

AD-A267 094



NASA Contractor Report 191448

ICASE Report No. 93-16

# ICASE



## A SIMPLE PARALLEL PREFIX ALGORITHM FOR COMPACT FINITE-DIFFERENCE SCHEMES

Xian-He Sun  
Ronald D. Joslin



This document has been approved  
for public release and sale; its  
distribution is unlimited.

NASA Contract No. NAS1-19480  
April 1993

Institute for Computer Applications in Science and Engineering  
NASA Langley Research Center  
Hampton, Virginia 23681-0001

Operated by the Universities Space Research Association



National Aeronautics and  
Space Administration  
Langley Research Center  
Hampton, Virginia 23681-0001

47000? **93-16585**



3125

93 7 22 01 2

# A Simple Parallel Prefix Algorithm for Compact Finite-Difference Schemes \*

*Xian-He Sun*

*Ronald D. Joslin*

ICASE  
Mail Stop 132C

Theoretical Flow Physics Branch  
Mail Stop 156

NASA Langley Research Center  
Hampton, VA 23681-0001

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

## Abstract

A compact scheme is a discretization scheme that is advantageous in obtaining highly accurate solutions. However, the resulting systems from compact schemes are tridiagonal systems that are difficult to solve efficiently on parallel computers. Considering the almost symmetric Toeplitz structure, a parallel algorithm, *simple parallel prefix* (SPP), is proposed. The SPP algorithm requires less memory than the conventional LU decomposition and is highly efficient on parallel machines. It consists of a prefix communication pattern and AXPY operations. Both the computation and the communication can be truncated without degrading the accuracy when the system is diagonally dominant. A formal accuracy study has been conducted to provide a simple truncation formula. Experimental results have been measured on a MasPar MP-1 SIMD machine and on a Cray 2 vector machine. Experimental results show that the simple parallel prefix algorithm is a good algorithm for the compact scheme on high-performance computers.

---

\*This research was supported by the National Aeronautics and Space Administration under NASA contract NAS1-19480 while the first author was in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23681-0001.

# 1 Introduction

Recent technological advances have made it possible to build computers that contain thousands of processors and can obtain gigaflops ( $10^9$  floating-point operations per second) on real applications. Emerging parallel computers are designed to solve large problems and achieve better accuracy than could previously be obtained [19]. Parallel computers demand new models, new discretization methods, and new algorithms to explore the potential of high-performance computing.

Conventionally, partial differential equations (PDEs) are discretized by finite-difference or finite-element methods, and are solved by Gauss-Seidel, conjugate-gradient, or successive overrelaxation (SOR) methods. A new discretization method, the *compact finite-difference scheme* (or *compact-difference scheme*, *compact scheme*) was proposed by Kreiss and Olinger [10] and was later improved upon by Lele [14]. Compared with the traditional finite-difference scheme, the compact finite-difference scheme achieves higher accuracy with smaller difference stencils and leads to more accurate approximations because of the smaller coefficients on the truncation error. With these advantages, the compact scheme has quickly gained popularity. In practice, the resulting discretized system of the compact schemes are tridiagonal systems that can be solved efficiently on sequential machines. However, tridiagonal systems are difficult to solve efficiently on parallel computers. For example, to study the physics of compressible homogeneous turbulence, the CDNS (compressible direct simulation of Navier-Stokes) code, based on a sixth-order compact scheme and a third-order time discretization, was implemented on the Intel Delta [5]. After carefully choosing an existing tridiagonal solver, mapping the algorithm to the architecture, and overlapping communication with computation, the communication overhead consumed about 30 percent of the total execution time. Clearly, more efficient algorithms are needed to explore the potential of compact schemes on parallel computers.

In recent years, intensive research has been done to develop efficient tridiagonal solvers on parallel computers. A good survey can be found in references [15], [7], and [12]. Of the known tridiagonal solvers, both the recursive-doubling reduction method (RCD) developed by Stone [17], and the odd-even, or cyclic, reduction method (OER) developed by Hockney [8] are able to solve an  $n$ -dimensional tridiagonal system in  $O(\log(n))$  time using  $n$  processors. These methods are designed for fine-grain computing. Substructured methods developed by Lawrie and Sameh [13], Wang [24], and Sun, Zhang, and Ni [23] were designed for median-grain and coarse-grain computing (i.e., the case of  $p < n$  or  $p \ll n$ , where  $p$  is the number of processors available). Lawrie and Sameh's algorithm is designed for shared-memory machines; Wang's algorithm is designed for distributed-memory machines; and Sun et al. proposed three different algorithms, each of which may be a better choice depending on the problem and the machine. For compact schemes, the tridiagonal systems have a special structure that consists of diagonal dominance and are almost symmetric Toeplitz. For this special structure, a parallel tridiagonal solver for fine-grain computing, *the simple parallel*

*prefix* (SPP) algorithm, is proposed in this paper.

Compared with the popular tridiagonal solver for fine-grain computing known as the cyclic reduction method [8], the SPP method is simple to implement and computationally efficient. It requires only  $2 \cdot \log(n)$  AXPY (vector plus scalar times vector) operations and prefix communication patterns. If the tridiagonal system is diagonally dominant, then the AXPY operations can be truncated after a certain number of steps without degrading the accuracy. A formal accuracy analysis is conducted and simple formulas are provided to compute the number of AXPY operations necessary.

This paper is organized as follows. Section 2 will present the compact scheme and discretized tridiagonal systems. Section 3 will introduce three versions of the simple parallel prefix algorithm: the SPP for tridiagonal systems with the given special structure, the SPP for solving symmetric Toeplitz tridiagonal systems, and the SPP for solving almost symmetric Toeplitz systems. Accuracy analysis will be conducted in Section 4. Section 5 will give experimental results on a 16K processing elements (PEs) MasPar SIMD computer and on a Cray 2 supercomputer. Finally, Section 6 will give the conclusions.

## 2 Compact Finite-Difference Schemes

With the conventional finite-difference and finite-element discretization methods, as the order of the approximation increases, the required number of boundary and near-boundary relations and the required number of mesh points per derivative stencil increases accordingly. To achieve higher accuracy without the use of additional mesh points, the compact scheme [10] was introduced. As originally suggested by Kreiss and Olinger [10], and later discussed for fluid dynamics problems by Hirsh [6], the first and second derivatives for compact differences may be approximated by

$$f'_n = \left( \frac{D_0}{1 + \frac{1}{6}h_x^2 D_+ D_-} \right) f_n \quad \text{and} \quad f''_n = \left( \frac{D_+ D_-}{1 + \frac{1}{12}h_x^2 D_+ D_-} \right) f_n, \quad (1)$$

where

$$D_0 f_n = \frac{1}{2h_x} (f_{n+1} - f_{n-1}), \quad D_+ f_n = \frac{1}{h_x} (f_{n+1} - f_n),$$

$$D_- f_n = \frac{1}{h_x} (f_n - f_{n-1}),$$

and  $h_x$  is the mesh spacing, which is constant for simplicity. By multiplying (1) by the respective denominators, relations for the derivatives may be found, which yield

$$\frac{1}{6}f'_{n-1} + \frac{2}{3}f'_n + \frac{1}{6}f'_{n+1} = \frac{1}{2h_x} (f_{n+1} - f_{n-1}), \quad (2)$$

and

$$\frac{1}{12}f''_{n-1} + \frac{5}{6}f''_n + \frac{1}{12}f''_{n+1} = \frac{1}{h_x^2}(f_{n+1} - 2f_n + f_{n-1}). \quad (3)$$

These equations yield tridiagonal systems when the appropriate boundary conditions are applied.

To make an accurate comparison between the compact-difference (eqs. (2) and (3)) and the standard central-difference schemes, Taylor-series expansions are employed. As Hirsh has shown, the truncation error for the compact differences are

$$E(f'_n) = -\frac{1}{180}h_x^4 f^{(v)} \quad \text{and} \quad E(f''_n) = -\frac{1}{240}h_x^4 f^{(vi)}.$$

Similar error analyses for the central differences yield

$$E(f'_n) = -\frac{1}{30}h_x^4 f^{(v)} \quad \text{and} \quad E(f''_n) = -\frac{1}{90}h_x^4 f^{(vi)}.$$

Although both schemes are fourth-order accurate, the compact-difference scheme should lead to more accurate approximations as a result of the smaller coefficients on the truncation error. Similar results hold for other higher order approximations.

As yet, no mention has been made about the boundary treatment for the compact scheme. At the boundaries, Hirsh [6] experimented with a variety of boundary conditions, and Adams [1] suggested a boundary relation that includes near-boundary derivatives in the formulation. The boundary conditions used by both Hirsh and Adams retained the tridiagonal nature of the system. To demonstrate the SPP algorithm, fourth-order one-sided finite differences will be used for boundary conditions.

