The Card Game SET

Benjamin Lent Davis and
Diane Maclagan

This column is a place for those bits of contagious mathematics that travel from person to person in the community, because they are so elegant, suprising, or appealing that one has an urge to pass them on.

Contributions are most welcome.

Please send all submissions to the Mathematical Entertainments Editor, **Ravi Vakil,** Stanford University, Department of Mathematics, Bldg. 380, Stanford, CA 94305-2125, USA
e-mail: vakil@math.stanford.edu

The card game called SET[1] is an extremely addictive, fast-paced game found in toy stores nationwide. Although children often beat adults, the game has a rich mathematical structure linking it to the combinatorics of finite affine and projective spaces and the theory of error-correcting codes. Last year an unexpected connection to Fourier analysis was used to settle a basic question directly related to the game of SET, and many related questions remain open.

The game of SET was invented by population geneticist Marsha Jean Falco in 1974. She was studying epilepsy in German Shepherds and began representing genetic data on the dogs by drawing symbols on cards and then searching for patterns in the data. After realizing the potential as a challenging puzzle, with encouragement from friends and family she developed and marketed the card game. Since then, SET has become a huge hit both inside and outside the mathematical community.

SET is played with a special deck of cards (Fig. 1). Each SET card displays a design with four attributes—number, shading, color, and shape—and each attribute assumes one of three possible values, given in Table 1.

Table 1

Number:	{One, Two, Three}
Shading:	{Solid, Striped, Open}
Color:	{Red, Green, Purple}
Shape:	{Ovals, Squiggles, Diamonds}

A SET deck has eighty-one cards, one for each possible combination of attributes. The goal of the game is to find collections of cards satisfying the following rule.

The SET rule: Three cards are called a SET if, with respect to each of the four attributes, the cards are either all the same or all different.

For example, Figure 2 illustrates a green SET. All cards have the same shape (ovals), the same color (green), and the same shading (solid), and each card has a different number of ovals. On the other hand, Figure 3, also green, fails to be a SET, because there are two oval cards and one squiggle card. Thus the cards are neither all the same nor all different with respect to the shape attribute.

To play the game, the SET deck is shuffled and twelve cards are dealt to a table face-up (Fig. 4). All players simultaneously search for SETS. The first player to locate a SET removes it, and three new cards are dealt. The player with the most SETS after all the cards have been dealt is the winner.

Occasionally, there will not be any SETS among the twelve cards initially dealt. To remedy this, three extra cards are dealt. This is repeated until a SET makes an appearance. This prompts the following SET-theoretic question.

Question. How many cards must be dealt to guarantee the presence of a SET?

Figure 5 shows a collection of twenty cards containing no SETS. A brute-force computer search shows that this is as large as possible, as any collection of twenty-one cards must contain a SET.

There is a wonderful geometric reformulation of this Question as follows. Let $\mathbb{F}_3$ be the field with three elements, and consider the vector space

Figure 1. Typical SET cards.

[1] SET is a trademark of SET Enterprises, Inc. The SET cards are depicted here with permission. SET game play is protected intellectual property.

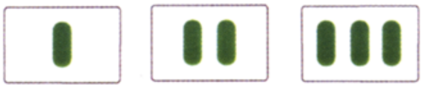

Figure 2. A SET.

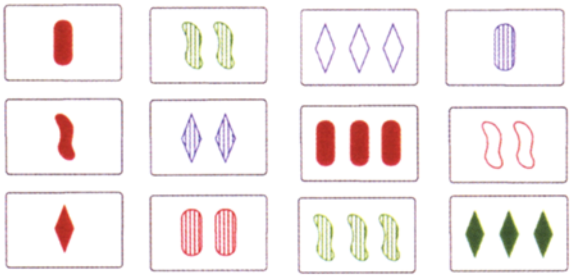

Figure 4. Can you find all five SETS? (Or all eight for those readers with black-and-white photocopies.)

$\mathbb{F}_3^4$. A point of $\mathbb{F}_3^4$ is a 4-tuple of the form (x_1, x_2, x_3, x_4), where each coordinate assumes one of three possible values. Using the table of SET attributes (Table 1), SET cards correspond to points of $\mathbb{F}_3^4$, and vice-versa.

$(2, 1, 3, 2) \longleftrightarrow$ Two Solid Purple Squiggles $\longleftrightarrow$  .

Under this correspondence, three cards form a SET if and only if the three associated points of $\mathbb{F}_3^4$ are collinear. To see this, notice that if α, β, γ are three elements of $\mathbb{F}_3$, then $\alpha + \beta + \gamma = 0$ if and only if $\alpha = \beta = \gamma$ or $\{\alpha, \beta, \gamma\} = \{0, 1, 2\}$. This means that the vectors a, b, and c are either all the same or all different with respect to each coordinate exactly when $a + b + c = 0$. Now $a + b + c = 0$ in $\mathbb{F}_3^4$ means that $a - b = b - c$, so the three points are collinear. Note that this argument works when $\mathbb{F}_3^4$ is replaced by $\mathbb{F}_3^d$ for any d. From this point of view, players of SET are searching for lines contained in a subset of $\mathbb{F}_3^4$. We summarize this rule as follows.

The Affine Collinearity Rule. Three points $a, b, c \in \mathbb{F}_3^d$ represent collinear points if and only if $a + b + c = 0$.

We define a *d-cap* to be a subset of $\mathbb{F}_3^d$ not containing any lines, and ask the following.

Equivalent Question. What is the maximum possible size of a cap in $\mathbb{F}_3^4$?

In this form the question was first answered, without using computers, by Giuseppe Pellegrino [19] in 1971. Note that this was three years before the game of SET was invented! He actually answered a more general question about "projective SET," which we explain in the last section.

Although SET cards are described by four attributes, from a mathematical perspective there is nothing sacred about the number four. We can play a three-attribute version of SET, for example by playing with only the red cards. Or we can play a five-attribute version of SET by using scratch-and-sniff SET cards with three different odors. In general, we define an *affine* SET *game of dimension d* to be a card game with one card for each point of $\mathbb{F}_3^d$, where three cards form a SET if the corresponding points are collinear.

