Analyzing Ownership Sets in HPF

Pramod G. Joisha *Prithviraj Banerjee*

Technical Report No. CPDC–TR–9906–012
© 1999 Center for Parallel and Distributed Computing
June 1999

Center for Parallel and Distributed Computing
Department of Electrical & Computer Engineering,
Technological Institute,
2145 Sheridan Road,
Northwestern University,
IL 60208–3118.

Analyzing Ownership Sets in HPF*

Pramod G. Joisha Prithviraj Banerjee

Center for Parallel and Distributed Computing, Electrical and Computer Engineering Department, Technological Institute, 2145 Sheridan Road, Northwestern University, IL 60208–3118.
Phone: (847) 467-4610, Fax: (847) 491-4455
Email: `[pjoisha, banerjee]@ece.nwu.edu`

Abstract

Ownership sets are fundamental to the partitioning of program computations across processors by the owner-computes rule. These sets arise due to the mapping of arrays onto processors. In this report, we focus on how ownership sets can be efficiently determined in the context of the HPF language, and show how the structure of these sets can be symbolically characterized in the presence of arbitrary array alignment and array distribution directives. Our starting point is a system of equalities and inequalities due to Ancourt et al. that captures the array mapping problem in HPF. We arrive at a refined system that enables us to efficiently solve for the ownership set using the Fourier-Motzkin Elimination technique, and that requires the course vector as the only auxiliary vector. The formulation makes it possible to enumerate the elements of the ownership set exactly once, a feature that is very beneficial when such sets are applied to handle `DO` loops qualified by HPF's `INDEPENDENT` directive.

We develop important and general properties pertaining to HPF alignments and distributions, and show how they can be used to eliminate redundant communication due to array replication. Polynomial-time schemes that determine whether the ownership set of a particular processor with respect to some array is the empty set, or whether the ownership set of every processor with respect to some array is the empty set, are presented. We show how distribution directives with unspecified processor meshes can be efficiently handled at compile time. We also show how to avoid the generation of communication code when pairs of array references are ultimately mapped onto the same processors. Experimental data demonstrating the improved code performance that the latter optimization enables is presented and discussed.

Keywords HPF, array alignment, array distribution, ownership set, Fourier-Motzkin Elimination technique, parallelizing compiler

*This research was partially supported by the National Science Foundation under Grant NSF CCR–9526325, and in part by DARPA under Contract F30602–98–2–0144.

1 Introduction

In languages such as High Performance FORTRAN (HPF) [KLS+94], array mappings guide the partitioning of program computations across processors. They are specified by the programmer in terms of annotations called *directives*. The actual mapping process typically involves two steps: Arrays are first *aligned* with a template and templates are then *distributed* over a virtual mesh of processors. The alignment operation, performed via the `ALIGN` directive, assigns every array element to at least a single template cell. The distribution operation, done using the `DISTRIBUTE` directive, associates every template cell with exactly one processor. In this way, array elements are eventually mapped onto processors.

In an automated code generation scenario, the compiler decides the processors on which to execute the various compute operations occurring in a program. In allocating program computations to processors, the compiler uses the mapping information associated with the data. A possible scheme, known as the *owner-computes rule* [Wol96], is to allow only the owner of the left-hand side reference in an assignment statement to execute the statement. By the owner-computes rule, expressions that use data located on the same processor can be evaluated locally on that processor, without the need for inter-processor communication. When the need to transfer remotely located data arises, the compiler produces the relevant communication code. Hence, the owner-computes rule leads to the notion of an array's *ownership set*, which is the set of all its elements mapped onto a processor by virtue of the alignment and distribution directives.

Since the assignment of computations to processors is determined by the allocation of data to processors, one of the aims of the HPF compilation problem is to find a suitable way of representing the mapping information. Given such a representation, the next issue that must be addressed is how can it be used to realize the owner-computes rule. Does the proposed framework provide insights into the nature of the array mapping problem? Does the representation reveal general properties that can be leveraged to generate efficient code? In this report, we investigate these questions in the context of a recent representation proposed by Ancourt et al [ACIK95].

1.1 An Example

Figure 1 shows a four-point stencil computation that occurs over a two-dimensional grid of 1022×1022 points. The value at every grid point is updated with the average of its four neighbors. Once the entire grid is updated this way, the process is repeated with the new values.

The example introduces an abstract two-dimensional array `T` of size 1024×1024, called the template, against which the arrays `A` and `B` are aligned. The alignment directives align the array elements `A(i, j)` and `B(i, j)` with the template cell `T(i, j)`. The `DISTRIBUTE` directive maps blocks of template cells in `T` onto a two-dimensional virtual processor mesh `P`. The extent of a block along a dimension is determined by the extents of `T` and `P` along that dimension. Since the extents along each dimension in this case are 1024 and 2 respectively, a block size of $(1024/2) \times (1024/2) = 512 \times 512$ is used. Thus, the cells of the template `T` in the region $\mathtt{T}(512p_1 + l_1,\ 512p_2 + l_2)$ get mapped onto the processor $\mathtt{P}(p_1,\ p_2)$, where $0 \leq l_1, l_2 < 512$ and $0 \leq p_1, p_2 < 2$. We could have also chosen a $\mathtt{CYCLIC}(B)$ distribution along a particular processor dimension instead of a `BLOCK` distribution; in this case, the distribution is allowed to "wrap around" the processor dimension. Because `A(i, j)` and `B(i, j)` are aligned with `T(i, j)`, the `(BLOCK, BLOCK)` distribution results in each of the four quadrants in `A` and `B` being mapped onto the corresponding "quadrant processor" in `P`. Hence, a specific processor would execute only those statement instances in Figure 1 whose left-hand sides lie in the quadrant of `A` or `B` owned by it.

1.2 Related Work

The problem of array alignment and array distribution has been extensively studied and numerous structures that describe the mapping of arrays to processors have been suggested and examined [CK87, Bal90, AL93, SPB94, TR96, AMC98]. Early representations focused on `BLOCK` distributions alone and were incapable of conveniently describing the general $\mathtt{CYCLIC}(B)$ distribution. This deficiency was addressed in subsequent work by using techniques ranging from finite state machines, virtual processor meshes to set-theoretic methods

```
       program example

       REAL A(2:1023, 2:1023)
       REAL B(1:1024, 1:1024)
       INTEGER i, j, k

!HPF$ TEMPLATE T(1:1024, 1:1024)
!HPF$ PROCESSORS P(2, 2)
!HPF$ ALIGN A(i, j) WITH T(i, j)
!HPF$ ALIGN B(i, j) WITH T(i, j)
!HPF$ DISTRIBUTE T(BLOCK, BLOCK) ONTO P

!C     Arrays A and B are initialized here.

       do k = 0, 99
          do j = 2, 1023
             do i = 2, 1023
                A(i, j) = (B(i-1, j)+B(i+1, j)+B(i, j-1)+B(i, j+1))/4
             enddo
          enddo
          do j = 2, 1023
             do i = 2, 1023
                 B(i, j) = A(i, j)
             enddo
          enddo
       enddo

       end
```

FIGURE 1: A Four-point Stencil Computation in HPF

[CGL+95, GKM+93, Mid95]. However, these schemes primarily concentrated on enumerating local memory access sequences and handling array expressions. A generalized view of the HPF mapping problem was subsequently presented by Ancourt et al. [ACIK95] who showed how a system of equalities and inequalities could be used to mathematically express the *regular* alignment and distribution of arrays to processors. These systems were then used to formulate ownership sets and compute sets for loops qualified by the `INDEPENDENT` directive, and parametric solutions for the latter were provided based on the Hermite Normal Form [ACIK95].

1.3 Contributions

In this report, we investigate the ownership set formulation in the Ancourt et al. framework and show how it can be refined to a form that requires a course vector as the only auxiliary vector and that also enables the efficient enumeration of its constituent elements. This property is desirable when ownership sets are applied to handle `DO` loops qualified by HPF's `INDEPENDENT` directive. Our approach to solving for the ownership set is based on the Fourier-Motzkin Elimination (FME) technique, and in that respect, we deviate from [ACIK95]. We also formulate an efficient polynomial-time test using which redundant communication due to array replication can be avoided. We present a sufficient condition called the *mapping test* that eliminates the need for generating communication code for certain right-hand side array references in an assignment statement. These two optimizations in turn depend upon whether the ownership set of a *particular* processor with respect to some array is the empty set, or whether the ownership set of *every* processor with respect to some array is the empty set. We discuss how these decisions can be arrived at in polynomial time, once the symbolic representation of the ownership set is known. The mapping test often results in marked performance improvements and we substantiate this by presenting experimental data. Finally, we discuss how our schemes permit the efficient handling at compile time of distributions in which the processor meshes are unspecified. The techniques mentioned in this report have been incorporated into a new version of the PARADIGM compiler [JB99] using *Mathematica*[b] as the symbolic manipulation engine.

1.4 Outline

The rest of this report is organized as follows. In § 2, we describe previous research that forms the foundations of our work. We discuss in this section how information pertaining to alignments and distributions can be compactly expressed as systems of equalities and inequalities. The ownership set is formally defined in this section. We refine the ownership set formulation in § 3 and prove an important equivalence relation. In § 4, we present the replication test and explain how it can be used to avoid redundant communication due to array replication. We also show in § 4 how to ascertain in polynomial time whether the ownership set is empty or nonempty, given its symbolic FME solution. In § 5, we describe the handling of distributions that lack an explicitly specified processor mesh. The mapping test is derived in § 6 and we illustrate its workings using three examples. Finally, in § 7, we report and analyze experimental data that demonstrates the performance benefits of the mapping test optimization.

2 Preliminaries

Figure 2 shows an artificial code fragment comprising a declaration and a set of HPF directives. Since the first dimension of A and the single subscript-triplet expression conform, this fragment is equivalent to that shown in Figure 3. The dummy variables i, j, k and l in Figure 3 satisfy the constraints $-1 \leq i \leq 20$, $3 \leq j \leq 40$, $0 \leq k \leq 20$ and $0 \leq l \leq 99$ respectively.

```
REAL A(-1:20, 3:40, 0:20)

!HPF$ TEMPLATE T(0:99, 0:99, 0:99)
!HPF$ PROCESSORS P(1:9, 1:9)
!HPF$ ALIGN A(:, *, k) WITH T(2*k+1, 2:44:2, *)
!HPF$ DISTRIBUTE T(*, CYCLIC(4), BLOCK(13)) ONTO P
```

FIGURE 2: Original Fragment

```
REAL A(-1:20, 3:40, 0:20)

!HPF$ TEMPLATE T(0:99, 0:99, 0:99)
!HPF$ PROCESSORS P(1:9, 1:9)
!HPF$ ALIGN A(i, j, k) WITH T(2*k+1, (i+1)*2+2, l)
!HPF$ DISTRIBUTE T(*, CYCLIC(4), BLOCK(13)) ONTO P
```

FIGURE 3: Modified Fragment

The alignment directives in Figure 3 can be compactly expressed through the following collection of equalities and inequalities [ACIK95]:

$$\hat{\mathcal{R}}\boldsymbol{t} = \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}, \tag{1}$$

$$\boldsymbol{a_l} \leq \boldsymbol{a} \leq \boldsymbol{a_u}, \tag{2}$$

$$\boldsymbol{0} \leq \boldsymbol{t} \leq \boldsymbol{u_T} - \boldsymbol{l_T}. \tag{3}$$

Similarly, the distribution directives in Figure 3 can be represented by the following system [ACIK95]:

$$\hat{\pi}\boldsymbol{t} = \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \boldsymbol{l}, \tag{4}$$

$$\hat{\lambda}\boldsymbol{c} = \boldsymbol{0}, \tag{5}$$

$$\boldsymbol{0} \leq \boldsymbol{p} < \hat{P}\boldsymbol{1}, \tag{6}$$

$$\boldsymbol{0} \leq \boldsymbol{l} < \hat{C}\boldsymbol{1}, \tag{7}$$

where the column vector $\boldsymbol{t}$ in Equation (4) satisfies Inequality (3). For the given example, the various matrices and vectors are

$$\hat{\mathcal{R}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{\mathcal{A}} = \begin{pmatrix} 0 & 0 & 2 \\ 2 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{s_0} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \quad \boldsymbol{a_l} = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix}, \quad \boldsymbol{a_u} = \begin{pmatrix} 20 \\ 40 \\ 20 \end{pmatrix}, \quad \boldsymbol{l_T} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{u_T} = \begin{pmatrix} 99 \\ 99 \\ 99 \end{pmatrix},$$

[b] *Mathematica* is a registered trademark of Wolfram Research, Inc.

and

$$\hat{\pi} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \hat{C} = \begin{pmatrix} 4 & 0 \\ 0 & 13 \end{pmatrix}, \qquad \hat{P} = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}, \qquad \hat{\lambda} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

While Equation (1) is a consequence of the `ALIGN` directive, Equations (4) and (5) are a result of the `DISTRIBUTE` directive. While Equation (4) dictates the mapping of template cells onto processors, Equation (5) indicates whether a particular processor dimension has a `BLOCK` or a `CYCLIC` distribution associated with it. Constraints on the array bounds vector $\boldsymbol{a}$ and the template cell vector $\boldsymbol{t}$ are given by Inequalities (2) and (3) respectively. Finally, Inequalities (6) and (7) describe the constraints that the processor identity vector $\boldsymbol{p}$ and the offsets vector $\boldsymbol{l}$ must satisfy. We shall use x, y and z to represent the number of dimensions of the *alignee* [KLS+94], template and processor mesh respectively. Using this notation, the column vectors $\boldsymbol{a}$ and $\boldsymbol{t}$ consist of x and y elements respectively, while the column vectors $\boldsymbol{p}$, $\boldsymbol{l}$ and $\boldsymbol{c}$—called the *course vector*—have z elements each.

