A sequence approach to linear perfect hash families*

S. G. Barwick and Wen-Ai Jackson
School of Pure Mathematics, University of Adelaide,
Adelaide 5005, Australia

December 22, 2005

Abstract

A linear (q^d, q, t)-perfect hash family of size s in a vector space V of order q^d over a field F of order q consists of a set $\phi_1, \ldots, \phi_s$ of linear functionals from V to F with the following property: for all t subsets $X \subseteq V$ there exists $i \in \{1, \ldots, s\}$ such that ϕ_i is injective when restricted to F. A linear (q^d, q, t)-perfect hash family of minimal size $d(t-1)$ is said to be *optimal.* In this paper we extend the theory for linear perfect hash families based on sequences developed by Blackburn and Wild. We develop techniques which we use to construct new optimal linear $(q^2, q, 5)$-perfect hash families and $(q^4, q, 3)$-perfect hash families. The sequence approach also explains a relationship between linear $(q^3, q, 3)$-perfect hash families and linear $(q^2, q, 4)$-perfect hash families.

1 Introduction to Perfect Hash Families

Perfect hash families were introduced by Mehlhorn [13] in 1984 as part of compiler design. Perfect hash families have also proved useful in a large variety of applications, in particular, there have been a number of recent applications to cryptography. For example, to threshold cryptography (see Blackburn, Burmester, Desmedt and Wild [8] and Blackburn [6]), to broadcast encryption (see Fiat and Naor [10]) and to improve explicit constructions of secure frameproof codes, key distribution patterns, group testing algorithms, cover free families and separating systems (see Stinson, van Trung and Wei [14]).

Let s, t, n, q be positive integers and let V be a set of size n and let F be a set of size q. A function $\phi: V \to F$ *separates* a subset X of V if ϕ is an injection when restricted to X. An (n, q, t)-*perfect hash family* of *size* s is a set $S = \{\phi_1, \ldots, \phi_s\}$ of s functionals from V to F with the property that for all t-subsets $X \subseteq V$, at least one of $\phi_1, \ldots, \phi_s$ separates X.

*This work was supported by the Australian Research Council

We say that S is a *linear* perfect hash family if F can be identified with a finite field $\mathrm{GF}(q)$ and V can be identified with a vector space over $\mathrm{GF}(q)$ in such a way that S is a set of linear functionals under this identification. Thus in the linear case, q is a prime power and $n = q^d$ for some non-negative integer d. This paper deals solely with linear perfect hash families and throughout we use q to denote a prime power. Linear perfect hash families also have a geometric interpretation which was used in [3, 4] to construct linear perfect hash families.

The following result from Blackburn and Wild [9] gives a bound on the size of a linear perfect hash family.

Theorem 1.1 ([9]) *Let d and t be integers such that $d \geq 2$ and $t \geq 2$ and let q be a prime power. If S is a linear (q^d, q, t)-perfect hash family, then $|S| \geq d(t-1)$.*

If $|S| = d(t-1)$ then S is called *optimal.* Blackburn and Wild give conditions for the existence of optimal linear perfect hash families.

Theorem 1.2 ([9]) *An optimal linear (q^d, q, t)-perfect hash family S exists if $q \geq \binom{t}{2}^{d(t-1)}$.*

Perfect hash families are an interesting combinatorial structure with practical applications. They are hard to construct, and there are limited known constructions of linear perfect hash families. Blackburn and Wild [9] give a general construction of an optimal linear perfect hash family which works for q much larger than the bound of Theorem 1.2 and of a certain form, namely $q = q_0^{\alpha_1 \alpha_2 \cdots \alpha_{d(t-2)}}$ where q_0 is any prime power and each $\alpha_i \geq d$. Blackburn [7] gives a construction of optimal linear $(p^2, p, 4)$-perfect hash families where p is a prime, $p = 11$ or $p \geq 17$. Wang and Xing [15] construct linear perfect hash families but their constructions are not optimal. In [3] the authors use geometric techniques to show that optimal linear $(q^2, q, 4)$-perfect hash families exist if and only if $q = 11$ or $q \geq 17$ (q a prime power) and constructions are given for each such q. In [4] geometric techniques are used to show that optimal linear $(q^3, q, 3)$-perfect hash families exist if and only if $q \geq 11$ (q a prime power) and constructions are given for each such q. The authors also give constructions of optimal linear $(q^2, q, 5)$-perfect hash families for restricted values of q. In [1, 2], recursive algorithms for constructing perfect hash families are given. These algorithms need as input perfect hash families with small parameters. This gives motivation for constructing small perfect hash families.

In this paper we investigate the sequence representation of linear perfect hash families developed by Blackburn and Wild [9] in more detail. We generalise their results with the aim of using a sequence approach to construct optimal linear perfect hash families. Known constructions methods of perfect hash families are ad hoc and this sequence approach gives a general method which can be used to construct optimal linear perfect hash families. We illustrate how this approach can be used to construct the optimal linear $(q^2, q, 4)$- and $(q^3, q, 3)$-perfect hash families found in [3, 4]. Further, the sequence approach explains a correspondence between these two cases. We then use these techniques to construct new optimal linear $(q^2, q, 5)$- and $(q^4, q, 3)$-perfect hash families.

2 The sequence approach to linear perfect hash families

In this section we introduce the sequence approach used by Blackburn and Wild in [9]. In their paper they used sequences to prove Theorem 1.2, that is, to show that optimal linear perfect hash families exist if q is large enough. The proof of this result is a probabilistic argument and does not use the sequences to produce constructions of perfect hash families. Our approach here is to study the sequences with the aim of characterizing when they are perfect hash families in order to develop techniques to construct perfect hash families. Consequently we need to study these sequences in more detail than previously done, in particular we need to generalise the results of [9] to results that hold for smaller q; and we need to obtain some more specialised results on these sequences. Some of the preliminary results here are from [9], but we have included some short proofs to assist in understanding the theory. We mostly use the notation of [9]. As a lot of notation and definitions are involved, we include a notation index in the Appendix for easy reference by the reader.

2.1 The sequence approach to the characterization of optimal linear perfect hash families

I'VE PUT $q \geq 2$ BELOW INSTEAD OF $>$, I THOUGHT IT WAS MORE NATURAL.

Let V be a vector space of dimension $d > 0$ over $\mathrm{GF}(q)$, where $q \geq 2$. Let V^* be the dual space of V, consisting of the set of linear functionals $\phi\colon V \to \mathrm{GF}(q)$. If $p \in V$ and $\phi \in V^*$, we will use the notation p^ϕ to represent ϕ acting on p. Given a set $S = \{\phi_1, \ldots, \phi_k\}$ of k linear functionals, we may order them in some arbitrary way to produce the *associated functional sequence* $\Phi = (\phi_1, \ldots, \phi_k) \in (V^*)^k$. In [9], conditions were found on Φ so that S (or equivalently Φ) is a perfect hash family. To explain these conditions we need to introduce the following notation and concepts from [9].

Note that the $t = 1$ case is trivial and if $q < t$ then no functional can separate t points. Hence let $2 \leq t \leq q$ and consider $(V^*)^t$, a dt-dimensional vector space over $\mathrm{GF}(q)$. Let $\Psi = (\psi_1, \ldots, \psi_t) \in (V^*)^t$. Let $P = (p_1, \ldots, p_t)$ be a sequence of t elements of V. We can regard

Ψ as a linear map $\Psi\colon V^t \to \mathrm{GF}(q)$ over $\mathrm{GF}(q)$ where $P^\Psi = \sum_{i=1}^t p_i^{\psi_i}$.

Further, P defines a subspace U_P of $(V^*)^t$ given by

$$U_P = \left\{\Psi = (\psi_1, \ldots, \psi_t) \in (V^*)^t : P^\Psi = 0\right\}.$$

Let $T = \{1, 2, \ldots, t\}$ and define T^2 to be $\{(a, b) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq a < b \leq t\}$. An element (a, b) of T^2 is a *pair* on T, and a sequence c of pairs (possibly repeated) from T^2 is said to be a *pair sequence on* T. For a pair (a, b) on T and all $\phi \in V^*$, define the vector $w_{(a,b),\phi} \in (V^*)^t$ by

$$w_{(a,b),\phi} = (\psi_1, \ldots, \psi_t) \quad \text{where} \quad \psi_i = \begin{cases} \phi & \text{if } i = a, \\ -\phi & \text{if } i = b, \\ 0 & \text{if } i \neq a, b. \end{cases}$$

So for example, $w_{(1,2),\phi} = (\phi, -\phi, 0, \ldots, 0)$. We use $w_{(a,b),\phi}$ to determine whether ϕ separates a t-subset $\{p_1, \ldots, p_t\}$ of V. We let $P = (p_1, \ldots, p_t)$ be an arbitrary ordering of the t-subset and say that ϕ *separates* the sequence P if ϕ separates the set $\{p_1, \ldots, p_t\}$.

Result 2.1 ([9]) *Let $P = (p_1, \ldots, p_t)$ be a sequence of t distinct elements of V. Let $\phi \in V^*$. Then ϕ fails to separate P if and only if $w_{(a,b),\phi} \in U_P$ for some $(a, b) \in T$.*

Proof The linear functional ϕ fails to separate P if and only if ϕ maps two elements of P to the same element of $\mathrm{GF}(q)$. That is, if and only if there exists $(a, b) \in T$ such that $p_a^\phi = p_b^\phi$, that is, $w_{(a,b),\phi} \in U_P$. □

So, for example, if $w_{(1,2),\phi} \in U_P$, then $p_1^\phi + p_2^{-\phi} + p_3^0 + \ldots + p_t^0 = 0$, that is $p_1^\phi = p_2^\phi$, so ϕ maps p_1 and p_2 to the same element of $\mathrm{GF}(q)$. Thus ϕ fails to separate P.

For $(a, b) \in T$, we define the subspace $V_{(a,b)}$ of $(V^*)^t$ by

$$V_{(a,b)} = \{w_{(a,b),\phi} : \phi \in V^*\}.$$

Note that $V_{(a,b)}$ is a vector space of dimension d over $\mathrm{GF}(q)$. We are interested in t-subsets $\{p_1, \ldots, p_t\}$ of V which necessarily have distinct elements. However, when we consider an arbitrary t-sequence $P = (p_1, \ldots, p_t)$ from V^t, P may have repeated elements. The subspace $V_{(a,b)}$ gives us a way to determine whether P has repeated elements.

Result 2.2 ([9]) *Let $P = (p_1, \ldots, p_t)$ be a sequence of t elements of V. Then P has a repeated element $p_a = p_b$ (for some a, b, $1 \leq a < b \leq t$) if and only if $V_{(a,b)} \subseteq U_P$.*

Proof We have $p_a = p_b$ if and only if $p_a^\phi = p_b^\phi$ for all $\phi \in V^*$, if and only if $w_{(a,b),\phi} \in U_P$ for all $\phi \in V^*$, if and only if $V_{(a,b)} \subseteq U_P$. □

Let $c = ((a_1, b_1), \ldots, (a_k, b_k))$ be a k-pair sequence on T and let $\Phi = (\phi_1, \ldots, \phi_k) \in (V^*)^k$ be a k-functional sequence. Define the subspace $W_{c,\Phi}$ of $(V^*)^t$ to be

$$W_{c,\Phi} = \left\langle w_{(a_i,b_i),\phi_i} : 1 \leq i \leq k \right\rangle .$$

So, for example, if $k = 2$ and $c = ((1, 2), (1, 3))$, $\Phi = (\phi_1, \phi_2)$, then $W_{c,\Phi} = \langle(\phi_1, -\phi_1, 0, \ldots, 0), (\phi_2, 0, -\phi_2, 0, \ldots, 0)\rangle$.

