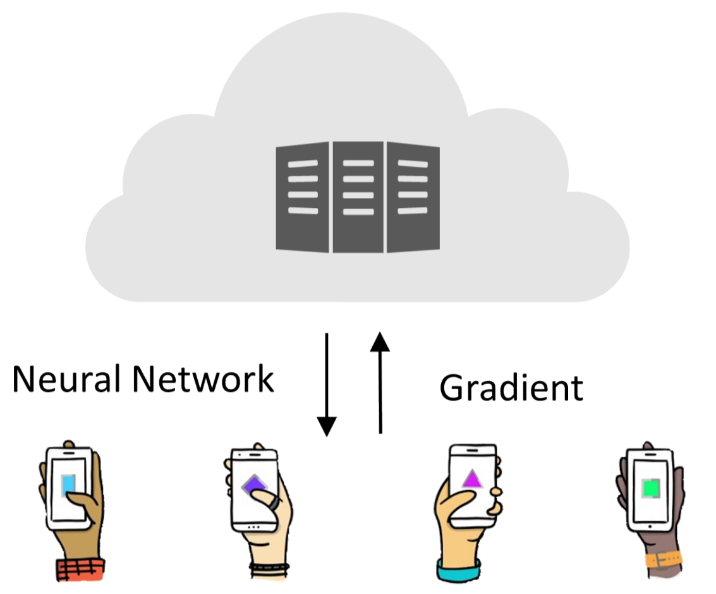
Neural Tangent Kernel Empowered Federated Learning

Kai Yue,[1] Richeng Jin,[1] Ryan Pilgrim,[2] Chau-Wai Wong,[1] Dror Baron,[1] Huaiyu Dai[1]
[1]North Carolina State University, USA [2]Independent Scholar

