Th P b m F d g Ou

F

2 M mum V um E p d M h d

V

D t tion of outli i an old p obl m Giv n a t of data (ob vation) find out whi h ob vation a outli i on ou o dubiou data Fo xampl a data t that m a u ap ha a t i ti might hav m a u m nt fo a h ap u h a kull i body w ight and olo of hid If w th w two dino au and fiv f hm n into th xp im nt w would xp t to find th data point to b outli

B

Y Y k^h n

p $m < n$

y

n

m

$m +$

m n

3 Th P g am

V

Sin with th a li mall ampl th di tan will hav an i gula patt n th algo ithm fi t al ulat xp t d di tan f om imulat d data th imulat d di tan a u d to no mali th di tan obtain d lat on f om th al data Thi no mali ation liminat wid va iation in Mahalanobi di tan al ulat d fo mall ub t of th data

3 1 General Characteristics

F

B

3 2 Major Data Structures

y

nexp

p nvar

calibvol

ordis label

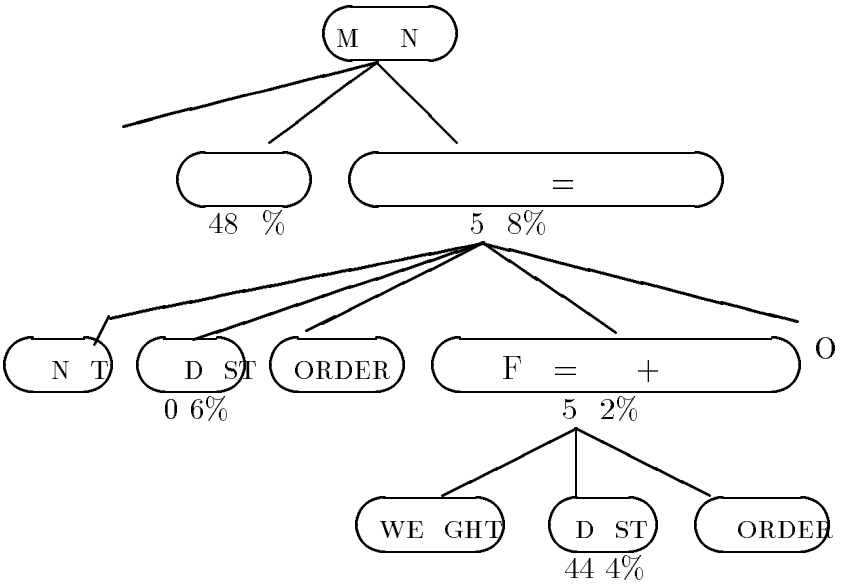
nexp

ifor nvar+ nexp;

: wtin(i) i

$\leq i \leq$ nexp

3 3 Call Graph



F

C

nvar nexp

% nexp

F ;

;

;

F

B

4 P g P m a

4 1 Execution Timing

[3] F

F F

-

;

4 2 Getting the Right Results

W

D ST

V

[3] Thi data t tak n f om [4] on i t of ight va iabl ha a t i ing th ompo ition of 87 ontain of milk

5 Pa a za S a gy

DO 11

%

DO 11 W

%

5 1 The Main Loop

[4] DO 11

:

V
iwt
ordis
stanvol *
stanvol
isymb
label
idum

*

:

volmin, fmsub, wtin[], det, nlow, ifor, i, vol, j, dism[] iy

idum

[4] Pa all li ation A i tanT Tool f om G o gia T h i d ib d in [1]

W

idum

B

wtin

wtin(1:nexp) wtin(1:nexp,1:NNSIM)

nstalac : i

wtin(1:nexp,i) j wtin(1:nexp,j)

nexp

5 2 A Parallel Random Number Generator

:

$$x_{n+\ i} \ = \ ax_{n\,i} + b_i \ mod \ m$$

F $_i\ b$

fl F

```
r = prng_next(me)
```

$\leq$ me $\leq$

5 3 CALIB

F

W

6 R u

6 1 The Effect of I/O

W

;

W

:

#	

A

W

:

F

	: : : :

W

:

; O HE K OW T
;

W

:

fl

Collating Strategy and Total Runtime

	:	:
	:	:
	:	:

%

nexp2fl

F

6 2 Private vs Replicated Arrays

F

W

W ?

