



Access balancing in storage systems by labeling partial Steiner systems

Yeow Meng Chee¹ · Charles J. Colbourn² · Hoang Dau³ · Ryan Gabrys⁴ · Alan C. H. Ling⁵ · Dylan Lusi² · Olgica Milenkovic⁴

Received: 28 December 2019 / Revised: 14 July 2020 / Accepted: 30 July 2020 / Published online: 14 August 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Storage architectures ranging from minimum bandwidth regenerating encoded distributed storage systems to declustered-parity RAIDs can employ dense partial Steiner systems to support fast reads, writes, and recovery of failed storage units. To enhance performance, popularities of the data items should be taken into account to make frequencies of accesses to storage units as uniform as possible. A combinatorial model ranks items by popularity and assigns data items to elements in a dense partial Steiner system so that the sums of ranks of the elements in each block are as equal as possible. By developing necessary conditions in terms of independent sets, we demonstrate that certain Steiner systems must have a much larger difference between the largest and smallest block sums than is dictated by an elementary lower bound. In contrast, we also show that certain dense partial $S(t, t + 1, v)$ designs can be labeled to realize the elementary lower bound. Furthermore, we prove that for every admissible order v , there is a Steiner triple system $(S(2, 3, v))$ whose largest difference in block sums is within an additive constant of the lower bound.

Keywords Steiner system · Steiner triple system · Independent set · Access balancing

Mathematics Subject Classification 05B07 · 05B40 · 68R05 · 94B60

1 Introduction

Distributed storage systems [15,35], systems for batch coding [36], and multiserver private information retrieval systems [18] have each employed combinatorial designs for data placement, so that elements of the design are associated with data items and blocks with storage units. In these contexts, the most common types of designs employed are t -designs and t -

Communicated by M. Buratti.

The work was supported by NSF Grants CCF 1816913 (CJC) and 1814298 (OM).

✉ Charles J. Colbourn
charles.colbourn@asu.edu

Extended author information available on the last page of the article

packings. A t -(v, k, λ) *packing* is a pair $(X, \mathcal{B})$, where X , the point set, is a v -set and $\mathcal{B}$ is a collection of k -subsets (*blocks*) of X such that every t -subset of X is contained in at most λ blocks. The packing is a t -(v, k, λ) *design* when every t -subset of X is a subset of exactly λ blocks. A t -($v, k, 1$) design is a *Steiner system*, denoted by $S(t, k, v)$. A 2-($v, 3, 1$) design is a *Steiner triple system of order v* , denoted by $STS(v)$. When $\lambda = 1$, a t -($v, k, 1$) packing is also referred to as a *partial $S(t, k, v)$* or *partial Steiner system*.

When data items are of the same size, and data is placed on storage units using a t -design, placement of data is uniform across the storage units. Indeed in t -(v, k, λ) design, every point appears in exactly $r = \frac{\lambda \binom{v-1}{t-1}}{\binom{k-1}{t-1}}$ blocks; this is the *replication number* of the design. In order to understand why Steiner systems can be employed in data placement, we outline some examples. Large-scale distributed storage systems (DSS) must address potential loss of storage units, while not losing data. One solution is to replicate each data item and distribute these replicas among multiple storage nodes; systems such as the Hadoop Distributed File System and the Google File System employ this strategy [9]. One can further mitigate information loss by sensibly organizing the data. For example, exact Minimum Bandwidth Regenerating (MBR) codes [15] consist of two subcodes, an outer MDS code along with an inner *fractional repetition code* (FRC) that support redundancy and repairability, respectively. To make this precise, an (n, k, d) -DSS with $k \leq d \leq n$ consists of n storage nodes in which a read can be accomplished by access to k nodes and a failed node recovered by access to d nodes. A fractional repetition code $\mathcal{C}$ [15] with repetition degree ρ for an (n, k, d) -DSS is a collection $\mathcal{C}$ of n subsets $V_1, V_2, \dots, V_n$ of a set V , $|V| = v$, and of cardinality d each, satisfying the condition that each element of V belongs to exactly ρ different sets in the collection. The *rate* of the FRC is $\min_{I \subseteq [n], |I|=k} |\bigcup_{i \in I} V_i|$. To optimize the rate and ensure correct repetition and repair, we require that $|V_i \cap V_j| \leq 1$ whenever $i \neq j$. When $\rho = \frac{v-1}{d-1}$, such an FRC is a Steiner 2-($v, d, 1$) design with replication number ρ , where the set of (coded) file chunks V is the set of points and the set of storage nodes $\{V_1, \dots, V_n\}$ is the set of blocks of the design.

Steiner systems also prove useful for applications needing both high data availability and throughput, such as transaction processing. The storage systems underlying these applications require uninterrupted operation, satisfying user requests for data even in the event of disk failure and repairing these failed disks, on-line, in parallel. Continuous operation alone is not sufficient, because such systems cannot afford to suffer significant loss of performance during disk failures. *Declassified-parity RAID*s (DPRAIDS) are designed to satisfy these requirements [8,23]. Like standard RAIDs (short for “Redundant Arrays of Inexpensive Disks”), DPRAIDS handle disk failure by using parity-encoded redundancy, in which subsets of the stored data (called *parity stripes*) are XORed together to store a single-error-correction code. Unlike standard RAIDs, however, all disks in the DPRAID cooperate in the reconstruction of all the data units on a single failed disk. One can represent a DPRAID as a t -(v, k, λ) design $(X, \mathcal{B})$, with X ($|X| = v$) being the set of disks in the array, and $\mathcal{B}$ being the set of all parity stripes, each of size k . Then each disk occurs in the same number c of parity stripes, guaranteeing that the reconstruction effort is distributed evenly.

Although designs arise naturally in balancing data placement, little attention has been paid to the relative popularity of the data items. However, one can exploit popularity information in order to improve the relative equality of access among the storage units. Dau and Milenkovic [12] formulate a number of problems to address access balancing, by labeling the points of the underlying design. In order to introduce their problems and results, we first present more definitions and known results concerning designs.

Although storage systems handle “hot” (frequently accessed) and “cold” (infrequently accessed) data categories differently, typically they do not take the long-term popularity of

the data items within each category into account, which may result in unbalanced access frequencies to the storage units. Access balancing can be achieved in part by selecting an appropriate packing or design, and by appropriate association of data items with elements of the packing or design. Dau and Milenkovic [12] propose a combinatorial model that ranks data items by popularity, and then strives to ensure that the sums of the ranks of the data elements in each block are not too small, not too large, or not too different from block to block. In Sect. 2 we summarize their model, state elementary bounds on various block sums, and provide a small but important improvement in the lower bound on the smallest possible difference among the block sums in a Steiner triple system. In Sect. 3 we establish a close connection between such block sums and the size of a maximum independent set of elements in the packing or design. For certain designs, this connection can be used to show that, no matter how data items are associated with the elements of the design, the block sums must be far from the values dictated by the elementary bounds from Sect. 2. Indeed, in order to approach the elementary bounds, one must select designs or packings with very specific properties; we pursue this in Sect. 4. Our results indicate the need to find specific $S(t, k, v)$ designs, or at least ‘dense’ t -($v, k, 1$) packings, to match the elementary bounds more closely. In Sect. 5, we explore a construction of t -($v, t + 1, 1$) packings that asymptotically match the bounds and contain almost the same number of blocks as the full Steiner system $S(t, t + 1, v)$. Completion of the dense t -($v, t + 1, 1$) packings to a Steiner system $S(t, t + 1, v)$ appears problematic for general t ; doing so without dramatically changing the block sums appears to be even more challenging. Nevertheless, in Sect. 6, we pursue this to establish, for every admissible order v , the existence of a Steiner triple system of order v whose difference in block sums is at most an additive constant more than the elementary lower bound.

A preliminary version of this research, without proofs, appears in [7].

