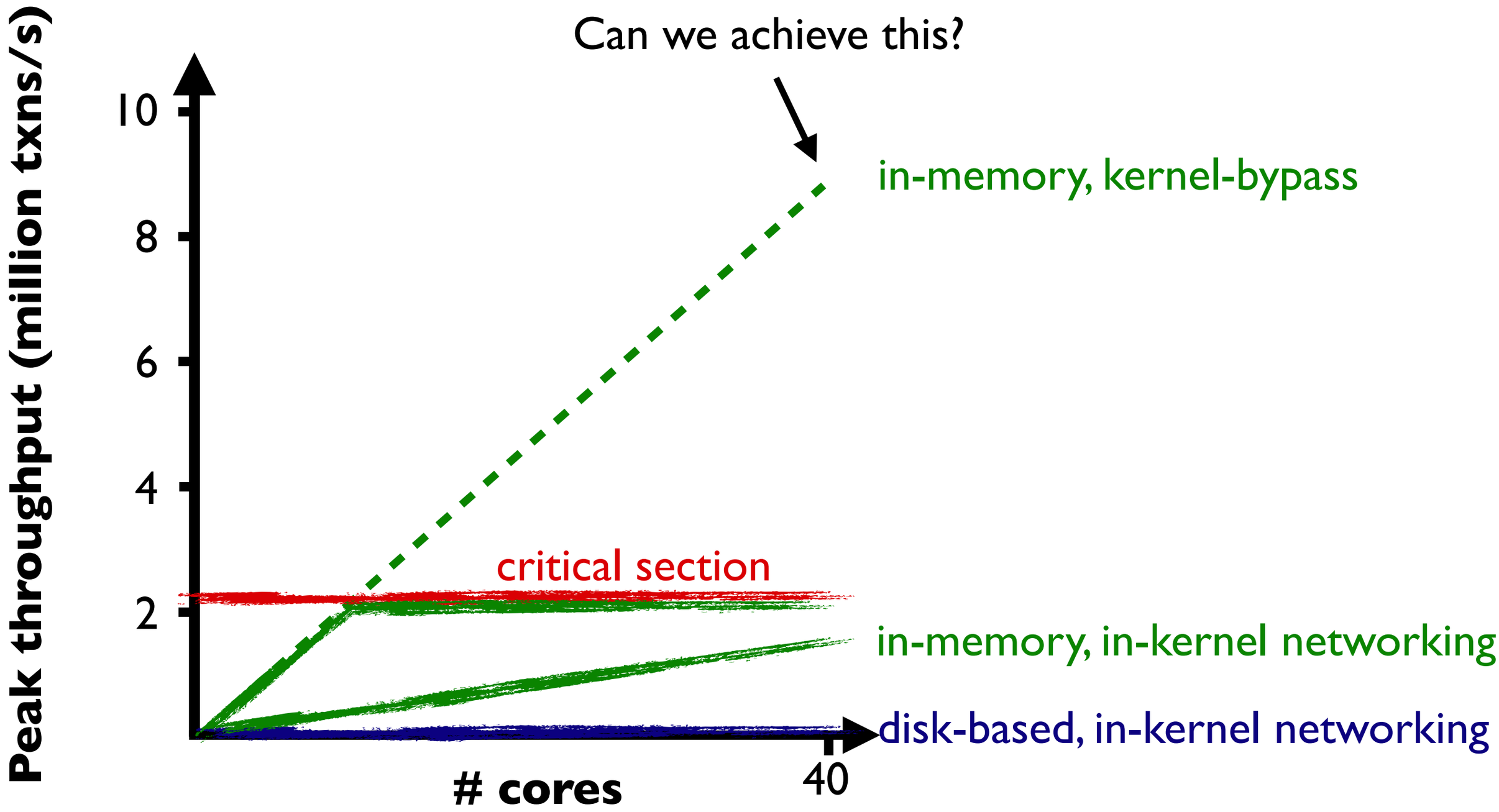
# Meerkat: Multicore-Scalable Replicated Transactions Following the Zero Coordination Principle

