

Sparse Linear Algebra CoSP2 Proxy Application for Electronic Structure Calculations

CoSP2 represents a sparse linear algebra parallel algorithm for calculating the density matrix in electronic structure theory. The algorithm is based on a recursive second-order Fermi-Operator expansion method (SP2) and is tailored for density functional based tight-binding calculations of non-metallic material systems. This SP2 algorithm is part of the Los Alamos Transferable Tight-binding for Energetics (LATTE) code, based on a matrix expansion of the Fermi operator in a recursive series of generalized matrix-matrix multiplications. The computational cost scales linearly for shared memory with the system size, which is achieved by using the ELLPACK-R (ELL) sparse matrix data format and OpenMP multi-threading. The CoSP2 implementation uses hierarchical parallelism, MPI between nodes and OpenMP on node.

We achieve efficient shared memory parallelism for generalized sparse matrix-matrix multiplications by using the ELL data structure. The ELL format allows sparse matrix storage in a regular manner (see Fig. 1). Each sparse matrix is described by three arrays: 1) a 2-dimensional array containing the numerical values, 2) a 2-dimensional array containing the column indices, and 3) a vector containing the number of non-zero entries per row. The row-wise data storage makes a parallel implementation of a sparse matrix-matrix multiplication fairly straightforward. Fig. 2 illustrates the calculation of a sparse matrix-matrix multiplication, x^2 , which is the operation governing the cost of the SP2 algorithm. The matrix elements of each row i of X are multiplied with the corresponding column vector elements. However, the matrix X is symmetric and we therefore multiply the corresponding row vector elements and accumulate into a temporary row buffer (RowBuf _{i}). The temporary row buffer vector represents the result of a new row i in x^2 , which is stored in the compressed ELL format. Using a distributed/shared memory architecture this scheme can be parallelized over the rows.

Pseudo code for the Second-order spectral projection (SP2) algorithm is shown below.

```
Estimate  $\epsilon_{\max}$  and  $\epsilon_{\min}$ 
 $X = (\epsilon_{\max}I - H)/(\epsilon_{\max} - \epsilon_{\min})$ 
TraceX = Tr[X]
BreakLoop = 0
i=0
while BreakLoop = 0 do
    i=i+1
    Exchange halo information between MPI ranks
    TraceXold = TraceX
     $X_{\text{tmp}} = X^2$ 
    TraceXtmp = Tr[Xtmp]
    if  $| \text{TraceXtmp} - N_{\text{occ}} | - | 2\text{TraceX} - \text{TraceXtmp} - N_{\text{occ}} | > \text{IdemTol}$  then
```

