

# GPU with Tensor Cores

- ▶ Tensor cores (NVIDIA) operate  $\approx 16$  times faster than normal computation (for some applications)
- ▶ Use 16-b, half-precision arithmetic, a floating-point format that uses a fraction of the digits compared with double (64-b) precision

## Mixed Precision: A Strategy for New Science Opportunities

**Jonathan Hines**  
Oak Ridge National  
Laboratory

**Editors:**  
**James J. Hack**  
Oak Ridge National Laboratory  
jhack@ornl.gov

**Michael E. Papka**  
Argonne National Laboratory  
papka@anl.gov

Since the days of vector supercomputers, computational scientists have relied on high-precision arithmetic to accurately solve problems. But changes to hardware, spurred by the demand for more computing capability and growth in machine learning, have researchers considering lower precision alternatives.