

01 Motivation

The problem

Clustering compresses weight values, but full index maps still scale with model size. Repeatedly transferring those assignments can dominate communication, while stale mappings can prevent convergence.

Contributions

- **Bidirectional savings:** exchange codebooks for both server-to-client and client-to-server communication.
- **Index-free rounds:** update weights by nearest-center mapping without transmitting assignments.
- **Periodic calibration:** restore assignments to manage the communication-accuracy trade-off.

Two parts of a compressed model

A codebook stores K cluster centers; an index map assigns P weights to them. Sending centers alone avoids the P-sized assignment payload.

What the reduction measures

Data transmission reduction (DTR) is FedAvg bits divided by method bits. Larger values mean less communication; they do not establish latency, energy or client wall-clock savings.

Scope

F1 and F2 control periodic index exchange; clustering adds local work. Calibration is needed for the reported accuracy, and repeated-run uncertainty and end-to-end runtime are unreported.

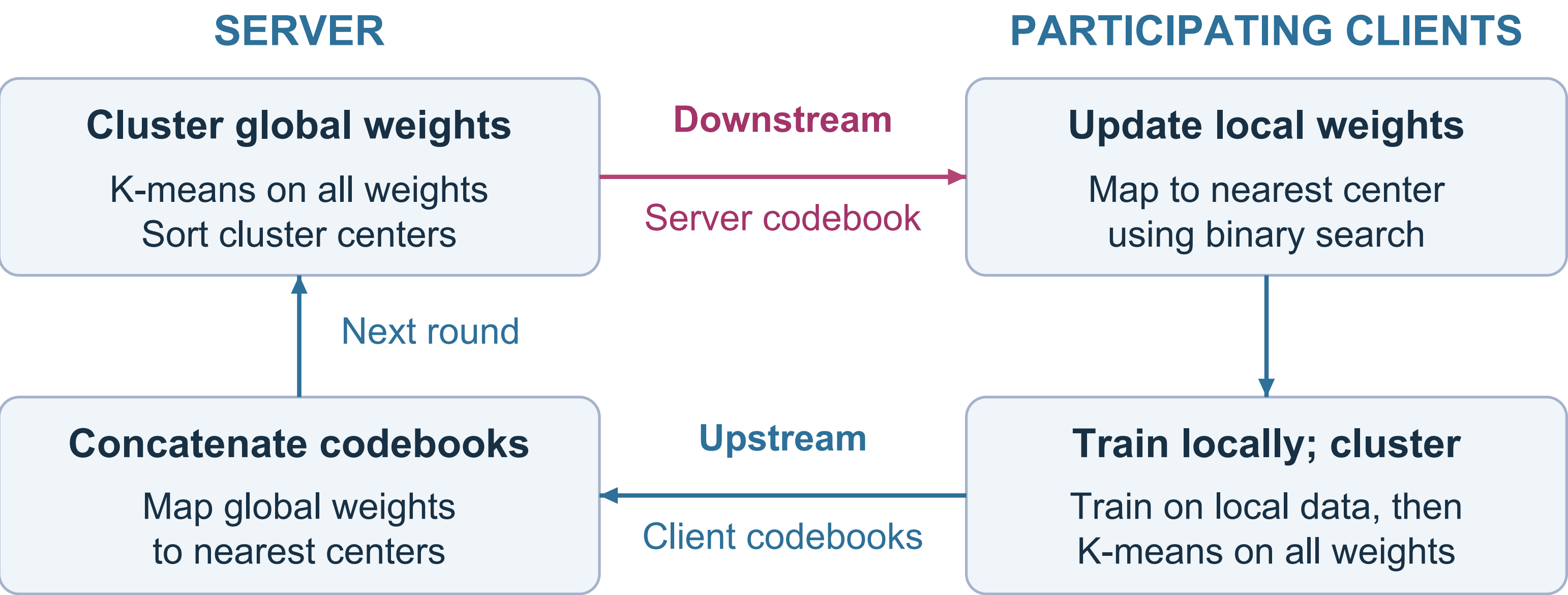
Notation

independent and identically distributed (IID)

02 FedCode

Codebook exchange with periodic calibration

ORDINARY CODEBOOK ROUNDS



Codebook = cluster centers; ordinary rounds omit weight indices.

PERIODIC CALIBRATION — SELECTED ROUNDS

Transfer codebook + weight indices in the relevant direction.

Clients reconstruct weights from indices and centers.

The server reconstructs client models, then applies FedAvg.

Technical core

$$j^* = \arg \min_j |w_i - c_j|, \quad w_i \leftarrow c_{j^*}$$

w_i : existing weight; c_j : received center. Ordinary rounds choose the nearest center; the server uses concatenated client codebooks.

$$\text{DTR} = \frac{2Pb}{2Kb + (F_1 + F_2)P\lceil \log_2 K \rceil}$$

P weights, K centers, b bits/weight; F1 and F2 are calibration frequencies. Less index transfer reduces traffic.



Paper



Code

03 Evaluation

14.6×

joint communication reduction

IID CIFAR-100; ResNet-20.

58.07%

FedCode test accuracy

FedAvg: 61.22% on the same workload.

Compress both communication directions

Method	Accuracy (%)	Down DTR	Up DTR	Joint DTR
FedAvg	61.22	—	—	—
OMC-T	61.21	—	2.9	1.4
FedCompress	59.67	4.11	4.11	4.11
FedZip	57.98	—	19.8	1.9
FedCode (Ours)	58.07	24.2	10.4	14.6

IID CIFAR-100 · ResNet-20 · 10 clients · full participation · 100 rounds · 4 local epochs. 64 centers; F1=0.2, F2=0.5.

Calibration controls the trade-off

F1 / F2	Accuracy Δ (pp)	Joint DTR	Data (MB)
1 / 1	-0.86	5.3	394.21
0.2 / 0.5	-3.15	14.6	143.92
0.1 / 0.1	-21.64	44.3	47.35
0 / 0	-46.88	264.7	7.9

Same setting; accuracy change vs FedAvg. FedAvg communication: 2102 MB.

Takeaway

FedCode achieves **14.6× joint communication reduction** at 58.07% accuracy, 3.15 points below FedAvg. Without calibration, communication falls further but accuracy drops **46.88 points** below FedAvg.