A DISTRIBUTED DYNAMIC LOAD BALANCER FOR ITERATIVE APPLICATIONS

Harshitha Menon (gplkrsh2@illinois.edu)
Laxmikant Kale (kale@illinois.edu)
University of Illinois at Urbana Champaign

MOTIVATION

PROPOSED WORK

EVALUATION

MOTIVATION

Load Imbalance

MOTIVATION

Dynamic Load Imbalance

DYNAMIC LOAD BALANCER SHOULD…

Perform good load balance

Incur minimum overhead

Be profitable!

LOAD BALANCERS

Centralized

Distributed

Hierarchical

CENTRALIZED LB

Global view of the system

Bottleneck

DISTRIBUTED LB

Local view of the system

Scalable

Poor load balance

HIERARCHICAL LB

Subgroup of processors

Decisions at the root

Scalable

MOTIVATION

CentralizedLB	DistributedLB
Global view	Limited view
Bottleneck	Scalable
Good Load balance	Poor Load balance

GRAPEVINE-LB

GRAPEVINE-LB

Fully distributed

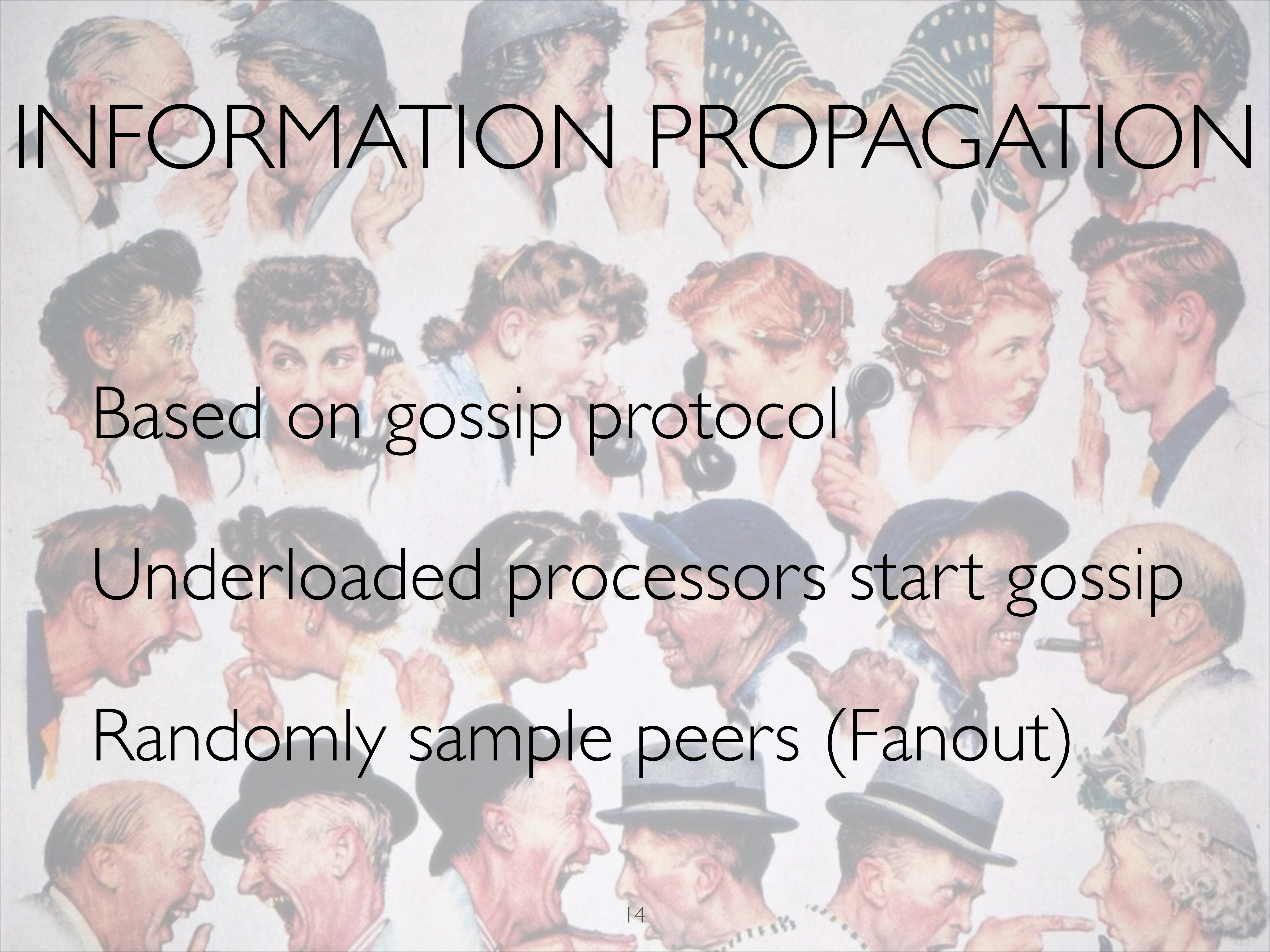
Partial information about global state

Scalable and good quality

GRAPEVINE-LB

1. Information Propagation
2. Load Transfer

INFORMATION PROPAGATION



Based on gossip protocol

Underloaded processors start gossip

Randomly sample peers (Fanout)

