BLAS and SCALAPACK results on the MEIKO CS2-HA

Michel J. Dayde, ENSEEIHT-IRIT
dayde@enseeiht.fr

Architecture of the MEIKO CS2-HA at CERFACS

- 16 computational nodes
- node :
  - two 100MHz Cypress HyperSPARCs (peak 100 MFlops each)
  - Cache size = 256 KB per processor
  - 128 MB of memory per node
- peak perf. = 3.2 GFlops
- Network :
  - multistage switch network
  - bidirectional links (50MB each direction)
  - 10 microsec. latency for remote store
  - 3.2 Gbytes/sec peak network aggregate bandwidth
  - two partitions : 4 nodes and 10 nodes

Software

- UNIX - SUNOS 5.3
- Compilers : C, C++, FORTRAN (77, 90), KAP
- PVM 3.2(MEIKO) - CSN Comm. Library - Tagged Message Passing (MEIKO) PARMACS, NX/2, MPI (with MPE).
- TOTALVIEW Parallel debugger

Experiments with PBLAS and SCALAPACK V1.0

- BLACS on top of NX
- Using RISC BLAS
  - Built around matrix-matrix multiplication (GEMM)
  - Available on anonymous ftp at ftp.enseeiht.fr
    cd pub/numerique/BLAS/RISC