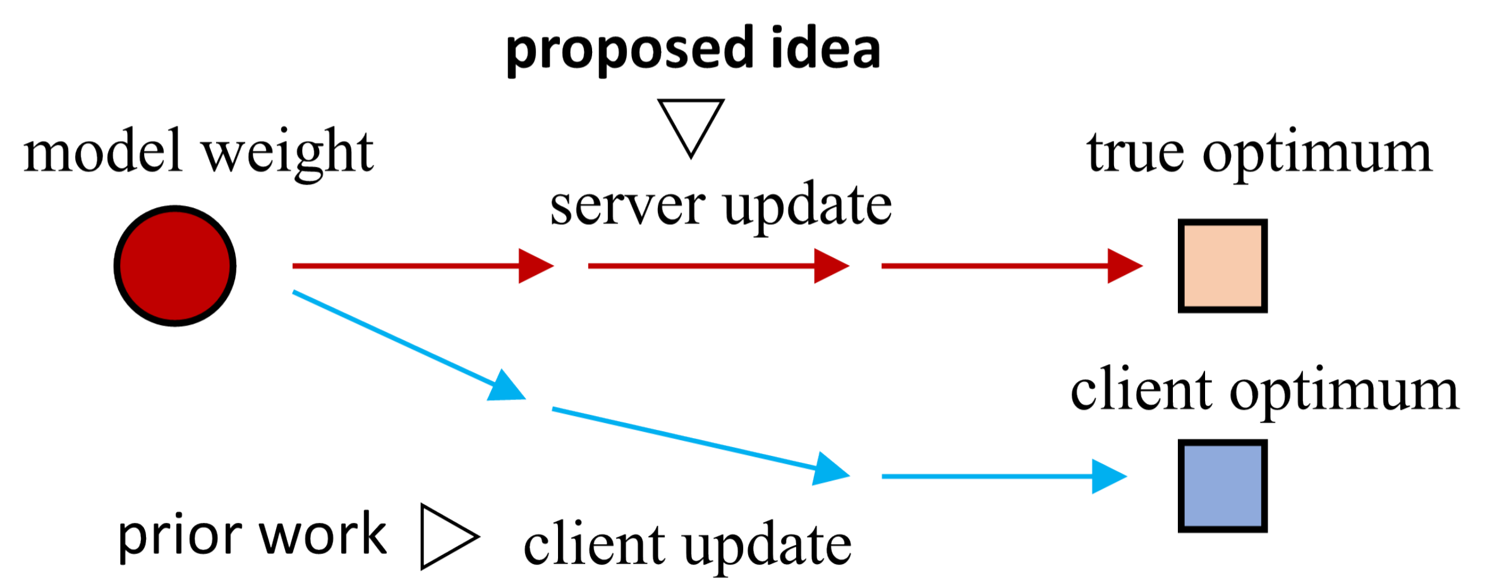
Introduction & Motivation

- Federated learning (FL): privacy-preserving paradigm that enables client collaboration
- Challenges:
  - client data deviate from IID
  - communication cost
  - privacy protection
  - stragglers

  . . .

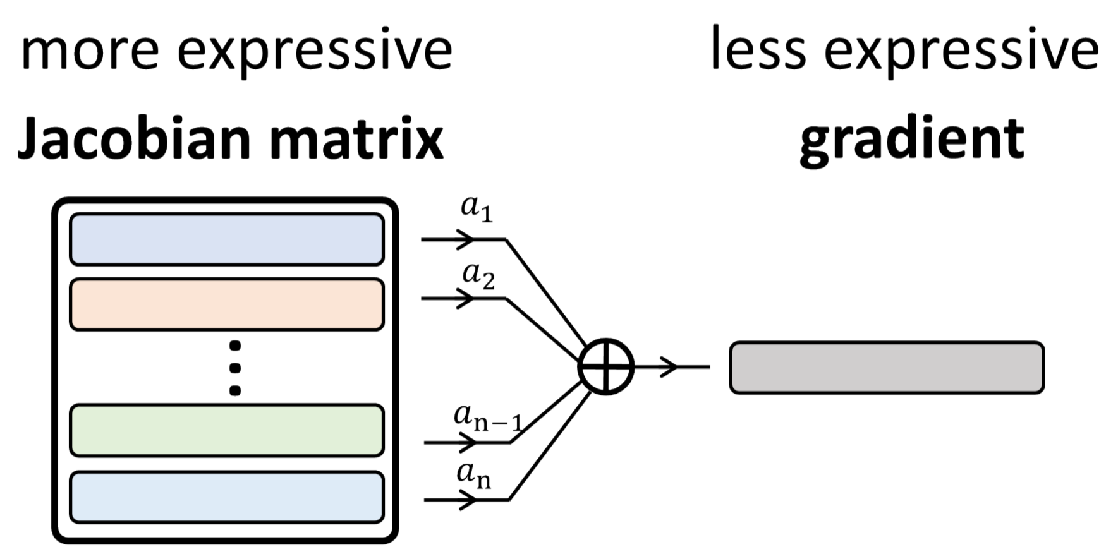


- FedAvg (McMahan et al., 2016)
  - server broadcasts global model
  - client updates & uploads local model
- Local update on client side → client optimum