The ownership set of a processor $\boldsymbol{p}$, which is defined with respect to an array `X`, denotes those elements of `X` that are finally mapped onto $\boldsymbol{p}$ by virtue of the alignment and distribution directives. In set-theoretic notation, this set is

$$\begin{aligned}\Delta_p(X) = \big\{\boldsymbol{a} | \exists \boldsymbol{c}, \boldsymbol{l}, \boldsymbol{t} \text{ such that} \\ &\hat{\mathcal{R}}\boldsymbol{t} = \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}, \\ &\hat{\pi}\boldsymbol{t} = \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \boldsymbol{l}, \\ &\hat{\lambda}\boldsymbol{c} = \boldsymbol{0}, \\ &\boldsymbol{a_l} \le \boldsymbol{a} \le \boldsymbol{a_u}, \\ &\boldsymbol{0} \le \boldsymbol{t} \le \boldsymbol{u_T} - \boldsymbol{l_T}, \\ &\boldsymbol{0} \le \boldsymbol{l} < \hat{C}\boldsymbol{1}\big\},\end{aligned} \tag{8}$$

where $\boldsymbol{0} \le \boldsymbol{p} < \hat{P}\boldsymbol{1}$. For instance, we could evaluate Equation (8) to obtain $\Delta_p(B)$ for the example in Figure 1. If we assume a `(CYCLIC(256), CYCLIC(256))` distribution (instead of the `(BLOCK, BLOCK)` distribution shown in the figure), this turns out to be

$$\begin{aligned}\Delta_p(B) = \Big\{ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} | \exists \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \begin{pmatrix} l_1 \\ l_2 \end{pmatrix}, \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \text{ such that} \\ &\lceil -\tfrac{255}{512} - \tfrac{p_1}{2} \rceil \le c_1 \le \lfloor \tfrac{1023}{512} - \tfrac{p_1}{2} \rfloor, \\ &\lceil -\tfrac{255}{512} - \tfrac{p_2}{2} \rceil \le c_2 \le \lfloor \tfrac{1023}{512} - \tfrac{p_2}{2} \rfloor, \\ &\max\big(0, -256(2c_1 + p_1)\big) \le l_1 \le \min(255, 1023 - 512c_1 - 256p_1), \\ &\max\big(0, -256(2c_2 + p_2)\big) \le l_2 \le \min(255, 1023 - 512c_2 - 256p_2), \\ &\max(0, 512c_1 + l_1 + 256p_1) \le t_1 \le \min(1023, 512c_1 + l_1 + 256p_1), \\ &\max(0, 512c_2 + l_2 + 256p_2) \le t_2 \le \min(1023, 512c_2 + l_2 + 256p_2), \\ &\max(1, 1 + t_1) \le a_1 \le \min(1024, 1 + t_1), \\ &\max(1, 1 + t_2) \le a_2 \le \min(1024, 1 + t_2)\Big\}\end{aligned}$$

where

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \le \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} < \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

3 Ownership Sets Revisited

To solve Equation (8) symbolically, we apply the Fourier-Motzkin Elimination (FME) technique [Ban93], eliminating the variables corresponding to the unknown vectors $\boldsymbol{c}$, $\boldsymbol{l}$, $\boldsymbol{t}$ and $\boldsymbol{a}$. During the elimination process, the vector $\boldsymbol{p}$, which denotes the processor's identity in a Cartesian processor mesh, is treated like any other constant; it therefore manifests in the bound expressions of the solution system. When resolved at run time to a particular processor's identity, the solution system will represent that processor's ownership set. Figure 4

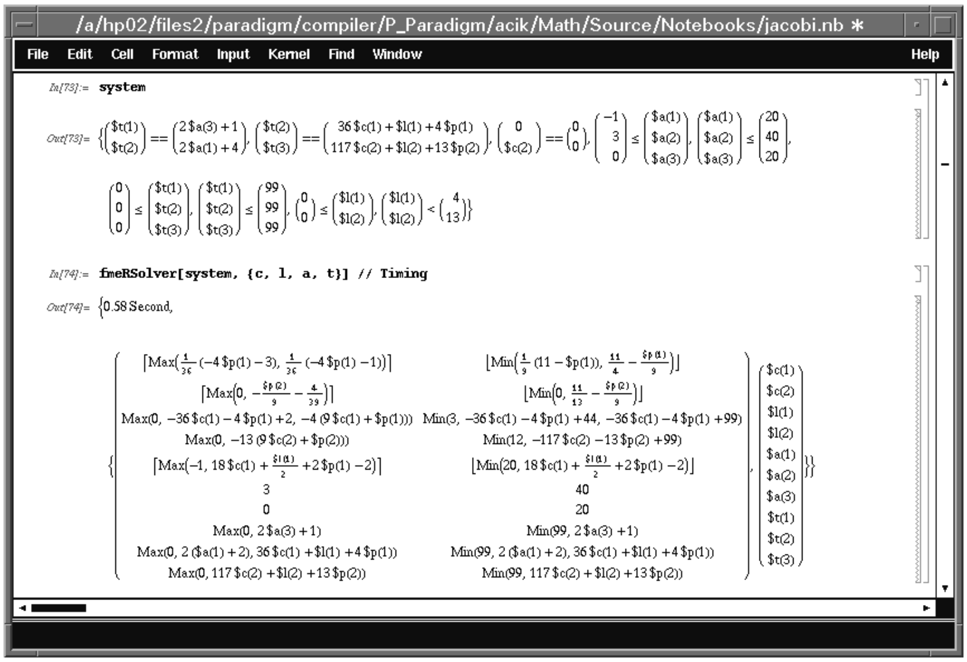

FIGURE 4: An FME Solver in *Mathematica*

shows the outcome when the FME technique is applied on the system corresponding to $\Delta_p(A)$ for the example in Figure 3. We represent the chosen elimination order by the *elimination vector* $\boldsymbol{\epsilon}$. The elimination vector $\boldsymbol{\epsilon}$ is a column vector whose last element denotes the first unknown to be eliminated and whose first element denotes the last unknown to be eliminated. In Figure 4, $\boldsymbol{\epsilon}^T = (\boldsymbol{c}^T, \boldsymbol{l}^T, \boldsymbol{a}^T, \boldsymbol{t}^T)$.

Note that the actual region of interest is the set of points corresponding to the solution vector $\boldsymbol{a}$. We can scan the elements of $\Delta_p(A)$ by constructing a loop nest in which the outermost loop corresponds to `$c(1)` (i.e., c_1) having $\lceil \max\big((-4\texttt{\$p(1)} - 3)/36, (-4\texttt{\$p(1)} - 1)/36\big) \rceil$ as its lower bound and $\lfloor \min\big((11 - \texttt{\$p(1)})/9, (11/4) - (\texttt{\$p(1)}/9)\big) \rfloor$ as its upper bound. Nested within `$c(1)` are the loops corresponding to `$c(2)`, `$l(1)`, `$l(2)`, `$a(1)`, `$a(2)`, `$a(3)`, `$t(1)`, `$t(2)`, `$t(3)` and in that order. Thus, the innermost loop that corresponds to `$t(3)` (i.e., t_3) would have $\max(0, 117\texttt{\$c(2)} + \texttt{\$l(2)} + 13\texttt{\$p(2)})$ as its lower bound and $\min(99, 117\texttt{\$c(2)} + \texttt{\$l(2)} + 13\texttt{\$p(2)})$ as its upper bound. Within the body of such a loop nest, `$a(1)` and `$a(2)` (i.e., a_1 and a_2) will be the subscripts of the elements of the array `A` owned by processor $\boldsymbol{p}$.

3.1 Refinement

Since the elements of $\boldsymbol{c}$, $\boldsymbol{l}$ and $\boldsymbol{t}$ serve only as auxiliary variables in Equation (8), an important question that crops up at this juncture is whether formulations for $\Delta_p(X)$ exist that require fewer auxiliary variables. A lesser number of unknowns translates to a smaller loop nest depth and hence a more efficient scan of $\Delta_p(X)$. This is especially important when compute sets derived from ownership sets are used to partition loops [JB99]. Reducing the number of unknowns also improves the overall timing of the FME solver.

Lemma 3.1: Consider the following definition for $\Delta'_p(X)$:

$$\begin{aligned}\Delta'_p(X) = \{\boldsymbol{a} \mid \exists \boldsymbol{c}, \boldsymbol{t} \text{ such that } & \\ & \hat{\mathcal{R}}\boldsymbol{t} = \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}, \\ & \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} \leq \hat{\pi}\boldsymbol{t} < \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1}, \\ & \hat{\lambda}\boldsymbol{c} = \boldsymbol{0}, \\ & \boldsymbol{a_l} \leq \boldsymbol{a} \leq \boldsymbol{a_u}, \\ & \boldsymbol{0} \leq \boldsymbol{t} \leq \boldsymbol{u_T} - \boldsymbol{l_T}\},\end{aligned} \tag{9}$$

where $\boldsymbol{0} \leq \boldsymbol{p} < \hat{P}\boldsymbol{1}$. Then,

$$\Delta_p(X) = \Delta'_p(X).$$

Proof.
Suppose $\boldsymbol{a'} \in \Delta'_p(X)$. Then there exists a $\boldsymbol{c'}$, $\boldsymbol{t'}$ such that $\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \leq \hat{\pi}\boldsymbol{t'} < \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1}$. Hence,

$$\boldsymbol{0} \leq \hat{\pi}\boldsymbol{t'} - (\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p}) < \hat{C}\boldsymbol{1}.$$

Setting $\boldsymbol{l'} = \hat{\pi}\boldsymbol{t'} - (\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p})$, we get $\hat{\pi}\boldsymbol{t'} = \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} + \boldsymbol{l'}$. Consequently, $\boldsymbol{a'} \in \Delta_p(X)$ because $\boldsymbol{c'}$, $\boldsymbol{l'}$, $\boldsymbol{t'}$ together with $\boldsymbol{a'}$ satisfy the system in Equation (8). Hence,

$$\Delta'_p(X) \subseteq \Delta_p(X). \tag{3.1.1}$$

Similarly, if $\boldsymbol{a} \in \Delta_p(X)$, then there exists a $\boldsymbol{c}$, $\boldsymbol{l}$, and $\boldsymbol{t}$ that together with $\boldsymbol{a}$ satisfy the system in Equation (8). Since

$$\begin{aligned}&\hat{\pi}\boldsymbol{t} = \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \boldsymbol{l}, \\ &\boldsymbol{0} \leq \boldsymbol{l} < \hat{C}\boldsymbol{1},\end{aligned}$$

we get $\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} \leq \hat{\pi}\boldsymbol{t} < \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1}$. Thus, $\boldsymbol{a} \in \Delta'_p(X)$ because $\boldsymbol{c}$, $\boldsymbol{a}$ and $\boldsymbol{t}$ satisfy the system in Equation (9). Therefore,

$$\Delta_p(X) \subseteq \Delta'_p(X). \tag{3.1.2}$$

Relations (3.1.1) and (3.1.2) therefore imply that $\Delta'_p(X) = \Delta_p(X)$.

The first improvement that can be done is removing the offsets vector $\boldsymbol{l}$. This can be accomplished in a straightforward manner and Lemma 3.1 shows how. The system in Lemma 3.1 is an improvement over that in Equation (8) because there is a reduction in the number of unknowns (by $\|\boldsymbol{l}\|$[c]) and in the number of inequalities (by $2\|\boldsymbol{l}\|$).

3.1.1 Removing t

A key observation to make in Equation (9) is that the replicated dimensions of a template do not affect the elements of $\boldsymbol{a}$. This is because, by premultiplying $\boldsymbol{t}$ with $\hat{\mathcal{R}}$, rows that correspond to replicated dimensions in $\boldsymbol{t}$ get elided. What is a "replicated dimension"? We refer to those dimensions of a template that contain a $*$ or an unmatched dummy variable in the alignment specification as the template's *replicated dimensions* (see [KLS+94]).The remaining dimensions are called its *aligned dimensions.*

The dimensions of a template that are not mapped onto any processor dimension are said to be *collapsed*; the remaining dimensions of the template are referred to as its *distributed dimensions.* If a template dimension is collapsed and not replicated, that template dimension can only affect the values of $\boldsymbol{a}$ through the equality

[c]The notation $\|\boldsymbol{v}\|$ indicates the number of rows in the column vector $\boldsymbol{v}$.

$\hat{\mathcal{R}}\boldsymbol{t} = \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}$ and the constraint $\boldsymbol{0} \le \boldsymbol{t} \le \boldsymbol{u_T} - \boldsymbol{l_T}$. To then find the corresponding elements of $\boldsymbol{a}$, we need only consider the pair

$$\begin{aligned}
&\boldsymbol{0} \le \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \le \boldsymbol{u_T} - \boldsymbol{l_T},\\
&\boldsymbol{a_l} \le \boldsymbol{a} \le \boldsymbol{a_u}.
\end{aligned}$$

Similarly, by considering

$$\begin{aligned}
&\hat{\mathcal{R}}\boldsymbol{t} = \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T},\\
&\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} \le \hat{\pi}\boldsymbol{t} < \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1},
\end{aligned}$$

template dimensions that are both aligned and distributed can be eliminated. Therefore, it is worth investigating whether a new system can be constructed for the ownership set that has a lesser number of unknowns than that required by Equation (9). As Lemma 3.2 will show, such a formulation indeed does exist.

Lemma 3.2: Consider the following definition for $\Delta_p^*(X)$:

$$\begin{aligned}
\Delta_p^*(X) = \big\{\boldsymbol{a} | \exists \boldsymbol{c} \text{ such that}&\\
&\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p}) \le \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}),\\
&\boldsymbol{0} \le \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} \le \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}),\\
&\hat{\lambda}\boldsymbol{c} = \boldsymbol{0},\\
&\boldsymbol{a_l} \le \boldsymbol{a} \le \boldsymbol{a_u},\\
&\boldsymbol{0} \le \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \le \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T})\big\},
\end{aligned} \tag{10}$$

where $\boldsymbol{0} \le \boldsymbol{p} < \hat{P}\boldsymbol{1}$. Then,

$$\Delta_p'(X) = \Delta_p^*(X).$$

Proof.
Suppose $\boldsymbol{a^*} \in \Delta_p^*(X)$; assume $\boldsymbol{c^*}$ to be the corresponding course vector for which Equation (10) holds for $\boldsymbol{a^*}$. Consider

$$\boldsymbol{t^*} = \hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a^*} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p}).$$

Then,[d]

$$\hat{\mathcal{R}}\boldsymbol{t^*} = \hat{\mathcal{A}}\boldsymbol{a^*} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}. \tag{3.2.1}$$

Next, from Equation (10), we have after premultiplying throughout by $\hat{\pi}$

$$\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p}) \le \hat{\pi}\big(\boldsymbol{t^*} - (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p})\big) \le \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

This gives us

$$\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p} \le \hat{\pi}\boldsymbol{t^*} \le \hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p} + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Now, $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T \le \hat{I}$. Since $\boldsymbol{0} \le \hat{C}\boldsymbol{1} - \boldsymbol{1}$, we get $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}) \le \hat{C}\boldsymbol{1} - \boldsymbol{1}$. Hence,

$$\hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p} \le \hat{\pi}\boldsymbol{t^*} < \hat{C}\hat{P}\boldsymbol{c^*} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1}. \tag{3.2.2}$$

[d] Note that $\hat{\mathcal{R}}\hat{\mathcal{R}}^T = \hat{I}$ where $\hat{I}$ is the $m \times m$ identity matrix. However, $\hat{\mathcal{R}}^T\hat{\mathcal{R}}$ is a $y \times y$ matrix in which the only nonzero elements are the principal diagonal elements; the ith principal diagonal element is 1 if and only if the ith template dimension is an aligned dimension. Similarly, $\hat{\pi}\hat{\pi}^T$ is the $z \times z$ identity matrix whereas $\hat{\pi}^T\hat{\pi}$ is a $y \times y$ matrix in which the only nonzero elements are the principal diagonal elements; the ith principal diagonal element is 1 if and only if the ith template dimension is a distributed dimension. Square matrices in which the only nonzero elements are those that reside on the principal diagonal are usually termed *square diagonal*.

