Scaling Across GPU architectures r = 1, n = 409603.5 **BASE** SHARED 3.0 CAT **COARSE MCELL PACK** 0.5 A100 H100 (Volta) (Ampere) (Hopper)

GPU