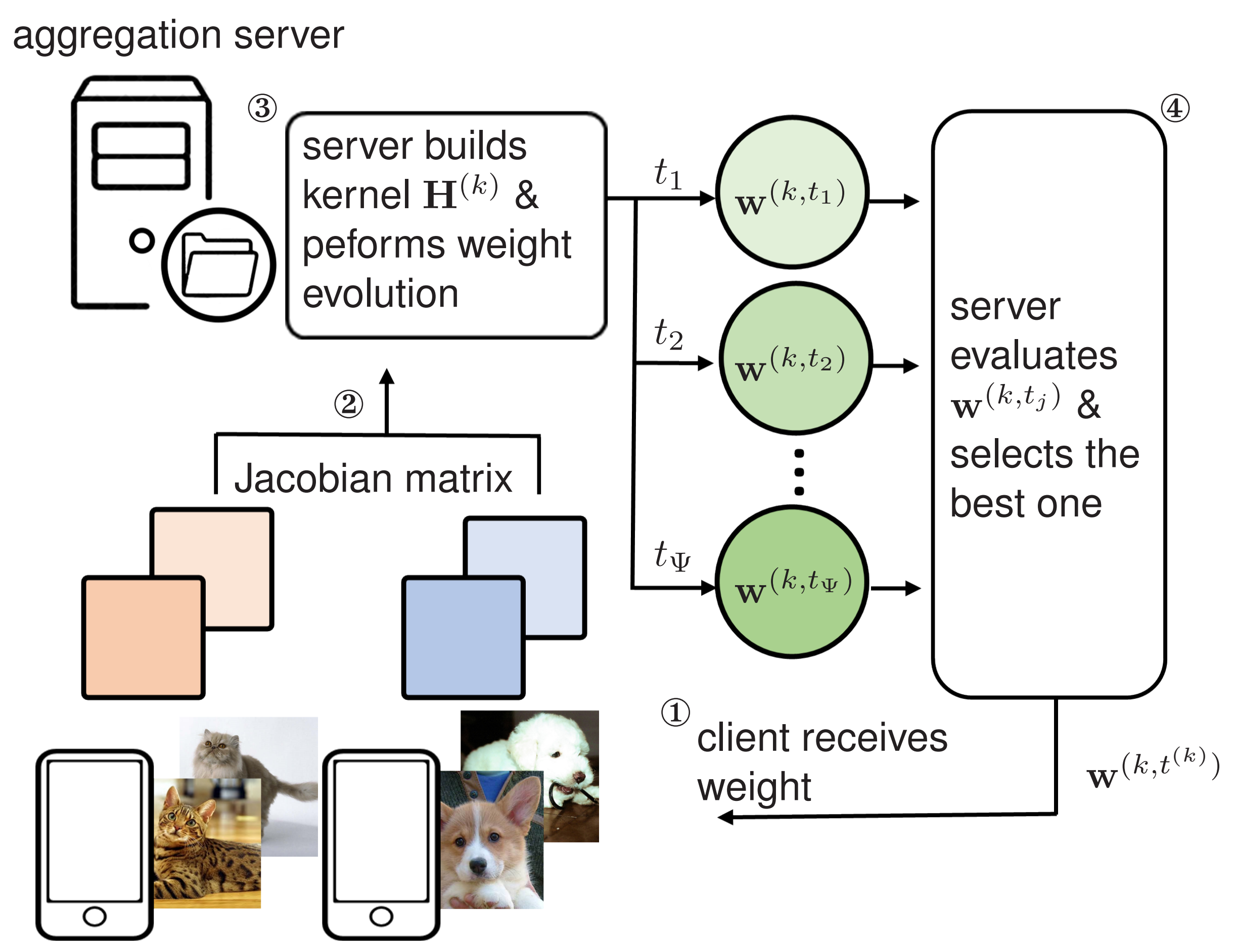


Proposed Neural Tangent Kernel (NTK) Empowered Federated Learning (FL)

