



ε -Almost collision-flat universal hash functions and mosaics of designs

Moritz Wiese¹ · Holger Boche¹

Received: 7 June 2023 / Revised: 11 October 2023 / Accepted: 12 October 2023 /
Published online: 18 November 2023
© The Author(s) 2023

Abstract

We introduce, motivate and study ε -almost collision-flat universal (ACFU) hash functions $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$. Their main property is that the number of collisions in any given value is bounded. Each ε -ACFU hash function is an ε -almost universal (AU) hash function, and every ε -almost strongly universal (ASU) hash function is an ε -ACFU hash function. We study how the size of the seed set $\mathcal{S}$ depends on ε , $|\mathcal{X}|$ and $|\mathcal{A}|$. Depending on how these parameters are interrelated, seed-minimizing ACFU hash functions are equivalent to mosaics of balanced incomplete block designs (BIBDs) or to duals of mosaics of quasi-symmetric block designs; in a third case, mosaics of transversal designs and nets yield seed-optimal ACFU hash functions, but a full characterization is missing. By either extending $\mathcal{S}$ or $\mathcal{X}$, it is possible to obtain an ε -ACFU hash function from an ε -AU hash function or an ε -ASU hash function, generalizing the construction of mosaics of designs from a given resolvable design (Gnilke et al. in Des. Codes Cryptogr. 86(1):85–95, 2017). The concatenation of an ASU and an ACFU hash function again yields an ACFU hash function. Finally, we motivate ACFU hash functions by their applicability in privacy amplification.

Keywords Universal hash function · Mosaic of designs · Balanced incomplete block design · Privacy amplification

Communicated by C. J. Colbourn.

This work was presented in part at the 2023 IEEE International Symposium on Information Theory and at the 29th Nordic Congress of Mathematicians. A short and preliminary version of this paper is [31]. It mainly contains: two of the lower bounds on the size of an ACFU hash function (Theorem 3.4) without proof; a less general version of Theorem 4.1, which shows how to construct an ACFU hash function from an almost universal hash function; all examples without the discussion of their design-theoretic properties; a discussion of the application of ACFU hash functions to privacy amplification which is more detailed than here (Section 5) and gives a better bound on the attained security level.

✉ Moritz Wiese
wiese@tum.de

Holger Boche
boche@tum.de

¹ Technical University of Munich, Chair of Theoretical Information Technology and BMBF Research Hub 6G-Life, Arcisstraße 21, München 80333, Germany

Mathematics Subject Classification 05B05 · 68R05 · 94A60

1 Introduction

Let $\mathcal{X}, \mathcal{S}, \mathcal{A}$ be finite sets. For $\varepsilon \geq 0$, we call a function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ an ε -almost collision-flat universal (ε -ACFU) hash function if

(ACFU1) for every $x \in \mathcal{X}$ and every $\alpha \in \mathcal{A}$,

$$|\{s : f(x, s) = \alpha\}| = \frac{|\mathcal{S}|}{|\mathcal{A}|},$$

(ACFU2) for all distinct $x, x' \in \mathcal{X}$ and every $\alpha \in \mathcal{A}$,

$$|\{s : f(x, s) = f(x', s) = \alpha\}| \leq \frac{\varepsilon |\mathcal{S}|}{|\mathcal{A}|}.$$

We call f nontrivial if $2 \leq |\mathcal{A}| < |\mathcal{X}|$. The set $\mathcal{S}$ is called the *seed set* of f ; occasionally, we will call $\mathcal{X}$ the *point set* of f . In this paper, we are going to motivate ε -ACFU hash functions and study their properties.

There are two well-known related types of hash functions which we will mention and use frequently. An ε -almost universal (ε -AU) hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ satisfies

(AU) for all distinct $x, x' \in \mathcal{X}$,

$$|\{s : f(x, s) = f(x', s)\}| \leq \varepsilon |\mathcal{S}|.$$

This definition goes back to Stinson [24] as a generalization of *universal hash functions* due to Carter and Wegman [7], where $\varepsilon = 1/|\mathcal{A}|$. Finally, an ε -almost strongly universal (ε -ASU) hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ satisfies the properties

(ASU1) for every $x \in \mathcal{X}$ and every $\alpha \in \mathcal{A}$,

$$|\{s : f(x, s) = \alpha\}| = \frac{|\mathcal{S}|}{|\mathcal{A}|}$$

(so this is the same as (ACFU1)),

(ASU2) for all distinct $x, x' \in \mathcal{X}$ and all $\alpha, \alpha' \in \mathcal{A}$,

$$|\{s : f(x, s) = \alpha, f(x', s) = \alpha'\}| \leq \frac{\varepsilon |\mathcal{S}|}{|\mathcal{A}|}.$$

Strongly universal (SU) hash functions, where the above properties hold with $\varepsilon = 1/|\mathcal{A}|$, were defined by Wegman and Carter [28], and the above definition of ε -ASU hash functions is again due to Stinson [24]. We also refer to the $\mathcal{S}$ sets of AU and ASU hash functions as their seed sets.

Remark 1 Sometimes ε -AU hash functions are defined in terms of the parameter $\varepsilon|\mathcal{A}|$, e.g., [13, 25]. Our convention is the same as in [24].

For a hash function f , the event that $f(x, s) = f(x', s)$ for distinct x, x' is usually referred to as a *collision*. If f is an AU hash function, it is not important in which value of f a collision occurs, only the total number of collisions has to be bounded. In contrast, for an ACFU hash function, we are also interested in the value of f when a collision occurs. In the range of numbers $|\{s : f(x, s) = f(x', s) = \alpha\}|$, where $x \neq x'$ are fixed and where α ranges over all of $\mathcal{A}$, there should be no peaks (valleys are allowed). Hence the name “collision-flat”.

The following relations between the three types of hash functions are obvious.

Lemma 1 *Let $\varepsilon \geq 0$.*

- (1) *Every ε -ACFU hash function is an ε -AU hash function.*
- (2) *Every ε -ASU hash function is an ε -ACFU hash function.*

Our motivation for studying ACFU hash functions comes from their applicability in *privacy amplification*. This is a key step in *secret key generation*, where two parties A and B want to generate a shared secret key from correlated random observations and public discussion, while an adversary can observe their public communication and possibly also makes observations correlated to the key-generating parties' ones [6, Chap. 4]. The privacy amplification step takes place once the two parties have agreed on a shared random variable X . The result of privacy amplification is a key shared by A and B which is secret with respect to the adversary. We formalize this setting in Sect. 5.

The standard choice of function to use in privacy amplification so far has been an arbitrary AU hash function [3, 4]; seeded extractors are also suitable [17]. These functions need additional randomness, a *seed*, as a second input (in the definitions above, from the set $\mathcal{S}$), and they guarantee security against all adversaries for which the entropy of X conditional on their own observation is sufficiently large. It is desirable to make the seed small, since its generation is costly and it must be known by both A and B.

For general AU hash functions and extractors, key security so far has only been proved under the assumption that the adversary has no other information about the key than the knowledge of the underlying joint probability distribution together with its observations. A stronger security test goes as follows: The adversary has the above information, and additionally, it has to distinguish between two arbitrary possible key values. Security (*key indistinguishability*) is declared if the adversary's decision performance is no more than negligibly better than random guessing for all such value pairs. In all cases we know of where this stronger security criterion has been applied and where security is guaranteed against all X with sufficiently high conditional entropy, the AU hash functions which were used for privacy amplification or in a related problem, the wiretap channel problem, actually are ACFU hash functions [14, 30]. (These functions will be considered below as examples.) For this reason, ACFU hash functions are worth a closer look. In Sect. 5, we sketch how key indistinguishability can be proven in a privacy amplification scenario using ACFU hash functions.

As we have said already, the second argument s to an ACFU hash function will be chosen uniformly at random. Since randomness is expensive, our first question about ACFU hash functions (Sect. 3) is how small the set $\mathcal{S}$ can be, given ε and the cardinalities of $\mathcal{X}$ and $\mathcal{A}$. We derive three lower bounds, each of which is relevant in different regimes depending on the relation of ε , $|\mathcal{X}|$, $|\mathcal{A}|$. We also discuss which structure an ACFU hash function attaining equality in any of the bounds has.

It turns out that if ε is optimal (i.e., minimal given $|\mathcal{X}|$ and $|\mathcal{A}|$), then an ε -ACFU hash function attains equality in the corresponding lower bound if and only if its underlying structure is that of a mosaic of balanced incomplete block designs (BIBDs) [11]. The two other lower bounds on the seed size hold for general ε . One of them holds for small ε , but may be void if $|\mathcal{A}|$ is large relative to $|\mathcal{X}|$ (roughly $|\mathcal{A}|^2 > |\mathcal{X}|$). If this bound applies, the ACFU hash functions satisfying equality correspond precisely to the duals of mosaics of quasi-symmetric BIBDs [30]. In the remaining bound, we do not have a full characterization of the ACFU hash functions attaining equality, but we show that equality is attained by mosaics of transversal designs or of nets. Due to these facts, we believe that ACFU hash functions also are interesting objects of study in themselves. In [30], the authors studied mosaics of balanced incomplete block designs and of group divisible designs and applied them to privacy

amplification as well as to another problem from information-theoretic security, the wiretap channel. The results of the present paper embed such mosaics in the wider picture of ACFU hash functions. The necessary design-theoretic definitions are collected in Sect. 2.

In addition to studying how small $\mathcal{S}$ can be for given ε and $|\mathcal{X}|, |\mathcal{A}|$, we also give three methods for constructing ACFU hash functions in Sect. 4. By the first one, based on an extension of the seed set with the help of a quasigroup (latin square), one obtains an ε -ACFU hash function from another function if and only if the latter is an ε -AU hash function. This can be seen as a generalization of the method proposed in [11] for constructing a mosaic of BIBDs from a resolvable BIBD, and also as a generalization of a method used frequently to construct ASU hash functions from AU hash functions. The second method is the dual of the first and produces an ε -ACFU hash function if and only if the original function is an ε -ASU hash function. In the case where the quasigroup in fact is an abelian group, we characterize those ACFU hash functions which can be derived by both methods as “double extensions” of ε -balanced functions. By the third construction method, one obtains an ε -ACFU hash function by concatenating an ε_1 -ACFU hash function and an ε_2 -ASU hash function, similar to methods used by Stinson [24] for constructing AU and ASU hash functions from other AU or ASU hash functions. For all these methods, we discuss if and how seed optimality of the original functions can lead to seed optimality of the resulting functions.

For a practical application of ACFU hash functions, it will be important to find examples which have a low computational complexity. All our examples can be computed efficiently, although their complexities differ in concrete detail. We will not consider this issue further; more on this was said in [30] on the examples given there, some of which we will meet later on in this paper (Examples 1, 3 and 4).

