Performance of mixed-precision GEMM
Problem size: N x N x N

- CUDAnative
- cuBLAS
- CUTLASS (WMMA)
- CUTLASS (mma.m8n8k4)

TFLOPS vs N