

APR14-2014-020176

Abstract for an Invited Paper  
for the APR14 Meeting of  
the American Physical Society

### **Porting Legacy LQCD Applications to GPUs**

M.A. CLARK, NVIDIA

The exponential growth of floating point power in GPUs, combined with high memory bandwidth, has given rise to an attractive platform upon which to deploy HPC applications. When it comes to legacy applications there is a danger that entire codebases have to be rewritten to fully embrace this computational power. In this session we discuss how to efficiently port legacy lattice quantum chromodynamics (LQCD) applications, e.g., MILC and Chroma, onto GPUs avoiding this rewriting overhead. The approach taken is a community-wide library (QUDA) which provides high-performance implementations for the time-critical LQCD algorithms, which can be linked into any legacy lattice QCD application, providing instant GPU acceleration. We discuss some of the bleeding-edge strategies taken by QUDA to maximize performance, including the use of communication reducing algorithms, mixed-precision methods and an aggressive auto-tuning methodology. While algorithms and routines that are not offloaded to QUDA will typically not be time-critical, they can potentially limit the overall speedup due to the onset of Amdahl's law. We discuss various compile-and-run strategies to circumvent this, including the use of OpenACC directives or retargeting the underlying domain-specific language (DSL) to generate GPU code directly from the original source.