

# POLYNOMIAL IDENTITIES FOR TERNARY INTERMOLECULAR RECOMBINATION

MURRAY R. BREMNER

ABSTRACT. The operation of binary intermolecular recombination, originating in the theory of DNA computing, permits a natural generalization to  $n$ -ary operations which perform simultaneous recombination of  $n$  molecules. In the case  $n = 3$ , we use computer algebra to determine the polynomial identities of degree  $\leq 9$  satisfied by this trilinear nonassociative operation. Our approach requires computing a basis for the nullspace of a large integer matrix, and for this we compare two methods: (i) the row canonical form, and (ii) the Hermite normal form with lattice basis reduction. In the conclusion, we formulate some conjectures for the general case of  $n$ -ary intermolecular recombination.

## 1. INTRODUCTION

Let  $X$  be a finite set, and let  $X^*$  be the free semigroup on  $X$ ; thus  $X^*$  consists of all finite nonempty strings of elements of  $X$ ,

$$X^* = \{x_1x_2 \cdots x_k \mid x_1, x_2, \dots, x_k \in X \ (k \geq 1)\},$$

together with the binary operation of concatenation,

$$(x_1x_2 \cdots x_k)(x_{k+1}x_{k+2} \cdots x_{k+\ell}) \mapsto x_1x_2 \cdots x_{k+\ell}.$$

Let  $n$  be a positive integer, and let  $(X^*)^n$  be the set of all  $n$ -tuples of elements of  $X^*$ . Let  $R(X, n)$  be the vector space with basis  $(X^*)^n$  over the field  $\mathbb{Q}$  of rational numbers. On  $R(X, n)$  we introduce a multilinear nonassociative operation  $\{a_1, a_2, \dots, a_n\}$ , called  **$n$ -ary intermolecular recombination**, and defined on basis elements as follows: for  $a_i = (a_{i1}, a_{i2}, \dots, a_{in}) \in (X^*)^n$  ( $i = 1, \dots, n$ ) we set

$$(1) \quad \{a_1, a_2, \dots, a_n\} = \sum_{\sigma \in S_n} (a_{\sigma(1),1}, a_{\sigma(2),2}, \dots, a_{\sigma(n),n}),$$

where  $S_n$  is the symmetric group on  $\{1, 2, \dots, n\}$ . Each  $a_i$  can be regarded as the abstract representation of a molecule divided into  $n$  consecutive submolecules  $a_{i1}, a_{i2}, \dots, a_{in}$ . (For example, if  $X = \{A, C, G, T\}$  represents the four bases found in DNA, then each  $a_i$  represents a DNA sequence partitioned into  $n$  subsequences.) With this interpretation, the sum defining  $\{a_1, a_2, \dots, a_n\}$  expresses the results of recombining the  $n$  molecules in all possible ways which preserve the positions of the  $n$  submolecules in each molecule. It is clear that  $n$ -ary intermolecular recombination satisfies the **complete symmetry identity** in degree  $n$ :

$$(2) \quad \{a_{\pi(1)}, a_{\pi(2)}, \dots, a_{\pi(n)}\} = \{a_1, a_2, \dots, a_n\},$$

for every  $\pi \in S_n$ ; thus  $\{a_1, a_2, \dots, a_n\}$  depends only on the multiset of arguments, not the order of the arguments.

In the case  $n = 2$ , the operation of binary intermolecular recombination was introduced by Landweber and Kari [4]. Bremner [1] used computer algebra to

show that every polynomial identity of degree  $\leq 5$  satisfied by this operation is a consequence of commutativity (binary complete symmetry) and the **binary recombination identity**

$$(3) \quad 2\{\{\{a, b\}, c\}, d\} - \{\{\{a, b\}, d\}, c\} - \{\{\{a, c\}, b\}, d\} - \{\{\{b, c\}, a\}, d\} \\ + \{\{a, b\}, \{c, d\}\} = 0.$$

Sverchkov [5] has recently shown that every identity, with no restriction on the degree, satisfied by binary intermolecular recombination is a consequence of commutativity and the binary recombination identity. In related work, Bremner, Piao and Richards [3] have shown that every polynomial identity satisfied by the zygotic algebras of simple Mendelian inheritance is a consequence of commutativity and the binary recombination identity.

In the present paper we consider the case  $n = 3$ ; we use computer algebra to determine the polynomial identities of degree  $\leq 9$  satisfied by ternary intermolecular recombination. A monomial which involves  $d$  applications of an  $n$ -ary operation has degree  $d(n-1) + 1$ ; thus polynomial identities for ternary intermolecular recombination exist only in odd degrees.

## 2. BINARY INTERMOLECULAR RECOMBINATION

In this section we recall the results of Bremner [1] on polynomial identities for binary intermolecular recombination, and show how these results may also be obtained (and slightly improved) using the Hermite normal form with lattice basis reduction.