From Equation (10), we also have $\mathbf{0} \le \hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \le \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T})$. Thus, $\mathbf{0} \le \hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T})$. That is,

$$(\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}) \le \hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}),$$
$$\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}) \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T}) + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}).$$

Since $\mathbf{0} \le \hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}$ and because $\hat{0} \le \hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}}$, we get $\mathbf{0} \le (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p})$. Taking this and Equation (3.2.1) into account, the previous inequalities become

$$\mathbf{0} \le \boldsymbol{t}^* \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T}) + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}).$$

But from Equation (10), $(\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}) \le (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T\hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T})$. Thus,

$$\mathbf{0} \le \boldsymbol{t}^* \le \big(\hat{\mathcal{R}}^T\hat{\mathcal{R}} + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T\hat{\pi}\big)(\boldsymbol{u_T} - \boldsymbol{l_T}).$$

Again, since $\hat{0} \le \hat{\mathcal{R}}^T\hat{\mathcal{R}} + (\hat{I} - \hat{\mathcal{R}}^T\hat{\mathcal{R}})\hat{\pi}^T\hat{\pi} \le \hat{I}$ and because $\boldsymbol{u_T} - \boldsymbol{l_T} \ge \mathbf{0}$, the previous constraint becomes

$$\mathbf{0} \le \boldsymbol{t}^* \le \boldsymbol{u_T} - \boldsymbol{l_T}. \tag{3.2.3}$$

Also from Equation (10), we have

$$\hat{\lambda}\boldsymbol{c}^* = \mathbf{0}, \tag{3.2.4}$$
$$\boldsymbol{a_l} \le \boldsymbol{a}^* \le \boldsymbol{a_u}. \tag{3.2.5}$$

Thus, because Equations (3.2.1) and (3.2.4) and Inequalities (3.2.2), (3.2.3) and (3.2.5) satisfy the system in Equation (9), $\boldsymbol{a}^* \in \Delta'_p(X)$. Hence,

$$\Delta^*_p(X) \subseteq \Delta'_p(X). \tag{3.2.6}$$

Similarly, if $\boldsymbol{a'} \in \Delta'_p(X)$, then there exists a $\boldsymbol{c'}$, $\boldsymbol{t'}$ that together with $\boldsymbol{a'}$ satisfy the system in Equation (9). Therefore, $\hat{\mathcal{R}}\boldsymbol{t'} = \hat{\mathcal{A}}\boldsymbol{a'} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}$. Hence,

$$\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\boldsymbol{t'} = \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a'} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}).$$

Since from Equation (9) $\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \le \hat{\pi}\boldsymbol{t'} < \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} + \hat{C}\mathbf{1}$, we get

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p}) \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\boldsymbol{t'} \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} + \hat{C}\mathbf{1} - \mathbf{1}).$$

Now $\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi} = \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}$.[e] Therefore,

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p}) \le \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a'} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \le \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} + \hat{C}\mathbf{1} - \mathbf{1}). \tag{3.2.7}$$

From Equation (9), we also have the following:

$$\hat{\pi}\boldsymbol{t'} - \hat{C}\mathbf{1} < \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \le \hat{\pi}\boldsymbol{t'},$$
$$\mathbf{0} \le \hat{\pi}\boldsymbol{t'} \le \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}).$$

From the above two inequalities, we get

$$-\hat{C}\mathbf{1} + \mathbf{1} \le \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \le \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}).$$

Consider the inequality $-\hat{C}\mathbf{1} + \mathbf{1} \le \hat{C}(\hat{P}\boldsymbol{c'} + \boldsymbol{p})$. Since $\hat{C}$ always has a well-defined inverse $\hat{C}^{-1}$, and because $\hat{C}^{-1} > \hat{0}$ is always true, we can premultiply $-\hat{C}\mathbf{1} + \mathbf{1} \le \hat{C}(\hat{P}\boldsymbol{c'} + \boldsymbol{p})$ throughout by $\hat{C}^{-1}$:

$$\hat{C}^{-1}\mathbf{1} - \mathbf{1} - \boldsymbol{p} \le \hat{P}\boldsymbol{c'}.$$

Since $\boldsymbol{p} < \hat{P}\mathbf{1}$ and because $\hat{P}$ also has a well-defined inverse $\hat{P}^{-1}$ that satisfies $\hat{P}^{-1} > \hat{0}$, we get

$$\hat{P}^{-1}(\hat{C}^{-1}\mathbf{1} - \mathbf{1} - \hat{P}\mathbf{1} + \mathbf{1}) \le \hat{P}^{-1}\hat{P}\boldsymbol{c'}.$$

[e] If $\hat{A}$ and $\hat{B}$ are two conforming square diagonal matrices, then the commutative law for matrix multiplication holds (i.e., $\hat{A}\hat{B} = \hat{B}\hat{A}$).

The last inequality simplifies to

$$\hat{P}^{-1}\hat{C}^{-1}\mathbf{1} - \mathbf{1} \leq \boldsymbol{c'}.$$

This clearly implies that the elements of $\boldsymbol{c'}$ are restricted to nonnegative integers only. Taken along with the fact that $\boldsymbol{p} \geq \mathbf{0}$, we get $\hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \geq \mathbf{0}$. Thus,

$$\mathbf{0} \leq \hat{C}\hat{P}\boldsymbol{c'} + \hat{C}\boldsymbol{p} \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}). \tag{3.2.8}$$

Since from Equation (9) $\mathbf{0} \leq \boldsymbol{t'} \leq \boldsymbol{u_T} - \boldsymbol{l_T}$, we get $\mathbf{0} \leq \hat{\mathcal{R}}\boldsymbol{t'} \leq \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T})$. Therefore,

$$\mathbf{0} \leq \hat{\mathcal{A}}\boldsymbol{a'} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \leq \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T}). \tag{3.2.9}$$

Again from Equation (9), we have

$$\hat{\lambda}\boldsymbol{c'} = \mathbf{0}, \tag{3.2.10}$$

$$\boldsymbol{a_l} \leq \boldsymbol{a'} \leq \boldsymbol{a_u}. \tag{3.2.11}$$

Since Equation (3.2.10) and Inequalities (3.2.7), (3.2.8), (3.2.9) and (3.2.11) fulfill the system in Equation (10), $\boldsymbol{a'} \in \Delta_p^*(X)$. Thus,

$$\Delta'_p(X) \subseteq \Delta_p^*(X). \tag{3.2.12}$$

Relations (3.2.6) and (3.2.12) taken together imply that $\Delta'_p(X) = \Delta_p^*(X)$.

The new system defined in Equation (10) has exactly the same number of inequalities as the system in Equation (9); however, the gain is a reduction of $||\boldsymbol{t}||$ in the number of unknowns.

3.1.2 Further Refinements

From the standpoint of efficiently scanning the ownership set, the system in Equation (10) has further scope for improvement. If we were to consider those processor dimensions j along which replicated template dimensions are distributed, the corresponding c_js are only constrained by $\mathbf{0} \leq \hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{p} \leq \hat{\pi}(\boldsymbol{u_T}-\boldsymbol{l_T})$. This means that for a given $\boldsymbol{a}$, there could exist more than one $\boldsymbol{c}$ for which the system in Equation (10) holds. Hence, if the solution system for Equation (10) were used to scan the ownership set, members could get enumerated more than once. To remedy this problem, we extend the system in Equation (10) by the equation $(\hat{I} - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c} = \mathbf{0}$. The following lemma ensures the validity of such a transformation.

Lemma 3.3: Consider the following definition for $\Delta''_p(X)$:

$$\begin{aligned}\Delta''_p(X) = \big\{\boldsymbol{a} | \exists \boldsymbol{c} \text{ such that} \\ &\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\mathbf{1} - \mathbf{1}), \\ &\mathbf{0} \leq \hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}), \\ &(\hat{I} - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c} = \mathbf{0}, \\ &\hat{\lambda}\boldsymbol{c} = \mathbf{0}, \\ &\boldsymbol{a_l} \leq \boldsymbol{a} \leq \boldsymbol{a_u}, \\ &\mathbf{0} \leq \hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \leq \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T})\big\},\end{aligned} \tag{11}$$

where, $\mathbf{0} \leq \boldsymbol{p} < \hat{P}\mathbf{1}$. Then, for every $\boldsymbol{a} \in \Delta''_p(X)$, there exists exactly one $\boldsymbol{c}$ that satisfies the system in Equation (11).

Proof.
Suppose $\boldsymbol{a''} \in \Delta''_p(X)$. Therefore, there exists a $\boldsymbol{c''}$ that together with $\boldsymbol{a''}$ fulfills the system in Equation (11). Since every equality and inequality in Equation (10) occurs in Equation (11), $\boldsymbol{c''}$ and $\boldsymbol{a''}$ also satisfy the system in Equation (10). Therefore, $\boldsymbol{a''} \in \Delta_p^*(X)$. Hence,

$$\Delta''_p(X) \subseteq \Delta_p^*(X). \tag{3.3.1}$$

Consider an $\boldsymbol{a}^* \in \Delta_p^*(X)$. Thus, there exists a $\boldsymbol{c}^*$ such that

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Premultiplying throughout by $\hat{\pi}$, and after rearranging, we get

$$\hat{C}\hat{P}\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}^* + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{p} \leq \hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{C}\hat{P}\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}^* + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Let

$$\boldsymbol{c''} = \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}^*. \tag{3.3.2}$$

Therefore, after premultiplying by $\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$, the previous constraint becomes

$$\begin{aligned}
&\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\hat{P}\boldsymbol{c''} + \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{p} \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}),\\
&\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\hat{P}\boldsymbol{c''} + \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).
\end{aligned}$$

Now

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T = \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T. \tag{3.3.3}$$

Also,

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T = \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T. \tag{3.3.4}$$

The last pair of inequalities therefore become

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c''} + \hat{C}\boldsymbol{p}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c''} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}). \tag{3.3.5}$$

From Equation (10), we have $\boldsymbol{0} \leq \hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}$. Premultiplying throughout by $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$ and using Equation (3.3.2), we get

$$\boldsymbol{0} \leq \hat{C}\hat{P}\boldsymbol{c''} + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{p}.$$

But $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{p} \leq \hat{C}\boldsymbol{p}$. Hence,

$$\boldsymbol{0} \leq \hat{C}\hat{P}\boldsymbol{c''} + \hat{C}\boldsymbol{p}. \tag{3.3.6}$$

Again from Equation (10),

$$\hat{C}\hat{P}\boldsymbol{c}^* \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}. \tag{3.3.7}$$

Premultiplying throughout by $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$ and using Equation (3.3.2), we get

$$\hat{C}\hat{P}\boldsymbol{c''} \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\big(\hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}\big). \tag{3.3.8}$$

Since $\boldsymbol{0} \leq \hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{p}$, we also obtain $-\hat{P}^{-1}\boldsymbol{p} \leq \boldsymbol{c}^*$. But $\boldsymbol{p} \leq \hat{P}\boldsymbol{1} - \boldsymbol{1}$. Therefore,

$$-\boldsymbol{1} + \hat{P}^{-1}\boldsymbol{1} \leq \boldsymbol{c}^*,$$

which means that

$$\boldsymbol{0} \leq \boldsymbol{c}^*. \tag{3.3.9}$$

Using Inequality (3.3.9) in Inequality (3.3.7), we get

$$\boldsymbol{0} \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}.$$

Because $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T \leq \hat{I}$, the last inequality leads to $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\big(\hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}\big) \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}$. Hence, from Inequality (3.3.8), $\hat{C}\hat{P}\boldsymbol{c''} \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}) - \hat{C}\boldsymbol{p}$. Combining this with Inequality (3.3.6), we therefore get

$$\boldsymbol{0} \leq \hat{C}\hat{P}\boldsymbol{c''} + \hat{C}\boldsymbol{p} \leq \hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T}). \tag{3.3.10}$$

Using Equation (3.3.2), we also get

$$(\hat{I} - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c''} = (\hat{I} - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}^* = (\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c}^* = \boldsymbol{0}. \tag{3.3.11}$$

From Equation (10), we have $\hat{\lambda}\boldsymbol{c}^* = \boldsymbol{0}$. Premultiplying by $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$ and by using Equation (3.3.2), we get

$$\hat{\lambda}\boldsymbol{c''} = \boldsymbol{0}. \tag{3.3.12}$$

From Equation (10), we get the remaining constraints satisfied by $\boldsymbol{a}^*$:

$$\boldsymbol{a_l} \leq \boldsymbol{a}^* \leq \boldsymbol{a_u}, \tag{3.3.13}$$

$$\boldsymbol{0} \leq \hat{\mathcal{A}}\boldsymbol{a}^* + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T} \leq \hat{\mathcal{R}}(\boldsymbol{u_T} - \boldsymbol{l_T}). \tag{3.3.14}$$

Since Equations (3.3.11) and (3.3.12) and Inequalities (3.3.5), (3.3.10), (3.3.13) and (3.3.14) fulfill the system in Equation (11), $\boldsymbol{a}^* \in \Delta_p''(X)$. Therefore,

$$\Delta_p^*(X) \subseteq \Delta_p''(X). \tag{3.3.15}$$

Relations (3.3.1) and (3.3.15) taken together imply that $\Delta_p^*(X) = \Delta_p''(X)$.

