Roofline Model for Dense Loop-over-GEMM kernel: Roofline Gemmforge 1200 cuTensor 1000 Performance (GFLOPs/s) 800 600 42.9% 91.8% 400 91.4% 23.6% 200 39.5% 39.7% 47.0% 20.1% 2.4% 0 Kernel Type