\mathcal{R}}(c_m) + 1 \geq e_{\mathcal{R}}(c_t) + 1 = e_{\mathcal{R}^*}(c_t). \end{aligned}$$

- For all $i \leq k$ we have $e_{\mathcal{R}^*}(c_i) \leq e_{\mathcal{R}^*}(c_t)$: For $i = t$, we have equality, while for other $i \leq k$ we have $e_{\mathcal{R}^*}(c_i) = e_{\mathcal{R}}(c_i) \leq e_{\mathcal{R}}(c_k) = e_{\mathcal{R}}(c_t) = e_{\mathcal{R}^*}(c_t)-1 \leq e_{\mathcal{R}^*}(c_t)$.

Proof of Claim 3.13. The basic idea of the proof is to use Lemma 3.10 to add a small colour element to $\mathcal{R}$. However we do not apply the lemma to $\mathcal{R}$ directly (in order to avoid medium colours from being affected). Instead we add "dummy elements" $d_{i,j}$ of medium colours to independent sets from $\mathcal{R}$ in order to obtain a new family $\mathcal{S}$ with the property that every $S \in \mathcal{S}$ contains every medium colour. These dummy elements need not come from M — we enlarge M by adding as many new medium colour dummy elements as are needed in an arbitrary fashion. Then, at the end of the proof, we delete the dummy colours to end up with a family $\mathcal{R}^*$ satisfying the claim. Formally: Let M' be a coloured matroid of rank $n+hn$ formed by adding a set of hn new elements $D = \{d_{i,j} : i \in [m+1, m+h], j \in [n]\}$, which are independent of everything (meaning that the independent sets in M' are $I \cup J$ where I is independent in M, while J is any subset of D). Give the element $d_{i,j}$ the colour i for all j (this colour is necessarily medium since $i \in [m+1, m+h]$). For each rainbow independent set $R_j \in \mathcal{R}$ (with $j = 1, \ldots, (1-\epsilon)n$), let $S_j = R_j \cup \{d_{i,j} : i = m+1, \ldots, m+h \text{ and } c_i \notin C(R_j)\}$, noting that S_j is rainbow. Now let $\mathcal{S} = (S_j : j = 1, \ldots, (1-\epsilon)n)$ to get our new family of disjoint rainbow independent sets (the sets are all disjoint because each $d_{i,j}$ is in at most one set, namely S_j). Note that now each rainbow independent set in $\mathcal{S}$ contains each medium colour, while on the small/large colours, the sets in $\mathcal{S}$ are the same as those in $\mathcal{R}$.

Let $\mathcal{T}$ be formed from $\mathcal{S}$ by deleting all large colour elements. Notice that $A(\mathcal{T})$ has parts of size $|\mathcal{T}| = (1-\epsilon)n$ and $|C(M)| = n$ with all small colours having $d_{A(\mathcal{T})}(c) \geq \epsilon n$ (since they occur in less than $e_{\mathcal{R}}(c_m) \leq (1-2\epsilon)n$ of the rainbow independent sets of $\mathcal{R}$ — and the same holds for $\mathcal{S}, \mathcal{T}$ since small colours are unchanged for these families), all medium colours having $d_{A(\mathcal{T})}(c) = 0$ (since every $T \in \mathcal{S}$ contains every medium colour by construction of $\mathcal{S}$ — and the same holds for $\mathcal{T}$, since medium colours are unchanged when building it), and all large colours having $d_{A(\mathcal{T})}(c) = (1-\epsilon)n \geq \epsilon n$ (since none of the $T \in \mathcal{T}$ contain large colour elements). Define $Y_{\text{small/large}}$ to be the set of small/large colours. By Lemma 3.5