A cap of the maximum possible size is called a *maximal cap*. It is natural to ask for the size of a maximal cap in $\mathbb{F}_3^d$, as a function of the dimension d. We denote this number by a_d, and the known values are given in Table 2.

Table 2

d	1	2	3	4	5	6	7
a_d	2	4	9	20	45	$112 \le a_6 \le 114$	?

The values of a_d in dimensions four and below can be found by exhaustive computer search. The search space becomes unmanageably large starting in dimension five. Yves Edel, Sandy Ferret, Ivan Landjev, and Leo Storme recently created quite a stir by announcing the solution in dimension five [6]. We shall spend some time working our way up to their solution.

There are many other possible generalizations of the game of SET. For example, we could add another color, shape, form of shading, and number to the cards, to make the cards correspond to points of $\mathbb{F}_4^4$. Here, however, several choices need to be made about the SET rule. Is a SET a collection of cards where every attribute is all the same or all different, or is it a collection of collinear points? In $\mathbb{F}_4^4$, there are four points on a line, so do we require three or four collinear points to form a SET? Furthermore, if we choose the collinearity criterion, then collinearity of SET cards is sensitive to the choice of which color, shape, etc. corresponds to which element of $\mathbb{F}_4^4$. Because of these complications we will restrict our attention here to caps (line-free collections) in $\mathbb{F}_3^d$.

We can exhibit caps graphically using the following scheme. Let us consider the case of dimension $d = 2$. A two-attribute version of SET may be realized by playing with only the red ovals. The vector space $\mathbb{F}_3^2$ can be

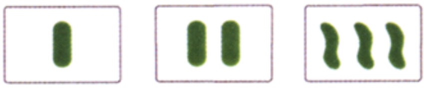

Figure 3. Not a SET.

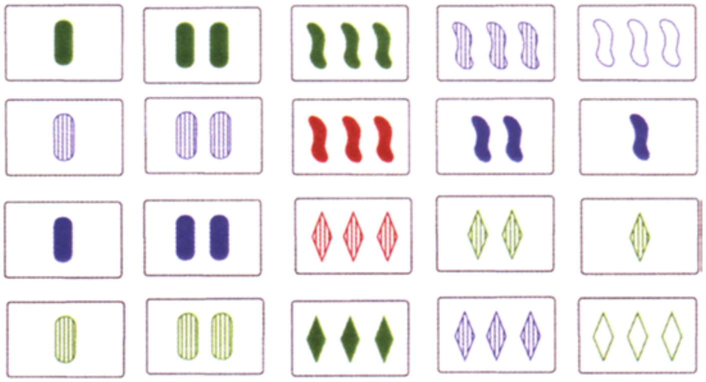

Figure 5. Twenty cards without a SET.

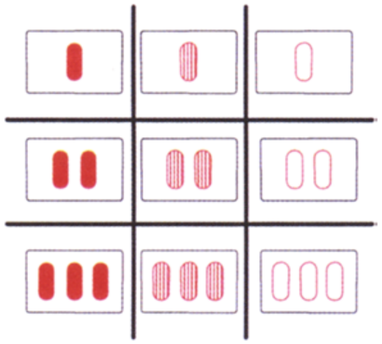

Figure 6. The correspondence between 2-attribute SET and $\mathbb{F}_3^2$.

graphically represented as a tic-tac-toe board as in Figure 6. We indicate a subset S of $\mathbb{F}_3^2$ by drawing an "X" in each square of the tic-tac-toe board corresponding to a point of S. The lines contained in S are almost plain to see: most of them appear as winning tic-tac-toes, while a few meet an edge of the board and "loop around" to the opposite edge. Check that the two lines in Figure 7 correspond to SETS in Figure 4.

Figure 8 contains pictures of some low-dimensional maximal caps. In dimensions one through four, the caps are visibly symmetrical, and each cap contains embedded copies of the maximal caps in lower dimensions. No such pattern is visible in the diagram of the 5-cap. It is natural to ask if the maximal caps in Figure 8 are the only ones in each dimension. In a trivial sense, the answer is 'no', since we can make a new cap by permuting the colors of an old cap. There are many other permutations of $\mathbb{F}_3^d$ guaranteed to produce new caps from old. Permutations of $\mathbb{F}_3^d$ taking caps to caps are exactly