–

i *i*

ffi F

	RUN1	RUN2	RUN3
wtin, dsim ordis, stanvol label, iwt			
	:	:	:
	:	:	:
26			
26	:	:	:
26			

A :

26 26

6 3 Speedup

F

:

D ST

	: :
	: :
	: :

:

D ST

W

DO

11

_

:

- ?
- ?
- ? W
[5]
- ?

?

B

P _ONE_SP_ ONG

6 4 Granularity

P _ONE_SP_ ONG

[5]Th figu 1/7 om f om ali ing that a on it ation b gin x ution th y t m i tting up th th ad fo th n xt on Th i not nough al ulation to ov om pa all l ov h ad wh n v n it ation fini h b fo th 8th an b t up

!

:

F

S E

:

GR

S E

ffi GR

PS_ONE_SP_ ONG

7 E a g g h App ca

$O\ rn^2$ r n

:

d infini um

F

V

8 mm T

F P T

; F

P T

P T D ST

P T

B

F

F det

; F w s

F

F

F

P T F

F

;

;

B

;

W

W

W ;

9 mm h KSR

F

B

F

;

B

ffi
W

0 Ack w dg m

F
B Y
F

R f c

B :
IEEE Sof w r – J
Workshop on
D An ysis nd Robus n ss J –J
B

S is ics nd Compu ing
J
S is ics : –

Journ of P r nd Dis ribu d Compu ing : –

B Z
Journ Am ric n S is ic Associ ion

Appendix A: Abbreviated Symbol Table

Variable		Scope	Size	Comments
aa		Task Local		For original Ran. # Gen.
calib:calibvol	output->	Parameter	100	See calibvol
calib:idum	<-in/out->	Parameter		See idum
calib:med	<-input	Parameter		
calib:nexp	<-input	Parameter		
calib:nvar	<-input	Parameter		
calib:sumdis		Local	100	Accumulates shortest sits
calib:volmin	output->	Parameter		
calib:y		Local	1000	Random data points
calibvol		Local	100	Calibrated volume/size
det		Task Local		Unused
dism		Local	100	Distances for
dist:det	output->	Parameter		current subset
dist:dism	output->	Parameter	100	See dism
dist:fmsub	<-input	Parameter		Size of subset
dist:nexp	<-input	Parameter		See nexp
dist:nvar	<-input	Parameter		See nvar
dist:wtin	<-input	Parameter	10	Subset membership
dist:y	<-input	Parameter	1000	See y
fmsub		Task Local		Size of current starting subset
gasdev:idum		Parameter		See idum
gaussj:a	<-in/out->	Parameter	100	
gaussj:ifault	output->	Parameter		
gaussj:n	<-input	Parameter		
gaussj:pivinv		Local		
idat		Task Local		Which data set to process
idum		Task Local		"seed" for Rand. # Gen.
ifor		Task Local		Current sete size (nlow..nexp)
init:idum	<-in/out->	Parameter		Random No. seed
init:nexp	<-input	Parameter		See nexp
init:nvar	<-input	Parameter		See nvar
init:wtin	output->	Parameter	100	Subset membership
isymb		Task Local	100	0 if point is close, 1 if far
iwt		Task Local	100	

```
label                    Task Local  100     Original pos. of points
ludcmp:a    <-input      Parameter    100    See y
ludcmp:d     output->    Parameter
ludcmp:indx <-input      Parameter   10      See nvar
ludcmp:n    <-input      Parameter           See nexp
med                      Task Local          # points in 1/2 volume ellipsoid
milkin:nexp output->     Parameter           Large data set
milkin:nvar output->     Parameter              "
milkin:y    output->     Parameter   1000       "
nexp                     Task Local          Number of observations
nlow                     Task Local          One more than prev sample size
nnsim                    Local               Parameter: # simulations to run
nsim                     Task Local          Used by Ran # Gen to initialize
nstalac                  Task Local          Which run, 1:100
nvar                     Task Local          Variables/observation
order:label  <-in/out->  Parameter   100     Index of orig. pos.
order:nobs   <-input     Parameter           # of data (nexp)
order:ordt    output->   Parameter    100     Sorted distances
order:t      <-input     Parameter    100     Unsorted data
ordis                    Local        100     Ordered distances
ran2:idum                Parameter           See idum
stout:dism   <-input     Parameter    100     See dism
stout:idat   <-input     Parameter           Whether small or large data used
stout:ifor   <-input     Parameter           See ifor
stout:isymb  output->    Parameter    100     See isymb
stout:nexp   <-input     Parameter           See nexp
volmin                   Task Local          Min vol calculated so far
weight:ifor  <-input     Parameter           Subset size (m)
weight:label <-input     Parameter    100     Li=1 if i-th distance
weight:nexp              Parameter              is one of minimum ones
weight:wtin   output->   Parameter    100     New sample, memb. func.
woodin:nexp   output->   Parameter    --      The smaller dataset
woodin:nvar   output->   Parameter              "
woodin:y      output->   Parameter   1000       "
wtin                     Local        100      Subset membership function
y                        Local                  1000     The input data
```

