Phorma: perfectly hashable order restricted multidimensional arrays

Lauro Lins\textsuperscript{a}, Sōstenes Lins\textsuperscript{b}, Sílvio Melo\textsuperscript{b}

\textsuperscript{a}Centro de Informática, Universidade Federal de Pernambuco, UFPE, Recife, Brazil
\textsuperscript{b}Departamento de Matemática, UFPE, Brazil

Received 26 July 2002; received in revised form 28 February 2003; accepted 22 March 2003

Abstract

In this paper we propose a simple and efficient data structure yielding a perfect hashing of quite general arrays. The data structure is named phorma, which is an acronym for perfectly hashable order restricted multidimensional array.

MSC: 05A05; 05C90; 06F99

Keywords: Perfect hash function; Digraph; Implicit enumeration; Nijenhuis–Wilf combinatorial family

1. Motivation

Let $a = a_1a_2 \ldots a_n$ and $z = z_1z_2 \ldots z_n$ be $n$-sequences of positive integers, $z \leq a$, meaning $z_i \leq a_i$, $i = 1, 2, \ldots, n$. Suppose that $f(z)$ is a symmetric function on the variables $z_i$, that is, the value of $f(z)$ does not change if the coordinates of $z$ are permuted in an arbitrary way. To store the function $f$, it is enough to allocate space for the values of $f(z)$, where $z_i \geq z_{i+1}$, $1 \leq i \leq n - 1$. Thus, we need to enumerate the $z$’s satisfying $z \leq a$ and the boolean function

$$B^{\geq}_{\text{sym}} = (z_1 \geq z_2) \land (z_2 \geq z_3) \land \cdots \land (z_{n-1} \geq z_n).$$

The motivation for this work is to enumerate and give a perfect hash function [2,4] for multidimensional arrays which have order restrictions on their entries. The simplest example of this situation is when the restrictions are given by $B^{\geq}_{\text{sym}}$. We show
that quite general boolean functions can take the place of $B_n^{\geq}$ and that the large class of enumerative/perfect hash associated problems can be put under a common framework.

To exemplify the appearance of a more complex boolean function, consider the problem of efficiently enumerate all the $L$-shaped pieces with vertices which fit in a $(p \times q)$ integer grid. This is a typical situation treated in [7]. An $L$-shaped piece is a rectangle $R$ from which we have removed a smaller rectangle $r \subseteq R$. Moreover $R$ and $r$ have a corner in common. By effecting rotations, translations and reflections we may suppose that our $L$-shaped piece has a corner in the origin and the common vertex to $r$ and $R$ is the vertex opposite to the origin in rectangle $R$. Positioned in this way, the $L$-piece is represented by a quadruple of positive integers $(X;Y;x;y) = (x_1x_2x_3x_4 \leq a_1a_2a_3a_4 = (p,q,p,q)$, as in Fig. 1.

The geometry imposes the restrictions: (1) $X \geq x$; (2) $Y \geq y$. Symmetry considerations enable us to partition the set of $a$-bounded $L$-pieces into equivalent classes and to distinguish a set $A$ of representatives for these classes. For the occupancy purposes in [7], the $L$-pieces $(X,Y,x,y)$ and $(Y,X,y,x)$ must be considered equivalent. This implies the restrictions: (3) $X \geq Y$ and (4) $X = Y \Rightarrow x \geq y$. In terms of occupancy, $(X,Y,X,y)$ with $y < X$, which is a degenerated $L$, can (and must) be replaced by the rectangle $(X,Y,X,Y)$. Analogously, $(X,Y,x,Y)$ with $x < X$ must be replaced by $(X,Y,X,Y)$. In this way, the equivalence $(X = x) \Leftrightarrow (Y = y)$ holds. The equivalence is rewritten as two opposite implications in the disguised form: (5) $(X \neq x) \lor (Y = y)$ and (6) $(Y \neq y) \lor (X = x)$. Restrictions (1)–(6) are gathered in a boolean expression $B_L$ in terms of the $x_i$’s:

$$B_L = (x_1 \geq x_3) \land (x_2 \geq x_4) \land (x_1 \geq x_2) \land ((x_1 \neq x_2) \lor (x_3 \geq x_4))$$

$$\land ((x_1 \neq x_3) \lor (x_2 = x_4)) \land ((x_2 \neq x_4) \lor (x_1 = x_3)).$$

So, we want to enumerate the 4-sequences $x = x_1x_2x_3x_4$ of positive integers $a \leq a$ and satisfying $B_L$. If, as it is typically needed in packing problems, $a$ is of order $(120,100,120,100) = (120,100)^2$ then we have 23,094,225 $x$’s that satisfies $B_L$ in a total of 144,000,000 possibilities. If $a = (7,5)^2$, then there is a total of 190 $x$’s in 1225 possibilities. The valid 190 $x$’s are in 1–1 correspondence with the $st$-paths in the digraph of Fig. 5.
2. The definition of phorma and the objective of the work

Let \( \mathbb{N} \) be the set of natural numbers, \( \mathbb{N}^* = \mathbb{N} \setminus \{0\} \) and \( N = \{1, 2, \ldots, n\} \). For \( 1 \leq m \leq n \), define \( M = \{1, \ldots, m\} \). Let \( Y^X \) be the set of all functions from \( X \) into \( Y \). Throughout this work, \( \alpha = \alpha_1 \ldots \alpha_n \) is an \( n \)-sequence of positive integers, that is, \( \alpha \in (\mathbb{N}^*)^N \). The relation \( \rho' \leq \rho \) for sequences \( \rho' \) and \( \rho \) of equal length means that \( \rho'_i \leq \rho_i \), for each \( i \)-term of the sequences.

An \( n \)-composition \( \delta = \delta_1 \ldots \delta_m \) is an element of \( (\mathbb{N}^*)^M \) such that \( \sum_{1 \leq m \leq n} \delta_m = n \). The set of \( n \)-compositions is denoted by \( C^n \). Given \( \alpha \), let \( m_\alpha \) be the number of distinct entries in \( \alpha \) and \( m_\delta \) be the length of \( \delta \). Let \( \tilde{\alpha} = \tilde{\alpha}_1 \ldots \tilde{\alpha}_{m_\alpha} \in C^n \) denote the \( n \)-composition, where \( \tilde{\alpha}_i \) is the number of occurrences of the \( i \)th smallest entry of \( \alpha \).