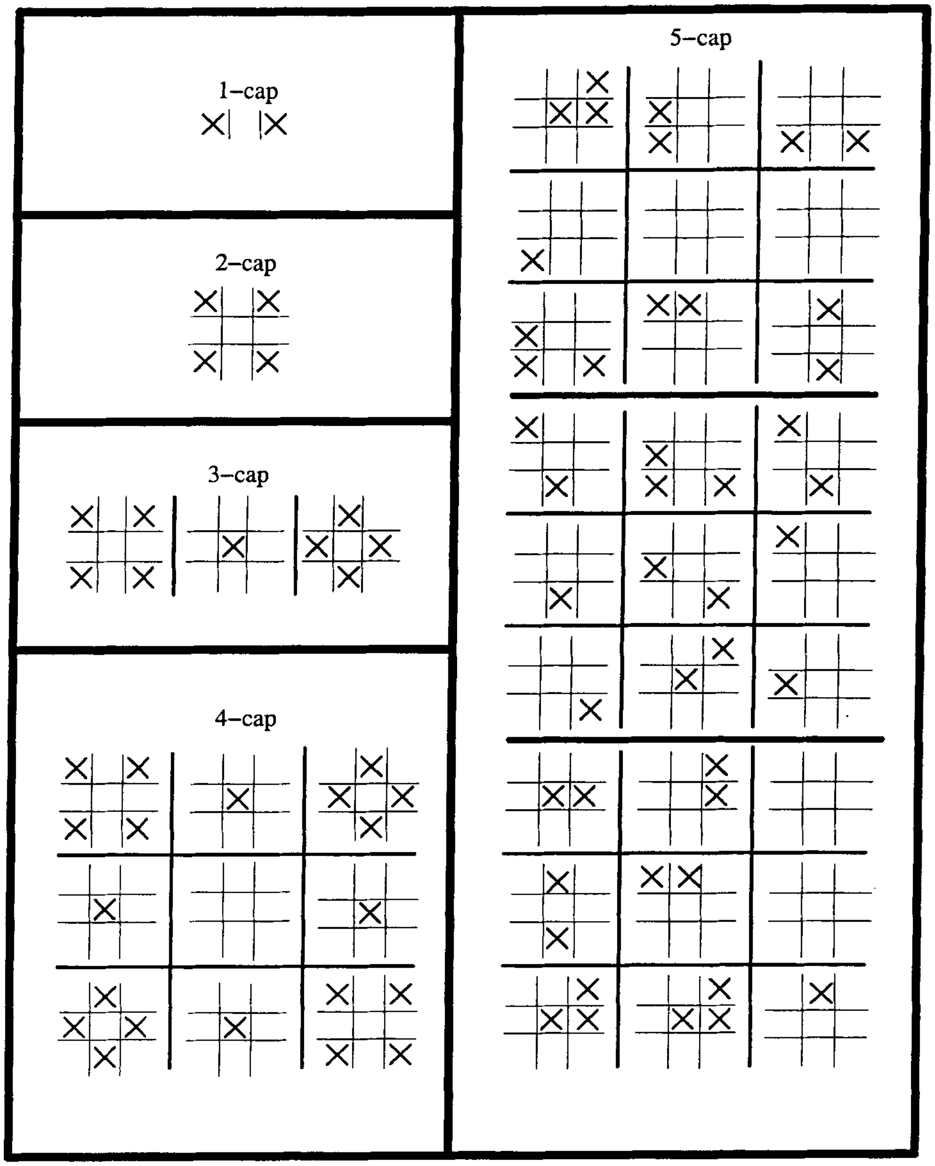

Figure 8. Low-dimensional maximal caps.

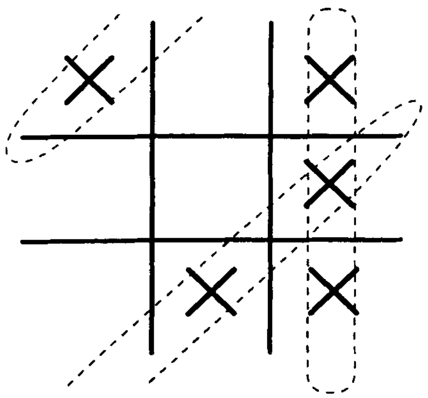

Figure 7. This collection of points contains two lines which are indicated by dashed curves.

those taking lines to lines, and such a permutation is called an *affine transformation.* Another characterization of affine transformations is that they are the permutations of $\mathbb{F}_3^d$ of the form

$$\sigma(v) = Av + b,$$

where A is an invertible $d \times d$-matrix with entries in $\mathbb{F}_3$, b is an arbitrary vector of $\mathbb{F}_3^d$, and v is a vector in $\mathbb{F}_3^d$. We say that two caps are of the same *type* if there is an affine transformation taking one to the other. For example, consider the affine transformation $\sigma(x, y) = (-x - y, -x + y - 1)$ taking a vector $(x, y) \in \mathbb{F}_3^2$ to another vector in $\mathbb{F}_3^2$. Applying this to a 2-cap gives another 2-cap of the same type. This is illustrated in Figure 9, where we have declared the center square of the tic-tac-toe board to be the origin of $\mathbb{F}_3^2$.

It is known that in dimensions five and below there is exactly one type of maximal cap. An affine transformation taking a cap to itself is called a *symmetry* of the cap. Although it is not obvious from Figure 8, the maximal 5-cap does have some symmetries. In fact, its symmetry group is *transitive*, meaning that, given two points of the 5-cap, there is always a symmetry taking one

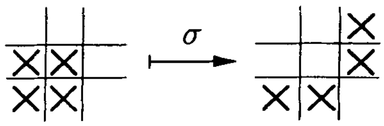

Figure 9. Two 2-caps of the same type.

to the other. Michael Kleber reports that the stabilizer of a point in the 5-cap is the semidihedral group of order 16.

The symmetry group is useful for reducing the number of cases that need to be checked in exhaustive computer searches for maximal caps, thus greatly speeding up run times. To see this idea in action, check out Donald Knuth's SET-theoretic computer programs [17].

Combinatorics

We can make some progress on computing the size of maximal caps using only counting arguments.

Proposition 1. *A maximal 2-cap has four points.*

Proof. We have exhibited a 2-cap with four points. The proof proceeds by contradiction. Suppose that there exists a 2-cap with five points, x_1, x_2, x_3, x_4, x_5. The plane $\mathbb{F}_3^2$ can be decomposed as the union of three horizontal parallel lines as in Figure 10.

Each line contains at most two points of the cap. Thus, there are two horizontal lines that contain two points of the cap, and one line, H, that contains exactly one point of the cap. Without loss of generality, let x_5 be this point. There are exactly four lines in the plane containing the point x_5, which we denote H, L_1, L_2, L_3. This is illustrated in Figure 11.

Since the line H contains none of the points $x_1, \ldots, x_4$, by the pigeon-hole principle two of these points x_r and x_s must lie on one line L_i. This shows that the line L_i contains the points x_r, x_s and x_5, which contradicts the hypothesis that the five points are a cap. □

We can apply the method of Proposition 1 to compute the size of a maximal cap in three dimensions.