Appendix B: Static Call Graph and Loop Listing

```
***MAIN*** 114 lines MAIN calls:
1 call(s) to read     (ask user which data set to use)
1 call(s) to woodin   (routine to input small data set)
1 call(s) to milkin   (routine to input large data set)
1 call(s) to ran2     ("prime" the random number generator)
1 call(s) to calib    (calculate estimated mean distance for sample sizes)
  2 call(s) to write
  2 call(s) to order
  2 call(s) to dist
  1 call(s) to weight
  1 call(s) to init
  1 call(s) to gasdev
    2 call(s) to ran2
1 call(s) to init     (find the initial subset of nvar+1 datapoints)
  1 call(s) to ran2   (do it randomly)
3 call(s) to order    (sort the M. distances)
3 call(s) to dist     (compute M. distances for all nexp points)
  1 call(s) to ludcmp     (LU decomposition)
  1 call(s) to write
  1 call(s) to gaussj     (Gaussian elimination)
1 call(s) to weight   (get the m+1 minimum distances)
1 call(s) to stout    (output a stalactite)
  1 call(s) to write
6 call(s) to write    (various outputs from main routine)

Loop listing for the Main Routine (* = trivially parallelizable)
--------------------------------

(lines 41-42) *       do 20 isim=1,nsim
     1 call(s) to ran2
(lines 48-99)         do 11 nstalac=1,maxstal
     4 call(s) to write
     1 call(s) to init
     2 call(s) to dist
     2 call(s) to order
     1 call(s) to weight
     1 call(s) to stout
```

```
    (lines 64-86)         do 10 ifor=nlow,nexp
         1 call(s) to weight
         1 call(s) to dist
         1 call(s) to stout
         1 call(s) to order
         1 call(s) to write
        (lines 70-71) *       do 18 i=1,nexp
        (lines 82-83) *       do 12  j=1,nexp
(lines 104-106)         do 13 i=1,nexp
(lines 110-112) *       do 14 i=nlow,nexp
 7 simple loop(s)

Loop listing for CALIB (78 lines)
---------------------------------

(lines 145-146) *       do 15,i=1,nexp
(lines 147-194)         do 11 nstalac=1,ncalib
     1 call(s) to write
     1 call(s) to init
     2 call(s) to dist
     2 call(s) to order
     1 call(s) to weight
     1 call(s) to gasdev
    (lines 151-154)         do 1 i=1,nexp
         1 call(s) to gasdev
        (lines 152-154) *       do 1 j=1,nvar
             1 call(s) to gasdev
    (lines 165-186)         do 10 ifor=nlow,nexp
         1 call(s) to weight
         1 call(s) to dist
         1 call(s) to order
        (lines 174-176)         do 17 ix=1,nexp
        (lines 182-183) *       do 12  j=1,nexp
    (lines 189-193) *       do 3 i=nlow,nexp
(lines 198-201) *       do 16 i=nlow,nexp
(lines 202-205) *       do 18 i=nlow,nexp
     1 call(s) to write
 10 simple loop(s)
```

```
Loop listing for DIST (55 lines)
--------------------------------

(lines 298-301)         do 2 i=1,nvar
    (lines 300-301) *       do 2 j=1,nvar
(lines 302-307)         do 3 ix=1,nexp
    (lines 303-306)         do 4 i=1,nvar
        (lines 305-306) *       do 4 j=1,nvar
(lines 308-311)         do 5 i=1,nvar
    (lines 310-311) *       do 5 j=1,nvar
(lines 315-318)         do 8 i=1,nvar
    (lines 316-318) *       do 8 j=1,nvar
(lines 326-334)         do 6 ix=1,nexp
    (lines 328-330)         do 7 i=1,nvar
        (lines 329-330) *       do 7 j=1,nvar
(lines 340-342)         do 1,i=1,nvar
 13 simple loop(s)
```