# Distributed storage systems are getting faster

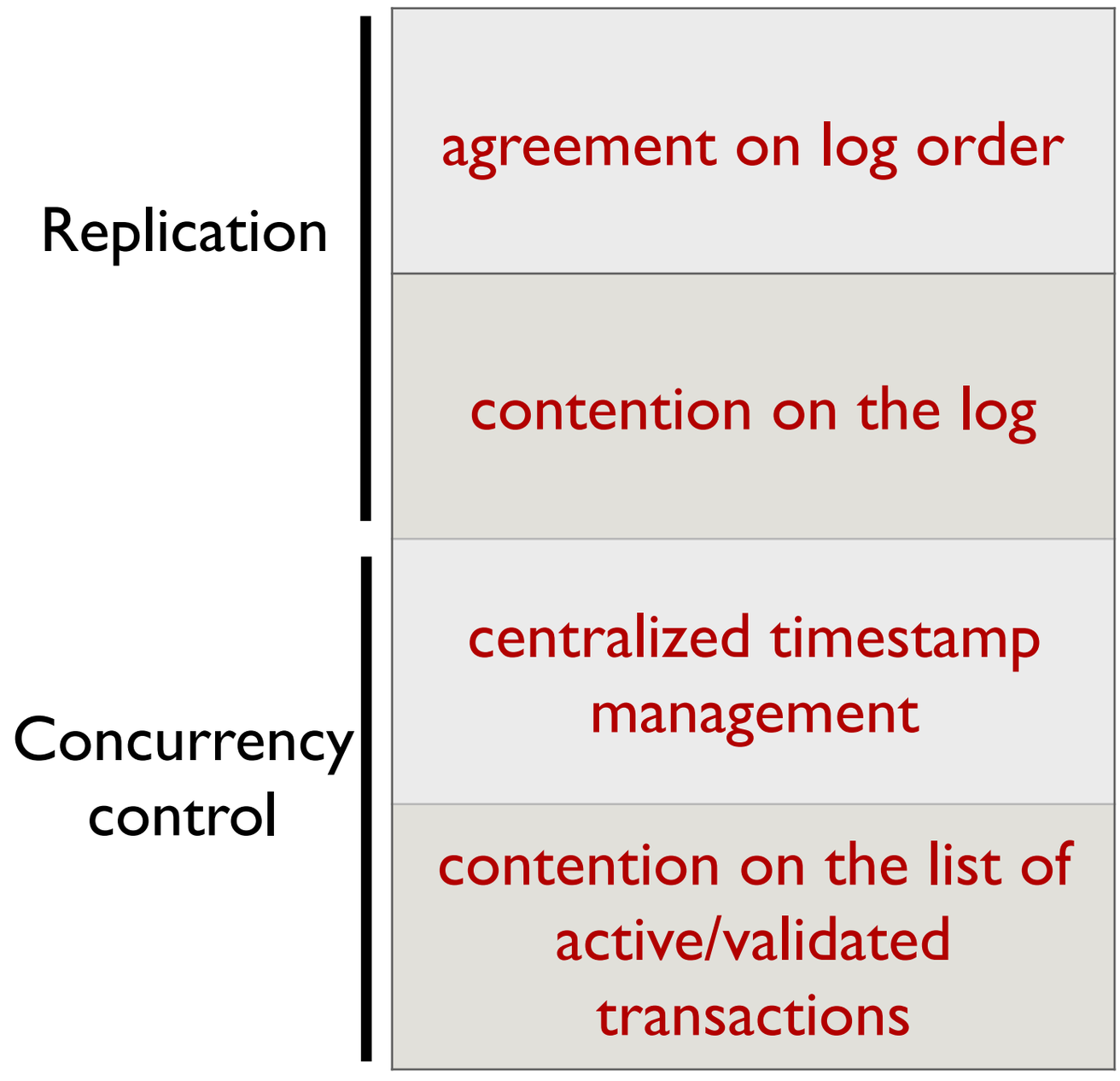


# The Zero-Coordination Principle:

When transactions do not conflict:

- No writes to memory shared with other cores (P1)
- No cross-replica coordination (P2)

# Common ways in which existing systems violate ZCP



| | Other systems | Meerkat |
|---|---|---|
| Replication | agreement on log order | decentralized agreement on transaction status |
| | contention on the log | per-core record of transactions |
| Concurrency control | centralized timestamp management | clients pick the commit timestamp |
| | contention on the list of active/validated transactions | key-parallel OCC |

# Meerkat's approach

Get rid of the log!
Use a decentralized approach instead.