We now need to prove the second part of the lemma: For every $\boldsymbol{a} \in \Delta_p''(X)$, there exists exactly one $\boldsymbol{c}$ for which Equation (11) holds. Suppose that for an $\boldsymbol{a} \in \Delta_p''(X)$, $\boldsymbol{c}$ satisfies the system in Equation (11). Hence

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Premultiplying throughout by $\hat{\pi}$, and because $(\hat{I} - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi})\boldsymbol{c} = \boldsymbol{0}$ by Equation (11), we get

$$\hat{C}\hat{P}\boldsymbol{c} \leq \boldsymbol{\kappa} \leq \hat{C}\hat{P}\boldsymbol{c} + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}), \tag{3.3.16}$$

where

$$\boldsymbol{\kappa} = \hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) - \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{p}.$$

If there exists another course vector $\boldsymbol{c}^*$ that together with $\boldsymbol{a}$ also fulfills the system in Equation (11), then $\boldsymbol{c}^*$ would also satisfy Inequality (3.3.16). That is,

$$\hat{C}\hat{P}\boldsymbol{c}^* \leq \boldsymbol{\kappa} \leq \hat{C}\hat{P}\boldsymbol{c}^* + \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}). \tag{3.3.17}$$

Using Inequalities (3.3.16) and (3.3.17), we get

$$-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}) \leq \hat{C}\hat{P}(\boldsymbol{c}^* - \boldsymbol{c}) \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}). \tag{3.3.18}$$

Now $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$ is a square diagonal matrix in which the principal diagonal elements are either 0 or 1. Hence, from Inequality (3.3.18), we can conclude that $\boldsymbol{c} = \boldsymbol{c}^*$. That is, for every $\boldsymbol{a} \in \Delta_p''(X)$, there exists exactly one $\boldsymbol{c}$ for which Equation (11) holds.

An important implication of Lemma 3.3 is that if the FME technique is now applied to the system in Equation (11), then, whatever be the order of elimination of the unknown variables (corresponding to the elements of $\boldsymbol{c}$ and $\boldsymbol{a}$), the associated loop nest will scan every member of the ownership set (i.e., $\boldsymbol{a}$) exactly once. To see why this is so, let $\boldsymbol{\epsilon}$ represent one such elimination order. Suppose $\theta = \|\boldsymbol{\epsilon}\| = x + z$ where x and z denote the number of dimensions of the alignee and processor mesh respectively. The integer bound expressions returned by the FME solver can be used to construct a loop nest that scans $\Delta_p''(X)$. The outermost loop in this nest matches ϵ_1, while the innermost loop matches ϵ_θ. Consider an iteration point $\boldsymbol{\rho}$ of such a loop nest and let $\boldsymbol{\varrho}$ be any other iteration point of the same loop nest. Thus, $\boldsymbol{\rho}$ and $\boldsymbol{\varrho}$ also represent solutions to the system in Equation (11). Since every iteration point of a loop nest is distinct, let $\boldsymbol{\rho}$ and $\boldsymbol{\varrho}$ differ in the ith position. If $\epsilon_i \equiv a_n$, then the $\boldsymbol{a}$ that corresponds to $\boldsymbol{\rho}$ obviously differs from the $\boldsymbol{a}$ that corresponds to $\boldsymbol{\varrho}$. If instead $\epsilon_i \equiv c_j$, then the $\boldsymbol{c}$ that corresponds to $\boldsymbol{\rho}$ differs from the $\boldsymbol{c}$ that corresponds to $\boldsymbol{\varrho}$. But from Lemma (3.3), only one $\boldsymbol{c}$ can satisfy the system for a given $\boldsymbol{a}$. Thus, the corresponding values for $\boldsymbol{a}$ in $\boldsymbol{\rho}$ and $\boldsymbol{\varrho}$ must also be different. That is, the $\boldsymbol{a}$ associated with the iteration point $\boldsymbol{\rho}$ must be different from the $\boldsymbol{a}$ associated with any other iteration point $\boldsymbol{\varrho}$ of the same loop nest. In other words, every member of the ownership set gets enumerated exactly once.

3.1.3 Comparisons

From a mathematical perspective, the systems in Equations (8), (9), (10), and (11) are all identical—they refer to the same ownership set in all cases. However, they differ in such metrics as the number of inequalities to be solved, the number of variables to be eliminated (consequently, the number of auxiliary parameters

Table 1: Comparisons of the Ownership Set Formulations

System	Number of inequalities[f]	Number of unknowns[g]
Equation (8)	$2(x+y+3z+m)$	$x+y+2z$
Equation (9)	$2(x+y+2z+m)$	$x+y+z$
Equation (10)	$2(x+y+2z+m)$	$x+z$
Equation (11)	$2(x+y+3z+m)$	$x+z$

required to scan the set), and finally, in the way the members of the set get enumerated. Table 1 summarizes these metrics. As can be seen from the table, the original system in Equation (8) requires the elimination of the largest number of variables and handles the largest number of inequalities. If a loop nest were generated from its FME solution system to scan $\Delta_p(X)$, it is not guaranteed that every member will get enumerated exactly once. This inability to ensure the "uniqueness" of each enumerated member has serious repercussions if the ownership set is used to generate other sets. For instance, we could use the compute sets defined in [ACIK95] to handle loops qualified by the `INDEPENDENT` directive. In [ACIK95], these sets were defined using a formulation of the ownership set similar to that in Equation (8). If the FME solution system to such a compute set formulation were scanned, certain iterations in the set could get enumerated more than once for certain alignment/distribution combinations. However, if the formulation in Equation (11) is used, this problem can be avoided. The FME approach that we adopt to solve for these sets is quite different from the approach in [ACIK95] where a parametric solution based on the Hermite Normal Form was exploited for the same purpose.

In Figure 5, we show the result of applying the FME technique to the formulation in Equation (11). Though the system has the same number of inequalities as the formulation in Equation (8), the number of unknowns to be eliminated is lesser by 5 variables, and this results in a significant reduction in the time required to solve the system from 0.58 seconds to 0.25 seconds.

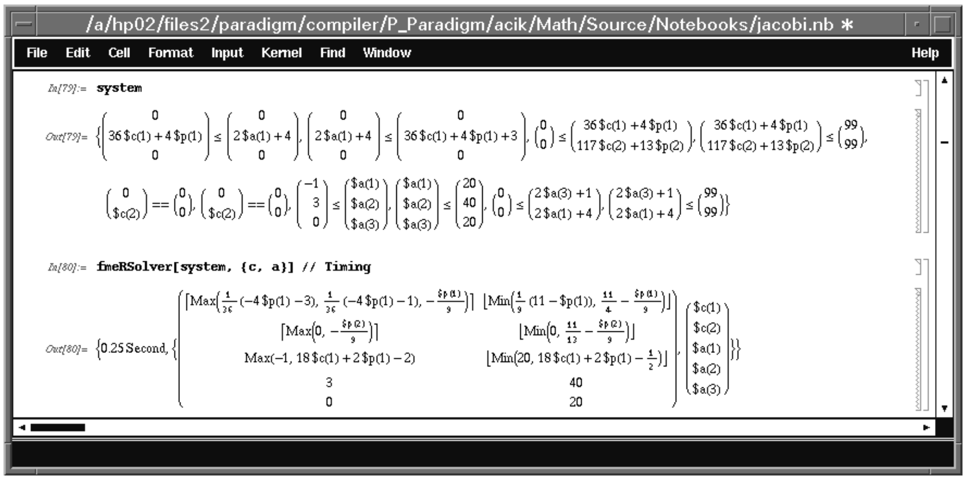

Figure 5: The FME Solver Applied on the Modified Ownership Set Formulation

[f]When the FME technique is applied, equalities of the form $f(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a}) = g(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a})$ get replaced by a pair of inequalities of the form $f(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a}) \leq g(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a})$ and $f(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a}) \geq g(\boldsymbol{c},\boldsymbol{l},\boldsymbol{t},\boldsymbol{a})$.

[g]To recapitulate, x is the number of dimensions of the alignee, y is the number of dimensions of the template, z is the number of dimensions of the processor mesh, and m is the number of aligned dimensions of the template.

3.2 An Equivalence Relation

It is interesting to enquire into the nature of ownership sets across processors. That is, for an arbitrary alignment/distribution combination, can these sets partially overlap? Or, are they equal or disjoint? The answers to these questions can be used to devise an efficient run-time test that avoids redundant communication due to array replication (see § 4).

Lemma 3.4: If $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q})=\mathbf{0}$, and $\Delta''_p(X)\neq\emptyset$, $\Delta''_q(X)\neq\emptyset$, then $\Delta''_p(X)=\Delta''_q(X)$.

Proof.
Since $\Delta''_p(X)\neq\emptyset$, consider an element $\boldsymbol{a''}\in\Delta''_p(X)$. Then there exists a $\boldsymbol{c}$ such that

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{p})\leq\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a''}+\boldsymbol{s_0}-\hat{\mathcal{R}}\boldsymbol{l_T})\leq\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{p}+\hat{C}\mathbf{1}-\mathbf{1}).$$

Premultiplying the above with $\hat{\pi}$ and since $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}=\boldsymbol{c}$, $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{p}=\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{q}$, we get after some rearrangement:

$$\hat{C}\hat{P}\boldsymbol{c}+\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{q}\leq\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a''}+\boldsymbol{s_0}-\hat{\mathcal{R}}\boldsymbol{l_T})\leq\hat{C}\hat{P}\boldsymbol{c}+\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{q}+\hat{C}\mathbf{1}-\mathbf{1}).$$

Premultiplying by $\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$, and by using Equations (3.3.3) and (3.3.4), we can simplify the above to

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{q})\leq\hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a''}+\boldsymbol{s_0}-\hat{\mathcal{R}}\boldsymbol{l_T})\leq\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{q}+\hat{C}\mathbf{1}-\mathbf{1}).\tag{3.4.1}$$

We also have $\mathbf{0}\leq\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{p}\leq\hat{\pi}(\boldsymbol{u_T}-\boldsymbol{l_T})$. Premultiplying this by $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$, and by using the fact $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{\pi}=\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}$, this becomes

$$\mathbf{0}\leq\hat{C}\hat{P}\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}+\hat{C}\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{p}\leq\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T}-\boldsymbol{l_T}).$$

Substituting $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}$ by $\boldsymbol{c}$, $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{p}$ by $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{q}$, we get,

$$\mathbf{0}\leq\hat{C}\hat{P}\boldsymbol{c}+\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{q}\leq\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T}-\boldsymbol{l_T}).\tag{3.4.2}$$

Now, since $\Delta''_q(X)\neq\emptyset$, there exists at least one $\boldsymbol{b''}$ in $\Delta''_q(X)$. Let $\boldsymbol{c}^*$ be the corresponding course vector that together with $\boldsymbol{b''}$ satisfies the system in Equation (11). We then have

$$\mathbf{0}\leq\hat{C}\hat{P}\boldsymbol{c}^*+\hat{C}\boldsymbol{q}\leq\hat{\pi}(\boldsymbol{u_T}-\boldsymbol{l_T}).$$

Premultiplying the above throughout by $(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)$, and by using the fact $(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c}^*=\mathbf{0}$, we get

$$\mathbf{0}\leq(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{C}\boldsymbol{q}\leq(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{\pi}(\boldsymbol{u_T}-\boldsymbol{l_T}).\tag{3.4.3}$$

From Inequality (3.4.2), we also have

$$(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{C}\boldsymbol{q}\leq\hat{C}\hat{P}\boldsymbol{c}+\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\hat{C}\boldsymbol{q}+(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{C}\boldsymbol{q}\leq\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}(\boldsymbol{u_T}-\boldsymbol{l_T})+(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\hat{C}\boldsymbol{q}.$$

By using Inequality (3.4.3), the last constraint leads to the following:

$$\mathbf{0}\leq\hat{C}\hat{P}\boldsymbol{c}+\hat{C}\boldsymbol{q}\leq\hat{\pi}(\boldsymbol{u_T}-\boldsymbol{l_T}).\tag{3.4.4}$$

We also have

$$(\hat{I}-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T)\boldsymbol{c}=\mathbf{0},\tag{3.4.5}$$

$$\hat{\lambda}\boldsymbol{c}=\mathbf{0},\tag{3.4.6}$$

$$\boldsymbol{a_l}\leq\boldsymbol{a''}\leq\boldsymbol{a_u},\tag{3.4.7}$$

$$\mathbf{0}\leq\hat{\mathcal{A}}\boldsymbol{a''}+\boldsymbol{s_0}-\hat{\mathcal{R}}\boldsymbol{l_T}\leq\hat{\mathcal{R}}(\boldsymbol{u_T}-\boldsymbol{l_T}).\tag{3.4.8}$$

Since Equations (3.4.5) and (3.4.6), and Inequalities (3.4.1), (3.4.4), (3.4.7) and (3.4.8) fulfill the system in Equation (11) for $\Delta''_q(X)$, we have $\boldsymbol{a''}\in\Delta''_q(X)$. Therefore,

$$\Delta''_p(X)\subseteq\Delta''_q(X).\tag{3.4.9}$$

Beginning with $\Delta_q''(X)$, we can similarly show that

$$\Delta_q''(X) \subseteq \Delta_p''(X). \tag{3.4.10}$$

Relations (3.4.9) and (3.4.10) taken together imply that $\Delta_p''(X) = \Delta_q''(X)$.

The expression $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T$ is a square diagonal matrix of size $z \times z$. It is easy to see that the principal diagonal elements of this matrix are either 0 or 1. It is also easy to see that the jth principal diagonal element is 0 if and only if the template dimension distributed on the jth processor dimension is a replicated one. To review, replicated dimensions are those dimensions of a template that contain either a `*` or an unmatched dummy variable in the alignment specification. Thus, if an array `X` is aligned to a template `T` that is then distributed onto a processor mesh `P`, Lemma 3.4 states that the ownership sets with respect to `X` of two processors $\boldsymbol{p}$ and $\boldsymbol{q}$ in `P` will be the same if (1) $\boldsymbol{p}$ and $\boldsymbol{q}$ own at least one element of the alignee array `X`, and (2) $\boldsymbol{p}$ and $\boldsymbol{q}$ match in at least those dimensions on which the aligned dimensions of the template `T` are distributed.

Lemma 3.5: If $\Delta_p''(X) \cap \Delta_q''(X) \neq \emptyset$, then $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q}) = \boldsymbol{0}$.