2 Point labelings and block sums

Let $D = (V, \mathcal{B})$ be a t -(v, k, λ) packing. A *point labeling* of D is a bijection $\text{rk} : V \mapsto \{0, \dots, v - 1\}$; our interpretation is that rk maps an element to its rank by popularity. The *reverse* $\overline{\text{rk}}$ of a point labeling rk has $\overline{\text{rk}}(i) = v - 1 - \text{rk}(i)$ for each $i \in \{0, \dots, v - 1\}$; the *reversal* of a point-labeled packing is one having the reverse of the point labeling. With respect to a specific point labeling rk , define $\text{sum}(B, \text{rk}) = \sum_{x \in B} \text{rk}(x)$ when $B \in \mathcal{B}$. Then define

$$\text{MinSum}(D, \text{rk}) = \min(\text{sum}(B, \text{rk}) : B \in \mathcal{B});$$

$$\text{MaxSum}(D, \text{rk}) = \max(\text{sum}(B, \text{rk}) : B \in \mathcal{B});$$

$$\text{DiffSum}(D, \text{rk}) = \text{MaxSum}(D, \text{rk}) - \text{MinSum}(D, \text{rk});$$

$$\text{RatioSum}(D, \text{rk}) = \text{MaxSum}(D, \text{rk}) / \text{MinSum}(D, \text{rk}).$$

Following [12], one primary objective is to choose point labelings to maximize the MinSum and/or to minimize one of the other three. Access balancing is concerned primarily with minimizing the DiffSum or RatioSum ; because of the similarity between these two entities we often focus on the DiffSum . Let $\mathcal{R}_D$ denote the set of all point labelings of D . Noting that

$\text{MaxSum}(D, \text{rk}) = k(v - 1) - \text{MinSum}(D, \overline{\text{rk}})$, we define

$$\text{MinSum}(D) = \max(\text{MinSum}(D, \text{rk}) : \text{rk} \in \mathcal{R}_D);$$

$$\text{MaxSum}(D) = k(v - 1) - \text{MinSum}(D);$$

$$\text{DiffSum}(D) = \min(\text{DiffSum}(D, \text{rk}) : \text{rk} \in \mathcal{R}_D);$$

$$\text{RatioSum}(D) = \min(\text{RatioSum}(D, \text{rk}) : \text{rk} \in \mathcal{R}_D).$$

If the storage system dictates the data layout and data items have the same size, we are free to permute the data items; this is captured by the selection of the point labeling rk . If we are also free to choose the t -($v, k, 1$) packing that underlies the data layout, we may select a packing to improve the sum metrics defined. In order to capture this, let $\mathcal{D}_{t,k,v,b}$ denote the set of all t -($v, k, 1$) packings having exactly b blocks. Then define

$$\text{MinSum}(t, k, v, b) = \max(\text{MinSum}(D) : D \in \mathcal{D}_{t,k,v,b});$$

$$\text{MaxSum}(t, k, v, b) = k(v - 1) - \text{MinSum}(t, k, v, b);$$

$$\text{DiffSum}(t, k, v, b) = \min(\text{DiffSum}(D) : D \in \mathcal{D}_{t,k,v,b});$$

$$\text{RatioSum}(t, k, v, b) = \min(\text{RatioSum}(D) : D \in \mathcal{D}_{t,k,v,b}).$$

When $b = \binom{v}{t}$, the packing is a Steiner system $S(t, k, v)$; in these cases we omit b from the notation to get $\text{MinSum}(t, k, v)$ and similarly for all other entities.

Theorem 1 [12] *When D is a Steiner system $S(t, k, v)$,*

$$\text{MinSum}(D) \leq \text{MinSum}(t, k, v) \leq \frac{1}{2}(v(k - t + 1) + k(t - 2));$$

$$\text{MaxSum}(D) \geq \text{MaxSum}(t, k, v) \geq \frac{1}{2}(v(k + t - 1) - kt);$$

$$\text{DiffSum}(D) \geq \text{DiffSum}(t, k, v) \geq (v - k)(t - 1);$$

$$\text{RatioSum}(D) \geq \text{RatioSum}(t, k, v) \geq \frac{v(k + t - 1) - kt}{v(k - t + 1) + k(t - 2)}.$$

When $k = t + 1$, $\text{MinSum}(D) \leq (v - 1) + \binom{t}{2}$, $\text{MaxSum}(D) \geq t(v - 1) - \binom{t}{2}$, $\text{DiffSum}(D) \geq (t - 1)(v - t - 1)$, and $\text{RatioSum}(D) \geq \text{RatioSum}(t, t + 1, v) \geq \frac{t(v-1)-\binom{t}{2}}{(v-1)+\binom{t}{2}}$.

When in addition $t = 2$ (D is a Steiner triple system), the stronger bounds $\text{DiffSum}(D) \geq v$ and $\text{RatioSum}(D) \geq 2$ hold.

Theorem 1 provides bounds on the metrics across all Steiner systems $S(t, k, v)$ and all point labelings of them. In previous work, the focus has been on the MinSum (or equivalently, by reversal, the MaxSum). Dau and Milenkovic [12] use the Bose [3] and Skolem [22,37] constructions of Steiner triple systems to establish the existence of an $\text{STS}(v)$ D with $\text{MinSum}(D) = v$, the largest possible by Theorem 1 (Brummond [4] establishes a similar result for Kirkman triple systems). They accomplish this by specifying a particular point labeling that meets the MinSum bound, but unfortunately the labeling chosen yields a MaxSum near $\frac{8}{3}v$, a DiffSum near $\frac{5}{3}v$, and a RatioSum near $\frac{8}{3}$, far from the bounds of $2v$, v , and 2 , respectively. The reversal of this labeling yields a MinSum far from optimal, the same DiffSum , and a larger RatioSum .

One might hope to improve the DiffSum and RatioSum by choosing a different labeling or by choosing a different Steiner system $S(t, k, v)$. In Sect. 3, we show that certain $S(t, k, v)$ s cannot meet any of the bounds in Theorem 1.

2.1 An easy case

Before embarking on the practical cases with $t \geq 2$, we provide a complete solution for Steiner systems with $t = 1$, which are partitions of a set of size ks into s sets, each of size k .

Lemma 1 *Let k be a nonnegative integer with $k \neq 1$ and let s be a positive integer. When $k(s-1)$ is even, $\text{MinSum}(1, k, ks) = \text{MaxSum}(1, k, ks) = \frac{1}{2}k(ks-1)$; moreover, $\text{DiffSum}(1, k, ks) = 0$ and $\text{RatioSum}(1, k, ks) = 1$. When $k(s-1)$ is odd,*

$$\begin{aligned}\text{MinSum}(1, k, ks) &= \frac{1}{2}(k(ks-1) - 1); \\ \text{MaxSum}(1, k, ks) &= \frac{1}{2}(k(ks-1) + 1); \\ \text{DiffSum}(1, k, ks) &= 1; \\ \text{RatioSum}(1, k, ks) &= 1 + \frac{2}{k(ks-1) - 1}\end{aligned}$$

Proof The sum of all elements is $\binom{ks}{2}$, so the average block sum is $\frac{1}{2}k(ks-1)$. This is an integer whenever k is even or s is odd. Now we provide constructions. Suppose that there is a point-labeled $S(1, k, ks)$ with MinSum m and MaxSum M , having blocks $B_0, \dots, B_{s-1}$. Add s to the label of each element, and adjoin new elements i and $(k+2)s-1-i$ to block B_i , for $0 \leq i \leq s-1$. This yields a point-labeled $S(1, k+2, (k+2)s)$ having MinSum $m+ks+(k+2)s-1$ and MaxSum $M+ks+(k+2)s-1$. It follows that if the Lemma holds for an $S(1, k, ks)$, it also holds for an $S(1, k+2, (k+2)s)$. Therefore we need only treat cases when $k \in \{0, 3\}$. When $k = 0$, there are no elements in the system and the statements hold trivially.