In degree 4, there are 15 commutative nonassociative multilinear monomials, 12 with association type  $\{\{\{a, b\}, c\}, d\}$  and 3 with association type  $\{\{a, b\}, \{c, d\}\}$ . We order these monomials first by association type, and then lexicographically by the permutation of the variables:

$$\begin{aligned} & \{\{\{a, b\}, c\}, d\}, \quad \{\{\{a, b\}, d\}, c\}, \quad \{\{\{a, c\}, b\}, d\}, \quad \{\{\{a, c\}, d\}, b\}, \\ & \{\{\{a, d\}, b\}, c\}, \quad \{\{\{a, d\}, c\}, b\}, \quad \{\{\{b, c\}, a\}, d\}, \quad \{\{\{b, c\}, d\}, a\}, \\ & \{\{\{b, d\}, a\}, c\}, \quad \{\{\{b, d\}, c\}, a\}, \quad \{\{\{c, d\}, a\}, b\}, \quad \{\{\{c, d\}, b\}, a\}, \\ & \{\{a, b\}, \{c, d\}\}, \quad \{\{a, c\}, \{b, d\}\}, \quad \{\{a, d\}, \{b, c\}\}. \end{aligned}$$

We expand each monomial by three applications of binary intermolecular recombination. For example, writing

$$a = (a_1, a_2), \quad b = (b_1, b_2), \quad c = (c_1, c_2), \quad d = (d_1, d_2),$$

we obtain

$$\begin{aligned} \{\{\{a, b\}, c\}, d\} &= (a_1, d_2) + (b_1, d_2) + 2(c_1, d_2) + (d_1, a_2) + (d_1, b_2) + 2(d_1, c_2), \\ \{\{a, b\}, \{c, d\}\} &= (a_1, c_2) + (a_1, d_2) + (b_1, c_2) + (b_1, d_2) + (c_1, a_2) + (c_1, b_2) \\ &\quad + (d_1, a_2) + (d_1, b_2). \end{aligned}$$

Each expansion is a linear combination (allowing zero coefficients) of 12 pairs, which we order lexicographically:

$$\begin{aligned} & (a_1, b_2), \quad (a_1, c_2), \quad (a_1, d_2), \quad (b_1, a_2), \quad (b_1, c_2), \quad (b_1, d_2), \\ & (c_1, a_2), \quad (c_1, b_2), \quad (c_1, d_2), \quad (d_1, a_2), \quad (d_1, b_2), \quad (d_1, c_2). \end{aligned}$$

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 2 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 & 2 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 2 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 2 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 2 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 2 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 & 2 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 2 & 2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$
TABLE 1. The expansion matrix  $E$  in degree 4 for  $n = 2$ 

We store the expansions in the  $12 \times 15$  matrix  $E$  in which entry  $(i, j)$  contains the coefficient of pair  $i$  in the expansion of monomial  $j$ ; see Table 1. The coefficient vectors of the polynomial identities satisfied by binary intermolecular recombination in degree 4 are the nontrivial linear combinations of the columns of  $E$  that give the zero vector; that is, the nonzero vectors in the nullspace of  $E$ .

There are two ways to compute a basis for the nullspace of a matrix  $E$  with entries in the ring  $\mathbb{Z}$  of integers:

- (a) Compute the row canonical form of  $E$  over  $\mathbb{Q}$ , set the free variables equal to the standard basis vectors in  $\mathbb{Q}^d$  where  $d$  is the dimension of the nullspace, and solve for the leading variables.
- (b) Compute the Hermite normal form  $H$  of the transpose  $E^t$  together with a transform matrix  $U$  for which  $UE^t = H$ . The bottom  $d$  rows of  $U$  form a basis for the integer nullspace lattice of  $E$ . Then use the LLL algorithm for lattice basis reduction on the bottom  $d$  rows of  $U$ . (This method was first applied to polynomial identities by Bremner and Peresi [2].)

For the matrix  $E$  of Table 1, the RCF over  $\mathbb{Q}$  is given in Table 2, and the corresponding nullspace basis vectors (sorted by increasing length) are given in Table 3. The squared Euclidean norms of these basis vectors are

$$6, 8, 8, 12, 20, 22, 34, 52, 70.$$

For each basis vector, we apply all permutations of  $a, b, c, d$  to the corresponding identity  $I$  and store the results in a  $24 \times 15$  matrix  $M(I)$ ; the rank of  $M(I)$  is the dimension of the  $S_4$ -module of identities which are consequences of  $I$ . Rows 2 and 3 are the shortest vectors (squared norm 8) for which the corresponding identities generate the entire nullspace, and row 3 corresponds to the binary recombination identity.