The next result involving $W_{c,\Phi}$ shows when $\{\phi_1, \ldots, \phi_k\}$ separates the t-subset $\{p_1, \ldots, p_t\}$ of V.

Result 2.3 ([9]) *Let $P = (p_1, \ldots, p_t)$ be a sequence of t distinct elements of V and let $S = \{\phi_1, \ldots, \phi_k\}$ be a set of k linear functionals with associated k-functional sequence Φ. Then S fails to separate P if and only if there exists a k-pair sequence c on T such that $W_{c,\Phi} \subseteq U_P$.*

Proof The set S fails to separate P if and only if every functional $\phi_i \in \Phi$ maps two distinct elements of P to the same element of GF(q). That is, if and only if for each $\phi_i \in \Phi$ there exists $(a_i, b_i) \in T$ such that $p_{a_i}^{\phi_i} = p_{b_i}^{\phi_i}$, that is, $w_{(a_i,b_i),\phi_i} \in U_P$. This occurs if and only if there exists a pair sequence c on T with $W_{c,\Phi} \subseteq U_P$. □

Note that if a set $\{\phi_1, \ldots, \phi_k\}$ fails to separate a sequence $(p_1, \ldots, p_t)$ of distinct elements, then it fails to separate any supersequence $(p_1, \ldots, p_t, p_{t+1})$.

In the next Lemma we will use the subspace Y of $(V^*)^t$ defined as:

$$Y = \{(\psi_1, \ldots, \psi_t) \in (V^*)^t : \sum_{i=1}^{t} \psi_i = 0\}.$$

Note that Y has dimension $d(t-1)$ over GF(q) (since $\psi_1, \ldots, \psi_{t-1}$ can be chosen arbitrarily, and then $\psi_t = -(\psi_1 + \ldots + \psi_{t-1})$).

Lemma 2.4 *Suppose S is a set of $d(t-1)$ linear functionals with associated functional sequence Φ. Let c be a $d(t-1)$-pair sequence on T. Then $W_{c,\Phi} \subseteq Y$, and so* rank $W_{c,\Phi} \leq d(t-1)$*. In particular, if* rank $W_{c,\Phi} = d(t-1)$*, then $W_{c,\Phi} = Y$ and so $V_{(a,b)} \subseteq W_{c,\Phi}$ for all pairs (a, b) on T.*

Proof As $w_{(a,b),\phi} \in Y$ for all pairs (a, b) on T and $\phi \in V^*$, it follows that $W_{c,\Phi} \subseteq Y$. The remaining statements follow from the definitions. □

The following result was stated in [9] without proof (the proof is very similar to the proof of [9, Theorem 4]).

Result 2.5 ([9]) *Suppose S is a linear (q^d, q, t)-perfect hash family of optimal size $d(t-1)$. Then every d elements of S are independent over GF(q).*

We now describe the geometrical interpretation of linear perfect hash families as introduced in [9] to provide an easy proof for Therem 2.6 below. For an introduction to finite projective geometry see [11]. Let π_∞ be the hyperplane at infinity in the projective space PG(d, q). We identify the elements of V with the affine points of PG(d, q). For any linear functional $\phi \in V^*$ and $\gamma \in$ GF(q), the points $v \in V$ with $v^\phi = \gamma$ form a hyperplane of PG(d, q) and so ϕ corresponds to a parallel class of hyerplanes. There is a correspondence between such parallel classes of hyperplanes in PG(d, q) and the hyperplanes of π_∞. Denote by $[\phi]$ the hyperplane in π_∞ corresponding to the parallel class determined by ϕ. Perfect hash family results now relate to properties in PG(d, q). For example, Result 2.5 states that every d linear functionals are independent, so the the hyperplanes corresponding to the linear functionals form a dual arc in π_∞. Two points p_1, p_2 are separated by ϕ if and only if they belong to different hyperplanes of the parallel class determined by ϕ, if and only if the line p_1p_2 does not meet π_∞ in a point of $[\phi]$.

Example 1 *Suppose $t = 2$. Consider the geometrical interpretation of linear perfect hash families as described above. Then, any set of d independent hyperplanes $\{\phi_1, \ldots, \phi_d\}$ of π_∞ form a $(q^d, d, 2)$ perfect hash family. To see this, consider two distinct affine points P_1, P_2. $P_1P_2 \cap \pi_\infty$ is a point. So $\{\phi_1, \ldots, \phi_d\}$ is a perfect hash family if and only if $\cap_{i=1}^d \phi_i$ does not contain a point, if and only if $\phi_1, \ldots, \phi_d$ are independent.*

If $q = 2$ then as $2 \leq t \leq q$, we have $t = 2$ and this is discussed above.

Theorem 2.6 *Suppose $2 \leq t \leq q$. If Φ is a a linear (q^d, d, t) perfect hash family, then for each t' $(2 \leq t' < t)$, and $d(t'-1)$ subset Φ' of Φ is a (q^d, d, t') perfect hash family.*

Proof If $t = 2$ there is nothing to prove. So suppose firstly that $t' = t - 1 \geq 2$. Suppose that Φ' is not a $(q^d, d, t-1)$ perfect hash family. So there exists distinct points $P' = (p_1, \ldots, p_{t-1})$ and pair sequence c' of length $d(t-2)$ with $W_{c',\Phi'} \subseteq U_{P'}$.

We now use the geometrical representation. Consider the d linear functionals $\phi_1, \ldots, \phi_d$ in $\Phi \backslash \Phi'$. As they are independent (Result 2.5), the intersection $[\phi_1] \cap [\phi_2] \cap \cdots \cap [\phi_{d-1}]$ of the hyperplanes of π_∞ corresponding to the first $d-1$ linear functionals is a point $h \in \pi_\infty$ say, which is not contained in the final linear functional $[\phi_d]$. Now the line p_1h intersects $\langle p_2, [\phi_d] \rangle$ in a point g, say.

Suppose firstly that $g \neq p_1, p_2$. So we have $p_1 g \cap \pi_\infty \in [\phi_1] \cap [\phi_2] \cap \cdots \cap [\phi_{d-1}]$ and $p_2 g \cap \pi_\infty \in [\phi_d]$. If $g \in \{p_3, \ldots, p_{t-1}\}$ then Φ does not separate the $t-1$ distinct points $p_1, \ldots, p_{t-1}$. Otherwise Φ does not separate the t distinct points $p_1, \ldots, p_{t-1}, g$. In both cases this contradicts Φ being a perfect hash family.

If $g = p_1$ then let p_t be any point of $p_1 h \backslash \{p_1, h\}$. So $\phi_1, \ldots, \phi_{d-1}$ does not separate p_1, p_t and ϕ_d does not separate p_1, p_2. Hence using $W_{c',\Phi'} \subseteq U_{P'}$ we can conclude Φ does not separate the distinct elements from $p_1, \ldots, p_t$, contradicting Φ being a perfect hash family.

Similarly, if $g = p_2$ then $\phi_1, \ldots, \phi_{d-1}$ does not separate p_1, p_2. Let $p_t \in \langle p_2, [\phi_d] \rangle \backslash \{p_2, \pi_\infty\}$. So ϕ_d does not separate p_2, p_t. Hence Φ does not separate the distinct elements from $p_1, \ldots, p_t$.

The remaining values of t' follow by repeating the argument. □

Example 2 *We now discuss the case $t = 3$. If we consider three affine (non-collinear) points P_1, P_2, P_3, then the points $Q_3 = P_1P_2 \cap \pi_\infty$, $Q_2 = P_1P_3 \cap \pi_\infty$ and $Q_1 = P_2P_3 \cap \pi_\infty$ are collinear. Thus $\{\phi_1, \ldots, \phi_{2d}\}$ is not a perfect hash family if we can find a partition of $\phi_1, \ldots, \phi_{2d}$ into sets A_1, A_2, A_3 where $|A_1| = |A_2| = d-1$ with*

$$R_1 = \bigcap_{\phi \in A_1} \phi, \qquad R_2 = \bigcap_{\phi \in A_2} \phi, \qquad \theta = \bigcap_{\phi \in A_3} \phi$$

and $R_1R_2 \cap \theta \neq \emptyset$ (R_1, R_2 are points by Theorem 2.6 and Result 2.5).

Thus, for every such partition A_1, A_2, A_3, we need $\langle R_1, R_2, \theta \rangle = \pi_\infty$. We see later (Sections 5,6) that there are other conditions to consider.

We now examine the case when $d = 2$, where the condition now becomes no three functionals of a $(q^2, 2, 3)$ perfect hash family are dependent and so the four functionals are distinct under the geometrical representation.

In the following lemmas we relate pair sequence properties with the rank properties of $W_{c,\Phi}$.

The next result determines the maximal rank of $W_{c,\Phi}$ for any k-functional sequence Φ and k-pair sequence c. We will show later that Φ is a perfect hash family if and only if the rank of $W_{c,\Phi}$ is this maximal value for all relevant k-pair sequences c. We will use the subspace $R = \{(p, \ldots, p) : p \in V\} \subseteq V^t$. Note that R is a vector space of dimension 1 over V, and consequently of dimension d over GF(q).

Lemma 2.7 *Let Φ be a k-functional sequence and let c be any k-pair sequence of T^k. Then we have* rank $W_{c,\Phi} \leq d(t-1)$. *Further,* rank $W_{c,\Phi} < d(t-1)$ *if and only if there exists $P \in V^t \setminus R$ with $W_{c,\Phi} \subseteq U_P$.*

Proof Let $\Phi = (\phi_1, \ldots, \phi_k)$, $c = ((a_1, b_1), \ldots, (a_k, b_k))$ and let $[W_{c,\Phi}]$ be the $k \times t$ matrix with rows $w_{(a_i,b_i)\phi_i}$, $1 \leq i \leq k$. So the row space of $[W_{c,\Phi}]$ is equal to $W_{c,\Phi}$. Let $P = (p_1, \ldots, p_t) \in V^t$. As each $w_{(a_i,b_i),\phi_i}\colon V^t \to \mathrm{GF}(q)$ is a linear map, we can regard $[W_{c,\Phi}]\colon V^t \to \mathrm{GF}(q)^k$ as a linear map over GF(q), where $P^{[W_{c,\Phi}]} = [W_{c,\Phi}]P^{Tr} = (P^{w_{(a_1,b_1),\phi_1}}, \ldots, P^{w_{(a_k,b_k),\phi_k}})$ (where ^{Tr} represents matrix transposition). Hence over GF(q), $[W_{c,\phi}]$ satisfies the dimension theorem:

$$\dim \text{kernel}([W_{c,\Phi}]) + \dim \text{image}([W_{c,\Phi}]) = \dim V^t.$$

Now dim $V^t = dt$ and $R \subseteq \text{kernel}([W_{c,\Phi}])$, so dim kernel$([W_{c,\Phi}]) \geq d$. Hence

$$\dim \text{image }([W_{c,\Phi}]) = \text{rank } W_{c,\Phi} \leq d(t-1)$$

with equality if and only if ker $([W_{c,\Phi}]) = R$. That is, if and only if $W_{c,\Phi} \not\subseteq U_P$ for any $P \notin R$. □

As the row space of $[W_{c,\Phi}]$ is equal to $W_{c,\Phi}$, they have the same rank, so from now on, we use $W_{c,\Phi}$ for $[W_{c,\Phi}]$.

