# Extrapolating memory behavior under strong scaling

Catherine Olschanowsky, Laura Carrington, Michael Laurenzano, Mustafa Tikir, Gabriel Marin, Allan Snavely

Performance models offer the ability to study and manipulate the complex interactions between applications and compute resources. This poster presents an overview of ongoing work to extend current modeling approaches to include strong scaling predictions and to find optimization opportunities in serial portions of code. Strong scaling predictions center around communication simulations using discrete event simulation. The simulations depend on accurately predicting the duration of CPU bursts, achieved by characterizing the memory access behavior during the bursts. We present the current status of our strong scaling predictions along with some initial data collected from PFLOTRAN.

Tuning the performance of parallel applications requires not only achieving good scalability as the number of processors increases, but also maximizing the application's performance on each individual node. We used modeling to understand the unfulfilled performance potential due to a mismatch between an application's characteristics and the resources present on the target architecture. We model an application's instruction execution cost and memory hierarchy utilization, and we identify performance problems arising from insufficient instruction-level parallelism or poor data locality. We present some preliminary tuning results, including two code transformations based on this analysis.