An \( n \)-phorma is a triple \( P = (a, B, C) \) satisfying: (i) \( a = a_1 a_2 \ldots a_n \in (\mathbb{N}^*)^N \); (ii) \( B \) is a boolean function whose literals of \( B \) are of type \( (\alpha, \cdot ) \), where \( \alpha \in (\mathbb{N}^*)^N \) and \( \cdot \in \{=, \geq, <, >, =, \neq\} \); (iii) \( C \subseteq C^n \) is a given set of \( n \)-compositions. The term \( n \)-phorma is an acronym for an \( n \)-dimensional perfectly hashable order restricted multidimensional array.

The objective of this paper is to enumerate the set

\[ A(P) = A(a, B, C) = \{ \alpha \mid \alpha \leq a, \ \alpha \text{ satisfies } B, \ \tilde{\alpha} \in C \}. \]

In the particular case when \( B \) is the empty boolean function, then there are no \( B \)-restrictions and \( A(a, B, C) \) is the subset of \( (\mathbb{N}^*)^N \) consisting of all sequences \( \alpha \leq a \), \( \tilde{\alpha} \in C \).

We construct a bijection \( h : A(P) \rightarrow \{0, 1, \ldots, |A(P)| - 1\} \), so that both \( h \) and \( h^{-1} \) are efficiently computable. Such functions are called perfect hash functions [2,4]. Their usefulness is well known.

As far as we know the problem of finding perfect hash functions for these quite general multidimensional arrays have not been considered before in the literature, whence the lack of more specific references and bibliography. Our solution is based on the theory of combinatorial families developed in [8]. Here we call these families \( NW \)-families and recall their definition in Section 4. The central idea is to associate a digraph to a collection of combinatorial objects in such a way that each object in the family is in 1–1 correspondence with a path in the digraph. A more detailed account of these combinatorial families appears in [9].

From a phorma \( (a, B, C) \) a digraph \( G(a, B, C) \) with a single source \( s \) and a single sink \( t \) can be constructed so that the elements in \( A(a, B, C) \) are in 1–1 correspondence with the \( st \)-paths. Indeed, \( G(a, B, C) \) is an NW-family [8] encoding \( A(a, B, C) \) with a simple perfect hash function \( h \). We briefly review these families in Section 4. The digraph \( G((7, 5)^2, B_L, C^4) \) associated to the phorma \( (7, 5)^2, B_L, C^4 \) is shown in Fig. 5. In this example, the set \( C \) of 4-compositions is the whole set \( C^4 \).

3. More applications of phormas

The need to impose order restrictions on arrays appears frequently and in many cases it is not difficult to express these restrictions as a phorma. For a larger example, consider the 7-phormas arising from the generation of T-shaped pieces. In Fig. 2 we show the
three kinds of such a piece. They are composed of a 3-block and a 3D L-piece. In the case of the $T_z$-piece, the L is truncated in one of its legs along the $z$-direction. These pieces are the 3D counterpart for the 2D L-shaped piece and they play an important role in 3D packing problems. They are described by seven parameters, which in the case of the $T_z$-piece are, $(x;X;y;Y;z;Z m;Z)$. To enumerate the $T$-pieces contained in a $(p \times q \times r)$-block was the motivating idea to formalize the notion of phorma. The need to effect this enumeration appears in [5,6].

As an example, for the $T_z$-piece, the restrictions coming from the geometry and the symmetry on the seven parameters $xXyYzZmZ = \{VT_1 \, VT_2 \, VT_3 \, VT_4 \, VT_5 \, VT_6 \, VT_7\}$ are of three types:

1. $(X \geq x); (Y \geq y); (Z \geq Z m \geq z)$;
2. $(X \geq Y); (X = Y) \Rightarrow (x \geq y)$;
3. $(x = X) \Rightarrow (z = Z m); (y = Y) \Rightarrow (x = X) \land (z = Z)$.

The first type of restrictions is obvious. The second type expresses the fact that the $T_z$-piece can be rotated around a vertical axis without modifying its containment properties. The $X$- and $Y$-directions are equivalent. Other axis of rotations, implying similar restrictions, could be used if the boxes to be packed into the $T_z$-piece could change its vertical. The third type of restrictions deals with the degenerated cases, in which the $T_z$-piece becomes a simpler piece. In terms of a phorma type boolean function, the restrictions translate as a boolean function $B_z^7$ with the following nine clauses:

$$B_z^7 = (x_2 \geq x_1) \land (x_4 \geq x_3) \land (x_7 \geq x_6) \land (x_6 \geq x_5) \land (x_2 \geq x_4)$$

$$\land((x_2 \neq x_4) \lor (x_1 \geq x_3)) \land ((x_1 \neq x_2) \lor (x_5 = x_6))$$

$$\land((x_3 \neq x_4) \lor (x_1 = x_2)) \land ((x_3 \neq x_4) \lor (x_5 = x_7)).$$

In this case, once more, $C$ is the whole set of 7-compositions $C^7$. If, just to be specific, $a=(15^217^219^3)$, then $|A(a,B_z^7,C^7)|=7,510,130$, while $15^217^219^3=446,006,475$. The amount of memory required to store the digraph $G(a,B_z^7,C^7)$ is logarithmically smaller than $|A(a,B_z^7,C^7)|$ (see Fig. 6) and its construction takes only a few seconds of computer time. Along the same line, we can derive boolean functions $B_x^7$ and $B_y^7$ for the other $T$-pieces $T_x$ and $T_y$ shown in Fig. 2. The three $T$-pieces are inequivalent.
under reflections and rotations which maintain the vertical direction. They play a complementary role in 3D packing problems in which the vertical direction of the boxes to be packed must be preserved.

We briefly mention another application of phorma: finding all the solutions for Cube It. Let \( x < y < z \) be real numbers. Consider the problem of finding all maximum packings of \((x \times y \times z)\)-bricks into a cube of side \( x + y + z \). If \( y + z < 3x \), then 27 is an upper bound on the number of bricks that can be packed, see [3]. There exists a phorma \((3^{81}, B_{\text{Cube}}^{3}, c_{27}^{3})\) of dimension 81 such that \( A(3^{81}, B_{\text{Cube}}^{3}, c_{27}^{3})\) has 1008 elements coinciding with the 1008 distinct solutions for the problem of packing the maximum of 27 boxes. In this case, \( a \) is the sequence of 81 repetitions of 3, \( a = 3333 \ldots 3 \) and \( C = \{c_{27}^{3}\} \), where \( c_{27}^{3} = (27, 27, 27) \). The expression for \( B_{\text{Cube}}^{3} \) and its justification are too long to be included in this paper. A higher dimensional analogue of \( B_{\text{Cube}}^{3} \) relates to an interesting open problem which is the subject of ongoing research: how to pack \( 5^5 = 3125 \) \((a \times b \times c \times d \times e)\)-boxes into a 5-cube of side \( a + b + c + d + e \). Our implementation (not yet optimized) of the phorma \((3^{81}, B_{\text{Cube}}^{3}, c_{27}^{3})\) found the 1008 solutions in about a day of computer time. What is interesting to mention, is that there are no symmetries in these 1008 solutions. So, their set can be partitioned into 21 classes of 48 elements each, corresponding to the symmetry group of the cube. Representatives of these 21 classes are given in Fig. 3. The bricks orientations are