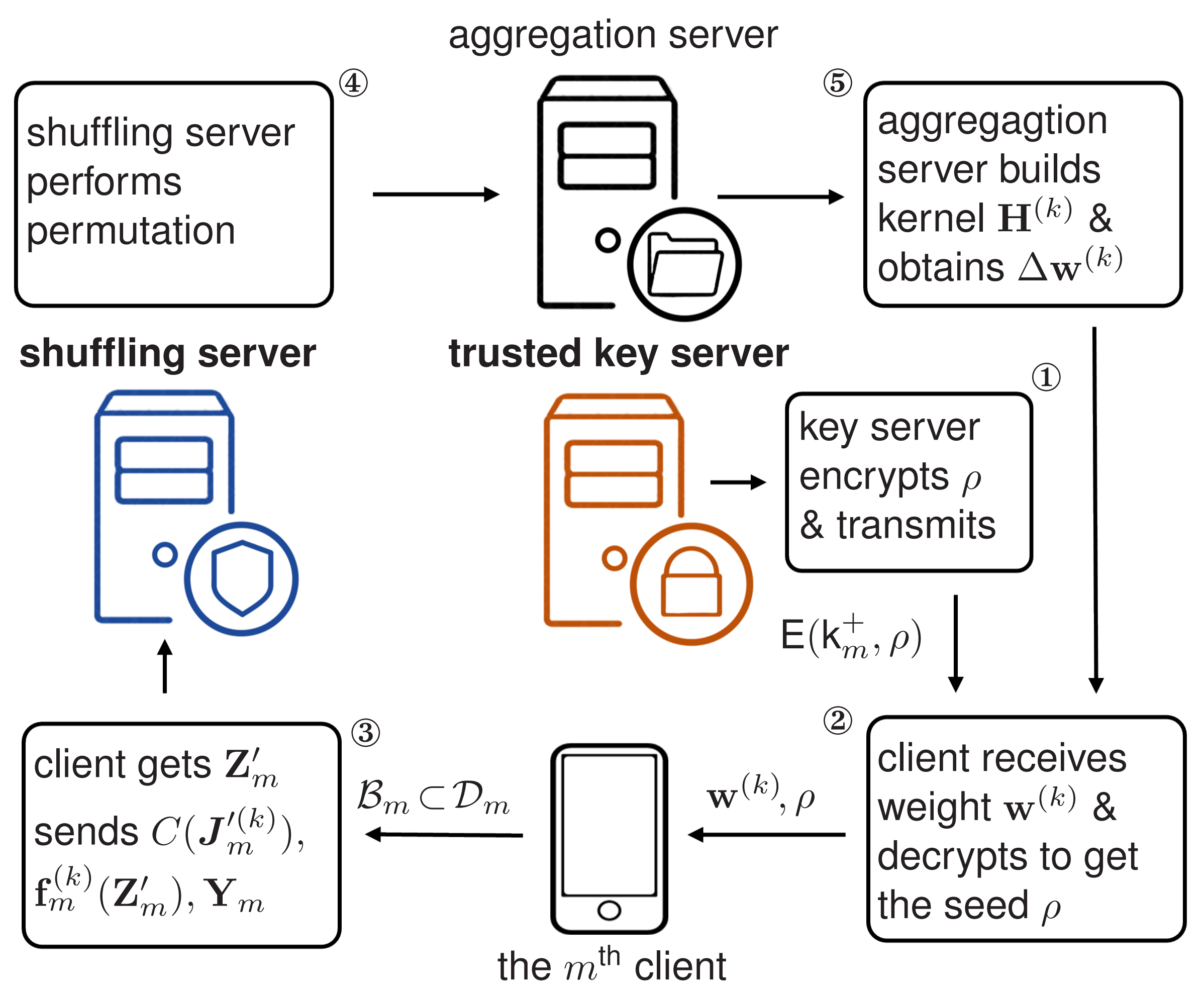
- **NTK-FL**: combine Jacobian matrices & construct a global kernel $\mathbf{H}^{(k)}$, perform NTK evolution on server
  - More *expressive* compared to gradient in FedAvg
  - Perform *multi-step server update* & choose best candidate
- **CP-NTK-FL**: add **c**ommunication-efficient & **p**rivacy-preserving features, such as projection and subsampling



NTK-FL


aggregation server
③ server builds kernel $\mathbf{H}^{(k)}$ & peforms weight evolution
② Jacobian matrix
t_1
t_2
t_Ψ
$\mathbf{w}^{(k,t_1)}$
$\mathbf{w}^{(k,t_2)}$
$\mathbf{w}^{(k,t_\Psi)}$
④ server evaluates $\mathbf{w}^{(k,t_j)}$ & selects the best one
① client receives weight
$\mathbf{w}^{(k,t^{(k)})}$

