ON GENERALIZATIONS OF SEPARATING AND SPLITTING FAMILIES

DANIEL CONDON, SAMUEL COSKEY, LUKE SERAFIN, AND CODY STOCKDALE

ABSTRACT. The work in this article concerns two different types of families of finite sets: separating families and splitting families. These families have applications in combinatorial search, coding theory, and cryptography as well as many other fields. We define generalizations of these two notions, naming them \( n \)-separating families and \( n \)-splitting families. We discuss the basic properties of these notions and compare them with existing notions of separation. We also give lower and upper bounds on the minimal sizes of \( n \)-separating families. For \( n \)-splitting families this turns out to be much harder; we provide partial results and outline open questions.

§1. INTRODUCTION

Separating families, also called separating systems, play a major role in several areas of applied combinatorics. Before discussing this motivation, let us record the definition: If \( X \) is a finite set and \( A, B \subseteq X \), we say that \( A \) separates \( B \) if both \( A \cap B \neq \emptyset \) and \( A^c \cap B \neq \emptyset \). We say that \( F \subseteq \mathcal{P}(X) \) is a separating family if for all \( B \subseteq X \) there exists \( A \in F \) such that \( A \) separates \( B \). In fact, it is equivalent to consider separating only pairs \( b \subseteq X \), as we prove in section 2.

Separating families were first studied in [Ren61] in connection with probabilistic questions about boolean algebras. Separating families have found applications in many areas, including combinatorial search, boolean logic, and coding theory. There are numerous similar notions of families that also have significant applications. For most efficient use in algorithms and applications to extremal questions, families of minimum size are desirable, so work has been done on bounding the minimum sizes of various types of separating and splitting families.

Numerous generalizations of the notion of a separating family have been studied in a variety of contexts. One of the most important is the following: \( F \) is an \((i, j)\)-separating family if for all \( P, Q \subseteq X \) such that \(|P| \leq i \) and \(|Q| \leq j \), there exists \( A \in F \) such that \( P \subseteq A \) and \( Q \cap A = \emptyset \) or vice versa. Thus ordinary separating is equivalent to \((1, 1)\)-separating. Applications of \((i, j)\)-separating families arise in automata theory, for instance see [Har65]. Bounds of the minimum size of \((i, j)\)-separating families given by [PK84] are used to produce bounds on the minimum sizes of new kinds of families introduced in this paper.

We also consider splitting families, which are closely related to separating families. Let us record the definition of a splitting family. If \( X \) is a finite set and \( A, B \subseteq X \), we say \( A \) splits \( B \)
if \(|A \cap B| = \lfloor |B|/2 \rfloor\) or \(|A \cap B| = \lceil |B|/2 \rceil\). We say that \(\mathcal{F} \subseteq \mathcal{P}(X)\) is a splitting family if for all \(B \subseteq X\) there exists \(A \in \mathcal{F}\) such that \(A\) splits \(B\). We allow the rounding to go either way for convenience. Some other authors have more strict rounding rules, for example see [RH12].

Splitting families have a slightly less illustrious history than separating families. They first appeared in Coppersmith’s algorithm which efficiently computes the discrete logarithm in the low Hamming weight case [Sti02]. Coppersmith’s algorithm only requires families that split sets of a specific size. Such families are studied in more detail in [LLyR04] and [DSL+07]. As far as we know, families that split all subsets of \([k]\) have not been previously studied.

In this paper, we define and study generalizations of separating families and splitting families, which we call \(n\)-separating families and \(n\)-splitting families respectively. We provide bounds on their minimum sizes and establish relationships with other notions of separating and splitting. We describe an application of the generalization of separating families to error correcting codes, and we are confident that both generalizations will find other applications similar to those of separating families and splitting families.

If \(X\) is a finite set and \(\mathcal{F} \subseteq \mathcal{P}(X)\), then \(\mathcal{F}\) is an \(n\)-separating family if and only if for all separable collections \(B_1, B_2, \ldots, B_n \subseteq X\), there exists \(A \in \mathcal{F}\) that separates each \(B_i\). We discuss which collections of sets are separable in Section 3.

Similarly, if \(\mathcal{F} \subseteq X\), then \(\mathcal{F}\) is an \(n\)-splitting family if and only if for all splittable collections \(B_1, B_2, \ldots, B_n \subseteq X\), there exists \(A \in \mathcal{F}\) that splits each \(B_i\). In each of the previous definitions, we defined \(\mathcal{F}\) to be a collection of subsets of the finite set \(X\). For the rest of this article, we let \(X = [k] = \{1, 2, \ldots, k\}\) for convenience. We discuss partial results on which collections of sets are splittable in Section 4 with a complete characterization in the case \(n \leq 3\).

We now briefly outline the conclusions proved in this paper. After discussing which collections of sets are separable, we define \(n\)-separating families and establish the relationship between \(n\)-separating families and the well-studied notion of \((i, j)\)-separating families. We then establish the following lower and upper bounds on the minimal size of an \(n\)-separating family.

**Theorem.** The minimal size of an \(n\)-separating family is \(\Omega(2^n \log k)\) and \(O(n2^n \log k)\).

Splitting families turn out to be more challenging to handle than separating families. We give a characterization of splittability only for \(n \leq 3\). We provide a new lower bound on the size of a splitting family in Section 4. We also establish an upper bound on the minimal size of a 2-splitting family.

**Theorem.** The minimal size of a 2-splitting family is \(\Omega(k)\) and \(O(k^2)\).

Unfortunately the proof of the above theorem does not immediately generalize to the case of \(n\)-splitting. But if the key result Theorem 4.10 can be generalized to this case, then we would obtain the following.
Conjecture. The minimal size of an $n$-splitting family is $O\left(f(n)k^{n/2+1}\right)$.

Acknowledgement: This article represents a part of the work done in the Boise State University 2014 math REU program. The work was supported by NSF grant DMS 1359425 and the Boise State University mathematics department.

§2. SEPARATING FAMILIES

In this section we discuss separating families on a finite set $[k]$. We present previously known facts about separating families and give the minimum size of a separating family with an example. We then state a linear time algorithm for determining whether a collection of subsets of $[k]$ is in fact a separating family. A classification of separating families is also given and we conclude the section with a method of determining the number of inequivalent separating families of size $m$ on $[k]$.

Definition 2.1. Let $X$ be any set. A family $F \subseteq \mathcal{P}(X)$ is a separating family on $X$ if and only if for every $B \subseteq X$ with $|B| \geq 2$ there exists $A \in F$ such that both $A \cap B \neq \emptyset$ and $A^c \cap B \neq \emptyset$.

Here we say $A$ separates $B$. We show that it is in fact enough to only consider separating pairs.

Proposition 2.2. A family $F \subseteq \mathcal{P}(X)$ is a separating family if and only if for every $b \subseteq X$ with $|b| = 2$ there exists $A \in F$ such that $|A \cap b| = 1$.

Proof. Clearly if $F$ separates all sets of cardinality at least two then it separates pairs. Conversely, suppose $F$ separates pairs, and consider $B \subseteq X$ with $|B| \geq 2$. Choose $x, y \in B$ with $x \neq y$; then $\{x, y\}$ is a pair so there exists $A \in F$ such that both $|A \cap B| = 1$ and $|A^c \cap B| = 1$, and hence $A \cap B \neq \emptyset$ and $A^c \cap B \neq \emptyset$ so $B$ is separated by $A \in F$. \hfill \Box

Separating families are most efficiently used in applications if they are of minimum size. Many of the results in this paper deal with giving bounds on the minimum sizes of various types of families. The minimum size of an ordinary separating family is well-known and typically attributed to Rényi.

Theorem 2.3. If $F$ is a separating family over $[k]$ of minimal size, then $|F| = \lceil \log k \rceil$.

Note that while we are mostly concerned with the finite case and the previous theorem is stated in terms of the finite number $k$, it also holds true for any cardinal $\kappa$, with $\lceil \log k \rceil$ replaced by $\min\{\lambda : 2^\lambda \geq \kappa\}$. The method of proof for theorem 2.3 relies on the matrix representation of separating families which we now describe.

