August 2017

Porting of the DBCSR library for Sparse Matrix-Matrix Multiplications to Intel Xeon Phi systems

Iain BETHUNE [a,1], Andreas GLÖSS [b,2], Jürg HUTTER [b,3], Alfio LAZZARO [b,4], Hans PABST [c,5] and Fiona REID [d,6]

[a] *Hartree Centre, Science and Technology Facilities Council, United Kingdom*
[b] *University of Zurich, Department of Chemistry, Switzerland*
[c] *Intel Semiconductor AG, Switzerland*
[d] *EPCC, The University of Edinburgh, United Kingdom*

Abstract. Multiplication of two sparse matrices is a key operation in the simulation of the electronic structure of systems containing thousands of atoms and electrons. The highly optimized sparse linear algebra library DBCSR (Distributed Block Compressed Sparse Row) has been specifically designed to efficiently perform such sparse matrix-matrix multiplications. This library is the basic building block for linear scaling electronic structure theory and low scaling correlated methods in CP2K. It is parallelized using MPI and OpenMP, and can exploit GPU accelerators by means of CUDA. We describe a performance comparison of DBCSR on systems with Intel Xeon Phi Knights Landing (KNL) processors, with respect to systems with Intel Xeon CPUs (including systems with GPUs).

We find that the DBCSR on Cray XC40 KNL-based systems is 11%-14% slower than on a hybrid Cray XC50 with Nvidia P100 cards, at the same number of nodes. When compared to a Cray XC40 system equipped with dual-socket Intel Xeon CPUs, the KNL is up to 24% faster.

Keywords. sparse matrix-matrix multiplications, vectorization, multi-threading, MPI parallelization, accelerators, Intel Xeon Phi, Knights Landing

1. Introduction

Multiplication of two sparse matrices (SpGEMM) is a key operation in the simulation of the electronic structure of systems containing thousands of atoms and electrons [1]. Examples of such systems include electronic devices, complex interfaces, macromolecules and large disordered systems, with applications in the fields of renewable energy and electronics. The theory that enables such studies is linear scaling Density Functional

[1] E-mail: iain.bethune@stfc.ac.uk
[2] E-mail: andreas.gloess@chem.uzh.ch
[3] E-mail: hutter@chem.uzh.ch
[4] E-mail: alfio.lazzaro@chem.uzh.ch
[5] E-mail: hans.pabst@intel.com
[6] E-mail: f.reid@epcc.ed.ac.uk