2 Preliminaries on (mosaics of) incidence structures

An *incidence structure* is a triple $D = (\mathcal{X}, \mathcal{S}, I)$, where $\mathcal{X}$ and $\mathcal{S}$ are finite sets and I is an incidence relation on $\mathcal{X} \times \mathcal{S}$. $\mathcal{X}$ is called the *point set* and $\mathcal{S}$ the *block index set* of D . We also say that D is an incidence structure on $\mathcal{X}$. The sets

$$\{x \in \mathcal{X} : x \text{ incident with } s\},$$

where $s \in \mathcal{S}$, are usually referred to as the *blocks* of D . The *dual* of D is the incidence structure $\tilde{D}$ on $\mathcal{S}$ with block index set $\mathcal{X}$ and s incident with x in $\tilde{D}$ if and only if x is incident with s in D .

An incidence structure D is *resolvable* if the block index set can be partitioned into *parallel classes* of constant size in such a way that for each parallel class, a point x is incident with a unique element of the class. (Note that when considering the partition of $\mathcal{X}$ into the blocks corresponding to a given parallel class, some of these blocks may be empty.) If D is resolvable, one can index the block indices of each parallel class by a set $\mathcal{A}$, and the block index set $\mathcal{S}$ of D can be written as a Cartesian product $\mathcal{S} = \mathcal{H} \times \mathcal{A}$, where $\mathcal{H}$ is an index set for the parallel classes.

Two incidence structures $D = (\mathcal{X}, \mathcal{S}, I)$ and $D' = (\mathcal{X}', \mathcal{S}', I')$ are called *isomorphic* if there exist bijective mappings $\phi : \mathcal{X} \rightarrow \mathcal{X}'$ and $\psi : \mathcal{S} \rightarrow \mathcal{S}'$ such that any $x \in \mathcal{X}$ and $s \in \mathcal{S}$ are incident in D if and only if $\phi(x)$ is incident with $\psi(s)$ in D' .

A *balanced incomplete block design (BIBD)* on a finite set $\mathcal{X}$ is an incidence structure $(\mathcal{X}, \mathcal{S}, I)$ for which there exist positive integers k and λ such that

- (1) each $s \in \mathcal{S}$ is incident with exactly k elements of $\mathcal{X}$,

- (2) for any two distinct $x, x' \in \mathcal{X}$, there are exactly λ elements of $\mathcal{S}$ incident with both x and x' .

If $|\mathcal{X}| = v$, then such a BIBD is called a $\text{BIBD}(v, k, \lambda)$. We call a BIBD *nontrivial* if $1 < k < v$.

In a $\text{BIBD}(v, k, \lambda)$ with $|\mathcal{S}| = b$, a point x is incident with exactly r block indices s , where r satisfies

$$bk = vr. \quad (2.1)$$

Another important relation is

$$\lambda(v - 1) = r(k - 1). \quad (2.2)$$

A BIBD is *quasi-symmetric* if there are two distinct numbers μ_1, μ_2 (the *intersection numbers*) such that two blocks intersect in either μ_1 or μ_2 points. By definition, the only BIBDs which are both resolvable and quasi-symmetric are the *affine designs*; they satisfy

$$b = v + r - 1. \quad (2.3)$$

A BIBD is *symmetric* if the intersection of any two distinct blocks has constant size; in this case, the point and the block index sets have the same cardinality. More details on BIBDs can be found in [5], quasi-symmetric designs are treated in depth in [22].

Another type of design which will play a role in this paper is the *net* [10]. It satisfies

- (1) To every point (block index) there exist two block indices (points) not incident with it,
- (2) two points are incident with at most one common block index,
- (3) if a point x is not incident with a block index s , then there exists one and only one block index s' incident with x such that the blocks pertaining to s and s' have empty intersection.

Note that a net is resolvable. A point of a net is incident with a constant number of block indices, which means that nets satisfy (2.1) as well. By definition, a *transversal design* is the dual of a net [5], so its point set can be partitioned into point classes such that two distinct points are joined by a unique block index if and only if they are contained in different point classes.

Let $\mathcal{A}$ be a finite set. A *mosaic of incidence structures* on $\mathcal{X}$ is a family $M = (D_\alpha)_{\alpha \in \mathcal{A}}$ such that for some set $\mathcal{S}$,

- (1) each D_α is an incidence structure with point set $\mathcal{X}$ and block index set $\mathcal{S}$,
- (2) every pair $(x, s) \in \mathcal{X} \times \mathcal{S}$ is incident in a unique D_α .

The incidence structures D_α are called the *members* of M . We call M *nontrivial* if $2 \leq |\mathcal{A}| < |\mathcal{X}|$.

The *sum* ΣM of M is the incidence structure on $\mathcal{X}$ with block index set $\mathcal{S} \times \mathcal{A}$ where x is incident with (s, α) if and only if x is incident with s in D_α . The *dual* of M is the mosaic $\tilde{M} = (\tilde{D}_\alpha)_{\alpha \in \mathcal{A}}$ on $\mathcal{S}$ with block index set $\mathcal{X}$, where each $\tilde{D}_\alpha$ is the dual of D_α . By a *mosaic of BIBD* (v, k, λ) , we mean a mosaic each of whose members is a $\text{BIBD}(v, k, \lambda)$. Mosaics of BIBDs were defined in [11]. Mosaics of incidence structures each of whose members is a BIBD, though with different parameters, have appeared earlier in the literature [9]. Mosaics of group divisible designs and duals thereof were defined in [30].

The connection of mosaics of incidence structures with functions is the following. Let $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ be any function. Then each inverse image $f^{-1}(\alpha)$ determines an incidence relation D_α on $\mathcal{X}$ with block index set $\mathcal{S}$, in such a way that x is incident with s if and only

if $f(x, s) = \alpha$. Thus every $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ determines a unique mosaic $M(f)$ of incidence structures on $\mathcal{X}$. Conversely, every mosaic $M = (D_\alpha)_{\alpha \in \mathcal{A}}$ gives rise to a unique function $f_M : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$, where $\mathcal{X}$ is the point set and $\mathcal{S}$ the block index set of the mosaic.

We note the following connection of mosaics with resolvability.

Lemma 2 *For any finite set $\mathcal{X}$, there is a one-to-one correspondence between the mosaics of incidence structures M on $\mathcal{X}$ and the resolvable incidence structures D on $\mathcal{X}$ such that $\Sigma M = D$.*

Proof Let $M = (D_\alpha)_{\alpha \in \mathcal{A}}$ be a mosaic on $\mathcal{X}$ with block index set $\mathcal{S}$. For each $s \in \mathcal{S}$, the set of pairs $\{(s, \alpha) : \alpha \in \mathcal{A}\}$ forms a parallel class for ΣM . Conversely, let D be a resolvable incidence structure on $\mathcal{X}$ with block index set $\mathcal{H} \times \mathcal{A}$, where $\mathcal{H}$ is an index set for the parallel classes of D and $\mathcal{A}$ is an index set for the elements of each parallel class. For each $\alpha \in \mathcal{A}$, define the incidence structure D_α on $\mathcal{X}$ with block index set $\mathcal{H}$ by letting a point x be incident with block index h in D_α if and only if x is incident with (h, α) . Clearly, $(D_\alpha)_{\alpha \in \mathcal{A}}$ is a mosaic of incidence structures on $\mathcal{X}$. $\square$

3 Structure of ACFU hash functions

In order to simplify our notation when we analyze ε -ACFU hash functions in terms of mosaics of incidence structures, we introduce two types of blocks associated with a function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$. For $x \in \mathcal{X}$ and $\alpha \in \mathcal{A}$, we define

$$B_{x,\alpha} = \{s \in \mathcal{S} : f(x, s) = \alpha\} \quad (3.1)$$

and for $s \in \mathcal{S}$ and $\alpha \in \mathcal{A}$, we set

$$B_{s,\alpha} = \{x \in \mathcal{X} : f(x, s) = \alpha\}. \quad (3.2)$$

We will call a set of any of these types a block. Note that the blocks $B_{s,\alpha}$ ($s \in \mathcal{S}$) are precisely the blocks of the α -th member D_α of $M(f)$, whereas the blocks $B_{x,\alpha}$ ($x \in \mathcal{X}$) are the blocks of the dual of D_α . The function f is an ε -ACFU hash function if the $B_{x,\alpha}$ have constant size and if the intersection $B_{x,\alpha} \cap B_{x',\alpha}$ for distinct $x, x' \in \mathcal{X}$ has cardinality at most $\varepsilon|\mathcal{S}|/|\mathcal{A}|$.

We start our analysis by observing that ε cannot be arbitrarily small for an ε -ACFU hash function.

Lemma 3 *For any ε -ACFU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$,*

$$\varepsilon \geq \frac{|\mathcal{X}| - |\mathcal{A}|}{|\mathcal{A}|(|\mathcal{X}| - 1)}. \quad (3.3)$$

If equality holds, then $\Sigma M(f)$ (the sum of the mosaic determined by f , see Sect. 2) is a resolvable BIBD on $\mathcal{X}$.

If $|\mathcal{X}|$ and $|\mathcal{A}|$ are given, we call the quantity on the right-hand side of (3.3) the *optimal* ε . The proof of Lemma 3 follows immediately from the first part of Lemma 1 together with the following results of Sarwate [21] and Stinson [23].

Lemma 4 ([21], p. 42 and [23], Theorem 1.1) *For any ε -AU hash function, ε satisfies (3.3).*

Lemma 5 ([23], Theorem 2.1)

- (1) If $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ is an ε -AU hash function where ε equals the right-hand side of (3.3), then $\Sigma M(f)$ is a resolvable BIBD.
- (2) Conversely, let D be a resolvable BIBD on $\mathcal{X}$. Then, the function f_M determined by the mosaic M corresponding to D by Lemma 2 is an ε -AU hash function with ε satisfying equality in (3.3).

We call an ε -ACFU hash function with optimal ε an *optimally collision-flat universal (OCFU) hash function*. An ε -AU hash function with optimal ε is called *optimally universal (OU)* by Sarwate and Stinson.

3.1 General ε

We start the analysis of the structure of ACFU hash functions by giving two bounds on the size of $\mathcal{S}$. Then we give criteria for equality and discuss the relation between the bounds.

Theorem 1 For a nontrivial ε -ACFU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$, it holds that

$$|\mathcal{S}| \geq 1 + \frac{|\mathcal{X}|(|\mathcal{A}| - 1)^2}{\varepsilon|\mathcal{A}|(|\mathcal{X}| - |\mathcal{A}|) + |\mathcal{A}|^2 - |\mathcal{X}|} \quad (3.4)$$

and

$$|\mathcal{S}| \geq \frac{|\mathcal{A}|}{\varepsilon}. \quad (3.5)$$

Proof The bound (3.5) is easy to see: If $\varepsilon|\mathcal{S}|/|\mathcal{A}| < 1$, then $|\{s : f(x, s) = f(x', s) = \alpha\}| = 0$ for every α and all distinct x, x' . In particular, each of the functions $x \mapsto f(x, s)$ would be injective, meaning that $|\mathcal{A}| \geq |\mathcal{X}|$. But this is impossible since we assume that f is nontrivial.