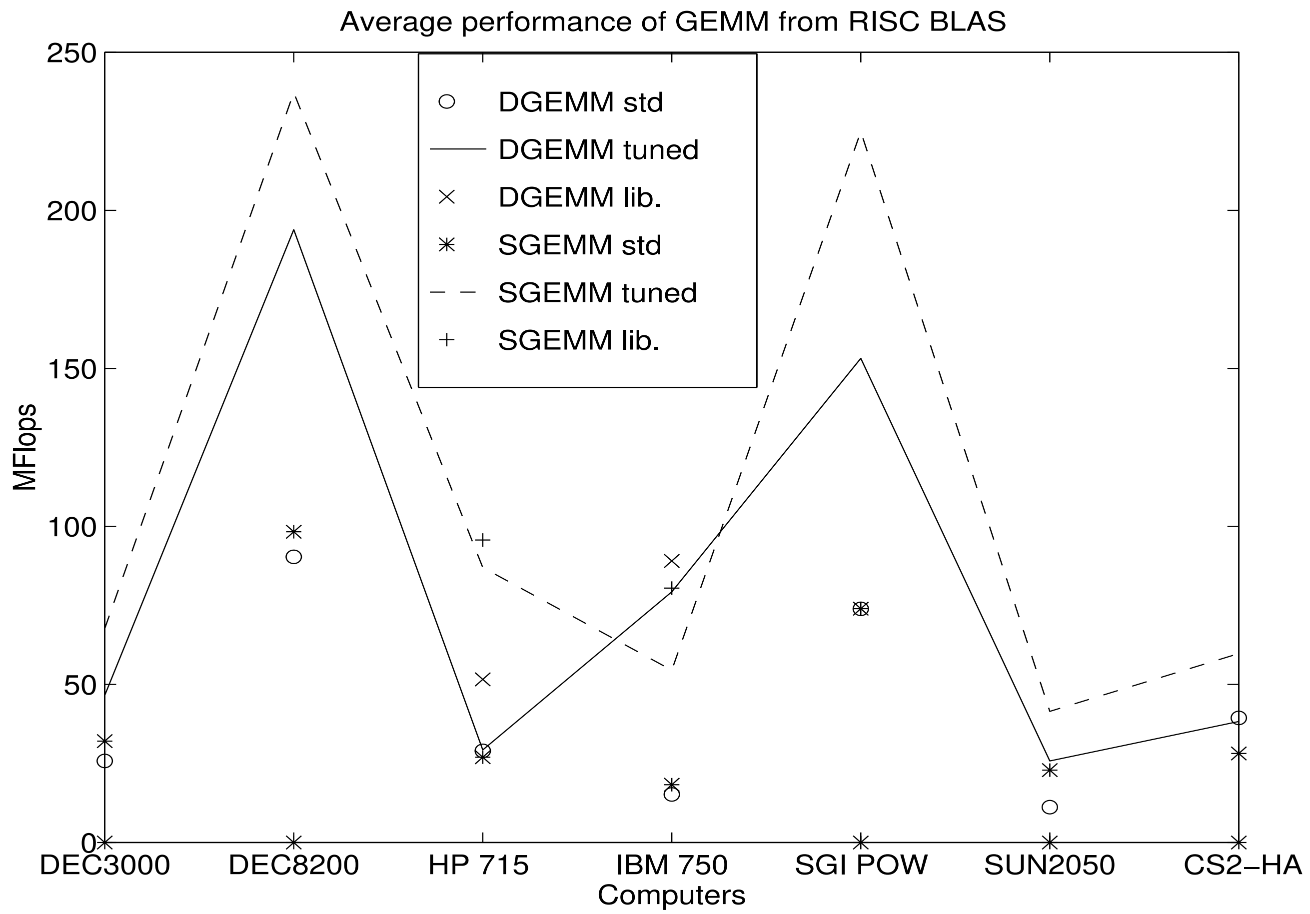
RISC BLAS

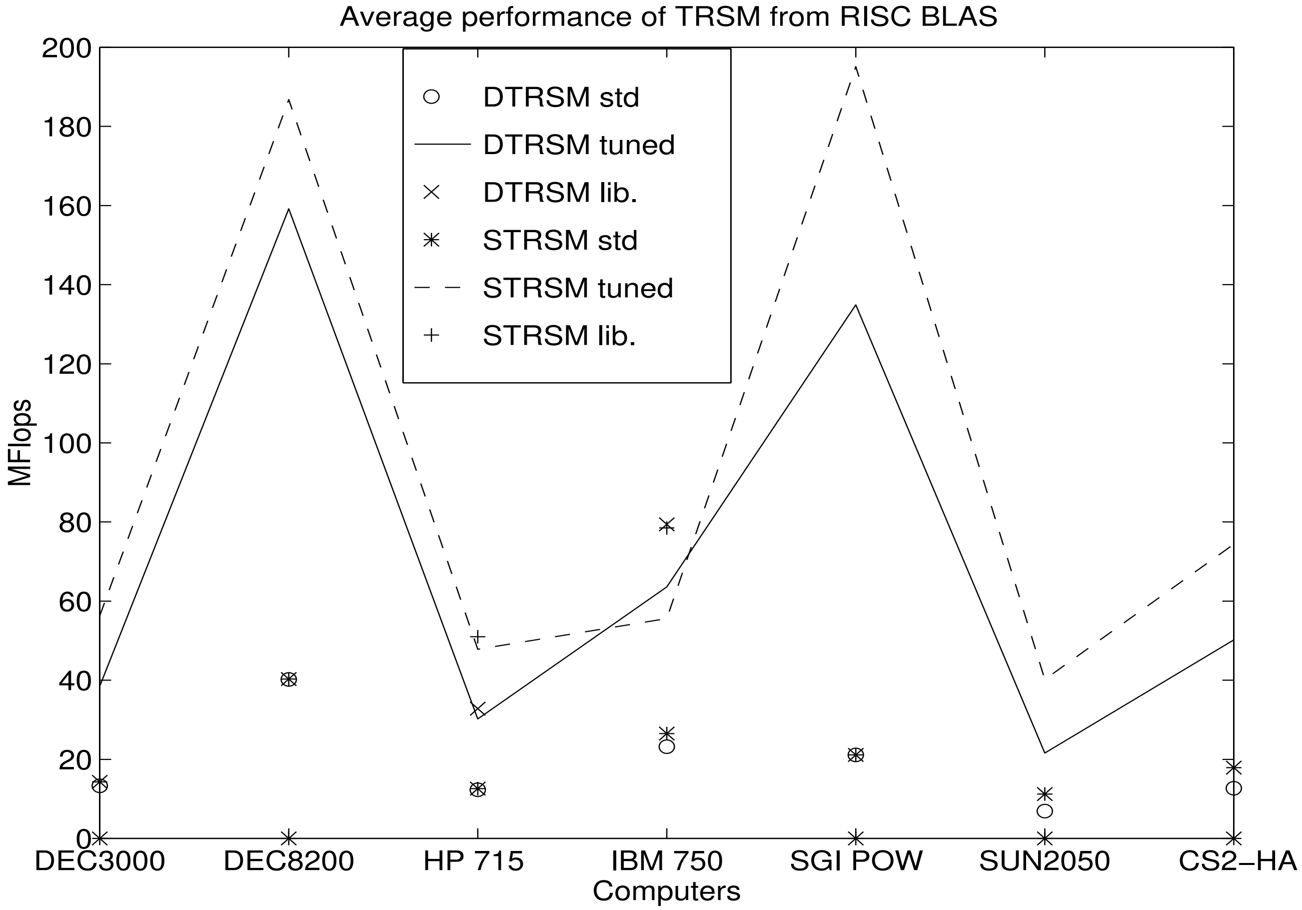
M.Dayde, ENSEEIHT-IRIT
I. Duff, RAL and CERFACS
P. Morere, ENSEEIHT-IRIT
A. Petitet, CERFACS
A. Qrichi, CERFACS