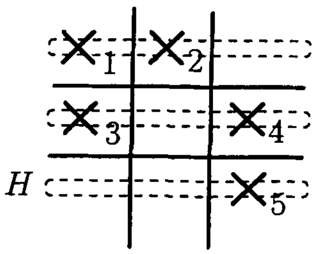

Figure 10. $\mathbb{F}_3^2$ **decomposed as the union of three parallel lines.**

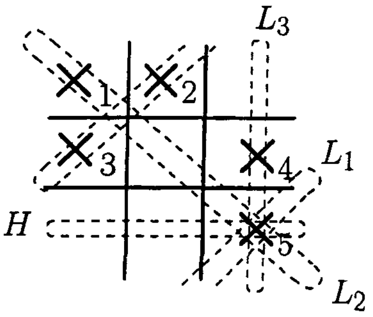

Figure 11. The four lines containing x_5.

Proposition 2. *A maximal 3-cap has nine points.*

Proof. We have exhibited a 3-cap with nine points. The proof proceeds by contradiction. Suppose that there exists a 3-cap with ten points. The space $\mathbb{F}_3^3$ can be decomposed as the union of three parallel planes. Since the intersection of any plane with the 3-cap is a 2-cap, Proposition 1 implies that no plane can contain more than four points of the cap. This means that the plane containing the fewest number of points must contain either two or three points, for if it contained four points we would need twelve points total, and one or zero points would mean at most nine points total. Call this plane H, and note that there are at least seven points of the cap, $x_1, \ldots, x_7$, not contained in H.

Let a and b be two points of the cap on the plane H. There are exactly four planes in the space $\mathbb{F}_3^3$ containing both a and b, which we denote H, M_1, M_2, M_3. Since H does not contain the points $x_1, \ldots, x_7$, by the pigeon-hole principle one of the planes M_i must contain three of these points x_r, x_s, x_t. This shows that the plane M_i contains a total of five points of the cap, which contradicts Proposition 1. □

Unfortunately, this method is not strong enough to prove that $a_4 = 20$. To do this, we employ another time-honored combinatorial technique, namely, counting the same thing in two different ways. By way of introduction, we will give another proof that $a_3 = 9$.

Proposition 3. *A maximal 3-cap has nine points.*

Proof. The proof is again by contradiction. Suppose that there exists a 3-cap C with ten points. The space $\mathbb{F}_3^3$ can be decomposed as the union of three parallel planes, H_1, H_2, H_3 in many different ways. Given such a decomposition, we obtain a triple of numbers,

$$\{|C \cap H_1|,\ |C \cap H_2|,\ |C \cap H_3|\},$$

called the *(unordered) hyperplane triple*, where $|C \cap H_i|$ is the size of $C \cap H_i$. Since a 2-cap has at most $a_2 = 4$ points, the only possible values for a hyperplane triple are $\{4, 4, 2\}$ or $\{4, 3, 3\}$. Let

a = the number of $\{4, 4, 2\}$ hyperplane triples,

b = the number of $\{4, 3, 3\}$ hyperplane triples.

How many different ways are there to decompose $\mathbb{F}_3^3$ as the union of three hyperplanes? On the one hand, there are $a + b$ ways. On the other hand, there is a unique line through the origin of $\mathbb{F}_3^3$ perpendicular to each family of three parallel hyperplanes, and we can count these lines as follows: Any nonzero point determines a line through the origin, and there are $3^3 - 1 = 26$ nonzero points. Since each line contains two nonzero points, there must be $26/2 = 13$ lines through the origin. Thus,

$$a + b = 13.$$

To obtain another equation in a and b, we will count *2-marked planes*, which are pairs of the form $(H, \{x, y\} \subset H \cap C)$, where H is a plane. It can be checked that there are exactly four planes containing any pair of distinct points. This is a special case of Proposition 4 which follows. Thus, there are $4\binom{10}{2} = 180$ 2-marked planes. On the other hand, for each $\{4, 4, 2\}$ hyperplane triple we count $\binom{4}{2} + \binom{4}{2} + \binom{2}{2} = 13$ 2-marked planes, and for each $\{4, 3, 3\}$ hyperplane triple we count $\binom{4}{2} + \binom{3}{2} + \binom{3}{2} = 12$ 2-marked planes. Hence,

$$13a + 12b = 180.$$

The only solution to these equations is $a = 24$, $b = -11$. This is a contradiction since a and b can only take nonnegative values. □

In the proof above we needed to count the number of hyperplanes containing a fixed pair of points, or in other words, containing a fixed line. To apply this method to maximal 4-caps, we will need to solve a generalization of this problem. Define a *k-flat* to be a k-dimensional affine subspace of a vector space.

Proposition 4. *The number of hyperplanes containing a fixed k-flat in* $\mathbb{F}_3^d$ *is given by*

$$\frac{3^{d-k}-1}{2}.$$

Proof. Let K be a k-flat containing the origin. Then the natural map

$$\mathbb{F}_3^{d-k} \to \mathbb{F}_3^d/K \cong \mathbb{F}_3^{d-k}$$

gives a bijection between hyperplanes of $\mathbb{F}_3^d$ containing K and hyperplanes of $\mathbb{F}_3^{d-k}$ containing the origin.

Each hyperplane containing the origin is determined by a nonzero normal vector, and there are exactly two nonzero normal vectors determining each hyperplane. Thus, there are half as many hyperplanes as there are nonzero vectors. Since there are $3^{d-k}-1$ nonzero vectors, there must be $(3^{d-k}-1)/2$ hyperplanes containing the origin. □

This lets us apply the ideas of Proposition 3 to calculate a_4.