we have

$$\begin{aligned}\sum_{c\in C(M)} k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c) &\geq \sum_{c\in Y_{\text{small/large}}} k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c)\\ &\geq \delta_{A(\mathcal{T}_\ell^{(r)})}(Y_{\text{small/large}})(|Y_{\text{small/large}}|-|\mathcal{T}|)-2k|Y_{\text{small/large}}|\\ &\geq \delta_{A(\mathcal{T})}(Y_{\text{small/large}})(|Y_{\text{small/large}}|-|\mathcal{T}|)-2k|Y_{\text{small/large}}|\\ &\geq \epsilon n(|Y_{\text{small/large}}|-|\mathcal{T}|)-2k|Y_{\text{small/large}}|\\ &= \epsilon n(n-h-(1-\epsilon)n)-2k(n-h)\\ &= \epsilon^2n^2(1-h/\epsilon n-2k/\epsilon^2 n+2kh/\epsilon^2n^2)\\ &\geq \epsilon^2n^2/2 \qquad (3.4)\end{aligned}$$

The first inequality used that $k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c)\geq 0$ always. The second inequality is exactly Lemma 3.5 (applied to the graph $A(\mathcal{T}_\ell^{(r)})$ with $X=\mathcal{T}$, $Y=C(M)$, $Y'=Y_{\text{small/large}}$) . The third inequality uses that $\mathcal{T}_\ell^{(r)}\leq\mathcal{T}$ by Observation 2.3, which implies that $A(\mathcal{T}_\ell^{(r)})\supseteq A(\mathcal{T})$ (and so each vertex has $d_{A(\mathcal{T}_\ell^{(r)})}(v)\geq d_{A(\mathcal{T})}(v)$). The third inequality uses that small/large colours c have $d_{A(\mathcal{T})}(c)\geq\epsilon n$, by the preceding paragraph. The first equality uses that the number of small/large colours is $|Y_{\text{small/large}}|=n-h$ and $|\mathcal{T}|=(1-\epsilon)n$. The second equality is rearrangement. The last inequality uses that n is sufficiently large compared to r_0,k,h,ϵ (more concretely, it follows from $n\geq 100^{r_0}$, and the definitions of r_0,k,h,ϵ). We claim that the following is true:

P: There is some $r\leq r_0$, some small colour c, some colour c element $e\notin E_{\mathcal{S}}(c)$, and some independent set $T\in\mathcal{T}_\ell^{(r)}$ with $T+e$ a rainbow independent set.

There are two cases.

- Suppose that for some $r\leq r_0$ there is some $T\in\mathcal{T}_\ell^{(r)}$ with $d_{A(\mathcal{T}_\ell^{(r)})}(T)\geq(1-\epsilon)n$. Equivalently $|T|\leq\epsilon n$. Since there are $>\epsilon n$ small colours, there is some small colour c absent from T. By definition of "small colour", the fact that small colours are unchanged going from $\mathcal{R}$ to $\mathcal{S}$, and Claim 3.12, we have $e_{\mathcal{S}}(c)=e_{\mathcal{R}}(c)\leq e_{\mathcal{R}}(c_m)\leq(1-2\epsilon)n$. By the augmentation property one of the $\geq 2\epsilon n$ colour c elements outside $\mathcal{S}$ is independent from T (let $I=E(c)\setminus E(\mathcal{S})$, noting that I is independent, since $E(c)$ is independent by the theorem's assumption. Now I,T are independent sets with $|I|>|T|$ and so there's some $e\in I$ with $T+e$ independent). Let e be such an element, and note that c,r,e,T satisfy P.
- Suppose that for all $r\leq r_0$ we have $d_{A(\mathcal{T}_\ell^{(r)})}(T)<(1-\epsilon)n$ for all $T\in\mathcal{T}_\ell^{(r)}$. Notice that in the ℓ-reduced families $\mathcal{T}_\ell^{(0)},\mathcal{T}_\ell^{(1)},\mathcal{T}_\ell^{(2)},\dots,\mathcal{T}_\ell^{(r_0)}$, no large colours occur (since these families are $\leq\mathcal{T}$ by Observation 2.3). Hence large colours have $d_{A(\mathcal{T}_\ell^{(r)})}(c)=|\mathcal{T}|=(1-\epsilon)n>d_{A(\mathcal{T}_\ell^{(r)})}(T)$ for all $T\in\mathcal{T}_\ell^{(r)},r\leq r_0$. Observation 3.2 applied with $G=A(\mathcal{T}_\ell^{(r)}),y=c$ shows that large colours have $k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c)=0$. Also, all medium colours occur on all independent sets of $\mathcal{T}_\ell^{(r)}$ (from Observation 2.5), which shows that medium colours are isolated vertices in $A(\mathcal{T}_\ell^{(r)})$, and so have $k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c)=0$. Thus the only colours of positive k-excess in any $A(\mathcal{T}_\ell^{(r)})$ are small.

