NIST Sparse BLAS: Preliminary Performance Studies

Karin A. Remington and Roldan Pozo*
National Institute of Standards and Technology

July 1996

1 Performance Results Summary

Performance measurements are given for simple matrix-vector and matrix-matrix multiplies for several sparsity patterns on a Sun Sparc 20 and an IBM RS6000. The measurements reflect compiler optimization (-O3 and loop unrolling) only. Typical performance for problems with large blocksize and multiple right-hand-sides is about 17 Mflops on a Sun Sparc 20 and about 27 Mflops on an IBM RS6000 Model 590. The "Lite" interface provides no measurable performance gain, except for some very small problems. The blocked schemes (BSR and VBR) begin to pay off when blocksize is greater than 5 or 10; for smaller blocksizes, the point-wise (CSR) scheme is more efficient.

2 Test Problem Generation

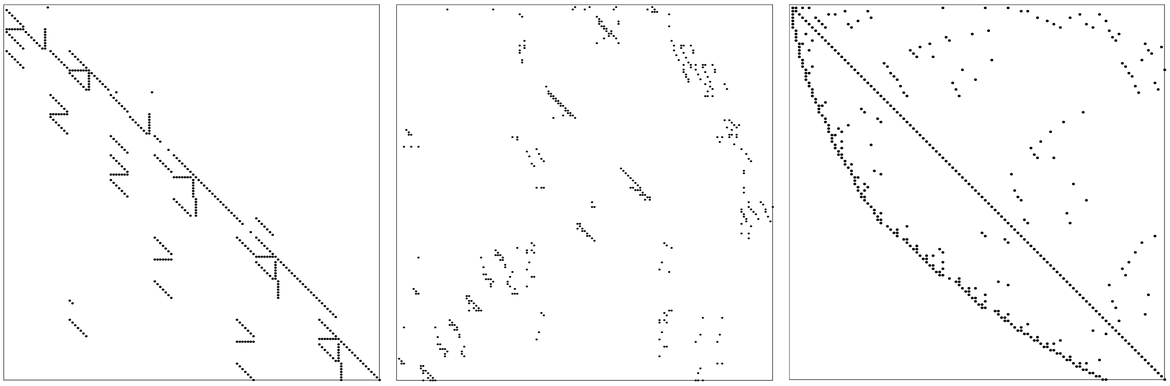
The test matrices used in the following tests were generated by reading sparsity patterns from Harwell-Boeing files, and using these patterns as the block structure for a matrix of given blocksize. The results shown are for matrix-vector and matrix-matrix multiplications only, rather than a full DAXPY, since we are interested in the efficiency of the *sparse* code. Source code for the performance testers is available, see Section 4.

The test results presented in Sections 3.1 and 3.2 are for the matrix patterns shown in Figure 1. The three Harwell-Boeing matrices used in this study are IMPCOL C, WEST0156, and GRE 115, and were chosen to give a variety of different memory access patterns. Size information for these matrices is summarized in Table 1.

*e-mail: [kremington,pozo]@nist.gov

NAME	size	nonzeros
IMPCOL C	137 by 137	411
WEST0156	156 by 156	371
GRE 115	115 by 115	421

Table 1: Test matrix size information



IMPCOL C WEST0156 GRE 115

Figure 1: Sparsity patterns of test matrices.

3 Performance results

Figures 2–7 show test measurements taken on a SUN Sparc 20, and Figures 8–13 represent the same measurements taken on an IBM RS6000 Model 590. The experimental parameters are blocksize and number of right-hand-sides. The machines were used under light load conditions, but not in dedicated mode. Each data point represents the average result of 4 runs with the same parameters.

Each graph plots three pairs of test results, one for each of the storage formats tested: Compressed Sparse Row (CSR), Block Compressed Row (BSR), and Variable Block Row (VBR). The pair of results represents the performance of the high-level "Toolkit" interface (described in [1]) and the lower-level "Lite" interface (described in the User's Guide for the NIST Sparse BLAS). Although the "Lite" interface was expected to improve performance by eliminating a layer of error checks and decision trees, the results indicate that the savings is minimal for problems of any substantial size. Further note that the point-wise oriented routines (for CSR) are more efficient than the block-oriented routines (for BSR and VBR) for problems with small blocksize. It is not until the blocksize reaches 5 to 10 than the block-oriented format begins to pay off. The performance of the block routines may improve if library calls to optimized dense level 3 BLAS are used. The present library uses fixed inlined versions of the dense BLAS to facilitate the automatic code generation scheme, and the possibility modularizing this part of the computation is being examined.