Let A be a subset of $T = \{1, \ldots, t\}$. If c is a pair sequence on T, define c_A to be the pair sequence on A where the pairs in c_A are only those pairs of c whose entries lie entirely in A. Similarly, if Φ is a functional sequence associated with c, let Φ_A be the functional sequence consisting of those elements of Φ corresponding to the pairs in c_A.

We remind the reader that a k-pair sequence $c = ((a_1, b_1), \ldots, (a_k, b_k))$ on T may contain repeated pairs. We say c is (d,t)-*limited* if for every t' $(2 \leq t' < t)$, and t'-subset A of T, the number of pairs $|c_A|$ in c_A is *strictly less* than $d(t'-1)$.

The concept of (d,t)-limited sequences is fundamental to our technique for constructing perfect hash families. This concept was not used in Blackburn and Wild [9]; it is important for our constructions as it considerably reduces the number of sequences to be considered in the construction of a perfect hash family.

SUE, THIS IS NEW

Lemma 2.8 *Consider a (q^d, d, t) perfect hash family with associated linear functional sequence Φ. Let c be a $d(t-1)$-pair sequence. Let c' a subsequence of c with one pair less. If c' is (d,t)-limited sequence then* rank $W_{c,\Phi} = d(t-1)$.

Proof Let Φ' be the subsequence of Φ corresponding to c'. By Lemma 2.7, rank $W_{c,\Phi} \leq d(t-1)$, so we suppose that rank $W_{c,\Phi} < d(t-1)$ and arrive at a contradiction. By Lemma 2.7 there exists $P = (p_1, \ldots, p_t) \notin R$ with $W_{c',\Phi'} \subseteq W_{c,\Phi} \subseteq U_P$. As Φ is a perfect hash family, P does not have distinct elements. So P has some repeated elements, but at least two different elements, as $P \notin R$. Suppose P has r $(1 < r < t)$ distinct elements $q_1, \dots, q_r$. Let $A_i \subseteq T$ be the set $\{j : p_j = q_i\}$, and let $a_i = |A_i|$, the number of times q_i occurs in P, so $1 \le a_i < t$. If no $a_i = 1$, set s to 0 otherwise suppose without loss of generality that $1 = a_1 = a_2 = \cdots = a_s$, and $a_i > 1$ for $i > s$. Note that $s < r$ and

$$s + \sum_{i=s+1}^{r} a_i = t. \tag{1}$$

Our aim is to show that the number $|c|$ of pairs in c is strictly bounded above by $d(t-1)$, contradicting c being of length $d(t-1)$, from which the theorem follows.

We partition the pairs of c' into $r+1$ categories $(0), (1), \dots, (r)$. The category (0) consists of the pairs (e, f) of c' where $e \in A_i$, $f \in A_j$ and $i \neq j$. The category (i) consists of the pairs (e, f) of c' where $e, f \in A_i$, $e \neq f$ $(1 \le i \le r)$.

We first consider the pairs in category (i) $(1 \le i \le r)$. If $a_i = 1$ then there are no pairs in this category (as a pair has two distinct entries). Otherwise, as c' is (d, t)-limited, we have

$$|c'_{A_i}| \le d(a_i - 1) - 1.$$

Now consider the pairs of category (0). We will define a new pair sequence c'_0. For each pair (e, f) of c' of category (0), define a corresponding pair (i, j) of c'_0, where $e \in A_i$, $f \in A_j$. By definition of category (0), $i \neq j$. (If $i > j$ write the pair as (j, i) without loss of generality.) Let Φ'_0 be those elements of Φ' corresponding to the pairs in category (0). As $W_{c',\Phi'} \subseteq U_P$, we have

$$W_{c'_0,\Phi'_0} \subseteq U_Q, \tag{2}$$

where $Q = (q_1, \dots, q_r)$.

If $|\Phi_0| \ge d(r-1)$, then by Result 2.6, Φ_0 is a (q^d, d, r) perfect hash family. As Q has distinct points, this contradicts (2). Hence $|\Phi'_0| \le d(r-1)-1$ and so the number $|c'_0|$ of pairs in category (0) is less than or equal to $d(r-1)-1$.

We are now ready to bound the length of c', using the bounds on the pairs in categories $(0), (1), \dots, (r)$. We have

$$\begin{aligned}
|c'| &\le d(r-1) - 1 + \sum_{i=s+1}^{r} (d(a_i - 1) - 1) \\
&= dr - d - 1 + \left(\sum_{i=s+1}^{r} d a_i \right) - (r-s)d - (r-s) \\
&= d \left(s + \sum_{i=s+1}^{r} a_i \right) - d - 1 - (r-s) \\
&= d(t-1) - 1 - (r-s) \qquad \text{by (1)} \\
&\le d(t-1) - 2
\end{aligned}$$

as $s < r$. A $|c| = |c'| + 1$, this contradicts c being of length $d(t-1)$. □

Lemma 2.9 *Let Φ be a k-functional sequence $(1 \leq k \leq d(t-1))$ with the property that every $d(t-2)$-subset of S is a $(q^d, q, t-1)$ perfect hash family. Let c be a k-pair sequence on T which is not (d,t)-limited. Then $V_{(a,b)} \subseteq W_{c,\Phi}$ for some pair (a,b) on T.*

Proof As c is not (d,t)-limited, let t' be the smallest value $(2 \leq t' < t)$ with a t'-subset A of T satisfying $|c_A| \geq d(t'-1)$. Choose any subsequence c' of c_A of length $d(t'-1)$, and let Φ' be the corresponding subsequence of Φ_A. By choice of t', c' is (d,t')-limited. As $t' < t$, by Theorem 2.6, Φ' is a (q^d, q, t') perfect hash family. Applying Lemma 2.8 with Φ' and c', we have rank $W_{c',\Phi'} = d(t'-1)$. By Lemma 2.4, there exists $a, b \in A$ with $V_{(a,b)} \subseteq W_{c',\Phi'} \subseteq W_{c,\Phi}$, as required. □

The following is our first theorem that characterises perfect hash families in terms of sequences. In [9] the forward implication was indicated, but we include the proof for completeness.

Theorem 2.10 *Suppose $2 \leq t \leq q$. Suppose S is a set of $d(t-1)$ linear functionals with associated functional sequence Φ. Then S is an (optimal) linear (q^d, q, t)-perfect hash family if and only if, for all $|S|$-pair sequences c we have either: (a) there exists a pair (a,b) on T with $V_{(a,b)} \subseteq W_{c,\Phi}$, or (b)* rank $W_{c,\Phi} = d(t-1)$.

Proof (⇒) Suppose that S is a perfect hash family and let c be a $|S|$-pair sequence. If c is not (d,t)-limited, then (a) holds by Theorem 2.6 and Lemma 2.9. If c is (d,t)-limited, then (b) holds by Lemma 2.8.

(⇐) Suppose that S is not a perfect hash family. Then there exists a set P of t distinct elements of V that S does not separate. Thus by Result 2.3, there exists a $|S|$-pair sequence c with $W_{c,\Phi} \subseteq U_P$. If c satisfies (a), then there exists $(a,b) \in T$ such that $V_{(a,b)} \subseteq W_{c,\Phi} \subseteq U_P$ and so by Result 2.2, P does not have distinct elements, a contradiction. If c satisfies (b), then by Lemma 2.4 we have $V_{(a,b)} \subseteq W_{c,\Phi} \subseteq U_P$ for all $(a,b) \in T$, and so by Result 2.2, P does not have distinct elements, a contradiction. Thus S is a perfect hash family. □

Corollary 2.11 *Suppose $2 \leq t \leq q$. Suppose S is a set of $d(t-1)$ linear functionals with associated functional sequence Φ. Then S is an (optimal) linear (q^d, q, t)-perfect hash family if and only if, for all $|S|$-pair sequences c there exists a pair (a,b) on T with $V_{(a,b)} \subseteq W_{c,\Phi}$.*

This is also new.

Lemma 2.12 *Suppose $2 \leq t \leq q$. Let Φ be a $d(t-1)$ linear functional sequence with the property that, for every $d(t-1)$ pair sequence c of the form $c = (c', (a,b))$, where c' be a (d,t)-limited pair sequence of length $d(t-1)-1$, we have* rank $W_{c,\Phi} = d(t-1)$. *Then Φ is a (q^d, d, t) perfect hash family.*

Proof We prove this by induction on t.

Suppose firstly that $t = 2$. In this case $T = \{1,2\}$ and $c = (1,2)^d$ (that is, the pair $(1,2)$ repeated d times). The condition that rank $W_{c,\phi} = d$ is equivalent to every d functions in Φ being linearly independent, is equivalent to Φ being a $(q^d, d, 2)$ perfect hash family (see Example 1).

Suppose now that the statement of the lemma holds for the value $t-1$ ($t-1 \geq 2$). We prove the statement for t.

Let c be a $d(t-1)$-pair sequence. If $c = (c', (a,b))$ where c' is (d,t)-limited, by assumption, rank $W_{c,\Phi} = d(t-1)$. We will show later that every $d(t-2)$-subset of Φ is a $(q^d, q, t-1)$ perfect hash family. So if $c = (c', (a,b))$ is not (d,t)-limited, by Lemma 2.9 there exists a pair (a,b) with $V_{(a,b)} \subseteq W_{c,\phi}$. Hence by Theorem 2.10, Φ is a (q^d, q, t) perfect hash family.

It remains to show that a $d(t-2)$ subsequence Ψ of Φ is a $(q^d, d, t-1)$ perfect hash family.

Take a $d(t-2)$ pair sequence $e = (e', (a,b))$ on $T' = \{1, \ldots, t-1\}$ where e' is $(d, t-1)$-limited. Now extend e' to a $d(t-1)$-sequence c' where

$$c' = (e', (t-2, t), (t-1, t)^{d-1}).$$

If we prove that c' is (d,t)-limited, then rank $W_{(c',(a,b)),\Phi} = d(t-1)$ (by induction assumption) and in particular rank $W_{e,\Psi} = d(t-2)$. From this we conclude by inductive hypothesis that Ψ is a $(q^d, d, t-1)$ perfect hash family, as required. It remains to show that c' is (d,t)-limited.

Note that as e' is $(d, t-1)$-limited, for any subset A' of T', we have

$$|c_{A'}| \leq d(|A'|-1) - 1.$$

Consider $A \subset T$. There are four cases to consider: (1) $t-1, t \in A$, (2) $t-1 \in A$, $t \notin A$, (3) $t-1 \notin A$, $t \in A$, and (4) $t-1, t \notin A$.

In Case (1), write $A = A' \cup \{t\}$, so $|A'| \geq 1$. If $|A'| > 1$ then

$$|c_A| \leq d(|A'|-1) - 1 + d = d(|A|-1) - 1$$

as required. Otherwise $|A'| = 1$ and $A = \{t-1, t\}$ and so $|c_A| = d(|A|-1) - 1$ as required.

In Case (2), $|c_A| = |e_A| \leq d(|A|-1)-1$ as required. In Case (3), write $A = A' \cup \{t\}$. If $|A'| > 1$, then $|c_A| \leq d(|A'|-1) - 1 + 1 \leq d(|A|-1) - 1$. If $|A'| = 1$ then $|c_A| \leq 1 \leq d(|A|-1) - 1$, as required.

In Case (4), $|c_A| = |e_A| \leq d(|A|-1) - 1$, as required.

Hence, c' is (d,t)-limited as required. □

Theorem 2.13 *Suppose $2 \leq t \leq q$. Suppose S is a set of $d(t-1)$ linear functionals with associated functional sequence Φ. Then S is a (q^d, d, t) perfect hash family if and only if both*

1. *Every $d(t-2)$-subset of S is a $(q^d, q, t-1)$ perfect hash family.*
2. *For all (d,t)-limited $|S|$-pair sequences c we have* $\operatorname{rank} W_{c,\Phi} = d(t-1)$.