Apply Lemma 3.8 to the reduced families $\mathcal{T}_\ell^{(r)}$ for $r = 0, 1, \ldots, r_0 - 1$ and for every colour whose k-excess is positive in $A(\mathcal{T}_\ell^{(r)})$. We claim that for at least one of these applications case (i) of Lemma 3.8 has to occur. Indeed, otherwise $e_{\mathcal{T}_\ell^{(r+1)}}(c) \le e_{\mathcal{T}_\ell^{(r)}}(c) - \frac{1}{2}k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c) + \ell n/k$ for all colours c and all $r = 0, 1, \ldots, r_0 - 1$ (for positive excess colours this will be from Lemma 3.8. For zero excess colours it happens because Observation 2.3 tells us that reducing a family always decreases the number of elements of each colour it contains). Summing over all colours this would give that for each $r = 0, \ldots, r_0 - 1$,

$$\begin{aligned} e(\mathcal{T}_\ell^{(r+1)}) &= \sum_{c\in C(M)} e_{\mathcal{T}_\ell^{(r+1)}}(c) \le \sum_{c\in C(M)} (e_{\mathcal{T}_\ell^{(r)}}(c) - \frac{1}{2}k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c) + \ell n/k) \\ &= e(\mathcal{T}_\ell^{(r)}) + \ell n^2/k - \sum_{c\in C(M)} \frac{1}{2}k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c) \\ &\le e(\mathcal{T}_\ell^{(r)}) + \ell n^2/k - \epsilon^2 n^2/4 \\ &= e(\mathcal{T}_\ell^{(r)}) - \epsilon^2 n^2/8 \end{aligned}$$

Here the second inequality is (3.4), while the last equation comes from "$k = 8\epsilon^{-2}\ell$". Adding these inequalities for $r = 0, 1, \ldots, r_0 - 1$ gives a contradiction:

$$\begin{aligned} e(\mathcal{T}_\ell^{(r_0)}) &\le e(\mathcal{T}_\ell^{(0)}) - r_0(\epsilon^2 n^2/8) \\ &= e(\mathcal{T}) - 100n^2/8 \le e(M') - 100n^2/8 = n(n+h) - 100n^2/8 < 0. \end{aligned}$$

Thus case (i) of Lemma 3.8 had to occur at some application, i.e., there is some $r \le r_0 - 1$, some colour c of positive k-excess in $A(\mathcal{T}_\ell^{(r)})$, and some independent set $T \in \mathcal{T}_\ell^{(r)}$ with at least $d_{A(\mathcal{T}_\ell^{(r)})}(c) + \frac{1}{2}k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c)$ colour c elements $e \notin E(\mathcal{T}_\ell^{(r)})$ with $T + e$ a rainbow independent set. Let $F \subseteq E(c) \setminus E(\mathcal{T}_\ell^{(r)})$ be the set of these elements, noting that $|F| \ge d_{A(\mathcal{T}_\ell^{(r)})}(c) + \frac{1}{2}k\text{-EX}_{A(\mathcal{T}_\ell^{(r)})}(c) > d_{A(\mathcal{T}_\ell^{(r)})}(c)$. Since medium/large colours have zero k-excess in $A(\mathcal{T}_\ell^{(r)})$, c is small. By the definition of the availability graph, we have

$$d_{A(\mathcal{T}_\ell^{(r)})}(c) = (1-\epsilon)n - |E_{\mathcal{T}_\ell^{(r)}}(c)| \ge |E_{\mathcal{S}}(c)| - |E_{\mathcal{T}_\ell^{(r)}}(c)| = |E_{\mathcal{S}\setminus\mathcal{T}_\ell^{(r)}}(c)|$$