It remains to treat cases with $k = 3$. First we treat the case when s is even, writing $s = 2\ell$. Form the triples $\{\{2i, 4\ell-1-i, 5\ell-1-i\} : 0 \leq i < \ell\}$, each having sum $9\ell-2 = \frac{1}{2}(3(6\ell-1)-1)$. Then adjoin the triples $\{\{2i+1, 3\ell-1-i, 6\ell-1-i\} : 0 \leq i < \ell\}$, each having sum $9\ell-1 = \frac{1}{2}(3(6\ell-1)+1)$. The result is a point-labeled $S(1, 3, 3s)$ with the required sums. Finally we treat the case when s is odd, writing $s = 2\ell+1$. Form the triples $\{\{i, 3\ell+1+i, 6\ell+2-2i\} : 0 \leq i \leq \ell\}$, each having sum $9\ell+3 = \frac{1}{2}3(3(2\ell+1)-1)$. Adjoin the triples $\{\{\ell+1+i, 2\ell+1+i, 6\ell+1-2i\} : 0 \leq i < \ell\}$, each again having sum $9\ell+3$. The result is a point-labeled $S(1, 3, 3s)$ with the required sums. $\square$

2.2 Improved bounds for STSs

There is an STS(7) with $\text{MinSum} = 6$ and $\text{MaxSum} = 13$ with blocks 016, 024, 035, 123, 145, 256, and 346 (here we write abc for $\{a, b, c\}$). There is an STS(9) with $\text{MinSum} = 9$ and $\text{MaxSum} = 18$ with blocks 018, 027, 036, 045, 126, 135, 147, 234, 258, 378, 468, and 567. However, we establish that these are the *only* two Steiner triple systems with $\text{DiffSum} = v$, and indeed the *only* STS(v) with $\text{RatioSum} = 2$ is the STS(9). We first prove a useful lemma.

Lemma 2 *A 2- $(x, 3, 1)$ packing on $\{0, \dots, x-1\}$ with MaxSum $x-1$ has at most $\lfloor \phi(x)/3 \rfloor$ triples, where*

$$\phi(x) = \left\lfloor \frac{x(x-1)}{4} \right\rfloor - \left\lfloor \frac{x}{12} \right\rfloor - \begin{cases} 0 & \text{if } x \equiv 0, 1, 2, 3, 4, 6, 7, 10 \pmod{12} \\ 1 & \text{if } x \equiv 5, 8, 9, 11 \pmod{12} \end{cases}$$

Proof We determine an upper bound on $\phi(x)$, the number of pairs that could appear in triples of the packing. In total there are $\binom{x}{2}$ pairs; of these, exactly $\lfloor \frac{x}{2} \rfloor$ have sum equal to

$x - 1$. For each pair $\{a, b\}$ with $a + b < x - 1$, the pair $\{x - 1 - a, x - 1 - b\}$ has sum $2x - 2 - (a + b) > x - 1$. Hence the number of pairs with sum at most $x - 1$ is $\frac{x^2}{4}$ when x is even, and $\frac{x^2-1}{4}$ when x is odd. Not all of these can appear together in a packing, as follows. Let $a \in \{0, \dots, \lfloor \frac{x-2}{3} \rfloor\}$. Consider the pairs $P_a = \{\{a, x - 1 - b\} : a \leq b \leq 2a\}$. To place a pair of P_a in a triple of sum at most $x - 1$, the third element must be from $\{0, \dots, a\}$, but it cannot be a . By the pigeonhole principle, at least one pair of P_a cannot be in a triple of the packing, reducing the number of pairs available by $\lfloor \frac{x+1}{3} \rfloor$. Hence the number $\phi(x)$ of pairs available is as given in the statement. $\square$

Theorem 2 *Let D be a Steiner triple system of order $v \geq 13$. Then $\text{DiffSum}(D) \geq v + 1$ and $\text{RatioSum}(D) > 2$.*

Proof Let $v = 2x + 1$, and consider an STS($2x + 1$) D on elements $\{0, \dots, 2x\}$, noting that $x \geq 6$. Partition $\{0, \dots, 2x\}$ into three classes $V_0 = \{0\}$, $V_s = \{1, \dots, x\}$, and $V_\ell = \{x + 1, \dots, 2x\}$. Let $m = \text{MinSum}(D)$ and $M = \text{MaxSum}(D)$. Suppose to the contrary that $\text{DiffSum}(D) = v$. Then $(m, M) \in \{(v - 3, 2v - 3), (v - 2, 2v - 2), (v - 1, 2v - 1), (v, 2v)\}$. If $m \leq v - 2$, the reversal of D has $\text{MinSum}(D) \in \{v - 1, v\}$, so we suppose that $m \in \{v - 1, v\}$. Because $m \geq v - 1$, all triples containing 0 contain one element of V_s and one of V_ℓ , as follows. Consider the pair $\{0, w\}$ with $w \in V_s$. The third element y completing its triple satisfies $y \geq 2x - w \geq 2x - x = x$. Now $y \neq x$ because $\{0, x, x\}$ cannot be a triple, so $y > x$, and hence $y \in V_\ell$. This accounts for all triples involving 0.

Call a pair *mixed* if it contains an element of V_s and one from V_ℓ , *pure* otherwise. Similarly a triple is *pure* if it lies entirely on V_s or V_ℓ , *mixed* when it has two from one and one from the other. The number of mixed triples can be calculated as follows. There are $x(x - 1)$ mixed pairs not contained in triples containing 0, and each must be contained in a mixed triple. Because each mixed triple contains two mixed pairs, there are exactly $\frac{1}{2}x(x - 1)$ mixed triples. Each mixed triple covers one pure pair. Hence the number of pure pairs to be covered by pure triples is $x(x - 1) - \frac{1}{2}x(x - 1) = \frac{1}{2}x(x - 1)$, and there are $\frac{1}{6}x(x - 1)$ pure triples.

Form a collection $\mathcal{D}_s$ of triples on $\{0, \dots, x - 1\}$ by including $\{a, b, c\}$ whenever $\{a + 1, b + 1, c + 1\}$ is a pure triple on V_s ; then $\mathcal{D}_s$ contains triples each having sum at least $m - 3$. The reversal $\mathcal{E}_s$ then has each sum at most $3x - m$. Form a collection $\mathcal{D}_\ell$ of triples on $\{0, \dots, x - 1\}$ by including $\{a, b, c\}$ whenever $\{2x - a, 2x - b, 2x - c\}$ is a pure triple on V_ℓ , so that $\mathcal{D}_\ell$ contains triples each having sum at least $6x - M$. The reversal $\mathcal{E}_\ell$ then has each sum at most $M - 3x - 3$.

Case 1. $\text{MinSum}(D) = v$ and hence $\text{MaxSum}(D) = 2v$. Then $\mathcal{E}_s$ and $\mathcal{E}_\ell$ both have maximum sum at most $x - 1$. Applying Lemma 2 to $\mathcal{E}_s$ and to $\mathcal{E}_\ell$, D can contain at most $2\phi(x)$ pairs in pure triples, but $2\phi(x) < \frac{1}{2}x(x - 1)$, which yields the contradiction. (Only when $v = 9$ would there be no contradiction.)

Case 2. $\text{MinSum}(D) = v - 1$ and hence $\text{MaxSum}(D) = 2v - 1$. Then $\mathcal{E}_s$ has maximum sum x and $\mathcal{E}_\ell$ has maximum sum $x - 2$. Because no pair involving element $x - 1$ can appear in a triple of $\mathcal{E}_\ell$, the number of pairs covered by triples is at most $\phi(x - 1)$ by Lemma 2. By a similar argument, the number of pairs covered by triples of $\mathcal{E}_s$ is at most $\phi(x + 1)$ by Lemma 2. Because $\frac{(x-1)(x-2)}{4} + \frac{x(x+1)}{4} = \frac{x(x-1)}{2} + \frac{1}{2}$, $\phi(x - 1) + \phi(x + 1) < \frac{1}{2}x(x - 1)$, which yields the contradiction. (Only when $v = 7$ would there be no contradiction.) $\square$

3 Independent Sets

Let $D = (V, \mathcal{B})$ be a t -(v, k, λ) packing. An *independent set* in D is a subset $X \subseteq V$ such that there is no $B \in \mathcal{B}$ with $B \subseteq X$. An independent set X is *maximal* if there is no independent set Y with $X \subset Y$, and *maximum* if there is no independent set Y such that $|Y| > |X|$. The *independence number* of D , denoted $\alpha(D)$, is the size of a maximum independent set. There is a close connection between the independence number of a packing and the quality of any of its labelings.