Proof The forward direction follows immediately from Lemma 2.8. The reverse direction follows from Lemma 2.9 and Theorem 2.10. □

We are now able to prove the main theorem which characterizes when a functional sequence Φ is a perfect hash family in terms of the rank of $W_{c,\Phi}$ and (d,t)-limited pair sequences.

Theorem 2.14 *Suppose $2 \leq t \leq q$. Suppose S is a set of $d(t-1)$ linear functionals with associated functional sequence Φ. Then S is an (optimal) linear (q^d, q, t)-perfect hash family if and only if, for all $|S|$-pair sequences of the form $c = (c', (a,b))$ where c' is (d,t)-limited, and all pairs (a,b) on T, we have*

$$\operatorname{rank} W_{c,\Phi} = d(t-1).$$

Proof The forward direction follows immediately from Theorem 2.6 and Lemma 2.8. The backward direction is Lemma 2.12. □

2.2 Methods to construct linear perfect hash families

These results give us two approaches to constructing perfect hash families. Both these methods involve finding $d(t-1)$ functionals that satisfy Theorem 2.14.

Before describing these methods, we first note that if c_1, c_2 are two k-pair sequences, then we can determine whether they are isomorphic under the induced action of the symmetric group S_t acting on the collection of k-pair sequences. If c_1 and c_2 are isomorphic, then for any k-functional sequence Φ, $W_{c_1,\Phi}$ is a column permutation of $W_{c_2,\Phi}$ and so $\operatorname{rank} W_{c_1,\Phi} = \operatorname{rank} W_{c_2,\Phi}$.

Thus, to show that a $d(t-1)$-functional sequence Φ satisfies Theorem 2.14, it is sufficient to show that it satisfies the rank condition for one representative of each isomorphism group of the d-limited $d(t-1)$-pair sequences under the induced action of S_t. We will call *representative* k-pair sequences the collection of one sequence from each isomorphism group. In the next sections, we present some examples and show how to find the representative pair sequences.

Method 1 This method involves guessing a $d(t-1)$-functional sequence Φ, and then checking if it is a linear (q^d, q, t)-perfect hash family. To do this we check that for each representative $d(t-1)$-pair sequence c with a subsequence c' of length $d(t-1)-1$ which is (d, t)-limited, we have rank $W_{c,\Phi} = d(t-1)$. Then by Theorem 2.14, Φ is an optimal linear (q^d, q, t)-perfect hash family.

This method is not very practical; the second method gives us a technique for building up a perfect hash family.

Method 2 This method has two main steps. The first step is to find an appropriate $d(t-1)-1$ functional sequence Φ'. The second step is to consider all (d, t)-limited sequences c' of length $d(t-1)-1$, and all pairs (a, b) and find the functions ϕ for which

$$\text{rank } W_{c',\Phi'} = \text{rank } W_{(c',(a,b)),(\Phi',\phi)}.$$

We formally show later that any other function ϕ will complete Φ' to a perfect hash family.

We say a $(d(t-1)-1)$ functional sequence Φ' is (d, t)-*suitable* if every $d(t-2)$ subset of Φ' is $(q^d, d, t-2)$ perfect hash family and rank $W_{c',\Phi'} = d(t-1)-1$ for all (representative) (d, t)-limited sequences c' of length $d(t-1)-1$.

Let (a, b) be any pair on T. If ϕ is a linear functional with rank $W_{c',\Phi'} =$ rank $W_{(c',(a,b)),(\Phi',\phi)}$, then we say that ϕ is *excluded.* In the case that $(c', (a, b))$ is (d, t)-limited, we say that ϕ is *limit-excluded.*

Now suppose $(c', (a, b))$ is not (d, t)-limited. As c' is (d, t)-limited, this means there exists $A \subset T$ with $|(c', (a, b))_A| = d(|A|-1)$. If ϕ is a linear functional with rank $W_{c',\Phi'} =$ rank $W_{(c',(a,b)),(\Phi',\phi)}$, then (Φ_A, ϕ) is not a $(q^d, d, |A|-1)$ perfect hash family, for A with $|A| < t$. We call ϕ *sub-excluded* (reflecting a *sub*set of a perfect hash family also being a perfect hash family). Then the sub-excluded functions are the functions which are excluded because adding them would result in a functional sequence contradicting Theorem 2.6.

It is important to note that by definition $W_{c,\Phi}$ and its rank is dependent on the value of q.

Theorem 2.15 *Let* $2 \leq t \leq q$. *Let* Φ' *be a* (d, t)*-suitable functional sequence of length* $d(t-1)-1$. *For any linear functional* ϕ *which is neither limit-excluded nor sub-excluded,* (Φ', ϕ) *forms a*

(q^d, d, t) perfect hash family. Further, every (q^d, d, t) perfect hash family can be constructed in this way.

Proof As Φ' is (d,t)-suitable, rank $W_{c',\Phi'} = d(t-1)-1$ for all (d,t)-limited sequences c'. If ϕ is neither limit-excluded nor sub-excluded then rank $W_{(c',(a,b)),(\Phi',\phi)} =$ rank $W_{c',\Phi'} + 1 = d(t-1)$. Hence, by Theorem 2.14, (Φ', ϕ) is a (q^d, d, t) perfect hash family.

Take any (q^d, d, t) perfect hash family Φ, and let $\Phi = (\Phi', \phi)$, where Φ' is the sequence of the first $d(t-1)-1$ elements of Φ and ϕ is the remaining linear functional. Then Φ' is (d,t)-suitable by Theorems 2.6 and 2.14. The functional ϕ is not excluded by Theorem 2.14. □

We now summarise Method 2.

To find a (q^d, d, t) perfect hash family

Assumed knowledge: Existence of $(q^d, d, t-1)$ perfect hash families

Suitability Phase: Find a (d,t) suitable $d(t-1)-1$ functional sequence Φ', that is, so that

(a) Every $d(t-2)$ subsequence of Φ' is a $(q^d, d, t-1)$ perfect hash family.

(b) Check for every (d,t)-limited $d(t-1)-1$ sequence c', that we have rank $W_{c',\phi'} = d(t-1)-1$.

Exclusion Phase:

(a) For every $d(t-2)-1$ subsequence of Φ', calculate the excluded functions. These are the sub-excluded functions of Φ'.

(b) For every (representative) (d,t)-limited $d(t-1)-1$ pair sequence c', calculate the limit-excluded functions of Φ'.

In previous sections, we have used c and c' to represent pair sequences of length $d(t-1)$ and $d(t-1)-1$ respectively. In the construction of perfect hash families, we are only concerned with sequences of length $d(t-1)-1$. Hence from now on we will omit the $'$, and use c and Φ to denote sequences of length $d(t-1)-1$.

2.3 Sequence condition for every d functionals independent

Let (a,b) be a pair on T. If c is a (d,t)-limited sequence of length $d(t-1)-1$ with $(a,b)^{d-1} \in c$, then the condition that rank $W_{(c,(a,b)),(\Phi,\phi)} = d(t-1)$ over all such sequences c implies that every

d linear functionals in a perfect hash family is linearly independent (cf Result 2.5). It simplifies the search for perfect hash families by reducing the number of functionals to eliminate.

In the case $d = 2$, an optimal linear (q^2, q, t) perfect hash family S has size $2(t-1)$. A linear functional $\phi \in V^*$ can be represented by $\phi = [a, b]$ where $a, b \in \mathrm{GF}(q)$ (we will use square brackets to denote linear functionals). Suppose that c' is a (d, t)-limited sequence of length $d(t-1)-1$. If a pair $(a, b)^{d-1} = (a, b)$ occurs in c', then a condition that ϕ is not sub-excluded is the condition that every two functions in a perfect hash family are independent. Hence, when constructing a perfect hash family if $\phi = [a, b] \in S$ then $\alpha\phi = [\alpha a, \alpha b]$ is a sub-excluded linear functional. Hence, without loss of generality, we only consider linear functionals of the form $\phi = [1, \alpha]$ where $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$ (where $\phi = [1, \infty]$ is used to represent the linear functional $[0, 1]$).

3 Case $d = 2,\ t = 4$

We will use the sequence theory developed above to show how to construct optimal linear $(q^2, q, 4)$-perfect hash families. We take a brief look at this case for three reasons. Firstly, this case was considered using a geometrical approach in [3] and we want to illustrate the difference between the geometric approach and the sequence approach. Secondly, we want to explain the correspondence between this case and a case of $d = 3$, $t = 3$ in Section 5. Thirdly, when considering the $d = 2$, $t = 5$ case in Section 4, we will see that (as implied by Theorem 2.6) the results for $t = 4$ appear as part of the conditions for the $t = 5$ case.

3.1 Constructing $(q^2, q, 4)$-perfect hash families

We will use Theorem 2.15 to construct a perfect hash family. As mentioned in Section 2.3, we will assume the linear functionals are of the form $\phi = [1, \alpha]$ where $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$. We have $T^2 = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq a < b \leq 4\} = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$. A $(2, 4)$-limited pair sequence of length $d(t-1) = 6$ uses each pair on T exactly once. This covers the limit-excluded functions. For the sub-excluded functions, a pair in c is repeated. However, this case is already dealt with in Section 2.3, since we take our linear functions to be of the form $[1, \alpha]$.

The method begins with a $(2, 4)$-suitable 5-functional sequence Φ and all the representative $(2, 4)$-limited 5-pair sequences c on T. As $|T^2| = 6$, there is a unique way to complete a $(2, 4)$-limited 5-pair sequence c to a $(2, 4)$-limited 6-pair sequence $(c, (a_6, b_6))$. For each such representative 5-pair sequence c, we form $W_{c,\Phi}$ and row reduce to calculate the limit-excluded linear functional ϕ corresponding to the remaining pair (a_6, b_6) on T. Once we have calculated all the limit-excluded linear functionals, any of the remaining linear functionals in V^* of the form $[1, \alpha]$, $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$ can be added to Φ to form a perfect hash family.

We first count the number of representative $(2,4)$-limited 5-pair sequences on T. The action of the symmetric group S_4 on $1, 2, 3, 4$ induces an action on T. For example, the mapping $\sigma = (123)$ (that is, $1 \mapsto 2 \mapsto 3 \mapsto 1$) maps the pair $(2,3)$ on T onto $(3,1) \equiv (1,3)$. Without loss of generality, we can assume that a representative 5-pair sequence contains the pairs $\mathcal{Q} = \{(1,2),(1,3),(1,4),(2,3),(2,4)\}$. So the subgroup $\langle (1,2),(3,4) \rangle$ of S_4 of size 4 fixing the set $\mathcal{Q}$. The number of representative 5-pair sequences consisting of the pairs $\{(1,2),(1,3),(1,4),(2,3),(2,4)\}$ under S_4 is therefore $5!/4 = 30$.