To prove (3.4), we use the “variance method” also used by Stinson [24] to find lower bounds on $\mathcal{S}$ in the case of ε -AU and ε -ASU hash functions. Fix an arbitrary $\alpha \in \mathcal{A}$ and $s \in \mathcal{S}$ and define $\lambda_\sigma = |B_{s,\alpha} \cap B_{\sigma,\alpha}|$ (recall the notation (3.1)). Then obviously,

$$\sum_{\sigma \neq s} 1 = |\mathcal{S}| - 1.$$

Note that $|\mathcal{S}| - 1 \geq 1$ since we can always assume $\varepsilon \leq 1$, so that (3.5) implies $|\mathcal{S}| \geq |\mathcal{A}| \geq 2$. Also, counting pairs (x, σ) such that $\sigma \neq s$ and $x \in B_{s,\alpha} \cap B_{\sigma,\alpha}$, and using property (ACFU1), we find that

$$\sum_{\sigma \neq s} \lambda_\sigma = \sum_{x: f(x,s)=\alpha} |\{\sigma \neq s : f(x, \sigma) = \alpha\}| = |B_{s,\alpha}| \left(\frac{|\mathcal{S}|}{|\mathcal{A}|} - 1 \right).$$

Moreover, counting triples (σ, x, x') with $\sigma \neq s$ as well as $x \neq x'$ and $x, x' \in B_{s,\alpha} \cap B_{\sigma,\alpha}$, we find using property (ACFU2) that

$$\begin{aligned} & \sum_{\sigma \neq s} \lambda_\sigma (\lambda_\sigma - 1) \\ &= \sum_{x: f(x,s)=\alpha} \sum_{x': f(x',s)=\alpha} |\{\sigma \neq s : f(x, \sigma) = f(x', \sigma) = \alpha\}| \\ &\leq \left(\frac{\varepsilon|\mathcal{S}|}{|\mathcal{A}|} - 1 \right) |B_{s,\alpha}| (|B_{s,\alpha}| - 1). \end{aligned}$$

The mean of the λ_σ is

$$\bar{\lambda} = \frac{|B_{s,\alpha}| \left(\frac{|\mathcal{S}|}{|\mathcal{A}|} - 1 \right)}{|\mathcal{S}| - 1}.$$

Hence

$$\begin{aligned} 0 &\leq \sum_{\sigma \neq s} (\lambda_\sigma - \bar{\lambda})^2 \\ &= \sum_{\sigma \neq s} \lambda_\sigma^2 - (|\mathcal{S}| - 1) \bar{\lambda}^2 \\ &\leq \left(\frac{\varepsilon |\mathcal{S}|}{|\mathcal{A}|} - 1 \right) |B_{s,\alpha}| (|B_{s,\alpha}| - 1) + |B_{s,\alpha}| \left(\frac{|\mathcal{S}|}{|\mathcal{A}|} - 1 \right) - \frac{|B_{s,\alpha}|^2 \left(\frac{|\mathcal{S}|}{|\mathcal{A}|} - 1 \right)^2}{|\mathcal{S}| - 1}. \end{aligned}$$

Solving this for $|\mathcal{S}|$, one sees that it implies that

$$|\mathcal{S}| \geq 1 + \frac{(|\mathcal{A}| - 1)^2 |B_{s,\alpha}|}{\varepsilon |\mathcal{A}| (|B_{s,\alpha}| - 1) + |\mathcal{A}| - |B_{s,\alpha}|}. \quad (3.6)$$

In order to make sure that the denominator does not vanish, we observe that if this were the case, we would have

$$\varepsilon = \frac{|B_{s,\alpha}| - |\mathcal{A}|}{|\mathcal{A}| (|B_{s,\alpha}| - 1)},$$

which in the light of Lemma 3 would mean that $|B_{s,\alpha}| = |\mathcal{X}|$. But then ε would satisfy equality in (3.3) and f would be an OU hash function, so $|B_{s,\alpha}| = |\mathcal{X}|/|\mathcal{A}|$. However since f is nontrivial, it is impossible that $|\mathcal{A}| = 1$. Hence the right-hand side of (3.6) is well-defined.

We obtain the claimed bound by observing that for every $s \in \mathcal{S}$, there must be a value α such that $|B_{s,\alpha}| \geq |\mathcal{X}|/|\mathcal{A}|$, and inserting this in (3.6). $\square$

The proof of (3.4) provides us with enough information to precisely characterize the structure of those ACFU hash functions which attain equality.

Corollary 1 *If $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ is a nontrivial ε -ACFU hash function satisfying equality in (3.4), then the dual of its mosaic $M(f)$ is a mosaic of quasi-symmetric BIBD(v, k, λ), where*

$$v = |\mathcal{S}|, \quad k = \frac{|\mathcal{S}|}{|\mathcal{A}|}, \quad \lambda = \frac{|\mathcal{X}|(|\mathcal{S}| - |\mathcal{A}|)}{|\mathcal{A}|^2(|\mathcal{S}| - 1)},$$

and the intersection numbers are 0 and $\varepsilon|\mathcal{S}|/|\mathcal{A}|$.

Conversely, if M is the dual of a mosaic of nontrivial quasi-symmetric BIBD(v, k, λ) with intersection numbers 0 and μ , then f_M is a nontrivial ε -ACFU hash function with

$$\varepsilon = \frac{\mu}{k}.$$

Proof Let $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ be an ε -ACFU hash function. Equality in (3.4) holds if and only if in the proof of (3.4), all inequalities are equalities irrespective of the choice of s and α . Hence, equality holds if and only if the following conditions are satisfied:

- (1) $|B_{s,\alpha}|$ is constant and equal to $|\mathcal{X}|/|\mathcal{A}|$;

(2) for all distinct $s, s' \in \sigma$ and all α ,

$$|B_{s,\alpha} \cap B_{s',\alpha}| = \bar{\lambda}$$

for $\bar{\lambda}$ as in the proof of Theorem 1; another way of stating this is that any two distinct s and s' are contained in precisely $\bar{\lambda}$ different blocks of the form $B_{x,\alpha}$;

(3) for any distinct x, x' for which the corresponding blocks $B_{x,\alpha}$ and $B_{x',\alpha}$ have nonempty intersection,

$$|B_{x,\alpha} \cap B_{x',\alpha}| = \frac{\varepsilon|S|}{|\mathcal{A}|}.$$

The first two conditions imply that $\bar{\lambda}$ equals the λ from the statement of the corollary. Together with the constant size of the blocks $B_{x,\alpha}$, which is guaranteed by property (ACFU1), condition (2) says that for each α , the incidence relation D_α on the point set S with block set $\{B_{x,\alpha} : x \in \mathcal{X}\}$ is a BIBD(v, k, λ) with parameters as claimed. The third of the above conditions for equality implies that any two distinct blocks $B_{x,\alpha}$ and $B_{x',\alpha}$ intersect in either 0 or $\varepsilon|S|/|\mathcal{A}|$ points. If the intersection size were equal to $\varepsilon|S|/|\mathcal{A}|$ for all distinct blocks, then the D_α would be symmetric, in particular, $b = v$. This is impossible by Corollary 2 below. Hence, D_α must be quasi-symmetric.

Now assume we are given a mosaic $\tilde{M} = (\tilde{D}_\alpha)_{\alpha \in \mathcal{A}}$ of quasi-symmetric BIBD(v, k, λ) on the point set S and with block index set $\mathcal{X}$, with intersection numbers 0 and $\mu > 0$. It is well-known that

$$\mu = \frac{(k-1)(\lambda-1)}{r-1} + 1 \quad (3.7)$$

(e.g., [22, Proposition 3.17]). Let M be the dual of $\tilde{M}$ and f_M the function induced by M . Then, using the notation (3.2) with $f = f_M$, it holds for any $\alpha \in \mathcal{A}$ and distinct points $x, x' \in \mathcal{X}$ that

$$|\{s \in S : f_M(x, s) = f_M(x', s) = \alpha\}| = |B_{x,\alpha} \cap B_{x',\alpha}| \leq \mu.$$

Defining ε as in the statement then makes f_M an ε -ACFU hash function. Inserting this in the right-hand side of (3.4) and solving for $|S| = v$ using (3.7) shows that equality in (3.4) is satisfied. Obviously, f_M is nontrivial since $|\mathcal{A}| = v/k \geq 2$. $\square$

We have not been able to characterize those ACFU hash functions f satisfying equality in (3.5). Combinatorially, equality is equivalent to the statement that $|B_{x,\alpha} \cap B_{x',\alpha}| \leq 1$ for all $x \neq x'$. This immediately implies $|B_{s,\alpha} \cap B_{s',\alpha}| \leq 1$ for $s \neq s'$, so if the $B_{s,\alpha}$ have constant size, then the dual $\tilde{M}(f)$ of $M(f)$ also gives rise to an ACFU hash function satisfying equality in (3.5). If the members $\tilde{D}_\alpha$ of $\tilde{M}(f)$ in addition are BIBDs (i.e., $|B_{x,\alpha} \cap B_{x',\alpha}| = 1$ for all distinct x, x'), then they are quasi-symmetric, and in this case f satisfies both bounds of Theorem 1, see Example 1. So to find examples where only (3.5) is satisfied, other types of designs need to be considered. ACFU hash functions whose mosaics consist of transversal designs or nets are given in Example 4.

When does each of the two bounds given in Theorem 1 apply? For the purpose of this discussion, given $|\mathcal{X}|$ and $|\mathcal{A}|$, let us call ε *feasible* if it satisfies the inequality of Lemma 3 and if $\varepsilon \leq 1$ (which we can assume without loss of generality). Let us also say that the inequality (3.4) *applies* to ε if the right-hand side of (3.4) is larger than that of (3.5); otherwise, we say that (3.5) *applies*.

Lemma 6 Let $|\mathcal{X}| > |\mathcal{A}| > 1$. The bound (3.4) applies to those feasible ε satisfying

$$\frac{|\mathcal{X}| - |\mathcal{A}|}{|\mathcal{A}|(|\mathcal{X}| - 1)} \leq \varepsilon \leq \frac{|\mathcal{X}| - |\mathcal{A}|^2}{|\mathcal{X}| - |\mathcal{A}|}.$$

This set is nonempty if and only if

$$|\mathcal{X}| \geq \frac{|\mathcal{A}|}{2} \left(|\mathcal{A}| + \sqrt{(|\mathcal{A}| + 3)(|\mathcal{A}| - 1)} + 1 \right). \quad (3.8)$$

Proof The right-hand side of (3.5) is larger than that of (3.4) if and only if ε satisfies

$$\frac{|\mathcal{X}| - |\mathcal{A}|^2}{|\mathcal{X}| - |\mathcal{A}|} \leq \varepsilon \leq 1.$$

Comparing this with the optimal ε from Lemma 3, we obtain the criterion (3.8). $\square$

The case which interests us most is where $\varepsilon \approx 1/|\mathcal{A}|$ (see Sect. 5). Then if (3.5) applies to ε , the cardinality of $\mathcal{S}$ cannot be much smaller than $|\mathcal{A}|^2$. But the right-hand side of (3.8) is approximately equal to $|\mathcal{A}|^2$ for large $|\mathcal{A}|$, so in this case, $|\mathcal{S}|$ will usually be larger than $|\mathcal{X}|$. On the other hand, the lower bound (3.4) is at most $|\mathcal{X}|$, and Example 2 below shows that $|\mathcal{S}| < |\mathcal{X}|$ is indeed possible since it satisfies equality in (3.4) with $\varepsilon = 1/|\mathcal{A}|$. However, for optimal ε , the situation is special and the seed set is strictly larger than the point set, as we will see in the next subsection.

