Efficient Strongly Universal and Optimally Universal Hashing
(Extended Abstract)

Philipp Woelfel

Lehrstuhl Informatik 2, Universität Dortmund, 44221 Dortmund, Germany
woelfel@Ls2.cs.uni-dortmund.de

Abstract. New hash families are analyzed, mainly consisting of the hash functions

$$h_{a,b} : \{0,\ldots,u-1\} \to \{0,\ldots,r-1\}, \quad x \mapsto \big((ax+b) \bmod (kr)\big) \operatorname{div} k.$$

Universal classes of such functions have already been investigated in [5, 6], and used in several applications, e.g. [3, 9]. The new constructions which are introduced here, improve in several ways upon the former results. Some of them achieve a smaller universality parameter, i.e., two keys collide under a randomly chosen function with a smaller probability. In fact, an optimally universal hash class is presented, which means that the universality parameter achieves the minimum possible value. Furthermore, the bound of the universality parameter of a known, almost strongly universal hash family is improved, and it is shown how to reduce the size of a known class, retaining its properties. Finally, a new composition technique for constructing hash classes for longer keys is presented. Its application leads to efficient hash families which consist of linear functions over the ring of polynomials over $\mathbb{Z}_m$.

1 Introduction

Since its introduction by Carter and Wegman [4, 18], the concept of universal hashing has proven to be very successful. A large number of applications, partly in areas which are remote from the original problem, are known. They range from complexity theoretical investigations over message authentication to standard applications like dictionary implementations or integer sorting.

Let two finite sets U (universe) and R (range) be given, as well as a family H of hash functions, which map U to R. H is called ε–almost universal (ε–AU), if two keys $x_1 \neq x_2$ from U collide under a randomly chosen hash function with a probability of at most ε. It is called ε–almost strongly universal (ε–ASU), if the probability that x_1, x_2 are hashed to arbitrary fixed values $y_1, y_2 \in R$ is at most ε (for precise definitions see Section 2.1.) Applications for ε–AU classes can be found e.g. in [4, 10, 8, 6, 3, 12], and Wigderson [19] gives a bibliography of 36 papers concerning the topic of pairwise independence, i.e., strongly universal hashing. In recent years, a new type of hash family was considered. The so called

M. Kutyłowski, L. Pacholski, T. Wierzbicki (Eds.): MFCS'99, LNCS 1672, pp. 262–272, 1999.

ε–Δ universal families are a type of ε–AU families, which were mainly used for message authentication [11, 13] or the construction of ε–ASU classes [17].

There are several important properties of hash families. They should be easy to implement, and their cardinality should be as small as possible, since it determines the amount of random bits that are needed to choose a function. Furthermore, the universality parameter ε has great influence on the performance of the underlying application. E.g. in the dynamic dictionary described in [8], hash functions are chosen at random from an ε–AU hash class, until they satisfy certain properties. The expected number of hash functions that have to be tested, highly depends upon the universality parameter ε. Also the expected memory consumption is influenced by this parameter, because the sizes of the hash tables increase, if "bad" hash functions are chosen.

The time needed to evaluate functions from a universal hash class is another important factor. Experiments in [7] have shown that in the mentioned dictionary implementation, a large amount of computing time is used for the evaluation of hash functions. We therefore informally say, a hash class is *efficient* if its functions can be evaluated efficiently.

While many constructions of AU, ASU and Δ universal hash classes are known, they differ very much in how far they satisfy the above properties. Some families, as those mainly used in message authentication, achieve only a constant universality parameter (e.g. $\varepsilon = 2^{-64}$). For many other applications this is not enough. They require c/r–AU or c/r^2–ASU hash classes, where c is a constant to be minimized. Most known families achieving such bounds require prime numbers of size $|U|$, or arithmetic in finite fields, while others need matrix multiplication or convolution over some finite field. Such solutions are either inefficient, or introduce the problem of providing prime numbers or irreducible polynomials. As discussed in [12, 2, 1], this may be inconvenient and inefficient for applications (e.g. dictionaries or integer sorting), where the size of the universe is determined at runtime.

To overcome these restrictions, functions of the type $h_{a,b} : \{0, \dots , u - 1\} \to \{0, \dots , r - 1\}$, $x \mapsto \big((ax + b) \bmod (kr)\big) \operatorname{div} k$ were used. The so called "multiplicative class" [6] is $2/r$–AU, and consists of functions $h_{a,0}$, where $0 < a < u$ is odd and $u = kr$ is a power of 2. In [5], the concept was generalized. The "linear class" consists of functions $h_{a,b}$ for $0 \leq a, b < kr$ and $k \geq u - 1$. It was shown, that this family is $1/r^2$–ASU for u, k and r being powers of the same prime and $5/(4r^2)$–ASU otherwise. Both classes have already been found useful in applications, such as [3, 9], and experimental studies in [7] have shown that the multiplicative class performs well in dynamic dictionary implementations.

