Parallelism strategies for H10Ra6xMelism strategies for H100 SXMPZexideLieuenstyrategies for H100 SXMPZexideLieuenstyrategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMPzexideLieuenstyrategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMPzexideLieuenstyrategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMPzexideLieuenstyrategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMParallelism strategies for H100 SXMPzexideLieuenstyrategies for H100 SXMParallelism strategies for H100 SXMParallelis Tensor parallelism Tensor parallelism Tensor parallelism 0.7 0.7 0.7 0.7 Data parallelism Data parallelism Data parallelism Pipeline parallelism Pipeline parallelism Pipeline parallelism 0.6 -0.6 -0.6 -0.6 0.6 Expert parallelism Expert parallelism Expert parallelism 0.5 0.5 0.5 0.5 0.5 fraction 6.0 fraction 7.0 Tensor parallelism Tensor parallelism ဨႍ 0.4 -၉ 0.4 Data parallelism Data parallelism Parallelism .0 .c Parallelism allelism ... Parallelism Pipeline parallelism Pipeline parallelism 0.3 0.3 Expert parallelism 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^{36} 10^{28} 10^{32} 10^{27} 10^{30} 10^{33} 10^{27} 10^{33} 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)