INFORMATION PROPAGATION

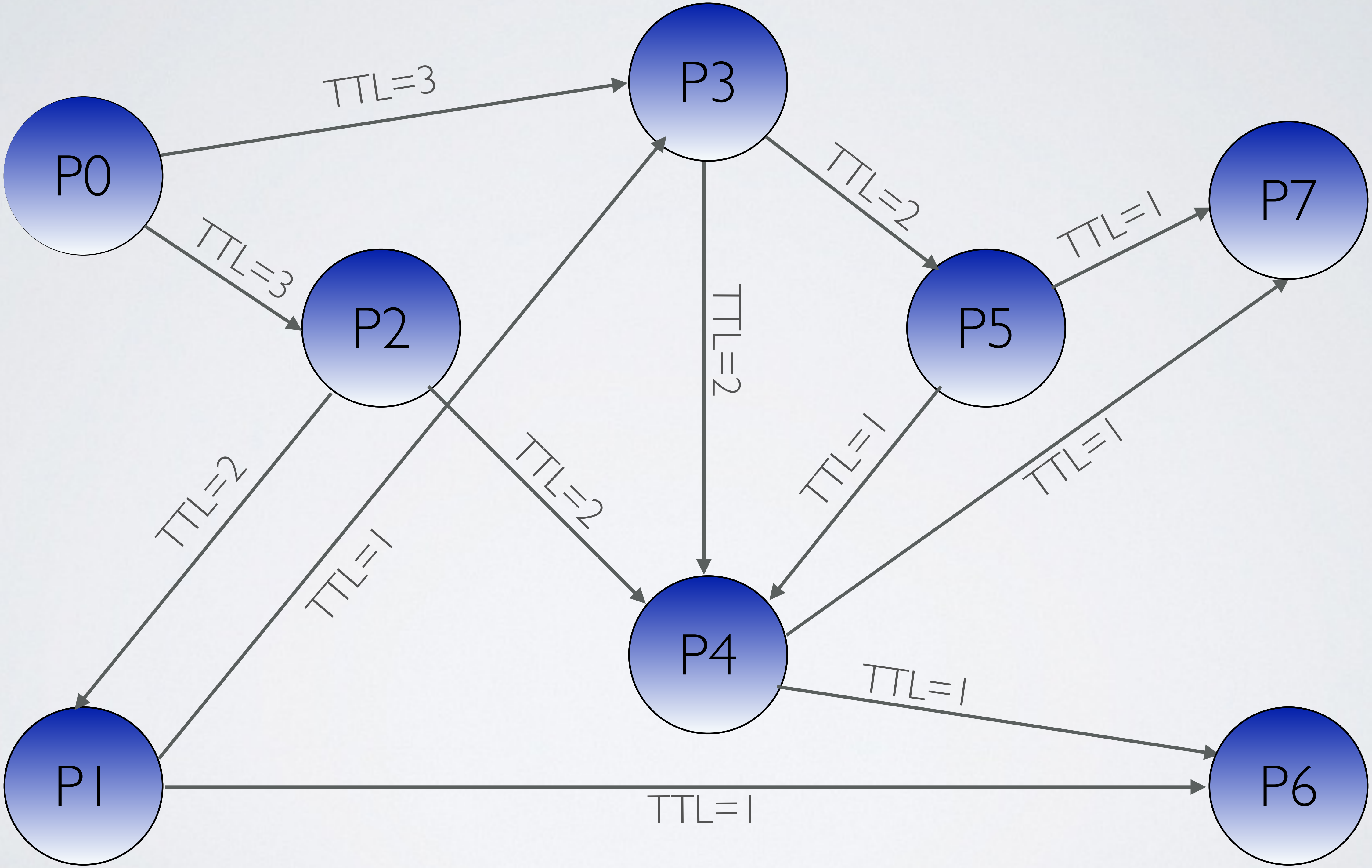
On receiving load information

- Updates its knowledge
- Forwards to random peers

No explicit synchronization

- TTL (Time To Live)