For the matrix  $E$  of Table 1, the Maple command

```
HermiteForm( Transpose(E), output='U', method='integer[reduced]' ):
```

produces the transform matrix  $U$  of Table 4. The bottom 9 rows of  $U$  are a reduced basis for the integral nullspace lattice of  $E$ ; every one of these vectors has squared norm 6, the same as the shortest vector from Table 3. Even for this small matrix,

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & -3 & -3 & -5 & 0 & -3 & -2 & 1 & -2 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 3 & 3 & 5 & 1 & 4 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & -1 & 1 & 0 & -2 & -2 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 2 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -2 & -1 & -3 & -1 & -3 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 1 & 3 & 1 & 3 & 1 & 1 & 1 \end{bmatrix}$$

TABLE 2. The row canonical form of the matrix  $E$  for  $n = 2$ 

$$\begin{bmatrix} -1 & 1 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 2 & -1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 2 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & -3 & 0 & -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 3 & -3 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -3 & 1 & -1 & 2 & 0 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -4 & 2 & -2 & 3 & 0 & -3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 5 & -5 & 0 & -1 & 3 & 0 & -3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

TABLE 3. The canonical basis of the nullspace of  $E$  for  $n = 2$ 

$$\begin{bmatrix} 1 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 & 1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 & 1 & -1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & -1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

TABLE 4. A transform matrix  $U$  for which  $UE^t = H$  for  $n = 2$ 

the LLL algorithm has produced a basis of the nullspace with significantly shorter vectors. The identity corresponding to row 9 of  $U$  generates the entire nullspace:

$$(4) \quad \begin{aligned} & \{ \{ \{ a, b \}, c \}, d \} - \{ \{ \{ a, c \}, b \}, d \} - \{ \{ \{ b, c \}, d \}, a \} - \{ \{ \{ b, d \}, a \}, c \} \\ & + \{ \{ \{ b, d \}, c \}, a \} + \{ \{ a, b \}, \{ c, d \} \} = 0. \end{aligned}$$

Identity (4) has one more term than the binary recombination identity (3), but its coefficient vector is shorter.

Identities (3) and (4) both have only one term in the second association type, so both imply that every monomial in the second type can be expressed as a linear combination of monomials in the first type. For example, identity (4) can be written in the form

$$\begin{aligned} \{\{a, b\}, \{c, d\}\} &= -\{\{\{a, b\}, c\}, d\} + \{\{\{a, c\}, b\}, d\} + \{\{\{b, c\}, d\}, a\} \\ &\quad + \{\{\{b, d\}, a\}, c\} - \{\{\{b, d\}, c\}, a\}. \end{aligned}$$

This identity can be used to reduce the number of association types that we need to consider when studying identities of higher degree.

### 3. TERNARY INTERMOLECULAR RECOMBINATION

We now present the main results of this paper: a complete and minimal set of polynomial identities of degree  $\leq 9$  for ternary intermolecular recombination.

**3.1. Degree 3.** Every polynomial identity in degree 3 satisfied by ternary intermolecular recombination is a consequence of the complete symmetry identity. Indeed, equation (2) implies that there is only one monomial in degree 3, namely  $\{a, b, c\}$ , and hence the only possibly polynomial identity in degree 3 is  $\{a, b, c\} = 0$  for all  $a, b, c$ ; this clearly does not hold.

**3.2. Degree 5.** In degree 5 there are three distinct association types for a ternary operation:

$$\{\{a, b, c\}, d, e\}, \quad \{a, \{b, c, d\}, e\}, \quad \{a, b, \{c, d, e\}\}.$$

Equation (2) implies that we need to consider only the first type, and that we need to consider only 10 distinct monomials in this type; we order these monomials lexicographically:

$$\begin{aligned} &\{\{a, b, c\}, d, e\}, \quad \{\{a, b, d\}, c, e\}, \quad \{\{a, b, e\}, c, d\}, \quad \{\{a, c, d\}, b, e\}, \\ &\{\{a, c, e\}, b, d\}, \quad \{\{a, d, e\}, b, c\}, \quad \{\{b, c, d\}, a, e\}, \quad \{\{b, c, e\}, a, d\}, \\ &\{\{b, d, e\}, a, c\}, \quad \{\{c, d, e\}, a, b\}. \end{aligned}$$

Any polynomial identity in degree 5 satisfied by ternary intermolecular recombination is a linear combination of these 10 monomials.

**Lemma 1.** *Every polynomial identity of degree 5 satisfied by ternary intermolecular recombination is a consequence of the complete symmetry identity in degree 3.*

*Proof.* Consider the expansions of the 10 monomials in degree 5 using two applications of ternary intermolecular recombination. For example, since

$$\begin{aligned} \{a, b, c\} &= (a_1, b_2, c_3) + (a_1, c_2, b_3) + (b_1, a_2, c_3) \\ &\quad + (b_1, c_2, a_3) + (c_1, a_2, b_3) + (c_1, b_2, a_3), \end{aligned}$$

we obtain

$$\begin{aligned} \{\{a, b, c\}, d, e\} &= 2(a_1, d_2, e_3) + 2(a_1, e_2, d_3) + 2(d_1, a_2, e_3) + 2(d_1, e_2, a_3) \\ &\quad + 2(e_1, a_2, d_3) + 2(e_1, d_2, a_3) + 2(b_1, d_2, e_3) + 2(b_1, e_2, d_3) \\ &\quad + 2(d_1, b_2, e_3) + 2(d_1, e_2, b_3) + 2(e_1, b_2, d_3) + 2(e_1, d_2, b_3) \\ &\quad + 2(c_1, d_2, e_3) + 2(c_1, e_2, d_3) + 2(d_1, c_2, e_3) + 2(d_1, e_2, c_3) \\ &\quad + 2(e_1, c_2, d_3) + 2(e_1, d_2, c_3). \end{aligned}$$

