Gradient Norm Comparison: Max/Min Gradient Norm Ratio (Layer Imbalance) Optimizer 10^{14} **ADAMW ADAM** SGD 10^{12} GGD Max/Min Gradient Norm Ratio GGD LW **NOVOGRAD** 10^{10} 10⁸ 10^{6} 10^{4} 10^{2} 200 400 600 800 1000 0 **Iteration**