Multi-dimensional MPI communications, where MPI communications have to be performed in each dimension of a Cartesian communicator, have been frequently used in many of today's high performance computing applications. While single-dimensional MPI communications have been extensively studied and optimized, little optimization has been performed for multi-dimensional MPI communications.

In this research, we consider all communications in all dimensions together and optimize the total communication time in all dimensions. We demonstrate that the default process-to-core mappings in today's state-of-the-art MPI implementations provided by MPI_Init() and MPI_Cart_create() are often sub-optimal for multi-dimensional communications. We propose an application-level tile-based multicore-aware process-to-core re-mapping scheme that is capable of achieving optimal performance for multi-dimensional communications. By re-mapping computational processes to hardware cores using a tiling technique, the communications are re-distributed to minimize the inter-node communications. The proposed technique improves the performance by up to 80% over the default Cartesian topology built by Cray's MPI implementation MPT 3.1.02 on the world's current third fastest supercomputer, Jaguar, at Oak Ridge National Laboratory.