In this paper, we construct several new hash families with improved properties, mainly consisting of the functions $h_{a,b}$. Among them are ε–Δ universal and ε–ASU classes for small ε, which have significantly smaller cardinality than the linear class. Also, a new proof yields an improved bound for the universality parameter of the linear class, in the case that u, k and r are not powers of the same prime (Section 3). New constructions of AU hash classes are presented in Section 4. The probably most important results are an $1/r$–AU and an optimally universal hash class. For the case, that kr is a power of 2, there are no other constructions of this kind known, which are comparable in efficiency, without sacrificing a reasonable size. Actually, there has been mentioned only one direct construction of an optimally universal hash class in literature, that is suitable for implementation. It requires though, arithmetic over finite fields [16]. Section 5 considers linear functions over polynomials. The resulting hash classes have properties similar to those given in the Sections 3 and 4, and are very suitable for hashing longer keys without the need for implementing long integer arithmetic.

2 Preliminaries

2.1 Definitions

Let U and R be two finite sets with $u := |U|$, $r := |R|$ and $1 < r \leq u$. A function $h : U \to R$ is said to be a hash function with *universe* U and *range* R. Various types of families of hash functions have been studied. For the following definitions of the most important types, we consider H to be a (multi-)set of functions from U to R and h to be chosen randomly from H (according to the uniform distribution on H).

1. H is *ε-almost universal* (short: ε–AU), if $\text{Prob}\big(h(x_1) = h(x_2)\big) \leq \varepsilon$ for all $x_1 \neq x_2$ in U. An $1/r$–AU class is also called *universal*, and an ε–AU class with $\varepsilon = (u-r)/(ur-r)$ is called *optimally universal*.
2. H is *ε-almost strongly universal* (short: ε–ASU), if for all $y_1, y_2 \in R$ and $x_1 \neq x_2$ in U the following two conditions are satisfied:
   (a) $\text{Prob}\big(h(x_1) = y_1\big) = 1/r$.
   (b) $\text{Prob}\big(h(x_1) = y_1 \wedge h(x_2) = y_2\big) \leq \varepsilon$.
   If H is $1/r^2$–ASU, it is also called *strongly universal* (short: SU).
3. Let R be an abelian group. H is *ε-Δ universal* (short: ε–ΔU), if $\text{Prob}\big(h(x_2) - h(x_1) = d\big) \leq \varepsilon$ for all $d \in R$ and $x_1 \neq x_2$ in U. If H is $1/r$–ΔU, it is also called *Δ universal* (short: ΔU).

The term optimally universal was introduced in [14], where it was shown that for any class H of hash functions, there exist two keys $x_1 \neq x_2$ in U, which collide under at least $|H|(u-r)/(ur-r)$ functions from H. It is also well known, that for ASU and ΔU classes, universality parameters of $1/r^2$ and $1/r$ respectively are the best possible.

2.2 Notation

In the rest of this paper, we will consider R to be the additive group modulo r and U to be the u-element subset $\{0, \ldots, u-1\}$ of the ring $M = \mathbb{Z}_m$, where $m = kr \geq u$ for some integer k. If not stated otherwise, additions and multiplications between two elements from M are over this ring. An addition over the group R

will be denoted by $\oplus$. Furthermore, for $x \in M$ we define $Mx = \{ax \mid a \in M\}$. Finally, for two integers x, y we let $\gcd(x, y)$ be the greatest common divisor of x and y over $\mathbb{Z}$.

For $a, b \in M$ define the mappings

$$g_b : M \to R,\ x \mapsto (x + b) \operatorname{div} k \quad \text{and} \quad h_{a,b} : U \to R,\ x \mapsto g_b(ax).$$

In other words, $h_{a,b}(x)$ could be written as $\big((ax + b) \bmod m\big) \operatorname{div} k$. For multisets $A, B \subseteq M$, the hash class $\mathcal{H}_{k,A,B}$ consists of the functions $h_{a,b}$ with $a \in A$ and $b \in B$.