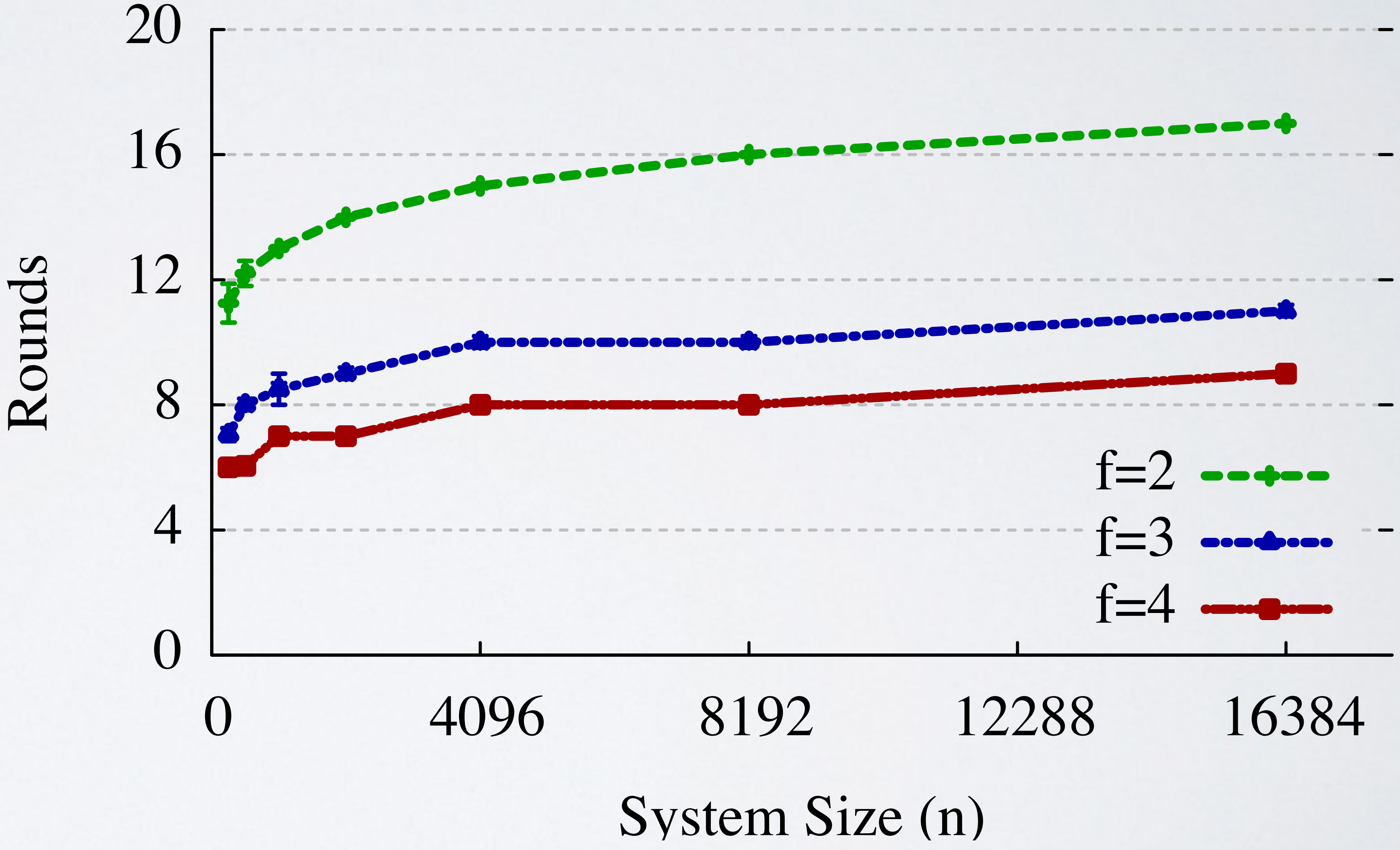
8 Processors
Fanout 2
TTL 3



INFORMATION PROPAGATION

Number of rounds taken to propagate single update

$$r = O(\log_f n)$$



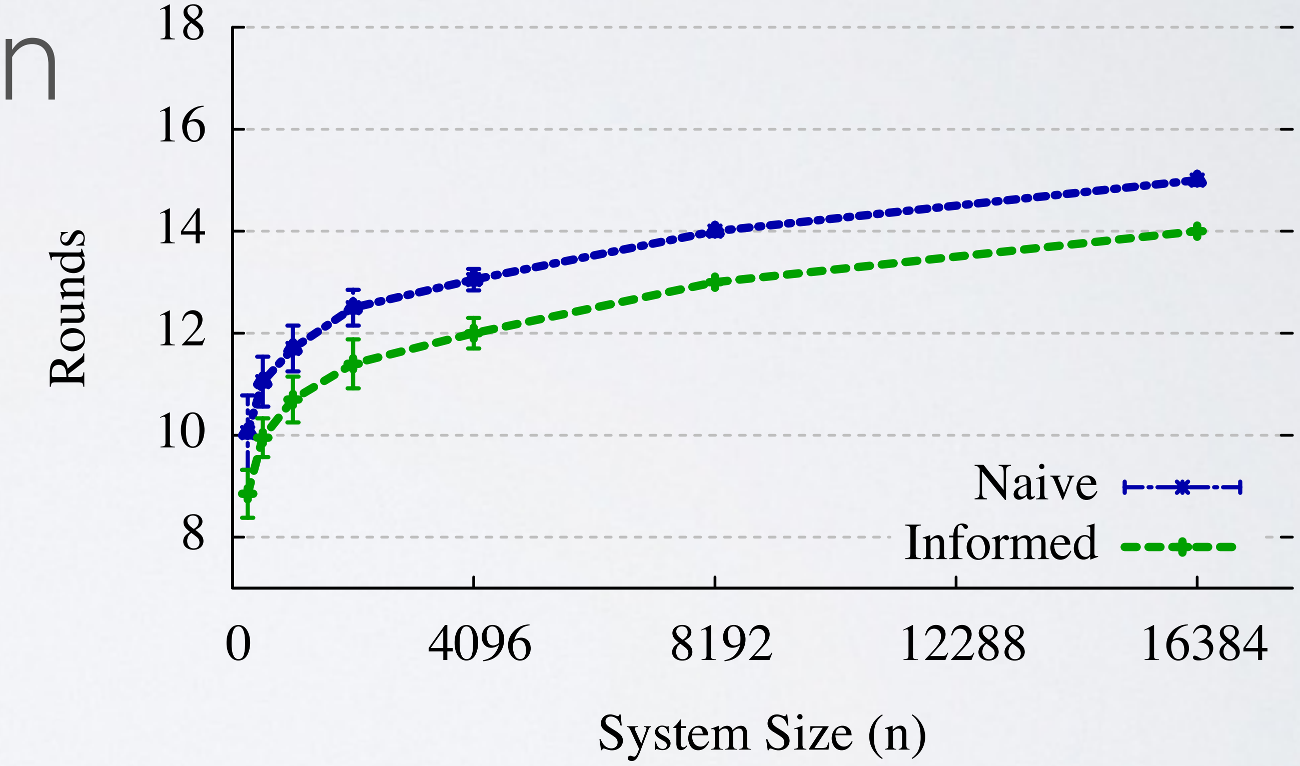
INFORMATION PROPAGATION

Naïve

- Random selection

Informed

- Biased selection
- Incorporate current knowledge



LOAD TRANSFER

Distributed

Naïve transfer

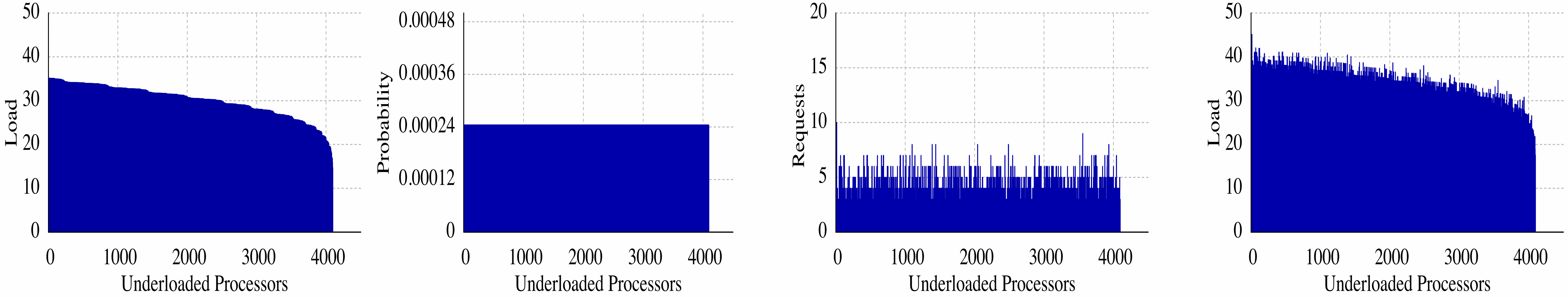
- Select processors uniformly at random