3.2 Optimal ε

We now consider the case where ε is the optimal one from Lemma 3, i.e., we deal with the case of OCFU hash functions. Stinson's result, Lemma 5, shows that an OCFU hash function f has a corresponding $\Sigma M(f)$ which is a BIBD. The next result characterizes OCFU hash functions by the members of $M(f)$.

Theorem 2 If $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ is an OCFU hash function, then $M(f)$ is a mosaic of BIBDs (v, k, λ) with

$$v = |\mathcal{X}|, \quad k = \frac{|\mathcal{X}|}{|\mathcal{A}|}, \quad \lambda = \frac{\varepsilon|\mathcal{S}|}{|\mathcal{A}|}, \quad b = |\mathcal{S}|, \quad r = \frac{|\mathcal{S}|}{|\mathcal{A}|}.$$

Conversely, if $M = (D_\alpha)_{\alpha \in \mathcal{A}}$ is a mosaic of BIBD (v, k, λ) on $\mathcal{X}$ with block index set $\mathcal{S}$, then f_M is an OCFU hash function, and $|\mathcal{X}|, |\mathcal{S}|, |\mathcal{A}|$ satisfy the above relations.

Proof Let $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ be an OCFU hash function. Since $\Sigma M(f)$ is a BIBD, all blocks $B_{s,\alpha}$ have fixed size k . Property (ACFU1) also ensures that the dual blocks $B_{x,\alpha}$ have constant size $|\mathcal{S}|/|\mathcal{A}|$, which we denote by r .

We need to check that two points meet in precisely λ blocks for suitable λ . To do this, fix $\alpha \in \mathcal{A}$ and $x \in \mathcal{X}$ and define the weights

$$\kappa_i = |\{x' \in \mathcal{X} : x' \neq x, |B_{x,\alpha} \cap B_{x',\alpha}| = i\}|.$$

Set

$$L = \lfloor \varepsilon |\mathcal{S}| / |\mathcal{A}| \rfloor = \left\lfloor \frac{r(k-1)}{|\mathcal{X}| - 1} \right\rfloor,$$

the largest possible value of i . Note that

$$\sum_{i=0}^L \kappa_i = |\mathcal{X}| - 1.$$

By counting pairs (x', s) with $x' \neq x$ satisfying $s \in B_{x,\alpha} \cap B_{x',\alpha}$, we obtain

$$\frac{1}{|\mathcal{X}| - 1} \sum_{i=0}^L i \kappa_i = \frac{r(k-1)}{|\mathcal{X}| - 1}.$$

This implies that $\kappa_L = |\mathcal{X}| - 1$ and $\kappa_i = 0$ for $i < L$. It follows that $|B_{x',\alpha} \cap B_{x,\alpha}| = L$ for all α and $x \neq x'$. Hence D_α is a BIBD(v, k, λ). The fact that

$$\lambda = L = \frac{r(k-1)}{|\mathcal{X}| - 1} = \frac{\varepsilon|\mathcal{S}|}{|\mathcal{A}|}$$

follows from (2.2).

In the other direction, given a mosaic M of BIBDs, it is straightforward to check that f_M is an OCFU hash function and that the parameters are related as claimed in the statement. $\square$

Corollary 2 For an OCFU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$,

$$|\mathcal{S}| \geq \frac{|\mathcal{A}|(|\mathcal{X}| - 1)}{|\mathcal{A}| - 1}. \quad (3.9)$$

Equivalently, in a mosaic of BIBD(v, k, λ), it holds that $b \geq v + r - 1$.

Proof In a mosaic M of BIBD(v, k, λ), the block size k of each member D_α of M divides the size v of the point set. Since each D_α is a BIBD, the result proved independently by Roy [20] and Mikhail [18] applies, stating that $b \geq v + r - 1$ in this case. With $a = v/k$, this is equivalent to $(a - 1)b/a = (a - 1)r \geq v - 1$ (recall (2.1)), which can be transformed into inequality (3.9) for f_M . $\square$

The bound given in the corollary is tight, as will be seen in Example 1 below. However, it can only be attained in the regime where $|\mathcal{X}| \geq |\mathcal{A}|^2$, since the right-hand side of (3.5) is strictly larger than the right-hand side of (3.9) if $|\mathcal{X}| < |\mathcal{A}|^2$. For the latter case, an OCFU function is given in Example 3.

We also note that the corollary implies that the block index set of a nontrivial mosaic of BIBDs is strictly larger than its point set. Therefore the member BIBDs cannot be symmetric. This proves the claim that the designs achieving equality in (3.4) are truly quasi-symmetric, completing the proof of Corollary 1.

3.3 Examples

Example 1 It was shown in [11] that a mosaic of BIBDs (D_α) can be constructed from any resolvable BIBD D in such a way that every D_α is isomorphic to D . Starting with an affine design, this gives a mosaic of designs satisfying equality in Corollary 2. We recall the explicit form of a corresponding OCFU function, originally given in [30]. Let q be a prime power and $\mathbb{F}_q$ the field with q elements. For any positive integer t , let D be the affine designs on $\mathbb{F}_q^t$ with blocks given by the hyperplanes (cosets of $(t - 1)$ -dimensional subspaces) of $\mathbb{F}_q^t$. Every $(t - 1)$ -dimensional subspace of $\mathbb{F}_q^t$ can be identified with the solution space of the equation

$\sum_i h_i x_i = 0$, where h is a unique nonzero vector in $\mathbb{F}_q^t$ whose first nonzero component is 1. Denote the set of such vectors by $\mathcal{H}$ and define $f : \mathbb{F}_q^t \times (\mathcal{H} \times \mathbb{F}_q) \rightarrow \mathbb{F}_q$ by

$$f(x; h, \beta) = \sum_i h_i x_i + \beta.$$

For fixed $\alpha \in \mathbb{F}_q$, as the pair (h, β) ranges over all possible values, the preimages $f(\cdot; h, \beta)^{-1}(\alpha) = B_{h, \beta; \alpha}$ range over all hyperplanes of $\mathbb{F}_q^t$. Hence the incidence structure D_α on $\mathcal{X}$ formed by the block set $\{B_{h, \beta; \alpha} : h \in \mathcal{H}, \beta \in \mathbb{F}_q\}$ and the $\in$ relation is isomorphic to D , and the family $M = (D_\alpha)_{\alpha \in \mathbb{F}_q}$ is a mosaic of BIBDs. Thus f is an OCFU hash function.

The example shows that the bound from Corollary 2 is tight (not surprisingly, given (2.3)). In case $t = 2$, which corresponds to a mosaic of affine planes, equality holds in (3.5) as well.

Example 2 Consider the dual $\tilde{M}$ of the mosaic M of affine BIBDs considered in the previous example. Since two distinct non-parallel hyperplanes meet in q^{t-2} points, $f = f_{\tilde{M}}$ is a $1/|\mathcal{A}|$ -ACFU hash function. Since an affine design is quasi-symmetric, f attains the bound (3.4). If f derives from the dual of a mosaic of affine planes (i.e., $t = 2$), then it also satisfies equality in (3.5). Written as a function with the same notation as in the previous example, f satisfies the formula

$$f : (\mathcal{H} \times \mathbb{F}_q) \times \mathbb{F}_q^t \rightarrow \mathbb{F}_q, \quad f(h, \beta; x) = \sum_{i=1}^t h_i x_i + \beta.$$

Example 3 Any OCFU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ whose underlying mosaic $M(f)$ is a mosaic of $\text{BIBD}(v, k, 1)$ (with $v = |\mathcal{X}|$, $k = |\mathcal{X}|/|\mathcal{A}|$) satisfies equality in (3.5). In this case, $|\mathcal{A}|^2$ has to be at least as large as $|\mathcal{X}|$ by the discussion after Corollary 2. An example was studied in [30, p. 608], based on the resolvable designs arising from Denniston's construction of maximal arcs in projective space over $\mathbb{F}_2$. Unfortunately, the corresponding OCFU hash function does not have a nice closed form, but it was shown in [30] to be polynomial-time computable. The size of $|\mathcal{A}|$ ranges between $\sqrt{|\mathcal{X}|}$ and $|\mathcal{X}|$.

Example 4 We would also like to have $1/|\mathcal{A}|$ -ACFU hash functions with a small seed set in the range where (3.5) is the relevant lower bound for the seed size, i.e., where $|\mathcal{A}|(|\mathcal{A}| + 1) \geq |\mathcal{X}|$. Again let q be a prime power, let $\mathcal{H}$ be a subset of $\mathbb{F}_q$, set $\mathcal{X} = \mathcal{H} \times \mathbb{F}_q$ as well as $\mathcal{S} = \mathbb{F}_q^2$ and $\mathcal{A} = \mathbb{F}_q$. Then define [30, p. 610]

$$f(h, y; s_1, s_2) = s_2 - h s_1 + y.$$

Since for distinct (h, y) and (h', y') and any $\alpha \in \mathbb{F}_q$ there exists at most one (s_1, s_2) such that $f(h, y; s_1, s_2) = f(h', y'; s_1, s_2) = \alpha$, this function satisfies equality in (3.5). (If we set $\mathcal{H} = \mathbb{F}_q$, additionally extend it by the symbol ∞ and define $f(\infty, y, s_1, s_2) = s_1 + y$, then we obtain the case $t = 2$ from Example 2.)

The combinatorial structure of f is as follows. We can partition $\mathcal{X}$ into the $|\mathcal{H}|$ point classes $\{(h, y) : y \in \mathbb{F}_q\}$. If two distinct points (h, y) and (h', y') are from the same point class, then no block of $M(f) = (D_\alpha)_{\alpha \in \mathcal{A}}$ contains both of them. If they come from different point classes, then for each $\alpha \in \mathcal{A}$, there exists a unique block from D_α containing them both. That means that each D_α is isomorphic to the same transversal design, and the members of $M(f)$ even share the same point class partition.

By passing to the dual $\tilde{f} : \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{A}$, we obtain a seed-optimal $1/|\mathcal{H}|$ -ACFU hash function. $M(\tilde{f}) = \tilde{M}(f)$ is a mosaic of nets.