We choose a general 5-functional sequence Φ and consider the following 5-pair sequence c:

$$\begin{array}{rcllllll} \Phi & = & ([1,v], & [1,w], & [1,x], & [1,y], & [1,z]) \\ c & = & ((1,2), & (1,3), & (1,4), & (2,3), & (2,4)). \end{array}$$

We need to construct $W_{c,\Phi}$. Recall that $w_{(1,2),\phi} = (\phi, -\phi, 0, 0)$, thus the first row of $W_{c,\Phi}$ is $(1, v, -1, -v, 0, 0, 0, 0)$. Similarly we obtain all the $w_{c,\phi}$ and so $W_{c,\Phi}$ written as a matrix is:

$$W_{c,\Phi} = \left[\begin{array}{cc|cc|cc|cc} 1 & v & -1 & -v & 0 & 0 & 0 & 0 \\ 1 & w & 0 & 0 & -1 & -w & 0 & 0 \\ 1 & x & 0 & 0 & 0 & 0 & -1 & -x \\ 0 & 0 & 1 & y & -1 & -y & 0 & 0 \\ 0 & 0 & 1 & z & 0 & 0 & -1 & -z \end{array}\right].$$

We need Φ to be $(2,4)$-suitable. From Example 2, every 4 elements of v, w, x, y, z are distinct. So we need the rank of $W_{c,\Phi}$ to be 5 for all $(2,4)$-limited 5-pair sequences c. By inspection the matrix above has rank 5 if v, w, x, y, z are distinct. Now, different representative 5-pair sequences correspond to permuting the pairs in c and so correspond to permuting v, w, x, y, z in this matrix. Hence for any representative 5-pair sequence c, rank $W_{c,\Phi} = 5$. Thus any set of 5 distinct values of v, w, x, y, z results in a $(2,4)$-suitable 5-functional sequence Φ.

We now perform row operations on $W_{c,\Phi}$ to find the functional $[1, \theta]$ such that $w_{(3,4),[1,\theta]}$ is in the row space of $W_{c,\Phi}$. That is, we perform row operations to construct a row $[0, 0, 0, 0, 1, \theta, -1, -\theta]$. For the above matrix, the linear functional corresponding to the final pair $(3,4)$ is $[1, \theta]$ where

$$\theta \ = \ -\frac{xy(v - w - z) + (x + y)wz - vwz}{xy - (x + y)v + zv - zw + wv}.$$

Thus we have rank $W_{c,\Phi} =$ rank $W_{(c,(3,4)),(\Phi,[1,\theta])} = 5$ and so we cannot add the linear functional $[1, \theta]$ to our set $\{[1,v],[1,w],[1,x],[1,y],[1,z]\}$ to obtain a perfect hash family. That is, $[1, \theta]$ is a limit-excluded linear functional. (Note by choice of the functionals being of the form $[1, \alpha]$, the sub-excluded linear functionals are exactly the elements $\{[1, v], [1, w], [1, x], [1, y], [1, z]\}$).

There are 30 representative 5-pair sequences, so there are 30 limit-excluded functionals. However, we see that θ is symmetrical in x and y (with corresponding pairs $(1, 4)$ and $(2, 3)$) and so two sequences with the pairs $(1, 4)$ and $(2, 3)$ swapped yield the same limit-excluded linear functional $[1, \theta]$. This gives us a total of 15 representative 5-pair sequences to consider. These can be calculated directly by hand; Table 1 lists 15 representative 5-pair sequences. (Note also that θ is symmetrical in w and z but this does not reduce the number of sequences to be considered.)

Table 1: The 15 representative sequences $c_1, \ldots, c_{15}$ for $d = 2$, $t = 4$

c_1	$((1,2),(1,3),(1,4),(2,3),(2,4))$	c_9	$((1,3),(1,4),(2,3),(1,2),(2,4))$
c_2	$((1,2),(1,3),(1,4),(2,4),(2,3))$	c_{10}	$((1,3),(1,4),(2,3),(2,4),(1,2))$
c_3	$((1,2),(1,3),(2,4),(1,4),(2,3))$	c_{11}	$((1,3),(1,4),(2,4),(1,2),(2,3))$
c_4	$((1,3),(1,2),(1,4),(2,3),(2,4))$	c_{12}	$((1,3),(1,4),(2,4),(2,3),(1,2))$
c_5	$((1,3),(1,2),(1,4),(2,4),(2,3))$	c_{13}	$((1,3),(2,4),(1,2),(1,4),(2,3))$
c_6	$((1,3),(1,2),(2,4),(1,4),(2,3))$	c_{14}	$((1,3),(2,4),(1,4),(1,2),(2,3))$
c_7	$((1,3),(1,4),(1,2),(2,3),(2,4))$	c_{15}	$((1,3),(2,4),(1,4),(2,3),(1,2))$
c_8	$((1,3),(1,4),(1,2),(2,4),(2,3))$		

To construct a perfect hash family, we choose the 5-functional sequence to be

$$\Phi = ([0, 1], [1, 0], [1, 1], [1, a], [1, b]).$$

We can do this without loss of generality as the group $\mathrm{GL}(2, q)$ of linear functionals of $\mathrm{GF}(q)^2$ is 3-transitive on a one dimensional subspace (see [12]). By the suitability of Φ, we need $a \neq b$ and $a, b \neq 0, 1, \infty$.

We want to find the limit-excluded functional for each representative 5-pair sequence c in Table 1. Rather than performing row operations on a new $W_{c,\Phi}$ for each sequence c, we use the $W_{c,\Phi}$ matrix above, and note that changing the sequence c is equivalent to permuting v, w, x, y, z in the matrix. Hence it is equivalent to permuting v, w, x, y, z in the excluded functional $[1, \theta]$. So for the first sequence $c_1 = ((1, 2), (1, 3), (1, 4), (2, 3), (2, 4))$, we have $v = \infty$, $w = 0$, $x = 1$, $y = a$, $z = b$, hence $\theta = -a/(b - a - 1)$. Thus we obtain the limit-excluded linear functional $\phi_1 = [1, -a/(b - a - 1)]$ which appears against c_1 in Table 2. For the second sequence $c_2 = ((1, 2), (1, 3), (1, 4), (2, 4), (2, 3))$, we have $v = \infty$, $w = 0$, $x = 1$, $y = b$, $z = a$ and so $\theta = -b/(a - b - 1)$, giving the limit-excluded linear functional $\phi_2 = [1, -b/(a - b - 1)]$. Similarly, we obtain ϕ_3. For the remaining 12 sequences c_4 to c_{15}, the pair $(1, 3)$ corresponds to ∞, that is $w = \infty$. Thus in these cases θ becomes $-(-xy + zx - zv + yz)/(-z + v)$. So, for example, $c_4 = ((1,3),(1,2),(1,4),(2,3),(2,4))$, so $w = \infty$, $v = 0$, $x = 1$, $y = a$, $z = b$ and hence $\theta = -(-a+b+ab)/(-b)$. Table 2 lists the 15 limit-excluded linear functionals obtained.

Table 2: The 15 excluded linear functionals for $d = 2$, $t = 4$

c_1	$\phi_1 = [1, -a/(-a+b-1)]$	c_9	$\phi_9 = [1, b(a-1)/(a-b)]$
c_2	$\phi_2 = [1, -b/(-b+a-1)]$	c_{10}	$\phi_{10} = [1, a(b-1)/(b-a)]$
c_3	$\phi_3 = [1, ab/(a+b-1)]$	c_{11}	$\phi_{11} = [1, (a-b)/(a-1)]$
c_4	$\phi_4 = [1, (ab-a+b)/b]$	c_{12}	$\phi_{12} = [1, (b-a)/(b-1)]$
c_5	$\phi_5 = [1, (ab+a-b)/a]$	c_{13}	$\phi_{13} = [1, ab]$
c_6	$\phi_6 = [1, -ab+a+b]$	c_{14}	$\phi_{14} = [1, b/a]$
c_7	$\phi_7 = [1, b(a-1)/(b-1)]$	c_{15}	$\phi_{15} = [1, a/b]$
c_8	$\phi_8 = [1, a(b-1)/(a-1)]$		

We note that these 15 limit-excluded linear functionals correspond to the 15 linear functionals calculated via geometrical means in [3]. Constructions there used these 15 excluded linear functionals to show the following existence result.

Theorem 3.1 *Optimal linear $(q^2, q, 4)$-perfect hash families exist for all prime powers q except when $q = 2, 3, 4, 5, 7, 8, 9, 13$.*

4 Case $d = 2$, $t = 5$

An optimal linear $(q^2, q, 5)$-perfect hash family has size $d(t-1) = 8$. Theorem 1.2 shows that they exist for $q \geq 10^8$. In this section we prove an existence bound that is much smaller. Further, we construct optimal linear $(q^2, q, 5)$-perfect hash families for much smaller q than the known examples.

We first list the known constructions of optimal linear $(q^2, q, 5)$-perfect hash families. They are all based on chains of subfields. The Blackburn-Wild [9] construction gives optimal linear $(q^2, q, 5)$-perfect hash families for q of the form $q = q_0^{\alpha_1\alpha_2\cdots\alpha_{d(t-2)}}$ where q_0 is any prime power and each $\alpha_i \geq d$. In [4], the authors construct optimal linear $(q^2, q, 5)$-perfect hash families for $q = r^2$, r a prime, $r \geq 31$; and for $q = r^4$, r a prime, $r \geq 11$ $(r \neq 13)$.

4.1 An Existence Bound

We use Method 2 from Section 2.2. As noted in Section 2.3, we only need to consider linear functionals of the form $[1, \alpha]$, $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$ when constructing our perfect hash family.

We are interested in $(2,5)$-limited $(d(t-1)-1)$-pair sequences from $T = \{(1,2),\ (1,3),\ (1,4),\ (1,5),\ (2,3),\ (2,4),\ (2,5),\ (3,4),\ (3,5),\ (4,5)\}$. So we only need to consider sequences of length 7 where each pair from T occurs at most once. Since $|T| = 10$, we can complete this sequence in 10 ways, but in only 3 ways without repeating a pair.

We wish to count the number of representative $(2,5)$-limited 7-pair sequences so that we can get a bound on the number of excluded functionals. Note that different sequences may result in the same excluded functional, so this count gives us an upper bound. Let c be a $(2,5)$-limited 7-pair sequence from T. Then, under the action of the symmetry group S_5, the remaining three pairs in T are one of four types: A, B, C or D which are given in Table 3.

Table 3: Sequence Types for $d = 2$ and $t = 5$

Case	remaining 3 pairs
A	$\{(1,2),(1,3),(1,4)\}$
B	$\{(1,2),(1,3),(4,5)\}$
C	$\{(1,2),(1,3),(2,3)\}$
D	$\{(1,2),(1,3),(2,4)\}$

We now provide an upper bound on the total number of sequences of each type.

A. This sequence is not $(2,5)$-limited as taking $A = \{2,3,4,5\}$ results in $|c_A| = 2(|A|-1) = 6$. So we can ignore this case.

B. To preserve a sequence of type B, we can permute $2,3$ and $4,5$. Hence sequences of type B are still type B under the action of the permutation group $\langle(23),(45)\rangle$ of size 4. This group has 4 elements, hence there are $7!/4 = 1260$ representative sequences of type B.

C. Sequences of type C remain type C under the action of the group $S_3 \times S_2$ of size 12 permuting $1,2,3$ and $4,5$. Hence there are $7!/12 = 420$ representative sequences of type C.

D. A sequence of type D remains a sequence of type D under the action of the permutation group $\langle(12)(34)\rangle$ of size 2. Hence there are $7!/2 = 2520$ representative sequences of type D.

For each $(2,5)$-limited sequence c above, consider repeating a pair (a,b) already occuring in the sequence. The condition generated by this pair, by the discussion in Section 2.3, is every two functionals are independent. This is already accounted for by considering only functionals of the form $[1,\alpha]$. Thus, we only need to consider adding the three pairs not already occuring in the sequence.