Note that the functions $h_{a,b}$ can be evaluated very efficiently on ordinary computer architectures. Particularly, if $m = kr$ is a power of 2, then the modulo operation and the division can be replaced by a bitwise "and" and a bitwise shift. In this case, the time needed for the evaluation of the functions $h_{a,b}$ is very close to the time needed for one multiplication. Another slight reduction of the evaluation time can be achieved by choosing b to be 0 for all functions, since then no addition is required. Generally though, this leads to an increase of the universality parameter. Since in most applications the size of a hash table can be increased by a factor of at most 2 without much effort, having $m = 2^n$ is probably the most important case. In this case the functions $h_{a,b}$ can be evaluated in $O(n \log n \log\log n)$ steps on Turing machines and in depth $O(\log n)$ and size $O(n \log n \log\log n)$ by circuits with fan-in 2, using the Schönhage-Strassen multiplication method [15]. As discussed in [5], circuits for hash classes that involve prime number arithmetic or finite field arithmetic lack uniformity and are larger.

3 ΔU and SU Hashing

We first state a lemma which is crucial for the proofs in the next two sections.

Lemma 1. *Let $d \in M$ and $x_1, x_2 \in U$ with $\delta = x_2 - x_1 \neq 0$ and $\gamma = |\gcd(\delta, m)|$. Then*

(a) $M\delta \;=\; \{i\gamma \mid 0 \le i < m/\gamma\}$

(b) $$\operatorname*{Prob}_{a \in M}\big(ax_2 - ax_1 = d\big) \;=\; \operatorname{Prob}\big(a\delta = d\big) \;=\; \begin{cases} \gamma/m, & \text{if } d \in M\delta. \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Observing that $M\delta$ is the principal ideal generated by δ, (a) follows from basic algebra. For the proof of (b), consider the mapping $\varphi_\delta : M \to M\delta,\ a \mapsto a\delta$. Since φ_δ is a surjective group homomorphism, the number of $a \in M$ satisfying $a\delta = d$ is 0 if $d \notin M\delta$, and otherwise is $\left|\varphi_\delta^{-1}(d)\right| = |M|/|M\delta| = \gamma$. □

3.1 Homogeneous Functions

We now consider the $\epsilon-\Delta$U hash families $\mathcal{H}_{k,M,\{0\}}$ for $k \geq u-1$. They consist of the functions $h_{a,0}$ and can – as noted in the introduction – be evaluated without addition. I.e., $h_{a,0}(x) = \big((ax) \bmod m\big) \operatorname{div} k$. As we show below, the universality parameter of $\mathcal{H}_{k,M,\{0\}}$ is $2/r$ if kr is a prime power and $3/r$ otherwise.

For the latter case, it can be bounded more precisely as follows. Let $\Gamma_{u,r,k} = \max\{0, \gamma = \gcd(x, kr) \mid 0 \le x < u, \gamma \nmid k\}$. For $k \ge u-1$, which is the case to be considered here, we have $0 \le \Gamma_{u,r,k}/k \le 1$. Moreover, $\Gamma_{u,r,k}/k$ converges to 0 with increasing k, and if kr is a prime power, this term equals 0 in any case.

Theorem 1.

1. *If $k \ge u-1$, then $\mathcal{H}_{k,M,\{0\}}$ is c/r-ΔU, where $c = 2 + \Gamma_{u,r,k}/k \le 3$.*
2. *If $m = kr$ is a power of the prime p and $k \ge u/p$, then $\mathcal{H}_{k,M,\{0\}}$ is $2/r$-ΔU.*

Proof. Let $d \in R$ and $x_1, x_2 \in U$ with $\delta = x_2 - x_1 \neq 0$ and $\gamma = |\gcd(\delta, m)|$ (obviously $\gamma \le k$). For arbitrary $y_1, y_2 \in M$, it is clear that $g_0(y_2) - g_0(y_1) = d$ requires $k(d-1) < y_2 - y_1 < k(d+1)$. Therefore, one obtains by Lemma 1 that

$$\operatorname{Prob}\big(g_0(ax_2) - g_0(ax_1) = d\big) \le \sum_{\substack{y \in M\delta \\ k(d-1)<y<k(d+1)}} \gamma/m \le \lceil 2k/\gamma \rceil \cdot \frac{\gamma}{kr}.$$

This is at most $(2 + \Gamma_{u,r,k}/k)/r$. □

We now use this result to construct an ε/r–ASU hash class with ε as given in the above theorem. The following construction method is described in [17]. Consider a hash family H of functions $U \to R$, where $(R, \oplus)$ is an abelian group. For any $h \in H$ and any $y \in R$ let $\varphi_{h,y} : U \to R$ be defined by $\varphi_{h,y}(x) = h(x) \oplus y$. It is easy to see that if H is ε–ΔU, then the family of functions $\varphi_{h,y}$ for $(h, y) \in H \times R$ is ε/r–ASU.