- Transfer load until load below threshold

Informed transfer

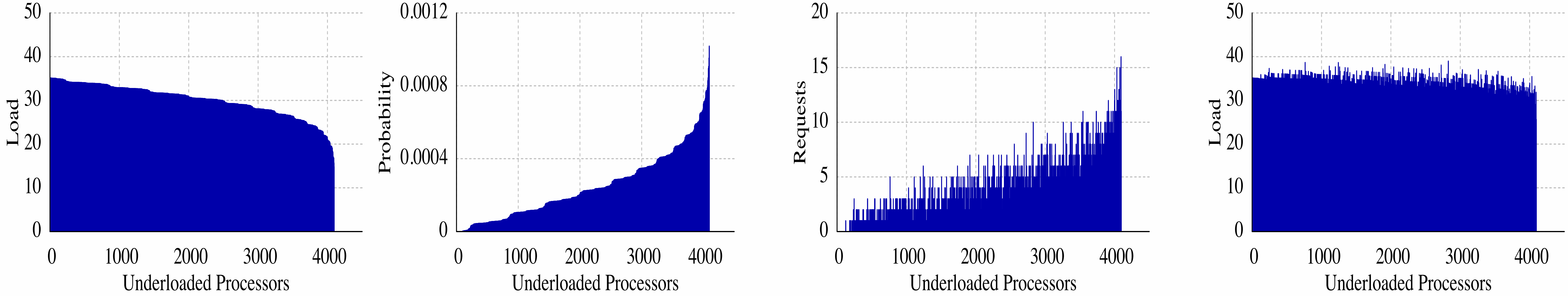
- Select processors based on load

$$p_i = \frac{1}{z} \times \left(1 - \frac{L_i}{L_{avg}}\right)$$

Naïve transfer



Informed transfer



Initial load **Probabilities assigned** **Work transferred** **Final load**

QUALITY OF LOAD BALANCE

Partial information sufficient

Tunable using TTL

TUNABLE PARAMETERS

TTL (Time To Live)