<table>
<thead>
<tr>
<th>sol 1</th>
<th>sol 2</th>
<th>sol 3</th>
<th>sol 4</th>
<th>sol 5</th>
<th>sol 6</th>
<th>sol 7</th>
<th>sol 8</th>
<th>sol 9</th>
<th>sol 10</th>
<th>sol 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acC</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
</tr>
<tr>
<td>bCA</td>
<td>bCe</td>
<td>bCe</td>
<td>bCe</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
</tr>
<tr>
<td>ABb</td>
<td>AAb</td>
<td>AAb</td>
<td>AAb</td>
<td>Bb</td>
<td>Bb</td>
<td>Bb</td>
<td>Bb</td>
<td>Bb</td>
<td>Bb</td>
<td>Bb</td>
</tr>
<tr>
<td>CBb</td>
<td>CBb</td>
<td>CBb</td>
<td>CBb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
</tr>
<tr>
<td>aAC</td>
<td>ACa</td>
<td>ACa</td>
<td>ACa</td>
<td>aC</td>
<td>aC</td>
<td>aC</td>
<td>aC</td>
<td>aC</td>
<td>aC</td>
<td>aC</td>
</tr>
<tr>
<td>BAc</td>
<td>BCa</td>
<td>BCa</td>
<td>BCa</td>
<td>BC</td>
<td>BC</td>
<td>BC</td>
<td>BC</td>
<td>BC</td>
<td>BC</td>
<td>BC</td>
</tr>
<tr>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
<td>Cb</td>
</tr>
<tr>
<td>CAa</td>
<td>CAa</td>
<td>CAa</td>
<td>CAa</td>
<td>CB</td>
<td>CB</td>
<td>CB</td>
<td>CB</td>
<td>CB</td>
<td>CB</td>
<td>CB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sol 12</th>
<th>sol 13</th>
<th>sol 14</th>
<th>sol 15</th>
<th>sol 16</th>
<th>sol 17</th>
<th>sol 18</th>
<th>sol 19</th>
<th>sol 20</th>
<th>sol 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
<td>acB</td>
</tr>
<tr>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
<td>bC</td>
</tr>
<tr>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
<td>BAa</td>
</tr>
<tr>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
<td>CBd</td>
</tr>
<tr>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
<td>aAB</td>
</tr>
<tr>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
<td>ACA</td>
</tr>
<tr>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
<td>BAc</td>
</tr>
<tr>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
<td>CaA</td>
</tr>
<tr>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
<td>CAC</td>
</tr>
</tbody>
</table>

Fig. 3. All the solutions for cube it.
given by the conventions: \( a \mapsto yzx; A \mapsto zyx; b \mapsto xzy; B \mapsto zxy; c \mapsto xyz; C \mapsto yxz. \) The parameters of \( G(3^8, B_{10}^3, c_2^3) \) are listed in Fig. 6. In particular, \( H(3^8, B_{10}^3, c_2^3) \) has only 4 vertices and the whole difficulty is to find \([A(3^8, B_{10}^3, c_2^3)]\) which in this case coincides with \( A(3^8, B_{10}^3, c_2^3). \)

4. NW-families

The following concept, introduced in [8], is the central tool for our hashing scheme. A Nijenhuis–Wilf combinatorial family, or simply an NW-family, is an acyclic digraph \( G \) whose vertex set is denoted by \( V(G) \), having the properties below:

1. \( V(G) \) has a partial order (for \( x, y \in V(G), y \preceq x \) if there is a directed path from \( x \) to \( y \)) with a unique minimal element \( t \). For each \( v \in V(G) \) the set \( \{ x \in V(G) \mid x \preceq v \} \) is finite and includes \( t \).
2. Every vertex \( v \), except \( t \) has a strictly positive out-valence \( \rho(v) \). For each \( v \in V(G) \), the set \( E(v) \) of outgoing edges has a \( v \)-local rank-label \( \ell_v \), \( 0 \leq \ell_v(e) \leq \rho(v) - 1, e \in E(v) \).

A path starting at \( v \) and ending in \( t \) is encoded by the sequence of label-ranks of the sequence of its edges. Such a path is called an object of \( v \) [8]. The beauty of this scheme is that we can perform various tasks on the family in an abstract way, without referring to the actual encoding/decoding of the objects as paths. An NW-family is especially suited to deal with the following 5 tasks. Tasks 1–4 are from [8]. Task 0 is emphasized here because of its applicability to the phorma: we need to calibrate the cardinality of \( A(a, B, C) \) by choosing \( a \) in an adequate way.

**Task 0. Counting:** What is the family’s cardinality? Algorithm: Given \( v \in V(G) \), let \( |v| = \sum |\text{head}(e)| e \in E(G), \text{tail}(e) = v \}. \) From this formula, \( |v| \) is easily obtained by recursion. It is convenient to store it as an attribute of \( v \in V(G) \) in a pre-processing phase, or compilation time.

**Task 1. Sequencing:** Given an object in the family, construct the “next” object. Algorithm: The next path of a given path \( \pi \) in coded form is, in coded form, the lexicographic successor of \( \pi \).

**Task 2. Ranking (perfect hashing):** Given an object \( \omega \) in the family, find the integer \( h(\omega) \) such that \( \omega \) is the \( h(\omega) \)th element in the order induced by Task 1. Algorithm: Let an element-path \( \pi \) of order \( v \) of an NW-family, \( \pi = (e_1, e_2, \ldots, e_p) \) be given. The rank of \( \pi \) is defined as \( h(\pi) = \sum_{i=1}^{p} \chi(e_i) \), where \( \chi(e) = \sum \{|\text{head}(f)| \mid f \in E(v) \}, f < \ell_v(e) \).