Here the inequality uses that $|E_{\mathcal{S}}(c)| \le (1-\epsilon)n$ which happens because there are $(1-\epsilon)n$ independent sets in $\mathcal{S}$. The last equation uses that $\mathcal{T}_\ell^{(r)} \le \mathcal{S}$ which comes from $\mathcal{T} \le \mathcal{S}$ and Observation 2.3. Since $|F| > d_{A(\mathcal{T}_\ell^{(r)})} \ge |E_{\mathcal{S}\setminus\mathcal{T}_\ell^{(r)}}(c)|$, there must be at least one $e \in F \setminus E_{\mathcal{S}\setminus\mathcal{T}_\ell^{(r)}}(c)$. Since F is disjoint from $E(\mathcal{T}_\ell^{(r)})$, this tells us that $e \notin E(\mathcal{S})$. Thus we have found a colour c element $e \notin E(\mathcal{S})$ with $T + e$ a rainbow independent set as required by P.

Now, having established that P is true, let r, c, e, T be as in P. Let $S \in \mathcal{S}$ with $T \subseteq S$. By Lemma 3.10, there is a family of disjoint rainbow independent sets $\mathcal{S}^*$ with $E(\mathcal{S}^*) = \{e\} \cup (E(\mathcal{S}) \setminus X)$, for some $X \subseteq E(\mathcal{S}) \setminus E(\mathcal{T})$ with $|X| \le 2^{r+1}$. Since dummy elements $d_{i,j} \in D$ all have medium colours and e is small (by P), we have

$e \notin D$. Similarly $X \cap D = \emptyset$ because $X \subseteq E(\mathcal{S}) \setminus E(\mathcal{T})$ and everything in $E(\mathcal{S}) \setminus E(\mathcal{T})$ is large. Let $\mathcal{R}^*$ be $\mathcal{S}^*$ with all dummy elements $d_{i,j} \in D$ deleted, noting that we have $E(\mathcal{R}^*) = E(\mathcal{S}^*) \setminus D = (\{e\} \cup (E(\mathcal{S}) \setminus X)) \setminus D = \{e\} \cup ((E(\mathcal{S}) \setminus D) \setminus X) = \{e\} \cup (E(\mathcal{R}) \setminus X)$ (for the last equation, using that $\mathcal{S}$ was constructed from $\mathcal{R}$ by adding dummy elements). Note that since $e \notin E(\mathcal{S})$ and $\mathcal{R} \leq \mathcal{S}$, we have $e \notin E(\mathcal{R})$. Since e is small and X contains only large elements, $\mathcal{R}^*$ satisfies the claim. □

□

4. Concluding remarks

One apparent strengthening of the results in this paper is to just ask for the colours to be size n independent sets (as opposed to bases). This follows by applying Theorem 1.2 to a suitably defined matroid.

Corollary 4.1. Let $B_1, \dots, B_n$ be disjoint independent sets of size n in a matroid M (of any rank). Then there are $n - o(n)$ disjoint rainbow independent sets in $B_1 \cup \dots \cup B_n$ of size $n - o(n)$.

Proof. Let M' be the family of independent sets of size $\leq n$ in M. Notice that M' is also a matroid (for any two independent sets $I, I' \in M'$ with $|I| > |I'|$, we have $I, I' \in M$ and so there is some $x \in I \setminus I'$ with $I' \cup \{x\}$ independent in M. But using $I \in M'$, we have $|I' \cup \{x\}| = |I'| + 1 \leq |I| \leq n$ and so $I' \cup \{x\} \in M'$ also), and that M' has rank n. Now $B_1, \dots, B_n$ are bases in M' and the result follows from Theorem 1.2. □

It is easy to work out the bounds our proof gives: it produces $n - \frac{Cn}{\sqrt{\log n}}$ disjoint rainbow independent sets of size $n - \frac{Cn}{\sqrt{\log n}}$ (for some fixed large constant C). It would be interesting to improve this. Additionally, it would be nice to prove qualitatively stronger asymptotic versions of the conjecture. The following problems are natural goals.