Definition 2.4. Let $F = \{A_1, \ldots, A_m\}$ be a separating family of subsets of $[k]$. Then the matrix representation of $F$ is the matrix $M \in \mathcal{M}_{m,k}$ that satisfies $M_{ij} = 1$ if and only if $j \in A_i$. 
In other words, $M$ represents $F$ if and only if its rows are precisely the characteristic vectors of the elements of $F$. Since $F$ is unordered, the matrix representation is well-defined only up to permutations of its rows.

We are now in position to prove the following lemma.

**Lemma 2.5.** A family $F \subseteq \mathcal{P}[k]$ is separating if and only if its corresponding matrix $M$ has distinct columns.

**Proof.** Suppose columns $i$ and $j$ of $M$ are equal. Then for every $A \in F$, either $\{i, j\} \subseteq A$ or $\{i, j\} \cap A = \emptyset$, so $F$ does not separate $\{i, j\}$.

Conversely, suppose $M$ has distinct columns. Then for every pair $\{i, j\}$ the columns $i, j$ disagree at some row, say row $n$, in which case the set corresponding to row $n$ separates $\{i, j\}$. □

**Proof of Theorem 2.3.** Since there are $2^{\lceil \log k \rceil}$ many binary strings of length $\lceil \log k \rceil$, in order for the matrix $M$ corresponding to $F$ to have distinct columns it must contain at least $\lceil \log k \rceil$ many rows. Furthermore, this suffices since we may simply choose $k$ many distinct columns of length $\lceil \log k \rceil$ and put them into a matrix $M$ which represents a separating family over $[k]$ of size $\lceil \log k \rceil$. □

**Example 2.6.** $F = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 5, 7\}\}$ is a separating family on $[8]$ with $m = 3$. This separating family has the following matrix representation where a block with a • corresponds to a 1 and an empty block corresponds to a 0.

![Matrix Representation](image)

**Theorem 2.7.** There is an algorithm to recognize separating families that runs in linear time in the size of the matrix representing the family.

**Proof.** Given $M \in M_{m,k}$, we first sort the columns of $M$ using radix sort. We then check whether any pair of adjacent columns are equal. If any adjacent pair of columns are identical, then the family is not splitting. Both the sort and the comparisons require $O(mk)$ many bit comparisons. □

We now describe a classification of separating families.

**Definition 2.8.** Two separating families $F$ and $G$ are said to be equivalent (written $F \equiv G$) if and only if $G$ can be obtained from $F$ by means of the following operations:

- replace an element of $F$ with its complement;
- permute the elements of $[k]$.  

Since both of these operations are invertible, they generate a natural group acting on the collection of separating families over $[k]$ of cardinality $m$. This group is a natural quotient of the following group acting on $m \times k$ matrices:

**Definition 2.9.** The group $\text{CCR}_{m,k}$, or simply CCR if $m$ and $k$ are understood, is the permutation group acting on $\mathcal{M}_{m,k}$ and generated by the following operations:

- Complementation: replace any row $\vec{r}$ with $\vec{1} - \vec{r}$ (corresponds to taking the complement of an element of $\mathcal{F}$; here $\vec{1}$ denotes to the vector of 1's);
- Column permutation (corresponds to permuting the elements of $[k]$); and
- Row permutation (corresponds to reordering the elements of $\mathcal{F}$).

Since separating families are unordered, the CCR group modulo row permutations acts on separating families in a natural way, and two separating families are equivalent if and only if they lie in the same orbit of this action.

To classify separating families up to equivalence, it turns out to be very useful to represent such families as subsets of a Hamming cube. We denote the $m$-dimensional Hamming cube by $Q_m$, and its symmetry group by $H_m$.

**Definition 2.10 (Cube representation).** Let $M$ be an $m \times k$ matrix which represents a separating family. Then the cube representation of $M$ is the subset $C_M$ of the Hamming cube consisting of those vectors which are the columns of $M$. That is, $C_M$ is the set \{Me_1, \ldots, Me_k\}.

This cube representation is useful because two matrices $M_1, M_2$ represent equivalent separating families if and only if there exists a symmetry of the Hamming cube bearing the cube representation of $M_1$ to the cube representation of $M_2$, as we shall presently prove.

Let $\text{Col}_{m,k}$ denote the column permutation subgroup of $\text{CCR}_{m,k}$. For $M_1, M_2 \in \mathcal{M}_{m,k}$ say $M_1 \sim_{\text{Col}} M_2$ if and only if there exists $\pi \in \text{Col}_{m,k}$ such that $\pi M_1 = M_2$. Note that elements of $\mathcal{M}_{m,k} / \sim_{\text{Col}}$ are in natural bijective correspondence with $[\mathcal{F}_2^m]^k$. The natural action of $\text{CCR}_{m,k}$ on $\mathcal{M}_{m,k}$ induces an action of $\text{CCR}_{m,k} / \text{Col}_{m,k}$ on $\mathcal{M}_{m,k} / \sim_{\text{Col}}$.

Also, the cube group $H_m$ acts on $[Q_m]^k$ by means of the natural action of $H_m$ on $Q_m$ applied to each element of a set.

**Theorem 2.11.** Assume $2^m \leq k$. The natural action of $\text{CCR}_{m,k} / \text{Col}_{m,k}$ on $\mathcal{M}_{m,k} / \sim_{\text{Col}}$ is equivariant with the natural action of the cube group $H_m$ on $[Q_m]^k$.

**Proof.** We define an equivariance $(\phi, f)$, where $\phi : \text{CCR}_{m,k} / \text{Col}_{m,k} \to H_m$ and $f : \mathcal{M}_{m,k} / \sim_{\text{Col}} \to [Q_m]^k$. Let $G = \text{CCR}_{m,k} / \text{Col}_{m,k}$ and $M = \mathcal{M}_{m,k}$. For $\pi \in \text{CCR}_{m,k}$ let $\overline{\pi}$ denote its image in $G$ under projection. Note that $G$ is generated by the following elements:

- $\overline{\sigma_{ij}}$, where $\sigma_{ij}$ interchanges rows $i$ and $j$.
- $\overline{T_{ij}}$, where $T_{ij}$ composites row $i$. 

...
We define $\phi$ on this set of generators. Note that a rotation of $Q_m$ is uniquely determined by a permutation of the coordinates of its vertices, so $H_m$ is generated by rotations which interchange two coordinates and reflections across coordinate hyperplanes; let $r_{ij}$ be the rotation interchanging coordinates $i$ and $j$, and $c_i$ be the reflection across the hyperplane normal to the $i$th coordinate axis. For $1 \leq i, j \leq m$, let $\phi(T_{ij}) = r_{ij}$ and $\phi(T_i) = c_i$; clearly $\phi: G \to H_m$ is an isomorphism.

Now we define the bijection $f$. Given $M \in \mathcal{M}$, let $f(M)$ be the subset of the vertices of $Q_m$ which are columns of some element of $M$. Clearly $f: \mathcal{M} \to [Q_m]^k$ is a bijection.

It remains to verify that $(\phi, f)$ is an equivariance. Let $M \in \mathcal{M}$ and $\pi \in G$. It is easy to show that $f(\pi x) = \phi(\pi)f(x)$, which completes the proof. □

**Corollary 2.12.** Let $\mathcal{F}, \mathcal{G}$ be separating families over $[k]$ of cardinality $m$. Then $\mathcal{F} \equiv \mathcal{G}$ if and only if their cube representations lie in the same orbit of $H_m$ acting on $[Q_m]^k$ in the natural way.

**Proof.** Represent $\mathcal{F}, \mathcal{G}$ as matrices $M_1, M_2$, and represent these matrices (modulo column rearrangement) as subsets $S_1, S_2$ of $Q_m$. Since $\text{CCR}_{m,k}$ includes column permutations, $M_1, M_2$ are CCR-equivalent if and only if $S_1, S_2$ lie in the same orbit of the action of $H_m$ on $[Q_m]^k$. □

**Corollary 2.13.** Let $\text{sep}(m, k)$ be the number of non-equivalent separating families over $[k]$ of cardinality $m$. Then $\text{sep}(m, k) = \text{sep}(m, 2^m - k)$.