Many relevant fluid dynamics applications can make use of high-order compact-difference operators to numerically solve the governing systems of equations. For example, Burger's equation, the boundary-layer equations, and the driven cavity problem were solved by Hirsh [6] with compact-difference operators. Further, Joslin et al. [9] used the compact-difference equations (2) and (3) to numerically solve the fully nonlinear Navier-Stokes equations of fluid dynamics.

$$\frac{\partial u_i}{\partial x_i} = 0, \quad (4)$$

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j}, \quad (5)$$

where  $u_i = (u_1, u_2, u_3)$  are the velocity components that correspond to the streamwise, normal, and spanwise directions,  $x_i = (x_1, x_2, x_3)$ ;  $\nu$  is the kinematic viscosity,  $\rho$  is the density, and repeated indices infer a summation over the index. To demonstrate how compact-difference operators could be employed to solve the nonlinear PDE's (4) and (5), consider the model problem of the one-



### 3 The Simple Parallel Prefix Algorithm

We are interested in solving a tridiagonal linear system of equations

$$Ax = d. \quad (10)$$

In this system of equations,  $A$  is either a symmetric Toeplitz tridiagonal system of order  $n$

$$A = \begin{bmatrix} c & 1 & & & \\ 1 & c & 1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & 1 \\ & & & 1 & c \end{bmatrix} = [1, c, 1], \quad (11)$$

or an almost symmetric Toeplitz tridiagonal system of order  $n$

$$A = \begin{bmatrix} a & 0 & & & \\ 1 & c & 1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & 1 \\ & & & 1 & c \end{bmatrix}, \quad (12)$$

where  $x = (x_1, \dots, x_n)^T$  and  $d = (d_1, \dots, d_n)^T$  are  $n$ -dimensional vectors. We assume that matrix  $A$  is diagonally dominant (i.e.,  $|c| > 2$ ). Although we assume that  $A$ ,  $x$ , and  $d$  have real coefficients, the extension to the complex case is straightforward.

#### 3.1 The Simple Prefix Method

In this section, we study how to efficiently solve the system

$$\tilde{A}\tilde{x} = d \quad (13)$$

on parallel computers, where

$$\tilde{A} = \begin{pmatrix} a & 1 & & & \\ 1 & c & \ddots & & \\ & \ddots & \ddots & 1 & \\ & & & 1 & c \\ & & & & 1 \end{pmatrix} = a \begin{pmatrix} 1 & & & & \\ b & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & b & 1 \\ & & & & 1 \end{pmatrix} \begin{pmatrix} 1 & b & & & \\ & \ddots & \ddots & & \\ & & \ddots & b & \\ & & & \ddots & 1 \end{pmatrix} = a \cdot [b, 1, 0] \cdot [0, 1, b],$$

and  $a$  and  $b$  are the real solutions of:

$$\begin{cases} a + b = c \\ a \cdot b = 1 \end{cases} \quad (14)$$

Because  $a \cdot b = 1$  and  $|c| > 2$ , we may further assume that  $|a| > 1$  and  $|b| < 1$ . By equation (13),

$$\tilde{x} = a^{-1} \cdot [0, 1, b]^{-1} [b, 1, 0]^{-1} d = b \cdot [0, 1, b]^{-1} [b, 1, 0]^{-1} d.$$

Let  $L = [-b, 0, 0]$ . Then

$$[b, 1, 0] = [0, 1, 0] - [-b, 0, 0] = I - L$$

and

$$[b, 1, 0]^{-1} = (I + L + L^2 + \dots + L^{n-1}) \quad (15)$$

$$= (I + L^{2^{\lceil \log n \rceil - 1}})(I + L^{2^{\lceil \log n \rceil - 2}}) \dots (I + L^4)(I + L^2)(I + L). \quad (16)$$

Note that  $n$  is the dimension of matrix  $\tilde{A}$  and that equations (15) and (16) can be verified directly. The superscript of matrix  $L$  represents matrix multiplication

$$L^i = \begin{pmatrix} 0 & & & & \\ 0 & 0 & & & \\ (-b)^i & 0 & 0 & & \\ 0 & \ddots & \ddots & \ddots & \\ 0 & 0 & (-b)^i & 0 & 0 \end{pmatrix},$$

where the first nonzero element  $(-b)^i$  is at position  $(i + 1, 1)$ . Similarly, let  $U = [0, 0, -b]$ . Then,

$$[0, 1, b] = [0, 1, 0] - [0, 0, -b] = I - U,$$

and

$$[0, 1, b]^{-1} = (I + U + U^2 + \dots + U^{n-1})$$

$$= (I + U^{2^{\lceil \log n \rceil - 1}}) \dots (I + U^2)(I + U),$$

where

$$U^i = \begin{pmatrix} 0 & 0 & (-b)^i & 0 & 0 \\ & \ddots & \ddots & \ddots & 0 \\ & & 0 & 0 & (-b)^i \\ & & & 0 & 0 \\ & & & & 0 \end{pmatrix}.$$

The first nonzero element of  $U^i$  is at position  $(1, i + 1)$ . Thus, the solution of equation (13) is

$$\tilde{x} = b \cdot (I + U^{2^{\lceil \log n \rceil - 1}}) \dots (I + U) \cdot (I + L^{2^{\lceil \log n \rceil - 1}}) \dots (I + L)d \quad (17)$$

Equation (17) was first used by Chung and Shen [2] in solving  $b$ -spline curve fitting on a Cray X/MP computer.

Let  $v = (v_1, v_2, \dots, v_n)^T$  be an  $n$ -dimensional vector. Given the special structure of  $L^i$ , we find

$$(I + L^i)v = v + (-b)^i v_{(i)},$$

where

$$v_{(i)} = (0, \dots, 0, v_1, \dots, v_{n-i})^T$$

and  $v_1$  is the  $i + 1$  element of  $v_{(i)}$ . Similarly, given the special structure of  $U^i$ , we find

$$(I + U^i)v = v + (-b)^i v^{(i)},$$

where

$$v^{(i)} = (v_{i+1}, \dots, v_n, 0 \dots 0)^T.$$

Equation (17) shows that equation (13) can be solved with  $2 \cdot \lceil \log n \rceil$  AXPY operations. Because  $|b| < 1$ ,  $\|L^i\| \rightarrow 0$  and  $\|U^i\| \rightarrow 0$  when  $n \rightarrow \infty$ , the AXPY operation may be truncated without influencing the accuracy. Formulas will be given in Section 4 to compute the smallest truncation integer  $k$ . A sequential Fortran like code for solving equation (17) within truncation error is given in Figure 1.

```

Do 20  i = 1, k
      Do 20  j = n, 2i + 1, -1
          dj = dj + (-b)i dj-2i

Do 40  i = 1, k
      Do 40  j = 1, n - 2i
          dj = dj + (-b)i dj+2i

Do 60  j = 1, n
      x̄j = b · dj

```

Figure 1. The simple prefix method.

If  $n$  processing elements are available and  $d_j$  is stored in processor  $j$ , then the loop 20 and loop 40 will lead to prefix computations. Figure 2 shows the prefix computation pattern that correspond to loop 40 when  $n$  is equal to 8. Loop 20 leads to a similar prefix computation pattern, except that the communication is from right to left. Prefix (or recursive doubling) computation is a widely used computation model in scientific computing. Any linear recursive relation can be computed by recursive doubling [21]. A recursive-doubling algorithm exists for solving tridiagonal

systems and involves matrix-matrix multiplications [4, 23]. Compared with the existing recursive-doubling algorithm, the computation of our prefix method for solving tridiagonal systems (13) is simple. We call the prefix method given in Figure 1 the simple prefix method. Figure 3 shows the communication pattern of the widely used cyclic reduction method [8]. In a comparison of figures 2 and 3, we can see that the prefix method has a relatively simple communication pattern.

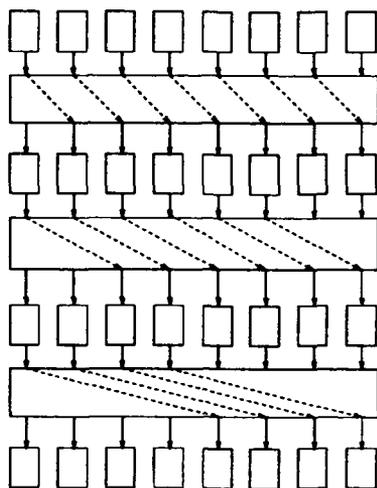


Figure 2. Communication of prefix computation.

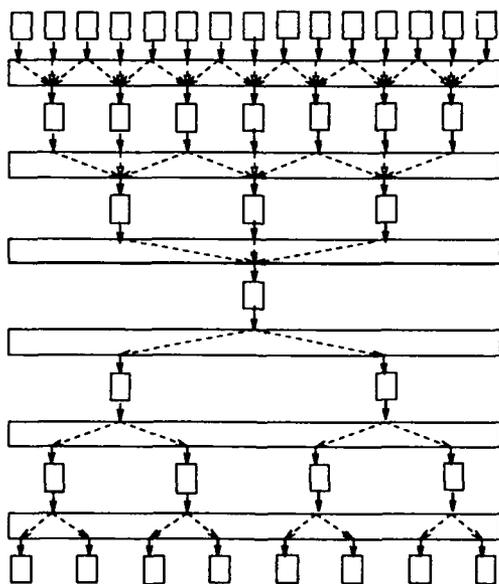


Figure 3. Communication of cyclic reduction method.