Corollary 1. *Let $B = \{ik \mid 0 \le i < r\}$.*

1. *If $k \ge u-1$, then $\mathcal{H}_{k,M,B}$ is c/r^2-ASU, where $c = 2 + \Gamma_{u,r,k}/k \le 3$.*
2. *If $m = kr$ is a power of the prime p and $k \ge u/p$, then $\mathcal{H}_{k,M,B}$ is $2/r^2$-ASU.*

Proof. $h_{a,ki}(x) = h_{a,0}(x) \oplus i$. □

3.2 Inhomogeneous Functions

Recall that the ε–ΔU hash class from Theorem 1 consists of functions $h_{a,0}$, which can be evaluated without addition. It can be expected though, that on most computer architectures the time for an extra addition is very small compared to the time needed for the multiplication and the divisions. We will therefore increase the class by adding functions $h_{a,b}$ with $b \neq 0$, to obtain better universality parameters. The resulting class is $9/(8r)$–ΔU, and in the prime power case even $1/r$–ΔU.

The universality parameter ε can be described more precisely by the following term. We obtain $\varepsilon = 1/r + \vartheta_k(\Gamma_{u,r,k})/r$, where for $0 \le \gamma$

$$\vartheta_k(\gamma) := \begin{cases} 0 & \text{if } \gamma = 0. \\ \dfrac{1}{4\lfloor k/\gamma \rfloor \cdot (\lfloor k/\gamma \rfloor + 1)} & \text{otherwise.} \end{cases}$$

Although this may look technical, the important properties can be described easily. For $k \geq u-1$, we obtain $\vartheta_k(\Gamma_{u,r,k}) \leq 1/8$. Further, since $\Gamma_{u,r,k}$ is at most $u-1$, we find that $\vartheta_k(\Gamma_{u,r,k})$ is in $O(1/k^2)$, and therefore by an increase of k converges quite fast to 0.

Lemma 2. *For $1 \leq i \leq n$ let (y_i, y_i') be pairs of elements from M where the $\delta_i = y_i - y_i'$ are pairwise distinct and form a set $\{z\gamma + C \mid 0 \leq z < m/\gamma\}$ for some $C \in M$ and some $\gamma \leq k$. Then for any $d \in R$ and randomly chosen $1 \leq i \leq n$, $0 \leq b < k$,*

$$\mathrm{Prob}\left(g_b(y_i) - g_b(y_i') = d\right) \leq \begin{cases} 1/r, & \text{if } \gamma \text{ divides } k. \\ \left(1 + \vartheta_k(\gamma)\right)/r, & \text{otherwise.} \end{cases}$$

Proof. Let $1 \leq j \leq n$. It can easily be shown that the number of $b \in \{0, \ldots, k-1\}$ satisfying $g_b(y_j) - g_b(y_j') = d$ is exactly $\delta_j \bmod k$ if $\delta_j \in \{k(d-1), \ldots, kd-1\}$ and $k - \delta_j \bmod k$ if $\delta_j \in \{kd, \ldots, k(d+1)-1\}$. Otherwise, no $b \in B$ will satisfy this condition.

Now assume w.l.o.g. that those δ_i in $\{k(d-1), \ldots, kd-1\}$ are $\delta_1 < \ldots < \delta_t$ and those in $\{kd, \ldots, k(d+1)-1\}$ are $\delta_{t+1} < \ldots < \delta_{t+t'}$. If we let

$$S = \sum_{j=1}^{t} \delta_j \bmod k + \sum_{j=t+1}^{t'} (k - \delta_j \bmod k),$$

then the probability of $g_b(y_i) - g_b(y_i') = d$ is $S/k \cdot \gamma/m = 1/r \cdot S\gamma/k^2$. In order to bound the sum S, we have to consider three cases, namely $t = t'$, $t = t' - 1 = \lfloor k/\gamma \rfloor$ and $t = t' + 1 = \lceil k/\gamma \rceil$.

First, assume $t = t'$, and note that $\gamma | k$ implies this case. Using the fact that $\delta_j \bmod k - \delta_{t+j} \bmod k = k - t\gamma$ $(1 \leq j \leq t)$, we obtain $S\gamma/k^2 = 2t(\gamma/k) - t^2(\gamma^2/k^2)$. And since the real function $F(x) = \lambda x - \tau x^2$ has a global maximum with value $\lambda^2/(4\tau)$, we have $S\gamma/k^2 \leq (2t)^2/(4t^2) = 1$.