Table 1 The lower bounds for optimal ε and $\varepsilon = 1/|\mathcal{A}|$

ε	AU	ACFU	ASU
Optimal	$\frac{ \mathcal{X} - 1}{ \mathcal{A} - 1}$	$\frac{ \mathcal{A} (\mathcal{X} - 1)}{ \mathcal{A} - 1}$	–
$\frac{1}{ \mathcal{A} }$	$\frac{ \mathcal{X} }{ \mathcal{A} }$	$\max \left\{ \mathcal{A} ^2, 1 + \frac{ \mathcal{X} (\mathcal{A} - 1)}{ \mathcal{A} } \right\}$	$1 + \mathcal{X} (\mathcal{A} - 1)$

For ASU hash functions, $\varepsilon \geq 1/|\mathcal{A}|$ [24]

3.4 Comparison with AU and ASU bounds

For completeness and comparison, we briefly discuss lower bounds on the size of the seed set of an AU or ASU function.

Lemma 7 ([24], Theorems 4.1 and 4.3) *Let $\varepsilon > 0$.*

(1) *For any ε -AU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$,*

$$|\mathcal{S}| \geq \frac{|\mathcal{X}|(|\mathcal{A}| - 1)}{\varepsilon|\mathcal{A}|(|\mathcal{X}| - |\mathcal{A}|) + |\mathcal{A}|^2 - |\mathcal{X}|}. \quad (3.10)$$

(2) *For any ε -ASU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$,*

$$|\mathcal{S}| \geq 1 + \frac{|\mathcal{X}|(|\mathcal{A}| - 1)^2}{\varepsilon|\mathcal{A}|(|\mathcal{X}| - 1) + |\mathcal{A}| - |\mathcal{X}|}. \quad (3.11)$$

We also have an additional simple lower bound for the seed size of ASU hash functions analogous to (3.5), which to our knowledge has not yet been stated explicitly anywhere.

Lemma 8 *Let $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ be an ε -ASU hash function. Then*

$$|\mathcal{S}| \geq \frac{|\mathcal{A}|}{\varepsilon}.$$

Proof Assume $f(x, s) = \alpha$. For any distinct $x' \in \mathcal{X}$, there must exist $\alpha' \in \mathcal{A}$ such that $f(x', s) = \alpha'$. Therefore $\varepsilon|\mathcal{S}|/|\mathcal{A}| \geq 1$. $\square$

It is simple to check that the relation between the ASU bounds is as follows.

Lemma 9 *For an ε -ASU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ with $|\mathcal{X}| > |\mathcal{A}|$, the right-hand side of the bound from Lemma 8 is larger than the one from (3.11) if and only if*

$$\frac{|\mathcal{X}| - |\mathcal{A}|}{|\mathcal{X}| - 1} \leq \varepsilon \leq 1.$$

Note that, if $|\mathcal{X}| > |\mathcal{A}|$, the left-hand side of the inequality of Lemma 9 is always at least $1/|\mathcal{A}|$. Hence given $|\mathcal{X}|$ and $|\mathcal{A}|$, there is always a range of ε sufficiently close to $1/|\mathcal{A}|$ where the bound from Lemma 7 applies.

Remark 2 The analogous lower bound $|\mathcal{S}| \geq 1/\varepsilon$ for ε -AU hash functions gives no new information, since $1/\varepsilon$ is always smaller than the right-hand side of (3.10), with equality if and only if $|\mathcal{X}| = |\mathcal{A}|^2$.

We are again interested in conditions for equality in the above bounds. For the OU case, Stinson gives the following criterion.

Lemma 10 ([23], Theorem 2.2) *If f is an OU hash function satisfying equality in (3.10), then $\Sigma M(f)$ is an affine BIBD. Conversely, if D is any affine BIBD, then the function f_M induced by the mosaic M corresponding to D by Lemma 2 is an OU hash function which satisfies equality in (3.10).*

No result is known to us which characterizes ε -AU hash functions satisfying equality in (3.10) for general ε .

For the ASU case of Lemma 7, van Trung determines the condition for equality. Our proof of Corollary 1 is similar to the proof of van Trung's result.

Lemma 11 ([27], Theorem 3.1) *An ε -ASU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ satisfies equality in (3.11) if $\Sigma \tilde{M}(f)$ is a resolvable quasi-symmetric design with one intersection number equal to zero, where $\tilde{M}(f)$ is the dual of $M(f)$. In the other direction, if a resolvable quasi-symmetric design D is given with one intersection number equal to 0, then there exists an ε such that $f_{\tilde{M}}$ is an ε -ASU hash function satisfying equality in (3.11), where M is the mosaic determined by D via Lemma 2 and $\tilde{M}$ is its dual.*

The similarity of this result to ours on ACFU hash functions, Corollary 1, is no coincidence, as we will see in the next section (Theorem 4).

For Lemma 8, we cannot characterize those ASU hash functions satisfying equality, but again there is a striking similarity with the ACFU situation (see the discussion after Corollary 1) which will be explained in the next section (Theorem 4 again). However, a class of designs from which one can construct such ASU hash functions are the nets. A net $\tilde{D}$ on the point set $\mathcal{S}$ obviously is a resolvable incidence structure, so by Lemma 2 it gives rise to a mosaic $\tilde{M} = (\tilde{D}_\alpha)_{\alpha \in \mathcal{A}}$. The blocks all have the same size and different blocks intersect in at most one point, so equality is satisfied in Lemma 8. Let $\mathcal{X}$ be the block index set. Then the dual M of $\tilde{M}$ induces a function f_M which is an ε -ASU hash function for $\varepsilon = |\mathcal{A}|/|\mathcal{S}|$. If the net is an affine plane, then f_M also satisfies equality in van Trung's bound.

Table 1 shows the bounds for the cases of optimal ε as well as for $\varepsilon = 1/|\mathcal{A}|$. (Due to the application of ACFU hash functions we have in mind (see Sect. 5), $\varepsilon \approx 1/|\mathcal{A}|$ is what interests us most.) Interestingly, we need to consider the case of OCFU hash functions separately, whereas Stinson's lower bound for ε -AU hash functions also covers OU hash functions.

4 Constructions of ε -ACFU hash functions

4.1 Extensions of AU and ASU hash functions

While ε -ACFU hash functions are new, ε -AU and ε -ASU hash functions are well-investigated concepts with many efficiently computable examples, so it would be attractive to be able to turn an ε -AU or ε -ASU hash function into an ε -ACFU hash function. It turns out that this is indeed possible.

Let $g : \mathcal{X} \times \mathcal{H} \rightarrow \mathcal{A}$ be an ε -AU hash function. Let L be a latin square with entries from $\mathcal{A}$ and rows and columns indexed by $\mathcal{A}$, too. This can equivalently be described as a quasigroup structure on $\mathcal{A}$ whose product $\circ$ is defined by the rule $\alpha \circ \beta = L(\alpha, \beta)$. We denote the unique solution γ of the equation $\gamma \circ \beta = \alpha$ by α/β . We can now define the function $\hat{g} : \mathcal{X} \times (\mathcal{H} \times \mathcal{A}) \rightarrow \mathcal{A}$ by

$$\hat{g}(x; h, \beta) = g(x, h) \circ \beta,$$

so its seed set is $\mathcal{S} = \mathcal{H} \times \mathcal{A}$. We call $\hat{g}$ the *seed extension* of g .

Theorem 3 *The function $g : \mathcal{X} \times \mathcal{H} \rightarrow \mathcal{A}$ is an ε -AU hash function if and only if its seed extension $\hat{g}$ is an ε -ACFU hash function. Each member of $M(\hat{g})$ is isomorphic to $\Sigma M(g)$.*

Proof The equation $\hat{g}(x; h, \beta) = \alpha$ means that $g(x, h) = \alpha/\beta$. Moreover, as β varies over $\mathcal{A}$, the unique solution α/β of $\gamma \circ \beta = \alpha$ assumes all possible values in $\mathcal{A}$. Thus for any $x, x' \in \mathcal{X}$,

$$|\{(h, \beta) : \hat{g}(x; h, \beta) = \hat{g}(x'; h, \beta) = \alpha\}| = |\{h : g(x, h) = g(x', h)\}|. \quad (4.1)$$

If $x = x'$, then the set in (4.1) is all of $\mathcal{H}$, whose size is $|\mathcal{S}|/|\mathcal{A}|$. Thus $\hat{g}$ always satisfies property (ACFU1). Now assume that $x \neq x'$. One sees immediately from (4.1) that g is an ε -AU hash function if and only if $\hat{g}$ is an ε -ACFU hash function.

Finally, consider the α -th member D_α of $M(\hat{g})$. Its blocks have the form

$$B_{h,\beta;\alpha} = \{x : \hat{g}(x; h, \beta) = \alpha\} = \{x : g(x, h) = \alpha/\beta\},$$

for $h \in \mathcal{H}$ and $\beta \in \mathcal{A}$. As h, β vary over all possible values, the blocks $B_{h,\beta;\alpha}$ vary over all blocks of $\Sigma M(g)$. This shows that D_α is isomorphic to $\Sigma M(g)$. $\square$

As a corollary, we obtain the result of Gnille, Greferath and Pavčević on the relation between resolvable BIBDs and mosaics of BIBDs.

Corollary 3 ([11], Theorem 3.4) *For any resolvable BIBD (v, k, λ) D , there exists a mosaic of BIBD (v, k, λ) each of whose members is isomorphic to D .*

The ACFU hash functions from Examples 1, 3 and 4 can be constructed from AU hash functions in this way. Note that if the OU hash function g satisfies equality in (3.10), then $\hat{g}$ necessarily satisfies equality in (3.9). If ε is strictly larger than the optimal one, an ε -AU hash function can almost never produce a seed-optimal $\hat{g}$. Equality in (3.4) would only be possible if the $|\mathcal{A}|$ -fold multiple of the right-hand side of (3.10) were smaller than the right-hand side of (3.4), but this can only hold for the trivial case of $\varepsilon = 1$. Equality in (3.5), i.e., the relation $|\mathcal{S}| = |\mathcal{A}|/\varepsilon$, would require $|\mathcal{H}| = 1/\varepsilon$. By Remark 2, this requires that (3.10) is satisfied as well and that $|\mathcal{X}| = |\mathcal{A}|^2$. By Lemma 10, the corresponding $\Sigma M(f)$ must be an affine plane.

Remark 3 Assume $\mathcal{X}$ and $\mathcal{A}$ are groups and that the function $g : \mathcal{X} \times \mathcal{H} \rightarrow \mathcal{A}$ is a group homomorphism in the first argument for every fixed $h \in \mathcal{H}$. In addition, assume that

$$|\{h : g(x, h) = \alpha\}| \leq \varepsilon |\mathcal{H}| \quad (4.2)$$

for every $\alpha \in \mathcal{A}$ and every $x \in \mathcal{X} \setminus \{e_{\mathcal{X}}\}$, where $e_{\mathcal{X}}$ is the neutral element of $\mathcal{X}$. Then g is an ε -AU hash function, since if α is the neutral element of $\mathcal{A}$, the left-hand side of (4.2) is the same as $|\{h : g(x_1, h) = g(x_2, h)\}|$ for any x_1, x_2 such that $x_1 x_2^{-1} = x$. Moreover, $\hat{g}$ is an ε -ASU hash function, where $\hat{g}$ is constructed with respect to the given group structure on $\mathcal{A}$, which can be seen by proceeding analogously to the proof of Theorem 3 and using (4.2).