**Proof.** Let $\mathscr{F}, \mathscr{G}$ be complete sets of distinct equivalence class representatives for the collections of separating families over $[k]$ with cardinalities $m$ and $2^m - k$, respectively, and $\mathcal{M}, \mathcal{N}$ be the corresponding subsets of $Q_m$. Then the map $A \mapsto A^c$ is a bijection between $\mathcal{M}$ and $\mathcal{N}$, so $|\mathscr{F}| = |\mathscr{G}|$. □

The number of distinct subsets of the $m$-cube up to cube symmetry, and hence the values of $\text{sep}(m, k)$, can be computed using Pólya theory. For two such calculations, see [Che93] and [HH68]. Unfortunately the formulas these articles produce each have an exponential number of terms. This leads us to conjecture that the following question has an affirmative answer.

**Question 2.14.** Is the computation of the values of $\text{sep}(m, k)$ NP-hard?

For more on the values of $\text{sep}(m, k)$, see the OEIS sequence A039754, and the monograph [Har65].

§3. $n$-SEPARATING FAMILIES

In this section we define the concept of $n$-separating families and discuss separability. We also investigate an application of $n$-separating families to error-correcting codes. We then define $(i, j)$-separating families and describe their relationships with $n$-separating families. As a consequence of this relationship with $(i, j)$-separating, we obtain a lower bound on the minimum
size of an \( n \)-separating family by applying a result from [FK84]. We then describe an application of \((n, 1)\)-disjunct families to combinatorial search and obtain a lower bound on the size of an \((n, 1)\)-separating family. We conclude this section by providing a probabilistic upper bound on the minimum size of an \( n \)-separating family.

It is first necessary to characterize separable collections of sets. Indeed, there exist collections that cannot simultaneously be separated by a single set. For example, \( \{1, 2\}, \{2, 3\}, \{3, 1\} \) cannot be simultaneously separated by any set.

**Definition 3.1.** A collection \( B_1, \ldots, B_n \) of subsets of \([k]\) is separable if and only if \( |B_i| \geq 2 \) for all \( i \in [n] \) and there exists \( A \subseteq [k] \) such that both \( A \cap B_i \neq \emptyset \) and \( A^c \cap B_i \neq \emptyset \) for each \( i \in [n] \).

**Definition 3.2.** A family \( F \subseteq \mathcal{P}[k] \) is \( n \)-separating if for every separable collection \( B_1, \ldots, B_n \subseteq [k] \), there exists \( A \in F \) such that both \( A \cap B_i \neq \emptyset \) and \( A^c \cap B_i \neq \emptyset \) for each \( i \in [n] \).

**Proposition 3.3.** A collection \( B_1, \ldots, B_n \) of subsets of \([k]\) is separable if and only if there exist pairs \( b_1 \subseteq B_1, \ldots, b_n \subseteq B_n \) such that the graph with edge set \( \{b_1, \ldots, b_n\} \) is bipartite.

**Proof.** (\( \Leftarrow \)) Suppose \( b_1 \subseteq B_1, \ldots, b_n \subseteq B_n \) are such that \( G = \{b_1, \ldots, b_n\} \) is the edge set of a bipartite graph. Fix a 2-colouring of the vertices of \( G \), \( f: V(G) \to \{0, 1\} \); then the set \( f^{-1}[0] \) separates the collection \( b_1, \ldots, b_n \), and hence also the collection \( B_1, \ldots, B_n \).

(\( \Rightarrow \)) Suppose a set \( A \) simultaneously separates \( B_1, \ldots, B_n \). For each \( i \in [n] \), let \( \alpha_i \in A \cap B_i \) and \( \beta_i \in A^c \cap B_i \), and define \( b_i = \{\alpha_i, \beta_i\} \) and \( G = \{b_1, \ldots, b_n\} \). Then the function \( f: V(G) \to \{0, 1\} \) given by \( f(\alpha_i) = 0 \) and \( f(\beta_i) = 1 \) for \( i \in [n] \) is a 2-colouring of \( G \), thereby constructing the desired graph. \( \square \)

**Remark 3.4.** It is worth mentioning that a collection \( B_1, \ldots, B_n \) is separable if and only if, when the collection is viewed as a hypergraph, it is 2-colorable. The problem of recognizing hypergraph 2-colorability is known to be NP-complete; see [Lov73].

We now provide a constructive upper bound for the minimum size of a \( 2 \)-separating family. This constructive upper bound is not optimal, as it is bested by the bound given in theorem 3.12. However, this bound is still noteworthy because it introduces an easily computable construction of a \( 2 \)-separating family given an ordinary separating family.

**Theorem 3.5.** We can construct a \( 2 \)-separating family of subsets of \( k \) whose size is \( O((\log k)^2) \).

This follows immediately from the following construction.

**Theorem 3.6.** Let \( F \) be a \( 1 \)-separating family on \([k]\), and \( F' = \{A \triangle B: A, B \in F\} \). Then \( F \cup F' \) is a \( 2 \)-separating family.

**Proof.** Let \( b_1, b_2 \) be two pairs in \([k]\) and \( A_1, A_2 \) be their respective separators in \( F \). Often, \( A_1 \) or \( A_2 \) will separate both pairs. Assuming neither does, we have that \( A_1 \) contains precisely one element
in $b_1$ and $A_2$ contains both or zero elements in $b_1$. Then $A_1 \triangle A_2$ contains precisely one element in $b_1$. By identical reasoning, $A_1 \triangle A_2$ contains precisely one element in $b_2$, and $A_1 \triangle A_2$ separates both $b_1$ and $b_2$ simultaneously.

Example 3.7. $\mathcal{F} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 5, 7\}, \{3, 4, 5, 6\}, \{2, 4, 5, 7\}, \{2, 3, 6, 7\}\}$ is a 2-separating family on $[8]$. This is the 2-separating family obtained from the previously described construction with the separating family given in example 2.6. This separating family has the following matrix representation.

\[
\begin{array}{cccc}
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\end{array}
\]

It is now a natural time to describe an application of $n$-separating families to error-correcting codes. They will be applied to produce Hamming $d$-codes. A Hamming $d$-code is an error-correcting code that can detect up to $d - 1$ errors and correct up to $\lfloor d/2 \rfloor - 1$ errors in a message. It may be represented as an $m \times k$ binary matrix whose columns have pairwise Hamming distances of at least $d$. As we now prove, the matrix representation of an $n$-separating family is an example of a Hamming $2^{n-1}$-code.

Theorem 3.8. If $F$ is $n$-separating, then the Hamming distance between any two columns in $F$ is at least $2^{n-1}$.

The proof will be given in a series of lemmata.

We first note that the matrix representation of $\mathcal{P}[k]$ contains each subset of $[k]$ as a row. It has $k$ columns and $2^k$ rows. Starting with a base case $\mathcal{P}[1] = (0, 1)^T$, we can construct $\mathcal{P}[k]$ inductively by starting with a $(k - 1) \times 2^k$ matrix with the rows of $\mathcal{P}[k - 1]$ repeated twice; that is, the first $2^{k-1}$ rows and the last $2^{k-1}$ rows are both $\mathcal{P}[k - 1]$. We can then add a $k^{th}$ column, such that the first half of its entries are 0’s and the last half are 1’s. The resulting matrix is a representation of $\mathcal{P}[k]$, which we will denote $M_k$.

Lemma 3.9. The columns of $M_k$ have mutual Hamming distances exactly $d = 2^{k-1}$.