Now assume $t = t' - 1 = \lfloor k/\gamma \rfloor$. We obtain

$$\begin{aligned} S &= \sum_{j=1}^{t} \left(\delta_j \bmod k + k - \delta_{t+j+1} \bmod k\right) + k - \delta_{t+1} \bmod k \\ &\leq k + t\left(2k - \gamma(t+1)\right), \end{aligned}$$

thus $S\gamma/k^2 \leq (2t+1)(\gamma/k) - t(t+1)(\gamma/k)^2$. For the same reason as above, this is at most $1 + 1/\left(4\lfloor k/\gamma \rfloor(\lfloor k/\gamma \rfloor + 1)\right)$.

The last case, $t = t' + 1 = \lceil k/\gamma \rceil$, follows from symmetry reasons. □

The functions $h_{a,b}$ with $a \in M$ and $b \in \{0, \ldots, k-1\}$ now give the desired ε–ΔU class. The proof is omitted here, since it mainly is the combination of Lemma 1 and Lemma 2.

Theorem 2. *Let $m = kr$ and $B = \{0, \ldots, k-1\}$.*

1. *If $k \geq u-1$, then $\mathcal{H}_{k,M,B}$ is c/r–ΔU, where $c = 1 + \vartheta_k(\Gamma_{u,r,k}) \leq 9/8$.*
2. *If m is a power of the prime p and $k \geq u/p$, then $\mathcal{H}_{k,M,B}$ is ΔU.*

Using the techniques from Section 3.1, it is now an easy task to construct a $9/(8r^2)$–ASU hash class. The resulting construction is the "linear class", which was first investigated by Dietzfelbinger [5]. He showed that it is $5/(4r^2)$–ASU and even SU for m being a prime power. Our bound therefore is somewhat tighter in the case of m being no prime power.

Corollary 2. *If $k \geq u-1$, then $\mathcal{H}_{k,M,M}$ is c/r^2–ASU, where $c = 1+\vartheta_k(\Gamma_{u,r,k}) \leq 9/8$.*

For the prime power case, we present now a new SU family, which has a smaller cardinality than the "linear class". While for the above construction, about $\log(u) + \log(r)$ random bits are necessary to choose the parameter b, approximately $\log(u)/2$ random bits can be spared by using the following hash family.

Theorem 3. *Let r be a power of the prime p and $k = p^K \geq u-1$.*

1. *If $B = \{ip^{\lceil K/2 \rceil} \mid 0 \leq i < p^{\lfloor K/2 \rfloor}\}$, then $\mathcal{H}_{k,M,B}$ is ΔU.*
2. *If $B = \{ip^{\lceil K/2 \rceil} \mid 0 \leq i < rp^{\lfloor K/2 \rfloor}\}$, then $\mathcal{H}_{k,M,B}$ is SU.*

The somewhat lengthy proof will be given in the full version of this paper.

4 Universal and Optimally Universal Hashing

4.1 Universal Hashing

Clearly, any ε–ΔU hash family is also ε–AU. In many situations though, one is only interested in finding hash classes where two arbitrary keys collide with low probability. The much stronger property of ε–Δ universality is often unnecessary. We are therefore interested in finding AU hash classes, that sacrifice the ASU or ΔU property for being more efficient and/or smaller in cardinality.

For the following constructions we will choose $k = u/r$. The classes are smaller than those from the former section, and can in many situations be evaluated more efficiently. Assume that a computer word consists of N bits, and that we have $u = 2^N$ and $r = 2^{N'}$ for some $N' < N$. On most processors, a multiplication modulo 2^N can be evaluated much faster than modulo $2^{N+N'}$, so that having $k = u/r$ (compared to $k \approx u$) leads to an improvement in efficiency. We start again with a homogeneous version, allowing the functions to be computed without addition.

Theorem 4. *Let $m = kr \geq u$ be a power of the prime p, and $A = \{ip+1 \mid 0 \leq i < m/p\}$. Then $\mathcal{H}_{k,A,\{0\}}$ is $2/r$–AU.*