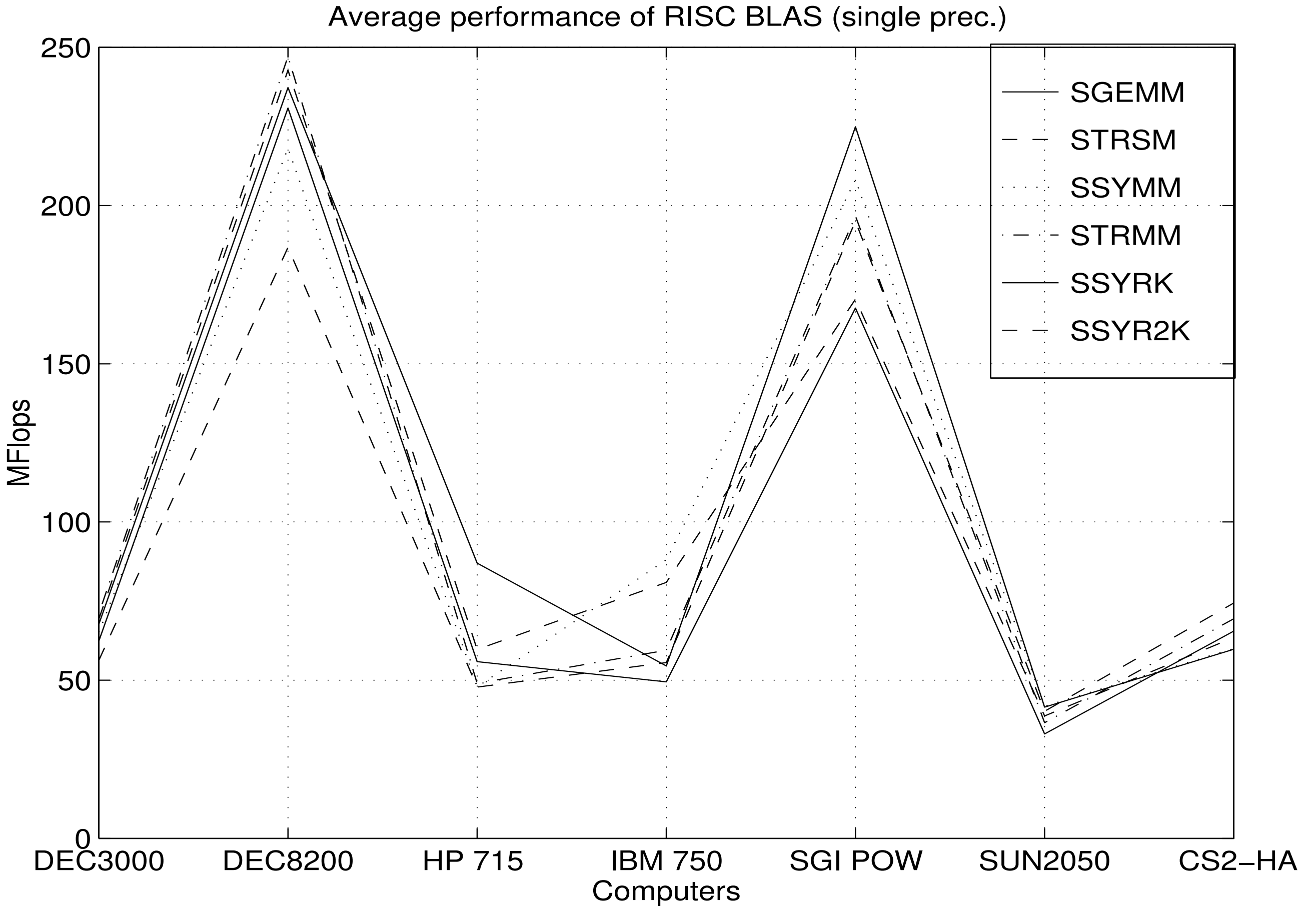
- Tuned BLAS for RISC processors using blocking, loop unrolling, and copying
- block size set according to cache size
- Triadic or not organization of operations
- Experiments on : DEC ALPHA, HP, IBM, SUN, MEIKO CS2-HA, SGI
- Currently includes real Level 3 BLAS and some Level 2 BLAS