### 3.2 Modification of Symmetric Toeplitz System

Our goal is to find the solution of equation (10). Modification is needed to convert the solution of equation (13) to the desired solution. The modification will be different for symmetric Toeplitz sys-

tems and for almost symmetric Toeplitz systems. For a given symmetric Toeplitz system, equation (13) is modified as

$$A = \tilde{A} + \Delta A = \tilde{A} + V E^T,$$

where

$$\Delta A = \begin{pmatrix} b & 0 & & & \\ 0 & 0 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & 0 & 0 \\ & & & 0 & 0 \end{pmatrix}, \quad (18)$$

and  $V = (b, 0, \dots, 0)^T$  and  $E = (1, 0, \dots, 0)^T$  are  $n$ -dimensional vectors. By the matrix-modification formula [16, 3, 23], equation (10) can be solved by

$$\begin{aligned} x &= A^{-1}d = (\tilde{A} + V E^T)^{-1}d \\ x &= \tilde{A}^{-1} - \tilde{A}^{-1}V(I + E^T \tilde{A}^{-1}V)^{-1}E^T \tilde{A}^{-1}d \\ &= \tilde{x} - \tilde{A}^{-1}V(I + E^T \tilde{A}^{-1}V)^{-1}\tilde{x}_1, \end{aligned}$$

where  $\tilde{x}_1$  is the first element of vector  $\tilde{x}$ . If the calculation approach given in [18] is followed, we have

$$(I + E^T \tilde{A}^{-1}V)^{-1} = \frac{1}{1 + \sum_{i=1}^n b^{2i}}$$

and

$$\tilde{A}^{-1}V = \left( \sum_{i=1}^n b^{2i}, \sum_{i=1}^{n-1} (-b)b^{2i}, \sum_{i=1}^{n-2} (-b)^2 b^{2i}, \dots, (-b)^{n+1} \right).$$

Thus

$$\begin{aligned} &\tilde{A}^{-1}V(I + E^T \tilde{A}^{-1}V)^{-1} \\ &= b^2 \left( \frac{1 - b^{2n}}{1 - b^{2(n+1)}}, (-b) \frac{1 - b^{2(n-1)}}{1 - b^{2(n+1)}}, \dots, (-b)^i \frac{1 - b^{2(n-i)}}{1 - b^{2(n+1)}}, \dots, (-b)^{n-1} \frac{(1 - b^2)}{1 - b^{2(n+1)}} \right)^T. \end{aligned} \quad (19)$$

The final solution is

$$x = \tilde{x} - \tilde{x}_1 z, \quad (20)$$

where vector  $z$  is the right side of equation (19). Because  $|b| < 1$ ,  $z$  can be truncated at some integer  $k_1$  without affecting the accuracy (see section 4). Furthermore, when  $n$  is large,  $b^{2(n-i+1)}$ ,  $i = 0, 1, \dots, k_1$  will be less than machine accuracy, and  $z$  reduces to

$$\tilde{z} = ((-b)^2, (-b)^3, \dots, (-b)^{k_1+2}, 0, \dots, 0)^T.$$



$\tilde{V} = (-1, 0, \dots, 0)^T$ , and  $\tilde{E} = (0, 1, 0, \dots, 0)^T$ . Following the matrix modification formula [23, 18], the solution to equation (10) becomes

$$\begin{aligned} x &= A^{-1}d = \tilde{A}^{-1}d - \tilde{A}^{-1}\tilde{V}(I + \tilde{E}^T\tilde{A}^{-1}\tilde{V})^{-1}\tilde{E}^T\tilde{A}^{-1}d \\ &= \tilde{x} - \tilde{A}^{-1}\tilde{V}(I + \tilde{E}^T\tilde{A}^{-1}\tilde{V})^{-1}\tilde{x}_2, \end{aligned}$$

where  $\tilde{x}_2$  is the second element of  $\tilde{x}$ . With this new modification,

$$\begin{aligned} (I + \tilde{E}^T\tilde{A}^{-1}\tilde{V})^{-1} &= \frac{a}{a - \sum_{i=0}^{n-2} (-b)b^{2i}}, \\ \tilde{A}^{-1}\tilde{V} &= -b\left(\sum_{i=0}^{n-1} b^{2i}, \sum_{i=0}^{n-2} (-b)b^{2i}, \dots, \sum_{i=0}^{n-j} (-b)^{j-1}b^{2i}, \dots, (-b)^{n-1}\right), \\ \tilde{A}^{-1}\tilde{V}(I + \tilde{E}^T\tilde{A}^{-1}\tilde{V})^{-1} &= (-b, \dots, (-b)^i \frac{1 - b^{2(n-i+1)}}{1 - b^{2n}}, \dots, (-b)^n \frac{1 - b^2}{1 - b^{2n}})^T. \end{aligned} \quad (22)$$

The final solution is

$$x = \tilde{x} - \tilde{x}_2 y, \quad (23)$$

where the vector  $y$  is the right side of equation (22). Like the symmetric case, only the addition of the first  $k_2$  elements in equation (23) may be needed for a given accuracy for some integer  $k_2$ , and  $|b|^{2(n-i)}$  may be less than machine accuracy for  $0 \leq i < k_2$  when  $n$  is large. Let

$$\tilde{y} = (-b, (-b)^2, \dots, (-b)^{k_2}, 0, \dots, 0)^T. \quad (24)$$

The modification computation for the almost symmetric Toeplitz system is given in Figure 6. The algorithm for solving the almost symmetric Toeplitz systems is similar to the algorithm for solving the symmetric Toeplitz system (see Figure 5), except in step 2 the new modification (Figure 6) is used to replace the symmetric Toeplitz system modification.

Do 100  $i = 1, k_2$   
 $x = \tilde{x}_i - \tilde{x}_2 \tilde{y}_i$

Figure 6. Modification for almost symmetric Toeplitz system.

## 4 Accuracy Analysis

In a previous section, the claim was made that the AXPY operations and modifications can be truncated without influencing the accuracy. In this section, we will study the truncation accuracy. For simplicity, we truncate loop 20 and loop 40 at the same step  $k$ . The norm used in this section is the  $L_1$ -norm.

### 4.1 Accuracy Study of the Simple Prefix Method

Let  $\tilde{x}, y$  be exact solutions as

$$\begin{aligned} y &= b \cdot [b, 1, 0]^{-1} d, \\ &= (I + L + L^2 + \dots + L^{n-1}) b \cdot d, \\ \tilde{x} &= b \cdot [0, 1, b]^{-1} [b, 1, 0]^{-1} \cdot d \\ &= (I + U + U^2 + \dots + U^{n-1}) \cdot y. \end{aligned}$$

Let  $y^*$  and  $\tilde{x}'$  be the solutions given by loop 20 and loop 40, respectively, as

$$y^* = (I + L + \dots + L^{k-1}) \cdot b \cdot d,$$

where  $k$  is a power of 2, and

$$\tilde{x}' = (I + U + \dots + U^{k-1}) y^*.$$

Also, we define  $\tilde{x}^*$  as a hypothetical solution by

$$\tilde{x}^* = (I + U + \dots + U^{k-1}) y.$$

The difference between  $\tilde{x}$  and  $\tilde{x}^*$  is

$$\begin{aligned} \tilde{x} - \tilde{x}^* &= (I + U + \dots + U^{n-1}) y - (I + U + \dots + U^{k-1}) y \\ &= (U^k + U^{k+1} + \dots + U^{n-1}) y \\ &= (U^k + \dots + U^{n-1}) (I - U) \tilde{x} \\ &= (U^k - U^n) \tilde{x}. \end{aligned}$$

Thus, we find

$$\frac{\|\tilde{x} - \tilde{x}^*\|}{\|\tilde{x}\|} \leq \|U^k\| \cdot \|I - U^{n-k}\| \leq |b|^k (1 + |b|^{n-k}). \quad (25)$$

Equation (25) gives the relative error between  $\tilde{x}$  and  $\tilde{x}^*$ . Because the difference between  $\tilde{x}^*$  and  $\tilde{x}'$  is

$$\begin{aligned}
 \tilde{x}^* - \tilde{x}' &= (I + U + \dots + U^{k-1})y - (I + U + \dots + U^{k-1})y^* \\
 &= (I + U + \dots + U^{k-1})(y - y^*) \\
 &= (I + U + \dots + U^{k-1})(L^k + \dots + L^{n-1}) * b \cdot d \\
 &= (I + U + \dots + U^{k-1})(L^k + \dots + L^{n-1})(I - L)(I - U)\tilde{x} \\
 &= (I + \dots + U^{k-1})(L^k - L^n)(I - U)\tilde{x},
 \end{aligned}$$

the norm becomes

$$\frac{\|\tilde{x}^* - \tilde{x}'\|}{\|\tilde{x}\|} \leq \frac{1 - |b|^k}{1 - |b|} \cdot |b|^k \cdot (1 + |b|^{n-k})(1 + |b|). \quad (26)$$

