ABSTRACT

OPTIMIZING PARALLEL APPLICATIONS

by

Shih-Hao Hung

Chair: Edward S. Davidson

While parallel computing offers an attractive perspective for the future, developing efficient parallel applications today is a labor-intensive process that requires an intimate knowledge of the machines, the applications, and many subtle machine-application interactions. Optimizing applications so that they can achieve their full potential on parallel machines is often beyond the programmer's or the compiler's ability; furthermore its complexity will not be reduced with the increasingly complex computer architectures of the foreseeable future.

In this dissertation, we discuss how application performance can be optimized systematically. We show how insights regarding machine-application pairs and the weaknesses in their delivered performance can be derived by characterizing the machine, the application, and the machine-application interactions. We describe a general performance tuning scheme that can be used for selecting and applying a broad range of performance tuning actions to solve major performance problems in a structured sequence of steps, and discuss the interrelationship among and between performance problems and performance tuning actions. To guide programmers in performance tuning, we developed a goal-directed performance tuning methodology that employs hierarchical performance bounds to characterize the delivered performance quantitatively and explain where potential performance is lost. To reduce the complexity of performance tuning, we developed an innovative performance modeling scheme to quickly derive machine-application interactions from abstract representations of the machine and application of interest.

Collectively, this dissertation unifies a range of research work done within the Parallel Performance Project at the University of Michigan over the past seven years and significantly improves the state-of-the-art in parallel application development environments.