row	triple	matrix entries									
1	$(a_1, b_2, c_3)$	0	0	0	0	0	2	0	0	2	2
2	$(a_1, b_2, d_3)$	0	0	0	0	2	0	0	2	0	2
3	$(a_1, b_2, e_3)$	0	0	0	2	0	0	2	0	0	2
5	$(a_1, c_2, d_3)$	0	0	2	0	0	0	0	2	2	0
6	$(a_1, c_2, e_3)$	0	2	0	0	0	0	2	0	2	0
9	$(a_1, d_2, e_3)$	2	0	0	0	0	0	2	2	0	0
17	$(b_1, c_2, d_3)$	0	0	2	0	2	2	0	0	0	0
18	$(b_1, c_2, e_3)$	0	2	0	2	0	2	0	0	0	0
21	$(b_1, d_2, e_3)$	2	0	0	2	2	0	0	0	0	0
33	$(c_1, d_2, e_3)$	2	2	2	0	0	0	0	0	0	0

TABLE 5. Submatrix for the proof of Lemma 1

Each term in these 10 expansions has the form  $(x_1, y_2, z_3)$  where  $(x, y, z)$  is a permutation of a 3-element subset of  $\{a, b, c, d, e\}$ . There are 60 such triples, which we order lexicographically:

$$\begin{array}{cccccc}
(a_1, b_2, c_3), & (a_1, b_2, d_3), & (a_1, b_2, e_3), & (a_1, c_2, b_3), & (a_1, c_2, d_3), & (a_1, c_2, e_3), \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
(e_1, c_2, a_3), & (e_1, c_2, b_3), & (e_1, c_2, d_3), & (e_1, d_2, a_3), & (e_1, d_2, b_3), & (e_1, d_2, c_3).
\end{array}$$

We construct the  $60 \times 10$  matrix  $E$  in which entry  $(i, j)$  contains the coefficient of triple  $i$  in the expansion of monomial  $j$ . The polynomial identities in degree 5 for ternary intermolecular recombination correspond to the nonzero vectors in the nullspace of  $E$ . It suffices to show that this matrix has full rank, and for this it suffices to find 10 rows for which the corresponding  $10 \times 10$  submatrix has full rank. Rows 1, 2, 3, 5, 6, 9, 17, 18, 21 and 33 produce the submatrix displayed in Table 5; the row canonical form of this submatrix is the identity matrix.  $\square$

**3.3. Degree 7.** We now consider 7 distinct molecules, each divided into 3 sub-molecules:

$$\begin{aligned}
a &= (a_1, a_2, a_3), & b &= (b_1, b_2, b_3), & c &= (c_1, c_2, c_3), & d &= (d_1, d_2, d_3), \\
e &= (e_1, e_2, e_3), & f &= (f_1, f_2, f_3), & g &= (g_1, g_2, g_3).
\end{aligned}$$

After performing three applications of ternary intermolecular recombination on these molecules with some order of operations and some permutation of molecules, we obtain a linear combination of triples of the form  $(x_1, y_2, z_3)$  where  $(x, y, z)$  is a permutation of a 3-element subset of  $\{a, b, c, d, e, f, g\}$ . The number of triples which can occur as terms in these linear combinations is therefore

$$3! \binom{7}{3} = 210.$$

We order these triples lexicographically by the permutation  $(x, y, z)$ .

Equation (2) implies that we need only two association types in degree 7,

$$\{\{a, b, c\}, d, e\}, f, g \quad \text{and} \quad \{a, b, c\}, \{d, e, f\}, g,$$

$$\begin{aligned}
 & 4[a_1, f_2, g_3] + 4[a_1, g_2, f_3] + 4[b_1, f_2, g_3] + 4[b_1, g_2, f_3] + 4[c_1, f_2, g_3] \\
 & + 4[c_1, g_2, f_3] + 12[d_1, f_2, g_3] + 12[d_1, g_2, f_3] + 12[e_1, f_2, g_3] + 12[e_1, g_2, f_3] \\
 & + 4[f_1, a_2, g_3] + 4[f_1, b_2, g_3] + 4[f_1, c_2, g_3] + 12[f_1, d_2, g_3] + 12[f_1, e_2, g_3] \\
 & + 4[f_1, g_2, a_3] + 4[f_1, g_2, b_3] + 4[f_1, g_2, c_3] + 12[f_1, g_2, d_3] + 12[f_1, g_2, e_3] \\
 & + 4[g_1, a_2, f_3] + 4[g_1, b_2, f_3] + 4[g_1, c_2, f_3] + 12[g_1, d_2, f_3] + 12[g_1, e_2, f_3] \\
 & + 4[g_1, f_2, a_3] + 4[g_1, f_2, b_3] + 4[g_1, f_2, c_3] + 12[g_1, f_2, d_3] + 12[g_1, f_2, e_3]
 \end{aligned}$$