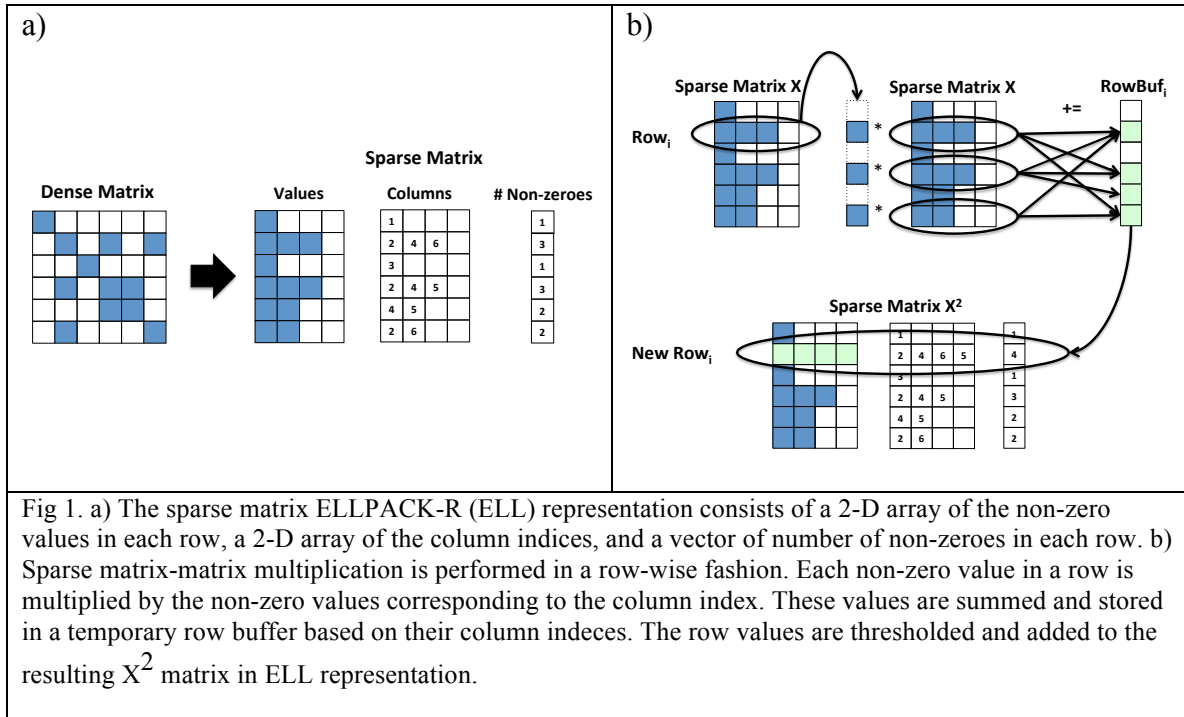
```

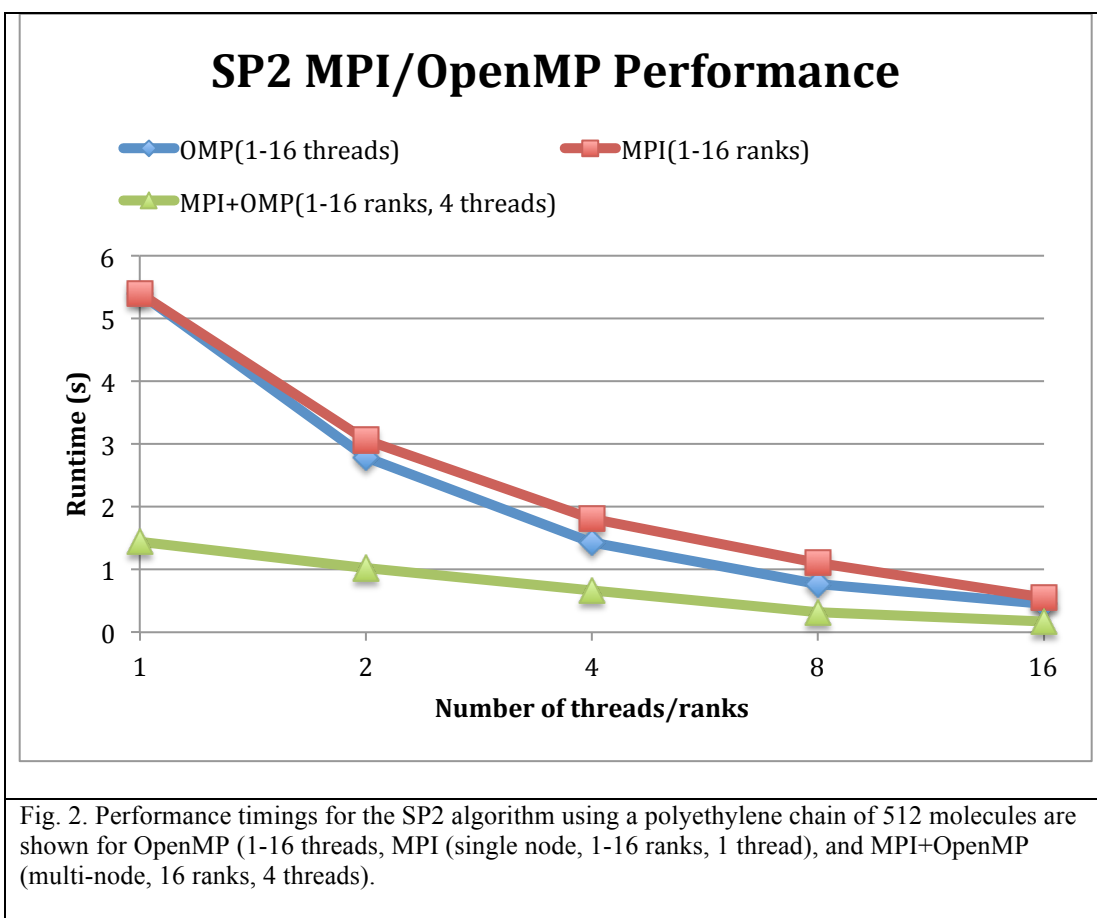
    X = 2X - Xtmp
    TraceX = 2TraceX - TraceXtmp
else
    X = Xtmp
    TraceX = TraceXtmp
end if
IdemErri = |TraceX - TraceXold|
if IdemErri-2 ≤ IdemErri and i > imin then
    BreakLoop = 1
end if
Exchange halo chunks of rows across MPI ranks
end while
P=X

```

The work decomposition across MPI ranks consists of splitting up the Hamiltonian matrix into chunks of rows, where each chunk is assigned to a different rank. Access to halo rows is needed during the computation, requiring an exchange of relevant halo chunks per iteration of the SP2 algorithm.

Performance timings are shown for a polyethylene chain of 1024 molecules (6144 atoms, 12,288 x 12,288 matrix). Fig. 2 shows timings for 1) OpenMP only with 1-16 threads on a single node; 2) MPI with 1 OpenMP thread, running on 1-16 ranks of one node; and 3) MPI with 4 OpenMP threads, running on 1-16 ranks (as separate nodes). We see similar performance for OpenMP and MPI running on a single node. Multi-node MPI+OpenMP shows improved performance due to hierarchical parallelism.





Background papers:

1. M. J. Cawkwell, E. J. Sanville, S. M. Mniszewski, A. M. N. Niklasson, 2012, Computing the density matrix in electronic structure theory on graphics processing units, *J. Chem. Theory Comput.*, 8, 4094–4101.
2. E. H. Rubensson, A. M. N. Niklasson, 2014, Interior eigenvalues from density matrix expansions in quantum mechanical molecular dynamics, *SIAM J. Sci. Comput.* Vol. 36, No. 2, pp. B147–B170.
3. P. Souvatzis, A. M. N. Niklasson, 2014, First principles molecular dynamics without self-consistent field optimization, *J. Chem. Physics* 140, 044117.
4. M. J. Cawkwell, A. M. N. Niklasson, 2012, Energy conserving, linear scaling Born-Oppenheimer molecular dynamics, *J. Chem. Physics* 137, 134105.
5. A. M. N. Niklasson, P., Steneteg, N. Bock, 2011, Extended Lagrangian free energy molecular dynamics, *J. Chem. Physics* 135, 164111.
6. A. M. N. Niklasson, 2008, Extended Born-Oppenheimer molecular dynamics, *PRL* 100, 123004.
7. A. M. N. Niklasson, 2002, Expansion algorithm for the density matrix, *Phys. Rev. B* 66, 155115.