Problem 4.2. Let $B_1, \dots, B_n$ be disjoint bases in a rank n matroid M. Show that there are $(1 - o(1))n$ disjoint rainbow bases.

Problem 4.3. Let $B_1, \dots, B_n$ be disjoint bases in a rank n matroid M. Show that $B_1 \cup \dots \cup B_n$ can be decomposed into $(1 + o(1))n$ disjoint rainbow independent sets.

A solution to any of the above problems, would give a strengthening of Theorem 1.2. Theorem 1.2 may be a good starting point for solving the above problems. It is not uncommon in combinatorics for non-trivial reductions between different kinds of asymptotic results to exist. Indeed in an earlier version of this paper the following problem was also posed "let $B_1, \dots, B_n$ be disjoint bases in a rank n matroid M. Show that there are n disjoint rainbow independent sets of size $(1 - o(1))n$". Kwan [12] solved this problem by finding a nice reduction to Theorem 1.2.

The results in this paper may eventually lead to a solution of Rota's Conjecture for sufficiently large n via *the absorption method*. Absorption is a technique for turning asymptotic results into exact ones. It has recently has found success in rainbow problems related to Rota's Conjecture [9]. Now that we have an asymptotic solution to the conjecture in Theorem 1.2, it seems promising to try and turn it into a exact solution using absorption.

Acknowledgment. The author would like to thank Matija Bucic, Matthew Kwan, and Benny Sudakov for discussions relating to this paper, particularly related to Lemma 3.4. Additionally, he would like to thank Matthew Kwan and an anonymous referee for reading drafts of the paper suggesting many improvements.

References

1. Matija Bucić, Matthew Kwan, Alexey Pokrovskiy, and Benny Sudakov, *Halfway to Rota's basis conjecture*, International Mathematics Research Notices (2018).
2. M Cheung, *Computational proof of Rota's basis conjecture for matroids of rank 4*, Unpublished manuscript. http://educ. jmu. edu/duceyje/undergrad/2012/mike. pdf (2012).
3. T. Chow, *Rota's basis conjecture: Polymath 12*, https://asone.ai/polymath/index.php?title=Rota%27s_conjecture (2017).
4. Sally Dong and Jim Geelen, *Improved bounds for Rota's Basis Conjecture*, Combinatorica **39** (2019), no. 2, 265–272.
5. Arthur A Drisko, *On the number of even and odd Latin squares of order $p+1$*, advances in mathematics **128** (1997), no. 1, 20–35.
6. Benjamin Friedman and Sean McGuinness, *Girth conditions and rota's basis conjecture*, Graphs and Combinatorics **39** (2023), no. 3, 51.
7. Jim Geelen and Peter J Humphries, *Rota's basis conjecture for paving matroids*, SIAM Journal on Discrete Mathematics **20** (2006), no. 4, 1042–1045.
8. Jim Geelen and Kerri Webb, *On Rota's basis conjecture*, SIAM Journal on Discrete Mathematics **21** (2007), no. 3, 802–804.
9. Stefan Glock, Daniela Kühn, Richard Montgomery, and Deryk Osthus, *Decompositions into isomorphic rainbow spanning trees*, Journal of Combinatorial Theory, Series B (2020).
10. David G Glynn, *The conjectures of Alon–Tarsi and Rota in dimension prime minus one*, SIAM Journal on Discrete Mathematics **24** (2010), no. 2, 394–399.
11. Rosa Huang and Gian-Carlo Rota, *On the relations of various conjectures on Latin squares and straightening coefficients*, Discrete Mathematics **128** (1994), no. 1-3, 225–236.
12. Matthew Kwan, private communication (2020).
13. D. H. J. Polymath, *Rota's basis conjecture online for matroids*, unpublished manuscript, https://www.overleaf.com/8773999gccdbdmfdgkm (2017).
14. Marcel Wild, *On Rota's problem about n bases in a rank n matroid*, Advances in Mathematics **108** (1994), no. 2, 336–345.

Department of Mathematics, University College London
Email address: dr.alexey.pokrovskiy@gmail.com