**Task 3. Unranking:** Given an integer \( r \), we need to construct the \( r \)th path from \( v \) to \( t \). Define \( \text{pred}_v(e) \) as the highest-rank edge of the set \( \{ f \in E(v) \mid \ell_v(f) < \ell_v(e) \} \), and let \( |\text{head}(\text{pred}_v(e))| = 0 \) if this set is empty. The required \( r \)th path \( \pi_r \) is generated as follows. Algorithm: \( \pi_r = \emptyset; r' \leftarrow 0; v' \leftarrow v; \text{repeat} \) append to \( \pi_r \), the highest-rank edge \( e \) of \( E(v') \) such that \( r' + |\text{head}(\text{pred}_v(e))| \leq r \); \( r' \leftarrow r' + |\text{head}(\text{pred}_v(e))|; v' \leftarrow \text{head}(e) \) until \( v' = t \).
Task 4. Getting random object: Choose an object uniformly at random from the given family. Algorithm: Let $\xi \in [0,1]$ be uniformly chosen at random; return the $(\lfloor m \cdot \xi \rfloor)$th object.

5. Reducing, sorting, a-roofing: the digraph $G(a, B, C)$

If $\alpha$ has $m \leq n$ distinct entries, let $M_{\alpha} = \{1, \ldots, m\}$. The reduction of $\alpha$, denoted by $|\alpha|$, is the unique surjection in $(M_{\alpha})^N$ which is order compatible with $\alpha$. That is, for $i \in N$, if $\alpha_i$ is the $i$th smallest entry in $\alpha$, then $|\alpha|_i = i$. Let also $\hat{\alpha}^\gamma$ denote the $m$-sequence of distinct entries of $\alpha$ in ascending order. We call $\hat{\alpha}^\gamma$ the sorting of $\alpha$. Given an ascending $m$-sequence $\gamma$, let $m_\gamma = m$.

**Proposition 1.** The $n$-vector of positive integers $\alpha$ is recoverable from $(|\alpha|, \hat{\alpha}^\gamma)$.

**Proof.** It is sufficient to observe that $\alpha_i = \hat{\alpha}^{|\alpha|}_i$. \qed

Since $\alpha$ induces the pair $(|\alpha|, \hat{\alpha}^\gamma)$ and, by Proposition 1, is recoverable from, it we can think of $\alpha$ as the pair $(|\alpha|, \hat{\alpha}^\gamma)$ and write $\alpha \equiv (|\alpha|, \hat{\alpha}^\gamma)$.

For $\alpha \in A(a, B, C)$ let the a-roof of $\alpha$ be $|\alpha|^a = \gamma^*$ where $\gamma^*$ is the lexicographically maximal increasing $m$-sequence with the property that $(|\alpha|, \gamma^*) \in A(a, B, C)$. In particular, $\hat{\alpha}^\gamma |\alpha|_i \leq |\alpha|^a_i = \gamma^*_i$, $i \in N$.

**Proposition 2.** The a-roof of $\alpha$, $|\alpha|^a$, does not depend on $\alpha$ itself but only on $|\alpha|$ and $a$, in the sense that $|\alpha|^a = [|\alpha|]^a$.

**Proof.** The a-roof $|\alpha|^a = \gamma^*_1^{\gamma^*_2} \cdots \gamma^*_m$ can be constructed as follows. Suppose that, for $1 \leq i \leq m$, $i$ occurs at positions $p_{i1}, \ldots, p_{ij}$ of $|\alpha|$. Then we must have $\gamma^*_m = \min\{a_{pi1}, a_{pi2}, \ldots, a_{p_{ij}}\}$, due to $a$-dominance. For $i = m - 1, m - 2, \ldots, 1$, the definition implies that $\gamma^*_i = \min\{a_{pi1}, \ldots, a_{pij}, \gamma^*_i+1\}$, by $a$-dominance and to insure the strict increase of $\gamma^*$. Since the construction only depended on $|\alpha|$ and $a$, the Proposition is proved. \qed

Given a phorma $(a, B, C)$ and the corresponding $A(a, B, C)$, three sets are defined

(i) $[A(a, B, C)] = \{|\alpha| \mid \alpha \in A(a, B, C)\}$,
(ii) $A^\hat{\alpha}(a, B, C) = \{\hat{\alpha}^\gamma \mid \alpha \in A(a, B, C)\}$,
(iii) $[A(a, B, C)]^a = \{|\alpha|^a \mid \alpha \in A(a, B, C)\}$.

Usually, but not necessarily (see the phorma $(381, B_{37}^{\text{Cube}}, c_3)$), $|[A(a, B, C)]|$ is much smaller than $|A(a, B, C)|$. By Proposition 2, $|[A(a, B, C)]^a| \leq |[A(a, B, C)]|$. In general this inequality is also not tight. See examples in Fig. 6. The perfect hash function that
is constructed for \(A(a, B, C)\) depends on the explicit enumeration of the set \([A(a, B, C)]\).

This set, in the case of our ongoing example, has nine elements,

\[
[A((7, 5)^2, B_L, C^4)] = \{1111, 2121, 2211, 3211, 3221, 4231, 4312, 4321\}.
\]

The \(a\)-roof set has only seven elements because of two duplicates

\[
[A((7, 5)^2, B_L, C^4)]^a = \{5, 57, 45, 457, 457, 3457, 3457, 3457\}.
\]

Given a phorma \((a, B, C)\) the digraph \(A(a, B, C)\) is defined as follows. Its vertex set is \(V(A(a, B, C)) = \{s\} \cup [A(a, B, C)] \cup [A(a, B, C)]^a\), where \(s\) is a single source. It is a simple graph, and so, each of its directed edges can be represented by an ordered pair of vertices. For each \([x] \in [A(a, B, C)]\) there are edges \((s, [x])\) and \(([x], [x]^a)\). These are all the edges of \(A(a, B, C)\), concluding its definition. The digraph \(A(a, B, C)\) is a subgraph of \(G(a, B, C)\). In Fig. 5, the edges of \(A((7, 5)^2, B_L, C^4)\) are depicted in dashed gray. The edges of its complement \(H((7, 5)^2, B_L, C^4)\) in \(G((7, 5)^2, B_L, C^4)\) (which we define next) are depicted in solid lines. The number near a vertex \(v\) (the first number, when there are two) is the number of \(v\)-paths in \(G((7, 5)^2, B_L, C^4)\).