As expected, performance improves for block-wise Level 3 operations (compare, for example, Figure 2 and 3). However, did not see an improvement when the number of right-hand-sides was raised from 10 to 20.

3.1 Sparse BLAS performance on SUN Sparc 20

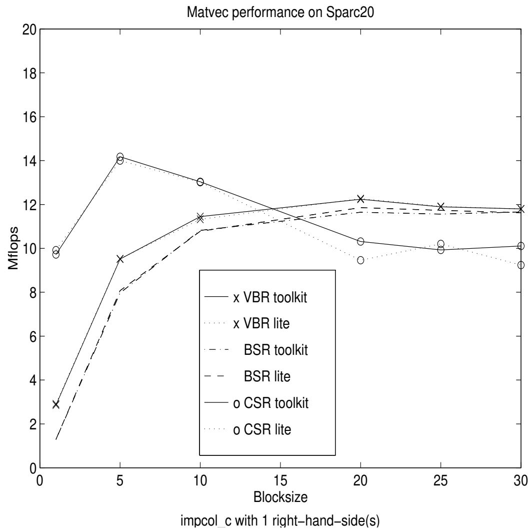


Figure 2: Sparc 20 results for IMPCOL C with 1 right-hand-side

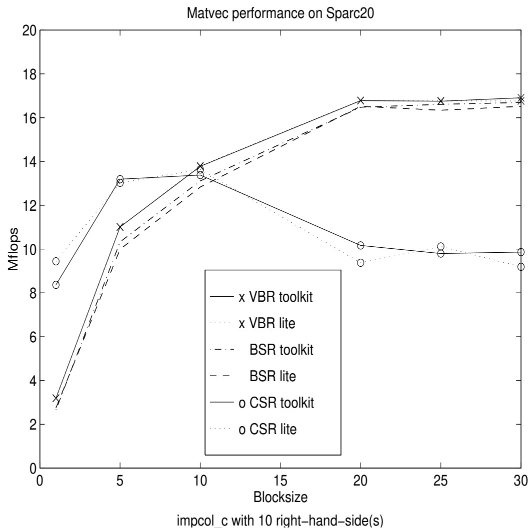


Figure 3: Sparc 20 results for IMPCOL C with 10 right-hand-sides

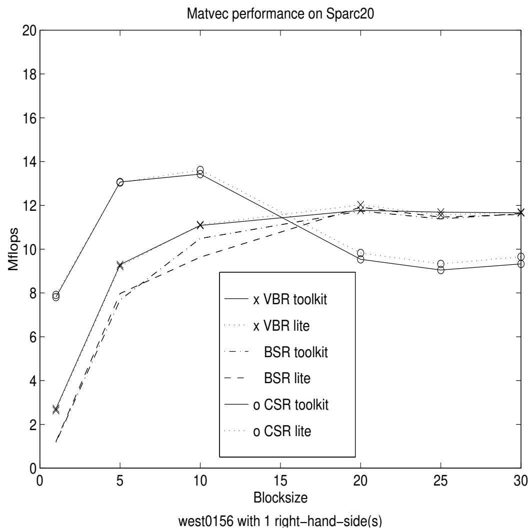


Figure 4: Sparc 20 results for WEST0156 with 1 right-hand-side

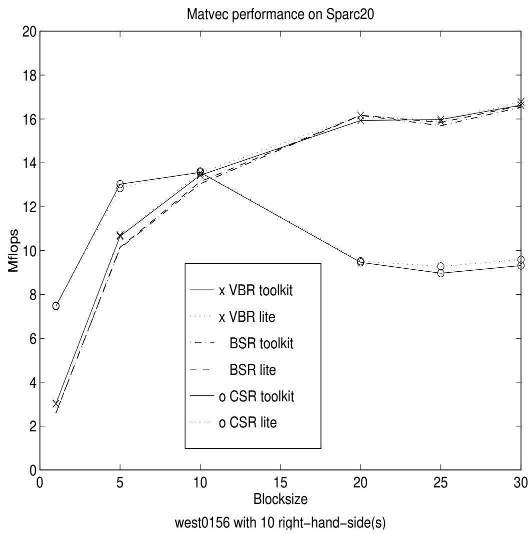