Proposition 5. *A maximal* 4*-cap has twenty points.*

Proof. We have exhibited a 4-cap with 20 points. The proof proceeds by contradiction. Suppose that there exists a 4-cap C with 21 points. Let x_{ijk} be the number of $\{i,j,k\}$ hyperplane triples of C. Since a 3-cap has at most $a_3 = 9$ points, there are only 7 possible hyperplane triples:

$$\{i, j, k\} = \{9, 9, 3\}, \{9, 8, 4\}, \{9, 7, 5\}, \{9, 6, 6\}, \{8, 8, 5\}, \{8, 7, 6\}, \{7, 7, 7\}.$$

The number of ways to decompose $\mathbb{F}_3^4$ as a union of three parallel hyperplanes is equal to the number of lines through the origin in $\mathbb{F}_3^4$, which is $(3^4-1)/2 = 40$. Thus,

$$(1)\quad x_{993} + x_{984} + x_{975} + x_{966} + x_{885} + x_{876} + x_{777} = 40.$$

To obtain another equation in x_{ijk}, let us count 2*-marked hyperplanes*, which are pairs of the form $(H, \{x, y\} \subset H \cap C)$, where H is a hyperplane. Using Proposition 4 above, we find that the number of hyperplanes containing a distinct pair of points is 13. Thus, there are $13\binom{21}{2} = 2730$ 2-marked hyperplanes. As in the proof of Proposition 3, there are

$$\left[\binom{9}{2}+\binom{9}{2}+\binom{3}{2}\right]x_{993} + \cdots + \left[\binom{7}{2}+\binom{7}{2}+\binom{7}{2}\right]x_{777}$$

2-marked hyperplanes. Explicitly computing each coefficient above yields the formula

$$(2)\quad 75x_{993} + 70x_{984} + 67x_{975} + 66x_{966} + 66x_{885} + 64x_{876} + 63x_{777} = 2730.$$

To obtain yet another equation in x_{ijk}, let us count 3*-marked hyperplanes*, which are pairs of the form $(H, \{x, y, z\} \subset H \cap C)$, where H is a hyperplane. Notice that, since $\{x, y, z\} \subset C$, the points x, y, and z cannot be collinear. There are 4 hyperplanes containing 3 distinct non-collinear points, thus, there are $4\binom{21}{3} = 5320$ 3-marked hyperplanes. Imitating our count of 2-marked hyperplanes above, we find that

$$(3)\quad 169x_{993} + 144x_{984} + 129x_{975} + 124x_{966} + 122x_{885} + 111x_{876} + 105x_{777} = 5320.$$

We now have three equations in seven variables, and so in principle there could be infinitely many solutions. Fortunately we are only interested in the nonnegative integer solutions. Adding 693 times equation (1) to three times equation (3), and then subtracting off 6 times equation (2), gives

$$5x_{984} + 8x_{975} + 9x_{966} + 3x_{885} + 2x_{876} = 0.$$

The only nonnegative solution to this equation is $x_{984} = x_{975} = x_{966} = x_{885} = x_{876} = 0$. But equation (2) minus 63 times equation (1) is

$$12x_{993} + 7x_{984} + 4x_{975} + 3x_{966} + 3x_{885} + x_{876} = 210.$$

This reduces to $12x_{993} = 210$, which contradicts x_{993} being an integer. □

This proof was improved from a previous version by conversations with Yves Edel. The method of counting marked hyperplanes via hyperplane triples gives the shortest known proof of $a_4 = 20$ that does not use an exhaustive computer search. Unfortunately, a straightforward application of this method fails to show that $a_5 = 45$. Part of the problem is that the new equations counting 4-marked hyperplanes require an additional variable to distinguish between the cases when four points are affinely dependent or independent. In the next section we describe another approach which computes a_5.

The Fourier Transform

The Fourier transform is an immensely useful tool for analyzing problems with associated symmetry groups. It is a natural construction in representation theory, and we refer the reader to the book of Fulton and Harris [7] for more about this fascinating subject. In this section we describe a Fourier transform method originated by Roy Meshulam [18] which was later used by Jürgen Bierbrauer and Yves Edel [1]. The following bound appears in these papers:

Proposition 6. *Let* $C \subset \mathbb{F}_3^d$ *be a d-cap such that any hyperplane intersects C in at most h points. Then*

$$p \le \frac{1+3h}{1+\frac{h}{3^{d-1}}},$$

where p is the size of C.

In particular, any hyperplane intersects a d-cap in a $(d-1)$-cap. Starting with the fact that $a_1 = 2$, we can inductively apply Proposition 6 to obtain

$$a_2 \le 4, \qquad a_3 \le 9, \qquad a_4 \le 21.$$

The bound $a_6 \le 114$ comes from applying Proposition 6 using $h = a_5 = 45$ and $d = 6$. Thus, for low-dimensional caps, Proposition 6 gives nearly sharp bounds. In contrast to other methods, Proposition 6 does not become more difficult to apply as the dimension grows larger.

Given a function $f: \mathbb{F}_3^d \to \mathbb{C}$, define the *Fourier transform* of f to be a new function $\hat{f}: \mathbb{F}_3^d \to \mathbb{C}$ defined by the formula

$$(4)\qquad \hat{f}(z) = \sum_{x\in\mathbb{F}_3^d} f(x)\xi^{z\cdot x},$$

where $\xi = e^{2\pi i/3}$. Given a set $S \subset \mathbb{F}_3^d$, the *characteristic function of* S is defined by the formula

$$\chi(x) = \begin{cases} 1 \text{ if } x \in S, \\ 0 \text{ if } x \notin S. \end{cases}$$

Knowing the characteristic function of S is exactly the same as knowing the set S. The Fourier transform of the characteristic function,

$$\tau(z) = \hat{\chi}(z) = \sum_{x \in \mathbb{F}_3^d} \chi(x)\xi^{z \cdot x} = \sum_{c \in S} \xi^{z \cdot c},$$

has a natural geometric interpretation as follows. Notice first that $\tau(0)$ is simply the size of the set S. Next, let z be a nonzero vector, and consider the three parallel hyperplanes H_0, H_1, H_2 normal to z, where

$$H_j = \{x \in \mathbb{F}_3^d \mid z \cdot x = j\}.$$

To each nonzero vector z we associate an *(ordered) hyperplane triple*

$$(h_0, h_1, h_2) = (|S \cap H_0|, |S \cap H_1|, |S \cap H_2|).$$

Proposition 7. *The complex number $\tau(z)$ encodes the same data as the ordered hyperplane triple (h_0, h_1, h_2) associated to z. In particular,*

$$\tau(z) = h_0\xi^0 + h_1\xi^1 + h_2\xi^2,$$

and

$$h_0 = \frac{2}{3}u + \frac{1}{3}p,$$

$$h_1 = \frac{1}{3}(p - u) + \frac{1}{\sqrt{3}}v,$$

$$h_2 = \frac{1}{3}(p - u) - \frac{1}{\sqrt{3}}v,$$

where $\tau(z) = u + iv$ and $p = \tau(0)$ is the size of S.