TABLE 6. Expansion of monomial  $\{\{a, b, c\}, d, e\}, f, g\}$

$$\begin{aligned}
 & 4[a_1, d_2, g_3] + 4[a_1, e_2, g_3] + 4[a_1, f_2, g_3] + 4[a_1, g_2, d_3] + 4[a_1, g_2, e_3] \\
 & + 4[a_1, g_2, f_3] + 4[b_1, d_2, g_3] + 4[b_1, e_2, g_3] + 4[b_1, f_2, g_3] + 4[b_1, g_2, d_3] \\
 & + 4[b_1, g_2, e_3] + 4[b_1, g_2, f_3] + 4[c_1, d_2, g_3] + 4[c_1, e_2, g_3] + 4[c_1, f_2, g_3] \\
 & + 4[c_1, g_2, d_3] + 4[c_1, g_2, e_3] + 4[c_1, g_2, f_3] + 4[d_1, a_2, g_3] + 4[d_1, b_2, g_3] \\
 & + 4[d_1, c_2, g_3] + 4[d_1, g_2, a_3] + 4[d_1, g_2, b_3] + 4[d_1, g_2, c_3] + 4[e_1, a_2, g_3] \\
 & + 4[e_1, b_2, g_3] + 4[e_1, c_2, g_3] + 4[e_1, g_2, a_3] + 4[e_1, g_2, b_3] + 4[e_1, g_2, c_3] \\
 & + 4[f_1, a_2, g_3] + 4[f_1, b_2, g_3] + 4[f_1, c_2, g_3] + 4[f_1, g_2, a_3] + 4[f_1, g_2, b_3] \\
 & + 4[f_1, g_2, c_3] + 4[g_1, a_2, d_3] + 4[g_1, a_2, e_3] + 4[g_1, a_2, f_3] + 4[g_1, b_2, d_3] \\
 & + 4[g_1, b_2, e_3] + 4[g_1, b_2, f_3] + 4[g_1, c_2, d_3] + 4[g_1, c_2, e_3] + 4[g_1, c_2, f_3] \\
 & + 4[g_1, d_2, a_3] + 4[g_1, d_2, b_3] + 4[g_1, d_2, c_3] + 4[g_1, e_2, a_3] + 4[g_1, e_2, b_3] \\
 & + 4[g_1, e_2, c_3] + 4[g_1, f_2, a_3] + 4[g_1, f_2, b_3] + 4[g_1, f_2, c_3]
 \end{aligned}$$

TABLE 7. Expansion of monomial  $\{\{a, b, c\}, \{d, e, f\}, g\}$

which contain respectively the following numbers of distinct monomials,

$$\frac{7!}{3!2!2!} = 210 \quad \text{and} \quad \frac{1}{2!} \cdot \frac{7!}{3!3!1!} = 70.$$

Within each association type, we order the monomials lexicographically by the permutation of the variables. Any polynomial identity in degree 7 satisfied by ternary intermolecular recombination is a linear combination of these 280 nonassociative monomials.

The expansion matrix  $E$  in degree 7 has 210 rows and 280 columns; entry  $(i, j)$  is the coefficient of triple  $i$  in the expansion of nonassociative monomial  $j$ . The expansions of the two association types are displayed in Tables 6 and 7. Each expansion has  $6^3 = 216$  terms; after collecting terms, the first type produces a linear combination of 30 triples with multiplicities 4 and 12, and the second type produces a linear combination of 54 triples with multiplicity 4. We use Maple to find that the matrix  $E$  has rank 35, and so its nullspace has dimension 245. We compute the row canonical form of  $E$ , and find that every entry of the RCF is an integer. We obtain the canonical basis of the nullspace of  $E$ ; this is a list of 245 vectors of dimension 280 with integer components. We sort these vectors by increasing Euclidean norm; the list of squared norms is displayed in Table 8.

4	4	4	4	4	4	4	4	6	6
6	6	6	6	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6
6	8	8	8	8	8	8	8	8	8
10	10	10	10	10	10	10	10	10	10
10	10	12	12	12	12	12	14	16	32
32	34	34	34	34	52	52	52	52	52
52	80	80	82	82	86	86	88	94	94
96	96	98	100	100	100	100	100	100	104
104	104	108	108	110	110	110	110	110	110
118	118	122	122	128	128	134	134	134	134
136	138	138	140	140	144	146	146	146	152
152	156	156	156	156	158	158	158	158	158
162	162	162	170	170	170	170	174	174	174
176	176	178	178	182	184	186	188	188	188
188	190	190	190	190	198	206	210	216	228
232	234	240	240	242	246	246	246	246	246
252	254	256	266	270	728	728	744	744	756
756	876	876	1210	1210	1216	1216	1244	1244	1350
1386	1386	1402	1402	1408	1412	1414	1414	1440	1470
1482	1484	1494	1516	1568	1592	1678	1758	1818	1824
1852	1894	1908	1910	1918	1946	1946	1982	2002	2026
2064	2174	2196	2226	2246	2272	2312	2324	2326	2340
2346	2366	2390	2396	2402	2418	2430	2472	2490	2562
2634	2640	2654	2668	2844					