Lemma 3 A t -(v, k, λ) packing D has MinSum at most $k\alpha(D) - \binom{k}{2}$, MaxSum at least $k(v - 1 - \alpha(D)) + \binom{k}{2}$, and DiffSum at least $k(v + k - 2 - 2\alpha(D))$.

Proof It suffices to prove the statement for MinSum. No matter how D is given a point labeling, on elements with ranks in $\{0, \dots, \alpha(D)\}$, there is a block. The sum of this block is at most $\sum_{i=1}^k (\alpha(D) - (i - 1))$. $\square$

Corollary 1 Meeting the bound on MinSum in Theorem 1 for a t -($v, k, 1$) packing D requires that

$$\alpha(D) \geq \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2}.$$

For example, Corollary 1 states that a necessary condition for a partial Steiner triple system D to have MinSum equal to v is that $\alpha(D) \geq \frac{v}{3} + 1$.

We refine this bound by using a second disjoint independent set. Suppose that a t -(v, k, λ) packing D contains two *disjoint* independent sets of sizes γ_D and δ_D , respectively, with $\gamma_D \geq \delta_D$; two disjoint independent sets form an *independent pair*. Set

$$\begin{aligned}\gamma'_D &= \min \left(\gamma_D, \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2} \right), \\ \delta'_D &= \min \left(\delta_D, \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2} \right).\end{aligned}$$

Two independent sets form a *maximum independent pair* when $\gamma'_D + \delta'_D$ is as large as possible.

Lemma 4 An $S(t, k, v)$, D , with a maximum independent pair of sizes (γ_D, δ_D) has DiffSum at least $k(v + k - 2 - \delta'_D - \gamma'_D)$.

Proof Suppose to the contrary that some point labeling of D has DiffSum less than $k(v + k - 2 - \delta'_D - \gamma'_D)$. Without loss of generality, choose such a labeling in which the smallest x for which a block appears on $\{0, \dots, x - 1\}$ also has a block on $\{v - x, \dots, v - 1\}$; reverse the labeling if necessary to do this. Let c be the smallest value for which a block appears on $\{0, \dots, c\}$, so that $\{0, \dots, c - 1\}$ forms an independent set. Proceed similarly to select d so that $\{v - d, \dots, v - 1\}$ is an independent set. If $c \geq \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2}$, set $c' = \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2}$ and otherwise set $c' = c$. Then for the chosen labeling of D , we find MinSum at most $kc' - \binom{k}{2}$, based on Theorem 1 and the argument in the proof of Lemma 3. In the same manner, if $d \geq \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2}$, set $d' = \frac{v(k - t + 1)}{2k} + \frac{k + t - 3}{2}$ and otherwise set $d' = d$. The chosen labeling has MaxSum at least $k(v - 1 - d') + \binom{k}{2}$. Hence the DiffSum for this labeling is at least $k(v + k - 2 - c' - d')$, so $c + d \geq c' + d' > \gamma'_D + \delta'_D$. This contradicts the requirement that the maximum independent pair have sizes (γ_D, δ_D) . $\square$

Corollary 2 *Meeting the bound on DiffSum in Theorem 1 for an $S(t, k, v)$ D requires that D have a independent pair of sizes $(\lceil \frac{v(k-t+1)}{2k} + \frac{k+t-3}{2} \rceil, \lceil \frac{v(k-t+1)}{2k} + \frac{k+t-3}{2} \rceil)$.*

In order to meet the bound in Corollary 2, the independent pair must be maximum. However, this does not require that either of the independent sets of the pair be maximum. Nor is it required that the sum of their sizes be as large as possible (see [28] for Steiner triple systems). For a Steiner triple system, for example, Corollary 2 asks only for two disjoint independent sets, each of size at least $\frac{v}{3} + 1$, for a combined size of $\frac{2v}{3} + 2$. Applying the $2v + 1$ construction [10] twice to an STS(v), we form an STS($4v + 3$) having a maximum independent pair of sizes $(2v + 2, v + 1)$; despite the fact that the combined size is over $\frac{3}{4}$ of the size of the STS, such a pair could not lead to a DiffSum that meets the bound of Theorem 1, because the second largest of the pair is too small.

Corollary 2 gives a necessary condition, not a sufficient one. Nevertheless, some bounds on the metrics can be stated.

Lemma 5 *When a t -($v, k, 1$) packing D has two disjoint independent sets of sizes α and β , there is a point labeling with $\text{MinSum}(D) \geq \alpha + \binom{k-1}{2}$ and (for the same labeling) $\text{MaxSum}(D) \leq k(v-1) - \beta - \binom{k-1}{2}$, so $\text{DiffSum}(D) \leq k(v-k) - \beta - \alpha$.*

Proof Any point labeling assigning labels $\{0, \dots, \alpha - 1\}$ to the points of the independent set of size α , labels $\{v - \beta, \dots, v - 1\}$ to the points of the independent set of size β , and labels $\{\alpha, \dots, v - \beta - 1\}$ to the remaining points, meets the stated bounds. $\square$

A Steiner system $S(t, k, v)$ is 2-chromatic if its elements can be partitioned into two classes, both being independent sets. When a 2-chromatic $S(3, 4, v)$ D exists (see, for example, [13,24,30]), Lemma 5 establishes that $\text{DiffSum}(D) \leq 3v - 17$.

Recall that Dau and Milenkovic [12] use the Bose and Skolem constructions of Steiner triple systems. In retrospect, this choice is well-justified because the Bose construction leads to maximum independent pairs of sizes $(\frac{v}{3} + 1, \frac{v}{3} + 1)$ when $v \equiv 3 \pmod{6}$ and the Skolem construction leads to maximum independent pairs of sizes $(\frac{v+2}{3} + 1, \frac{v+2}{3} + 1)$ when $v \equiv 1 \pmod{6}$.

Labeling for access balancing must focus on Steiner triple systems, and on t -($v, k, 1$) packings in general, having large sizes in maximum independent pairs. This choice is important, because not all such systems have even a single large independent set, as we explain next.

4 Small maximum independent sets

Can one choose an arbitrary t -($v, k, 1$) packing, and by cleverly choosing a point labeling optimize one or more of the sum metrics? If not, how far from the bound of Theorem 1 can the best point labeling be? In order to discuss these questions, define

$$\alpha_{\min}(t, k, v) = \min\{\alpha(D) : D \text{ is a } t\text{-(}v, k, 1\text{) packing}\}, \text{ and}$$

$$\alpha_{\min}^*(t, k, v) = \min\{\alpha(D) : D \text{ is an } S(t, k, v)\}.$$

When an $S(t, k, v)$ exists, $\alpha_{\min}(t, k, v) \leq \alpha_{\min}^*(t, k, v)$.

Erdős and Hajnal [16] establish that $\alpha_{\min}(2, 3, v) \geq \lfloor \sqrt{2v} \rfloor$; indeed a simple greedy algorithm produces an independent set of this size.

A t -($v, k, 1$) packing has each element in at most $\binom{v-1}{t-1} / \binom{k-1}{t-1} = \prod_{i=1}^{t-1} \frac{v-i}{k-i}$ blocks. Spencer [38] generalized Turán's theorem for graphs to obtain

$$\alpha_{\min}(t, k, v) \geq c_k \frac{v}{\left(\prod_{i=1}^{t-1} \frac{v-i}{k-i}\right)^{\frac{1}{k-1}}}$$

for c_k a constant independent of v . For partial Steiner triple systems, this asserts that $\alpha_{\min}(2, 3, v) \geq c \cdot v\sqrt{2}/\sqrt{v-1}$, a small improvement on the Erdős-Hajnal result. State-of-the-art lower bounds [17,25,26,39] all differ only by constant factors, and all rely heavily on a theorem about “uncrowded” hypergraphs.