We call τ the *(ordered) hyperplane triple function* of S. In the previous section our interest in hyperplane triples was ad hoc; we studied them because, in the end, it paid to do so. We now see that hyperplane triples arise naturally via the Fourier transform.

There is an amazing formula counting the number of lines contained in a set S.

Proposition 8. *Let S be a subset of $\mathbb{F}_3^d$ that contains p points and l lines. Then*

$$p + 6l = \frac{1}{3^d} \sum_{z \in \mathbb{F}_3^d} \tau^3(z),$$

where τ is the hyperplane triple function of S.

In [1], Bierbrauer and Edel use the formula above together with some clever estimates of $|\tau^3(z)|$ to prove the Fourier bound of Proposition 6. We refer the reader to their paper for more details. We now summarize the proof of Edel, Ferret, Landjev, and Storme [6] that $a_5 = 45$.

Proposition 9. *A maximal 5-cap has 45 points.*

Proof. Figure 8 contains a 5-cap with 45 points, so we only need to show that there is none with 46 points. Suppose for a contradiction that C is a 5-cap with 46 points. By the Fourier analysis bound of Proposition 6, if every hyperplane intersects C in at most 18 points, then C can have at most 45 points. Thus, there must be a hyperplane H intersecting C in 19 or 20 points. Deleting a point of C not on H produces a 5-cap with 45 points such that H is a hyperplane intersecting in 19 or 20 points. However, in [6] it is shown that every 5-cap with 45 points has no hyperplanes intersecting in 19 or 20 points. The proof exploits an ingenious identity in the equations for counting marked hyperplanes, together with an exhaustive computer search. □

Solidity

In this section we discuss what is known about high-dimensional maximal caps. In [3], A. Robert Calderbank and Peter Fishburn create very large high-dimensional caps via product constructions based on large low-dimensional caps. As a measure of the "largeness" of a cap, define the *solidity* of a d-cap C to be

$$\sigma(C) = \sqrt[d]{|C|},$$

and define the *asymptotic solidity* of maximal caps to be the supremum of the solidities of maximal caps,

$$\sigma = \sup_d \left\{\sqrt[d]{a_d}\right\}.$$

Thus, asymptotic solidity is at least the solidity of any particular cap. Since every d-cap has fewer than 3^d points, the asymptotic solidity is at most 3. On the other hand, the cap consisting of all 2^d points with all components 0 or 1 shows that the solidity is at least 2. The central open question is the following.

Question. Is the asymptotic solidity less than 3?

The definition of asymptotic solidity leaves open the possibility that for some low d there is a d-cap with high solidity, but for all larger d every d-cap has a substantially smaller solidity. This would make the name "asymptotic solidity" rather questionable, but the following proposition shows that this never happens.

Proposition 10. *Asymptotic solidity is the limit as $d \to \infty$ of the solidity of maximal d-caps.*

Proof. This would be a very short proof if we knew that the solidity of maximal caps σ_d was an increasing sequence. Unfortunately, this is not known, so we will take a more roundabout approach. We first note that given a d-cap C, we can construct a $2d$-cap C' with the same solidity. We do this by taking the product of C with itself: each point of C' has as its first d coordinates a point of C, and as its last d coordinates another point of C, so $|C'| = |C|^2$. Then $\sigma(C') = \sqrt[2d]{|C|^2} = \sigma(C)$. In fact we can also apply this construction to take the product of a d_1-cap C_1 and a d_2-cap C_2 to get a $(d_1 + d_2)$-cap C' with solidity $\sigma(C') = \sqrt[d_1+d_2]{|C_1||C_2|}$. For example, taking the repeated product of a d-cap with itself n times gives an (nd)-cap with the same solidity. So we can replace sup by lim sup in the definition of asymptotic solidity, justifying the "asymptotic" in the name.

We now note that this product construction shows that the function $f : \mathbb{N} \to \mathbb{N}$ defined by setting $f(d)$ to be the size a_d of a maximal d-cap satisfies

$$f(m + n) \geq f(m)f(n).$$

Then Fekete's Lemma (see, for example [20, Lemma 11.6]) implies that $\lim_{n\to\infty} f(n)^{1/n}$ exists. Since solidity is the lim sup of $f(n)^{1/n}$, this limit must equal σ, completing the proof. □

Calderbank and Fisburn use this product construction to show $\sigma > 2.210147$. In [3], they explicitly give two 6-caps, each with 112 points. They exploit these caps in a refined version of the above product construction to get a 13,500-cap with the required solidity. This result has been improved, with a simpler cap, by Yves Edel, who has constructed a 62-cap with solidity 2.214781, and a 480-cap with solidity 2.21739 [5].

Projective Caps

A basic property of an affine SET game is that each pair of cards is contained in a unique SET. Are there other SET-like card games with this property? Yes! In fact, there is a non-affine SET-like game with only seven cards. Consider the *Fano plane* in Figure 12.

The seven *points* of the Fano plane are indicated by the dots in the figure. Each *line* of the Fano plane consists either of the three dots lying on a line segment of the diagram, or the three dots lying on the circle. In the Fano plane, any pair of points determines a unique line, and every line has precisely three points. We define the *Fano* SET *game* to be a card game with one card for each point of the Fano plane, where three cards form a SET if the corresponding points of the Fano plane are collinear.