Appendix C: FORGE Timing Data

```
Viewing APR/aca2/aca.t

 >PSRTIM 20  Timing Profile for Routine aca

                                 INCLUSIVE    EXCLUSIVE DO-LOOP-LENGTH
 NEST  LOOP OR SUBPROG     %JOB:%ROUTNE %JOB:%ROUTNE     COUNT

    ACA                   100.0:100.0   0.0:  0.0           1
  1  MILKIN                  0.0:  0.0   0.0:  0.0           1
  1  DO 20 ISIM              0.0:  0.0   0.0:  0.0           1
  2   RAN2                   0.0:  0.0   0.0:  0.0          97
  1  CALIB                  48.1: 48.1   0.3:  0.3           1
  1  DO 11 NSTALAC          51.8: 51.8   0.0:  0.0           1
  2   INIT                   0.0:  0.0   0.0:  0.0         100
  2   DIST                   0.6:  0.6   0.5:  0.5         100
  2   ORDER                  0.0:  0.0   0.0:  0.0         100
  2   DO 10 IFOR            51.2: 51.2   1.3:  1.3         100
  3    WEIGHT                0.1:  0.1   0.1:  0.1        7700
  3    DIST                 44.4: 44.4  39.5: 39.5        7700
  3    DO 18 I               0.2:  0.2   0.2:  0.2        7700
  3    STOUT                 2.4:  2.4   2.4:  2.4        7700
  3    ORDER                 2.9:  2.9   2.9:  2.9        7700
  3    DO 12 J               0.0:  0.0   0.0:  0.0        2713
  1  DO 13 I                 0.0:  0.0   0.0:  0.0           1
  1  DIST                    0.0:  0.0   0.0:  0.0           1
  1  ORDER                   0.0:  0.0   0.0:  0.0           1
  1  DO 14 I                 0.0:  0.0   0.0:  0.0           1
```