CP-NTK-FL

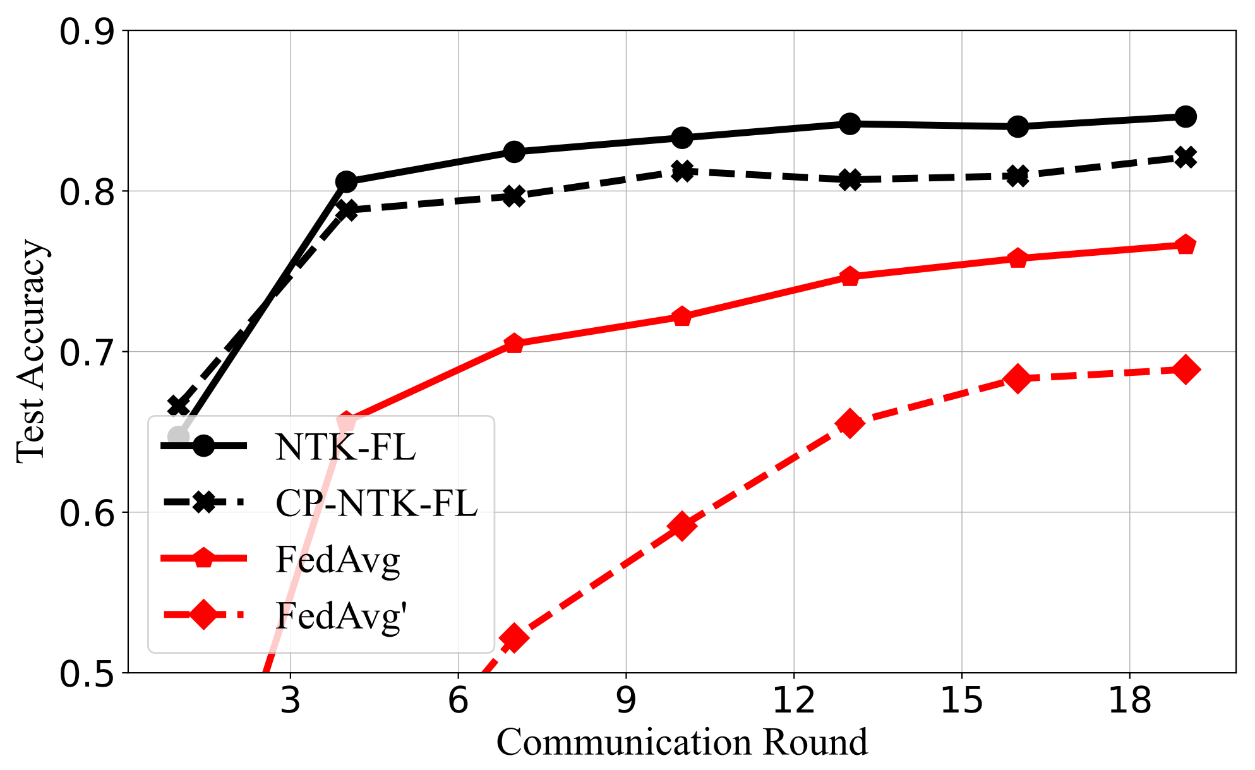

aggregation server
④ shuffling server performs permutation
⑤ aggregagtion server builds kernel $\mathbf{H}^{(k)}$ & obtains $\Delta\mathbf{w}^{(k)}$
shuffling server
trusted key server
① key server encrypts ρ & transmits
$\mathsf{E}(\mathsf{k}_m^+, \rho)$
② client receives weight $\mathbf{w}^{(k)}$ & decrypts to get the seed ρ
$\mathbf{w}^{(k)}, \rho$
$\mathcal{B}_m \subset \mathcal{D}_m$
③ client gets $\mathbf{Z}'_m$ sends $C(\boldsymbol{J}'^{(k)}_m)$, $\mathbf{f}^{(k)}_m(\mathbf{Z}'_m), \mathbf{Y}_m$
the m^{th} client