This statement was proved by Krawczyk in [15]; he calls g ε -balanced if it satisfies (4.2). This construction of ASU hash functions has been applied frequently in the literature. For instance, Stinson constructs a $1/|\mathcal{A}|$ -ASU hash function in this way in the proof of [24, Theorem 5.2]. Two specific examples are given in Examples 5 and 6.

Example 5 In an $m \times n$ Toeplitz matrix $T = (t_{ij})$ over the field $\mathbb{F}_q$ of size q , the entry t_{ij} only depends on $i - j$. Thus $T = T_h$ is determined by the vector h of the $m + n - 1$ entries in the first row and first column. Define

$$g : \mathbb{F}_q^n \times \mathbb{F}_q^{m+n-1} \rightarrow \mathbb{F}_q^m, \quad g(x, h) = T_h x.$$

It was shown in [16, Claim 2.2] that g satisfies (4.2) with $\varepsilon = 1/|\mathcal{A}|$ and that $\hat{g}$ is an ε -ASU hash function. In terms of seed length, both g and $\hat{g}$ are suboptimal.

Example 6 For any prime power q , define the function

$$g : \mathbb{F}_{q^n} \times \mathbb{F}_{q^n} \rightarrow \mathbb{F}_q^m, \quad g(x, h) = (hx)_m,$$

where $(hx)_m$ is the vector consisting of the first m components of a representation of hx as an n -dimensional vector over $\mathbb{F}_q$. This g satisfies (4.2) for $\varepsilon = q^{-m} = 1/|\mathcal{A}|$ and is linear in x with x regarded as an element of $\mathbb{F}_q^n$ for every h , hence $\hat{g}$ is a $1/|\mathcal{A}|$ -ASU hash function. For q prime, $\hat{g}$ was already defined in [7] and recognized as a $1/|\mathcal{A}|$ -ASU hash function in [28]. Stinson also uses it in [24, Theorem 5.2] to construct an ASU hash function.

By excluding $h = 0$, one obtains a function g_* for which all preimages $\{x : g_*(h, x) = \alpha\}$ have the same size. In fact, this modification makes g_* an OU hash function [1, Appendix B]. Hence, the corresponding $\hat{g}_*$ is another example of an OCFU hash function, although it satisfies neither (3.9) nor (3.5). $\hat{g}_*$ also is an ε -ASU hash function, but with $\varepsilon = q^{n-m}/(q^n - 1) > q^{-m} = 1/|\mathcal{A}|$.

The function g_* itself was used in [1, 2] to show semantic security for symmetric wiretap channels, an application related to the one we present in Sect. 5. A variant of $\hat{g}_*$, but with a larger seed, was used in [14, Remark 16, Lemma 21] in the context of information-theoretic security. From the practical viewpoint, g and g_* as well as their extended versions $\hat{g}$, $\hat{g}_*$ have the advantage that $|\mathcal{A}| = q^m$ can be any power of q between 1 and $|\mathcal{X}| = q^n$.

It is not only possible to turn an AU hash function into an ACFU hash function by extending its seed set. Let $g : \mathcal{Y} \times \mathcal{S} \rightarrow \mathcal{A}$ be an ε -ASU hash function and equip $\mathcal{A}$ with a quasigroup product denoted by $\circ$. Setting $\mathcal{X} = \mathcal{Y} \times \mathcal{A}$, we define the *point extension* of g by $\check{g} : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$,

$$\check{g}(y, \beta; s) = g(y, s) \circ \beta.$$

To connect this with the seed extension of functions, for any function $g : \mathcal{Y} \times \mathcal{S} \rightarrow \mathcal{A}$, let $\tilde{g} : \mathcal{S} \times \mathcal{Y} \rightarrow \mathcal{A}$ be defined by $\tilde{g}(s, y) = g(y, s)$ (the tilde makes sense since $M(\tilde{g}) = \tilde{M}(g)$, the dual of the mosaic of g .) Then we have the relation

$$\check{\tilde{g}} = \hat{g}.$$

The point extension of functions g where $\Sigma \tilde{M}(g)$ is a resolvable BIBD or GDD was already considered in [30].

Theorem 4 A function $g : \mathcal{Y} \times \mathcal{S} \rightarrow \mathcal{A}$ is an ε -ASU hash function if and only if its point extension $\check{g}$ is an ε -ACFU hash function. Each member of $M(\check{g})$ is isomorphic to the dual of $\Sigma \tilde{M}(g)$, where $\tilde{M}(g)$ is the dual of $M(g)$.

Proof Assume that g is ε -ASU. By the definition of $\check{g}$,

$$|\{s : \check{g}(y, \beta; s) = \check{g}(y', \beta'; s) = \alpha\}| = |\{s : g(y, s) = \alpha/\beta, g(y', s) = \alpha/\beta'\}|.$$

This expression equals $|\mathcal{S}|/|\mathcal{A}|$ if $(y, \beta) = (y', \beta')$ by (ASU1). If $y = y'$, but $\beta \neq \beta'$, then it equals zero; if $y \neq y'$, it is upper-bounded by $\varepsilon|\mathcal{S}|/|\mathcal{A}|$ by (ASU2). Hence, $\check{g}$ is an ε -ACFU function.

Now assume that $\check{g}$ is ε -ACFU. For any $y, y' \in \mathcal{Y}$ and $\alpha, \alpha' \in \mathcal{A}$, consider the set

$$\{s : g(y, s) = \alpha, g(y', s) = \alpha'\}.$$

For any pair $\beta, \beta' \in \mathcal{A}$, this is the same as

$$\{s : g(y, s) \circ \beta = \alpha \circ \beta, g(y', s) \circ \beta' = \alpha' \circ \beta'\}. \quad (4.3)$$

If $y = y'$ as well as $\alpha = \alpha'$ and $\beta = \beta'$, this set has the same cardinality as the set

$$\{s : \check{g}(y, \beta; s) = \alpha \circ \beta\}.$$

Thus the property (ACFU1) of $\check{g}$ gives us (ASU1) for g .

On the other hand, if $y \neq y'$, we can choose β and β' in such a way that $\alpha \circ \beta = \alpha' \circ \beta'$, so that the set (4.3) has the same cardinality as

$$\{s : \check{g}(y, \beta; s) = \check{g}(y', \beta'; s) = \alpha \circ \beta\},$$

which by (ACFU2) is upper-bounded by $\varepsilon|\mathcal{S}|/|\mathcal{A}|$. This implies (ASU2).

Finally, we prove the statement about the members of $M(\check{g})$. To see this, note that $\tilde{M}(\check{g}) = M(\check{\tilde{g}}) = M(\hat{\tilde{g}})$. By Theorem 3, the members of this mosaic are isomorphic to $\Sigma M(\tilde{g}) = \Sigma M(g)$. $\square$

If we are given a resolvable quasi-symmetric BIBD, then we know from Corollary 3 that one can construct a mosaic of quasi-symmetric BIBDs from this. Taking duals, one gets from a seed-optimal ε -ASU hash function (by Lemma 11) to a seed-optimal ε -ACFU hash function (by Corollary 1). Theorem 4 does this in a single step, and (necessarily) the lower bound (3.11) transforms into (3.4) in the right way. It is also obvious that an ε -ASU g satisfies equality in Lemma 8 if and only if $\check{g}$ satisfies equality in (3.5).

Examples 2 and 4 show ACFU hash functions which can be constructed as point extensions of ASU hash functions. The function from Example 4 can be represented both as the seed extension of an AU hash function and as the point extension of an ASU hash function. In fact, it is the typical example of such a function in the case where $\mathcal{A}$ is an abelian group. To see this, we extend Krawczyk's notion of ε -balancedness (see Remark 3) to arbitrary functions whose image lies in an abelian group. We say that a function $a : \mathcal{Y} \times \mathcal{H} \rightarrow \mathcal{A}$, where $\mathcal{A}$ is an abelian group, is ε -balanced if for any two distinct $y, y' \in \mathcal{Y}$ and any $\beta \in \mathcal{A}$, it satisfies

$$|\{h \in \mathcal{H} : a(y, h) - a(y', h) = \beta\}| \leq \varepsilon|\mathcal{H}|.$$

We also remark that if $f = \hat{g}_1 = \check{g}_2$ and if it maps into $\mathcal{A}$, then it must have the form $f : (\mathcal{Y} \times \mathcal{A}) \times (\mathcal{H} \times \mathcal{A}) \rightarrow \mathcal{A}$.

Proposition 1 *Let f be an ε -ACFU hash function. Then f can be represented as the seed extension of an ε -AU hash function $g_1 : (\mathcal{Y} \times \mathcal{A}) \times \mathcal{H} \rightarrow \mathcal{A}$ and as the point extension of an ε -ASU hash function $g_2 : \mathcal{Y} \times (\mathcal{H} \times \mathcal{A}) \rightarrow \mathcal{A}$ if and only if there exists an ε -balanced function $a : \mathcal{Y} \times \mathcal{H} \rightarrow \mathcal{A}$ such that*

$$f(y, \beta; h, \gamma) = a(y, h) + \beta + \gamma.$$

Proof If $f = \hat{g}_1 = \check{g}_2$, then for all y, h, β, γ ,

$$g_1(y, \beta; h) = g_2(y; h, \gamma) + \beta - \gamma.$$

Since the left-hand side does not depend on γ , there is an $a(y, h) \in \mathcal{A}$ such that $g_1(y, \beta; h) = a(y, h) + \beta$, which implies that f has the claimed form.

We check that the function a is ε -balanced. Let $y \neq y'$. Since g_1 is an ε -AU hash function, for any β, β' ,

$$|\{h : a(y, h) + \beta = a(y', h) + \beta'\}| \leq \varepsilon|\mathcal{H}|.$$

This inequality is precisely the definition of ε -balancedness since $\beta' - \beta$ can assume any value in $\mathcal{A}$.

It is straightforward to check the converse. $\square$

Clearly, in Example 4, the function a is multiplication of elements of $\mathcal{R}$ and of $\mathbb{F}_q$.

4.2 Concatenation

We can show a result similar to the results on the concatenation of almost (strongly) universal hash functions in [24].