Proof.
Let $\boldsymbol{a''} \in \Delta_p''(X) \cap \Delta_q''(X)$. Since $\boldsymbol{a''} \in \Delta_p''(X)$, there exists a $\boldsymbol{c}$ such that

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a''} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c} + \hat{C}\boldsymbol{p} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Also, since $\boldsymbol{a''} \in \Delta_q''(X)$, there exists a $\boldsymbol{c}^*$ such that

$$\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{q}) \leq \hat{\pi}^T\hat{\pi}\hat{\mathcal{R}}^T(\hat{\mathcal{A}}\boldsymbol{a''} + \boldsymbol{s_0} - \hat{\mathcal{R}}\boldsymbol{l_T}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\hat{P}\boldsymbol{c}^* + \hat{C}\boldsymbol{q} + \hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Thus,

$$-\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\big(\hat{C}\hat{P}(\boldsymbol{c}^* - \boldsymbol{c}) + \hat{C}(\boldsymbol{q} - \boldsymbol{p})\big) \leq \hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}).$$

Premultiplying the above with $\hat{\pi}$ and since $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c} = \boldsymbol{c}$, $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T\boldsymbol{c}^* = \boldsymbol{c}^*$, we get

$$-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}) \leq \hat{C}\big(\hat{P}(\boldsymbol{c}^* - \boldsymbol{c}) + \boldsymbol{\varphi}\big) \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{C}\boldsymbol{1} - \boldsymbol{1}), \tag{3.5.1}$$

where $\boldsymbol{\varphi} = \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{q} - \boldsymbol{p})$. From Inequality (3.5.1), we thus get

$$\hat{P}(\boldsymbol{c}^* - \boldsymbol{c}) + \boldsymbol{\varphi} = \boldsymbol{0},$$

or

$$\boldsymbol{\varphi} = \hat{P}(\boldsymbol{c} - \boldsymbol{c}^*). \tag{3.5.2}$$

Since $\hat{0} \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T \leq \hat{I}$ and $\boldsymbol{0} \leq \boldsymbol{p}, \boldsymbol{q} < \hat{P}\boldsymbol{1}$, we also have the following constraint for $\boldsymbol{\varphi}$:

$$-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{P}\boldsymbol{1} - \boldsymbol{1}) \leq \boldsymbol{\varphi} \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{P}\boldsymbol{1} - \boldsymbol{1}).$$

Using the last constraint and Equation (3.5.2), we therefore get

$$-\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{P}\boldsymbol{1} - \boldsymbol{1}) \leq \hat{P}(\boldsymbol{c}^* - \boldsymbol{c}) \leq \hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\hat{P}\boldsymbol{1} - \boldsymbol{1}).$$

From the above, $\boldsymbol{c}^* - \boldsymbol{c} = \boldsymbol{0}$; consequently, from Equation (3.5.2), $\boldsymbol{\varphi} = \boldsymbol{0}$.

The reverse is also true; if the ownership sets of two processors with respect to an array overlap, then their coordinates along those dimensions on which the aligned dimensions of the template are distributed must match. This is what Lemma 3.5 states. The above two lemmas can be used to prove the following theorem.

```
for each 0 ≤ q < P̂1                                for each 0 ≤ q < P̂1
    if p ≠ q then                                      if p ≠ q then
       send Δ_p(Y) ∩ ∇_q(S', Y(Ŝ_y α + a_{0_y})) to q     receive Δ_q(Y) ∩ ∇_p(S', Y(Ŝ_y α + a_{0_y})) from q
    endif                                              endif
endfor                                             endfor
```

FIGURE 6: Send and Receive Actions at a Processor $\boldsymbol{p}$

Theorem I. For a given array X, the ownership sets across processors must be either equal or disjoint. That is,

$$\Delta''_p(X) = \Delta''_q(X),$$
or
$$\Delta''_p(X) \cap \Delta''_q(X) = \emptyset.$$

Proof.
Suppose $\Delta''_p(X)$ and $\Delta''_q(X)$ are not disjoint. Then $\Delta''_p(X) \cap \Delta''_q(X) \neq \emptyset$. Hence, from Lemma 3.5, we get

$$\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q}) = \mathbf{0}. \tag{I.1}$$

Since we have assumed that $\Delta''_p(X) \cap \Delta''_q(X) \neq \emptyset$, then $\Delta''_p(X) \neq \emptyset$ and $\Delta''_q(X) \neq \emptyset$. By Lemma 3.4, this fact and Equation (I.1) therefore imply that

$$\Delta''_p(X) = \Delta''_q(X).$$

Thus, either $\Delta''_p(X) \cap \Delta''_q(X) = \emptyset$ or $\Delta''_p(X) = \Delta''_q(X)$ must be true.

Let us define a binary relation $\sim$ on a mapped array such that given two array elements $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, $\boldsymbol{\beta} \sim \boldsymbol{\gamma}$ if and only if $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are mapped onto the same processors. The rules of HPF ensure that for any legal ALIGN/DISTRIBUTE combination, every element of the mapped array will reside on at least one processor [KLS+94]. Hence $\sim$ must be reflexive. Also, if $\boldsymbol{\beta} \sim \boldsymbol{\gamma}$, $\boldsymbol{\gamma} \sim \boldsymbol{\beta}$ is obviously true. Therefore, $\sim$ is symmetric. Finally, if $\boldsymbol{\beta} \sim \boldsymbol{\gamma}$ and $\boldsymbol{\gamma} \sim \boldsymbol{\delta}$ are true, then from Theorem I, $\boldsymbol{\beta} \sim \boldsymbol{\delta}$. That is, $\sim$ is transitive. Hence, the ALIGN and DISTRIBUTE directives for a mapped array define an equivalence relation on that array.

4 The Replication Test

Let $\nabla_p\big(S', Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})\big)$ indicate those elements of a right-hand side array reference $Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})$ in an assignment statement S' that $\boldsymbol{p}$ views on account of its compute work in S'. Thus, $\boldsymbol{p}$ would have to potentially fetch these elements from a remote processor $\boldsymbol{q}$ and the set of elements to be received would be $\Delta_q(Y) \cap \nabla_p\big(S', Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})\big)$. Likewise, $\boldsymbol{p}$ would have to potentially send the elements in $\Delta_p(Y) \cap \nabla_q\big(S', Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})\big)$ to a remote processor $\boldsymbol{q}$ because $\boldsymbol{q}$ may in turn view the elements owned by $\boldsymbol{p}$. Pseudo-code fragments in Figure 6 illustrate how such data exchange operations can be realized.

In § 3.2, we saw that if $\Delta_p(Y) \neq \emptyset$, and $\Delta_q(Y) \neq \emptyset$, then $\Delta_p(Y)$ and $\Delta_q(Y)$ are equal if and only if $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q}) = \mathbf{0}$. This property can be used to avoid redundant communication due to array replication; the modified pseudo-code fragments in Figure 7 show this optimization.

Once the integer FME solution for the system of equalities and inequalities that describe the ownership set is obtained, computing whether $\Delta_p(Y)$ is the empty set for a given $\boldsymbol{p}$ incurs only an additional polynomial-time overhead. The key idea that enables this decision is that in the FME solution system for Equation (11), a

```
for each 0 ≤ q < P̂1
    if (Δ_q(Y) = ∅) ∨ (π̂R̂^T R̂π̂^T(q − p) ≠ 0) then
        send Δ_p(Y) ∩ ∇_q(S′, Y(Ŝ_y α + a_{0_y})) to q
    endif
endfor
```

```
for each 0 ≤ q < P̂1
    if (Δ_p(Y) = ∅) ∨ (π̂R̂^T R̂π̂^T(p − q) ≠ 0) then
        receive Δ_q(Y) ∩ ∇_p(S′, Y(Ŝ_y α + a_{0_y})) from q
    endif
endfor
```

FIGURE 7: Send and Receive Actions with the Replication Test Optimization

particular p_j will occur in the bound expressions for c_j.[h] That is, there will be an inequality pair of the form

$$f_j(p_j) \le c_j \le g_j(p_j)$$

in the solution system. In addition, there can *at most* at one more inequality pair in the solution system that also contains p_j in its bound expressions. This inequality pair will have the form $F_j(p_j, c_j) \le a_{n_j} \le G_j(p_j, c_j)$. Hence, if Equation (11) has a solution $\boldsymbol{a}$ for a given $\boldsymbol{p}$, each of the z disjoint inequality groups

$$\begin{gathered} f_j(p_j) \le c_j \le g_j(p_j), \\ F_j(p_j, c_j) \le a_{n_j} \le G_j(p_j, c_j), \end{gathered}$$

in the solution system must *independently* admit a solution. The task of checking whether each of these groups has a solution for a particular p_j is clearly of quadratic complexity. Hence, the complexity of ascertaining whether $\Delta_p(Y)$ is nonempty for a given $\boldsymbol{p}$ is polynomial. Since the complexity of evaluating the condition $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q}) \ne \boldsymbol{0}$ is $O(z)$, the overall run-time complexity of evaluating the Boolean predicates in Figure 7, given the integer FME solution system for the ownership set (known at compile time), becomes polynomial.

Observe that in the absence of replication, $\hat{\mathcal{R}}^T\hat{\mathcal{R}}$ is the identity matrix; in this situation, $\hat{\pi}\hat{\mathcal{R}}^T\hat{\mathcal{R}}\hat{\pi}^T(\boldsymbol{p}-\boldsymbol{q}) \ne \boldsymbol{0}$ if and only if $\boldsymbol{p} \ne \boldsymbol{q}$. Hence, in the absence of replication, the test degenerates to the usual $\boldsymbol{p} \ne \boldsymbol{q}$ condition.

5 Unspecified Processor Meshes

When a variable is eliminated from a system of inequalities by the FME technique, the method partitions the system into three sets: a set $\mathbb{S}_-$ in which the coefficients of the variable are negative, a set $\mathbb{S}_0$ in which the coefficients of the variable are zero and a set $\mathbb{S}_+$ in which the coefficients of the variable are positive [Ban93]. Therefore, for those variables that are to be eliminated by the technique, knowledge regarding the signs of their coefficients must be available. In the context of the system in Equation (11), this implies that the principal diagonal entries of the square diagonal matrices $\hat{C}$ and $\hat{P}$ can be symbolic. This is because these elements are known to be positive a priori, and their actual values are of no concern until run time. A useful consequence of this fact is that a processor mesh need not be provided in a `DISTRIBUTE` directive. Hence, the determination of the actual processor mesh can be postponed until run time.

For instance, suppose that the distribution directive in Figure 3 lacked a processor mesh—assume that it was `!HPF$ DISTRIBUTE T(*, CYCLIC(4), BLOCK(13))`. Then all that can be said about $\hat{P}$ is that it should have the form

$$\hat{P} = \begin{pmatrix} P_{11} & 0 \\ 0 & P_{22} \end{pmatrix},$$

where P_{11} and P_{22} are positive. If we apply the FME method on the system in Equation (11) with this knowledge of $\hat{P}$, we obtain the solution shown in Figure 8. Notice that on replacing P_{11} and P_{22} in this solution system by 9, we obtain the solution system shown in Figure 5.

Apart from being positive, since the third dimension of `T` is `BLOCK` distributed on the second dimension of `P`, the choice for P_{22} would have to be additionally constrained by the inequality $13P_{22} \ge 100$ (i.e., $C_{jj}P_{jj} \ge u_{k_j} - l_{k_j} + 1$ where k_j is the template dimension distributed in a `BLOCK` fashion on the jth processor dimension).

[h]The symbols p_j and c_j indicate elements in $\boldsymbol{p}$ and $\boldsymbol{c}$ respectively.

More generally, the principal diagonal elements of $\hat{P}$ would have to be so chosen so as to satisfy the following inequalities:

$$\hat{P}\mathbf{1} \geq \mathbf{1}, \tag{12}$$

$$\hat{\lambda}\hat{C}\hat{P}\mathbf{1} \geq \hat{\lambda}\hat{\pi}(\boldsymbol{u_T} - \boldsymbol{l_T} + \mathbf{1}). \tag{13}$$

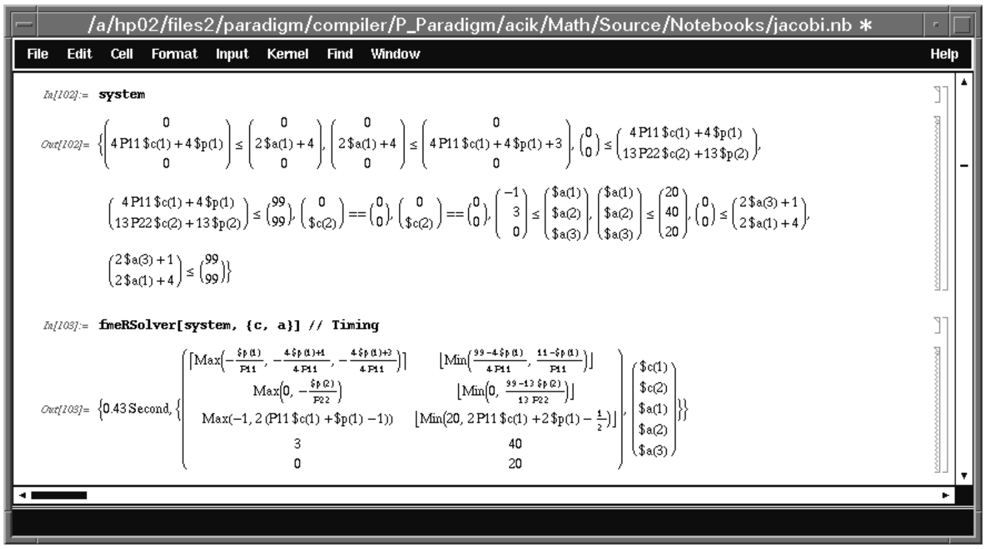

FIGURE 8: Solving the Ownership Set with a Symbolic $\hat{P}$

set η = number_of_processors()
set $\hat{P} = \hat{I}$
set $\hat{\lambda}_T\hat{\lambda}_S\hat{P}\mathbf{1} = \hat{\lambda}_T\hat{\lambda}_S\mathtt{max}(\hat{C}_T^{-1}\hat{\pi}_T(\boldsymbol{u_T} - \boldsymbol{l_T} + \mathbf{1}), \hat{C}_S^{-1}\hat{\pi}_S(\boldsymbol{u_S} - \boldsymbol{l_S} + \mathbf{1}))$
set $\hat{\lambda}_T(\hat{I} - \hat{\lambda}_S)\hat{P}\mathbf{1} = \hat{\lambda}_T(\hat{I} - \hat{\lambda}_S)\hat{C}_T^{-1}\hat{\pi}_T(\boldsymbol{u_T} - \boldsymbol{l_T} + \mathbf{1})$
set $\hat{\lambda}_S(\hat{I} - \hat{\lambda}_T)\hat{P}\mathbf{1} = \hat{\lambda}_S(\hat{I} - \hat{\lambda}_T)\hat{C}_S^{-1}\hat{\pi}_S(\boldsymbol{u_S} - \boldsymbol{l_S} + \mathbf{1})$
if $|\hat{P}| > \eta$ then
    print "Error: Insufficient number of processors."
    exit
else
    set $(\hat{I} - \hat{\lambda}_T)(\hat{I} - \hat{\lambda}_S)\hat{P}\mathbf{1} = \lfloor(\frac{\eta}{|\hat{P}|})^{1/\omega}\rfloor$
endif

FIGURE 9: Establishing $\hat{P}$ at Run Time

A possible strategy for determining the principal diagonal elements of $\hat{P}$ at run time is presented in Figure 9. In general, more than one distributee may be associated with a processor mesh and the handling of such a case is highlighted in the figure. The displayed pseudo-code fragment assumes that two templates `T` and `S` are distributed on the processor mesh `P`. The procedure `number_of_processors()` is a generic system inquiry function that returns the number of physical processors available in the lower-level processor arrangement. The quantity ω—determined at compile time—equals the number of processor dimensions on which template dimensions are distributed in a `CYCLIC`(B) fashion *considering both* `T` and `S`. That is, ω equals the rank of $(\hat{I} - \hat{\lambda}_T)(\hat{I} - \hat{\lambda}_S)$. Note that by this strategy, some of the physical processors may remain unused. More sophisticated schemes that utilize all of the physical processors and that also optimize with respect to some other criteria could be devised.