Robustness of CP-NTK-FL

- CP-NTK-FL: achieve target accuracy within the fewest communication rounds

optimization algorithms	comm. rounds	comm. cost (MB)
CP-NTK-FL	26	386
FedCOM	250	379
QSGD (4 bit)	614	465
FedAvg	284	1720

- Compression tools works better on CP-NTK-FL
  - use Top-k sparsification & subsampling



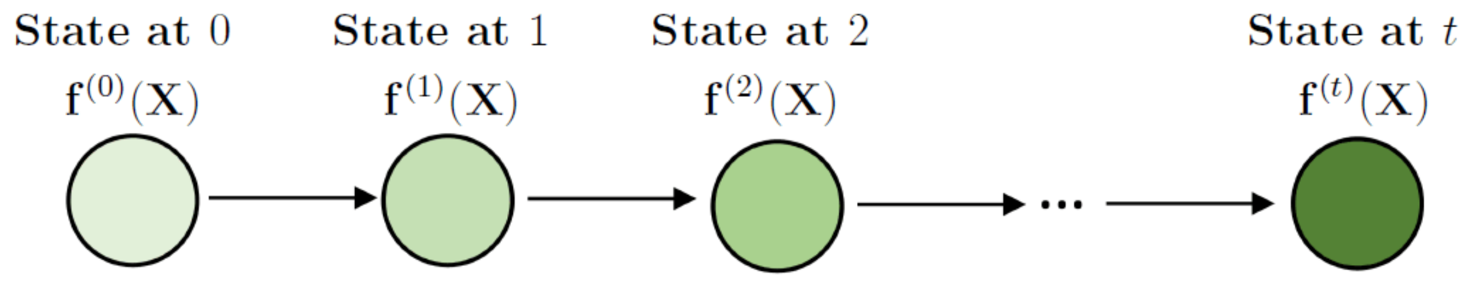
Neural Tangent Kernel

- Neural tangent kernel (NTK): capture training dynamics of a wide neural networks

$$\text{rate of change} \quad \frac{\mathrm{d}\mathbf{f}}{\mathrm{d}t} = \underset{\text{learning rate}}{\eta}\underset{\text{kernel matrix}}{\mathbf{H}}\left[\mathbf{Y} - \mathbf{f}^{(t)}(\mathbf{X})\right]$$