Proof. Let $x_1, x_2 \in U$ with $\delta = x_2 - x_1 \neq 0$. If δ is a multiple of k, then $k \leq a\delta \leq m - k$ for any $a \in A$, thus x_1 and x_2 do not collide at all. Assume therefore that δ is not a multiple of k, thus k is a multiple of $\gamma = |\gcd(p\delta, m)| = p \cdot |\gcd(\delta, m/p)|$. Then obviously $A\delta = \{i\gamma + \delta \mid 0 \leq i < m/\gamma\}$. So, similar to the proof of Theorem 1, one obtains

$$\mathrm{Prob}\left(g_0(ax_2) = g_0(ax_1)\right) \;\leq\; \sum_{\substack{y \in A\delta \\ -k<y<k}} \gamma/m \;\leq\; \lceil 2k/\gamma \rceil \cdot \frac{\gamma}{kr},$$

which is at most $2/r$. □

For m a power of 2, almost the same hash family was established in [6], and called the multiplicative class. Our construction generalizes the concept for arbitrary prime powers m. Note also, that if kr is a power of 2, it is is possible to use only half the functions. It can be proven, that in this case a can be chosen randomly from the set of odd numbers $\{1, 3, \ldots, m/4 - 1\}$ without increasing the universality parameter.

The construction of the universal class $\mathcal{H}^{\mathrm{univ}}$ which we present now, is completely new. Actually, no other universal class is known which is of reasonable cardinality and as efficient as ours for kr being a power of 2.

Theorem 5. *Let r be a power of the prime p, and $k = p^K \geq u/r$. Furthermore, let $A = \{ip + 1 \mid 0 \leq i < m/p\}$ and $B = \left\{ip^{\lceil K/2 \rceil} \mid 0 \leq i < p^{\lfloor K/2 \rfloor}\right\}$. Then $\mathcal{H}^{\mathrm{univ}} := \mathcal{H}_{k,A,B}$ is universal.*

More precisely, any $x_1 \neq x_2$ in U collide under a randomly chosen $h \in \mathcal{H}^{\mathrm{univ}}$ with the same probability $1/r$, if $|\gcd(x_2 - x_1, m)| < k$. Otherwise, they do not collide at all. We omit the proof, which is very similar to the proofs of Theorem 3 and 4.

4.2 Optimally Universal Hashing

As we said before, some specific keys do not collide under any function from $\mathcal{H}^{\mathrm{univ}}$. The idea of the following construction is to add other functions, which let exactly such keys collide. This leads to an improvement in the universality parameter. In fact, for u being a power of r (which itself is a prime power), the universality parameter takes its minimum possible value, i.e., the class is optimally universal. Note that there is an equivalence between optimally universal hash classes and resolvable balanced incomplete block designs (RBIBDs) — see [16]. Although many existence results of RBIBDs are known, not much attention has been paid in the literature, on how to efficiently evaluate them. Likewise, no very practical constructions of optimally universal hash classes have been found, yet. The hash family we present now, fills in the gap, since it is of reasonable size and its functions are easy to implement and efficient to evaluate. The proof is omitted due to space limits.

Theorem 6. *Let r be a power of the prime p, $m = r^t \geq u$ for some integer t and $k = p^K = r^{t-1}$. Further, let $A = \{(ip+1)\cdot r^j \mid 0 \leq j < t, 0 \leq i < m/(pr^j)\}$ and $B = \{ip^{\lceil K/2\rceil} \mid 0 \leq i < p^{\lfloor K/2 \rfloor}\}$. Then for $\mathcal{H}^{\text{opt}} := \mathcal{H}_{k,A,B}$ the following holds:*

1. *Any two distinct keys collide with the same probability $(m-r)/(mr-r) < 1/r$ under a randomly chosen function.*
2. *If $m = u$, then $\mathcal{H}^{\text{opt}}$ is optimally universal.*

Assuming that $m = u = r^t$, for $\mathcal{H}^{\text{opt}}$ we have $|A| = (u/p)\cdot(r^t-1)/(r^t - r^{t-1})$, which is less than $(u/p)\cdot r/(r-1)$. So, this class is not much larger than $\mathcal{H}^{\text{univ}}$, where $|A| = u/p$. Since the evaluation of the functions from both classes is equally efficient (provided that the word length for the multiplication is equal), $\mathcal{H}^{\text{opt}}$ may be preferred in many situations. In others though, namely those, where the smallest $r^t \geq u$ is much larger than u, $\mathcal{H}^{\text{univ}}$ might be the better choice.

5 Hashing Polynomials

In some situations, it is necessary to hash keys, which do not fit into a single computer word. In such cases, the evaluation of the functions $h_{a,b}$ may be inconvenient and inefficient, since multiplication of long integers would have to be implemented. If an ϵ–ΔU hash class H with universe U and range R is given, then the following construction allows us to hash v words from U to ρ words from R with a universality parameter of ϵ^ρ.