If equations (25) and (26) are combined, then the relative error between the exact solution  $\tilde{x}$  and the truncated solution  $\tilde{x}'$  is

$$\begin{aligned}
 \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} &\leq \frac{\|\tilde{x} - \tilde{x}^*\|}{\|\tilde{x}\|} + \frac{\|\tilde{x}^* - \tilde{x}'\|}{\|\tilde{x}\|} \\
 &\leq |b|^k(1 + |b|^{n-k})\left(1 + \frac{1 - |b|^k}{1 - |b|}(1 + |b|)\right) \\
 &\leq 2 \cdot |b|^k \cdot (1 + |b|^{n-k}).
 \end{aligned}$$

## 4.2 Final Accuracy of Symmetric Toeplitz Systems

The truncation error of the simple prefix method (Figure 1) will be carried into the modification step and will influence the accuracy of the final solution. Let  $x$  be the solution of a symmetric Toeplitz tridiagonal system (10) and  $x^*$  be the corresponding solution that has been modified with the truncated solution,

$$\begin{aligned}
 x &= \tilde{x} - \tilde{x}_1 z, \\
 x^* &= \tilde{x}' - \tilde{x}'_1 z,
 \end{aligned}$$

where  $z = b^2 \left( \frac{1-b^{2n}}{1-b^{2(n+1)}}, (-b)\frac{1-b^{2(n-1)}}{1-b^{2(n+1)}}, \dots, (-b)^{n-1}\frac{(1-b^2)}{1-b^{2(n+1)}} \right)^T$ ; see equation (19). The error generated by this truncation is

$$\begin{aligned}
 \|x - x^*\| &= \|(\tilde{x} - \tilde{x}') + (\tilde{x}_1 - \tilde{x}'_1)z\| \\
 &\leq \|\tilde{x} - \tilde{x}'\| + \|\tilde{x}_1 - \tilde{x}'_1\| \cdot \|z\| \\
 &\leq \|\tilde{x} - \tilde{x}'\|(1 + \|z\|).
 \end{aligned}$$

The norm of vector  $z$  can be computed directly as

$$\begin{aligned}
\|z\| &= \frac{b^2}{1-b^{2(n+1)}} \sum_{i=0}^{n-1} |(b)^i(1-b^{2(n-i)})| \\
&\leq \frac{b^2}{1-b^{2(n+1)}} \left( \sum_{i=0}^{n-1} |b|^i + \sum_{i=0}^{n-1} |b|^{2n-i} \right) \\
&\leq \frac{b^2}{1-b^{2(n+1)}} \cdot \frac{(1+|b|^{n+1})(1-|b|^n)}{(1-|b|)} \\
&\leq b^2 \cdot \frac{(1-|b|^n)}{(1-|b|^{n+1})(1-|b|)} \leq \frac{b^2}{1-|b|} = \frac{|b|}{|a|-1}.
\end{aligned}$$

Therefore,

$$\|x - x^*\| \leq \|\tilde{x} - \tilde{x}'\| \left(1 + \frac{|b|}{|a|-1}\right) \quad (27)$$

$$= \|\tilde{x} - \tilde{x}'\| \left(\frac{|c|-1}{|a|-1}\right) \quad (28)$$

$$= \|\tilde{x} - \tilde{x}'\| \left(\frac{1-|b|+|b|^2}{1-|b|}\right) \leq \|\tilde{x} - \tilde{x}'\| \left(\frac{1}{1-|b|}\right), \quad (29)$$

and

$$\frac{\|\tilde{x} - \tilde{x}^*\|}{\|x\|} \leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|x\|} \left(\frac{1-|b|+|b|^2}{1-|b|}\right) \quad (30)$$

$$= \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} \left(\frac{1-|b|+|b|^2}{1-|b|}\right). \quad (31)$$

The only unknown in the right side of equation (31) is  $\frac{\|\tilde{x}\|}{\|x\|}$ , where  $\tilde{x}$  is the solution of equation (13) and  $x$  is the solution of equation (10). For a symmetric Toeplitz system,

$$A = \tilde{A} + \Delta A,$$

where  $\Delta A$  is given by equation (18). If equations (10) and (13) are combined, we have

$$(\tilde{A} + \Delta A)x = \tilde{A}\tilde{x},$$

$$(I + \tilde{A}^{-1}\Delta A)x = \tilde{x},$$

and

$$\frac{\|\tilde{x}\|}{\|\tilde{x}\|} \leq \|I + \tilde{A}^{-1}\Delta A\|. \quad (32)$$

After several calculations, we find

$$\tilde{A}^{-1} = b \begin{pmatrix} 1 & -b & b^2 & \cdot & (-b)^{n-1} \\ & 1 & -b & \cdot & (-b)^{n-2} \\ & & \cdot & \cdot & \cdot \\ & & & 1 & -b \\ & & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & & \\ -b & 1 & & & \\ b^2 & -b & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (-b)^{n-1} & \cdot & \cdot & -b & 1 \end{pmatrix}$$

and

$$\tilde{A}^{-1} \Delta A = \begin{pmatrix} \sum_{i=1}^n b^{2i} & 0 & \cdot & \cdot & 0 \\ \sum_{i=1}^{n-1} (-b)b^{2i} & 0 & \cdot & \cdot & 0 \\ \sum_{i=1}^{n-2} (-b)^2 b^{2i} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (-b)^{n+1} & 0 & \cdot & \cdot & 0 \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \|\tilde{A}^{-1} \Delta A\| &\leq \sum_{i=0}^{n-1} \left| (-b)^i \frac{b^2(1-b^{2(n-i)})}{1-b^2} \right| \\ &\leq \frac{b^2}{1-b^2} \sum_{i=0}^{n-1} |b^i(1-b^{2(n-i)})| \\ &\leq \frac{b^2}{1-b^2} \cdot \frac{(1+|b|^{n+1})(1-|b|^n)}{(1-|b|)} \end{aligned}$$

and

$$\|I + \tilde{A}^{-1} \Delta A\| \leq 1 + \frac{b^2}{(1-b^2)} \cdot \frac{(1+|b|^{n+1})(1-|b|^n)}{(1-|b|)}.$$

The relative error of the solution to equation (10) is

$$\frac{\|x - x^*\|}{\|x\|} \leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} \cdot \frac{1}{1-|b|} \quad (33)$$

$$\leq \frac{|b|^k(1+|b|^{n-k})}{1-|b|} \left( 1 + \frac{(1-|b|^k)(1+|b|)}{1-|b|} \right) \left( 1 + \frac{b^2(1+|b|^{n+1})(1-|b|^n)}{(1-b^2)(1-|b|)} \right). \quad (34)$$

When the order of the tridiagonal system  $n$  is large,  $|b|^n$  may be less than machine accuracy. In this case, the inequality (34) becomes

$$\frac{\|x - x^*\|}{\|x\|} \leq \frac{|b|^k(1+|b|^{n-k})}{1-|b|} \left( 1 + \frac{(1-|b|^k)(1+|b|)}{1-|b|} \right) \left( 1 + \frac{b^2}{(1-b^2)(1-|b|)} \right), \quad (35)$$

when truncation is applied in the modification. In addition to the truncation error carried from step 1, the truncated modification will also generate truncation error. If

$$x' = \tilde{x}' - \tilde{x}'_1 \tilde{z},$$

(Figures 4 and (20)), then

$$x - x' = \tilde{x} - \tilde{x}' - (\tilde{x}_1 - \tilde{x}'_1)\tilde{z} + \tilde{x}_1(z - \tilde{z}).$$

The  $L_1$ -norm is given by

$$\begin{aligned} \|x - x'\| &\leq \|\tilde{x} - \tilde{x}'\| + \|\tilde{x} - \tilde{x}'\| \cdot \|\tilde{z}\| + \|x\| \cdot \|z - \tilde{z}\| \\ &\leq \|\tilde{x} - \tilde{x}'\|(1 + \|\tilde{z}\|) + \|x\| \cdot \|z - \tilde{z}\|, \end{aligned}$$

which leads to

$$\frac{\|x - x'\|}{\|x\|} \leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} (1 + \|\tilde{z}\|) + \|z - \tilde{z}\|. \quad (36)$$

In equation (36),

$$\begin{aligned} \|z - \tilde{z}\| &= \frac{b^2}{1 - b^{2(n+1)}} \sum_{i=k_1}^{n-1} |b^i(1 - b^{2(n-i)})| \\ &= \frac{b^2}{1 - b^{2(n+1)}} \sum_{j=0}^{n-1-k_1} |b^{i+k_1}(1 - b)^{2(n-j)}|, \end{aligned}$$

where  $j = i - k_1$ ,  $n_1 = n - k_1$ . Then

$$\begin{aligned} \|z - \tilde{z}\| &\leq \frac{|b|^{k_1+2}}{1 - b^{2(n+1)}} \cdot \frac{(1 + |b|^{n_1+1})(1 - |b|^{n_1})}{(1 - |b|)} \\ &\leq \frac{|b|^{k_1+2}}{1 - b^{2(n_1+1)}} \frac{(1 + |b|^{n_1+1})(1 - |b|^{n_1})}{(1 - |b|)} \\ &\leq \frac{|b|^{k_1+2}}{1 - |b|}. \end{aligned}$$

Note that  $\|\tilde{z}\| \leq \|z\|$ , so that we have the inequality

$$\begin{aligned} \frac{\|x - x'\|}{\|x\|} &\leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \frac{\|\tilde{x}\|}{\|x\|} (1 + \|z\|) + \frac{|b|^{k_1+2}}{1 - |b|} \\ &\leq \frac{|b|^k}{1 - |b|} \left(1 + \frac{(1 - |b|^k)(1 + |b|)}{1 - |b|}\right) \left(1 + \frac{b^2}{(1 - b^2)(1 - |b|)}\right) + \frac{|b|^{k_1+2}}{1 - b}. \end{aligned}$$

The error introduced by the truncated modification is insignificant if  $k_1 > k - 2$ . In practice, we can choose  $k_1 = k$  and use the inequality (34) to compute the truncation number  $k$ .