Computer	Clock (MHz)	Peak (MFlops)	Cache size	Block size Sgle	Dble	Org. of ops
DEC 3000/300-AXP	150.0	150.0	8 KB	24	16	TRIADIC
DEC 8200 5/300	300.0	600.0	8 KB	24	16	TRIADIC
HP 715/64	64.0	128.0	64 KB	72	52	TRIADIC
IBM RS/6000-750	62.5	125.0	32 KB	42	36	TRIADIC
MEIKO CS2-HA	100.0	100.0	256 KB	140	100	NOTRIADIC
SGI Power Challenge	75.0	300.0	16 KB	36	24	TRIADIC
SUN SPARC 20/50	50.0	50.0	16 KB	36	24	NOTRIADIC

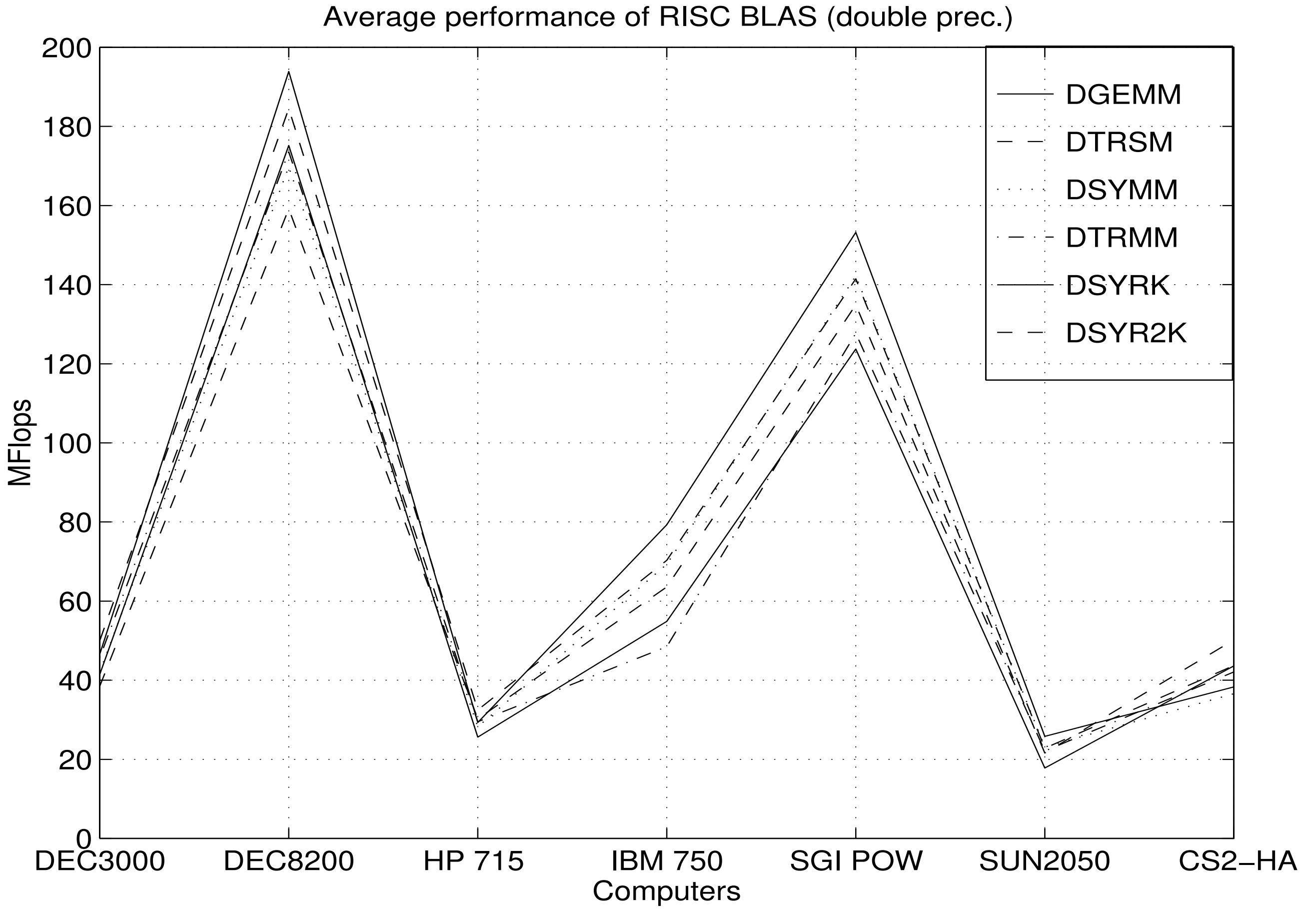
Table 1: Block size used in the blocked BLAS on the target computers.


Average performance of GEMM from RISC BLAS
MFlops
250
200
150
100
50
0
DGEMM std
DGEMM tuned
DGEMM lib.
SGEMM std
SGEMM tuned
SGEMM lib.
DEC3000
DEC8200
HP 715
IBM 750
SGI POW
SUN2050
CS2−HA
Computers