Let \(H^\infty\) be the set of all finite strictly increasing sequences of positive integers. The empty sequence is in \(H^\infty\) and is denoted by \(t\). Suppose \(\gamma = \gamma_1 \gamma_2 \ldots \gamma_m \in H^\infty\). We define an NW-family \(H_{\gamma}\), as follows. If \(\gamma_m > m\) let \(\gamma_m \leftarrow \gamma\) denote the increasing sequence of length \(m\) satisfying \(\gamma_m = \gamma_{m-1} - 1\) and \(\gamma' = \min\{\gamma_i - 1, \gamma_i\}\), for \(i = m, m-1, \ldots, 1\). If \(\gamma_m = m\), then \(\gamma\) does not exist. If \(\gamma \neq t\), let \(\gamma'\) be the sequence of length \(m - 1\) obtained from \(\gamma\) by removing its last entry: \(\gamma' = \gamma_1 \ldots \gamma_{m-1}\). If \(\gamma = t\), then \(\gamma'\) does not exist. Given \(\gamma, \gamma' \in H^\infty\), we say that \(\gamma \leq \gamma'\) if there is a sequence \((\gamma_i') = \gamma_1', \gamma_2', \ldots, \gamma_p' = \gamma)\), with \(\gamma' \in H^\infty\), such that, for each \(i = 1, 2, \ldots, p-1\), either \(\gamma_i' = \gamma_i \) or else \(\gamma_i' = \gamma_i + 1\). The relation \(\leq\) makes \(H^\infty\) a partially ordered set, or poset. For \(\gamma \in H^\infty\), let \(H_{\gamma}\) be the acyclic digraph whose vertex set is \(V(H_{\gamma}) = \{\gamma' \mid \gamma' \leq \gamma\}\). From each vertex \(\gamma' \in V(H_{\gamma})\) there are at most two outgoing edges: \((\gamma', \gamma'')\), of \(\gamma'\)-local rank-label 0, if \(\gamma''\) exists and \((\gamma', \gamma''')\), if \(\gamma'''\) exists. The \(\gamma'\)-local rank-label of this last edge is either 1, if \(\gamma''\) exists or 0 otherwise. This concludes the definition of \(H_{\gamma}\).

Given a path \(\pi\) from \(\gamma^*\) to \(t\) in \(H_{\gamma^*}\), a fall of \(\pi\) is a vertex \(\gamma\) such that the edge \((\gamma, \gamma'\)) is used by \(\pi\). Path \(\pi\) has exactly \(m_{\gamma^*}\) falls. In Fig. 4 the 4 falls of the path shown in thick edges are: 5678, 567, 34 and 3. The encoding/decoding of the increasing sequences \(\gamma \leq \gamma^*\) as paths in the NW-family \(H_{\gamma^*}\) is particularly simple:

**Proposition 3.** To a path \(\pi\) in \(H_{\gamma^*}\) from \(\gamma^*\) to \(t\) corresponds \(\gamma_{\pi} \leq \gamma^*\) consisting of the last coordinates of the \(\pi\) falls (in reverse order). Reciprocally, to \(\gamma \leq \gamma^*\), corresponds the unique path \(\pi_{\gamma}\) from \(\gamma^*\) to \(t\) such that the last entry of its \(i\)th fall coincides with the \(i\)th entry of \(\gamma\). Moreover, \(\pi_{\gamma_{\pi}} = \pi\), \(\gamma_{\pi_{\pi}} = \gamma\).
**Proof.** Straightforward from the definitions.

Given a path \( \pi \) from \( \gamma^* \) to \( t \) in \( H_{\gamma^*} \), a post-fall of \( \pi \) is a vertex \( \gamma' \) such that the edge \((\gamma, \gamma')\) is used by \( \pi \). Path \( \pi \) has at most \( m_{\gamma^*} \) falls. The set of post-falls of \( \pi \) is denoted \( \text{PostFall}(\pi) \). In Fig. 4, calling \( \pi \) the path shown in thick edges, we have \( \text{PostFall}(\pi) = \{4567, 456, 23, 2\} \) and their members are depicted as white vertices. The hash function \( h_{\gamma^*} \) in the NW-family \( H_{\gamma^*} \) takes a simple form:

**Proposition 4.** The perfect hash function associated with the NW-family \( H_{\gamma^*} \) is \( h_{\gamma^*}(\gamma) = \sum_{\gamma' \in \text{PostFall}(\pi)} |\gamma'| \).

**Proof.** The result is an specialization of the rank function of a generic NW-family to \( H_{\gamma^*} \). It follows directly from the definitions.

From this proposition it follows in Fig. 4 that \( h_{5679}(3478) = 35 + 20 + 3 + 2 = 60 \). The terms of the sum correspond to the orders of the white vertices, forming the set \( \text{PostFall}(\pi_{3478}) \).

Define \( H(a, B, C) = \bigcup\{H_{\gamma^*} \mid \gamma^* \in [A(a, B, C)]^a \} \). Actually, in this union we need only to take maximal \( \gamma^* \)'s. If \( \gamma' \preceq \gamma^* \), then \( H_{\gamma'} \) is a subgraph of \( H_{\gamma^*} \) and it is irrelevant for the union. The digraph \( H(7575, B_4, C_4) \) shown in Fig. 5, is formed by the union of 4 maximal \( \gamma^* \)'s: \( H_{3457} \cup H_{4567} \cup H_{457} \cup H_{57} \). In general, the *digraph of a phorma* \( P = (a, B, C) \) is defined as \( G(a, B, C) = \Lambda(a, B, C) \cup H(a, B, C) \).

In order to make \( G(a, B, C) \) an NW-family, we need to define the \( v \)-local rank labels of the \( v \)-outgoing edges for each vertex \( v \) of \( G(a, B, C) \). This can be accomplished by ordering lexicographically the elements of \( [A(a, B, C)] \) and ranking them in the ascending order: \( 0, 1, \ldots, ([A(a, B, C)] - 1 \). The edge \( (s, [x]) \) gets as \( s \)-local rank the same rank as \( [x] \). An edge of type \( ([x], [x]^a) \) gets \( [x] \)-local rank 0, because it is the unique \( [x] \)-outgoing edge. For \( \gamma \in V(H(a, B, C)) \) we have already defined the \( \gamma \)-local label-ranks. With these local ranks the two conditions of NW-family are
satisfied by \( G(a,B,C) \). It remains to verify that its \( st \)-paths encode the elements of \( A(a,B,C) \):

**Theorem 1** (Main Theorem). For every phorma \( P=(a,B,C) \) the \( st \)-paths of \( G(a,B,C) \) are in 1–1 correspondence with the elements of \( A(a,B,C) \).

