Parallelism strategies for H1070a634Melism strategies for H100 SXMPZexideLiesternstyrategies for H100 S74VaCGebilsen Lett/Latineligies for H100 SXMP2Grizolbeellskit/Latineligies for H100 SXMP2Grizo Tensor parallelism Tensor parallelism Tensor parallelism Tensor parallelism 0.7 0.7 0.7 0.7 0.7 -Data parallelism Data parallelism Data parallelism Data parallelism Pipeline parallelism Pipeline parallelism Pipeline parallelism Pipeline parallelism 0.6 -0.6 -0.6 -0.6 -0.6 Expert parallelism Expert parallelism Expert parallelism Expert parallelism 0.5 0.5 fraction 7.0 fraction . 7.0 fraction 6.4. Tensor parallelism ၉ 0.4 Data parallelism Parallelism .0 .0 Parallelism 0 .c ırallelism allelisn Pipeline parallelism allelis 0.3 0.3 Expert parallelism 0.2 0.2 0.2 0.2 0.2 0.1 0.1 0.1 0.1 0.1 0.0 0.0 0.0 0.0  $10^{28}$  $10^{32}$  $10^{27}$  $10^{33}$  $10^{30}$  $10^{32}$  $10^{27}$  $10^{30}$  $10^{33}$  $10^{27}$  $10^{33}$  $10^{36}$  $10^{36}$ Training compute (16-bit FLOP) Training compute (16-bit FLOP) Training compute (16-bit FLOP) Training compute (16-bit FLOP) Training compute (16-bit FLOP)