Figure 5: Sparc 20 results for WEST0156 with 10 right-hand-sides

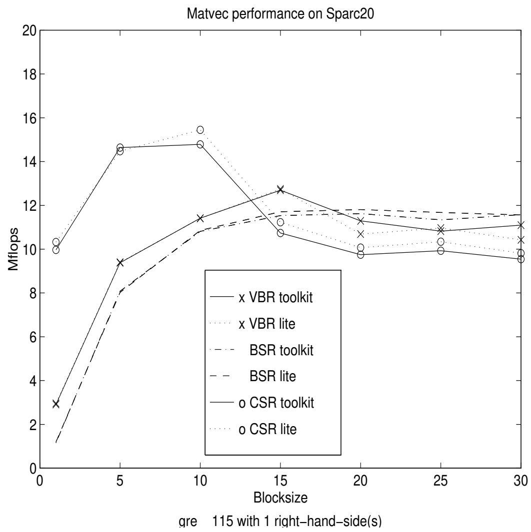


Figure 6: Sparc 20 results for GRE 115 with 1 right-hand-side

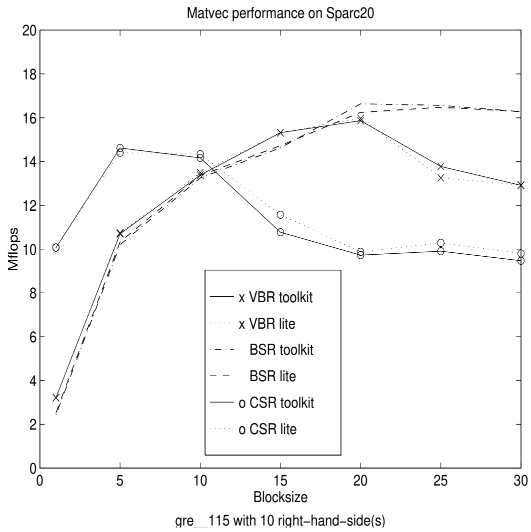


Figure 7: Sparc 20 results for GRE 115 with 10 right-hand-sides

3.2 Sparse BLAS performance on IBM RS6000 Model 590

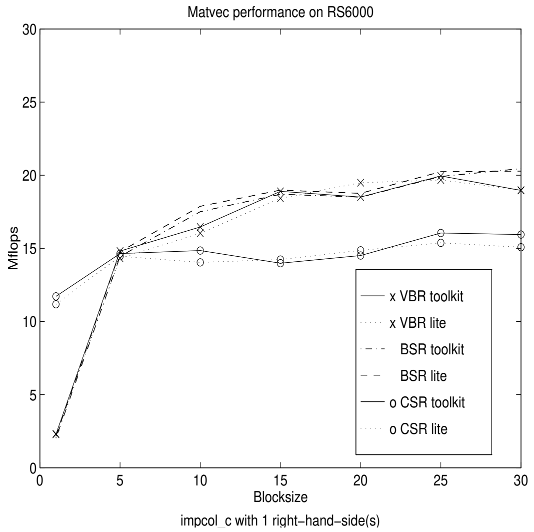


Figure 8: RS6000 results for IMPCOL C with 1 right-hand-side

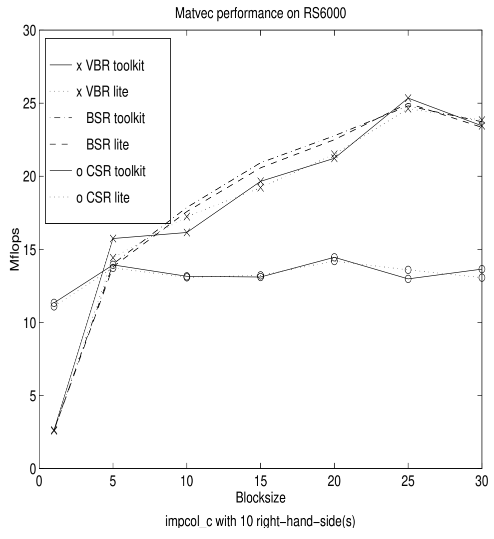


Figure 9: RS6000 results for IMPCOL C with 10 right-hand-sides

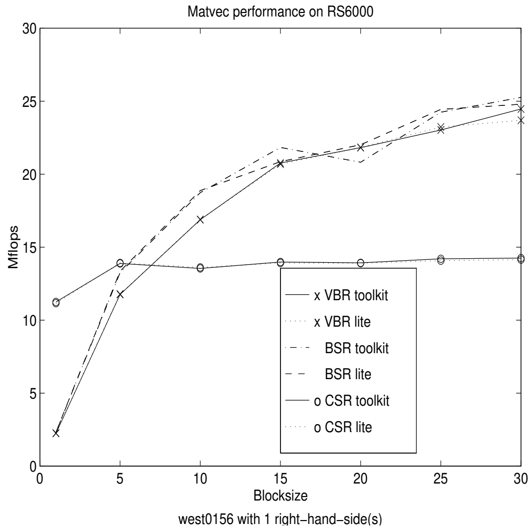


Figure 10: RS6000 results for WEST0156 with 1 right-hand-side

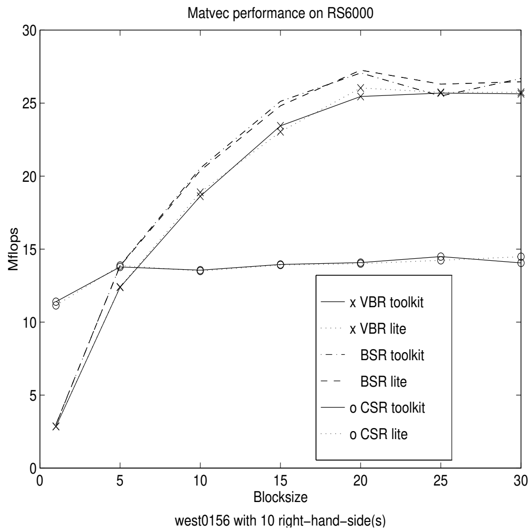


