BLAS Interface for Different Precisions

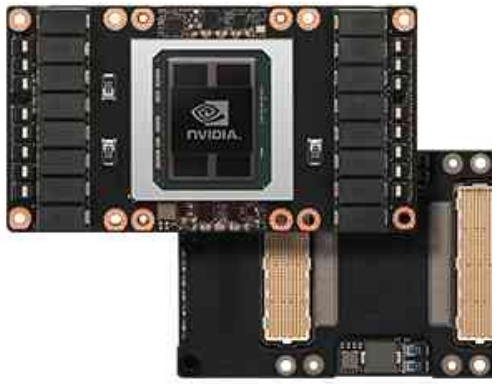
Jack Dongarra

XBLAS

- BLAS Technical Forum defined Extended and Mixed Prevision BLAS
  - Single, Double, Indigenous, Extra
  - BLAS_ZGEMM with argument PREC

Nvidia 16 bit BLAS

- cublasHgemm() has been added to support half-precision floating point (FP16).
- Nvidia's P100
  - 5.3 TeraFLOPS double-precision performance
  - 10.6 TeraFLOPS single-precision performance
  - 21.2 TeraFLOPS half-precision performance



Compiler Support?

- gcc support __fp16 type
  - but only as storage type - to compute they cast to float, etc

Proposing Something Along the Lines of…

- RGEMM_xx and CGEMM_xx
  - With the same calling sequence as SGEMM and CGEMM
- RGEMM_32 is equivalent to SGEMM
- RGEMM_64 is equivalent to DGEMM
- RGEMM_16 would be the 16-bit floating point version of GEMM.

- RGEMM_128 etc.
- Arbitrary? RGEMM_12