**Proof.** Given an \( \alpha \in A(a,B,C) \), let \( [\alpha]^* = \gamma^* \). Define \( \pi_{\alpha} = (s, [\alpha]) \circ ([\alpha], [\alpha]^*) \circ \pi_{\alpha^*} \). Reciprocally, given an \( st \)-path \( \pi \) in \( G(a,B,C) \), let \( \beta \) be the second vertex of \( \pi \), \( \gamma^* \) be its third vertex and \( \gamma \) be such that \( \pi = (s, \beta) \circ (\beta, \gamma^*) \circ \pi_{\gamma} \). Define \( \pi_{\pi} \equiv (\beta, \gamma) \). These definitions imply \( \pi_{\alpha} = \pi \) and \( \alpha_{\pi_{\alpha}} = \alpha \). \( \square \)

Given \( [A(a,B,C)] \) ordered lexicographically and \( \beta \in [A(a,B,C)] \) define \( \|\beta\| = \sum \{|\beta'| \text{ such that } \beta' < \beta\} \). In Fig. 5 the values of \( \|\beta\| \) appear as the second number near each vertex \( \beta \). The hash function \( h \) for a phorma assumes a particularly simple expression:

**Proposition 5.** Given \( \alpha \equiv ([\alpha], \hat{\alpha}^*) \in A(a,B,C) \), the perfect hash function \( h \) of \( G(a,B,C) \) is

\[
h(\alpha) = \|\alpha\| + h_{[\alpha]}(\hat{\alpha}^*).\]

**Proof.** This value of \( h(\alpha) \) follows from the general algorithm for ranking in an abstract NW-family, when specialized to phormas. \( \square \)
6. Implementation aspects

The need of the boolean function $B$ in a phorma $(a,B,C)$ is just to enable the enumeration of $|A(a,B,C)|$. If the size of this set is small, then an explicit list of its elements, $\{\beta^1,\beta^2,\ldots,\beta^m\}$, can be given in place of $B$. If this is not the case, then a convenient way to input a generic phorma type boolean function is by means of a tree $T(B)$ with three types of internal nodes: $\lor$-nodes, $\land$-nodes, $\neg$-nodes. The leaves of the tree correspond to the basic constituent boolean functions of type $\alpha_i \cdot \alpha_j$, where $\star \in \{\leq, \geq, <, >, =, \neq\}$. The $\neg$-nodes (negation operator) must have at most one child. Note that each subtree rooted at an internal $\lor$-node $v$ ($\lor \in \{\lor, \land, \neg\}$) is a boolean tree obtained by taking the $\lor$-operation of the boolean tree(s) corresponding to the children of $v$. Given an $\alpha$, it is possible to decide its B-satisfiability, by evaluating from the leaves up and arriving to the root of $T(B)$. See [1] for more details.

We also admit two ways of inputting $C$: by means of an explicit list of its elements, $\{\delta^1,\delta^2,\ldots,\delta^n\}$, if $z = |C|$ is small, or by a phorma type of boolean restrictions on the coordinates of the $\delta$'s. In this case, $C$ is itself a boolean expression with clauses of type $\langle \delta_i \rangle$. In the case $C = C_n^\infty$, this boolean expression is empty. We define an NW-family encoding $\bigcup_{n \in \mathbb{N}} \{C^n\}$: consider the digraph $L^\infty$, whose vertex set is the set of points in the plane which have positive integer coordinates. There are at most two edges from a point $(p,q) \in V(L^\infty)$, namely a west edge $((p,q),(p-1,q))$, if $p \geq 2$, and a southwest edge $((p,q),(p-1,q-1))$, if $p,q \geq 2$. The $(p,q)$-local rank-label of the first edge is 0, if it exists, and the $(p,q)$-local rank-label of the second edge is 1, if both edges exist. In the case that only the second edge exists, then the $(p,q)$-local rank-label is 0. Let $C^n_m$ be the subset of $C^n$ of $n$-compositions which have length $m$.

Proposition 6. The paths from $(n,m)$ to $(1,1)$ in $L^\infty$ are in 1–1 correspondence with the elements of $C^n_m$. Thus $L^\infty$ is an NW-family encoding the $n$-compositions for all $n \in \mathbb{N}$.

Proof. Let $\delta \in C^n_m$ be given. Construct a path $\pi_\delta$ from $(n,m)$ to $(0,0)$ in $L^\infty$ as follows. Let $\delta' \leftarrow \delta$ and $\pi' \leftarrow$ the empty path. Repeat $n$ times: if $\delta'_1 > 1$, then $\delta'_1 \leftarrow \delta'_1 - 1$, extend $\pi'$ with a west edge; if $\delta'_1 = 1$, then $\delta'$ becomes $\delta'$ without its first part; extend $\pi'$ with a southwest edge. After the $n$ iterations of this loop, $\delta'$ is the composition 1 of 1 in 1 part and define $\pi_\delta = \pi'$. Reciprocally, given a path $\pi$ from $(n,m)$ to $(1,1)$ in $L^\infty$, construct a $\delta_\pi \in C^n_m$ as follows. Let $\delta' \leftarrow 1$ and $\pi' \leftarrow \pi$. For $i = 1,2, \ldots, n$ do: if the $i$th edge of $\pi$ is a southwest edge, $\delta' \leftarrow (1,\delta')$; if the $i$th edge of $\pi$ is a west edge, let $\delta'_1 \leftarrow \delta'_1 + 1$. Define $\delta_\pi = \delta'$. These definitions imply $\delta_\pi = \delta$ and that $\pi_\delta = \pi$, establishing a 1–1 correspondence between $C^n_m$ and the paths from $(n,m)$ to $(1,1)$ in $L^\infty$.

By using Proposition 6 it is possible to generate in an efficient way the $\delta$'s satisfying the boolean expression $C$ via a $C$-restricted implicit enumeration based on $L^\infty$.

The crucial task to construct (at compiler time) the digraph $G(a,B,C)$ is to explicitly generate $|A(a,B,C)|$. Since $\alpha$ and $|\alpha|$ are order isomorphic, one possibility to produce $|A(a,B,C)|$ is to generate all the $n^n$ members of $N^N$ and to test each such sequence
for reducibility and $B$-satisfiability [2]. This simple minded approach is suitable for small dimension $n$. In our four-dimensional phorma $((7, 5)^2, B_L, C^4)$ there are only 256 tests to be made. When $n$ increases this simple minded method becomes inapplicable. For example, for the 7-phorma $(a, B_7, C^7)$ there are $7^7 = 823,543$ tests to be made and a better approach is needed to generate the 1134 elements of $[A(a, B'_7, C^7)]$ as well as the 20 elements of $[A(a, B, C^7)]^a$ (for $a = (15^2 17^2 19^3)$). The basic idea is to implement a $B$-restricted implicit enumerating scheme which takes into account only reduced sequences in generating the set $(A(a, B, C))$. This methodology extends substantially the realm of the phorma applicability.