Theorem 3 [1] *Let $\kappa \geq 2$ be a fixed integer. Let G be a $(\kappa + 1)$ -uniform hypergraph on n vertices. Then there are constants $t_0(\kappa)$ and $n_0(\kappa, \tau)$ so that whenever*

1. G is uncrowded (i.e., has no 2-, 3-, or 4- cycles);
2. the maximum degree $\Delta(G)$ satisfies $\Delta(G) \leq \tau^\kappa$ where $\tau \geq t_0(\kappa)$; and
3. $n \geq n_0(\kappa, \tau)$,

one has that

$$\alpha(G) \geq \frac{.98}{e} \cdot 10^{-5/\kappa} \cdot \frac{n}{\tau} \cdot (\ln \tau)^{1/\kappa}.$$

The lower bound for all t -($v, k, 1$) packings in Theorem 4 is obtained by selecting a large uncrowded set of the blocks and applying Theorem 3, while the upper bound is established using the Lovász Local Lemma.

Theorem 4 [14,32] *For fixed k and t , there are absolute constants c and d for which*

$$cv^{\frac{k-t}{k-1}} (\log v)^{\frac{1}{k-1}} \leq \alpha_{\min}(t, k, v) \leq dv^{\frac{k-t}{k-1}} (\log v)^{\frac{1}{k-1}}.$$

It is possible in principle that restricting to Steiner systems, rather than packings, one might observe different behaviour in the minima. However, Phelps and Rödl [29] establish that the bounds of Theorem 4 apply to Steiner triple systems, not just to partial ones; that is,

$$c\sqrt{v \ln v} \leq \alpha_{\min}^*(2, 3, v) \leq d\sqrt{v \ln v}$$

for absolute constants c and d . Grable, Phelps and Rödl [20] establish similar statements when $t \in \{2, 3\}$ for all $k > t$.

For the applications intended, it is of interest to find independent sets of (at least) the size guaranteed efficiently. For research in this vein, see [2,19]. Of course, one wants to find a pair of disjoint maximum independent sets whose total size is as large as possible, but this is NP-complete even for 3-uniform hypergraphs [27]. Remarkably, there is a polynomial time algorithm to determine whether an $S(3, 4, v)$ contains two independent sets, each of size $v/2$ [11], but the ideas used do not appear to generalize.

Nevertheless, the bounds on sizes of smallest maximum independent sets provide bounds on the best sum metrics one can hope to achieve. Combining Lemma 3 and the results in [20,29], some Steiner triple systems only have point labelings far from the bounds of Theorem 1:

Theorem 5 *For infinitely many orders v , there is an absolute constant c so that there exists an STS(v) D with $\text{MinSum}(D) \leq 3c\sqrt{v \ln v} - 3$ and $\text{MaxSum}(D) \geq 3v - 3c\sqrt{v \ln v}$, and hence $\text{DiffSum}(D) \geq 3v - 6c\sqrt{v \ln v} + 3$.*

We must focus on specific Steiner systems or packings, if we are to obtain sum metrics at or near the basic bounds.

5 Dense t -(v , $t + 1$, 1) packings

We establish next that one can obtain metrics close to the optimal when $k = t + 1$ for packings that contain all but a vanishingly small fraction of the blocks of an $S(t, t + 1, v)$ as $v \rightarrow \infty$. The independent set requirements indicate that we must have a maximum independent pair having large sizes. To accomplish this, we partition all $(t + 1)$ -subsets of $\mathbb{Z}_v$ according to their sum modulo v , and choose one class of the partition to form the blocks of the packing. The basic strategy dates back at least a century to Bussey [6], and perhaps earlier.

This is not a mere theoretical curiosity; as Chen et al. observe in [8], declustered-parity RAIDs do not in practice need to have their loads perfectly balanced, so one may omit some blocks from the design.

Theorem 6 *Let t and v be integers with $v > \binom{t+2}{2} + \binom{t+1}{2}$ so that v and $t + 1$ are relatively prime. For each of the following statements, there exists a t -(v , $t + 1$, 1) packing D on elements $\mathbb{Z}_v$, with point labeling rk being the identity function, having $\frac{\binom{v}{t+1}}{v} = \frac{v-t}{v} \frac{\binom{v}{t}}{\binom{t+1}{t}}$ blocks.*

- (1) $\text{MinSum}(D, rk) = v + \sigma$ and $\text{MaxSum}(D, rk) = tv + \sigma$ whenever $-\binom{t+2}{2} + 1 \leq \sigma < \binom{t+1}{2}$.
- (2) $\text{MinSum}(D, rk) = v + \binom{t+1}{2} - 1$.
- (3) $\text{MaxSum}(D, rk) = tv - \binom{t+2}{2} + 1$.
- (4) $\text{DiffSum}(D, rk) = (t - 1)v$.
- (5) $\text{RatioSum}(D, rk) = \frac{tv + \binom{t+1}{2} - 1}{v + \binom{t+1}{2} - 1}$.

Proof It suffices to prove statement (1); the other results follow directly from it.

Partition all $(t + 1)$ -subsets of $\mathbb{Z}_v$ into v classes $\{\mathcal{B}_\sigma : 0 \leq \sigma < v\}$ by placing set $S = \{x_1, \dots, x_{t+1}\}$ in class $\mathcal{B}_\sigma$ if and only if $\sigma \equiv \sum_{i=1}^{t+1} x_i \pmod{v}$. Because for any t -subset T of $\mathbb{Z}_v$ and each σ with $0 \leq \sigma < v$ there is a unique element s for which $\sigma \equiv s + \sum_{x \in T} x \pmod{v}$, each $\mathcal{B}_\sigma$ is a t -(v , $t + 1$, 1) packing.

Without restrictions on v , these v packings need not have the same number of blocks. We now use the restriction that $(v, t + 1) = 1$. Consider the orbits of $(t + 1)$ -subsets of $\mathbb{Z}_v$ under the cyclic action of $\mathbb{Z}_v$. When S is a $(t + 1)$ -subset of $\mathbb{Z}_v$ with sum σ , let $S + \alpha$ be the subset of $\mathbb{Z}_v$ obtained by adding α (modulo v) to each element in S . Then the orbit containing S is $\{S + \alpha : 0 \leq \alpha < v\}$. For $0 \leq \alpha < v$, the sum of $S + \alpha$ is $\sigma + (t + 1)\alpha \pmod{v}$. Now if $S + \alpha$ and $S + \beta$ have the same sum modulo v , $(t + 1)\alpha \equiv (t + 1)\beta \pmod{v}$, which can happen only when $\alpha \equiv \beta \pmod{v}$. Hence every orbit contains exactly v blocks, one in each of the v classes, and therefore each $\mathcal{B}_\sigma$ contains $\frac{\binom{v}{t+1}}{v}$ blocks.

Now we prove statement (1). First we treat the cases when $\sigma \geq 0$. Choose σ so that $0 \leq \sigma < \binom{t+1}{2}$ and consider the packing $D = (\mathbb{Z}_v, \mathcal{B}_\sigma)$. Suppose to the contrary that S is a $(t + 1)$ -subset of $\mathbb{Z}_v$ with smallest sum $\tau < v + \sigma$. When $\tau \equiv \sigma \pmod{v}$ and $\tau < v + \sigma$, it must happen that $\sigma = \tau = \sum_{x \in S} x$. But $\sum_{x \in S} x \geq \sum_{i=0}^t i = \binom{t+1}{2} > \sigma$, which is a contradiction. Hence $\text{MinSum}(D, rk) \geq v + \sigma$. Because $\sigma \geq 0$, $tv + \sigma$ is the largest integer less than $(t + 1)v$ that is congruent to σ modulo v , and hence $\text{MaxSum}(D, rk) \leq tv + \sigma$.

