Performance of mixed-precision GEMM Problem size: N x N x N CUDAnative cuBLAS CUTLASS (WMMA) 30 10 2<sup>8</sup> 2<sup>10</sup> 2<sup>12</sup>