TABLE 8. Squared norms of canonical basis vectors in degree 7

4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	4	4	4	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	10	10	10
10	10	10	10	10	10	10	10	12	12	12	12	12	12	12
12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
12	12	12	12	12	12	12	14	14	14	14	14	14	14	14
14	14	14	14	14	14	14	16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16	18	18	18	18	18	18	18
18	20	20	20	20	20	20	20	20	20	20	22	22	22	22
22	22	22	22	22	22	24	24	24	24	24	24	24	24	24
26	26	26	26	26	28	28	28	28	28	28	30	30	30	32
32	34	34	34	38										

TABLE 9. Squared norms of reduced basis vectors in degree 7

We now perform further computations, using modular arithmetic to save memory, to determine a set of generators for the nullspace as a module over the symmetric group  $S_7$ . (We use  $p = 101$ , the smallest prime greater than 100; any prime greater than the degree of the identities would produce the same dimensions.) For each basis vector in the nullspace, we apply all 5040 permutations of  $a, b, c, d, e, f, g$  to the corresponding identity, and store the results in a matrix of size  $5040 \times 280$ . We compute the row canonical form of this matrix; its nonzero rows form a basis of the  $S_7$ -module generated by the identity, and its rank is the dimension of this module. We process the nullspace basis vectors in order, saving previous results (the nonzero rows which form a basis of the  $S_7$ -module generated by the previous basis vectors). At each stage, an increase in the rank implies that the current identity is a new generator of the nullspace as a module over  $S_7$ . The results of these computations show that the canonical basis vectors with positions 1, 10 and 60 (after the basis vectors have been sorted by increasing Euclidean norm) represent polynomial identities which generate the entire nullspace as an  $S_7$ -module:

$$\begin{aligned}
I &= \{\{a, b, c\}, d, e\}, f, g\} - \{\{a, b, e\}, c, d\}, f, g\} - \{\{b, c, d\}, a, e\}, f, g\} \\
&\quad + \{\{b, d, e\}, a, c\}, f, g\}, \\
J &= \{\{a, b, e\}, c, g\}, d, f\} - \{\{a, b, e\}, f, g\}, c, d\} - \{\{a, c, e\}, b, g\}, d, f\} \\
&\quad + \{\{a, c, e\}, f, g\}, b, d\} + \{\{a, e, f\}, b, g\}, c, d\} - \{\{a, e, f\}, c, g\}, b, d\}, \\
K &= 2\{\{a, b, c\}, d, e\}, f, g\} - \{\{a, b, c\}, d, g\}, e, f\} - 2\{\{a, b, c\}, e, g\}, d, f\} \\
&\quad + 2\{\{a, b, d\}, c, e\}, f, g\} - 2\{\{a, b, d\}, c, g\}, e, f\} - \{\{a, b, d\}, e, g\}, c, f\} \\
&\quad - \{\{a, b, e\}, c, d\}, f, g\} + \{\{a, b, e\}, c, g\}, d, f\} + 3\{\{a, b, g\}, c, d\}, e, f\} \\
&\quad - \{\{a, c, d\}, b, e\}, f, g\} - \{\{b, c, d\}, a, e\}, f, g\} + \{a, b, e\}, \{c, d, g\}, f\}.
\end{aligned}$$

Every identity in the nullspace is a linear combination of permutations of these three identities, which have squared norms 4, 6 and 32; the second identity has been normalized so that it begins with a positive coefficient.

We can get significantly better results using the Hermite normal form and lattice basis reduction.

**Theorem 2.** *Every polynomial identity of degree 7 satisfied by ternary intermolecular recombination is a consequence of complete symmetry in degree 3 and these three identities in degree 7:*

$$\begin{aligned}
P &= \{\{a, c, f\}, e, g\}, b, d\} - \{\{a, c, g\}, e, f\}, b, d\} - \{\{c, e, f\}, a, g\}, b, d\} \\
&\quad + \{\{c, e, g\}, a, f\}, b, d\}, \\
Q &= \{\{a, d, e\}, b, g\}, c, f\} - \{\{a, d, e\}, f, g\}, b, c\} - \{\{b, d, e\}, a, g\}, c, f\} \\
&\quad + \{\{b, d, e\}, f, g\}, a, c\} + \{\{d, e, f\}, a, g\}, b, c\} - \{\{d, e, f\}, b, g\}, a, c\}, \\
R &= \{\{a, b, c\}, f, g\}, d, e\} + \{\{a, b, g\}, d, f\}, c, e\} + \{\{a, c, d\}, b, f\}, e, g\} \\
&\quad - \{\{a, c, d\}, b, g\}, e, f\} + \{\{a, c, d\}, f, g\}, b, e\} - \{\{a, c, f\}, b, d\}, e, g\} \\
&\quad + \{\{a, c, g\}, d, f\}, b, e\} - \{\{a, d, f\}, c, g\}, b, e\} - \{\{a, d, g\}, b, c\}, e, f\} \\
&\quad - \{\{a, f, g\}, b, c\}, d, e\} + \{\{b, c, d\}, a, g\}, e, f\} - \{a, d, g\}, \{b, c, f\}, e\}.
\end{aligned}$$

*Identities P and Q are independent: neither implies the other. Identity R implies both P and Q, but identities P and Q together do not imply R.*