Figure 11: RS6000 results for WEST0156 with 10 right-hand-sides

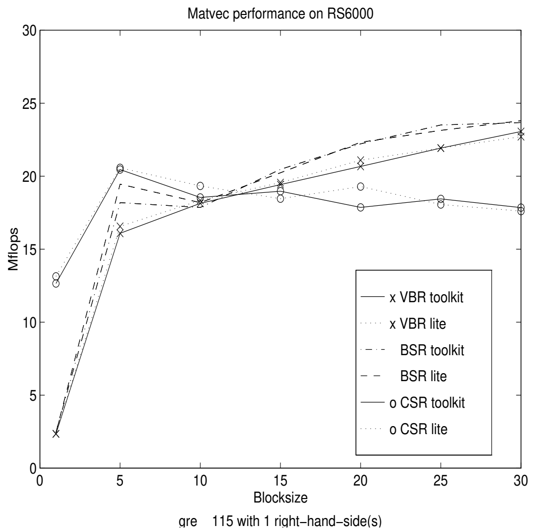


Figure 12: RS6000 results for GRE 115 with 1 right-hand-side

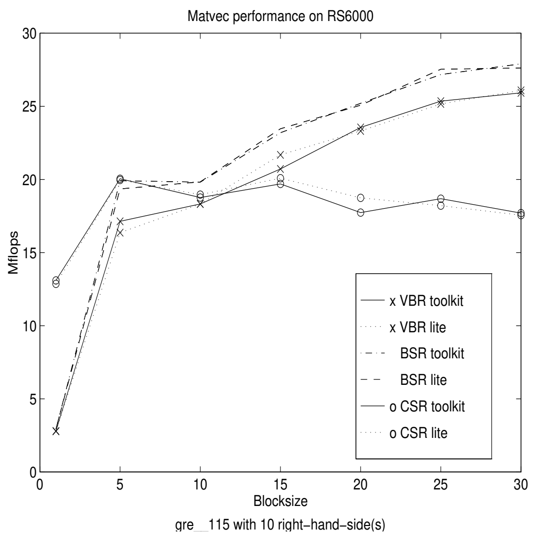


Figure 13: RS6000 results for GRE 115 with 10 right-hand-sides

4 Source code availability

Source code for the performance tests can be obtained via anonymous ftp to:

math.nist.gov/pub/karin/spblas/spblasperf.shar.gz

5 References

[1] S. Carney, M. Heroux, G. Li, R. Pozo, K. Remington, and K. Wu, *A revised proposal for a sparse blas toolkit*, http://www.cray.com/PUBLIC/APPS/SERVICES/algorithms.html, (1996).