There is a natural projective geometric construction of the Fano plane. Given a vector space $V = \mathbb{F}_q^{d+1}$, the *projective space of* V, $\mathbb{P}^d\mathbb{F}_q$, is an object tailored to encode the incidence structure of linear subspaces of V. In particular, the elements of the set $\mathbb{P}^d\mathbb{F}_q$ are just the one-dimensional subspaces of V. These are called the *projective points* of $\mathbb{P}^d\mathbb{F}_q$. Given two distinct one-dimensional subspaces, there is a unique two-dimensional subspace containing them. Thus, if we call the two-dimensional subspaces of V the *projective lines*, then we have the nice fact that any two projective points determine a unique projective line.

When $q = 2$, then each projective line contains exactly three projective points. To see this, notice that since the underlying field is $\mathbb{F}_2$, any one-dimensional subspace contains exactly two points: the zero vector, and the nonzero basis vector. Thus, there is a bijection between nonzero vectors and projective points. Let L be a two-dimensional subspace of $\mathbb{F}_2^{d+1}$ representing a projective line, and let $\{e, f\}$ be a vector space basis of L. Then L contains exactly four vectors: 0, e, f, and $e + f$. The nonzero vectors represent the three projective points of L. Amazingly, this gives rise to the same test for collinearity as in the affine case:

> **The Projective Collinearity Rule:** Three non-zero vectors $a, b, c \in \mathbb{F}_2^{d+1}$ represent collinear projective points if and only if $a + b + c = 0$.

The vector space $\mathbb{F}_2^3$ has eight vectors, and so the projective space $\mathbb{P}^2\mathbb{F}_2$ has seven projective points. In fact, $\mathbb{P}^2\mathbb{F}_2$ is the Fano plane. We define a *projective* SET *game of dimension* d to be a card game with one card for each projective point in $\mathbb{P}^d\mathbb{F}_2$, where three cards form a SET if the corresponding projective points are collinear. Then the Fano SET game is just the projective SET game of dimension two.

H. Tracy Hall [8] has devised for himself a deck of cards for a playable projective SET game of dimension five. The key step of his construction is to group the components of a vector of $\mathbb{F}_2^6$ into three pairs,

$$\begin{aligned} a &= (x_1, x_2, y_1, y_2, z_1, z_2) \\ &= ((x_1x_2), (y_1y_2), (z_1z_2)), \end{aligned}$$

and then to interpret each pair as the binary expansion of one of the integers 0,1,2, or 3. For example,

$$\begin{aligned} a &= (1, 0, 0, 1, 0, 0) \\ &= ((10), (01), (00)) = (2, 1, 0). \end{aligned}$$

To further encode the vector as a design on a card, we associate a symbol to each of the three integers 1, 2, or 3, and use the blank symbol for the integer 0. We print three such symbols on each card, one for each coordinate, and distinguish the symbols by printing them in three different fonts. In Figure 13 we do this using different families of symbols for the different coordinates: $\{\blacktriangle, \blacksquare, \bigstar\}$, $\{<, =, >\}$, and $\{\mathbb{Z}, \mathbb{R}, \mathbb{Q}\}$. Hall has a much cuter way to do this using the characters from a popular children's game.

With respect to this method of encoding vectors, the projective collinearity rule has the following translation.

> **The Projective SET Rule:** Three cards are called a SET, if each font appears in exactly one of the following three ways:
>
> - Not at all.
> - As the same symbol exactly twice, and not as any other symbol.
> - As all three symbols.

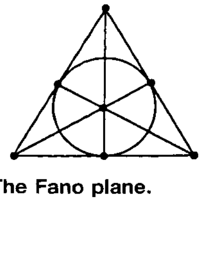

Figure 12. The Fano plane.

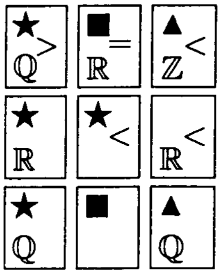

Figure 13. Can you find all four projective SETS?

Hall reports teaching, and then losing at, this game to his nine-year-old niece.

Are there still other SET-like games beyond affine and projective SET? Yes and no. A *Steiner triple system* is a set X together with a collection S of three-element subsets of X, such that, given any two elements $x, y \in X$, there is a unique triple $\{x, y, z\} \in S$. Interpreting elements of X as cards, and triples in S as SETS, then we obtain a SET-like game from any Steiner triple system. The affine and projective SET card games are examples of Steiner triple systems, but it turns out that there are many more exotic Steiner triple systems. Their study is a very rich subject, and the interested reader should look at the book of Charles Colbourn and Alexander Rosa [4].

A natural invariant attached to any Steiner triple system is its *symmetry group*. This is defined in precisely the same way we defined symmetry groups of affine SET games, namely, as the permutations of the points of X taking triples in S to triples in S. One way of studying Steiner triple systems is via their symmetry groups. A notable property of the symmetry groups of affine and projective SET games is that their symmetry groups are *2-transitive* on cards; that is, there is a symmetry taking any ordered pair of cards to any other ordered pair of cards. In particular, this means that, up to symmetry, there is only one type of SET. To capture this, let us define an *abstract* SET *game* to be any Steiner triple system where the symmetry group acts 2-transitively on points. We have the following deep theorem classifying abstract SET games, first conjectured in 1960 by Marshall Hall, Jr. [9].

Theorem 11. *The only abstract* SET *games are affine and projective* SET *games, in* $\mathbb{F}_3^d$ *and* $\mathbb{P}^d\mathbb{F}_2$, *respectively.*

This result is due to Jennifer Key and Ernest Shult [16], Hall [10], and William Kantor [15]. Interestingly, the proofs use part of the classification of finite simple groups.