Given a phorma $(a, B, C)$ and $\delta \in C$. Define

$$[A(a, B, C, \delta)] = \{x \in [A(a, B, C)] \mid \bar{x} = \bar{\delta}\}.$$  

As we know how to generate $\delta \in C$, the generation of $[A(a, B, C)]$ reduces to the generation of each $[A(a, B, C, \delta)]$, because ($\cup$ means disjoint union)

$$[A(a, B, C)] = \bigcup_{\delta \in C} [A(a, B, C, \delta)].$$

The $m$-dimensional grid digraph $J^m$ is the digraph whose vertices are the points of $\mathbb{R}^m$ with integer coordinates. There is an edge from $p = p_1 \ldots p_j \ldots p_m$ to $q = q_1 \ldots q_j \ldots q_m$ if $p_i = q_i$ except for $i = j$, where $p_j = q_j + 1$.

**Proposition 7.** An element of $[A(a, B, C, \delta)]$ corresponds to a path from the point $\delta$ to the origin in $J^m$.

**Proof.** Given $\beta = \beta_1 \beta_2 \ldots \beta_m \in [A(a, B, C, \delta)]$ we define a path named $\pi_\beta$ in digraph $J^m$ from $\delta$ to the origin as follows. Path $\pi_\beta$ starts at $\delta$ and its $i$th edge is the edge parallel to the $\beta_i$th axis. It follows from the definitions that $\pi_\beta$ finishes at the origin. \(\Box\)

From Proposition 7 a $B$-restricted implicit enumeration scheme based on paths in $J^m$, only produces reduced words. The construction of $[A(a, B, C)]$, $[A(a, B, C)]^a$, and as a consequence, the construction of the digraph $A(a, B, C)$ are efficiently performed in this way.

Now we turn our attention to the construction and storage of the digraph $H(a, B, C)$. Let $\mathcal{L}(r, m) = \{\gamma \in V(H(a, B, C)) \mid \gamma \in (\mathbb{N}^\ast)^m, \, \gamma_m = r\}$ and $[A(a, B, C)]^{a}_{\text{max}} = \{\gamma^* \in [A(a, B, C)]^a \mid \gamma^* \text{ maximal}\}$.

**Proposition 8.** $|\mathcal{L}(r, m)| \leq |[A(a, B, C)]^{a}_{\text{max}}|$.

**Proof.** For each element $\gamma \in \mathcal{L}(r, m)$ choose some $\gamma^* \in [A(a, B, C)]^{a}_{\text{max}}$ such that $\gamma \preceq \gamma^*$. This defines a function $f$ from $\mathcal{L}(r, m)$ to $[A(a, B, C)]^{a}_{\text{max}}$, given by $f(\gamma) = \gamma^*$. It is enough to prove that $f$ is injective. Let $\gamma$ and $\gamma'$ be distinct elements of $\mathcal{L}(r, m)$. Note that $\prec \gamma \neq \prec \gamma'$. Suppose that $\bar{\gamma}$ and $\bar{\gamma}'$ are such that $\gamma \preceq \bar{\gamma}$ and $\gamma' \preceq \bar{\gamma}'$. Then it follows that $\bar{\gamma} \neq \bar{\gamma}'$ because the first $m - 1$ entries of $\bar{\gamma}$ form $\prec \gamma$ and the first $m - 1$ entries of $\bar{\gamma}'$ form $\prec \gamma'$. So, $f$ is injective. \(\Box\)
Let \( a_* = \max \{ a_i \} \), \( n_* = \max \{ m \mid \exists \delta \in C \text{ with } m_\delta = m \} \) and \( v \) the number of non-empty \( \mathcal{D}(r, m) \)'s.

Proposition 9. \(|V(H(a, B, C))| \leq 1 + \left| \left[ A(a, B, C) \right]^{a_*}_{\max} \right| (a_* - (n_* - 1)/2) n_* \).

Proof. Clearly, \( v \leq (a_* - (n_* - 1)/2) n_* \). The term 1 is for the sink \( t \). The inequality follows from Proposition 8. \( \square \)

The bound given in Proposition 9 is not tight. In general, the maximum value of \(|\mathcal{D}(r, m)|\), \( \lambda \), tends to be much smaller than \(|\left[ A(a, B, C) \right]^{a_*}_{\max} \)|. A more informative parameter related to the size of \( H(a, B, C) \) is \( \mu \) defined as

\[
\mu = |V(H(a, B, C))|/v.
\]

For phormas arising in the realm of the applications that we have explored, \( \mu \) is rather small. Given a vertex \( \gamma \) of this digraph, \( \gamma \) and \( \overleftarrow{\gamma} \) are easily obtainable. So the edges of \( H(a, B, C) \) do not need to be stored. Each one of the \( v \) \( \mathcal{D}(r, m) \)'s is kept as a lexicographically ordered list indexed by an \((a_* \times n_*)\)-array. The \((r, m)\)-entry of this array is a pointer to the list \( \mathcal{D}(r, m) \). A binary search can then be used to locate a specific member of \( \mathcal{D}(r, m) \), when computing \( h \) and \( h^{-1} \).

The amount of work needed to compute \( h(\gamma) \) is basically proportional to \( m_\gamma \), the length of \( \gamma \). Indeed, from Proposition 4 we need only to find the \( m \) elements of the set \( PostFall(\pi_\gamma) \) and add their orders. These orders are stored at the construction of \( \gamma \). This makes the time for computing \( h(\alpha) \) independent of \( a_* \).

Fig. 6 displays basic parameters of various phormas. The following shortcuts are used: \( v_G = |V(G(a, B, C))| \), \( v_H = |V(H(a, B, C))| \), \( \alpha^d = \left| \left[ A(a, B, C) \right]^{a^d}_{\max} \right| \), \( \alpha^a = \left| \left[ A(a, B, C) \right]^{a}_{\max} \right| \). The last column of Fig. 6 is \( 10^4 \times d \), with \( d = |A(a, B, C)|/ (\prod_{i \in \mathbb{N}} a_i) \) the density of \( (a, B, C) \). It is interesting to observe how fast the densities of the symmetric phormas (the ones with \( B = B^{sym}_{sym} \)) go to zero as \( n \) increases. We present parameters for the phormas \( (9^n, B^{sym}_{sym}, C^n) \), \( 2 \leq n \leq 9 \). The boolean functions \( B^{sym}_{sym} \) for these phormas are obtained from \( B^{sym}_{sym} \) by replacing the inequalities \( \triangleright \) by the strict inequalities \( \triangleright \). Thus, only strictly decreasing sequences are permitted. Note that \( A(9^{10}, B^{sym}_{sym}, C^{10}) = \emptyset \).