This gives a total of 4200 representative $(2,5)$-limited 7-pair sequences, each can be completed to an 8-pair sequence in three ways. Hence there are 4200×3 representative $(2,5)$-limited 8-pair sequences. Each 8-pair sequence results in one excluded linear functional of the form $[1,\alpha]$. Hence there at most 4200×3 excluded linear functionals of the form $[1,\alpha]$. However, to prove existence of a perfect hash family, we first need to verify that there exists a $(2,5)$-suitable 7-functional sequence Φ. So we need to show that firstly, every 6 elements of Φ form a $(q^2,q,4)$ perfect hash family, and secondly, $W_{c,\Phi}$ has maximal rank 7 for any $(2,5)$-limited 7-pair sequence c. We study the form of $W_{c,\Phi}$ in the next section and will prove that a $(2,5)$-suitable sequence exists if $q > 95$. The next theorem gives a bound on the existence of optimal linear $(q^2,q,5)$-perfect hash families assuming a $(2,5)$-suitable sequence exists.

Theorem 4.1 *If* $q+1 > 7+4200\times 3 = 12607$*, and there exists a* $(2,5)$*-suitable sequence, then there exists an optimal linear* $(q^2,q,5)$*-perfect hash family.*

Proof We assume there exists a $(2,5)$-suitable 7-functional sequence Φ. We can use this sequence together with the 4200 representative $(2,5)$-limited 7-pair sequences (each of which has 3 excluded functionals) to calculate the $7 + 4200 \times 3$ linear functionals in V^* of the form $[1,\alpha]$ that cannot be added to Φ to form a perfect hash family. There are $q+1$ linear functionals of the form $[1,\alpha]$, $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$, so provided $q + 1 > 12607$, there are functionals in V^* that we can add to Φ to form a $(q^2,q,5)$-perfect hash family. □

In fact we will show that $(q^2,q,5)$-perfect hash families exist for much smaller q than given in Theorem 4.1, since the excluded linear functionals are not necessarily distinct.

4.2 Sequence Analysis for $d = 2$ and $t = 5$

REWRITTEN THIS SECTION

In order to construct linear $(q^2,q,5)$-perfect hash families, we use a similar technique to that used in Section 3.1. We now study in detail the representative $(2,5)$-limited 7-pair sequences from each of the cases B, C and D given in Table 3.

Case B. Consider the 7-functional sequence Φ and 7-pair sequence c:

$$\begin{array}{rcl} \Phi & = & ([1,s],\quad [1,u],\quad [1,v],\quad [1,w],\quad [1,x],\quad [1,y],\quad [1,z]) \\ c & = & ((1,4),\quad (1,5),\quad (2,3),\quad (2,4),\quad (2,5),\quad (3,4),\quad (3,5)). \end{array}$$

Row reducing $W_{c,\Phi}$ enables us to calculate the excluded linear functionals for each of the 3 missing pairs $(4,5),(1,2),(1,3)$. They are:

$$B_{(4,5)} = \left[1, -\tfrac{-vwz-wxy+wzx+wzy-yzx+yvx}{-vy+zv-wz+vw-vx+xy}\right],$$

$$B_{(1,2)} = \left[1, \tfrac{-wuzv+zuvs+zuwy-vwsx+xwzs+xysv-xzys+wuvx-uvsx-wuxy+yusx-uysv-swuz+swuv}{zsv+wzy-zys-vsx-wxy+ysx-yzx+yvx+wzx+yuz-uvy-wuz+wuv-vwz)}\right],$$
$$B_{(1,3)} = \left[1, \tfrac{-uysv+yusx+swuv+yzsv-xzys-wzsv+xwzs+zuvs-uvsx-yuzv+yuvx-swuz+zuwy-wuxy}{-uvy+uxy+wuv-wux-yzx-vwz+wzx+zsv-vsx+yvx-wzs+wsx+wzy-wxy}\right].$$

Case C: Consider the 7-functional sequence Φ and 7-pair sequence c:

$$\begin{array}{rcl} \Phi & = & ([1,s], \quad [1,u], \quad [1,v], \quad [1,w], \quad [1,x], \quad [1,y], \quad [1,z]) \\ c & = & ((1,4), \quad (1,5), \quad (2,4), \quad (2,5), \quad (3,4), \quad (3,5), \quad (4,5)). \end{array}$$

The excluded linear functionals for the 3 missing pairs $(1,2),(1,3),(2,3)$ are:

$$C_{(1,2)} = \left[1, \tfrac{-(-uzv-wus+wuv+uvs-svw+wzs)}{-uv-zs+zv+ws+uz-wz}\right],$$
$$C_{(1,3)} = \left[1, \tfrac{-usx+uzx+uys-uxy+ysx-zys}{-(-ys+zy-uz+ux+zs-zx)}\right],$$
$$C_{(2,3)} = \left[1, \tfrac{-(-yzv+yvx-vwx+yvw-wxy+wzx)}{zy-wz+wx-vy+zv-zx}\right].$$

Case D: Consider the 7-functional sequence Φ and 7-pair sequence c:

$$\begin{array}{rcl} \Phi & = & ([1,s], \quad [1,u], \quad [1,v], \quad [1,w], \quad [1,x], \quad [1,y], \quad [1,z]) \\ c & = & ((1,4), \quad (1,5), \quad (2,3), \quad (2,5), \quad (3,4), \quad (3,5), \quad (4,5)). \end{array}$$

The excluded linear functionals for the 3 missing pairs $(1,2),(1,3),(2,4)$ are:

$$D_{(1,2)} = \left[1, \tfrac{-uysv+uvsx+yuzv-uzvx-vwsx-wusx-wuzv+wuzx+wuvx+wuys-wuxy+wysx+wzsv-wzys}{yuz+wys-wsx-wzy+wzx-uzv+uvx-uxy-ysv+ysx+yzv+zsv-zvx-zys}\right],$$
$$D_{(1,3)} = \left[1, \tfrac{-usx+uzx+uys-uxy+ysx-zys}{-(-ys+zy-uz+ux+zs-zx)}\right],$$
$$D_{(2,4)} = \left[1, \tfrac{-vwx-yzv+zvx+vwz+wxy-wzx}{xy+wy-wx-vy+zv-zy}\right].$$

The linear functional $B_{(4,5)}$ is the linear functional sub-excluded with $A = \{2,3,4,5\}$. More specifically, in the case $d = 2, t = 4$ with the 5-functional sequence and 5-pair sequence as follows:

$$\begin{array}{rcl} \Phi'' & = & ([1,v], \quad [1,w], \quad [1,x], \quad [1,y], \quad [1,z]) \\ c'' & = & ((2,3), \quad (2,5), \quad (3,4), \quad (3,5), \quad (4,5)). \end{array}$$

CAN YOU CHECK THIS?

A similar interpretation applies for the remaining sub-excluded linear functionals $C_{(1,2)}$, $C_{(1,3)}$, $C_{(2,3)}$, $D_{(1,3)}(= C_{(1,3)})$ and $D_{(2,4)}$ (with A equal to $\{1,2,4,5\}$, $\{1,3,4,5\}$, $\{2,3,4,5\}$, $\{1,2,3,5\}$ and $\{2,3,4,5\}$ respectively). These cases correspond to the fact that given any 5-functional subsequence of Φ, the 15 excluded linear functionals given in Table 2 for the case $d = 2, t = 4$ are sub-excluded functionals here. Thus the total number of (possibly distinct) sub-excluded functions are $\binom{7}{5} \times 15 = 350$.

The remaining three excluded functionals $B_{(1,2)}, B_{(1,3)}$ and $D_{(1,2)}$ are limit-excluded. This gives a revised total of $1260 \times 2 + 2520 + 350 = 5380$ excluded functions (without counting the original 7 functions in Φ).

Note that the symmetry group $\langle(23)\rangle$ maps the sequence $((1,4),(1,5),(2,3),(2,4),(2,5),(3,4),$ $(3,5))$ to $((1,4),(1,5),(2,3),(3,4),(3,5),(2,4),(2,5))$, hence $B_{(1,3)}$ is $B_{(1,2)}$ with the variables w and y swapped and x and z swapped.

4.3 The existence of $(2,5)$-suitable sequences

In this section we prove that there exists a $(2,5)$-suitable sequence Φ if $q > 95$. We first need to check that every 6 linear functionals of Φ form a $(q^2,q,4)$-perfect hash family. Secondly, we need to show that in Cases B, C and D, rank $W_{c,\Phi} = 7$. However, it is easy to see that in Cases B, C, D, the matrix $W_{c,\Phi}$ always has rank 7 if s,u,v,w,x,y,z are distinct.

For example, for Case B, we arrange the pairs in the order shown below, writing $*$ instead of the linear functions, we can consider $W_{c,\Phi}$:

$$W_{c,\Phi} = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \left[\begin{array}{c|cccc} * & & & * & \\ * & & & & * \\ \hline & * & * & & \\ & * & & * & \\ 0 & * & & & * \\ & & * & * & \\ & & * & & * \end{array}\right] \end{array}$$

The bottom five rows have rank 5 because every 6 functions form a $(q^2,2,4)$ perfect hash family. Thus the whole matrix has rank 7. A similar argument holds for cases C and D.

The next lemma proves the existence of $(2,5)$-suitable sequences for $q > 95$. Moreover, the proof gives a technique which we will use to construct suitable sequences.

Lemma 4.2 *If $q > 95$, there exists a $(2,5)$-suitable sequence.*

Proof Recall by Section 2.3 we can consider the linear functionals to be of the form $[1,\alpha]$, $\alpha \in \mathrm{GF}(q) \cup \{\infty\}$. Suppose q is a prime power greater than 95, then by Theorem 3.1 there exists a linear $(q^2,q,4)$ perfect hash family $S = \{\phi_1,\ldots,\phi_6\}$. For each 5-subset of S, the associated 5-functional sequence Φ' leads to 15 excluded linear functionals (given by Table 2). Including the original six linear functionals $\phi_1,\ldots,\phi_6$, this accounts for $6+6\times 15 = 96$ functionals. Hence if $q+1 > 96$, there is a functional $\phi = [1,\alpha] \in V^*$ which is not excluded. Thus $S \cup \{\phi\}$ is a set of 7 functionals, such that every 6 form a $(q^2,q,4)$ perfect hash family (and as discussed above, the rank condition is automatically satisfied), and so the associated 7-functional sequence is $(2,5)$-suitable. □

4.4 Strategy for constructing perfect hash families for $d = 2$, $t = 5$

To construct a $(q^2, q, 5)$-perfect hash family, we start with a $(2, 5)$-suitable 7-functional sequence Φ. To build Φ, we use the technique described in the proof of Lemma 4.2. We begin with S'', an optimal linear $(q^2, q, 4)$-perfect hash family (constructions for these are given in [3]). Note that S'' consists of 6 linear functionals. For each 5-functional subsequence of S'' we calculate the 15 excluded linear functionals given in Table 2. We then choose any non-excluded functional from V^* to add to S'' to give us a $(2, 5)$-suitable sequence Φ. $(2, 5)$-suitable sequences were investigated in [4] and examples are found for all primes $q \geq 31$. For example, $\Phi = ([1, \infty], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 12])$ is a $(2, 5)$-suitable sequence for all primes $q > 79$.