### 4.3 Accuracy of Solving Almost Symmetric Toeplitz Systems

Following the similar arguments given in last section, error bounds can be obtained for almost symmetric Toeplitz systems. For almost symmetric systems, we have a different modification

vector  $y$  and a different modification matrix  $\Delta\tilde{A}$ . From equation (23), the exact solution to the almost symmetric Toeplitz system is

$$x = \tilde{x} - \tilde{x}_2 y,$$

where

$$y = (-b, (-b)^2 \frac{1 - b^{2(n-1)}}{1 - b^{2n}}, \dots, (-b)^n \frac{1 - b^2}{1 - b^{2n}}).$$

The norm of vector  $y$  is

$$\begin{aligned} \|y\| &= \sum_{i=1}^n \left| (-b)^i \frac{1 - b^{2(n-i+1)}}{1 - b^{2n}} \right| \\ &\leq \frac{1}{1 - b^{2n}} \left( \sum_{i=1}^n |b|^i + \sum_{i=1}^n b^{2(n+1)} \cdot |b|^{-i} \right) \\ &\leq \frac{1}{1 - b^{2n}} \left( \frac{|b|(1 - |b|^n)(1 + |b|^{n+1})}{(1 - |b|)} \right) \leq \frac{|b|}{1 - |b|}. \end{aligned}$$

Let  $x^*$  be the solution modified with the truncated solution as

$$x^* = \tilde{x}' - \tilde{x}'_2 y,$$

where  $\tilde{x}', \tilde{x}'_2$  are the same as defined in Section 4.1. Then, following a similar deduction for the inequality (29), we have

$$\begin{aligned} \|x - x^*\| &\leq \|\tilde{x} - \tilde{x}'\| (1 + \|y\|), \\ \frac{\|x - x^*\|}{\|x\|} &\leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} \left( 1 + \frac{|b|}{1 - |b|} \right) \\ &\leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} \left( \frac{1}{1 - b} \right). \end{aligned}$$

The above inequality is in the same form as inequality (31), except that the ratio  $\frac{\|\tilde{x}\|}{\|x\|}$  is different for the almost symmetric system. For the almost symmetric Toeplitz system,

$$x = \tilde{A} + \Delta\tilde{A},$$

where  $\Delta\tilde{A}$  is defined by equation (21). By direct calculation, we find

$$\tilde{A}^{-1} \Delta\tilde{A} = \begin{pmatrix} 0 & \sum_{i=0}^{n-1} -b \cdot b^{2i} & 0 & \cdot & 0 \\ 0 & \sum_{i=0}^{n-2} (-b)^2 b^{2i} & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & (-b)^n & 0 & \cdot & 0 \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \|\bar{A}^{-1}\Delta\bar{A}\| &\leq \sum_{i=0}^{n-1} \left| \frac{(-b)^{i+1}(1-b^{2(n-i)})}{(1-b^2)} \right| \\ &\leq \frac{|b|(1+|b|^{n+1})(1-|b|^n)}{(1-b^2)(1-|b|)}. \end{aligned}$$

For the almost symmetric system,

$$\frac{\|\tilde{x}\|}{\|x\|} \leq \|I + \bar{A}^{-1}\Delta\bar{A}\| \leq 1 + \|\bar{A}^{-1}\Delta\bar{A}\|.$$

(See inequality (32).) We conclude that

$$\frac{\|x - x^*\|}{\|x\|} \leq \frac{|b|^k(1+|b|^{n-k})}{1-|b|} \left( 1 + \frac{(1-|b|^k)(1+|b|)}{1-|b|} \right) \left( 1 + \frac{|b|(1+|b|^{n+1})(1-|b|^n)}{(1-b^2)(1-|b|)} \right) \quad (37)$$

$$\leq \frac{|b|^k}{1-|b|} \left( 1 + \frac{(1-|b|^k)(1+|b|)}{1-|b|} \right) \left( 1 + \frac{|b|}{(1-b^2)(1-|b|)} \right). \quad (38)$$

When compared with inequality (34), the numerator of the last term in inequality (38) is slightly increased.

If the modification is truncated to  $k_2$  (Figure 6), then

$$x' = \tilde{x}' - \tilde{x}'_1 \tilde{y}.$$

Following the derivation of inequality (36), we have

$$\frac{\|x - x'\|}{\|x\|} \leq \frac{\|\tilde{x} - \tilde{x}'\|}{\|\tilde{x}\|} \cdot \frac{\|\tilde{x}\|}{\|x\|} \cdot (1 + \|\tilde{y}\|) + \|y - \tilde{y}\|, \quad (39)$$

where

$$\begin{aligned} \|y - \tilde{y}\| &= \sum_{i=k_2}^{n-1} \left| \frac{(-b)^{i+1}(1-b^{2(n-i)})}{1-b^{2n}} \right| \\ &\leq \frac{|b|}{1-b^{2n}} \sum_{i=k_2}^{n-1} |b^i(1-b^{2(n-i)})| \\ &\leq \frac{|b|^{k_2+1}}{1-b^{2n}} \frac{(1+|b|^{n-k_2-1})(1-|b|^{n-k_2})}{(1-|b|)} \\ &\leq \frac{|b|^{k_2+1}}{1-|b|}. \end{aligned}$$

Thus,

$$\frac{\|x - x'\|}{\|x\|} \leq \frac{|b|^k}{1 - |b|} \left( 1 + \frac{(1 - |b|^k)(1 + |b|)}{1 - |b|} \right) \left( 1 + \frac{|b|}{(1 - b^2)(1 - |b|)} \right) + \frac{|b|^{k_2+1}}{1 - |b|}. \quad (40)$$

Similar to the symmetric case, the error introduced by modification truncation is insignificant if  $k_2 > k - 1$ . We could choose  $k_2 = k$  and use inequality (38) for error estimates.

## 5 Experimental Results

In this section, the performance of the SPP algorithm on the MasPar parallel computer and on the Cray vector computer will be presented.

### 5.1 Parallel Computing

The MasPar MP-1 is a distributed-memory massively parallel SIMD computer with a high-speed two-dimensional toroidal mesh topology. A control unit (ACU) has a direct connection to all the processing elements (PEs) and issues instructions at a 12.5 MHz clock rate. Each processing element in the array is a 4-bit custom load and storage processor with a minimum of 16 kilobytes of memory. Communication is relatively cheap on the MasPar. For example, on the MasPar M-1, a double-precision multiplication function is ten times more expensive than sending the product to an adjacent PE.

Table 1 gives the computation and communication count of the simple parallel prefix (SPP) algorithm. We assume that both the AXPY operations and the modification are truncated after  $k$  operations; the modification vector used is either equation (20) or (24), respectively. The  $b^{2^i}$ ,  $i = 1, \dots, k$ , are computed sequentially, or redundantly, on each PE. The modification vector  $\bar{z}$  or  $\bar{y}$  will be computed concurrently on different PEs. We count a power-function computation as 8 floating-point operations. So, the modification phase costs 10 parallel operations in total. The calculation of  $b$  (equation (14)) is not considered. In the computing phase,  $2 \cdot \log(k)$  parallel, one-to-one communications are required. We use  $\alpha$  to represent the one-to-one communication. On the MasPar, the one-to-one communication is achieved by using the *router* command. We use  $\beta$  to represent the broadcast. On the MasPar, the broadcast is achieved by transferring the local data to ACU and then distributing it to all PEs. One broadcast communication is needed in the modification phase. Because the tridiagonal systems that arise in the compact scheme have multiple right sides, the computation and communication count for solving multiple right-side systems is also listed in Table 1, where the computation of  $b^{2^i}$  and the modification vector are not considered. Note that  $n_1$  is the number of right sides.