7. Conclusion

We have defined a data structure generator which permits the perfect hash of order restricted multidimensional arrays \( A(a, B, C) \). The restrictions accord a general type of boolean functions \( B \) formed by order restricting pairs of entries of the array in arbitrary ways. The boolean function \( B \) is used in forming a reduced set \([A(a, B, C)]\), inducing a partition of \( A(a, B, C) \). An \([\alpha] \in [A(a, B, C)]\) corresponds to a member subset \([A(a, B, C), [\alpha]]\) of this partition. The elements of \([A(a, B, C, \delta)]\), \( \delta \in C \), are in 1–1 correspondence with paths from \( \delta \) to the origin in the \( m_\delta \)-dimensional integer grid digraph \( J^{m_\delta} \), and can be efficiently found in a \( B \)-restricted implicit enumeration scheme which produces only reduced sequences. To generate all \( c \in C \), which might be itself a boolean function on the \( \delta \)'s, we use the NW-family \( R^{\infty} \) in a \( C \)-restricted implicit
Fig. 6. Parameters values for $G(a,B,C)$ of various phormas.

| Phorma | $v_G$ | $v_H$ | $|A||$ | $|A|$ | $\alpha^a$ | $\alpha^{a\max}$ | $\lambda$ | $\mu$ | $10^4d$ |
|--------|--------|--------|--------|------|-------|-------------|--------|------|--------|
| $9^2B_{B_1}^{C_2}$ | 20/19 | 17/17 | 2/1 | 45/35 | 2/1 | 1/1 | 1.0000 | 5556/4444 |
| $9^3B_{B_2}^{C_3}$ | 29/24 | 24/22 | 4/1 | 165/84 | 3/1 | 2/1 | 1.0000 | 2263/1152 |
| $9^4B_{B_3}^{C_4}$ | 39/27 | 30/25 | 8/1 | 495/126 | 4/1 | 3/1 | 1.0000 | 754/192 |
| $9^5B_{B_4}^{C_5}$ | 52/28 | 35/26 | 16/1 | 1287/126 | 5/1 | 4/1 | 1.0000 | 218/21 |
| $9^6B_{B_5}^{C_6}$ | 72/27 | 39/25 | 32/1 | 3003/84 | 6/1 | 5/1 | 1.0000 | 57/1.58 |
| $9^7B_{B_7}^{C_7}$ | 107/24 | 42/22 | 64/1 | 6435/36 | 7/1 | 6/1 | 1.0000 | 13/0.0752 |
| $9^8B_{B_8}^{C_8}$ | 173/19 | 44/17 | 128/1 | 12870/9 | 8/1 | 7/1 | 1.0000 | 3/0.0030 |
| $9^9B_{B_9}^{C_9}$ | 302/12 | 45/10 | 256/1 | 24310/1 | 9/1 | 8/1 | 1.0000 | 0.6/10^{-5} |
| $9^{10}B_{B_{10}}^{C_{10}}$ | 557 | 45 | 511 | 43758 | 9 | 8 | 1.0000 | 0.1 |
| $(7,5)^2B_{B_C}$ | 32 | 22 | 9 | 190 | 7 | 4 | 2 | 1.0476 | 1551 |
| $(40,30)^2B_{B_C}$ | 164 | 154 | 9 | 245670 | 7 | 4 | 2 | 1.0621 | 1706 |
| $(50,40)^2B_{B_C}$ | 204 | 194 | 9 | 652910 | 7 | 4 | 2 | 1.0486 | 1632 |
| $(60,50)^2B_{B_C}$ | 244 | 234 | 9 | 1420325 | 7 | 4 | 2 | 1.0400 | 1578 |
| $(99,50)^2B_{B_C}$ | 400 | 390 | 9 | 5196500 | 7 | 4 | 2 | 1.1404 | 2121 |
| $(100,50)^2B_{B_C}$ | 404 | 394 | 9 | 5317825 | 7 | 4 | 2 | 1.1420 | 2127 |
| $10^2B_{B_C}$ | 1184 | 49 | 1134 | 237325 | 7 | 6 | 1 | 1.0000 | 237 |
| $15^2B_{B_C}$ | 1219 | 84 | 1134 | 3853200 | 7 | 6 | 1 | 1.0000 | 226 |
| $20^2B_{B_C}$ | 1254 | 119 | 1134 | 28226800 | 7 | 6 | 1 | 1.0000 | 221 |
| $25^2B_{B_C}$ | 1289 | 154 | 1134 | 132916875 | 7 | 6 | 1 | 1.0000 | 218 |
| $30^2B_{B_C}$ | 1324 | 189 | 1134 | 472460925 | 7 | 6 | 1 | 1.0000 | 216 |
| $15^217^2B_{B_C}$ | 1262 | 127 | 1134 | 7510130 | 20 | 13 | 3 | 1.1651 | 168 |
| $25^227^29B_{B_C}$ | 1332 | 197 | 1134 | 204089675 | 20 | 13 | 3 | 1.1006 | 184 |
| $10^250^212^2B_{B_C}$ | 1201 | 66 | 1134 | 300270 | 17 | 10 | 2 | 1.0645 | 9 |
| $3^8P_{B_C}^H$ | 1013 | 4 | 1008 | 1008 | 1 | 1 | 1 | 1.0000 | 10^{-32} |

The whole scheme is summarized by two facts: (i) an $z \in A(a,B,C)$ induces three pieces of information, $[z]$, $\hat{z}^a$ and $[z]^a$ and is recoverable from the first two, $z \equiv ([z], \hat{z}^a)$; (ii) this decomposition reflects in the rank formula for a perfect hashing of $A(a,B,C)$: $h(z) = ||z|| + h([z]) \times (\hat{z}^a)$. This encoding scheme has the power of perfectly addressing huge and quite intricate arrays $A(a,B,C)$ by means of the logarithmically smaller NW-family $G(a,B,C)$. This general type of perfect hash scheme does
not seem to have been treated before in the literature. In particular, its use in database systems is a possible source of relevant applications and remains to be investigated.

Acknowledgements

The authors thank A. Bondy for bringing reference [3] to their attention. They also thank three anonymous referees for helpful comments improving the legibility of the paper. The financial support of CNPq, (contract no. 30.1103/80) in the case of the second author, is acknowledged.

References