Next we address the cases when $-\binom{t+2}{2} + 1 \leq \sigma < 0$. Let $\omega = v + \sigma$, and consider the packing $D = (\mathbb{Z}_v, \mathcal{B}_\omega)$. Then $\text{MinSum}(D, rk) \geq \omega = v + \sigma < v$, because of the congruence requirement. For $\text{MaxSum}(D, rk)$, suppose to the contrary that S is a $(t + 1)$ -subset of $\mathbb{Z}_v$ with largest sum $\tau > tv + \sigma = (t - 1)v + \omega$. Then $\tau = tv + \omega$. Now

$\Sigma_{i=1}^{t+1} (v-i) = (t+1)v - \binom{t+2}{2} \geq \Sigma_{x \in S} x$. Hence $\omega \leq v - \binom{t+2}{2}$ so $\sigma \leq -\binom{t+2}{2}$, which is a contradiction.

Statements (2), (4), and (5) follow by taking $\sigma = \binom{t+1}{2} - 1$. Statement (3) follows by taking $\sigma = -\binom{t+2}{2} + 1$. $\square$

Not surprisingly, the packings so produced contain large independent sets. For example, when $\sigma = 0$, the elements $\{0, \dots, \lfloor \frac{v}{t+1} \rfloor\}$ form an independent set.

Theorem 6 yields packings that are dense in the following sense. When an $S(t, t+1, v)$ exists, it has $\binom{v}{t+1}$ blocks; the packings considered have a $\frac{v-t}{v}$ fraction of this number. Hence for fixed t the fraction of t -sets left uncovered by the packing approaches 0 as $v \rightarrow \infty$. Moreover, the bounds established for dense t -($v, t+1, 1$) packings on MinSum and MaxSum match the values from Theorem 1 (which are best for Steiner systems). On the other hand, as $v \rightarrow \infty$ and t is fixed, the ratio of DiffSum of the packing to the bound approaches 1, and the RatioSum approaches its bound of $t-1$. By generalizing to partial systems, Theorem 6 applies to all parameters that are large enough, whether or not an $S(t, t+1, v)$ exists.

Although Theorem 6 establishes a DiffSum of $(t-1)v$ for certain dense t -($v, t+1, 1$) packings, this may not be the best possible, as Theorem 1 ensures only that $(v-k)(t-1)$ is a lower bound on the DiffSum. Theorem 7 gives evidence that the bound may not be the best possible, by producing a packing that achieves a smaller DiffSum than that of Theorem 6 when $t=3$, but is nearly as dense.

Theorem 7 *When $v > 18$ is even, there is a 3 -($v, 4, 1$) packing D with $\frac{v-4}{v-1} \binom{v}{3}$ blocks, having $\text{MinSum}(D) \geq v+2$, $\text{MaxSum}(D) \leq 3v-6$, and hence $\text{DiffSum}(D) \leq 2v-8$.*

Proof Write $v = 2s$. We form D on elements $\{0, \dots, 2s-1\}$, with blocks

1. $\{\{a, b, c, s+d\} : 0 \leq a < b < c < s, 0 \leq d < s, a+b+c+d \equiv 2 \pmod{s}\}$, and
2. $\{\{s+a, s+b, s+c, d\} : 0 \leq a < b < c < s, 0 \leq d < s, a+b+c+d \equiv s-6 \pmod{s}\}$.

This forms a 3 -($v, 4, 1$) packing with the specified number of blocks. Because $a+b+c+d \in \{s+2, 2s+2, 3s+2\}$, blocks of the first class have sum in $\{2s+2, 3s+2, 4s+2\}$. Similarly, because $a+b+c+d \in \{s-6, 2s-6, 3s-6\}$, blocks of the second class have sum in $\{4s-6, 5s-6, 6s-6\}$. Hence $\text{MinSum}(D) \geq 2s+2 = v+2$ and $\text{MaxSum}(D) \leq 6s-6 = 3v-6$. $\square$

6 Sums and Steiner triple systems

For the intended applications in storage systems, it remains desirable to employ a Steiner system, rather than a dense packing, when possible. In what follows, we extend Theorem 6 to produce Steiner triple systems in which the sum metrics are close to optimal.

Building on the construction in Theorem 6, Schreiber [33] and Wilson [40] demonstrate that for certain values of v , the packing can be completed to an STS(v) (see also [21,31]). Their interest was to construct so-called ‘large sets’ of Steiner triple systems, so they imposed a stronger condition that we need. We consider the cyclic group $\mathbb{Z}_n$ for odd n . When x and n are relatively prime, the elements of $\mathbb{Z}_n$ are partitioned into classes under multiplication by x : a and b are in the same class if and only if $a \equiv x^\ell b \pmod{n}$ for some nonnegative integer ℓ . These classes are the *cycles* of $\mathbb{Z}_n$ for x .

When $n \equiv 1, 5 \pmod{6}$, no cycle has size two because for every nonzero $x \in \mathbb{Z}_n$, $x \equiv (-2)^2 x \pmod{n}$ requires that $1 \equiv 4 \pmod{p}$ for some prime divisor p of n . To treat the labeling and block sums, we employ a technical lemma:

Lemma 6 *Let $n \equiv 1, 5 \pmod{6}$. Every pair in $\{a, b\} : a, b \in \mathbb{Z}_n \setminus \{0\}, b \equiv -2a \pmod{n}\}$ has $(n+1)/2 \leq a+b \leq (n-1)/2 + n$.*

Proof Consider such a pair $a, b \in \mathbb{Z}_n$ with $b \equiv -2a \pmod{n}$. We examine two cases:

Case 1: $1 \leq a \leq (n-1)/2$. Then $b = n - 2a$ and hence $(n+1)/2 \leq a+b \leq n-1$.

Case 2: $(n+1)/2 \leq a \leq n-1$: Then $b = 2n - 2a$ and hence $n+1 \leq a+b \leq n+(n-1)/2$. $\square$

Now we re-prove the Schreiber-Wilson result, in order to focus on the block sums. (In [33,40], the $STS(v)$ is constructed, but the point labelling is not.)

Theorem 8 *Let $v \equiv 1, 3 \pmod{6}$. Suppose that for every nonzero $x \in \mathbb{Z}_{v-2}$, the cycle for -2 containing x has even size. Then there is an $STS(v)$, D , with $\text{MinSum}(D) \geq v-2$, $\text{MaxSum}(D) \leq 2v+2$, and hence $\text{DiffSum}(D) \leq v+4$ and $\text{RatioSum}(D) \leq \frac{2v+2}{v-2}$.*

Proof Let $n = v-2$. Using the proof of Theorem 6, construct a $2-(v, 3, 1)$ packing $\mathcal{B}_0$ on $\mathbb{Z}_{v-2}$ (points $v-2$ and $v-1$ appear in no triples). Each triple in $\mathcal{B}_0$ has sum $v-2$ or $2v-4$ at present. The pairs left uncovered on $\mathbb{Z}_{v-2}$ by any triple are $\mathcal{E}_0 = \{\{x, -2x\} : x \in \mathbb{Z}_{v-2} \setminus \{0\}\}$, each having sum between $(v-1)/2$ and $(v-3)/2 + v-2$ by Lemma 6.

Because the cycle for -2 containing x has even size for every nonzero $x \in \mathbb{Z}_n$, the pairs in $\mathcal{E}_0$ can be partitioned into two 1-factors, F_1 and F_2 , on $\mathbb{Z}_n \setminus \{0\}$.

To form an $STS(v)$ on $\mathbb{Z}_v$ with block set $\mathcal{C}$, employ the mapping $\phi : \mathbb{Z}_{v-2} \mapsto \mathbb{Z}_v \setminus \{(v-1)/2, (v+1)/2\}$ defined by $\phi(x) = x$ when $0 \leq x \leq (v-3)/2$ and $\phi(x) = x+2$ when $(v-1)/2 \leq x < v-2$. Then $\mathcal{C}$ is formed as follows.