Two sample matrices are chosen to illustrate the performance of the SPP algorithm and to verify the theoretical error bounds given in the previous section. Both of the matrices are almost

Table 1. Computation and communication count of the Simple Prefix Algorithm

System	Best sequential	SPP	
		Computation	Communication
Single system	$8n-7$	$5 \cdot \log(k) + 10$	$2 \cdot \log(k) \cdot \alpha + \beta$
Multiple right sides	$(5n-3) * n1$	$(4 \cdot \log(k) + 2) * n1$	$(2 \cdot \log(k) \cdot \alpha + \beta) * n1$

symmetric Toeplitz matrices that arise in the compact schemes. One matrix is

$$A_1 = [1, 3, 1] - \Delta \tilde{A}.$$

The other matrix is

$$A_2 = [1, 10, 1] - \Delta \tilde{A}.$$

Equation (18) defines  $\Delta \tilde{A}$ . The corresponding solution of equation (14) is  $b = \frac{3-\sqrt{5}}{2}$  and  $b = \frac{7-\sqrt{45}}{2}$  for  $A_1$  and  $A_2$ , respectively. The error is measured relative to the LU solution. The accuracy comparison for solving system  $A_1$  is given in Figure 7. In this implementation, no truncation is implemented in the modification phase, and the prediction formula used is equation (38). In solving  $A_2$ , the modification is applied at the modification stage with  $k_2 = k$ , and the prediction formula used is equation (39). The accuracy comparison for solving system  $A_2$  is given in Figure 8. From Figures 7 and 8, we can see that the theoretical bound matches the measured results well.

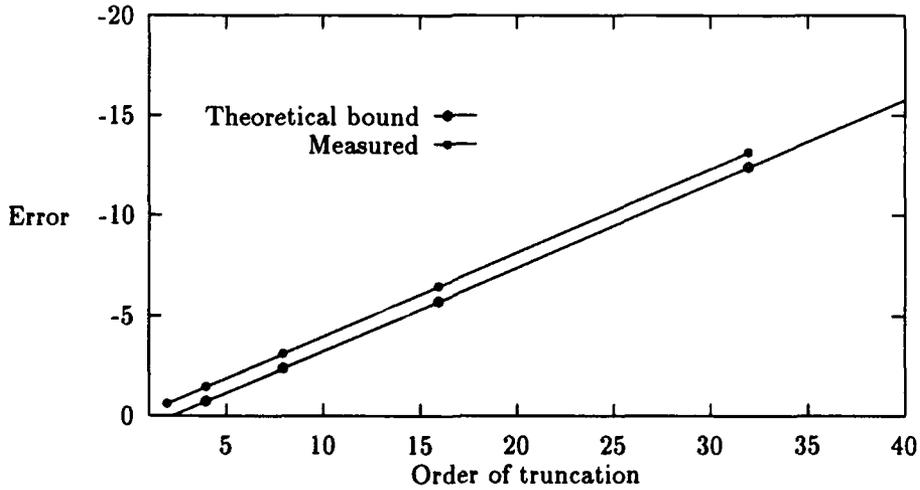


Figure 7. Measured and predicted accuracy for solving matrix  $A_1$ .

Figure 9 shows the speedup of the SPP algorithm over the best sequential algorithm for solving

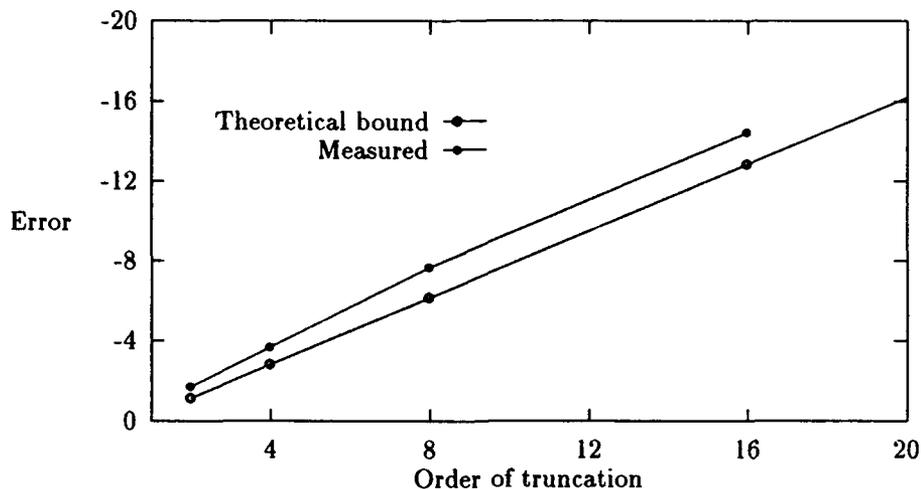


Figure 8. Measured and predicted accuracy for solving matrix  $A_2$ .

a single system. The best sequential algorithm is based on the LU decomposition, and is fine-tuned to take advantage of the almost symmetric Toeplitz structure. All computations are double precision. Truncation numbers  $k = 64$  and  $k = 16$  are chosen to achieve double-precision ( $10^{-16}$ ) and single-precision ( $10^{-7}$ ) accuracy, respectively. The order of matrix is the same as the number of PEs available. Because of the high-speed communication, with  $n$  equal to  $1k$ ,  $2k$ ,  $4k$ ,  $8k$ , and  $16k$ , the execution time is not noticeably changed in parallel processing. The sequential algorithm is implemented on a single PE. Because of memory limitations, only small systems are solved by the sequential algorithm. The data used in Figure 9 is predicted based on the small-size timing. Figure 10 shows the corresponding speedup of solving a system with 1024 right sides. The factorization of the matrix is not included in timing for solving the system with multiple right sides. The speedup is slightly higher for solving multiple right-side systems.

Because the order of the matrix increases linearly with the number of PEs available, the speedups given by Figures 9 and 10 are memory-bounded speedup [20]. From table 1, the problem size, in terms of floating-point operations, is a linear function of the order of the matrix. The linear memory-bounded speedups given by Figures 9 and 10 indicate a linear speed increase. In accordance with the isospeed metric of scalability [22], the SPP algorithm is perfectly scalable on the SIMD MasPar machine. Because the isoefficiency metric [11] is equivalent to the isospeed metric when serial execution has a fixed speed [22], the SPP algorithm is also perfectly scalable with the isoefficiency metric. The reader may refer to [22] and [11] for more information regarding scalability of parallel algorithm-machine combinations.

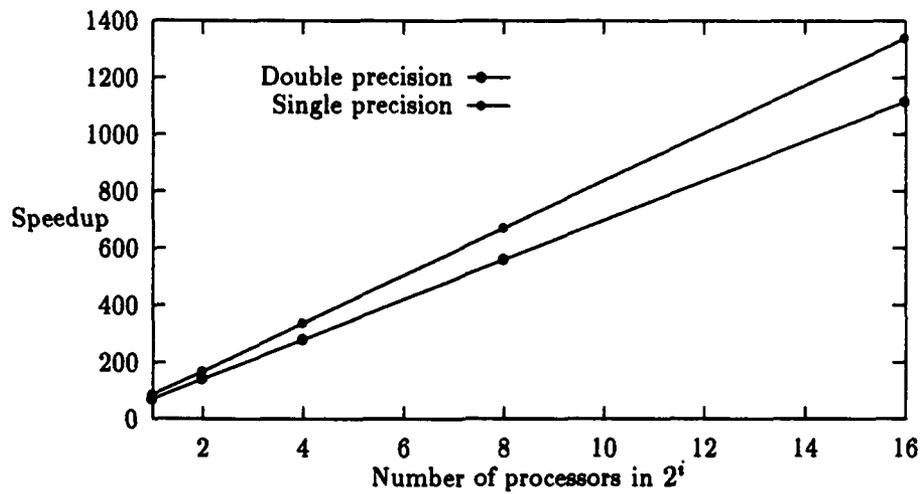


Figure 9. Speedup over the best sequential algorithm on single system.

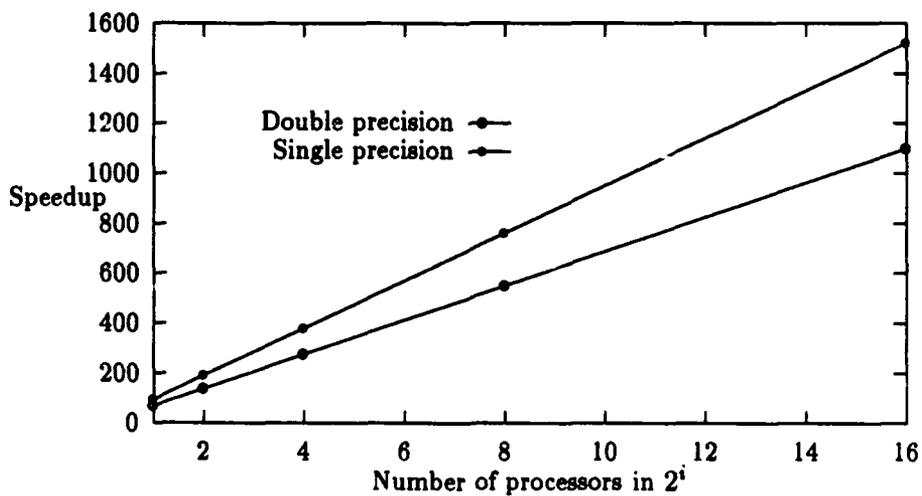


Figure 10. Speedup over the best sequential algorithm on system with 1024 right sides.