6 The Mapping Test

Consider an assignment statement S' bearing affine subscript expressions and contained in a loop nest characterized by the loop iteration vector $\boldsymbol{\alpha}$:

$$X(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x}) = \cdots + Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y}) + \cdots$$

The communication sets $\Delta_p(Y) \cap \nabla_q\big(S', Y(\hat{S}_y\boldsymbol{\alpha}+\boldsymbol{a}_{\mathbf{0}_y})\big)$ and $\Delta_q(Y) \cap \nabla_p\big(S', Y(\hat{S}_y\boldsymbol{\alpha}+\boldsymbol{a}_{\mathbf{0}_y})\big)$ (shown in Figures 6 and 7) that would be generated for the above assignment statement take into account the relative alignments and distributions of the left-hand side and right-hand side array references. If these communication sets are empty, no communication will occur at run time. However, the overhead of checking at run time whether a particular processor should dispatch a section of its array to some other processor that views it exists, irrespective of whether data is actually communicated or not. This could result in the expensive run-time cost of communication checks, even in cases where it could be avoided such as when the elements of $X(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x})$ and $Y(\hat{S}_y\boldsymbol{\alpha}+\boldsymbol{a}_{\mathbf{0}_y})$ are ultimately mapped onto the same processor. If the compiler could detect such situations, it could refrain from generating communication code for such array reference pairs. This is what the mapping test attempts to do.

6.1 A Sufficient Condition

Theorem II. Define two vectors $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ for the assignment statement S':

$$\boldsymbol{\zeta} = \hat{\pi}_T\hat{\mathcal{R}}_x^T\big(\hat{\mathcal{A}}_x(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x}) + \boldsymbol{s}_{\mathbf{0}_x} - \hat{\mathcal{R}}_x\boldsymbol{l}_{\boldsymbol{T}}\big),$$
$$\boldsymbol{\xi} = \hat{\pi}_S\hat{\mathcal{R}}_y^T\big(\hat{\mathcal{A}}_y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y}) + \boldsymbol{s}_{\mathbf{0}_y} - \hat{\mathcal{R}}_y\boldsymbol{l}_{\boldsymbol{S}}\big).$$

Here we assume that `T` and `S` are the templates to which `X` and `Y` are respectively aligned, and that both `T` and `S` are distributed onto a processor mesh `P`. If $\hat{\mu} = \hat{\pi}_T\hat{\mathcal{R}}_x^T\hat{\mathcal{R}}_x\hat{\pi}_T^T$ and $\hat{\nu} = \hat{\pi}_S\hat{\mathcal{R}}_y^T\hat{\mathcal{R}}_y\hat{\pi}_S^T$, then

$$\begin{aligned} &\Delta_r(Y) \neq \emptyset \;\forall\; \mathbf{0} \leq \boldsymbol{r} < \hat{P}\mathbf{1} \\ &\wedge\, \hat{\nu} \leq \hat{\mu} \\ &\wedge\, \hat{\nu}(\lfloor \hat{C}_T^{-1}\boldsymbol{\zeta}\rfloor \bmod \hat{P}\mathbf{1}) = \lfloor \hat{C}_S^{-1}\boldsymbol{\xi}\rfloor \bmod \hat{P}\mathbf{1} \end{aligned} \tag{14}$$

is a sufficient condition for the array reference $Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})$ to be identically mapped onto the same processor that owns $X(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x})$.

Proof.
Suppose $X(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x}) \in \Delta''_p(X)$; as discussed earlier, a legal `ALIGN/DISTRIBUTE` combination will admit at least one such $\boldsymbol{p}$. Thus, there exists a $\boldsymbol{c}$ such that from the system in Equation (11), we have

$$\hat{\pi}_T\hat{\mathcal{R}}_x^T\big(\hat{\mathcal{A}}_x(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x}) + \boldsymbol{s}_{\mathbf{0}_x} - \hat{\mathcal{R}}_x\boldsymbol{l}_{\boldsymbol{T}}\big) = \hat{\pi}_T\hat{\mathcal{R}}_x^T\hat{\mathcal{R}}_x\hat{\pi}_T^T(\hat{C}_T\hat{P}\boldsymbol{c} + \hat{C}_T\boldsymbol{p}) + \boldsymbol{l}, \tag{II.1}$$

where

$$\mathbf{0} \leq \boldsymbol{l} \leq \hat{\pi}_T\hat{\mathcal{R}}_x^T\hat{\mathcal{R}}_x\hat{\pi}_T^T(\hat{C}_T\mathbf{1} - \mathbf{1}). \tag{II.2}$$

By virtue of Equation (11), $\hat{\mu}\boldsymbol{c} = \boldsymbol{c}$. Therefore, after some rearranging, Equation (II.1) becomes $\hat{C}_T(\hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p}) + \boldsymbol{l} = \boldsymbol{\zeta}$. From Inequality (II.2), $\mathbf{0} \leq \hat{C}_T^{-1}\boldsymbol{l} \leq \hat{\mu}(\mathbf{1} - \hat{C}_T^{-1}\mathbf{1})$. Therefore,

$$\hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} \leq \hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} + \hat{C}_T^{-1}\boldsymbol{l} \leq \hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} + \hat{\mu}(\mathbf{1} - \hat{C}_T^{-1}\mathbf{1}). \tag{II.3}$$

Since $\mathbf{1} \le \hat{C}_T\mathbf{1}$, we have $\mathbf{0} < \hat{C}_T^{-1}\mathbf{1} \le \mathbf{1}$. Thus,

$$\mathbf{0} \le \mathbf{1} - \hat{C}_T^{-1}\mathbf{1} < \mathbf{1}. \tag{II.4}$$

Since $\hat{0} \le \hat{\mu} \le \hat{I}$ is always true, we get $\hat{0} \le \hat{\mu}(\mathbf{1} - \hat{C}_T^{-1}\mathbf{1}) \le \mathbf{1} - \hat{C}_T^{-1}\mathbf{1}$ from the above. Using Inequality (II.4) again, we also get $\mathbf{0} \le \hat{\mu}(\mathbf{1} - \hat{C}_T^{-1}\mathbf{1}) < \mathbf{1}$. Consequently, Inequality (II.3) becomes

$$\lfloor \hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} \rfloor \le \lfloor \hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} + \hat{C}_T^{-1}\boldsymbol{l} \rfloor < \lfloor \hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} + \mathbf{1} \rfloor,$$

or $\hat{P}\boldsymbol{c} + \hat{\mu}\boldsymbol{p} = \lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor$. Since $\mathbf{0} \le \hat{\mu}\boldsymbol{p} \le \boldsymbol{p} < \hat{P}\mathbf{1}$, we therefore have

$$\hat{\mu}\boldsymbol{p} = \lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}. \tag{II.5}$$

That is, if $X(\hat{S}_x\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_x}) \in \Delta_p''(X)$, then $\boldsymbol{p}$ must fulfill Equation (II.5). Similarly, if $Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y}) \in \Delta_q''(Y)$, then $\boldsymbol{q}$ must satisfy Equation (II.6):

$$\hat{\nu}\boldsymbol{q} = \lfloor \hat{C}_S^{-1}\boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1}. \tag{II.6}$$

It is also given that $\Delta_r''(Y) \ne \emptyset$ for all $\mathbf{0} \le \boldsymbol{r} < \hat{P}\mathbf{1}$. Thus, from Lemma 3.4, every $\boldsymbol{r}$ that fulfills Equation (II.6) also owns $Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y})$. We now need to verify whether a $\boldsymbol{p}$ that is a solution for Equation (II.5) is also a solution for Equation (II.6). Since we are given that $\hat{\nu}(\lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \lfloor \hat{C}_S^{-1}\boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1}$, using Equations (II.5) and (II.6) we get

$$\hat{\mu}\hat{\nu}\boldsymbol{p} = \hat{\nu}\boldsymbol{q}. \tag{II.7}$$

But we are also given that $\hat{\nu} \le \hat{\mu}$. Therefore, multiplying by $\hat{\nu}$ and noting that $\hat{\nu}^2 = \hat{\nu}$, we get

$$\hat{\nu} \le \hat{\mu}\hat{\nu}. \tag{II.8}$$

Again, since $\hat{\mu} \le \hat{I}$, we get on multiplying by $\hat{\nu}$,

$$\hat{\mu}\hat{\nu} \le \hat{\nu}. \tag{II.9}$$

Hence, Inequalities (II.8) and (II.9) imply that $\hat{\mu}\hat{\nu} = \hat{\nu}$. Therefore, Equation (II.7) becomes $\hat{\nu}\boldsymbol{p} = \hat{\nu}\boldsymbol{q}$. Thus, $\boldsymbol{p}$ also fulfills Equation (II.6). In other words, $Y(\hat{S}_y\boldsymbol{\alpha} + \boldsymbol{a}_{\mathbf{0}_y}) \in \Delta_p''(Y)$ is also true.

The sufficient condition stated in Theorem II can be established at compile time. Predicate (14) constitutes the actual mapping test. Verifying whether $\hat{\nu} \le \hat{\mu}$ and $\hat{\nu}(\lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \lfloor \hat{C}_S^{-1}\boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1}$ is an $O(z)$ operation. Establishing whether $\Delta_r(Y) \ne \emptyset$ for all $\mathbf{0} \le \boldsymbol{r} < \hat{P}\mathbf{1}$ is a polynomial-time operation, given the symbolic representation of the ownership set (see § 4). Thus, the overall time complexity for verifying the requirements of Theorem II is polynomial, once the FME solution system for the ownership set is known.

The impact of the mapping test on run times can often be dramatic. To illustrate the savings, the run times for the ADI benchmark, for arrays of sizes $4 \times 1024 \times 2$ on a 1×4 mesh of processors, with and without this optimization were 0.51 and 64.79 seconds respectively! The large value of 64.79 seconds arose due to three assignment statements that were the sinks of loop-independent flow dependencies and that were enclosed within a triply nested loop spanning an iteration space of $2048 \times 2 \times 1022$ points. Each of these three assignment statements included right-hand side array references that were finally distributed onto the same processor as the corresponding left-hand side array reference. Hence, in all, eighteen communication checks (nine for `MPI_SEND` and another nine for `MPI_RECV`) per iteration were eliminated.

6.2 The Matrix Multiplication Benchmark

To demonstrate how the mapping test works, we begin with the Matrix Multiplication benchmark (from the Livermore Kernel 21 [McM86]), in which each of the templates `T` and `S` are simultaneously collapsed and

```
      program matmul

      REAL A(1024, 1024), B(1024, 1024), C(1024, 1024)
      INTEGER i, j, k

!HPF$ TEMPLATE S(1024, 1024)
!HPF$ TEMPLATE T(1024, 1024)
!HPF$ PROCESSORS P(2, 2)
!HPF$ ALIGN A(i, *) WITH S(i, *)
!HPF$ ALIGN B(*, j) WITH S(*, j)
!HPF$ ALIGN C(i, j) WITH T(i, j)
!HPF$ DISTRIBUTE S(BLOCK, BLOCK) ONTO P
!HPF$ DISTRIBUTE T(BLOCK, BLOCK) ONTO P

!C
!C    We assume that the matrices A, B and C
!C    are already initialized.
!C

      do j = 1, 1024
         do k = 1, 1024
            do i = 1, 1024
               C(i, j) = C(i, j)+A(i, k)*B(k, j)
            enddo
         enddo
      enddo

      end
```

FIGURE 10: Matrix Multiplication

replicated. The mapping test correctly determines that none of the right-hand side array references in the benchmark's single assignment statement require any communication code to be generated.