Given a $(2, 5)$-suitable 7-functional sequence Φ, we want to calculate the excluded linear functionals for each representative $(2, 5)$-limited 7-pair sequence. In this case, there are 4200 representative $(2, 5)$-limited 7-pair sequences, which is too many to calculate by hand as we did in the $d = 2, t = 4$ case. However, using a computer, we can easily find the excluded functionals for *all* the $(2, 5)$-limited 7-pair sequences. We choose a $(2, 5)$-suitable 7-functional sequence, for example, let $\Phi = ([1, \infty], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 12])$ (for $q > 79$). There are 7! ways of assigning s, u, v, w, x, y, z to be one of $\infty, 0, 1, 2, 3, 5, 12$. These different permutations correspond to choosing different 7-pair sequences c in the matrix $W_{c,\Phi}$. For each of these permutations, we calculate the nine excluded linear functionals $B_{(4,5)}, \ldots, D_{(2,4)}$ (given in Section 4.2). This method will generate all the excluded linear functionals, and we can then choose a non-excluded linear functional (distinct from the original seven) from V^* to add to Φ to form a perfect hash family.

For simplicity we looked for constructions in the case q a prime, $q < 50,000$. However, the excluded functionals in Section 4.2 can be used to construct $(q^2, q, 5)$-perfect hash families for any (large enough) value of q. We found many constructions of optimal linear $(q^2, q, 5)$-perfect hash families. In summary, using the $(2, 5)$-suitable sequence

$$\Phi = \{[0, 1], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 12]\},$$

we found constructions of $(q^2, q, 5)$-perfect hash families for primes $q = 359$, 397, 401, 433, 439, 443, 449, 461, and for all primes $467 \leq q \leq 50,000$, $q \neq 541$. The larger q is, the more $(q^2, q, 5)$-perfect hash families containing Φ there are. We conjecture that optimal linear $(q^2, q, 5)$-perfect hash families containing Φ will exist for all primes $q \geq 467$, $q \neq 541$.

The smallest example we found for $d = 2$ and $t = 5$ is a $(311^2, 311, 5)$-perfect hash family $S = \{[0, 1], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 16], [1, 87]\}$. In fact, starting with the $(q^2, q, 4)$-perfect hash family $S'' = \{[0, 1], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5]\}$, we found examples for $d = 2$ and $t = 5$ for all primes $347 \leq q \leq 50,000$. Our search was by no means complete, as we only used a few suitable sequences to base our search on. It seems likely that using a different starting suitable sequence may lead to examples of optimal linear $(q^2, q, 5)$-perfect hash families for $q < 311$. For the interested reader, Table 4 lists the smallest values of q for which we found examples.

Table 4: Small values of q for which optimal linear $(q^2, q, 5)$-perfect hash families $S = \{[1, s], [1, u], [1, v], [1, w], [1, x], [1, y], [1, z], [1, i]\}$ exist.

q	s, u, v, w, x, y, z	i
311	$\infty, 0, 1, 2, 3, 5, 16$	$87, 264$
317	$\infty, 0, 1, 2, 3, 5, 19$	179
331	$\infty, 0, 1, 2, 3, 5, 14$	$75, 295$
337	$\infty, 0, 1, 2, 3, 5, 20$	268
347	$\infty, 0, 1, 2, 3, 5, 20$	227
359	$\infty, 0, 1, 2, 3, 5, 12$	329
367	$\infty, 0, 1, 2, 3, 5, 26$	334

We also wanted to find a general example of an optimal linear $(q^2, q, 5)$-perfect hash family. $S = \{[0, 1], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 12], [1, 44]\}$ is an optimal linear $(q^2, q, 5)$-perfect hash family for primes q, $2400 \leq q \leq 50,000$, except for $q = 2543$, 2579 and 2843. (In these cases $\{[0, 1], [1, 0], [1, 1], [1, 2], [1, 3], [1, 5], [1, 12], [1, 46]\}$ is a perfect hash family.) This set S is also a perfect hash family for many smaller values of q. We conjecture that S is a perfect hash family for all primes $q > 2843$.

5 Case $d = 3, \ t = 3$

An optimal linear $(q^3, q, 3)$-perfect hash family has size $d(t-1) = 6$. We start by considering the representative $(3, 3)$-limited 5-pair sequences from $T^2 = \{(1, 2), (1, 3), (2, 3)\}$. Using the action of the symmetric group S_3 on 1,2,3 we may assume the missing pair is $(2, 3)$. To be $(3, 3)$-limited, each pair occurs at most twice. So there are $(5!/(2!2!))$ $(3, 3)$-limited sequences as the pairs $(1, 2)$ and $(1, 3)$ each occur exactly twice. However, the group of size 2 swapping 2 and 3 fixes the missing pair $(2, 3)$. Hence there are $5!/(2!2! \times 2) = 15$ sequences to consider.

To build a $(q^3, q, 3)$-perfect hash family, we need to start with a $(3, 3)$-suitable 5-functional sequence Φ. To be suitable, we need any $d(t-2) = 3$ functionals in Φ to form a $(q^3, 3, 2)$ perefect hash family, that is, every three functionals are independent. This means we can use a special form for Φ to make our calculations easier. We use the fact that without loss of generality, any five elements, no three dependent, in $\mathrm{GF}(q)^3$ can be written as $[1, \alpha, \alpha^2]$ for five distinct values of $\alpha \in \mathrm{GF}(q)$ (see [11], [4]). Hence we may consider the following 5-functional sequence Φ and the general $(3,3)$-limited representative 5-pair sequence c:

$$\begin{array}{rcccccc} \Phi & = & ([1,v,v^2], & [1,w,w^2], & [1,x,x^2], & [1,y,y^2], & [1,z,z^2]) \\ c & = & ((2,3), & (1,2), & (1,3), & (1,3), & (1,2)). \end{array}$$

So we have

$$W_{c,\Phi} = \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & v & v^2 & -1 & -v & -v^2 \\ 1 & w & w^2 & -1 & -w & -w^2 & 0 & 0 & 0 \\ 1 & x & x^2 & 0 & 0 & 0 & -1 & -x & -x^2 \\ 1 & y & y^2 & 0 & 0 & 0 & -1 & -y & -y^2 \\ 1 & z & z^2 & -1 & -z & -z^2 & 0 & 0 & 0 \end{array}\right].$$

Our choice of functionals ensures that any $d = 3$ elements in Φ are independent provided v, w, x, y, z are distinct. The remaining condition for Φ to be $(3,3)$-suitable, is that rank $W_{c,\Phi} = 5$ for all 3-limited 5-pair sequences c. Rearranging the rows of the matrix $W_{c,\Phi}$ as

$$\left[\begin{array}{cccc} 0 & * & & * \\ * & 0 & & * \\ * & 0 & & * \\ \hline * & * & \vdots & 0 \\ * & * & \vdots & 0 \end{array}\right]$$

we can see that the matrix $W_{c,\Phi}$ has rank 5 for any distinct v, w, x, y, z. This is true for any permutation of c, thus Φ is a $(3,3)$-suitable functional sequence for any distinct v, w, x, y, z.

By inspection, adding the pair $(1,2)$ to c results in a set of sub-excluded functionals which is $\langle [1,w,w^2], [1,z,z^2] \rangle$. A similar result holds for the pair $(1,3)$.

Now consider the pair $(2,3)$. It completes c to the only $(3,3)$-limited 6-pair sequence. Thus we need to perform row operations on $W_{c,\Phi}$ to find the limit-excluded linear functional corresponding to this last pair $(2,3)$. Row reductions give us two rows in the right form, namely $(0,0,0,1,v,v^2,-1,-v,-v^2)$ and $(0,0,0,a,b,c,-a,-b,-c)$ where $a = z+w-x-y$, $b = wz-xy$, $c = zw(x+y) - xy(z+w)$. Thus any linear functional which is a linear combination of these is excluded. A general linear functional has form $[\ell, m, n]$, so we exclude all such functionals with

$$\begin{vmatrix} \ell & m & n \\ 1 & v & v^2 \\ a & b & c \end{vmatrix} = 0. \qquad (3)$$

We now consider the special case where the final linear functional $[\ell, m, n]$ is of the form $[1, \theta, \theta^2]$. Assume that all parameters v, w, x, y, z, θ are distinct. This will ensure that any three functionals in our perfect hash family are independent (and so we need not consider the sub-excluded linear functionals) Solving equation (3) gives us the limit-excluded functionals $[1, \theta, \theta^2]$ where

$$\theta = -\frac{xy(v-w) + wz(x-v) + yz(w-x)}{z(v-w) + y(x-v) + v(w-x)}.$$

That is, $w_{(2,3),[1,\theta,\theta^2]}$ is in the row space of $W_{c,\Phi}$ if θ has this form.

We note that this is exactly the same expression for θ as in the $t = 2$, $d = 4$ case. This correspondence was noted in [4] where this case was studied using geometric methods, however the geometric interpretation did not explain the correspondence, whereas the sequence technique does. More specifically, the correspondence between the sequences is:

$$\begin{array}{lrclllll}
d = 3, t = 3 & \Phi & = & ([1, v, v^2], & [1, w, w^2], & [1, x, x^2], & [1, y, y^2], & [1, z, z^2]) \\
 & c & = & ((2,3), & (1,2), & (1,3), & (1,3), & (1,2)) \\
\hline
d = 2, t = 4 & \Phi & = & ([1, v], & [1, w], & [1, x], & [1, y], & [1, z]) \\
 & c & = & ((1,2), & (1,3), & (1,4), & (2,3), & (2,4))
\end{array}$$

So in this special case we can use the construction results from the $d = 2$, $t = 4$ case. This gives us constructions of $(q^3, q, 3)$-perfect hash families for all prime powers q except $q = 2, 3, 5, 7, 8, 9, 13$. Using the general form $[\ell, m, n]$ in equation (3) allows us to construct a $(q^3, q, 3)$-perfect hash family for the additional case $q = 13$. Constructions of $(q^3, q, 3)$-perfect hash families for prime powers $q \geq 11$ are given in [4].

6 Case $d = 4$, $t = 3$

An optimal linear $(q^4, q, 3)$ perfect hash family has $d(t-1) = 8$ elements. Once again we use Method 2 from Section 2.2. We have $T = \{(1,2), (1,3), (2,3)\}$, and we first count the number of representative $(4,3)$-limited 7-pair sequences from T. Since each pair from T occurs at most 3 times in our sequence, we have 9 pairs to choose from. We consider the different possibilities for the two pairs that are not in the sequence. Under the symmetric group S_3 acting on $\{1, 2, 3\}$ there are 2 possibilities for these remaining two pairs.

Case A. The remaining two pairs are $(2,3), (2,3)$. The number of $(4,3)$-limited representative 7-pair sequences c (with pairs from $\{(1,2), (1,2), (1,2), (1,3), (1,3), (1,3), (2,3)\}$) is: $7!/(3!3!1! \times 2) = 70$.

Case B. The remaining two pairs are $(1,3), (2,3)$. The number of $(4,3)$-limited representative 7-pair sequences c (with pairs from $\{(1,2), (1,2), (1,2), (1,3), (1,3), (2,3), (2,3)\}$) is: $7!/(3!2!2! \times 2) = 105$.

In both cases, an 8-pair sequence will lead to a subspace of excluded functionals. For this case, we restrict our attention to functionals of a special form. This will make it easier to find constructions, but will mean that we do not completely answer the existence question. The sub-exclusion condition is equivalent to every $d(t-2)$ functions of a $(q^4, q, 3)$-perfect hash family being independent, so we consider the special case using only linear functionals that have form $[1, \alpha, \alpha^2, \alpha^3]$ for $\alpha \in \mathrm{GF}(q)$. Note that this means the sub-exclusion condition is automatically satisfied provided we use distinct values of α for our functionals.