If we actually play projective SET we want to know how many cards need to be dealt to guarantee a SET. Just as in the affine case, we call a collection of points in $\mathbb{P}^d\mathbb{F}_2$ containing no three points on a projective line a cap. The problem of finding maximal caps for projective SET was solved in 1947 by Raj Chandra Bose. In [2], he showed that the maximal caps of $\mathbb{P}^d\mathbb{F}_2$ have 2^d points. Bose's interest in this problem certainly didn't stem from SET, as the game was not to be invented for another 27 years. Rather, he was coming at it from quite another direction, namely, the theory of *error-correcting codes*, which is the study of the flawless transmission of messages over noisy communication lines. As detailed in the book of Raymond Hill [14], there is a correspondence between projective caps in $\mathbb{P}^d\mathbb{F}_2$ and families of efficient codes. Specifically, if we form the matrix whose columns are vectors representing the projective points of the cap, then the kernel of this matrix is a linear code with Hamming weight four. The more points the cap contains, the more "code-words" the corresponding code has, and so this naturally motivates the problem of finding maximal projective caps. Bose completely solved this problem when $q = 2$, but, as in the affine case, things become much more difficult when $q = 3$. We denote by b_d the size of a maximal projective cap in $\mathbb{P}^d\mathbb{F}_3$. The known values of b_d are given in Table 3.

Table 3

d	1	2	3	4	5	6
b_d	2	4	10	20	56	?

The sizes in dimensions 2 and 3 are due to Bose [2], dimension 4 is due to Pellegrino [19], and dimension 5 is due to Hill [11, 12]. We note that we always have $a_d \leq b_d$, since there is a copy of $\mathbb{F}_3^d$ inside $\mathbb{P}^d\mathbb{F}_3$, so Pellegrino's result is the first proof that $a_4 \leq 20$.

Even though there is no abstract SET game with cards given by points of $\mathbb{P}^5\mathbb{F}_3$, there is still some interesting SET-theory associated with the study of maximal projective caps in this space. In particular, the 45-point affine cap in Figure 8 was constructed by deleting a hyperplane from the 56-point projective 5-cap given by Hill in Figure 4 of [13]. Uniqueness of this affine cap was shown in [6] to be a consequence of the uniqueness of the projective cap, which in turn was demonstrated by Hill in [12] by means of a code-theoretic argument.

Acknowledgments

We have been fascinated with SET since first spending many many hours playing it when we started graduate school together, and have had many helpful conversations with other SET-enthusiasts. We would never have spent so many hours thinking about the "SET-Problem" without Joe Buhler's goading remark that it was shocking that two Berkeley students couldn't work out the answer. Josh Levenberg helped then by writing our original brute-force program, before we came up with a non-computer proof of "20." Galen Huntington got us thinking along the right lines with the proof of Proposition 1. Thanks also to Michael Kleber for many interesting conversations on this topic, only a fraction of which have occurred during the writing of this article, and for informing us of Proposition 10. More recently Mike Zabrocki's computer demonstration at FPSAC 2001 reinspired us, and Sunny Fawcett has been invaluable with integer programming assistance and other ideas. Finally, we are indebted to Joe Buhler, Juergen Bierbrauer, and Yves Edel for extensive comments on an earlier draft of this paper.

REFERENCES

[1] J. Bierbrauer and Y. Edel. Bounds on affine caps. *J. Combin. Des.*, 10(2):111–115, 2002.

[2] R. C. Bose. Mathematical theory of the symmetrical factorial design. *Sankhyā*, 8: 107–166, 1947.

[3] A. R. Calderbank and P. C. Fishburn. Maximal three-independent subsets of $\{0, 1, 2\}^n$. *Des. Codes Cryptogr.*, 4(3):203–211, 1994.

[4] C. J. Colbourn and A. Rosa. *Triple systems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1999.

[5] Y. Edel. Extensions of generalized product caps. Preprint available from http://www.mathi.uni-heidelberg.de/~yves/.

[6] Y. Edel, S. Ferret, I. Landjev, and L. Storme. The classification of the largest caps in *AG*(5, 3). *Journal of Combinatorial Theory, Series A*, 99:95–110, 2002.

[7] W. Fulton and J. Harris. *Representation theory, A first course*. Readings in Mathematics. Springer-Verlag, New York, 1991.

[8] H. T. Hall. Personal communication.

[9] M. Hall, Jr. Automorphisms of Steiner triple systems. *IBM J. Res. Develop*, 4:460–472, 1960.

[10] M. Hall, Jr. Steiner triple systems with a doubly transitive automorphism group. *J. Combin. Theory Ser. A*, 38(2):192–202, 1985.

[11] R. Hill. On the largest size of cap in $S_{5,3}$. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* (8), 54:378–384 (1974), 1973.

[12] R. Hill. Caps and codes. *Discrete Math.*, 22(2):111–137, 1978.

[13] R. Hill. On Pellegrino's 20-caps in $S_{4,3}$. In *Combinatorics '81 (Rome, 1981)*, pages 433–447. North-Holland, Amsterdam, 1983.

[14] R. Hill. *A first course in coding theory*. The Clarendon Press Oxford University Press, New York, 1986.

[15] W. M. Kantor. Homogeneous designs and geometric lattices. *J. Combin. Theory Ser. A*, 38(1):66–74, 1985.

[16] J. D. Key and E. E. Shult. Steiner triple systems with doubly transitive automorphism groups: a corollary to the classification theorem for finite simple groups. *J. Combin. Theory Ser. A*, 36(1):105–110, 1984.

[17] D. Knuth. Programs setset, setset-all, setset-random. Available from http://sunburn.stanford.edu/~knuth/programs.html.

[18] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A*, 71(1):168–172, 1995.

[19] G. Pellegrino. Sul massimo ordine delle calotte in $S_{4,3}$. *Matematiche (Catania)*, 25:149–157 (1971), 1970.

[20] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition, 2001.

Department of Mathematics, and Computer Science, Saint Mary's College, Moraga, CA 94575, USA
e-mail address: bdavis@stmarys-ca.edu

Department of Mathematics, Stanford University, Stanford, CA 94305, USA
e-mail address: maclagan@math.stanford.edu