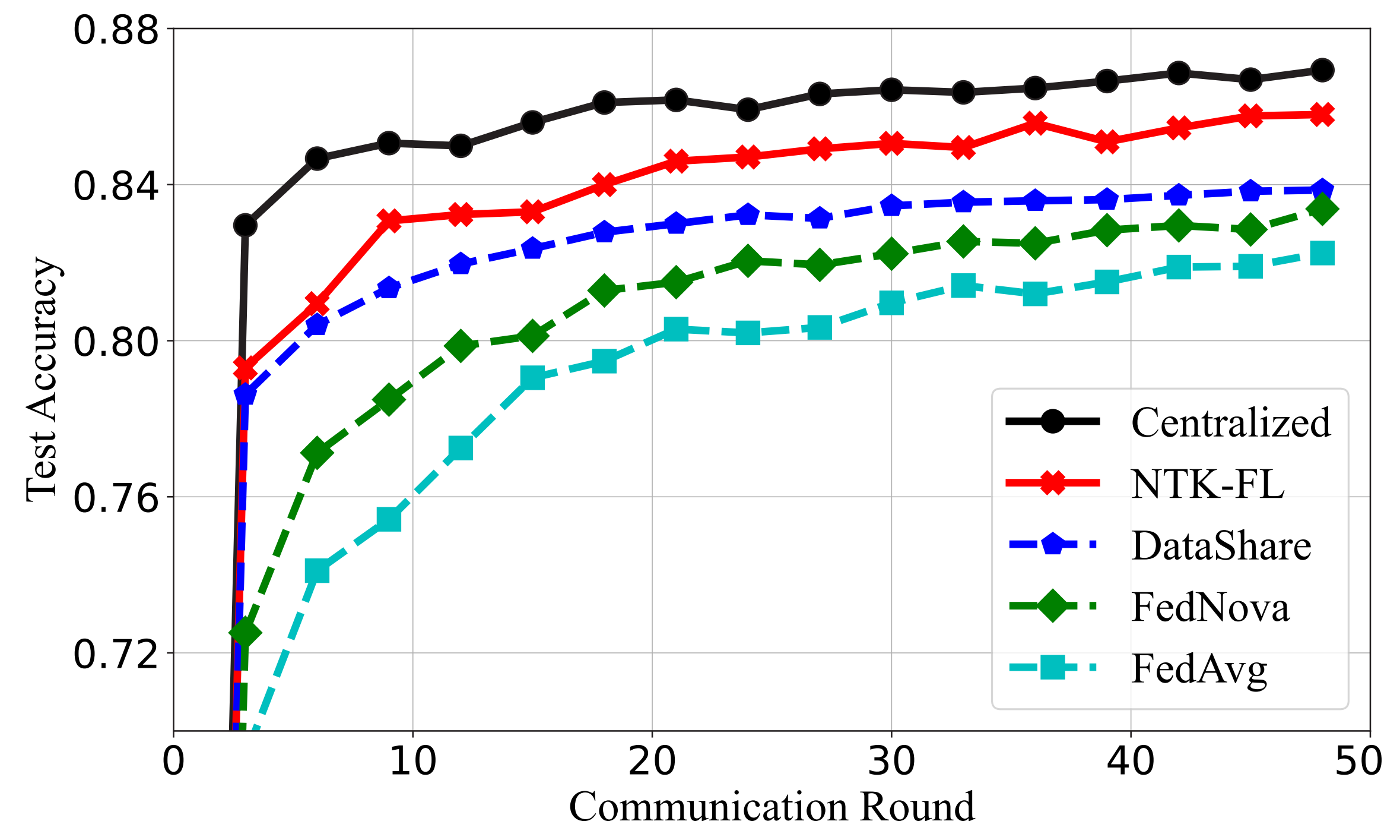
- Solution to the differential equation → model state