Proof. We proceed inductively, using the matrix construction described above. The two columns of $\mathcal{P}[2]$ have Hamming distance 2, so the base case holds. Then, hypothesize that the columns of $M_{k-1}$ have $d = 2^{k-2}$. Consider the $i^{th}$ and $j^{th}$ columns of $M_k$ with $i, j < k$. Their Hamming distance is twice that of the $i^{th}$ and $j^{th}$ columns of $M_{k-1}$ so by the inductive hypothesis $d = 2^{k-1}$. Consider the Hamming distance between the $i^{th}$ column and the $k^{th}$ column, where $i^{th}$ column of
$M_{k-1}$ had Hamming weight $x$. Then the $i^{th}$ and $k^{th}$ columns disagree on $x$ of the first $2^{k-1}$ rows, and on $2^{k-1} - x$ of the second $2^{k-1}$ rows, for a total of $d = x + 2^{k-1} - x = 2^{k-1}$ disagreements. \[\square \]

**Lemma 3.10.** If $F \subseteq \mathcal{P}[k]$ and for all $A \in \mathcal{P}[k]$, exactly one of $A$ or $A^c \in F$, then the Hamming distance between any two columns in the matrix representation of $F$ is $2^{k-2}$.

**Proof.** Fix $i, j$ distinct columns of the matrix representation of $\mathcal{P}[k]$. Consider removing any row $\ell$ from the matrix. If columns $i, j$ agree at position $\ell$, then their Hamming distance is not changed. If columns $i, j$ disagree at position $\ell$, then their Hamming distance is reduced by 1. Since $A, A^c$ are distinct for every $A \subseteq [k]$, $F$ has $2^{k-1}$ rows. In exactly half of these rows, columns $i, j$ will disagree. So the Hamming distance between these two columns is $2^{k-1} - \frac{1}{4}2^k = 2^{k-2}$. \[\square \]

**Lemma 3.11.** If $F$ is an $n$-separating family, then any set of $n + 1$ columns in $F$ will contain $\mathcal{P}[n + 1]$ up to complements.

**Proof.** Consider the following construction on a connected bipartite graph with $n + 1$ vertices and $n$ edges. Color $i$ of the vertices red and $n + 1 - i$ of them blue. Connect all red vertices to one particular blue vertex and all remaining blue vertices to one particular red vertex. We have $n + 1 - i + (i - 1) = n$ edges. Allowing $i$ to range up to $\lfloor \frac{n}{2} \rfloor$ we get the desired result. \[\square \]

**Proof of Theorem 3.8** Since $F$ is $n$-separating, any collection of $n$ columns from $F$ contains $\mathcal{P}([n + 1])$ up to complements. Thus the Hamming distance between any two columns in $F$ is at least $2^{(n+1)-2} = 2^{n-1}$. \[\square \]

We now provide an upper bound on the minimum size of an $n$-separating family.

**Theorem 3.12.** If $F$ is an $n$-separating family, then $|F| \leq \frac{2n \log k}{\log(1 - 2^{-n})}$. In particular, the minimal size of an $n$-separating family of subsets of $k$ is $O(2^n n \log k)$.

We shall prove this using the probabilistic method. For this, we make use of the following lemma, which gives an upper bound on the number of objects needed to complete $N$ tasks, provided that for each task, a randomly chosen object completes it with probability at least $p$.

**Lemma 3.13.** Suppose there is a set of $N$ tasks to be completed, and that for each task, a randomly chosen object completes it with probability at least $p$. Then the minimal size of a family $F$ of objects which completes all the tasks is

$$|F| \leq \frac{\log N}{-\log(1 - p)}.$$ 

**Proof.** Choose a particular task $\tau$. The probability that a randomly chosen object does not complete $\tau$ is at most $(1 - p)$, so the probability that no object from a collection of $m$ randomly chosen objects completes $\tau$ is at most $(1 - p)^m$. Thus the expected number of tasks left uncompleted by a collection of $m$ objects is $N(1 - p)^m$. We are thus looking for the least $m$ such that $N(1 - p)^m < 1$
(so there is at least one family of \(m\) objects which completes all the tasks). Solving this inequality for \(m\),

\[
N(1 - p)^m < 1
\]

\[
(1 - p)^m < \frac{1}{N}
\]

\[
m \log(1 - p) < - \log N
\]

\[
m < \frac{\log N}{\log(1 - p)},
\]

we obtain the inequality desired (as \(m = |\mathcal{F}|\)). \(\square\)

**Lemma 3.14.** The probability that a randomly chosen subset of \([k]\) simultaneously separates a given separable collection of \(n\) pairs has lower bound \(2^{-n}\).

**Proof.** Let \(\{b_1, \ldots, b_n\}\) be a separable collection of pairs, and \(G\) be the graph with edge set \(\{b_1, \ldots, b_n\}\). Let \(p_n\) be the probability that a random set \(A \subseteq [k]\) simultaneously separates all pairs in \(G\). We proceed by induction on \(n\). The base case \(n = 1\) is clear. For the inductive step, we break into cases.

Case 1: If \(b_{n+1}\) is disjoint from \(G\), then the event that \(b_{n+1}\) is separated is independent of the event that all edges in \(G\) are separated, and the probability that \(b_{n+1}\) is separated is \(\frac{1}{2}\). Thus \(p_{n+1} = \frac{1}{2}p_n \geq 2^{-(n+1)}\).

Case 2: Let \(b_{n+1} = \{x, y\}\) have one vertex, \(x\), in common with \(G\). \(b_{n+1}\) is separated if and only if \(x\) and \(y\) are either not both in \(A\) or both in \(A^c\), giving \(p_{n+1} = \frac{1}{2}p_n \geq 2^{-(n+1)}\).

Case 3: Both vertices of \(b_{n+1}\) are in \(G\), and both vertices of \(b_{n+1}\) lie in a connected component of \(G\). Then \(b_{n+1}\) will automatically be separated, giving \(p_{n+1} = p_n \geq 2^{-(n+1)}\).

Case 4: Both vertices of \(b_{n+1}\) are in \(G\), and \(b_{n+1}\) connects two components of \(G\). We remove an edge \(b_i\) without disconnecting any component of \(b_{n+1} \cup G\); as any acyclic graph has at least two leaves, this is always possible. The probability that \(b_{n+1} \cup (G \setminus b_i)\) is separated is at least \(2^{-n}\) by the inductive hypothesis. Replacing \(b_i\) results in either Case 1, Case 2, or Case 3, and gives us \(p_{n+1} \geq \frac{1}{2}p_n = 2^{-(n+1)}\). \(\square\)

**Remark 3.15.** The worst-case probability \(p_n = 2^{-n}\) is attained for a disjoint set of pairs (because the events that each pair is separated are independent). This worst case is in fact attained when the graph of the pairs forms a forest whose connected components are single edges.

**Proof of Theorem 3.12.** Since it suffices to separate collections of \(n\) pairs for \(n\)-separation, we notice there are \(\binom{k}{2}^n\) collections of pairs to be separated (tasks to be completed). Substituting the minimal probability from Lemma 3.14 into the bound provided in Lemma 3.13, we obtain \(\frac{-\log(\binom{k}{2}^n)}{-\log(1 - 2^{-n})}\) as an upper bound. Since \(\binom{k}{2} \leq k^2\), we realize the desired bound of \(|\mathcal{F}| \leq \frac{2n \log k}{-\log(1 - 2^{-n})}\). \(\square\)
We will now define the previously studied notion of an \((i, j)\)-separating family.

**Definition 3.16.** A family \(F\) of subsets of \([k]\) is \((i, j)\)-separating if and only if for all \(P, Q \subseteq [k]\) with \(|P| \leq i\) and \(|Q| \leq j\), there exists \(A \in F\) such that either \(P \subseteq A\) and \(A \cap Q = \emptyset\) or \(Q \subseteq A\) and \(A \cap P = \emptyset\).

In this section we establish the full set of relationships between the notions of \((i, j)\)-separating and the notions of \(n\)-separating. In particular, we prove the implications depicted in Figure 1 and also that no other implications hold.

The first implication is obvious from the definition.

**Proposition 3.17.** If \(F\) is \((i + 1, j)\)-separating then \(F\) is \((i, j)\)-separating. If \(F\) is \((i, j + 1)\)-separating then \(F\) is \((i, j)\)-separating.

We next establish two key implications between an \((i, j)\)-separating notion and an \(n\)-separating notion.

**Theorem 3.18.** Let \(F\) be a family of subsets of \(k\). If \(F\) is \((n, n)\)-separating then \(F\) is \(n\)-separating.