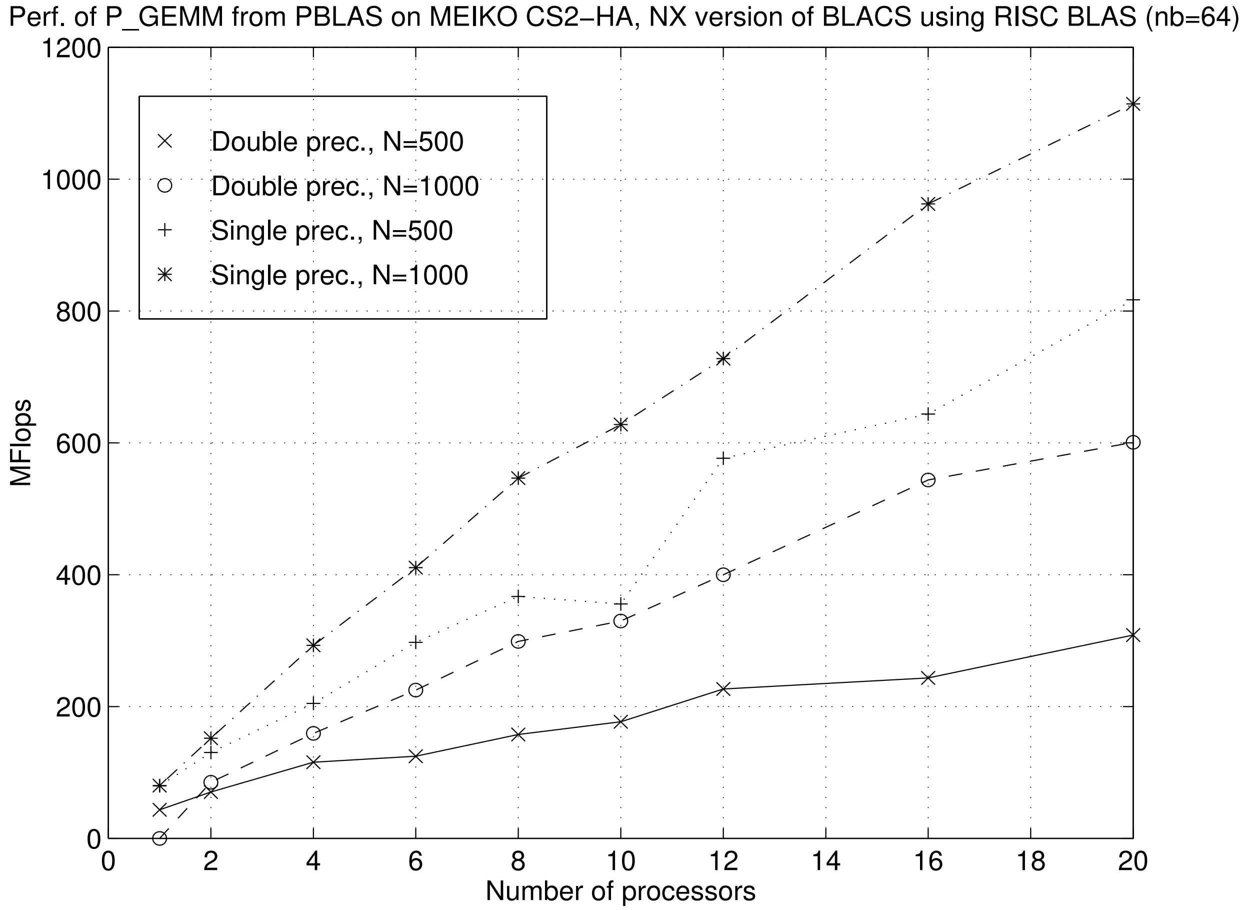

Average performance of TRSM from RISC BLAS
MFlops
Computers
DEC3000
DEC8200
HP 715
IBM 750
SGI POW
SUN2050
CS2−HA
0
20
40
60
80
100
120
140
160
180
200
DTRSM std
DTRSM tuned
DTRSM lib.
STRSM std
STRSM tuned
STRSM lib.