Fanout

Imbalance threshold

BACKGROUND

Application over-decomposed

Load balancer invoked periodically

Using Charm++ load balancing framework

Load balancing framework

- Instruments
- Collects statistics
- Migrates objects

EVALUATION

Applications

LeanMD (Strong scaling)

Adaptive Mesh Refinement (Strong scaling)

Machine:

IBM BG/Q, Mira

Comparison

GreedyLB, AmrLB, HierarchicalLB, DiffusionLB

METRICS

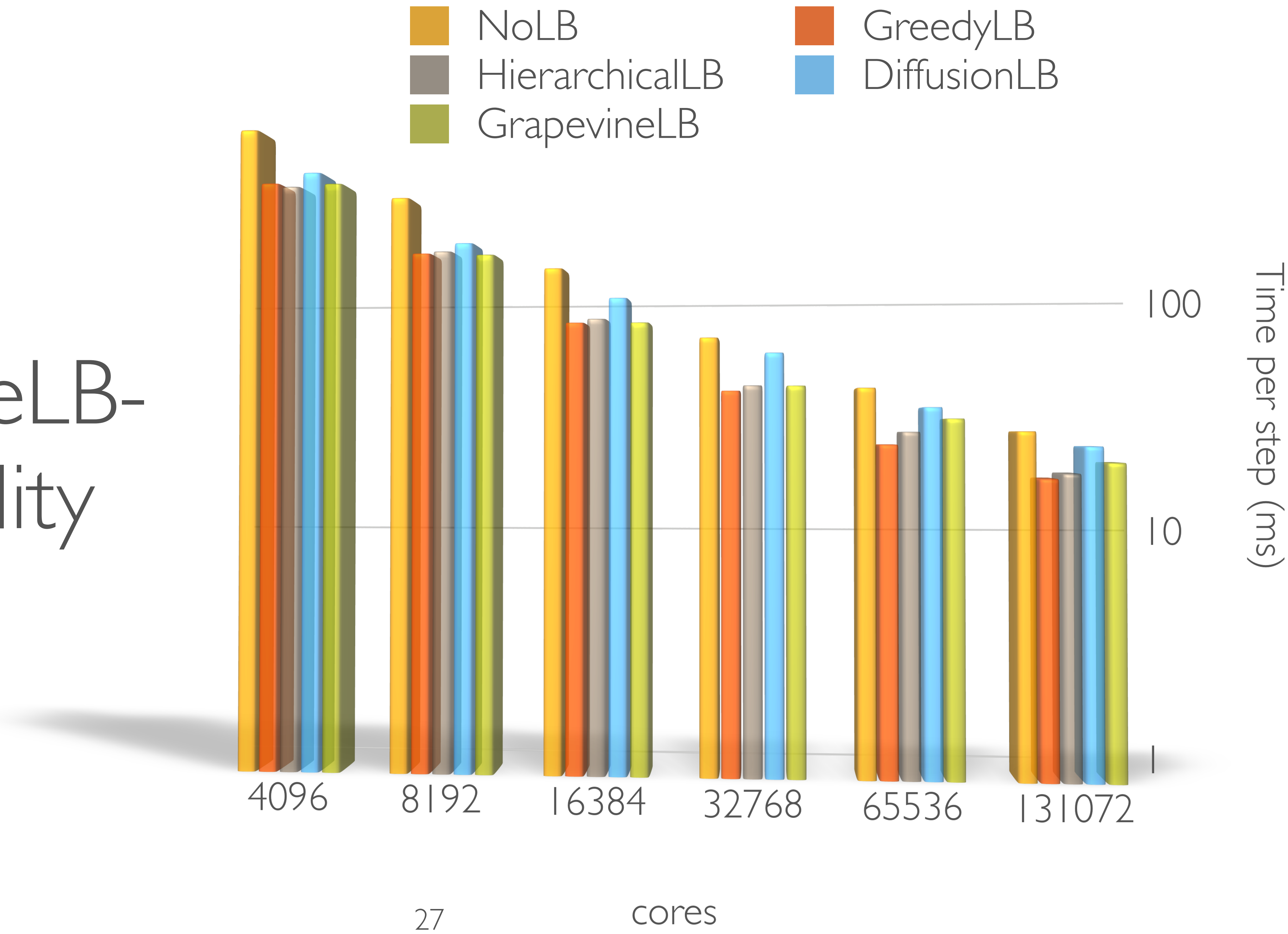
Time per step excluding LB time

Load balancing overhead

Total application time