**Proof.** Consider a separable collection of pairs \(b_1, b_2, \ldots, b_n\). Viewing these pairs as edges of a bipartite graph, note there are at most \(n\) vertices of either color. Since \(F\) is \((n, n)\)-separating, we can find an \(A \in F\) such that \(A\) contains exactly one element from each \(b_i\). \(\square\)

**Theorem 3.19.** If a family \(F\) is \((i + j - 1)\)-separating, then \(F\) is \((i, j)\)-separating.
Proof. Given two sets $P, Q$ with respective sizes $i, j$, we can construct a connected bipartite graph $G$ with vertex classes $P$ and $Q$ using $i + j - 1$ edges. Since $F$ separates $G$, some element of $F$ contains either $P$ or $Q$. To construct the graph, first draw an edge between $\min P$ and each element in $Q$, then draw an edge between every other element of $P$ and any element of $Q$. Note that this graph is bipartite, connected, and contains $i + j - 1$ edges. \qed

Theorem 3.20. No other implications hold between $(i, j)$-separating or $n$-separating notions aside from the (transitive) consequences of Propositions 3.17 and Theorems 3.18 and 3.19.

We prove this theorem in a sequence of lemmata that give counterexamples to any other implications.

Lemma 3.21. There exists an $n$-separating family which is not $(n, n)$-separating.

Proof. Fix distinct $A, B \subseteq [k]$ with $|A| = |B| = n$. Let $F = \bigcup_{i=1}^{n-1} [k]^i \cup ([k]^n \setminus \{A, B\})$. This family is not $(n, n)$-separating because $A, B$ are not separated by anything in the family. To see that $F$ is $n$-separating, consider a collection of $n$ pairs forming the edge set of a bipartite graph $G$ with partitions of sizes $i$ and $j$.

Case 1. Without loss of generality $i < n$. Let $P$ be the partition of $G$ containing $i$ vertices. Then $P \in F$ and $P$ separates all edges of $G$.

Case 2. $i = j = n$. Then the edge set $\{b_1, \ldots, b_n\}$ of $G$ is a collection of $n$ disjoint pairs. Let $\{\beta_1, \ldots, \beta_n\}$ be such that for each $s$, $\beta_s \in b_s$. If for some $\ell$, $b_\ell$ contains an element $x$ not in $A \cup B$, then provided $\beta_\ell = x$, we have that the set $\{\beta_1, \ldots, \beta_n\}$ is an element of $F$ and separates $G$. If there is $\ell$ such that $b_\ell \subseteq A$, then $\{\beta_1, \ldots, \beta_n\}$ necessarily separates $G$, and similarly if there is $\ell$ such that $b_\ell \subseteq B$. The final case is where for every $s$, $b_s$ intersects both $A$ and $B$. In this case let $b_i = \{a, b\}$ where $a \in A$ and note that $(A \setminus \{a\}) \cup \{b\}$ separates $G$. \qed

Lemma 3.22. If $i + j \geq n + 2$, then there exists an $n$-separating family which is not $(i, j)$-separating.

Proof. Fix $A, B \subseteq [k]$ such that $|A| = i$, $|B| = j$, and $A \cap B = \emptyset$. Let $F = \mathcal{P}[k] \setminus \{C \mid A \subseteq C \text{ or } B \subseteq C\}$. Since no set in $F$ contains $A$ or $B$, $A$ and $B$ are not $(i, j)$-separated by any set in $F$. To see $F$ is $n$-separating, let $b_1, b_2, \ldots, b_n$ be a separable collection of pairs and consider the bipartite graph $G$ with edge set $\{b_1, \ldots, b_n\}$, together with a 2-colouring $f : V(G) \to \{0, 1\}$. If neither $A$ nor $B$ is monochromatic, then $b_1, \ldots, b_n$ are separated by some set in $F$. Assume that the set of vertices of one color in $G$ contains $A$. Since $i + j \geq n + 2$, the vertices in $A \cup B$ do not lie in a single connected component of $G$. Let $\Xi$ be a component of $G$ with vertices in $A \cup B$, and define a new colouring $f'$ by

$$f'(v) = \begin{cases} 1 + f(v) \mod 2 & \text{if } v \in \Xi \\ f(v) & \text{otherwise.} \end{cases}$$
Under $f'$, no monochromatic set of vertices from $G$ contains either $A$ or $B$, so since $b_1, \ldots, b_n$ are separated by a monochromatic set with respect to any 2-colouring of $G$, they are separated by an element of $\mathcal{F}$. This completes the proof that $\mathcal{F}$ is $n$-separating.

\textbf{Lemma 3.23.} For $n > 1$, there exists an $(n-1,j)$-separating family which is not $n$-separating.

\textit{Proof.} Let $\mathcal{F} = [k]^{n-1}$. Note that $\mathcal{F}$ contains all sets of size $n-1$ and so is $(n-1,j)$-separating. However, since it contains no sets of cardinality $n$, it cannot separate disjoint collections of $n$ pairs, and so is not $n$-separating. □

\textbf{Lemma 3.24.} There exists an $n$-separating family which is not $(n+1)$-separating.

\textit{Proof.} The family $\mathcal{F} = [k]^n$ is clearly $n$-separating, but cannot split disjoint collections of $n+1$ pairs, so $\mathcal{F}$ is $n$-separating but not $(n+1)$-separating. □

\textbf{Lemma 3.25.} Let $i \leq j$. There exists an $(i,j)$-separating family which is not $(i+1,j)$-separating. And there exists an $(i,j)$-separating family which is not $(i,j+1)$-separating.

\textit{Proof.} Without loss of generality we may suppose that $i \leq j$ in each expression above. Then $[k]^i$ is $(i,j)$-separating but not $(i+1,j)$-separating. Also, $[k]^{k-j}$ is $(i,j)$-separating, but not $(i,j+1)$-separating. □

\textbf{Lemma 3.26.} Let $i < i' \leq j' < j$. Then we have the following:

(1) There exists an $(i,j)$-separating family that is not an $(i',j')$-separating family.

(2) There exists an $(i',j')$-separating family that is not an $(i,j)$-separating family.

\textit{Proof.} (1) Let $\mathcal{F} = [k]^i$. Let $B_i, B_j \subseteq [k]$ such that $|B_i| \leq i$ and $|B_j| \leq j$. There exists $A \in \mathcal{F}$ such that $B_i \subseteq \mathcal{F}$ and $A \cap B_j = \emptyset$, so $\mathcal{F}$ is $(i,j)$-separating. But no set in $\mathcal{F}$ will contain $i'$ (or $j'$) elements, so it is not $(i',j')$-separating.

(2) Let $A \in [k]$ be such that $|A| = i$ and set $\mathcal{F} = ([k]^i \setminus \{A' \mid A \subseteq A'\}) \cup [k \setminus A]^j$. Let $B_{i'}, B_{j'} \subseteq [k]$ such that $|B_{i'}| \leq i'$, $|B_{j'}| \leq j'$, and $B_{i'} \cap B_{j'} = \emptyset$. If $A \subseteq B_{i'}$, then there is $B \in \mathcal{F}$ such that $B_{i'} \subseteq B$ and so $B$ separates $B_i$ and $B_j$. If $A \not\subseteq B_{i'}$, then there is $B \in \mathcal{F}$ such that $B_{j'} \subseteq B$ and so $B$ separates $B_i$ and $B_{j'}$, so $\mathcal{F}$ is $(i',j')$-separating.

Let $B_j$ be such that $|B_j| = j$ and $A \cap B_j = \emptyset$. Since no set in $\mathcal{F}$ contains $A$, any set in $\mathcal{F}$ that would separate $A$ and $B_j$ must contain $B_j$. This is not possible since the sets in $\mathcal{F}$ have cardinality at most $j'$ and $j' < j$. Thus $\mathcal{F}$ is not $(i,j)$-separating.

\textit{Proof of Theorem 3.20} The preceding lemmata address all other possible implications between $(i,j)$-separating families and $n$-separating families, thus the proof is complete. □

Theorem 3.20 leads us to believe that $n$-separating is an interesting concept in its own right. There are two nontrivial implications with previously known concepts and other possible implications do not exist.
We now provide a lower bound on the minimum size of an $n$-separating family.