*Proof.* We apply the Maple command

```
HermiteForm( Transpose(E), output='U', method='integer[reduced]' ):
```

to the transpose (size  $280 \times 210$ ) of the expansion matrix  $E$ , and obtain a transform matrix  $U$  (size  $280 \times 280$ ). The bottom 245 rows of  $U$  are a basis for the integral nullspace lattice of  $E$ . We sort these vectors by increasing Euclidean norm; the squared norms are displayed in Table 9. It is clear by comparing Tables 8 and 9 that the basis vectors obtained in this way are generally much shorter than those obtained from the RCF. Identities  $P$ ,  $Q$  and  $R$  have coefficients  $\pm 1$  and the squared norms of the corresponding coefficient vectors are 4, 6 and 12. (These identities correspond to the reduced basis vectors in positions 1, 40 and 129 after the reduced basis vectors have been sorted by increasing Euclidean norm.) These three identities can be verified by expanding each term using three applications of ternary intermolecular recombination, and then checking that the results collapse to zero.

To prove the stated dependence and independence relations we proceed as follows. Given an identity  $I(a, b, c, d, e, f, g)$  we apply all 5040 permutations of  $a, b, c, d, e, f, g$  and store the coefficients of the resulting identities in a  $5040 \times 280$  matrix. The rank of this matrix is the dimension of the  $S_7$ -submodule of identities generated by  $I$ . To check whether another identity  $J(a, b, c, d, e, f, g)$  is implied by  $I$  we simply determine whether  $J$  is in the row space of this matrix. We find that identity  $P$  produces dimension 105, and identity  $Q$  produces dimension 127; stacking the two corresponding matrices together, we see that these two identities together produce dimension 155. However, identity  $R$  produces dimension 245, and this is also the rank produced by the three identities together. Since 245 is also the dimension of the nullspace of the expansion matrix, it follows that every identity in degree 7 is a consequence of the identity  $R$ , in the sense that every identity in degree 7 is a linear combination of permutations of  $R(a, b, c, d, e, f, g)$ .  $\square$

The identity  $R$  of Theorem 2 will be called the **ternary recombination identity**. The last term in  $R$  is the only term in identities  $P$ ,  $Q$  and  $R$  which has the second association type, so we have the following result.

**Corollary 3.** *Every polynomial identity of degree  $\leq 7$  satisfied by ternary intermolecular recombination is a consequence of complete symmetry and the following identity, which expresses any monomial in the second association type as a linear combination of monomials in the first association type:*

$$\begin{aligned} & \{\{a, b, c\}, \{d, e, f\}, g\} = \\ & \quad \{\{\{a, d, e\}, f, c\}, b, g\} + \{\{\{a, d, c\}, b, f\}, e, g\} + \{\{\{a, e, b\}, d, f\}, g, c\} \\ & \quad - \{\{\{a, e, b\}, d, c\}, g, f\} + \{\{\{a, e, b\}, f, c\}, d, g\} - \{\{\{a, e, f\}, d, b\}, g, c\} \\ & \quad + \{\{\{a, e, c\}, b, f\}, d, g\} - \{\{\{a, b, f\}, e, c\}, d, g\} - \{\{\{a, b, c\}, d, e\}, g, f\} \\ & \quad - \{\{\{a, f, c\}, d, e\}, b, g\} + \{\{\{d, e, b\}, a, c\}, g, f\}. \end{aligned}$$

(We have permuted the variables so that the term in the second association type has the identity permutation.)

**3.4. Degree 9.** The number of permutations of 3-element subsets of a set of 9 variables is clearly

$$3! \binom{9}{3} = 504.$$

This is the number of rows in the expansion matrix in degree 9.

Equation (2) implies that we need only four association types in degree 9:

$$\begin{aligned} & \{\{\{a, b, c, d, e, f, g, h, i\}\} \text{ with } \frac{9!}{3!2!2!2!} = 7560 \text{ monomials,} \\ & \{\{\{a, b, c, \{d, e, f, g, h, i\}\}\} \text{ with } \frac{1}{2!} \cdot \frac{9!}{3!3!1!2!} = 2520 \text{ monomials,} \\ & \{\{\{a, b, c, d, e, \{f, g, h, i\}\}\} \text{ with } \frac{9!}{3!2!3!1!} = 5040 \text{ monomials, and} \\ & \{\{a, b, c, \{d, e, f\}, \{g, h, i\}\}\} \text{ with } \frac{1}{3!} \cdot \frac{9!}{3!3!3!} = 280 \text{ monomials.} \end{aligned}$$

The total is 15400; this is the number of columns in the expansion matrix.

We initialize the expansion matrix and then use modular arithmetic (again with  $p = 101$ ) to compute its rank; the result is 84, so the dimension of the nullspace is 15316. To decide whether any nullspace vector represents a new identity in degree 9, we need to determine the dimension of the subspace of the nullspace which consists of all consequences of the ternary recombination identity  $R = R(a, b, c, d, e, f, g)$  from Theorem 2. To obtain consequences in degree 9, we can either

- (i) replace one of the variables  $x \in \{a, b, c, d, e, f, g\}$  by the triple  $\{x, h, i\}$  where  $h$  and  $i$  are two new variables, or
- (ii) embed the entire identity  $R$  in a triple  $\{R, h, i\}$ .