LEANMD:TIME PER STEP

GrapevineLB- good quality



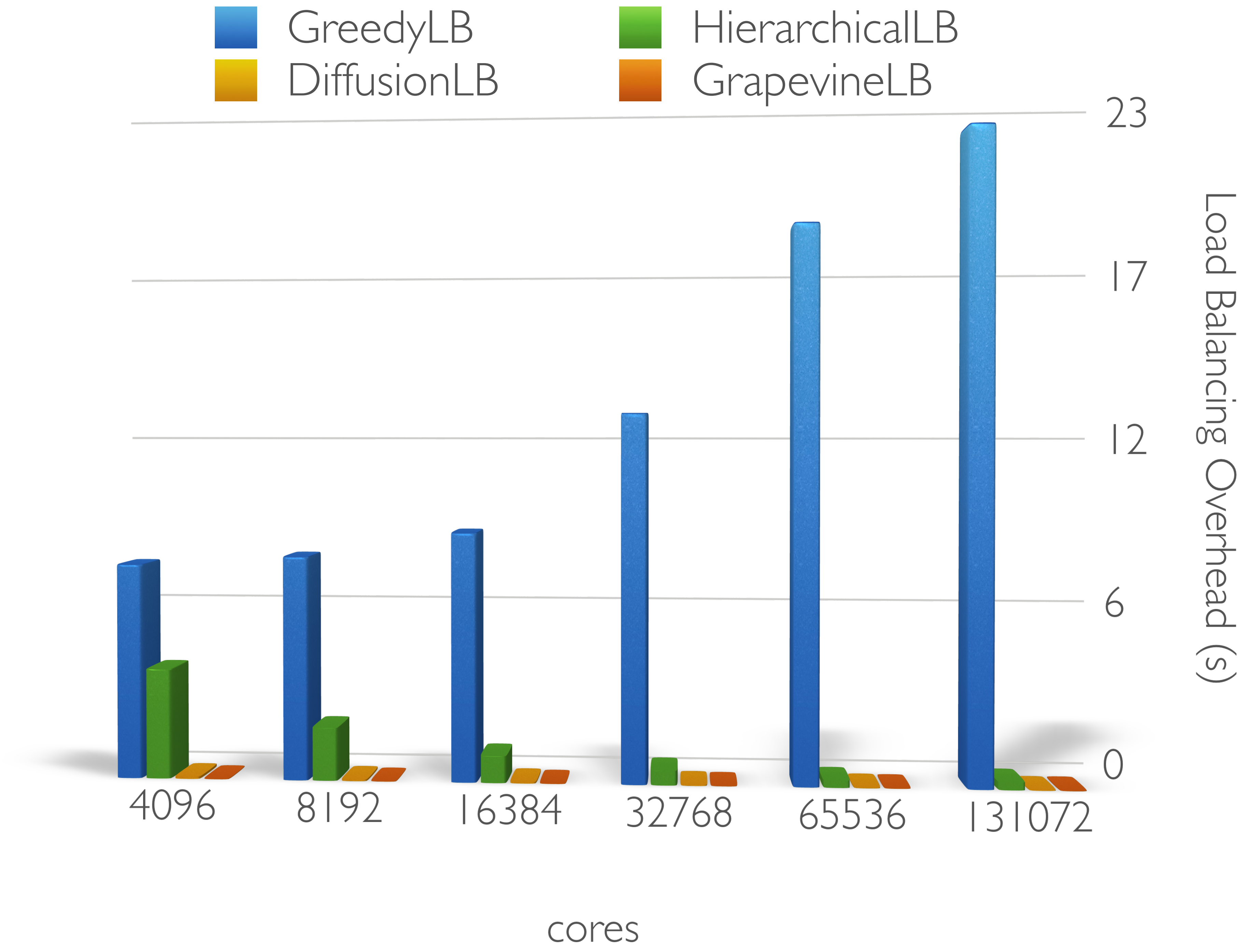
LEANMD: LB OVERHEAD

Centralized LB

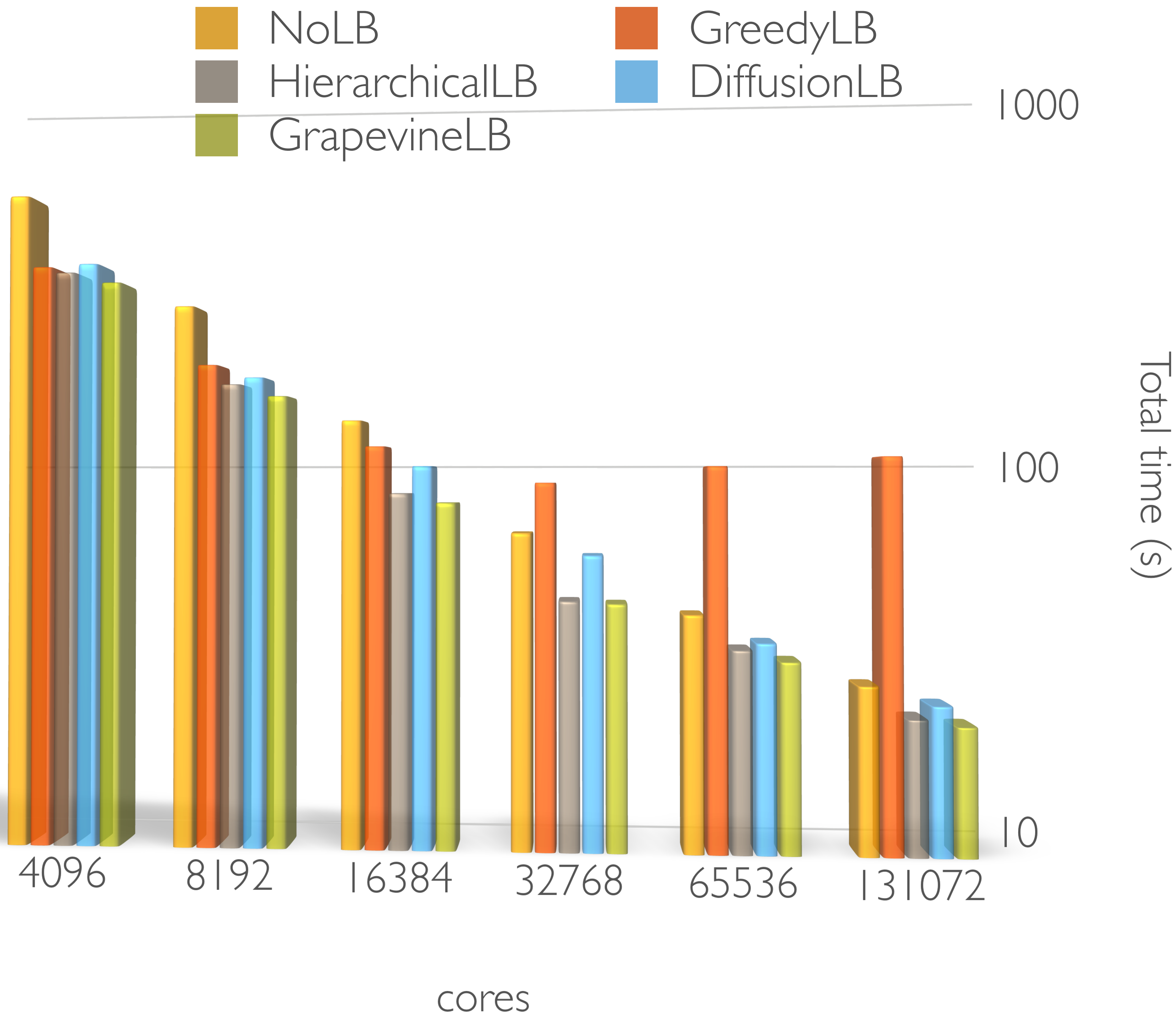
-high overhead

Distributed LBs

-low overhead



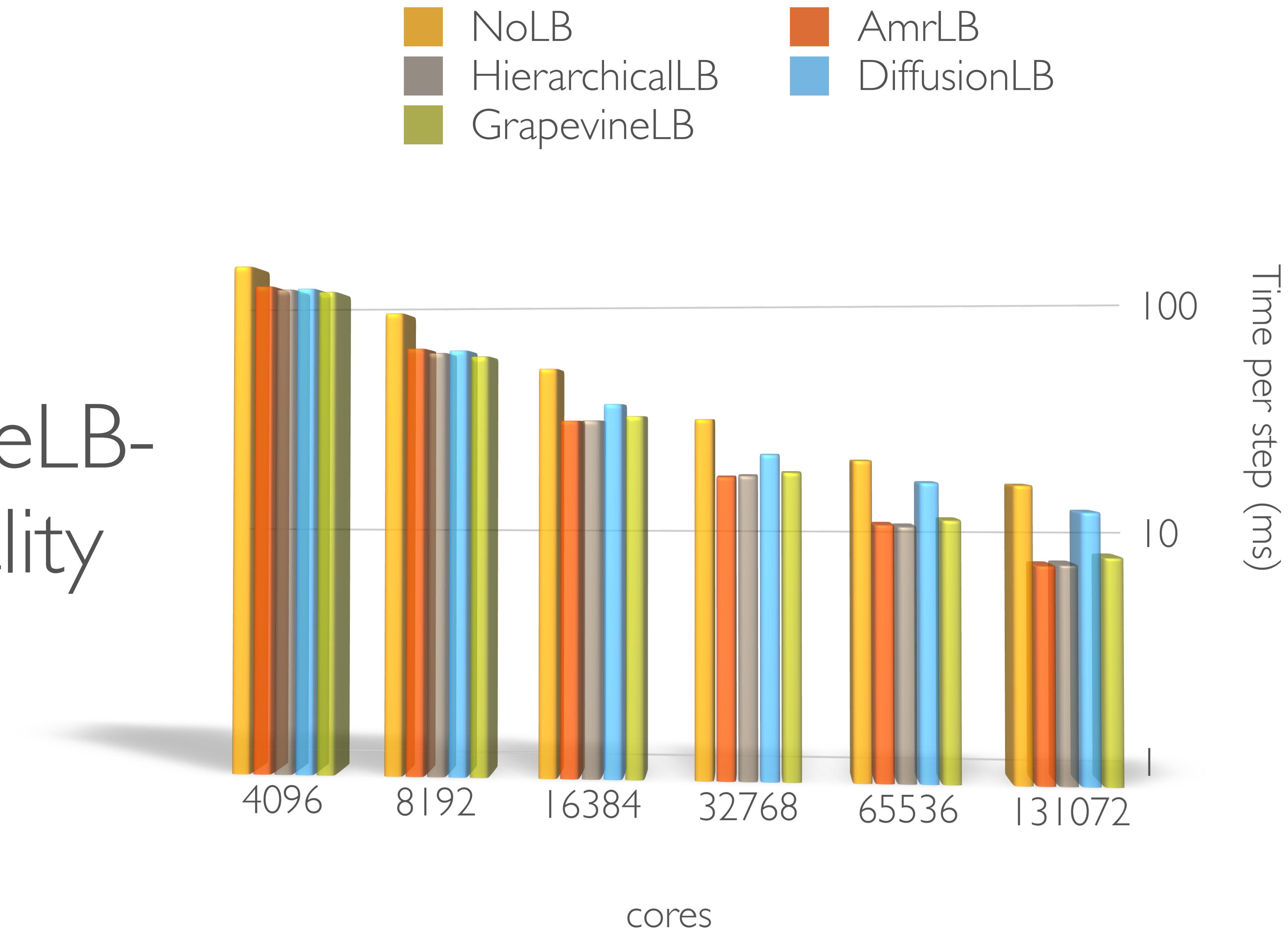