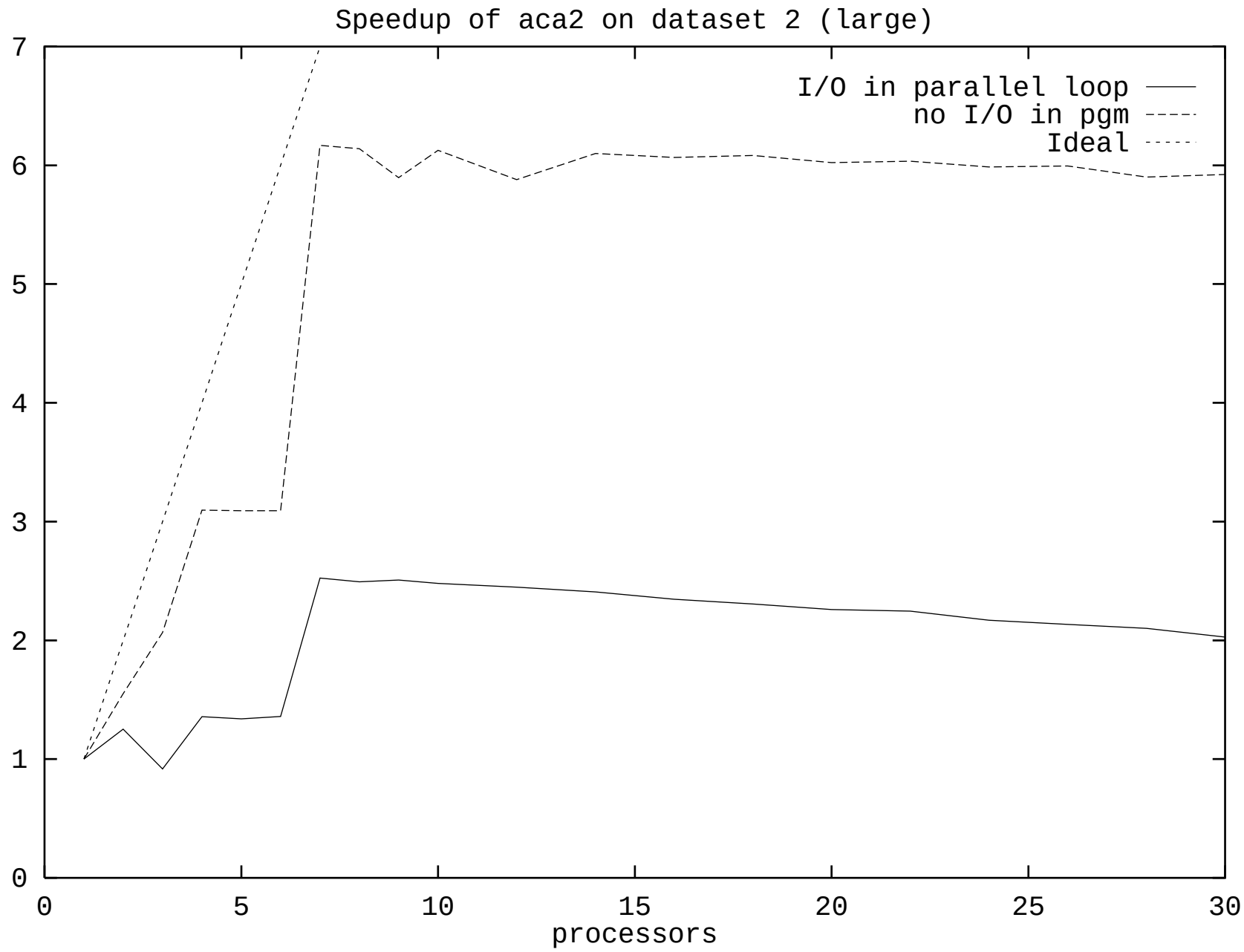
Appendix D: Binary Optimization

F

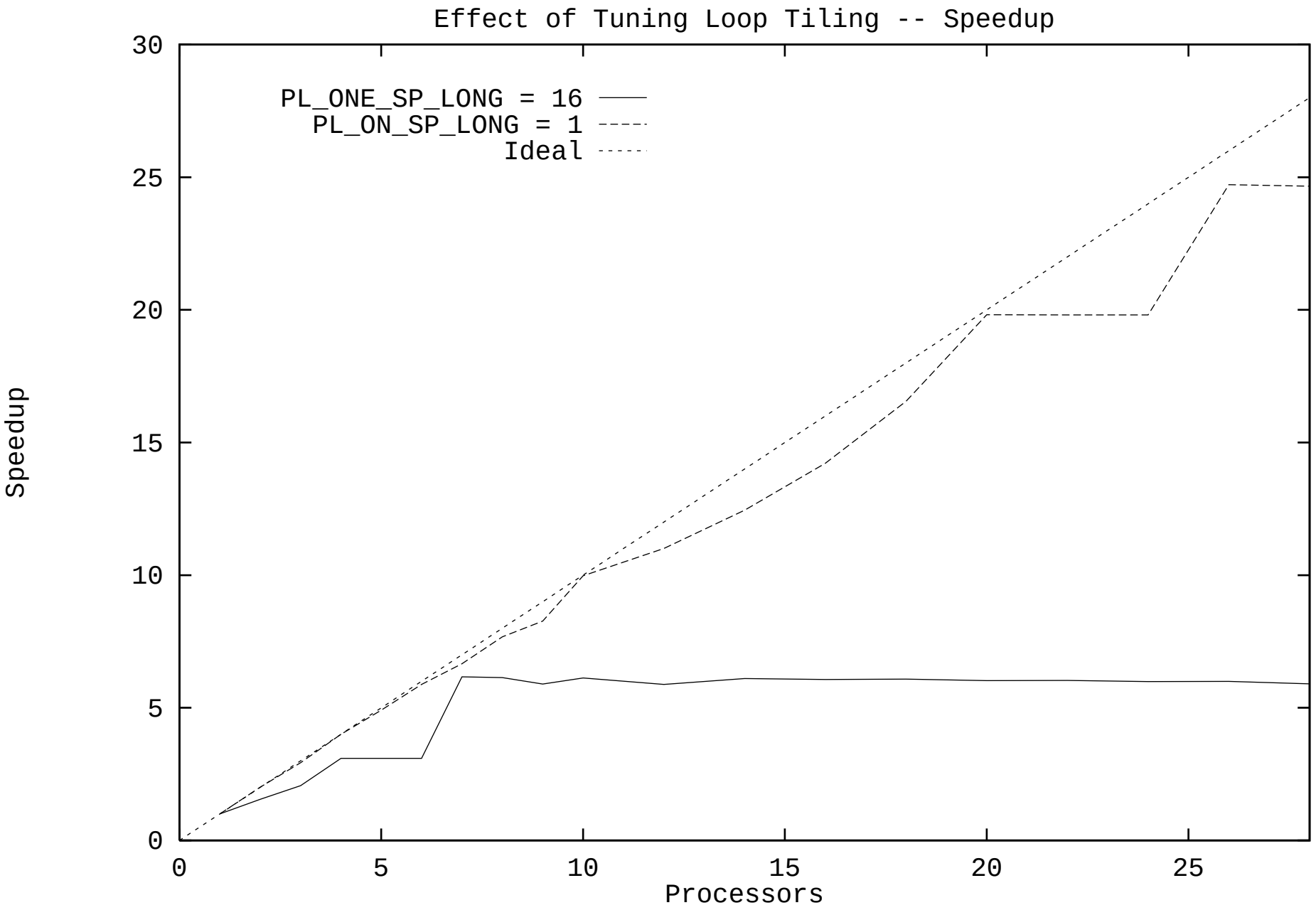
ROUT NE	STEP1	STEP2	STEP3	STEP4	STEP5	STEP6
RESU TS:		B	B			

:

B D ST



Appendix E: Speedup and I/O within Parallel Loop


Effect of Tuning Loop Tiling -- Speedup
PL_ONE_SP_LONG = 16
PL_ON_SP_LONG = 1
Ideal
Speedup
Processors
0
5
10
15
20
25
30