# Meerkat's decentralized approach

- Decentralized OCC
  - client picks a commit timestamp using loosely synchronized clocks
  - replicas **independently** check for conflicts

- Fast, decentralized consensus

**(fast path)** – client learns the fate of the transaction

**(slow path)** – client proposes to commit the transaction only if OCC checks successful at a majority

Correctness comes from quorum intersection + pairwise conflict checks; see paper

| | Other systems | Meerkat | |
|---|---|---|---|
| Replication | agreement on log order | decentralized agreement on transaction status | ZCP ✓ |
| | contention on the log | per-core record of transactions | |
| Concurrency control | centralized timestamp management | clients pick the commit timestamp | ZCP ✓ |
| | contention on the list of active/validated transactions | key-parallel OCC | |

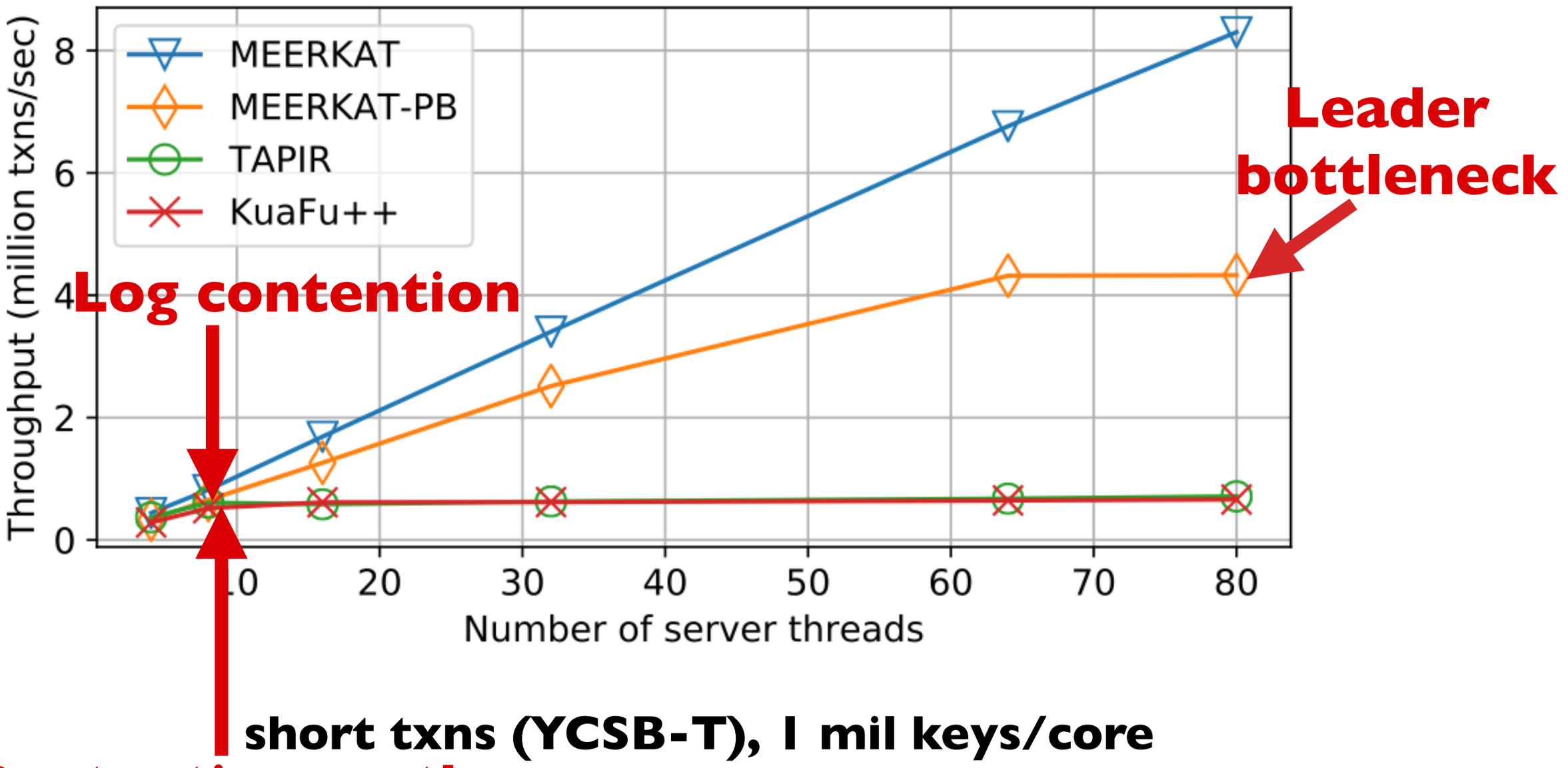
# Meerkat also has some nice performance properties

- Low latency (no leader)
  - commits transactions in 1 RTT
    (in the absence of conflicts and failures)
  - waits for replies from the fastest replicas
- Read from any replica
  - balance the workload