Consider the pair of references `C(i, j)` and `A(i, k)` in Figure 10. By inspection, we have

$$\hat{\mathcal{A}}_c = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \boldsymbol{s_{0_c}} = \begin{pmatrix}0\\ 0\end{pmatrix}, \quad \hat{\mathcal{R}}_c = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \boldsymbol{a_{l_c}} = \begin{pmatrix}1\\ 1\end{pmatrix}, \quad \boldsymbol{a_{u_c}} = \begin{pmatrix}1024\\ 1024\end{pmatrix},$$

$$\hat{\pi_T} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \hat{C_T} = \begin{pmatrix}512 & 0\\ 0 & 512\end{pmatrix}, \quad \hat{\lambda_T} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \boldsymbol{l_T} = \begin{pmatrix}1\\ 1\end{pmatrix}, \quad \boldsymbol{u_T} = \begin{pmatrix}1024\\ 1024\end{pmatrix},$$

$$\hat{\mathcal{A}}_a = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \quad \boldsymbol{s_{0_a}} = \begin{pmatrix}0\\ 0\end{pmatrix}, \quad \hat{\mathcal{R}}_a = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \quad \boldsymbol{a_{l_a}} = \begin{pmatrix}1\\ 1\end{pmatrix}, \quad \boldsymbol{a_{u_a}} = \begin{pmatrix}1024\\ 1024\end{pmatrix},$$

$$\hat{\pi_S} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \hat{C_S} = \begin{pmatrix}512 & 0\\ 0 & 512\end{pmatrix}, \quad \hat{\lambda_S} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \boldsymbol{l_S} = \begin{pmatrix}1\\ 1\end{pmatrix}, \quad \boldsymbol{u_S} = \begin{pmatrix}1024\\ 1024\end{pmatrix},$$

$$\hat{P} = \begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}.$$

Thus,

$$\boldsymbol{\zeta} = \hat{\pi}_T \hat{\mathcal{R}}_c^T \big(\hat{\mathcal{A}}_c(\hat{S}_c\boldsymbol{\alpha} + \boldsymbol{a_{0_c}}) + \boldsymbol{s_{0_c}} - \hat{\mathcal{R}}_c \boldsymbol{l_T}\big) = \begin{pmatrix}i-1\\ j-1\end{pmatrix},$$

$$\boldsymbol{\xi} = \hat{\pi}_S \hat{\mathcal{R}}_a^T \big(\hat{\mathcal{A}}_a(\hat{S}_a\boldsymbol{\alpha} + \boldsymbol{a_{0_a}}) + \boldsymbol{s_{0_a}} - \hat{\mathcal{R}}_a \boldsymbol{l_S}\big) = \begin{pmatrix}i-1\\ 0\end{pmatrix},$$

$$\hat{\mu} = \hat{\pi}_T \hat{\mathcal{R}}_c^T \hat{\mathcal{R}}_c \hat{\pi}_T^T = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \qquad \hat{\nu} = \hat{\pi}_S \hat{\mathcal{R}}_a^T \hat{\mathcal{R}}_a \hat{\pi}_S^T = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

To verify the first requirement, we need to determine the integer FME solution system for $\Delta_p(A)$. By applying the FME solver with (c_1, c_2, a_1, a_2) as the elimination order, we find this to be

$$\begin{aligned}\Delta_p(A) = \{ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} | \exists \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} & \text{ such that} \\ & \lceil \max(0, -\tfrac{511}{1024} - \tfrac{p_1}{2}, \tfrac{-p_1}{2}) \rceil \le c_1 \le \lfloor \min(0, \tfrac{1023}{1024} - \tfrac{p_1}{2}) \rfloor, \\ & \lceil \max(0, \tfrac{-p_2}{2}) \rceil \le c_2 \le \lfloor \min(0, \tfrac{1023}{1024} - \tfrac{p_2}{2}) \rfloor, \\ & \max(1, 1 + 1024c_1 + 512p_1) \le a_1 \le \min\big(1024, 512(1 + 2c_1 + p_1)\big), \\ & 1 \le a_2 \le 1024 \}. \end{aligned} \tag{15}$$

There are two constraints in Equation (15) that contain p_1 in their bound expressions and another constraint that contains p_2 in its bound expressions. Hence, by separately considering two disjoint inequality groups, we can determine in polynomial time that $\Delta_r(A) \neq \emptyset$ for all $\mathbf{0} \le \boldsymbol{r} < \hat{P}\mathbf{1}$. Further, since $\hat{\nu} \le \hat{\mu}$, the second requirement of the mapping test is also satisfied. In addition,

$$\lfloor \hat{C}_T^{-1} \boldsymbol{\zeta} \rfloor = \begin{pmatrix} \lfloor \frac{i-1}{512} \rfloor \\ \lfloor \frac{j-1}{512} \rfloor \end{pmatrix}, \qquad \lfloor \hat{C}_S^{-1} \boldsymbol{\xi} \rfloor = \begin{pmatrix} \lfloor \frac{i-1}{512} \rfloor \\ 0 \end{pmatrix}.$$

Therefore the third requirement is also satisfied since,

$$\hat{\nu}(\lfloor \hat{C}_T^{-1} \boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \begin{pmatrix} \lfloor \frac{i-1}{512} \rfloor \bmod 2 \\ 0 \end{pmatrix}, \qquad \lfloor \hat{C}_S^{-1} \boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1} = \begin{pmatrix} \lfloor \frac{i-1}{512} \rfloor \bmod 2 \\ 0 \end{pmatrix}.$$

We can thus conclude that `A(i, k)` is identically mapped onto the same processor that owns `C(i, j)`, for all values of the loop iteration vector $\boldsymbol{\alpha}$. We can similarly show that `B(k, j)` is also identically mapped onto the same processor that owns `C(i, j)`, for all iterations of the loop nest. Since the third right-hand side array reference `C(i, j)` is trivially mapped onto the same processor as the left-hand side array reference, no communication code needs to be generated for this benchmark.

How does the replication test fare on this benchmark? Note that the the Boolean predicate $\hat{\pi}_S \hat{\mathcal{R}}_a^T \hat{\mathcal{R}}_a \hat{\pi}_S^T (\boldsymbol{p} - \boldsymbol{q}) \neq \mathbf{0}$ simplifies to $p_1 \neq q_1$. Thus, though the replication test avoids *some* of the communication check overhead in the send and receive actions shown in Figure 7, the send and receive sets would still be computed and checked whenever p_1 and q_1 are not equal. However, because of the nature of the affine subscript expressions of the array references involved (i.e., `C(i, j)` and `A(i, k)`), the send and receive sets evaluate to empty sets even when p_1 and q_1 are not equal. Thus, in the case of this benchmark, the entire overhead of performing communication checks at run time can be avoided, and this is exactly what the mapping test detects.

6.3 A Synthetic Example

Consider the synthetic code shown in Figure 11. For this fragment, the various matrices and vectors required by the mapping test are

$$\boldsymbol{\zeta} = \begin{pmatrix} 1408 \\ 2i + 2j + 6 \end{pmatrix}, \qquad \boldsymbol{\xi} = \begin{pmatrix} 512 \\ 2i + 2j + 6 \end{pmatrix}, \qquad \hat{\mu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \hat{\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\hat{\nu}(\lfloor \hat{C}_T^{-1} \boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \begin{pmatrix} 4 \\ \lfloor \frac{2i+2j+6}{16} \rfloor \bmod 9 \end{pmatrix}, \qquad \lfloor \hat{C}_S^{-1} \boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1} = \begin{pmatrix} 4 \\ \lfloor \frac{2i+2j+6}{16} \rfloor \bmod 9 \end{pmatrix}.$$

Since $\hat{\nu} = \hat{\mu}$ and $\hat{\nu}(\lfloor \hat{C}_T^{-1} \boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \lfloor \hat{C}_S^{-1} \boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1}$, the second and third requirements of the mapping test are satisfied. By applying the FME solver with $(c_1, c_2, a_1, a_2, a_3, a_4)$ as the elimination order, we find the integer FME solution system for $\Delta_p(Y)$ to be

$$
\begin{aligned}
\Delta_p(Y) = \{ & \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} | \exists \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \text{ such that} \\
& \lceil \max(0, -\tfrac{127}{1024} - \tfrac{p_1}{8}, -\tfrac{123}{1024} - \tfrac{p_1}{8}, \tfrac{-p_1}{8}) \rceil \le c_1 \le \lfloor \min(0, \tfrac{511}{512} - \tfrac{p_1}{8}, \tfrac{1023}{1024} - \tfrac{p_1}{8}) \rfloor, \\
& \lceil \max(-\tfrac{5}{48} - \tfrac{p_2}{9}, -\tfrac{11}{144} - \tfrac{p_2}{9}, \tfrac{-p_2}{9}) \rceil \le c_2 \le \lfloor \min(\tfrac{511-16p_2}{144}, \tfrac{32-p_2}{9}) \rfloor, \\
& \max(1, -3 + 1024c_1 + 128p_1) \le a_1 \le \min\big(1019, 4(31 + 256c_1 + 32p_1)\big), \\
& 2 \le a_2 \le 120, \\
& \max(1, -3 + 144c_2 + 16p_2) \le a_3 \le \min\big(508, 4(3 + 36c_2 + 4p_2)\big), \\
& 1 \le a_4 \le 32 \}.
\end{aligned} \tag{16}
$$

Equation (16) consists of two disjoint inequality groups; we can therefore determine in polynomial time that $\Delta_r(Y) \neq \emptyset$ for all $\mathbf{0} \le \boldsymbol{r} < \hat{P}\mathbf{1}$. Therefore, from Theorem II, the right-hand side array reference is identically mapped onto the same processor that owns the left-hand side array reference. This is indeed the case as can be verified by visualizing the alignments and distributions. Notice that terms unknown at compile time are only manipulated symbolically by the mapping test. For instance, if instead of `Y(509, i+7k, 2j+2i+3, 7k)` we had `Y(509, i+7k, 3j+2i+3, 7k)` then,

$$
\lfloor \hat{C}_S^{-1} \boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1} = \begin{pmatrix} 4 \\ \lfloor \frac{3i+2j+6}{16} \rfloor \bmod 9 \end{pmatrix},
$$

and the test would have failed since nothing can be said at compile time about the equality of $(\lfloor (2i + 2j + 6)/16 \rfloor \bmod 9)$ and $(\lfloor (3i + 2j + 6)/16 \rfloor \bmod 9)$. For some values of i and j this may be true (say, $i = 3, j = 2$) while not for others (say, $i = 2, j = 2$). Thus, in such a situation, the test fails, and conservatively, communication code is generated.

```
      REAL X(682, 4:521, 512), Y(1019, 2:120, 508, 32)
      INTEGER i, j, k

!HPF$ TEMPLATE T(2:2048, 8:1042)
!HPF$ TEMPLATE S(0:1023, 0:10, 0:250, 0:512, 0:100)
!HPF$ PROCESSORS P(8, 9)
!HPF$ ALIGN X(i, j, *) WITH T(3*i, 2*j)
!HPF$ ALIGN Y(i, j, k, l) WITH S(i+3, *, 2*j+2, k+3, 3*l+4)
!HPF$ DISTRIBUTE T(CYCLIC(32), CYCLIC(16)) ONTO P
!HPF$ DISTRIBUTE S(BLOCK, *, *, CYCLIC(16), *) ONTO P

      do k = ...
         do j = ...
            do i = ...
               X(470, j+i+7, 3k+2i) = Y(509, i+7k, 2j+2i+3, 7k)
            enddo
         enddo
      enddo
```

FIGURE 11: Synthetic Code

6.4 The ADI Benchmark

The last example shows how the mapping test makes it possible for the compiler to avoid generating any communication code for the Automatic Differentiation and Integration (ADI) benchmark (from the Livermore Kernel 8 [McM86]), when the alignments and distributions are suitably chosen.

In essence, the ADI benchmark comprises six arrays of type `REAL`, of which three are one-dimensional and consist of 1024 elements each, while the remaining three are three-dimensional arrays consisting of $4 \times 1024 \times 2$ elements each. The only flow dependencies that this benchmark exhibits are three loop-independent ones, and all of the source-sink statement pairs are contained in the innermost loop. The right-hand side array references on which these dependencies terminate can be aligned and distributed onto the same processors that own the corresponding left-hand side array references, and it is this particular case that the mapping test optimizes.

```
      program adi

      REAL DU1(1024), DU2(1024), DU3(1024)
      REAL AU1(4, 1024, 2), AU2(4, 1024, 2), AU3(4, 1024, 2)
      INTEGER l, kx, ky
      ...
!HPF$ TEMPLATE T(4, 1024, 2)
!HPF$ ALIGN DU1(i) WITH T(*, i, *)
!HPF$ ALIGN AU1(i, j, k) WITH T(i, j, k)
!HPF$ PROCESSORS P(1, 16)
!HPF$ DISTRIBUTE T(BLOCK, CYCLIC(62), *) ONTO P

!C
!C    Initializations.
!C
      ...
!C
!C    The ADI kernel.
!C

      do l = 1, 2048
         do kx = 2, 3
            do ky = 2, 1023
               DU1(ky) = AU1(kx, ky+1, 1)-AU1(kx, ky-1, 1)
               ...
               AU1(kx, ky, 2) = AU1(kx, ky, 1)+a11*DU1(ky)+a12*DU2(ky)+
     1         a13*DU3(ky)+sig*(AU1(kx+1, ky, 1)-2*AU1(kx, ky, 1)+AU1(k
     2         x-1, ky, 1))
               ...
            enddo
         enddo
      enddo

      end
```

FIGURE 12: ADI

In the code excerpt shown in Figure 12, only one source-sink statement pair among the three is indicated. The other two are similar, with the array `DU1` replaced respectively by `DU2` and `DU3` in the source statements, and with the array `AU1` replaced respectively by `AU2` and `AU3` in the source and sink statements. We now consider the pair of array references `AU1(kx, ky, 2)` and `DU1(ky)` occurring in the second assignment statement of Figure 12. Thus,

$$\boldsymbol{\zeta} = \begin{pmatrix} kx-1 \\ ky-1 \end{pmatrix}, \qquad \boldsymbol{\xi} = \begin{pmatrix} 0 \\ ky-1 \end{pmatrix}, \qquad \hat{\mu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \hat{\nu} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since

$$\begin{aligned} \Delta''_p(DU1) = \Big\{ (a_1) \,|\, \exists \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \text{ such that} \\ &\max(0, -p_1) \le c_1 \le \lfloor \min(0, \tfrac{3}{4} - p_1) \rfloor, \\ &\lceil \max(-\tfrac{61}{992} - \tfrac{p_2}{16}, \tfrac{-p_2}{16}) \rceil \le c_2 \le \lfloor \tfrac{33-2p_2}{32} \rfloor, \\ &\max(1, 1 + 992c_2 + 62p_2) \le a_1 \le \min\big(1024, 62(1 + 16c_2 + p_2)\big) \Big\}, \end{aligned}$$

we see that the first requirement is fulfilled. In addition, since $\hat{\nu} \le \hat{\mu}$, the second requirement is also satisfied. Finally,

$$\lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor = \begin{pmatrix} \lfloor \frac{kx-1}{4} \rfloor \\ \lfloor \frac{ky-1}{62} \rfloor \end{pmatrix}, \qquad \lfloor \hat{C}_S^{-1}\boldsymbol{\xi} \rfloor = \begin{pmatrix} 0 \\ \lfloor \frac{ky-1}{62} \rfloor \end{pmatrix}.$$

Hence,

$$\hat{\nu}(\lfloor \hat{C}_T^{-1}\boldsymbol{\zeta} \rfloor \bmod \hat{P}\mathbf{1}) = \begin{pmatrix} 0 \\ \lfloor \frac{ky-1}{62} \rfloor \bmod 16 \end{pmatrix}, \qquad \lfloor \hat{C}_S^{-1}\boldsymbol{\xi} \rfloor \bmod \hat{P}\mathbf{1} = \begin{pmatrix} 0 \\ \lfloor \frac{ky-1}{62} \rfloor \bmod 16 \end{pmatrix}.$$

Thus, the third requirement is also satisfied. We can therefore conclude that `DU1(ky)` gets identically mapped onto the same processor that owns `AU1(kx, ky, 2)` for all values of the loop iteration vector $\boldsymbol{\alpha}$. It can be similarly shown that for the same alignment and distribution directives, the right-hand side array references in the other assignment statements on which flow dependencies terminate in this benchmark are also identically mapped onto the same processor as the respective left-hand side array references. Thus, no communication code needs to be generated for the benchmark.