This gives 8 identities in degree 9 which generate the  $S_9$ -module of all consequences of the identity  $R$  in degree 7:

$$\begin{aligned} R(\{a, h, i\}, b, c, d, e, f, g), & \quad R(a, \{b, h, i\}, c, d, e, f, g), \\ R(a, b, \{c, h, i\}, d, e, f, g), & \quad R(a, b, c, \{d, h, i\}, e, f, g), \\ R(a, b, c, d, \{e, h, i\}, f, g), & \quad R(a, b, c, d, e, \{f, h, i\}, g), \\ R(a, b, c, d, e, f, \{g, h, i\}), & \quad \{R(a, b, c, d, e, f, g), h, i\}. \end{aligned}$$

We need to “straighten” the terms of these identities: that is, we use the complete symmetry identity to replace each monomial by an equivalent monomial in one of the four association types listed above.

We now apply all  $9!$  permutations of  $a, b, c, d, e, f, g, h, i$  to each of the 8 consequences of  $R$  in degree 9, store these permuted identities in the rows of a matrix with 15400 columns, and compute the rank of the matrix. For  $\ell = 1, 2, \dots, 8$  the dimension of the subspace generated by the first  $\ell$  consequences of  $R$  is as follows:

$$14071, 15036, 15036, 15036, 15316, 15316, 15316, 15316.$$

Since 15316 is also the dimension of the nullspace of the expansion matrix, it follows that every polynomial identity in degree 9 satisfied by ternary intermolecular recombination is a consequence of the ternary recombination identity in degree 7. Hence there are no new identities in degree 9.

#### 4. CONCLUSION

The results of Bremner [1], Sverchkov [5] and the present paper are consistent with the following conjectures for the general case of  $n$ -ary intermolecular recombination.

**Conjecture 1.** *Every polynomial identity satisfied by  $n$ -ary intermolecular recombination in degree  $2n-1$  (that is, in which each term involves two applications of the  $n$ -ary operation) is a consequence of complete symmetry in degree  $n$ .*

**Conjecture 2.** *Every polynomial identity satisfied by  $n$ -ary intermolecular recombination in degree  $3n-2$  (that is, in which each term involves three applications of the  $n$ -ary operation) is a consequence of complete symmetry in degree  $n$  together with an identity  $R$  in degree  $3n-2$  in which every coefficient equals  $\pm 1$  and which implies that every monomial in the second association type,*

$$\{ \{ a_1, \dots, a_n \}, \{ a_{n+1}, \dots, a_{2n} \}, a_{2n+1}, \dots, a_{3n-2} \},$$

*equals a linear combination of monomials in the first association type,*

$$\{ \{ \{ a_1, \dots, a_n \}, a_{n+1}, \dots, a_{2n-1} \}, a_{2n}, \dots, a_{3n-2} \}.$$

**Conjecture 3.** *Every polynomial identity satisfied by  $n$ -ary intermolecular recombination, with no restriction on the degree, is a consequence of complete symmetry in degree  $n$  together with the identity  $R$  from Conjecture 2 in degree  $3n-2$ .*

Resolving these conjectures is the most important open problem in the theory of  $n$ -ary intermolecular recombination.

## 5. ACKNOWLEDGEMENTS

I thank the organizers (B. H. Li, R. C. Laubenbacher, J. J. P. Tian) of the Special Session on Biomathematics at the First Joint Meeting of the Shanghai Mathematical Society and the American Mathematical Society (Shanghai, China, December 17–21, 2008) for the invitation to give a presentation at the conference and to contribute a paper to this special issue.

This research was supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

## REFERENCES

- [1] M. R. Bremner: *Jordan algebras arising from intermolecular recombination*. SIGSAM Bulletin (Communications in Computer Algebra). Volume 39, Number 4, December 2005, Issue 154, Pages 106–117.
- [2] M. R. Bremner and L. A. Peresi: *An application of lattice basis reduction to polynomial identities for algebraic systems*. Linear Algebra and its Applications. Volume 430, Issues 2–3, 15 January 2009, Pages 642–659.
- [3] M. R. Bremner, Y. F. Piao and S. W. Richards: *Polynomial identities for Bernstein algebras of simple Mendelian inheritance*. Communications in Algebra. To appear.
- [4] L. Landweber and L. Kari: *The evolution of cellular computing: nature's solution to a computational problem*. Biosystems. Volume 52, Issues 1–3, October 1999, Pages 3–13.
- [5] S. R. Sverchkov: *Structure and representations of Jordan algebras arising from intermolecular recombination*. To appear in *Algebras, Representations and Applications*, Proceedings of the Conference (Maresias, Brazil, August 2007) in Honor of the Sixtieth Birthday of Ivan P. Shestakov. Edited by Vyacheslav Futorny, Victor Kac, Iryna Kashuba, and Efim Zelmanov. Contemporary Mathematics, American Mathematical Society, Volume 483.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF SASKATCHEWAN, CANADA  
*E-mail address:* `bremner@math.usask.ca`