**Theorem 3.27.** The least size of an $n$-separating family has lower bound $\Omega(2^n \log k)$.

**Proof.** By Theorem 3.18, every $n$-separating family is $(n/2, n/2)$-separating. Therefore we can use the lower bound on $(n/2, n/2)$-separating families from Fredman–Komlós [FK84].

**Remark 3.28.** There is an upper bound provided by Fredman and Komlós [FK84] for $(i, j)$-separating families. Taking this bound for $(n, n)$-separating families provides an upper bound for $n$-separating families by theorem 3.18. However, that bound is bested by our probabilistic upper bound from theorem 3.12.

§4. **$n$-splitting families and splittability**

In this section we introduce a particular type of separating family that is called a splitting family. Similar to the generalization of separating families to $n$-separating families, we generalize splitting families to $n$-splitting families. We then provide an upper bound on the minimum size of a splitting family that we obtain from a generalization of a construction of Coppersmith in [Sti02]. We then define splittability and the $n$-splitting families. We give partial results on the characterization of splittable configurations of sets for $n = 3$. We provide a lower bound for the size of a 2-splitting family and conjectures for bounds on $n$-splitting families.

If $A, B$ are any sets then $A$ splits $B$ if and only if $|A \cap B| = \lfloor |B|/2 \rfloor$ or $\lceil |B|/2 \rceil$. We remark that some authors require $|A \cap B| = \lfloor |B|/2 \rfloor$, for example [RH12]. We prefer our definition since it is technically handy and just as good for most applications.

**Definition 4.1.** A family $\mathcal{F}$ of subsets of $[k]$ is a splitting family if for all $B \subseteq [k]$ there exists $A \in \mathcal{F}$ such that $A$ splits $B$.

**Theorem 4.2.** The minimal size of a splitting family of subsets of $[k]$ has upper bound $\lceil k/2 \rceil$.

We prove this theorem with a generalization of Coppersmith’s construction described in [Sti02]. Our version allows for $k$ or $t$ to be odd.

**Proof.** Let $R$ be the right shift operator on finite binary sequences. Define $A_0 = 1^{[k/2]}0^{[k/2]} (|k/2|)$ 1’s followed by $|k/2|$ 0’s, and $A_n = R^n \ell_0$. Let $US_k = [A_i : 0 \leq i < k/2]$. We claim that $US_k$ is a splitting family over $[k]$.

Fix a set $B \subseteq [k]$. Define $v(i) = |A_i \cap B| - |([k] \setminus A_i) \cap B|$. If $v(0) \in \{-1, 0, 1\}$, then we are done. Suppose otherwise. Note that if $|k/2| \not\in B$, then $v(|k/2|) = -v(0)$ because $A_0$ is the complement of $A_{|k/2|}$ in $[k] \setminus \{ |k/2| \}$. And if $|k/2| \in B$, then $v(|k/2|) = -v(0) + 2$. Indeed, letting $z = |A_{|k/2|} \cap B|$ we see $v(|k/2|) = 2z - |B|$ and $v(0) = |B| - 2z + 2$. 
Next we consider $v(i + 1) - v(i) = (|A_i \cap B| - |A_{i+1} \cap B|) - (|A_i^c \cap B| - |A_{i+1}^c \cap B|)$. Note that $A_i$ and $A_{i+1}$ differ only because $(i + 1) \in A_i$ but $(i + 1) \notin A_{i+1}$, and $(i + 1 + [k/2]) \in A_{i+1}$ but $(i + 1 + [k/2]) \notin A_i$. If $(i + 1), (i + 1 + [k/2]) \in B$ or $(i + 1), (i + 1 + [k/2]) \notin B$, then $v(i + 1) - v(i) = 0$. If $(i + 1) \in B$ and $(i + 1 + [k/2]) \notin B$, then $v(i + 1) - v(i) = 2$. And if $(i + 1) \notin B$ and $(i + 1 + [k/2]) \in B$, then $v(i + 1) - v(i) = -2$.

Since $v(0) \notin \{-1, 0, 1\}$ we have $v(0) \leq -2$ and $v([k/2]) \leq 2$ or $v(0) \geq 2$ and $v([k/2]) \leq 0$. And since $v(i + 1)$ only differs from $v(i)$ by 0 or 2, there exists $i$ such that $v(i) \in \{-1, 0, 1\}$. Thus $A_i$ splits $B$.

We conjecture the upper bound given above is sharp, however at the moment we can only compute a lower bound of $\Omega(\sqrt{k})$. In order to obtain this estimate, we will use the volume method for computing lower bounds. This method is used together with more advanced techniques in [FK84] to obtain their lower bound on the size of $(i, j)$-separating families. The next result describing the volume method has been written in terms of objects and tasks, as was done for $3.13$.

**Lemma 4.3.** Suppose that there is a set of $N$ tasks to be completed, and each object completes at most $v$ of the tasks (the number of tasks completed by an object is called the object’s volume). Then for any family $\mathcal{F}$ of objects which jointly complete all the tasks,

$$|\mathcal{F}| \geq \frac{N}{v}$$

The proof of this lemma is completely trivial: A collection of $m$ many objects completes at most $mv$ many tasks, so a family of objects completing all the tasks must have size at least $\frac{N}{v}$.

We now apply the volume method to compute a lower bound for the size of a splitting family.

**Theorem 4.4.** The minimal size of a splitting family of subsets of $[k]$ is $\Omega(\sqrt{k})$.

**Proof.** For the purpose of asymptotics we may assume that $k$ is even. Then following [FK84] it is the case that the splitters of maximum volume are of size $k/2$. The volume of such a splitter is

$$v = \sum_i \left(\begin{array}{c} k/2 \\ i \end{array}\right) \left(\begin{array}{c} k/2 \\ i \end{array}\right) + \sum_i \left(\begin{array}{c} k/2 \\ i+1 \end{array}\right) \left(\begin{array}{c} k/2 \\ i+1 \end{array}\right) + \sum_i \left(\begin{array}{c} k/2 \\ i+1 \end{array}\right) \left(\begin{array}{c} k/2 \\ i \end{array}\right)$$

Note that the latter two terms are no larger than the first since the identity $2AB \leq A^2 + B^2$ implies

$$2\sum \left(\begin{array}{c} k/2 \\ i \end{array}\right) \left(\begin{array}{c} k/2 \\ i+1 \end{array}\right) \leq \sum \left(\begin{array}{c} k/2 \\ i \end{array}\right)^2 + \sum \left(\begin{array}{c} k/2 \\ i+1 \end{array}\right)^2 = 2\sum \left(\begin{array}{c} k/2 \\ i \end{array}\right)^2$$

It follows that the volume has upper bound

$$v \leq 3\sum \left(\begin{array}{c} k/2 \\ i \end{array}\right)^2$$
Using a well-known binomial coefficient identity together with a Stirling-type approximation, we see that
\[ v \leq 3 \binom{k}{k/2} \sim 2^k / \sqrt{k} \]
Meanwhile, the number of sets to be split is clearly \( N = 2^k \). Thus the volume bound is asymptotic to \( N/v = \Omega(\sqrt{k}) \), as desired. \( \square \)

We next study the generalization of splitting families we have called \( n \)-splitting families. The definition is analogous to that of \( n \)-separating families.

**Definition 4.5.** A collection \( B_1, \ldots, B_n \) of subsets of \([k]\) is **splittable** if there exists \( A \subseteq [k] \) which simultaneously splits all the \( B_i \).

**Definition 4.6.** A family \( F \) of subsets of \([k]\) is **\( n \)-splitting** if for every splittable collection \( B_1, \ldots, B_n \subseteq [k] \) there exists \( A \in F \) which simultaneously splits all the \( B_i \).