Proposition 1. *Let H be an ε-ΔU family of hash functions $U \to R$. Then for any two integers $1 \leq \rho \leq v$, there exists an ε^ρ-ΔU family of $|H|^{v+\rho-1}$ hash functions $U^v \to R^\rho$.*

Proof. For any $\underline{h} = (h_1, \dots, h_{v+\rho-1}) \in H^{v+\rho-1}$ let $\Phi_{\underline{h}} : U^v \to R^\rho$ map $(x_1, \dots, x_v)$ to $(y_1, \dots, y_\rho)$, where $y_k = \sum_{j=1}^{v} h_{k+j-1}(x_j)$. We show that the class consisting of all functions $\Phi_{\underline{h}}$, where $\underline{h} \in H^{v+\rho-1}$, is ε^ρ–ΔU. Consider two distinct keys $\underline{x} = (x_1, \dots, x_v), \underline{x'} = (x'_1, \dots, x'_v) \in U^v$, and let t be the highest index with $x_t \neq x'_t$. Furthermore, let $(d_1, \dots, d_\rho) = \Phi_{\underline{h}}(\underline{x'}) - \Phi_{\underline{h}}(\underline{x})$ for some randomly chosen $\underline{h} = (h_1, \dots, h_{v+\rho-1})$, and let $1 \leq i \leq \rho$. By the assumption, $h_{i+t-1}(x_t) - h_{i+t-1}(x'_t)$ takes any value in R with a probability of at most ε, and therefore so does d_i. Observing that $d_1, \dots, d_{i-1}$ are independent from the choice of h_{i+t-1}, the result follows immediately by induction over i. □

The direct application of the construction method to the hash classes from the former sections would already yield ΔU and – with the method described in Section 3 – ASU families. A slight modification of this technique though, will in our case lead to somewhat more efficient constructions.

Let the universe $\mathcal{U} = U^v$ with $U = \{0, \dots, u-1\}$ and the range $\mathcal{R} = R^\rho$ with $R = \{0, \dots, r-1\}$. As before, let $m = kr \geq u$ for some integer k and consider all operations over the ring $M = \mathbb{Z}_m$. Further, let $n \geq v$ be some fixed integer (to be determined later). For $\underline{a} = (a_0, \dots, a_{n-1}) \in M^n$ and $\underline{x} =$

$(x_0, \dots, x_{v-1}) \in M^v$, we define the convolution of $\underline{a}$ and $\underline{x}$, $\underline{a} * \underline{x}$, to be the n-element vector $\underline{y} = (y_0, \dots, y_{n-1}) \in M^n$, where $y_l = \sum_{i=0}^{\min(l,v-1)} a_{l-i}x_i$. For our hash functions, only the last ρ coordinates of the convolution need to be computed. For $\underline{a} \in M^n$ and $\underline{b} = (b_0, \dots, b_{\rho-1}) \in M^\rho$, let

$$g_{\underline{b}} : M^n \to M^\rho, \quad (x_0, \dots, x_{n-1}) \mapsto (y_0, \dots, y_{\rho-1}),$$

where $y_i = (x_{i+n-\rho} + b_i) \operatorname{div} k$ and

$$h_{\underline{a},\underline{b}} : \mathcal{U} \to \mathcal{R}, \qquad x \mapsto g_{\underline{b}}(a * x).$$

Note that the functions $h_{\underline{a},\underline{b}}$ are linear functions over the ring of polynomials over M of degree less than n.

The hash classes $\mathcal{H}^{\text{poly}}_{k,\mathcal{A},\mathcal{B}}$ for specific $\mathcal{A} \subseteq M^n$ and $\mathcal{B} \subseteq M^\rho$ consist of the hash functions $h_{\underline{a},\underline{b}}$ for $(\underline{a}, \underline{b}) \in \mathcal{A} \times \mathcal{B}$. Varying the parameters k, $\mathcal{A}$ and $\mathcal{B}$, we get similar to the former sections several ASU, ΔU and AU hash families. We summarize all results in the following theorem.

Theorem 7. *Let $k \geq u/p$, if r is a power of the prime p and $k \geq u-1$ otherwise. Let further $c_1 = \left(2 + \Gamma_{u,r,k}/k\right)^\rho$ and $c_2 = \left(1 + \vartheta_k(\Gamma_{u,r,k})\right)^\rho$ (recall that $c_1 \leq 3^\rho$ and $c_2 \leq (9/8)^\rho$, and if $m = kr$ is a prime power, then $c_1 = 2^\rho$ and $c_2 = 1$). Then the following holds:*

1. Let $n = v + \rho - 1$ and $\mathcal{A} = M^n$.
   (a) If $\mathcal{B} = \{0\}^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is c_1/r^ρ-ΔU.
   (b) If $\mathcal{B} = \{0, \dots, k-1\}^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is c_2/r^ρ-ΔU.
   (c) If $\mathcal{B} = \{ik \mid 0 \leq i < r\}^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is $c_1/r^{\rho+1}$-ASU.
   (d) If $\mathcal{B} = M^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is $c_2/r^{\rho+1}$-ASU.