Proposition 2 *If $f_1 : \mathcal{X}_1 \times \mathcal{S}_1 \rightarrow \mathcal{A}_1$ is an ε_1 -ASU hash function and $f_2 : \mathcal{A}_1 \times \mathcal{S}_2 \rightarrow \mathcal{A}_2$ is an ε_2 -ACFU hash function, then the function $f : \mathcal{X}_1 \times (\mathcal{S}_1 \times \mathcal{S}_2) \rightarrow \mathcal{A}_2$ defined by*

$$f(x_1; s_1, s_2) = f_2(f_1(x_1, s_1), s_2)$$

is an $(\varepsilon_1\varepsilon_2(|\mathcal{A}_1| - 1) + \varepsilon_1)$ -ACFU hash function.

Proof Let $x_1 \in \mathcal{X}_1$ and $\alpha_2 \in \mathcal{A}_2$. By (ACFU1), for each $\alpha_1 \in \mathcal{A}_1$, the number of $s_2 \in \mathcal{S}_2$ for which $f_2(\alpha_1, s_2) = \alpha_2$ equals $|\mathcal{S}_2|/|\mathcal{A}_2|$. The number of $s_1 \in \mathcal{S}_1$ for which $f_1(x_1, s_1) = \alpha_1$ equals $|\mathcal{S}_1|/|\mathcal{A}_1|$ by (ASU1). Hence

$$|\{(s_1, s_2) : f(x_1; s_1, s_2) = \alpha_2\}| = \frac{|\mathcal{S}_1|}{|\mathcal{A}_1|} \cdot \frac{|\mathcal{S}_2|}{|\mathcal{A}_2|} \cdot |\mathcal{A}_1| = \frac{|\mathcal{S}_1||\mathcal{S}_2|}{|\mathcal{A}_2|},$$

proving that f satisfies (ACFU1).

To check (ACFU2), let $\alpha_2 \in \mathcal{A}_2$ and choose distinct $x_1, x'_1 \in \mathcal{X}_1$. We need to consider two cases. Suppose first that $f_1(x_1, s_1) \neq f_1(x'_1, s_1)$. Then, $f_2(\cdot, s_2)$ needs to map these two distinct values to α_2 , and this is possible for at most $\varepsilon_2|\mathcal{S}_2|/|\mathcal{A}_2|$ values of s_2 by (ACFU2). If $f_1(x_1, s_1) = f_1(x'_1, s_1)$, then $|\mathcal{S}_2|/|\mathcal{A}_2|$ values of s_2 are possible. Since f_1 is an ASU hash function, for any pair α_1, α'_1 , there are at most $\varepsilon_1|\mathcal{S}_1|/|\mathcal{A}_1|$ possible s_1 such that $f_1(x_1, s_1) = \alpha_1$ and $f_1(x'_1, s_1) = \alpha'_1$. Therefore

$$\begin{aligned} & |\{(s_1, s_2) : f(x_1; s_1, s_2) = f(x'_1; s_1, s_2) = \alpha_2\}| \\ & \leq \frac{\varepsilon_1|\mathcal{S}_1|}{|\mathcal{A}_1|} \cdot \frac{\varepsilon_2|\mathcal{S}_2|}{|\mathcal{A}_2|} \cdot |\mathcal{A}_1|(|\mathcal{A}_1| - 1) + \frac{\varepsilon_1|\mathcal{S}_1|}{|\mathcal{A}_1|} \cdot \frac{|\mathcal{S}_2|}{|\mathcal{A}_2|} \cdot |\mathcal{A}_1| \\ & = (\varepsilon_1\varepsilon_2(|\mathcal{A}_1| - 1) + \varepsilon_1) \frac{|\mathcal{S}_1||\mathcal{S}_2|}{|\mathcal{A}_2|}. \end{aligned}$$

This completes the proof. $\square$

If we take f_1 to be an ε -ASU hash function with minimal ε , i.e., $\varepsilon = 1/|\mathcal{A}_1|$, and f_2 to be an OCFU hash function, then we obtain an $1/|\mathcal{A}|$ -ACFU hash function. However, the resulting seed set will always be larger than any of the bounds from Theorem 1. We have not checked the situation for other ε , but we think it unlikely that this method of concatenation yields ACFU hash functions with a minimal seed.

The lemma gives the possibility of constructing new ACFU hash functions, e.g., with larger ε than in most of the examples we have seen so far. (Although we do not have as yet any application for such functions.) For instance, Stinson [24] gives ASU hash functions with larger ε which can be used as the first function in concatenation.

5 Motivation from privacy amplification

In this section, we sketch how ACFU hash functions can be used in privacy amplification with the goal of establishing a uniformly distributed key between two parties whose values are indistinguishable to an adversary.

The concept of privacy amplification goes back to Bennett, Brassard and Robert [4] and Bennett, Brassard, Crépeau and Maurer [3]. In addition to the classical setup described in the introduction and below, it also plays an analogous role in quantum key distribution [26]. Quite a lot of research on privacy amplification has been made in classical information theory, and various suggestions how to measure the security of the secret key have been considered [8, 13, 17, 19]. We are going to use a very strict security measure, which appears first in a special setting related to the quantum BB84 protocol [12] and was not considered afterwards until recently [30], and for which ε -ACFU hash functions are a useful tool. Below, we will just present a simplified version of privacy amplification; for more details, see, e.g., [30]. An improved bound on the security of the method is given in [31].

Before we formalize the problem of privacy amplification, we recall the connection between a function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ and its mosaic $M(f) = (D_\alpha)_{\alpha \in \mathcal{A}}$ on $\mathcal{X}$ with block index set $\mathcal{S}$. For each α , we take N_α to be the incidence matrix of D_α ; in other words, N_α is a 01-matrix with rows indexed by $\mathcal{X}$ and columns by $\mathcal{S}$, and where

$$N_\alpha(x, s) = 1 \quad \text{if and only if} \quad f(x, s) = \alpha.$$

Clearly, if J denotes the all-ones matrix, then since $M(f)$ is a mosaic of incidence structures,

$$\sum_{\alpha} N_{\alpha} = J. \quad (5.1)$$

The function f is an ε -ACFU hash function if and only if

(1) denoting the all-ones vector of appropriate dimension by j ,

$$N_{\alpha} j = \frac{|\mathcal{S}|}{|\mathcal{A}|} j, \quad (5.2)$$

(2) and for $x \neq x'$,

$$(N_{\alpha} N_{\alpha}^T)(x, x') \leq \frac{\varepsilon |\mathcal{S}|}{|\mathcal{A}|}.$$

In particular, for any nonnegative vector $p \in \mathbb{R}^{\mathcal{X}}$,

$$\begin{aligned} p^T N_{\alpha} N_{\alpha}^T p &\leq \frac{|\mathcal{S}|}{|\mathcal{A}|} p^T p + \frac{\varepsilon |\mathcal{S}|}{|\mathcal{A}|} ((p^T j)^2 - p^T p) \\ &= \frac{|\mathcal{S}|}{|\mathcal{A}|} ((1 - \varepsilon) p^T p + \varepsilon (p^T j)^2). \end{aligned} \quad (5.3)$$

This inequality will be the key in the application of ε -ACFU hash functions to privacy amplification.

Now assume that X and Z are random variables on the finite alphabets $\mathcal{X}$ and $\mathcal{Z}$, respectively, with joint probability vector p_{XZ} . “Privacy amplification” means that we want to transform X into another random variable A whose probability distribution is close to uniform and about which an adversary observing Z knows as little as possible. For this transformation, we use an ε -ACFU hash function $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ with an associated family $(N_\alpha)_{\alpha \in \mathcal{A}}$ of 01-matrices. For the second argument of f , we choose an input uniformly at random, which may also be known to the adversary. This setting gives us a joint probability distribution on $\mathcal{X} \times \mathcal{Z} \times \mathcal{S} \times \mathcal{A}$ for the random variables X, Z, S, A ,

$$p_{XZSA}(x, z, s, \alpha) = p_{XZ}(x, z) \cdot \frac{1}{|\mathcal{S}|} \cdot N_\alpha(x, s).$$

For any $z \in \mathcal{Z}$, we define the vector $p_z \in \mathbb{R}^{\mathcal{X}}$ by

$$p_z(x) = p_{XZ}(x, z),$$

such that $p_z^T j = p_Z(z)$. The joint distribution of Z, S and A is

$$p_{ZSA}(z, s, \alpha) = \frac{1}{|\mathcal{S}|} \sum_x p_{XZ}(x, z) N_\alpha(x, s) = \frac{1}{|\mathcal{S}|} (p_z^T N_\alpha)(s). \quad (5.4)$$

In particular, by (5.2),

$$p_{ZA}(z, \alpha) = p_z^T j \cdot \frac{1}{|\mathcal{A}|} = p_Z(z) \cdot \frac{1}{|\mathcal{A}|}. \quad (5.5)$$

Hence Z and A are stochastically independent, and we even obtain a uniform distribution for A (not just an approximation).

Before we can show that the adversary knows little about A , we have to define what this should mean. We impose the strong requirement that the adversary, knowing S and Z , should not be able to distinguish any two values α, α' of which it knows that one is the true one. In order to formalize this, we recall the definition of conditional probabilities. If Y_1, Y_2 are any random variables with joint probability distribution $p_{Y_1 Y_2}$, then the conditional probability of Y_1 given $Y_2 = y_2$ is defined by

$$p_{Y_1|Y_2=y_2}(y_1) = p_{Y_1 Y_2}(y_1|y_2) = \frac{p_{Y_1 Y_2}(y_1, y_2)}{p_{Y_2}(y_2)}$$

if $p_{Y_2}(y_2) > 0$, otherwise it is undefined. We require that

$$\max_{\alpha, \alpha'} \|p_{ZS|A=\alpha} - p_{ZS|A=\alpha'}\|_1 \quad (5.6)$$

should be sufficiently small, where $\|\cdot\|_1$ is the ℓ_1 -norm on the set of probability vectors, or equivalently, the total variation distance on the space of probability measures. (It will depend on the application what “sufficiently” means. See Example 7.) By (5.4) and (5.5),

$$p_{ZS|A}(z, s|\alpha) = \frac{|\mathcal{A}|(p_z^T N_\alpha)(s)}{|\mathcal{S}|}.$$

Let $p_Z p_S$ denote the product of the distributions p_Z and p_S , such that $p_Z p_S(z, s) = p_Z(z)/|\mathcal{S}|$. Without loss of generality, we may assume that p_Z is everywhere positive. In order to bound (5.6), it is sufficient to find, for every $\alpha \in \mathcal{A}$, an upper bound for

$$\begin{aligned}
 \|P_{ZS|A=\alpha} - P_Z P_S\|_1 &= \sum_{z,s} P_Z(z) P_S(s) \left| \frac{P_{SZ|A=\alpha}(z, s)}{P_Z(z) P_S(s)} - 1 \right| \\
 &\leq \left(\sum_{z,s} P_Z(z) P_S(s) \left(\frac{P_{SZ|A=\alpha}(z, s)}{P_Z(z) P_S(s)} - 1 \right)^2 \right)^{1/2} \\
 &= \left(\sum_{z,s} \frac{P_{SZ|A=\alpha}(z, s)^2}{P_Z(z) P_S(s)} - 1 \right)^{1/2} \\
 &= \left(\frac{|\mathcal{A}|^2}{|S|} \sum_{z,s} \frac{(p_z^T N_\alpha)(s)^2}{p_z^T j} - 1 \right)^{1/2}.
 \end{aligned}$$

Using (5.3), the term under the square root satisfies

$$\begin{aligned}
 &\frac{|\mathcal{A}|^2}{|S|} \sum_z \frac{p_z^T N_\alpha N_\alpha^T p_z}{p_z^T j} - 1 \\
 &\leq |\mathcal{A}| \sum_z \frac{(1-\varepsilon)p_z^T p_z + \varepsilon(p_z^T j)^2}{p_z^T j} - 1 \\
 &= (1-\varepsilon)|\mathcal{A}| \sum_z \frac{p_z^T p_z}{p_z^T j} + |\mathcal{A}|\varepsilon - 1.
 \end{aligned}$$

By definition, the sum can be written in terms of a conditional Rényi 2-entropy,

$$\log \sum_z \frac{p_z^T p_z}{p_z^T j} = \log \sum_z P_Z(z) 2^{-H_2(X|Z=z)} = -H_2(X|Z).$$

(Note: There exists a different definition of conditional Rényi 2-entropy, see, e.g., [3].) Hence we obtain the following result.