We remark that every collection consisting of just two sets \( B_1, B_2 \) is 2-splittable. To see this, we can simply let \( D \subset B_1 \cap B_2 \) be such that \( |D| = \lceil |B_1 \cap B_2|/2 \rceil \), let \( E \subset B_1 \setminus B_2 \) be such that \( |E| = \lfloor |B_1 \setminus B_2|/2 \rfloor \), and let \( F \subset B_2 \setminus B_1 \) be such that \( |F| = \lfloor |B_2 \setminus B_1|/2 \rfloor \). Then it is easy to see that \( A = D \cup E \cup F \) is a splitter for both \( B_1 \) and \( B_2 \).

This fact together with the method of Theorem 4.4 gives the following lower bound on the size of 2-splitting families.

**Theorem 4.7.** The minimal size of a 2-splitting family of subsets of \([k]\) is \( \Omega(k) \).

**Proof.** This time we say that the volume of \( A \subseteq [k] \) is the number of (ordered) collections \( B_1, B_2 \) such that \( A \) simultaneously splits \( B_1, B_2 \). This time, it is easy to see that the splitters of maximum volume satisfy
\[ v \leq 9 \left( \sum_i \binom{k/2}{i} \binom{k/2}{i} \right)^2 \]
The number of ordered collections \( B_1, B_2 \) is \( (2^k)^2 \), so we obtain a bound of
\[ N/v = \Omega \left( (2^k)^2 / (2^k / \sqrt{k})^2 \right). \]
The latter expression is \( \Omega(k) \), as desired. \( \square \)

The sector-by-sector splitting technique described after Definition 4.6 does not work in general for collections of three or more sets, since for example \( B_1 = \{1, 2\}, B_2 = \{2, 3\}, B_3 = \{3, 1\} \) is simply not 3-splittable. The next result states that this is type of example essentially the only obstacle to 3-splittablity.

**Lemma 4.8.** The collection \( A, B, C \) is not 3-splittable if and only if \( |A \cap B \cap C^c| \) is odd and \( |A \cap B^c \cap C| \) is odd and \( |A^c \cap B \cap C| \) is odd and \( A \cap B \cap C \) is empty and there are no elements in \( A \cup B \cup C \) besides those in the sets above.
Proof. In the proof we will make numerous references to the seven regions of the Venn diagram of \(A, B, C\), and for convenience we label them according to the figure shown below.

We first show that if \(R_{AB}, R_{BC},\) and \(R_{AC}\) have odd size, and \(R_A = R_B = R_C = R_{ABC} = \emptyset\), then \(A, B, C\) is not 3-splittable. Indeed, suppose towards a contradiction that \(S\) simultaneously splits \(A, B, C\). Without loss of generality we can suppose \(|S \cap R_{AB}| > |R_{AB}|/2\). It follows that \(|S \cap R_{BC}| < |R_{BC}|/2\), and then that \(|S \cap R_{AC}| > |R_{AC}|/2\). It follows that \(S\) does not split \(R_{AB} \cup R_{AC}\), which is a contradiction because \(R_{AB} \cup R_{AC} = A\) in this case. For the converse, we show that if \(A, B, C\) is not of this form, then it is 3-splittable. We first consider the case when \(R_A = R_B = R_C = \emptyset\). If all four of the sectors \(R_{AB}, R_{BC}, R_{AC}, R_{ABC}\) are even, then we can build a splitter \(A\) by simply splitting each sector in half. If just one of these four sectors is odd we can simply round sector up or down. If just two of these four sectors is odd we can round one of them up and the other down. This leaves only the following three subcases shown in Figure 2.

\[
\text{Figure 2. Subcases 1, 2, and 3 from left to right. The symbols } e \text{ and } o \text{ denote even and odd sized sectors.}
\]

In subcases 1 and 2, we round the sector \(R_{ABC}\) up, and each of the other odd sectors down. In subcase 3, we know that \(|R_{ABC}| \geq 2\) (or else we are in the converse situation), and so we can build a splitter \(S\) with \(|S \cap R_{ABC}| = |R_{ABC}|/2 - 1\), and each intersection of \(S\) with \(R_{AB}, R_{BC}, R_{AC}\) rounded down.

We next consider the case when at least one of \(R_A, R_B, R_C\) is nonempty. If there exists a splitter \(S\) for the configuration \(A \setminus R_A, B \setminus R_B, C \setminus R_C\) is splittable, then we can build a splitter for \(A, B, C\)
by letting $S' \supset S$ and suitably rounding the intersection of $S'$ with $R_A, R_B, R_C$ up or down. Finally, if $A \setminus R_A, B \setminus R_B, C \setminus R_C$ is not splittable, then by the above analysis we must have $R_{AB}, R_{BC}, R_{AC}$ odd and $R_{ABC} = \emptyset$. Suppose for concreteness that $R_A \neq \emptyset$. Then we can build a splitter $S$ such that $S \cap R_{AB}$ is rounded down, $S \cap R_{BC}$ is rounded up, $S \cap R_{AC}$ is rounded down, and $|S \cap R_A| = |R_A|/2 - 1$. This completes the proof. 

We conjecture that the problem of deciding whether an arbitrary collection of sets is splittable is NP-complete.

We devote the rest of this section to establishing our upper bound on the minimum size of a $2$-splitting family, and stating a conjecture concerning the upper bound on the minimum size of an $n$-splitting family.

**Theorem 4.9.** The minimal size of a $2$-splitting family of subsets of $[k]$ is $O(k^2)$.

For this, we will need the following key result.

**Theorem 4.10.** Let $S, T \subseteq [k]$ of fixed size $s = |S|, t = |T|$. If $b = |S \cap T|$ then the number of splitters of $S$ and $T$ is a nondecreasing function of $b$.

Although the statement of Theorem 4.10 feels intuitive, the proof is somewhat technical and will be given as the conclusion of a series of lemmata. We first deal with the case where $s$ and $t$ are both even. To this end, we fix sets $S, T$ as in the statement of the theorem. Throughout the proof we will fix elements $x \in S \setminus T$ and $y \in T \setminus S$ (we may assume these exist) and consider two more pairs of sets. $S' = S \setminus \{x\}$ and $T' = T \setminus \{y\}$, and also $S'' = S$ and $T'' = (T \setminus \{y\}) \cup \{x\}$. Refer to Figure 3 to visualize these configurations.

**Lemma 4.11.** Suppose $s, t$ are both even.

- If $A$ is the family of sets that split $T$ and $S$ simultaneously and $B$ is the family of sets that split $T'$ and $S'$ simultaneously, then $4|A| = |B|$.
- If $C$ is the family of sets that splits $T''$ and $S''$ simultaneously, then $\frac{1}{2}|B| \leq |C|$.

*Proof.* There are $k - t - s + b$ elements in $[k] \setminus (T \cup S)$; we will ignore those elements as they multiply $|A|$ and $|B|$ both by the constant $2^{k-s-t+b}$, which does not affect our calculations. We proceed as if $t + s - b = k$.

Let $A \in A$, so $A$ splits $T$ and $S$. If $x \in A$ then $|A \cap T'| = \lfloor (t-1)/2 \rfloor$ and if $x \not\in A$ then $|A \cap T'| = \lceil (t-1)/2 \rceil$. The corresponding statement is true for $S'$, so $A$ splits $T'$ and $S'$, and hence $A \subseteq B$.

We can generate four distinct splitters of $T'$ and $S'$ from each splitter $A$ of $T$ and $S$. Let $A_1, A_2, A_3, A_4$ be sets such that $A_i \cap (T' \cup S') = A \cap (T' \cup S')$ and $x \in A_1 \cap A_2 \cap A_3 \cap A_4$ and $y \in A_1 \cap A_2 \cap A_3 \cap A_4$. Let $B$ be a splitter of $T$ and $S$ such that $B \neq A$. Then $B \cap (T' \cup S') \neq A \cap (T' \cup S')$. Then given splitters generated from $B$ and $A$, $A_i \cap (T' \cup S') = A \cap (T' \cup S') \neq B \cap (T' \cup S') = B_i \cap (T' \cup S')$,
so $A_i \neq B_i$. So the splitters generated from $B$ and $A$ are distinct, and there are 4 splitters in $B$ for each splitter in $A$. Thus $4|A| = |B|$, which concludes the proof of the first statement.