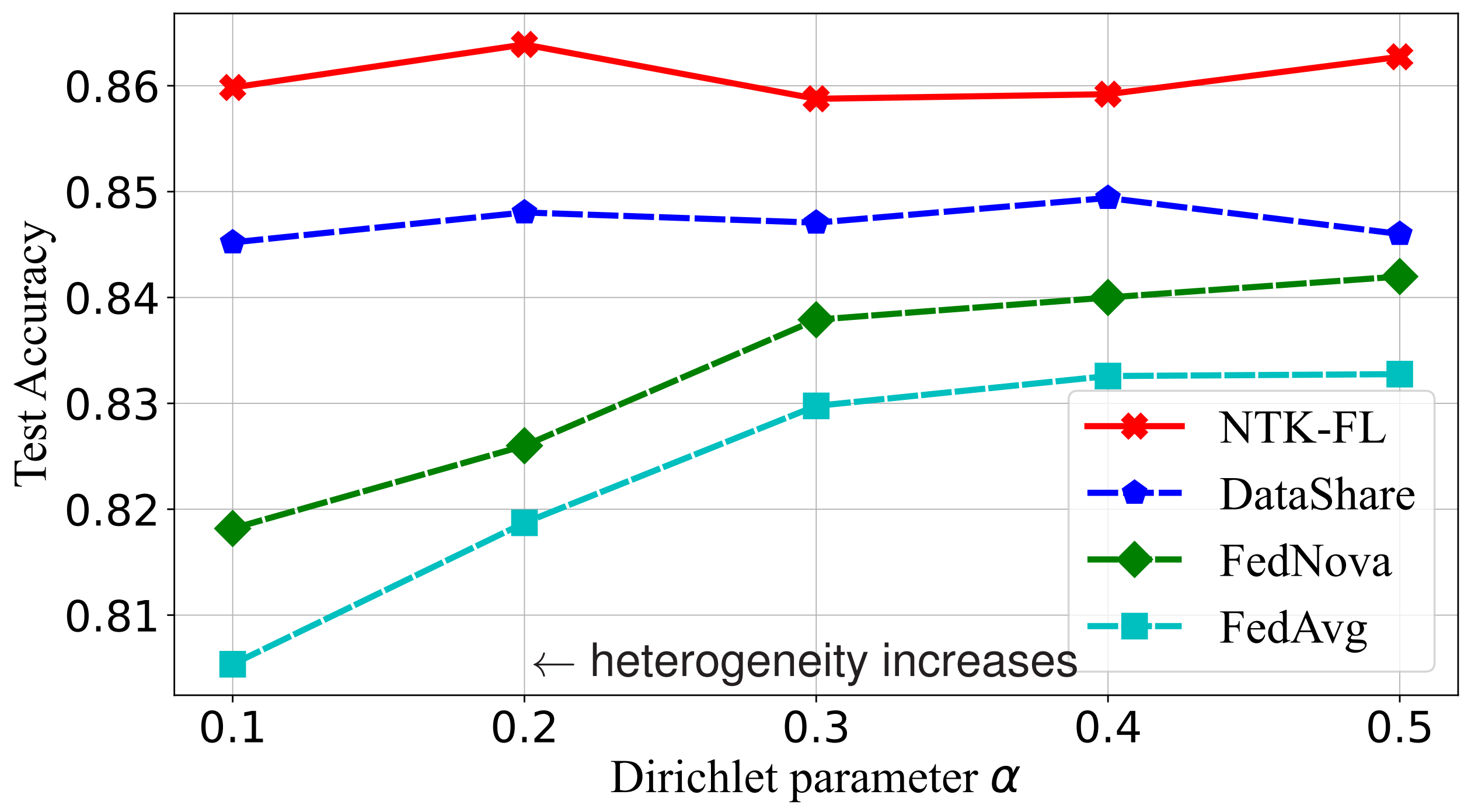
$$\mathbf{f}^{(t)}(\mathbf{X}) = \left(\mathbf{I} - e^{-\frac{\eta t}{N}\mathbf{H}^{(0)}}\right)\mathbf{Y} + e^{-\frac{\eta t}{N}\mathbf{H}^{(0)}}\mathbf{f}^{(0)}(\mathbf{X})$$


State at 0
$\mathbf{f}^{(0)}(\mathbf{X})$
State at 1
$\mathbf{f}^{(1)}(\mathbf{X})$
State at 2
$\mathbf{f}^{(2)}(\mathbf{X})$
State at t
$\mathbf{f}^{(t)}(\mathbf{X})$

Robustness of NTK-FL

- NTK-FL approaches centralized learning and is robust to different non-IID settings
  - Left: test accuracy v. communication round of different methods on non-IID Fashion-MNIST
  - Right: test accuracy with different degrees of heterogeneity (controlled by Dirichlet parameter α)





Conclusion

- Characteristics of NTK-FL highlighted
  - enable multi-step server update
  - ↘ negative influence of data heterogeneity
  - adaptively choose update steps
- Potential challenges
  - uplink communication cost
  - expressive information may leak more privacy
- CP-NTK-FL improves efficiency & privacy
- Please refer to our paper for more details

paper