## 5.2 Vector Computing

Vector computing is widely used at national laboratories, universities, and supercomputing centers for large-scale computing applications. For this reason, a CRAY-2S/4-128 at NASA Langley Research Center was also used to test the SPP algorithm on a vector machine. The Cray-2 notation "S" indicates that the memory is static rather than dynamic, and "4-128" indicates that the machine has 4 processors and 128 million 64-bit words of central memory. Each CPU is a register-to-register vector processor with a 4.1 nsec minor cycle clock that can generate 100-300 megaflops. The four processors can be used for a single problem (multi-tasking) to achieve over 1 gigaflop of performance.

The speed of a vector machine depends on the vector length, vector stride, and the computational richness of the do loops. Because the vector register length is 64 and the CPU is extremely fast in carrying out floating-point operations, once operands are in the registers, best performance can be obtained with do-loop which have lengths that are multiples of 64, which are computationally intensive, and which use unit stride (separation of memory between elements).

The chosen sample tridiagonal matrices are almost symmetric Toeplitz and correspond to the first and second derivative compact-difference operators (2) and (3). The diagonals and necessary coefficients are:

$$A_3 = [1, 4, 1] \quad \text{with } a, b = 2 \pm \sqrt{3}$$

$$A_2 = [1, 10, 1] \quad \text{with } a, b = 5 \pm 2\sqrt{6}$$

For the first experiment on the Cray, single tridiagonal solves were made. The test problem used here and in the rest of this section corresponds to  $f(x) = 3x^3 - 2x + 1$ , which has smooth exact derivatives  $f'(x) = 9x^2 - 2$  and  $f''(x) = 18x$ . Figure 11 shows the performance of the SPP in terms of CPU seconds and the matrix order compared with the LU decomposition for computing  $f'$  and  $f''$ . In addition to the reduced memory requirements of SPP compared to LU, the performance shown in Figure 11 clearly indicated that the SPP is faster on the vector machine than the conventional LU solver; the benefits increase with the operator size. The significant difference between the SPP and LU timing can be explained in light of vector operations versus scalar operations. The SPP approach can be vectorized over the direction of the solve; the LU approach must use scalar operations. For the SPP approach, note that the diagonal dominance of the second-derivative operator  $f''$  leads to faster computations compared with the first-derivative operator  $f'$ . This time reduction results from the truncation of the SPP approach to obtain a predetermined level of error ( $10^{-14}$ ), which is essentially machine precision. For the first-derivative operator ( $A_3$ ),  $k = 32$  and  $k_1 = 24$ ; for the second-derivative operator ( $A_2$ ),  $k = 16$  and  $k_1 = 16$ , where  $k$  and  $k_1$  are the truncation numbers on the solving and modification phases, respectively.

Real applications which use compact-difference operators require many tridiagonal solves that

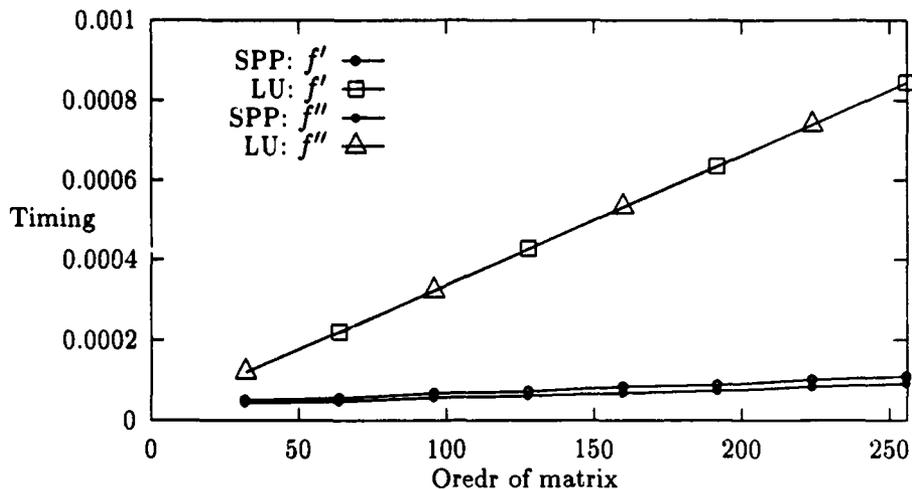


Figure 11. Timing of SPP and LU algorithms: single system.

correspond to time-marching algorithms and involve many right sides corresponding to the multidimensionality of the application. In this second evaluation, with the same accuracy  $10^{-14}$ , the performance of the SPP is compared with LU for multiple right sides. Shown in Figure 12 are CPU times for the SPP and LU for various orders of the second-derivative compact operator  $A_2$ . (Similar results were obtained with the operator  $A_3$  but are not shown.) For applications that use small operators ( $N < 96$ ), the LU solver is more efficient than SPP; for applications that use large operators ( $N > 96$ ), the SPP is much cheaper than the LU approach. This difference occurs because the LU approach vectorizes the do loop associated with the number of right sides, and the SPP vectorizes in the direction of the tridiagonal solve. With some creative programming, one could potentially vectorize the entire SPP approach with a single array, while the LU approach can vectorize over the right-side arrays.

In the final experiment, the ability of SPP to control truncation error is demonstrated. The highest order of accuracy in the solution is based on the truncation error of the compact-difference approaches in equations (2) and (3). As a result, to require machine-zero is overkill for the compact solver and leads to unnecessary computational cost. By using the inequality (40), the choice of truncation can be determined based on a desired error bound. Figure 13 shows the SPP results of truncations  $k = 8$  and  $k = 32$ , which correspond to errors  $10^{-5}$  and  $10^{-14}$ , respectively. If the accuracy of the SPP is relaxed, the computational cost decreases by a factor of 2.

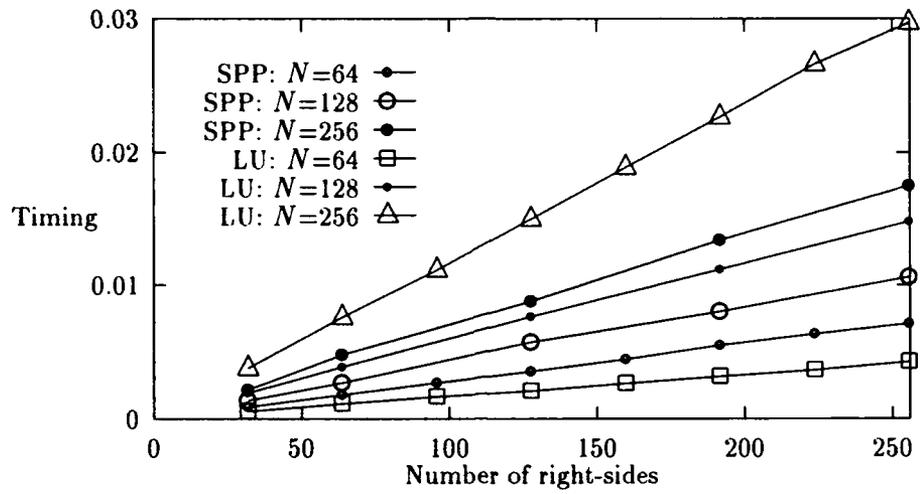


Figure 12. Timing of SPP and LU algorithms: multiple right sides.

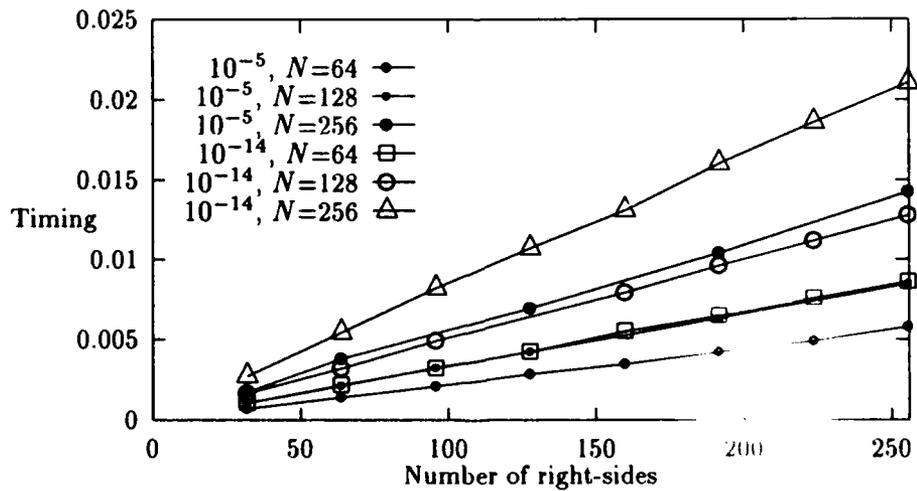


Figure 13. Timing of SPP with different accuracies.