- (1) When $\{x, y, z\} \in \mathcal{B}_0$, place $\{\phi(x), \phi(y), \phi(z)\}$ in $\mathcal{C}$;
- (2) For $i = 1, 2$, when $\{x, y\} \in F_i$, place $\{(v-3+2i)/2, \phi(x), \phi(y)\}$ in $\mathcal{C}$;
- (3) Place $\{0, (v-1)/2, (v+1)/2\}$ in $\mathcal{C}$.

Triples of $\mathcal{B}_0$ have sum $v-2$ or $2v-4$, so triples of type (1) in $\mathcal{C}$ have sum between $v-2$ and $v+2$, or between $2v-2$ and $2v+2$. A pair $\{x, y\} \in \mathcal{E}_0$ has $(v-1)/2 \leq x+y \leq (v-1)/2 + (v-3)$. Applying ϕ , we have $(v-1)/2 \leq \phi(x) + \phi(y) \leq (v-1)/2 + (v+1)$. Hence, each triple of type (2) in $\mathcal{C}$ has sum at least $v-1$ and at most $2v+1$. Finally, the single type (3) block has sum v . $\square$

Unlike the point labelings in [12], the labeling for the Schreiber-Wilson construction in Theorem 8 need not achieve the largest MinSum or smallest MaxSum. Nevertheless it yields a substantial improvement on earlier constructions with respect to the DiffSum and RatioSum, within an additive constant of the best bound possible for the DiffSum. Unfortunately, Theorem 8 requires that the order of $-2 \pmod{p}$ be even, and so applies to an infinite set of orders (see [34]) but not all admissible ones. (For example, it does not apply when $v-2$ is a multiple of 11, 59, 83, 107, 131, 179, 227, 251, 281, or 347, as in these cases the order of -2 is odd.) We remedy this next, using a result from [5], but obtaining slightly weaker bounds.

Theorem 9 *Whenever $v \equiv 1, 3 \pmod{6}$, there is an $STS(v)$, D , with $\text{MinSum}(D) \geq v-5$, $\text{MaxSum}(D) \leq 2v+2$, and hence $\text{DiffSum}(D) \leq v+7$ and $\text{RatioSum}(D) \leq \frac{2v+2}{v-5}$.*

Proof Form $\mathcal{B}_0$ over $\mathbb{Z}_{v-2}$ as in the proof of Theorem 8. Remove element 0 as well as all triples $\{\{0, x, v-2-x\} : 1 \leq x \leq (v-3)/2\}$ to form $\mathcal{D}_0$. Let $\mathcal{E}_0$ be the set of pairs

on $\mathbb{Z}_{v-2} \setminus \{0\}$ not covered by a triple of $\mathcal{D}_0$. The pairs in $\mathcal{E}_0$ form a 3-regular graph G on $\mathbb{Z}_{v-2} \setminus \{0\}$. By [5, Lemma 9], G can be partitioned into three 1-factors, F_1 , F_2 , and F_3 .

To form the STS(v) on $\mathbb{Z}_v$ with block set $\mathcal{C}$, we employ the mapping $\psi : \mathbb{Z}_{v-2} \setminus \{0\} \mapsto \mathbb{Z}_v \setminus \{(v-3)/2, (v-1)/2, (v+1)/2\}$ defined by $\psi(x) = x - 1$ when $1 \leq x \leq (v-3)/2$ and $\psi(x) = x + 2$ when $(v-1)/2 \leq x < v-2$. Then $\mathcal{C}$ is formed as follows.

- (1) When $\{x, y, z\} \in \mathcal{D}_0$, place $\{\psi(x), \psi(y), \psi(z)\}$ in $\mathcal{C}$;
- (2) For $i = 1, 2, 3$, when $\{x, y\} \in F_i$, place $\{(v-5+2i)/2, \psi(x), \psi(y)\}$ in $\mathcal{C}$;
- (3) Place $\{(v-3)/2, (v-1)/2, (v+1)/2\}$ in $\mathcal{C}$.

Triples of $\mathcal{B}_0$ have sum $v-2$ or $2v-4$, so triples of type (1) in $\mathcal{C}$ have sum between $v-5$ and $v-2$, or between $2v-1$ and $2v+2$. By Lemma 6, a pair $\{x, y\} \in \mathcal{E}_0$ has $(v-1)/2 \leq x+y \leq (v-1)/2 + (v-3)$. Applying ψ , we have $(v-1)/2 - 2 \leq \psi(x) + \psi(y) \leq (v-1)/2 + (v+1)$. Hence each triple of type (2) in $\mathcal{C}$ has sum at least $v-4$ and at most $2v+1$. The block of type (3) in $\mathcal{C}$ has sum $\frac{3v-3}{2}$. $\square$

Although the bounds are slightly weaker, Theorem 9 applies to all admissible orders for Steiner triple systems. In conjunction with Theorem 2, for all $v \equiv 1, 3 \pmod{6}$ with $v \geq 13$ one has $v+1 \leq \text{DiffSum}(2, 3, v) \leq v+7$ and $2 + \frac{1}{v} \leq \text{RatioSum}(2, 3, v) \leq 2 + \frac{12}{v-5}$.

For relatively small orders, one can search for point labelings with specified MinSum m and MaxSum M by starting with an empty packing. Repeatedly choose a pair that is in the fewest triples within the sum range that can still be added to the packing (placing no pair in two or more triples), and extend the packing to form packings by adding each of the possible triples containing the pair in turn. In our experience, often one is forced to add a specific triple for the chosen pair, and occasionally there are a few candidate triples, each resulting in a larger packing. Despite the limited number of candidates encountered, this yields an exhaustive search when run to completion. Using this approach, we constructed $S(2, 3, v)$ s with specified MinSum and MaxSum, as shown next:

Order v	MinSum	MaxSum	DiffSum	RatioSum
7	$v-1$	$2v-1$	v	$2 + \frac{1}{v-1}$
9	v	$2v$	v	2
13, 15, 19, 21, 25, 27	$v-1$	$2v$	$v+1$	$2 + \frac{2}{v-1}$
7, 15, 19, 21, 27	v	$2v+1$	$v+1$	$2 + \frac{1}{v}$
13, 25	v	$2v+2$	$v+2$	$2 + \frac{2}{v}$

It appears plausible that $\text{DiffSum}(2, 3, v) = v+1$ when $v \geq 13$. It also appears plausible that $\text{RatioSum}(2, 3, v) \in \{2 + \frac{1}{v}, 2 + \frac{2}{v}\}$ for every $v \neq 9$, but there is insufficient data to speculate on when it takes the larger value and when the smaller.

7 Concluding remarks

Because Theorem 6 achieves a DiffSum of $(t-1)v$ for dense t -($v, t+1, 1$) packings, one might hope that this difference can be realized for $S(t, t+1, v)$ Steiner systems. However, Theorem 2 establishes that this does not happen when $t = 2$ unless $v \in \{7, 9\}$, although

Theorem 9 is within an additive constant. The situation when $t = 3$ appears to be quite different. There is an $S(3, 4, 8)$ with blocks

$$\{0127, 0136, 0145, 0235, 0246, 0347, 0567, \\ 1234, 1256, 1357, 1467, 2367, 2457, 3456\},$$

having MinSum 10 and MaxSum 18. Adapting the construction in [13,30], one can produce an $S(3, 4, v)$ with MinSum $v + 2$, MaxSum $3v - 6$, and hence DiffSum $2v - 8$ whenever v is a power of 2. In these cases, the upper bound on the MinSum and the lower bound on the MaxSum from Theorem 1 are met simultaneously. We do not expect this to happen for all orders, because the smallest DiffSum for an $S(3, 4, v)$ when $v \in \{10, 14\}$ appears to arise from systems with MinSum $v + 1$ and MaxSum $3v - 5$. It may happen that for every admissible v , an $S(3, 4, v)$ with DiffSum strictly smaller than $2v$ exists. If so, completing the packing from Theorem 6 could not yield the smallest DiffSum. Nevertheless, the structure of independent sets must underlie appropriate constructions.

