



HPC Server Performance and Power Consumption for Atmospheric Modeling on GPUs Configured with Different CPU Platforms

Stan Posey, Peter Messmer, and Jeremy Appleyard
United States (sposey@nvidia.com)

Current trends in high performance computing (HPC) are moving towards the use of graphics processing units (GPUs) to achieve speedups through the extraction of fine-grain parallelism of application software. GPUs have been developed exclusively for computational tasks as massively-parallel co-processors to the CPU, and during 2014 the latest NVIDIA GPU architecture can operate with as many as three CPU platforms. In addition to the conventional use of the x86 CPU architecture with GPUs starting from the mid-2000's, the POWER and ARM-64 architectures have recently become available as x86 alternatives.

Today computational efficiency and increased performance per energy-cost are key drivers behind HPC decisions to implement GPU-based servers for atmospheric modeling. The choice of a server CPU platform will influence performance and overall power consumption of a system, and also the available configurations of CPU-to-GPU ratio. It follows that such system design configurations continue to be a critical factor behind scientific decisions to implement models at higher resolutions and possibly with an increased use of ensembles.

This presentation will examine the current state of GPU developments for atmospheric modeling with examples from the COSMO dycore and from various WRF physics, and for different CPU platforms. The examples provided will be relevant to science-scale HPC practice of CPU-GPU system configurations based on model resolution requirements of a particular simulation. Performance results will compare use of the latest available CPUs from the three available CPU architectures, both with and without GPU acceleration. Finally a GPU outlook is provided on GPU hardware, software, tools, and programmability for each of the available CPU platforms.