We consider Case A and the 7-functional sequence ϕ and 7-pair sequence c.

$$\begin{array}{rclllllll} \Phi & = & ([1,s,s^2,s^3] & [1,u,u^2,u^3] & [1,v,v^2,v^3] & [1,w,w^2,w^3] & [1,x,x^2,x^3] & [1,y,y^2,y^3] & [1,z,z^2,z^3]) \\ c & = & ((1,2) & (1,2) & (1,2) & (1,3) & (1,3) & (1,3) & (2,3)). \end{array}$$

We construct $W_{c,\Phi}$ and note that the matrix $W_{c,\Phi}$ has rank 7 provided s, u, v, w, x, y, z are distinct. We proceed as before, performing row reductions on $W_{c,\Phi}$ to find the excluded linear functionals. Completing the sequence with (12) or (13) repeates the condition that every four linear functionals are independent. Completing to (23) to obtain limit-excluded linear functions $(0, 0, 0, 0, 1, z, z^2, z^3, -1, -z, -z^2, -z^3)$, $(\underline{0}, \phi_1, -\phi_1)$, $(\underline{0}, \phi_2, -\phi_2)$, for certain ϕ_1, ϕ_2. Any linear combination of these 3 functionals is limit-excluded. However, we are considering the special case where we only want to add a functional to Φ that has form $[1, \alpha, \alpha^2, \alpha^3]$. To find the limit-excluded functionals of this form we solve

$$\begin{vmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ 1 & z & z^2 & z^3 \\ & \phi_1 & & \\ & \phi_2 & & \end{vmatrix} = 0.$$

There are 3 solutions for $\alpha \in \mathrm{GF}(q)$, one solution being $\alpha = z$ which corresponds to a functional in Φ (and so corresponds to a sub-excluded functional). Thus there are at most two limit-excluded functionals of the form $[1, \alpha, \alpha^2, \alpha^3]$.

We undertake a similar process for a sequence of type B. The matrix $W_{c,\Phi}$ has rank 7 for any distinct s, u, v, w, x, y, z, so we can conclude that Φ is $(4, 3)$-suitable for distinct s, u, v, w, x, yz. Here the two pairs missing from c are different, so we will get a different subspace of excluded functionals for each missing pair. Each of these subspaces excludes three linear functionals of the form $[1, \alpha, \alpha^2, \alpha^3]$. Two of these are functionals in Φ (and so are sub-excluded). Hence for each of the two pairs missing from c, we have one limit-excluded functional of the form $[1, \alpha, \alpha^2, \alpha^3]$.

In total there are 175 representative pair sequences, each with at most 2 limit-excluded linear functionals, so there are at most 350 limit-excluded linear functionals. Since there are q linear functionals of the form $[1, \alpha, \alpha^2, \alpha^3]$, $\alpha \in \mathrm{GF}(q)$, then provided $q > 7 + 350 + 1 = 358$, an optimal $(q^4, q, 3)$-perfect hash family will exist. We have proved:

Theorem 6.1 *Optimal linear $(q^4, q, 3)$-perfect hash families exist for all prime power $q > 358$.*

In order to construct optimal linear $(q^4, q, 3)$-perfect hash families, we begin with a $(4, 3)$-suitable 7-functional sequence Φ; by the above argument, any choice of distinct s, u, v, w, x, y, z for Φ will do. We chose $0, 1, 2, 3, 4, 5, 6$, that is,

$$\Phi = ([1, 0, 0, 0], [1, 1, 1, 1], [1, 2, 2^2, 2^3], [1, 3, 3^2, 3^3], [1, 4, 4^2, 4^3], [1, 5, 5^2, 5^3], [1, 6, 6^2, 6^3]).$$

For each of the 7! arrangements of $\{s, u, v, w, x, y, z\} = \{0, 1, 2, 3, 4, 5, 6\}$, we calculate the 2 solutions for α obtained from Case A, and the 2 solutions for α from Case B. Each arrangement excludes (at most) 4 values of α. Once these are calculated, any value $\beta \in \mathrm{GF}(q) \backslash \{0, 1, 2, 3, 4, 5, 6\}$ not excluded will give us a linear functional $[1, \beta, \beta^2, \beta^3]$ which we can add to Φ to make a perfect hash family. We did some computer searches and found that optimal linear $(q^4, q, 3)$-perfect hash families exist for q much smaller than the bound of 358. Our program searched for such perfect hash families for each prime $q < 2000$ and Table 5 lists an example of a perfect hash family containing S for q where they exist. Note that the larger q is, the more values of β that can be used to make a perfect hash family.

Result 6.2 *Let* $S = \{[1, 0, 0, 0], [1, 1, 1, 1], [1, 2, 2^2, 2^3], [1, 3, 3^2, 3^3], [1, 4, 4^2, 4^3], [1, 5, 5^2, 5^3], [1, 6, 6^2, 6^3]\}$. *For each prime* q, *and* $\beta \in GF(q)$, *we checked when* $S \cup \{[1, \beta, \beta^2, \beta^3]\}$ *is an optimal linear* $(q^4, q, 3)$*-perfect hash family. Table 5 lists an example of a perfect hash family* $S \cup \{[1, \beta, \beta^2, \beta^3]\}$ *for each prime* $q < 2000$ *where a* β *exists.*

Table 5: Examples of $(q^4, q, 3)$-perfect hash families $S \cup \{[1, \beta, \beta^2, \beta^3]\}$ for prime $q < 2000$

q	β
47	24
59	21
67	34
79	37
83	42
89	43
97,109	26
101,137	11
103,107,151,163,173,179,191,193,197,199,211,227,241,251,263,269	12
113,127,131,223,229,233,239,271,277,281,349,521,541,599	13
139,157,257	17
149	19
$283 \leq q \leq 2000$, $q \neq 349, 521, 541, 599$	12

Table 5 shows that

$$\{[1, 0, 0, 0], [1, 1, 1, 1], [1, 2, 2^2, 2^3], [1, 3, 3^2, 3^3], [1, 4, 4^2, 4^3], [1, 5, 5^2, 5^3],$$

$$[1, 6, 6^2, 6^3], [1, 12, 12^2, 12^3]\}$$

is an optimal linear $(q^4, q, 3)$ perfect hash family for all prime q, $599 < q < 2000$. We conjecture that this set will be a perfect hash family for all primes $q > 599$.

7 Conclusion

In this paper we have developed a technique which can be used to construct optimal linear (q^d, q, t)-perfect hash families. We have illustrated the technique by constructing new optimal linear $(q^2, q, 5)$- and $(q^4, q, 3)$-perfect hash families. This technique is straightforward to use for small d and t. It can be used for larger d, t, but the size of d and t will be limited by computational capacity.

References

[1] M. Atici, S. S. Magliveras, D. R. Stinson and W. -D. Wei. Some recursive constructions for perfect hash families. *J. Combin. Designs.* 4 (1996) 353-363.

[2] M. Atici, D. R. Stinson and W. -D. Wei. A new practical algorithm for the construction of a perfect hash function. *J. Combin. Math. Combin. Comput.* 35 (2000), 127-145.

[3] S. G. Barwick, W.-A. Jackson and C. T. Quinn. Optimal linear perfect hash families with small parameters. *J. Combin. Designs*, 12 (2004) 311-324.

[4] S. G. Barwick and W.-A. Jackson. Geometrical constructions of optimal linear perfect hash families. Preprint.

[5] A. Beutelspacher and U. Rosenbaum. *Projective Geometry: From Foundations to Applications.* Cambridge University Press, 1998.

[6] S. R. Blackburn. Combinatorics and threshold cryptography. Combinatorial designs and their applications, *CRC Research Notes in Mathematics*, 403 (1999) 49-70.

[7] S. R. Blackburn. Perfect hash families: probabilistic methods and explicit constructions. *J. Combin. Theory Ser A.* 92 (2000) 54-60.

[8] S. R. Blackburn, M. Burmester, Y. Desmedt and P. R. Wild. Efficient multiplicative sharing schemes. Advances in Cryptology – EUROCRYPT '96. *Lecture Notes in Computer Science* 1070 (1996) 107-118.

[9] S. R. Blackburn and P. R. Wild. Optimal linear perfect hash families, *J. Combin. Theory Ser. A*, 83 (1998) 233–250.

[10] A. Fiat and M. Naor. Broadcast Encryption. Advances in Cryptology – CRYPTO '93. *Lecture Notes in Computer Science* 773 (1994) 480-491.

[11] J. W. P. Hirschfeld. *Projective Geometry over Finite Fields.* Oxford Mathematical Monographs, Oxford, UK, 1998.

[12] D. R. Hughes and F. C. Piper. *Projective planes.* Springer-Verlag, Berlin-Heidelberg-New York, 1973.

[13] K. Mehlhorn. *Data Structures and Algorithms 1: Sorting and Searching.* Springer Verlag, Berlin, 1984.

[14] D. R. Stinson, T. van Trung and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *J. Statist. Plan. Infer.*, 86 (2000) 595–617.

[15] H. Wang and C. Xing. Explicit constructions of perfect hash families from algebraic curves over finite fields. *J. Combin. Theory* Ser A. 93 (2001) 112-124.

Appendix: Notation Index

This appendix contains an index of the notation and definitions introduced in Section 2.1.

$V = \mathrm{GF}(q)^d$

$V^* = \{\phi : V \to \mathrm{GF}(q)\}$

$P = (p_1, \ldots, p_t) \in V^t$

$U_P = \left\{(\psi_1, \ldots, \psi_t) \in (V^*)^t : \sum_{i=1}^t p_i^{\psi_i} = 0\right\}$

$T = \{1, \ldots, t\}$

$T^2 = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : 1 \le a < b \le t\}$

$w_{(a,b),\phi} = (\psi_1, \ldots, \psi_t) \in (V^*)^t$ where

$$\psi_i = \begin{cases} \phi & \text{if } i = a, \\ -\phi & \text{if } i = b, \qquad (\text{for } (a,b) \in T,\ \phi \in V^*) \\ 0 & \text{if } i \neq a, b \end{cases}$$

$V_{(a,b)} = \{w_{(a,b),\phi} : \phi \in V^*\} \subseteq (V^*)^t$, for $(a, b) \in T$

$W_{c,\Phi} = \left\langle w_{(a_i,b_i),\phi_i} : 1 \le i \le k \right\rangle \subseteq (V^*)^t$,
where $c = ((a_1, b_1), \ldots, (a_k, b_k)) \in T^k, \Phi = (\phi_1, \ldots, \phi_k) \in (V^*)^k$

$R = \{(p, \ldots, p) : p \in V\} \subseteq V^t$

$Y = \{(\psi_1, \ldots, \psi_t) \in (V^*)^t : \sum_{i=1}^{t} \psi_i = 0\}$

$\dim L$ refers to the dimension of the subspace L over $\mathrm{GF}(q)$

A *k-pair sequence* $c = ((a_1, b_1), \ldots, (a_k, b_k))$ on T is a sequence of k pairs from T.

A *k-functional sequence* $\Phi = (\phi_1, \ldots, \phi_k) \in (V^*)^k$ is a sequence of k functionals from V^*.

For $A \subseteq T$, c_A consists of those pairs (a_i, b_i) with $a_i, b_i \in A$. Φ_A consists of those functionals corresponding to the pairs in c_A.

A *(d, t)-limited* k-pair sequence is a k-pair sequence on T where for each proper subset A of T we have $|c_A| < d(|A| - 1)$.

A *(d, t)-suitable* sequence Φ is a $(d(t-1)-1)$-functional sequence such that every $d(t-2)$ elements for a $(q^d, d, t-1)$ perfect hash family, and for every (d, t)-limited $(d(t-1)-1)$-pair sequence c, we have rank $W_{c,\Phi} = d(t-1) - 1$.

The *representative* k-pair sequences is the collection of one sequence from each isomorphism group under the action of the symmetry group S_t.