2. Let $u = r$, $n = v$ and $\mathcal{A} = \{(k, a_1, \dots, a_{n-1}) \mid a_1, \dots, a_{n-1} \in M\}$.
   (a) If $\mathcal{B} = \{0\}^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is c_1/r^ρ-AU.
   (b) If $\mathcal{B} = \{0, \dots, k-1\}^\rho$, then $\mathcal{H}^{poly}_{k,\mathcal{A},\mathcal{B}}$ is c_2/r^ρ-AU.

Sketch of Proof. 1 (a) and (b) are proved by an induction like that in the proof of Proposition 1, but using the techniques from Theorem 1 and Lemma 2 respectively. For (c) and (d), recall the construction method described before Corollary 1. Finally, 2 (a) and (b) are obvious by considering the two cases that either two given vectors (keys) from the universe differ only in the ρ highest coordinates or also in lower ones.

Note that the AU classes require the words of the range and the universe to have the same length, i.e., $U = R$.

Acknowledgements

I want to thank Martin Hühne, who raised my interest in this topic, as well as Ingo Wegener, Martin Dietzfelbinger and Thomas Hofmeister for their helpful comments.

References

1. N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3:289–304, 1992.
2. N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Addendum to "simple constructions of almost k-wise independent random variables". *Random Structures and Algorithms*, 4:119–120, 1993.
3. A. Andersson, T. Hagerup, S. Nilsson, and R. Raman. Sorting in linear time? In *Proc. of 25th ACM STOC*, pages 427–436, 1995.
4. J. L. Carter and M. N. Wegman. Universal classes of hash functions. *J. Comp. Syst. Sci.*, 18:143–154, 1979.
5. M. Dietzfelbinger. Universal hashing and k-wise independent random variables via integer arithmetic without primes. In *Proc. of 13th STACS*, pages 569–580, 1996.
6. M. Dietzfelbinger, T. Hagerup, J. Katajainen, and M. Penttonen. A reliable randomized algorithm for the closest-pair problem. *J. Alg.*, 25:19–51, 1997.
7. M. Dietzfelbinger and M. Hühne. A dictionary implementation based on dynamic perfect hashing. *Proc. of 5th DIMACS Chal. Worksh.*, 1996. To appear.
8. M. Dietzfelbinger, A. Karlin, K. Mehlhorn, F. M. auf der Heide, H. Rohnert, and R. E. Tarjan. Dynamic perfect hashing: Upper and lower bounds. *SIAM J. Comput.*, 23:738–761, 1994.
9. Engelmann and Keller. Simulation-based comparison of hash functions for emulated shared memory. In *Proc. of 5th PARLE*, pages 1–11, 1993.
10. M. L. Fredman, J. Komlós, and E. Szemerédi. Storing a sparse table with $O(1)$ worst case access time. *J. Assoc. Comput. Mach.*, 31:538–544, 1984.
11. H. Krawczyk. LFSR-based hashing and authentication. In *Advances in Cryptology — CRYPTO '94*, pages 129–139, 1994.
12. Y. Matias and U. Vishkin. On parallel hashing and integer sorting. *J. Algorithms*, 12:573–606, 1991.
13. P. Rogaway. Bucket hashing and its application to fast message authentication. In *Advances in Cryptology — CRYPTO '95*, pages 29–42, 1995.
14. D. V. Sarwate. A note on universal classes of hash functions. *Inf. Proc. Letters*, 10:41–45, 1980.
15. A. Schönhage and V. Strassen. Schnelle Multiplikation großer Zahlen. *Computing*, 7:281–292, 1971.
16. D. R. Stinson. Combinatorial techniques for universal hashing. *J. Comp. Syst. Sci.*, 48:337–346, 1994.
17. D. R. Stinson. On the connections between universal hashing, combinatorial designs and error-correcting codes. Report TR95-052, ECCC, `ftp://ftp.eccc.uni-trier.de/pub/eccc/reports/1995/TR95-052/Paper.ps`, 1995.
18. M. N. Wegman and J. L. Carter. New classes and applications of hash functions. In *Proc. of 20th IEEE FOCS*, pages 175–182, 1979.
19. A. Wigderson. The amazing power of pairwise independence. In *Proc. of 26th ACM STOC*, pages 574–583, 1994.