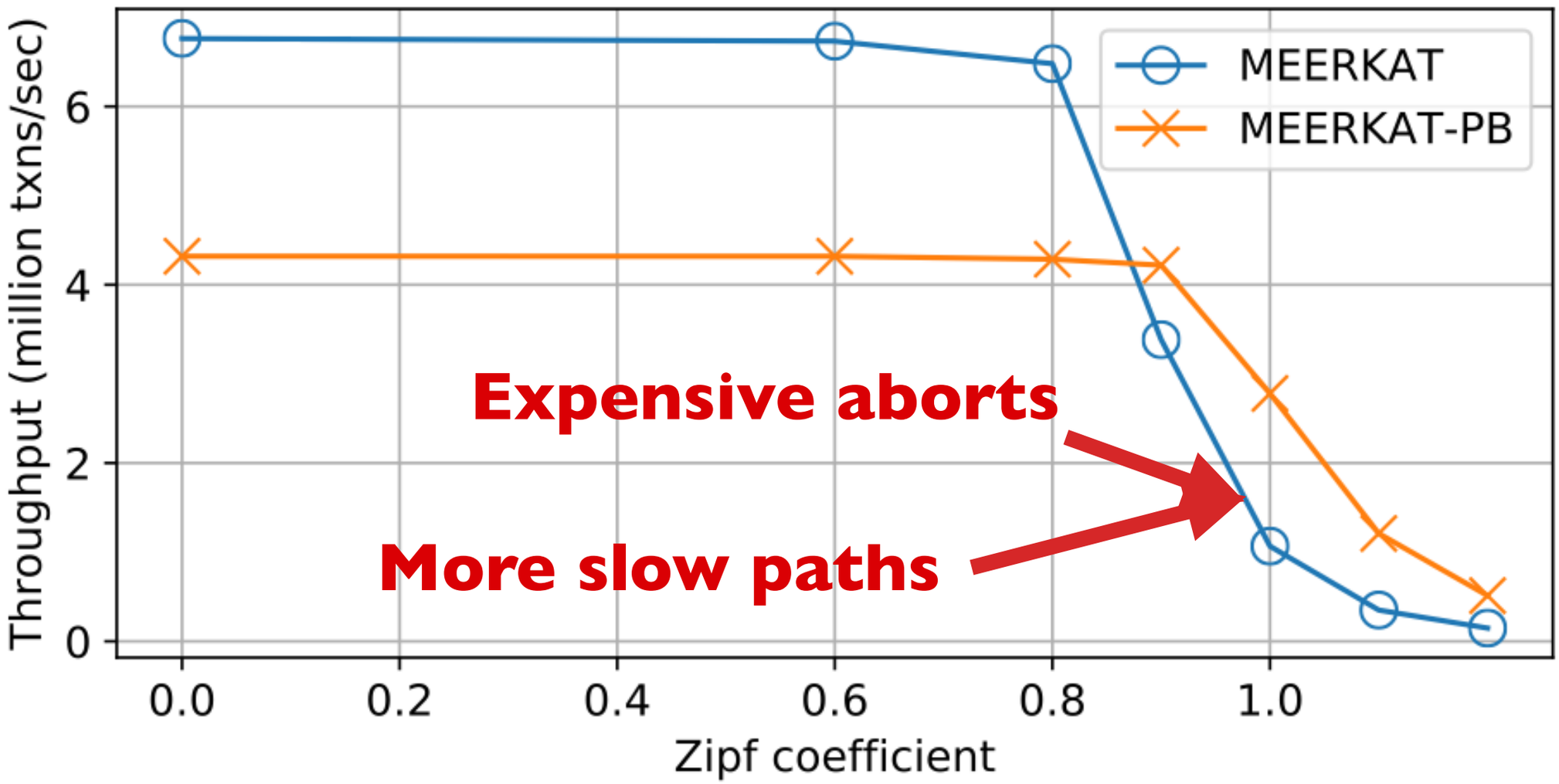
# Prototypes

| | **No cross-processor coordination** | **No cross-replica coordination** |
|---|---|---|
| KuaFu++ | X | X |
| TAPIR | X | ✓ |
| Meerkat-PB | ✓ | X |
| Meerkat | ✓ | ✓ |

# Meerkat scales near linearly when low contention (uniform)



# Meerkat performs well for low to medium contention




**short txns (YCSB-T), 1 mil keys/core, 64 hyperthreads**