Theorem 5 *Let $f : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{A}$ be an ε -ACFU hash function and let S be uniformly distributed on $\mathcal{S}$. Choose any $H \geq 0$. Then for any pair (X, Z) of random variables independent of S and satisfying $H_2(X|Z) \geq H$, the key $A = f(X, S)$ is uniformly distributed on $\mathcal{A}$ and satisfies*

$$\max_{\alpha, \alpha' \in \mathcal{A}} \|P_{ZS|A=\alpha} - P_{ZS|A=\alpha'}\| \leq 2 \left((1-\varepsilon)|\mathcal{A}|2^{-H} + |\mathcal{A}|\varepsilon - 1 \right)^{1/2}.$$

Generally, in order to achieve good security in the theorem, we need $\varepsilon \approx 1/|\mathcal{A}|$. For instance, consider the following example.

Example 7 Let a family $\{(X_n, Z_n) : n \geq 1\}$ of pairs of random variables be given. Assume that there exists a number $H > 0$ such that $H_2(X_n|Z_n) \geq nH$ for sufficiently large n . This is the case in the typical situation of secret-key generation from i.i.d. correlated sources where n indicates the number of observed source realizations. Let $\delta > 0$ and choose for every n an ε -ACFU hash function $f_n : \mathcal{X}_n \times \mathcal{S}_n \rightarrow \mathcal{A}_n$ such that X_n lives on $\mathcal{X}_n$ and with $\varepsilon \leq 1/|\mathcal{A}_n|$ and $\log |\mathcal{A}_n| \leq n(H - \delta)$. Then the theorem implies that (5.6) tends to 0 exponentially in n . In particular, this shows that the best-known key rate in the situation of an i.i.d. source is achievable even with our stronger-than-usual security measure, see, e.g., [6, pp. 151–159].

A related application of ACFU hash functions is to *wiretap channels*, see [30]. In this application, an additional requirement is that the blocks $\{x : f(x, s) = \alpha\}$ have constant size. We have seen in Sect. 3 that this poses no real restriction on the ACFU functions.

For application in privacy amplification and the wiretap channel problem, there exist functions which have a smaller seed than ACFU hash functions if $|\mathcal{S}| > |\mathcal{X}|$, namely $|\mathcal{S}| = |\mathcal{X}|$, but which also achieve the best known key or channel rates in the standard settings like the i.i.d. source setting from the example. They are based on mosaics of near-Ramanujan graphs, i.e., edge decompositions of a complete bipartite graph with equal-sized color classes into subgraphs each of which has a very small second-largest eigenvalue [29]. However, so far we do not know of any such mosaics whose corresponding functions are efficiently computable.

6 Open questions

After our extension results (Theorems 3 and 4), we discussed how the original function g and the generated $\hat{g}$ or $\check{g}$ relate with respect to equalities in the lower bounds on the seed sizes. What remained open was whether every seed-optimal OCFU hash function can be derived from a seed-optimal OU hash function. Formulated in terms of mosaics and designs, the question is: *Are the members of every mosaic of BIBDs resolvable?* In other words, is the method of Gnille, Geferath and Pavčević (Corollary 3) essentially the only way of constructing a mosaic of BIBDs? By Corollary 2, the members of a mosaics of $\text{BIBD}(v, k, \lambda)$ certainly need to satisfy the necessary condition $b \geq v + r - 1$ for resolvable designs.

If this question can be answered in the positive, then this also implies that dually, any ε -ACFU hash function with equality in (3.4) is the point extension of an ε -ASU hash function satisfying equality in (3.11). Another consequence would be that the sum of a mosaic of BIBDs is doubly-resolvable [5, Remark I.5.16].

More generally, a similar question can be posed about the structure of mosaics which neither in their “primal” nor their dual version consist of BIBDs. In terms of ACFU hash functions, this could in particular clarify the relation between seed-optimal ACFU hash functions with equality in (3.5) and seed-optimal ASU hash functions with equality in Lemma 8.

Acknowledgements M. Wiese and H. Boche were supported by the German Federal Ministry of Education and Research (BMBF) within the programme “Souverän. Digital. Vernetzt.”, project 6G-life, grant 16KISK002, and within the project NewCom under grant 16KIS1003K. H. Boche was additionally supported in part by BMBF through the project “Quantum Token Theory and Applications – Q.TOK” under grant 16KISQ037K.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bellare M., Tessaro S.: Polynomial-time, semantically-secure encryption achieving the secrecy capacity. [arxiv:1201.3160v2](https://arxiv.org/abs/1201.3160v2) (2012).
2. Bellare M., Tessaro S., Vardy A.: Semantic security for the wiretap channel. In: Safavi-Naini R., Canetti R. (eds.) CRYPTO 2012, vol. 7417, pp. 294–311. Lecture Notes in Computer Science. Springer, Berlin Heidelberg (2012).
3. Bennett C.H., Brassard G., Crépeau C., Maurer U.M.: Generalized privacy amplification. *IEEE Trans. Inform. Theory* **41**(6), 1915–1923 (1995).
4. Bennett C.H., Brassard G., Robert J.-M.: Privacy amplification by public discussion. *SIAM J. Comput.* **17**(2), 210–229 (1988).
5. Beth T., Jungnickel D., Lenz H.: Design Theory, vol. 69, 2nd edn Encyclopedia of mathematics and its applications. Cambridge University Press, Cambridge (1999).
6. Bloch M., Barros J.: Physical-Layer Security. Cambridge University Press, Cambridge (2009).
7. Carter J., Wegman M.N.: Universal classes of hash functions. *J. Comput. Syst. Sci.* **18**(2), 143–154 (1979).
8. Chou R.A., Bloch M.R.: Separation of reliability and secrecy in rate-limited secret-key generation. *IEEE Trans. Inform. Theory* **60**(8), 4941–4957 (2014).
9. Čustić A., Krčadinac V., Zhou Y.: Tiling groups with difference sets. *Electron. J. Comb.* **22**(2), 5954–5964 (2015).
10. Dembowski P.: Finite Geometries, vol. 44. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin (1968).
11. Gnilke O.W., Greferath M., Pavčević M.O.: Mosaics of combinatorial designs. *Des. Codes Cryptogr.* **86**(1), 85–95 (2017).
12. Hayashi M.: Upper bounds of eavesdropper’s performances in finite-length code with the decoy method. *Phys. Rev. A* **76**(1), 012329 (2007).
13. Hayashi M.: Security analysis of ϵ -almost dual universal₂ hash functions: Smoothing of min entropy versus smoothing of Rényi entropy of order 2. *IEEE Trans. Inform. Theory* **62**(6), 3451–3476 (2016).
14. Hayashi M., Matsumoto R.: Secure multiplex coding with dependent and non-uniform multiple messages. *IEEE Trans. Inform. Theory* **62**(5), 2355–2409 (2016).
15. Krawczyk H.: LFSR-based hashing and authentication. In: Advances in Cryptology – CRYPTO ’94, pp. 129–139. Springer, Berlin (1994).
16. Mansour Y., Nisan N., Tiwari P.: The computational complexity of universal hashing. *Theoret. Comput. Sci.* **107**(1), 121–133 (1993).
17. Maurer U., Wolf S.: Information-theoretic key agreement: From weak to strong secrecy for free. In: Preneel B. (ed.) Advances in Cryptology - EUROCRYPT 2000, vol. 1807, pp. 351–368. Lecture Notes in Computer Science. Springer, Berlin (2000).
18. Mikhail W.F.: An inequality for balanced incomplete block designs. *Ann. Math. Stat.* **31**(2), 520–522 (1960).
19. Renes J.M.: On privacy amplification, lossy compression, and their duality to channel coding. *IEEE Trans. Inform. Theory* **64**(12), 7792–7801 (2018).
20. Roy P.M.: A note on the resolvability of balanced incomplete block designs. *Calcutta Stat. Assoc. Bull.* **4**(3), 130–132 (1952).
21. Sarwate D.V.: A note on universal classes of hash functions. *Inform. Process. Lett.* **10**(1), 41–45 (1980).
22. Shrikhande M.S., Singhi S.S.: Quasi-Symmetric Designs. London Mathematical Society Lecture Note Series, vol. 164. Cambridge University Press, Cambridge (1991).
23. Stinson D.R.: Combinatorial techniques for universal hashing. *J. Comput. Syst. Sci.* **48**(2), 337–346 (1994).
24. Stinson D.R.: Universal hashing and authentication codes. *Des. Codes Cryptogr.* **4**(3), 369–380 (1994).
25. Tsurumaru T., Hayashi M.: Dual universality of hash functions and its applications to quantum cryptography. *IEEE Trans. Inform. Theory* **59**(7), 4700–4717 (2013).
26. van Assche G.: Quantum Cryptography and Secret-Key Distillation. Cambridge University Press, Cambridge (2006).
27. van Trung T.: A combinatorial characterization of certain universal classes of hash functions. *J. Comb. Des.* **2**(3), 161–166 (1994).
28. Wegman M.N., Carter J.: New hash functions and their use in authentication and set equality. *J. Comput. Syst. Sci.* **22**(3), 265–279 (1981).
29. Wiese M., Boche H.: Semantic security via seeded modular coding schemes and Ramanujan graphs. *IEEE Trans. Inform. Theory* **67**(1), 52–80 (2021).
30. Wiese M., Boche H.: Mosaics of combinatorial designs for information-theoretic security. *Des. Codes Cryptogr.* **90**(3), 593–632 (2022).

31. Wiese M., Boche H.: ε -almost collision-flat universal hash functions motivated by information-theoretic security. In: Proceedings of IEEE International Symposium on Information Theory (ISIT23), pp. 2589-2594 (2023).

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