References

1. Ajtai M., Komlós J., Pintz J., Spencer J., Szemerédi E.: Extremal uncrowded hypergraphs. *J. Comb. Theory Ser. A* **32**(3), 321–335 (1982).
2. Bertram-Kretzberg C., Lefmann H.: The algorithmic aspects of uncrowded hypergraphs. *SIAM J. Comput.* **29**(1), 201–230 (1999).
3. Bose R.C.: On the construction of balanced incomplete block designs. *Ann. Eugenics* **9**(4), 353–399 (1939).
4. Brummond, W.M.: Kirkman systems that attain the upper bound on the minimum block sum, for access balancing in distributed storage (2019). [arXiv:1906.02157](https://arxiv.org/abs/1906.02157).
5. Bryant D., Colbourn C.J., Horsley D., Wanless I.M.: Steiner triple systems with high chromatic index. *SIAM J. Discret. Math.* **31**(4), 2603–2611 (2017).
6. Bussey W.H.: The tactical problem of Steiner. *Am. Math. Monthly* **21**(1), 3–12 (1914).
7. Chee Y.M., Colbourn C.J., Dau H., Gabrys R., Ling A.C.H., Lusi D., Milenkovic O.: Access balancing in storage systems by labeling partial Steiner systems. In: 2020 IEEE International Symposium on Information Theory (2020).
8. Chen P.M., Lee E.K., Gibson G.A., Katz R.H., Patterson D.A.: RAID: high-performance, reliable secondary storage. *ACM Comput. Surv.* **26**(2), 145–185 (1994).
9. Cidon A., Rumble S.M., Stutsman R., Katti S., Ousterhout J.K., Rosenblum M.: Copysets: Reducing the frequency of data loss in cloud storage. In: *Usenix Annual Technical Conference*, pp. 37–48 (2013).
10. Colbourn C.J., Rosa A.: *Triple Systems*. Oxford Mathematical Monographs Clarendon Press, Oxford (1999).
11. Colbourn C.J., Colbourn M.J., Phelps K.T., Rödl V.: Colouring Steiner quadruple systems. *Discret. Appl. Math.* **4**(2), 103–111 (1982).
12. Dau H., Milenkovic O.: MaxMinSum Steiner systems for access balancing in distributed storage. *SIAM J. Discret. Math.* **32**(3), 1644–1671 (2018).
13. Doyen J., Vandensavel M.: Non isomorphic Steiner quadruple systems. *Bull. Soc. Math. Belg.* **23**, 393–410 (1971).
14. Duke R.A., Lefmann H., Rödl V.: On uncrowded hypergraphs. *Random Struct. Algorithms* **6**(2–3), 209–212 (1995).
15. El Rouayheb S., Ramchandran K.: Fractional repetition codes for repair in distributed storage systems. In: 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1510–1517. IEEE (2010).
16. Erdős P., Hajnal A.: On chromatic number of graphs and set-systems. *Acta Math. Acad. Sci. Hungar* **17**, 61–99 (1966).
17. Eustis A., Verstraëte J.: On the independence number of Steiner systems. *Comb. Probab. Comput.* **22**(2), 241–252 (2013).
18. Fazeli A., Vardy A., Yaakobi E.: Codes for distributed PIR with low storage overhead. In: 2015 IEEE International Symposium on Information Theory (ISIT), pp. 2852–2856. IEEE (2015).

19. Fundia A.D.: Derandomizing Chebyshev's inequality to find independent sets in uncrowded hypergraphs. *Random Struct. Algorithms* **8**(2), 131–147 (1996).
20. Grable D.A., Phelps K.T., Rödl V.: The minimum independence number for designs. *Combinatorica* **15**(2), 175–185 (1995).
21. Grannell M.J., Griggs T.S., Phelan J.S.: A new look at an old construction for Steiner triple systems. *ARS Comb.* **25**(A), 55–60 (1988).
22. Hanani H.: A note on Steiner triple systems. *Math. Scand.* **8**(1), 154–156 (1960).
23. Holland M., Gibson G.A.: Parity declustering for continuous operation in redundant disk arrays. *SIGPLAN Not.* **27**(9), 23–35 (1992).
24. Ji L.: A construction for 2-chromatic Steiner quadruple systems. *Eur. J. Comb.* **28**(6), 1832–1838 (2007).
25. Kostochka A., Mubayi D., Rödl V., Tetali P.: On the chromatic number of set systems. *Random Struct. Algorithms* **19**(2), 87–98 (2001).
26. Kostochka A., Mubayi D., Verstraëte J.: On independent sets in hypergraphs. *Random Struct. Algorithms* **44**(2), 224–239 (2014).
27. Lovász L.: Coverings and coloring of hypergraphs. In: *Proceedings of the Fourth Southeastern Conference on Combinatorics, Graph Theory, and Computing*. Florida Atlantic University, Boca Raton, pp. 3–12 (1973).
28. Lusi D., Colbourn C.J.: On the maximum double independence number of Steiner triple systems. *J. Comb. Des.* to appear (2020).
29. Phelps K.T., Rödl V.: Steiner triple systems with minimum independence number. *ARS Comb.* **21**, 167–172 (1986).
30. Phelps K.T., Rosa A.: 2-Chromatic Steiner quadruple systems. *Eur. J. Comb.* **1**(3), 253–258 (1980).
31. Quiring D.: A construction of disjoint Steiner triple systems. *J. Comb. Theory Ser. A* **27**(3), 407–408 (1979).
32. Rödl V., Šiňajová E.: Note on independent sets in Steiner systems. *Random Struct. Algorithms* **5**(1), 183–190 (1994).
33. Schreiber S.: Covering all triples on n marks by disjoint Steiner systems. *J. Comb. Theory Ser. A* **15**(3), 347–350 (1973).
34. Schreiber S.: Some mappings of periodic groups. *Israel J. Math.* **17**, 117–123 (1974).
35. Silberstein N., Etzion T.: Optimal fractional repetition codes based on graphs and designs. *IEEE Trans. Inf. Theory* **61**(8), 4164–4180 (2015).
36. Silberstein N., Gál A.: Optimal combinatorial batch codes based on block designs. *Des. Codes Cryptogr.* **78**(2), 409–424 (2016).
37. Skolem T.: Some remarks on the triple systems of Steiner. *Math. Scand.* **1**, 273–280 (1959).
38. Spencer J.: Turán's theorem for k -graphs. *Discret. Math.* **2**, 183–186 (1972).
39. Tian F., Liu Z.L.: Bounding the independence number in some (n, k, ℓ, λ) -hypergraphs. *Graphs Comb.* **34**(5), 845–861 (2018).
40. Wilson R.M.: Some Partitions of All Triples into Steiner Triple Systems, pp. 267–277. *Hypergraph Seminar* Springer, New York (1974).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Yeow Meng Chee¹ · Charles J. Colbourn²  · Hoang Dau³ · Ryan Gabrys⁴ · Alan C. H. Ling⁵ · Dylan Lusi² · Olgica Milenkovic⁴

Yeow Meng Chee
pvocym@nus.edu.sg

Hoang Dau
sonhoang.dau@rmit.edu.au

Ryan Gabrys
ryan.gabrys@gmail.com

Alan C. H. Ling
aling@emba.uvm.edu

Dylan Lusi
dlusi@asu.edu

Olga Milenkovic
milenkov@illinois.edu

- ¹ Industrial Systems Engineering and Management, National University of Singapore, Singapore, Singapore
- ² Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
- ³ Computer Science and Information Technology, RMIT University, Melbourne, Australia
- ⁴ Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA
- ⁵ Computer Science, University of Vermont, Burlington, VT, USA