## 6 Conclusion

A central goal of parallel processing is to achieve better, more accurate solutions. Because obtaining more accurate solutions, in general, means adding more discretization points, larger systems result and require greater computational power. The accuracy of a simulation solution is also bounded by the discretization scheme used. A clear requirement for obtaining a more accurate solution is to adopt discretization methods with high-order accuracy. Previously, a highly accurate discretization scheme, the *compact finite-difference scheme*, has been proposed. However, the almost symmetric Toeplitz tridiagonal systems that arise from compact schemes are sequential in nature and difficult to solve efficiently on parallel computers. In this paper, we have introduced a parallel algorithm, the *simple parallel prefix* (SPP) algorithm, for compact schemes.

The SPP algorithm is designed for fine-grain computing. With  $n$  processors, the SPP algorithm solves an  $n$ -dimensional system with  $2 \log(n) + 1$  AXPY operations. Two prefix communications are required in the solving phase and one broadcast communication is required in the modification phase. In comparison with existing tridiagonal solvers [17, 8], the SPP algorithm is simple in computing and simple in communication. It requires storage of only one  $\log(n)$ -dimensional vector for the computing phase and one  $n$ -dimensional vector for the modification phase. When the tridiagonal system is diagonally dominant, both the computing and the modification phases can be truncated without degrading the accuracy. Memory requirements will be further reduced when truncation is applied. A detailed accuracy analysis has been conducted to find the appropriate truncation number. Experimental results show that the SPP algorithm achieves a speedup greater than 1000 over the best sequential algorithm on a 16K PEs MasPar M-1 SIMD parallel computer. In addition to the good performance on the SIMD machines, the SPP algorithm also out performs the best sequential algorithm on a vector machine (Cray 2), even on systems with multiple right sides. Experimental and theoretical results show that the SPP algorithm is a good choice for compact schemes and for the emerging high-performance parallel computers.

The SPP algorithm is designed for almost symmetric Toeplitz tridiagonal systems. It can be modified for different boundary conditions and for cases where the number of processors  $p$  is less than the dimension of the system. However, generalization of the algorithm for general tridiagonal systems or for band systems is unlikely.

The work presented in this paper is a continuation of efforts to design efficient parallel solvers for compact scheme. An efficient solver, the PDD algorithm, for coarse- or median-grain computing has been proposed [18]. The PDD algorithm and the SPP algorithm can be combined on parallel machines with vector processing units.

## Acknowledgements

We would like to thank J. Gustafson and M. Carter of the Scalable Computing Laboratory at Ames Laboratory for providing the access and general help on the Ames 16k PEs MarPar parallel computer.

## References

- [1] ADAMS, Y. Highly accurate compact implicit methods and boundary conditions. *Journal of Computational Physics* 24 (1977), 10–22.
- [2] CHUNG, K.-L., AND SHEN, L.-J. Vectorized algorithm for b-spline curve fitting on Cary X-MP EA/16se. In *Proc. of Supercomputing '92* (1992), pp. 166–169.
- [3] DUFF, I., ERISMAN, A., AND REID, J. *Direct Methods for Sparse Matrices*. Clarendon Press, Oxford, 1986.
- [4] EGECIOGLU, O., KOC, D., AND LAUB, A. A recursive doubling algorithm for solution of tridiagonal systems on hypercube multiprocessors. *J. of Comp. and Appl. Math.* 27 (1989).
- [5] EIDSON, T., AND ERLEBACHER, G. Implementation of a fully-balanced periodic tridiagonal solver on a parallel distributed memory architecture. ICASE, NASA Langley Research Center, 1993. In preparation.
- [6] HIRSH, R. Higher order accurate difference solutions of fluid mechanics problems by a compact differencing technique. *J. Comput. Phys.* 19 (1975).
- [7] HO, C., AND JOHANSSON, S. Optimizing tridiagonal solvers for alternating direction methods on boolean cube multiprocessors. *SIAM J. Sci. Stat. Comput.* 11, 3 (1990), 563–592.
- [8] HOCKNEY, R. A fast direct solution of Poisson's equation using Fourier analysis. *J. ACM* 12 (1965), 95–113.
- [9] JOSLIN, R., STREETT, C., AND CHANG, C.-L. Validation of three-dimensional incompressible spatial direct numerical simulation code. NASA Technical Report, TP-3025, NASA Langley Research Center, July 1992.
- [10] KREISS, H., AND OLIGER, J. Methods for the approximate solution of time dependent problems. GARP Report No 10, 1973.
- [11] KUMAR, V., AND GUPTA, A. Analysis of scalability of parallel algorithms and architectures : A survey. In *Proc. of 1991 Int'l Conf. on Supercomputing* (June 1991).
- [12] LAMBIOTTE, J., AND VOIGT, R. The solution of tridiagonal linear systems on the cdc star-100 computer. *ACM Trans. Math. Soft.* 1, 4 (Dec. 1975), 308–329.
- [13] LAWRIE, D., AND SAMEH, A. The computation and communication complexity of a parallel banded system solver. *ACM Trans. Math. Soft.* 10, 2 (June 1984), 185–195.
- [14] LELE, S. Compact finite difference schemes with spectral-like resolution. *J. Comp. Phys.* 103, 1 (1992), 16–42.

- [15] ORTEGA, J., AND VOIGT, R. Solution of partial differential equations on vector and parallel computers. *SIAM Review* (June 1985).
- [16] SHERMAN, J., AND MORRISON, W. Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix. *Ann. Math. Stat.* 20, 621 (1949).
- [17] STONE, H. An efficient parallel algorithm for the solution of a tridiagonal linear system of equations. *J. of ACM* 20, 1 (Jan. 1973), 27-38.
- [18] SUN, X.-H. Application and accuracy of the parallel diagonal dominant algorithm. ICASE Technical Report, 93-6, ICASE, NASA Langley Research Center, 1993. A short version will appear in Proc. of ICPP '93.
- [19] SUN, X.-H., AND GUSTAFSON, J. Toward a better parallel performance metric. *Parallel Computing* 17 (Dec 1991), 1093-1109.
- [20] SUN, X.-H., AND NI, L. Another view on parallel speedup. In *Proc. of Supercomputing '90* (NY, NY, 1990), pp. 324-333.
- [21] SUN, X.-H., AND NI, L. A structured representation for parallel algorithm design on multi-computers. In *Proc. of the Sixth Conf. on Distributed Memory Computing* (April 1991).
- [22] SUN, X.-H., AND ROVER, D. Scalability of parallel algorithm-machine combinations. Technical Report, IS-5057, UC-32, Ames Laboratory, U.S. Department of Energy, 1991. Presented at Supercomputing '92 workshop.
- [23] SUN, X.-H., ZHANG, H., AND NI, L. Efficient tridiagonal solvers on multicomputers. *IEEE Transactions on Computers* 41, 3 (1992), 286-296.
- [24] WANG, H. A parallel method for tridiagonal equations. *ACM Trans. Math. Software* 7 (June 1981), 170-183.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 1993	3. REPORT TYPE AND DATES COVERED Contractor Report		
4. TITLE AND SUBTITLE A SIMPLE PARALLEL PREFIX ALGORITHM FOR COMPACT FINITE-DIFFERENCE SCHEMES			5. FUNDING NUMBERS C NAS1-19480 WU 505-90-52-01	
6. AUTHOR(S) Xian-He Sun Ronald D. Joslin			8. PERFORMING ORGANIZATION REPORT NUMBER ICASE Report No. 93-16	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Computer Applications in Science and Engineering Mail Stop 132C, NASA Langley Research Center Hampton, VA 23681-0001			10. SPONSORING / MONITORING AGENCY REPORT NUMBER NASA CR-191448 ICASE Report No. 93-16	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Langley Research Center Hampton, VA 23681-0001			11. SUPPLEMENTARY NOTES Langley Technical Monitor: Michael F. Card Final Report Submitted to Supercomputing '93	
12a. DISTRIBUTION / AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 61, 59			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) A compact scheme is a discretization scheme that is advantageous in obtaining highly accurate solutions. However, the resulting systems from compact schemes are tridiagonal systems that are difficult to solve efficiently on parallel computers. Considering the almost symmetric Toeplitz structure, a parallel algorithm, simple parallel prefix (SPP), is proposed. The SPP algorithm requires less memory than the conventional LU decomposition and is highly efficient on parallel machines. It consists of a prefix communication pattern and AXPY operations. Both the computation and the communication can be truncated without degrading the accuracy when the system is diagonally dominant. A formal accuracy study has been conducted to provide a simple truncation formula. Experimental results have been measured on a MasPar MP-1 SIMD machine and on a Cray 2 vector machine. Experimental results show that the simple parallel prefix algorithm is a good algorithm for the compact scheme on high-performance computers.				
14. SUBJECT TERMS high performance computing; parallel numerical algorithms; compact finite-difference scheme; SIMD computations; almost Toeplitz tridiagonal schemes			15. NUMBER OF PAGES 30	
17. SECURITY CLASSIFICATION OF REPORT Unclassified			16. PRICE CODE A03	
18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified		19. SECURITY CLASSIFICATION OF ABSTRACT		20. LIMITATION OF ABSTRACT