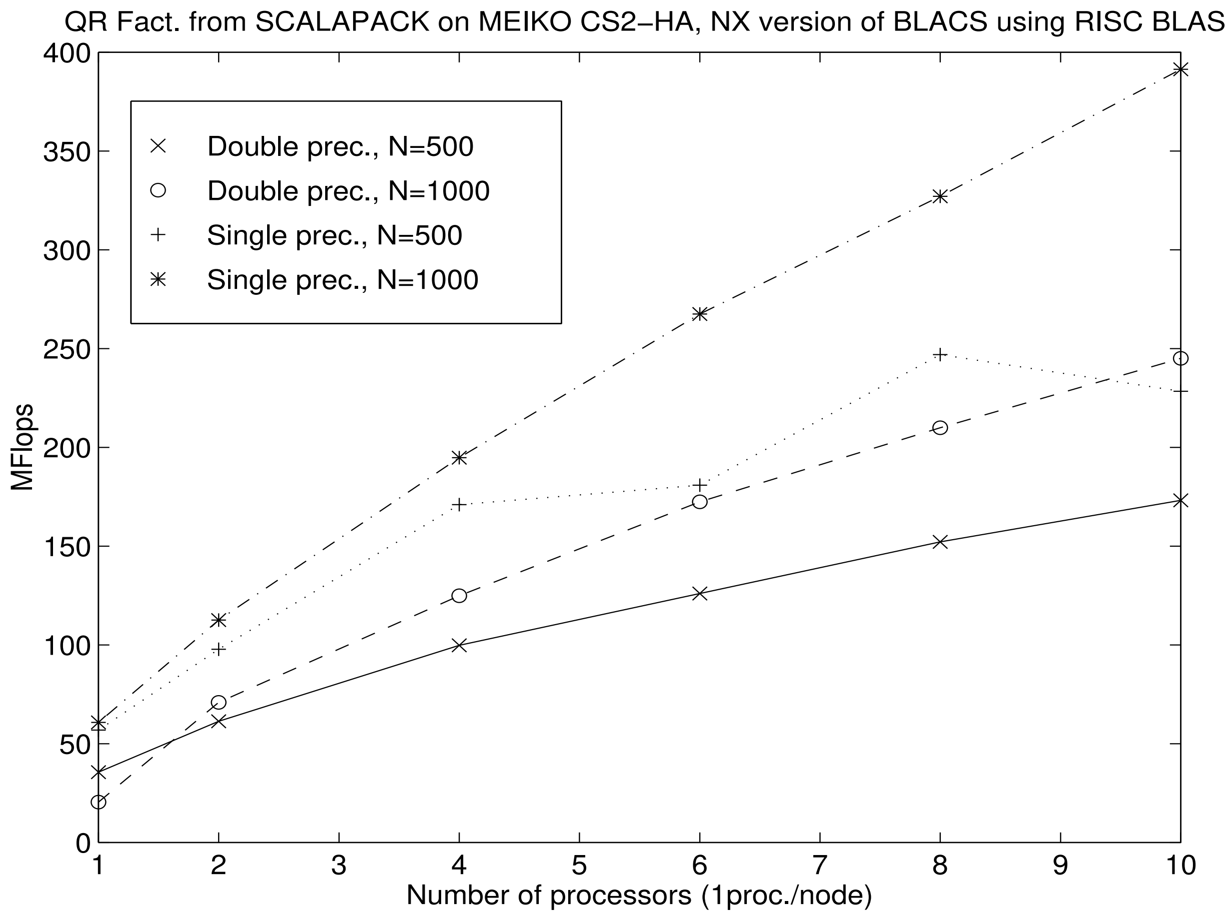


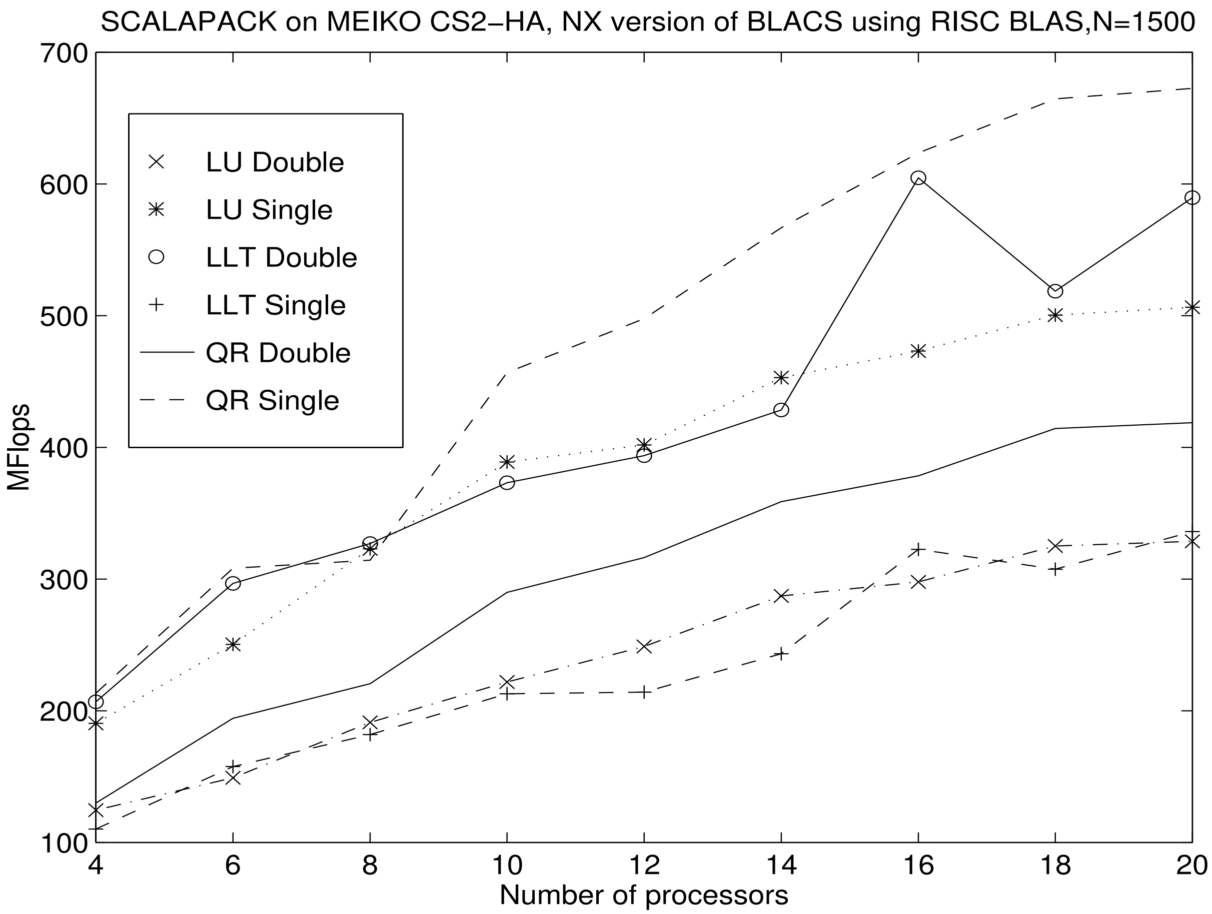
Average performance of RISC BLAS (single prec.)
SGEMM
STRSM
SSYMM
STRMM
SSYRK
SSYR2K
MFlops
250
200
150
100
50
0
DEC3000
DEC8200
HP 715
IBM 750
SGI POW
SUN2050
CS2−HA
Computers



Average performance of RISC BLAS (double prec.)
MFlops
Computers
DEC3000
DEC8200
HP 715
IBM 750
SGI POW
SUN2050
CS2−HA
DGEMM
DTRSM
DSYMM
DTRMM
DSYRK
DSYR2K
0
20
40
60
80
100
120
140
160
180
200


Perf. of P_GEMM from PBLAS on MEIKO CS2-HA, NX version of BLACS using RISC BLAS (nb=64)
Double prec., N=500
Double prec., N=1000
Single prec., N=500
Single prec., N=1000
MFlops
Number of processors
0
200
400
600
800
1000
1200
2
4
6
8
10
12
14
16
18
20


QR Fact. from SCALAPACK on MEIKO CS2−HA, NX version of BLACS using RISC BLAS
MFlops
Number of processors (1proc./node)
Double prec., N=500
Double prec., N=1000
Single prec., N=500
Single prec., N=1000
0
50
100
150
200
250
300
350
400
1
2
3
4
5
6
7
8
9
10


SCALAPACK on MEIKO CS2−HA, NX version of BLACS using RISC BLAS,N=1500
MFlops
Number of processors
LU Double
LU Single
LLT Double
LLT Single
QR Double
QR Single
100
200
300
400
500
600
700
4
6
8
10
12
14
16
18
20

Conclusions

- Good performance
- Probably more efficient to run parallel BLAS within a node (i.e. loop-level based) instead of 2 processes exchanging messages (using our parallel BLAS for shared memory multiproc.)
- Use of BLACS, PBLAS and SCALAPACK within sparse solvers (especially direct multifrontal) : alignment restrictions ?
- RISC BLAS
  - can be improved in many ways
  - make complex and double complex available