7 Mapping Test Measurements

Execution times and compilation times were measured for the PARADIGM (version 2.0) system with and without the mapping test optimization. For the sake of comparison, execution times and compilation times for the original sequential sources and the parallelized codes generated by `pghpf` (version 2.4) and `xlhpf` (version 1.03) were also recorded. `pghpf` and `xlhpf` are commercial HPF compilers from the Portland Group Inc., (PGI) and the International Business Machines (IBM) respectively. In the input codes to the `pghpf` compiler, `DO` loops were recast into `FORALL` equivalents where possible and were qualified with the `INDEPENDENT` directive when appropriate. The `FORALL` construct and the `INDEPENDENT` directive were not mixed in the inputs to `pghpf`, and the tabulated execution times correspond to the best of the two cases. All of the PARADIGM measurements were done in the presence of the replication test.

7.1 System Specifications

The IBM compilers `xlf` and `mpxlf` were used to handle FORTRAN 77 and FORTRAN 77+MPI sources respectively. The HPF sources were compiled using `xlhpf` and `pghpf`. The `-O` option, which results in the generation of optimized code, was always used during compilations done with `xlf`, `xlhpf`, `mpxlf` and `pghpf`. Compilation times were obtained by considering the source-to-source transformation effected by PARADIGM, *as well as* the source-to-executable compilation done using `mpxlf` (version 5.01). The source-to-source compilation times for PARADIGM were measured on an HP Visualize C180 with a 180MHz HP PA-8000 CPU, running HP-UX 10.20 and having 128MB of RAM. Compilation times for `pghpf`, `xlhpf` as well as `mpxlf` were measured on an IBM E30 running AIX 4.3 and having a 133MHz PowerPC 604 processor and 96MB of main memory. In those tables that tabulate the execution times, the RS6000 column refers to the sequential execution times obtained on the IBM E30. The parallel codes were executed on a 16-node IBM SP2 multicomputer, running AIX 4.3 and in which each processor was a 62.5MHz POWER node having 128MB of RAM. Inter-processor communication on the IBM SP2 was across a high performance adapter switch.

7.2 Alignments and Distributions

Measurements for the mapping test were taken across three benchmarks: ADI, Euler Fluxes (from FLO52 in the Perfect Club Suite) and Matrix Multiplication. For all the input samples, fixed templates and alignments were chosen; these are shown in Figure 13. Note that the most suitable alignments were chosen for the benchmark input samples. For the Matrix Multiplication and ADI benchmarks, these alignments resulted in communication-free programs independent of the distributions. For the Euler Fluxes benchmark, the distributions resulted in varying amounts of communication.

The `PROCESSORS` and the `DISTRIBUTE` directives were changed in every benchmark's input sample. The various distributions were chosen *arbitrarily*, the idea being to demonstrate the ability of the mapping test to handle any given alignment/distribution combination.

```
!Alignment directives for ADI.
!HPF$ TEMPLATE T(4, 1024, 2)
!HPF$ ALIGN DU1(i) WITH T(*, i, *)
!HPF$ ALIGN DU2(i) WITH T(*, i, *)
!HPF$ ALIGN DU3(i) WITH T(*, i, *)
!HPF$ ALIGN AU1(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN AU2(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN AU3(i, j, k) WITH T(i, j, k)
```

```
!Alignment directives for EFLUX.
!HPF$ TEMPLATE T(0:5001, 34, 4)
!HPF$ ALIGN FS(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN DW(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN W(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN X(i, j, k) WITH T(i, j, k)
!HPF$ ALIGN P(i, j) WITH T(i, j, *)
```

```
!Alignment directives for MATMUL.
!HPF$ TEMPLATE S(1024, 1024)
!HPF$ TEMPLATE T(1024, 1024)
!HPF$ ALIGN A(i, *) WITH S(i, *)
!HPF$ ALIGN B(*, j) WITH S(*, j)
!HPF$ ALIGN C(i, j) WITH T(i, j)
```

FIGURE 13: Alignment Directives for the ADI, Euler Fluxes and Matrix Multiplication Benchmarks

TABLE 2: Execution Times in Seconds

Benchmark	Mesh	Distribution	IBM AIX 4.3				
			RS6000	SP2			
			xlf v 5.01	pdm v 2.0		pghpf v 2.4	xlhpf v 1.03
				Optimized	Nonoptimized		
ADI	1×2	(BLOCK, BLOCK, *)	7.07	1.41	83.57	16.78	3.36
	1×4	(BLOCK, BLOCK, *)	6.79	0.51	64.79	12.67	2.35
	1×8	(BLOCK, CYCLIC(120), *)	7.90	4.34	609.44	20.94	-^{m}
EFLUX	1×2	(BLOCK, BLOCK, *)	71.85	19.03	19.44	231.67	33.79
	2×2	(BLOCK, *, CYCLIC)	71.40	13.47	13.83	274.93	3113.51
	8×1	(*, BLOCK, CYCLIC)	71.49	5.96	6.39	91.83	8.17
MATMUL	2×1	(BLOCK, BLOCK)^{n}	12.65	2.47	5.83	57.48	25.88
	2×2	(BLOCK, BLOCK)^{o}	104.96	10.06	17.20	224.29	100.37
	4×2	(BLOCK, CYCLIC(120))^{o}	104.74	5.71	205.00	123.59	-^{m}

7.3 Analysis

As Table 2 reveals, the benefits of the mapping test were most pronounced for the ADI benchmark, followed by the Matrix Multiplication benchmark. In the case of the Euler Fluxes benchmark, the mapping test eliminated six communication checks per iteration for the first input sample, and eight communication checks per iteration for the second and third input samples. In the absence of the mapping test, ten communication checks per iteration were generated. On account of loop-carried flow dependencies, the associated communication codes were hoisted immediately within the outermost loop. However, since the number of iterations of the outermost loop was a mere 100, the optimized compiled codes did not exhibit any significant improvement in run times. For all of the ADI benchmark input samples, the iteration space comprised of $2048 \times 2 \times 1022$ points, and the communication codes generated in the absence of the mapping test were hoisted immediately within the innermost loop. For the three Matrix Multiplication benchmark samples, the number of iteration points were $512 \times 512 \times 512$, $1024 \times 1024 \times 1024$ and $1024 \times 1024 \times 1024$ respectively and the single communication check that was generated in the absence of the mapping test was hoisted within the second innermost loop.

Given a sequential input source written using FORTRAN 77 and having HPF directives, PARADIGM produces an SPMD output consisting of FORTRAN 77 statements and procedure calls to the MPI library. The compilation of this SPMD code into the final executable is then performed using `mpxlf`. Since the mapping test eliminates the generation of communication code where possible, it also exerts an influence on the *overall* compilation times. That is, the application of the mapping test often results in the generation of a smaller intermediate SPMD code, and this improves on the back-end source-to-executable compilation time. In our setup, this was done using `mpxlf`. Note that applying the mapping test does not necessarily mean an increased time for the source-to-source compilation phase performed by PARADIGM. This is because though compilation in the presence of the mapping test involves the additional effort of identifying the candidate array reference pairs that are identically mapped, it however saves on the communication code generation part that would otherwise have to be done for the same array reference pairs. Hence, compilation times for the source-to-source compilation phase may in fact be more in the absence of the mapping test and this was found to be true for nearly all of the benchmark samples tested. However, as Table 3 also reveals, the back-end compilation times were nearly always more in the absence of the mapping test, and this was because of the larger intermediate SPMD code sizes handled.

TABLE 3: Compilation Times in Seconds

Benchmark	Mesh	Distribution	pdm v 2.0		mpxlf v 5.01		pghpf v 2.4	xlhpf v 1.03
			Optimized	Nonoptimized	Optimized	Nonoptimized		
ADI	1×2	(BLOCK, BLOCK, *)	6.00	9.00	2.00	7.00	7.00	5.00
	1×4	(BLOCK, BLOCK, *)	7.00	8.00	2.00	7.00	8.00	5.00
	1×8	(BLOCK, CYCLIC(120), *)	7.00	11.00	2.00	17.00	8.00	-[m]
EFLUX	1×2	(BLOCK, BLOCK, *)	11.00	12.00	6.00	8.00	7.00	6.00
	2×2	(BLOCK, *, CYCLIC)	10.00	14.00	4.00	11.00	9.00	12.00
	8×1	(*, BLOCK, CYCLIC)	11.00	12.00	6.00	22.00	9.00	5.00
MATMUL	2×1	(BLOCK, BLOCK)[n]	4.00	5.00	2.00	2.00	5.00	1.00
	2×2	(BLOCK, BLOCK)[o]	4.00	5.00	2.00	3.00	6.00	1.00
	4×2	(BLOCK, CYCLIC(120))[o]	5.00	5.00	2.00	3.00	6.00	-[m]

8 Summary

The preceding sections have shown certain basic and interesting properties that ownership sets exhibit, even in the presence of arbitrary alignments and distributions. Our approach to solving for the ownership set (and other sets derived from it) is based on integer FME solutions to the systems characterizing these sets. We also showed how the system of equalities and inequalities originally proposed in [ACIK95] can be refined to a form requiring the course vector as the only auxiliary vector. This refinement is beneficial to the FME approach. The fundamental property of ownership set equivalence is derived and we demonstrated how it can be used to eliminate redundant communication due to array replication. We also briefly described how to efficiently make decisions regarding the "emptiness" of an ownership set. Finally, we derived a sufficient condition that when true ensures that a right-hand side array reference of an assignment statement is available on the same processor that owns the left-hand side array reference, thus making it possible to avoid generating communication code for the pair.

The mapping test is a very useful optimization. Its positive effect was observable in the case of other benchmarks such as Jacobi, TOMCATV and 2-D Explicit Hydrodynamics (from the Livermore Kernel 18), and was significant in most situations. This was on account of the fact that typically, suitably chosen `ALIGN` and `DISTRIBUTE` directives perfectly align and distribute at least one pair of left-hand side and right-hand side

[m] `xlhpf` does not permit a `CYCLIC` blocking factor greater than 1.
[n] Arrays were of type `REAL`; array sizes were 512×512.
[o] Arrays were of type `REAL`; array sizes were 1024×1024.

array references in at least one assignment statement in the program, and such alignments and distributions are often valid whatever be the values that the loop iteration vector ranges through. Thus, by efficiently exploiting the ownership set, efficient SPMD code can be generated efficiently at compile time.

References

[ACIK95] Corinne Ancourt, Fabien Coelho, François Irigoin, and Ronan Keryell. "A Linear Algebra Framework for Static HPF Code Distribution". Technical Report A–278–CRI, Centre de Recherche en Informatique, École Nationale Supérieure des Mines de Paris, 35, rue Saint-Honoroé, F–77305 Fontainebleau cedex, France, November 1995.

[AL93] Saman P. Amarasinghe and Monica S. Lam. "Communication Optimization and Code Generation for Distributed-Memory Machines". In *Proceedings of the ACM SIGPLAN Programming Language Design and Implementation*, pages 126–138, Albuquerque, NM, USA, June 1993.

[AMC98] Vikram Adve and John Mellor-Crummey. "Using Integer Sets for Data-Parallel Program Analysis and Optimization". In *Proceedings of the ACM SIGPLAN Programming Language Design and Implementation*, pages 186–198, Montréal, Canada, June 1998.

[Bal90] V. Balasundaram. "A Mechanism for Keeping Useful Internal Information in Parallel Programming Tools—The Data Access Descriptor". *Journal of Parallel and Distributed Computing*, 9(2):154–170, June 1990.

[Ban93] Utpal Banerjee. **Loop Transformations for Restructuring Compilers: The Foundations**. Kluwer Academic Publishers, Norwell, MA 02061, USA, January 1993. ISBN 0–7923–9318–X.

[CGL+95] Siddhartha Chatterjee, John R. Gilbert, Fred J. E. Long, Robert S. Schreiber, and Shang-Hua Teng. "Generating Local Addresses and Communication Sets for Data-Parallel Programs". *Journal of Parallel and Distributed Computing*, 26(1):72–84, April 1995.

[CK87] David Callahan and Ken Kennedy. "Analysis of Interprocedural Side Effects in a Parallel Programming Environment". In *Proceedings of the 1st ACM International Conference on Supercomputing*, pages 138–171, Athens, Greece, May 1987.

[GKM+93] S. K. S. Gupta, S. D. Kaushik, S. Mufti, S. Sharma, C.-H. Huang, and P. Sadayappan. "On Compiling Array Expressions for Efficient Execution on Distributed-Memory Machines". In *Proceedings of the 22nd International Conference on Parallel Processing*; Editors: Alok Choudhary and P. Bruce Berra, volume II of *Software*, pages 301–305, St. Charles, IL, USA, August 1993. The Pennsylvania State University, CRC Press, Inc., Boca Raton, FL, USA.

[JB99] Pramod G. Joisha and Prithviraj Banerjee. "PARADIGM (Version 2.0): A New HPF Compilation System". In *Proceedings of the 13th International Parallel Processing Symposium & 10th Symposium on Parallel and Distributed Processing*, pages 609–615, San Juan, Puerto Rico, April 1999.

[KLS+94] Charles H. Koelbel, David B. Loveman, Robert S. Schreiber, Guy L. Steele Jr., and Mary E. Zosel. **The High Performance FORTRAN Handbook**. Scientific and Engineering Computation Series. The MIT Press, Cambridge, MA 02142, USA, 1994. ISBN 0–262–61094–9.

[McM86] F. M. McMahon. "The Livermore FORTRAN Kernels: A Computer Test of the Numerical Performance Range". Technical Report UCRL–55745, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA, USA, December 1986.

[Mid95] Samuel P. Midkiff. "Local Iteration Set Computation for Block-Cyclic Distributions". In *Proceedings of the 24th International Conference on Parallel Processing*; Editor: Constantine D. Polychronopoulos, volume II of *Software*, pages 77–84, Oconomowoc, WI, USA, August 1995. The Pennsylvania State University, CRC Press, Inc., Boca Raton, FL, USA.

[SPB94] Ernesto Yi-Hsin Su, Daniel J. Palermo, and Prithviraj Banerjee. "Processor Tagged Descriptors: A Data Structure for Compiling for Distributed-Memory Multicomputers". In *Proceedings of the 1994 International Conference on Parallel Architectures and Compilation Techniques*, pages 123–132, Montréal, Canada, August 1994.

[TR96] Ashwath Thirumalai and Jagannathan Ramanujam. "Efficient Computation of Address Sequences in Data-Parallel Programs Using Closed Forms for Basis Vectors". *Journal of Parallel and Distributed Computing*, 38(2):188–203, November 1996.

[Wol96] Michael J. Wolfe. **High Performance Compilers for Parallel Computing**. Addison-Wesley Publishing Company, Inc., Redwood City, CA 94065, USA, 1996. ISBN 0–8053–2730–4.