For the second statement, it suffices to show that at least $1/4$ of the elements of $B$ are in fact elements of $C$. To this end let $B \in B$, so $B$ is a splitter of $T'$ and $S'$. Note that $B$ is a splitter of $T''$ and $S''$ if and only if it rounds the same way to split both $T'$ and $S'$. That is $B \cap T' = t/2$ and $B \cap S' = s/2$; or $B \cap T' = t/2 - 1$ and $B \cap S' = s/2 - 1$. Furthermore, a splitter $B$ that rounds down (up) to split both $T'$ and $S'$ will also split $T''$ and $S''$ if and only if $x$ is in $B$ ($x$ is not in $B$). Thus precisely half of the sets that split both $T'$ and $S'$ also round the same way both times will also split $T''$ and $S''$.

Thus it now suffices to show that more than half of the splitters of both $T'$ and $S'$ round the same way for both sets. The number of splitters that round the same way each time is given by:

$$
\sum_{i=0}^{b} \binom{b}{i} \left[ \binom{t-b-1}{t/2-i} \binom{s-b-1}{s/2-i} + \binom{t-b-1}{t/2-i-1} \binom{s-b-1}{s/2-i-1} \right]
$$

On the other hand, the number of splitters that round differently each time is given by:

$$
\sum_{i=0}^{b} \binom{b}{i} \left[ \frac{t-b-1}{t/2-i} \frac{s-b-1}{s/2-i} + \frac{t-b-1}{t/2-i-1} \frac{s-b-1}{s/2-i-1} \right]
$$

Figure 3. A diagram showing the two steps of the proof when $s, t$ are even.
We shall show term-by-term that the first sum is greater than or equal to the second sum. Taking the $i$th term of the first sum minus the $i$th term in the second sum and factoring, we obtain that the following is equivalent to what we wish to show:

$$
\left[\frac{t - b - 1}{t/2 - i} - \frac{t - b - 1}{t/2 - i - 1}\right]\left[\frac{s - b - 1}{s/2 - i} - \frac{s - b - 1}{s/2 - i - 1}\right] \geq 0
$$

By the unimodality of the binomial coefficient, both of the terms in the above product are negative for $i < b/2$ and both are nonnegative for $i \geq b/2$, so the inequality is always true. Consequently at least $\frac{1}{4}$ of splitters in $B$ are also splitters in $C$, and $\frac{1}{4}|B| \leq |C|$.

For the case where $s$ is odd and $t$ is even, we assume that there exists $x \in [k] \setminus (T \cup S)$ and $y \in T \setminus S$ in order for this case to be reduced to the above, $s$ even case. Let $S' = S \cup \{x\}$ and $T' = (T \setminus \{y\}) \cup \{x\}$, so $|T'| = t$, $|S'| = s$, and $|T' \cap S'| = b + 1$.

**Lemma 4.12.** Suppose $s$ is odd and $t$ is even.

- If $A$ is the family of sets that split $T$ and $S$ simultaneously and $B$ is the family of sets that split $T$ and $S'$ simultaneously, then $|A| = 2|B|$.
- If $C$ is the family of sets which simultaneously split $S'$ and $T'$, then $2|B| \leq |C|$.

**Proof.** Let $A \subseteq A$. Then $A$ splits $S'$ if and only if it rounds $S$ down and contains $x$, or rounds $S$ up and omits $x$. $A$ rounds $S$ up or down independently of whether it contains $x$, and so there are two splitters in $A$ for each splitter in $B$, i.e. $|A| = 2|B|$. The second point in the above proof follows from Lemma 4.11.

We are now ready to complete the proof of the key result.

**Proof of Theorem 4.10.** It suffices to prove that if $|T| = |T'| = t$ and $|S| = |S'| = s$ and $|T \cap S| = b$ and $|T' \cap S'| = b + 1$, then there is an injection from the set $A$ of splitters of $T$ and $S$ to the set $B$ of splitters of $T'$ and $S'$. Let $|T| = t$, $|S| = s$, $|T'| = t - 1$, $|S'| = s - 1$. The case $t, s$ even is handled by Lemma 4.11 and the cases $t$ even, $s$ odd and $t$ odd, $s$ even are handled by Lemma 4.12. The remaining case is $t, s$ odd. For this case we adjoin an element to $T$ and reduce to the case $t$ even, $s$ odd, just as this case in turn was reduced to the case $t, s$ even.

Finally, we can conclude the upper bound calculation for the minimal size of a 2-splitting family.

**Proof of Theorem 4.9.** If $T \subseteq [k]$ with $|T| = t$, then by a Stirling-type approximation computed in Section 2 of [Sti02], the probability that $S$ is split by a uniformly random subset of $[k]$ has a lower bound of $c\sqrt{t}$ where $c$ is a constant.

Next, if also $S \subseteq [k]$ with $|S| = s$ then by Theorem 4.10 the probability $p_{S,T}$ that a uniformly random set simultaneously splits $S$ and $T$ is minimized when $S \cap T = \emptyset$. In this case, the event that $S$ is split and the event that $T$ is split are independent, and so $p_{S,T} = p_s p_t$. Minimizing over the possible sizes $s$ and $t$, we have that $p_{S,T}$ has lower bound $p \geq c^2/k$. We now wish to invoke
Lemma 3.13 with this value of $p$ and $N = (2^k)^2$. Thus, letting $m$ be the minimal size of a 2-splitting family of subsets of $[k]$, we have that

$$m \leq \frac{\log((2^k)^2)}{-\log(1 - c^2/k)} \leq \frac{2k}{-\log(1 - c^2/k)}$$

This latter expression is $O(k^2)$, as desired.

We close by conjecturing that the analog of Theorem 4.10 holds for configurations of $n$ sets as well. If this conjecture holds true, one can easily obtain an upper bound on the minimal size of an $n$-splitting family of $t$-subsets of $[k]$ of size $O(f(n)k^{n/2+1})$.

**Conjecture.** Let $B_1, \ldots, B_n$ be a collection of subsets of $[k]$. Then the number of splitters of $B_1, \ldots, B_n$ is minimized when the collection is pairwise disjoint.

**References**

[Che93] William Y. C. Chen. Induced cycle structures of the hyperoctahedral group. *SIAM J. Discrete Math.*, 6(3):353–362, 1993.

[DSL07] D. Deng, D. R. Stinson, P. C. Li, G. H. J. van Rees, and R. Wei. Constructions and bounds for $(m, t)$-splitting systems. *Discrete Math.*, 307(1):18–37, 2007.

[FK84] Michael L. Fredman and János Komlós. On the size of separating systems and families of perfect hash functions. *SIAM J. Algebraic Discrete Methods*, 5(1):61–68, 1984.

[Harr65] Michael A. Harrison. *Introduction to switching and automata theory*. McGraw-Hill Book Co., New York-Toronto-London, 1965.

[HH68] Michael A. Harrison and Robert G. High. On the cycle index of a product of permutation groups. *J. Combinatorial Theory*, 4:277–299, 1968.

[LLvR04] Alan C. H. Ling, P. C. Li, and G. H. J. van Rees. Splitting systems and separating systems. *Discrete Math.*, 279(1-3):355–368, 2004. In honour of Zhu Lie.

[Lov73] L. Lovász. Coverings and coloring of hypergraphs. In *Proceedings of the Fourth Southeastern Conference on Combinatorics, Graph Theory, and Computing* (Florida Atlantic Univ., Boca Raton, Fla., 1973), pages 3–12. Utilitas Math., Winnipeg, Man., 1973.

[Rén61] A. Rényi. On random generating elements of a finite Boolean algebra. *Acta Sci. Math. Szeged*, 22:75–81, 1961.

[RH12] Dongyoung Roh and Sang Geun Hahn. Constructions for uniform $(m, 3)$-splitting systems. *Math. Commun.*, 17(2):639–654, 2012.

[Sti02] D. R. Stinson. Some baby-step giant-step algorithms for the low Hamming weight discrete logarithm problem. *Math. Comp.*, 71(237):379–391 (electronic), 2002.