LEANMD: TOTAL TIME



Centralized-overhead exceeds benefit

GrapevineLB gives best performance

AMR: TIME PER STEP



GrapevineLB- good quality

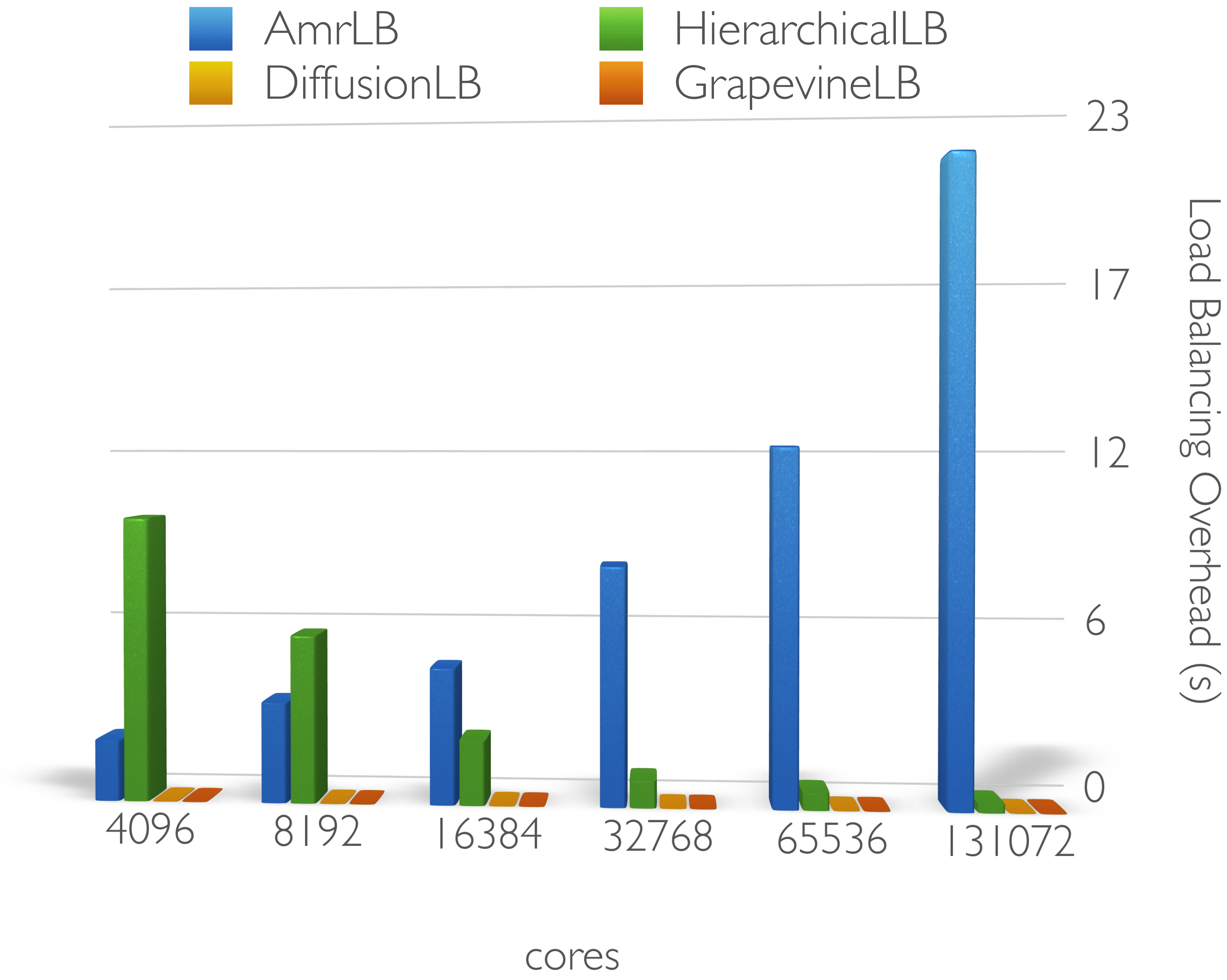
AMR - LB OVERHEAD

Centralized LB

-high overhead

Distributed LBs

-low overhead

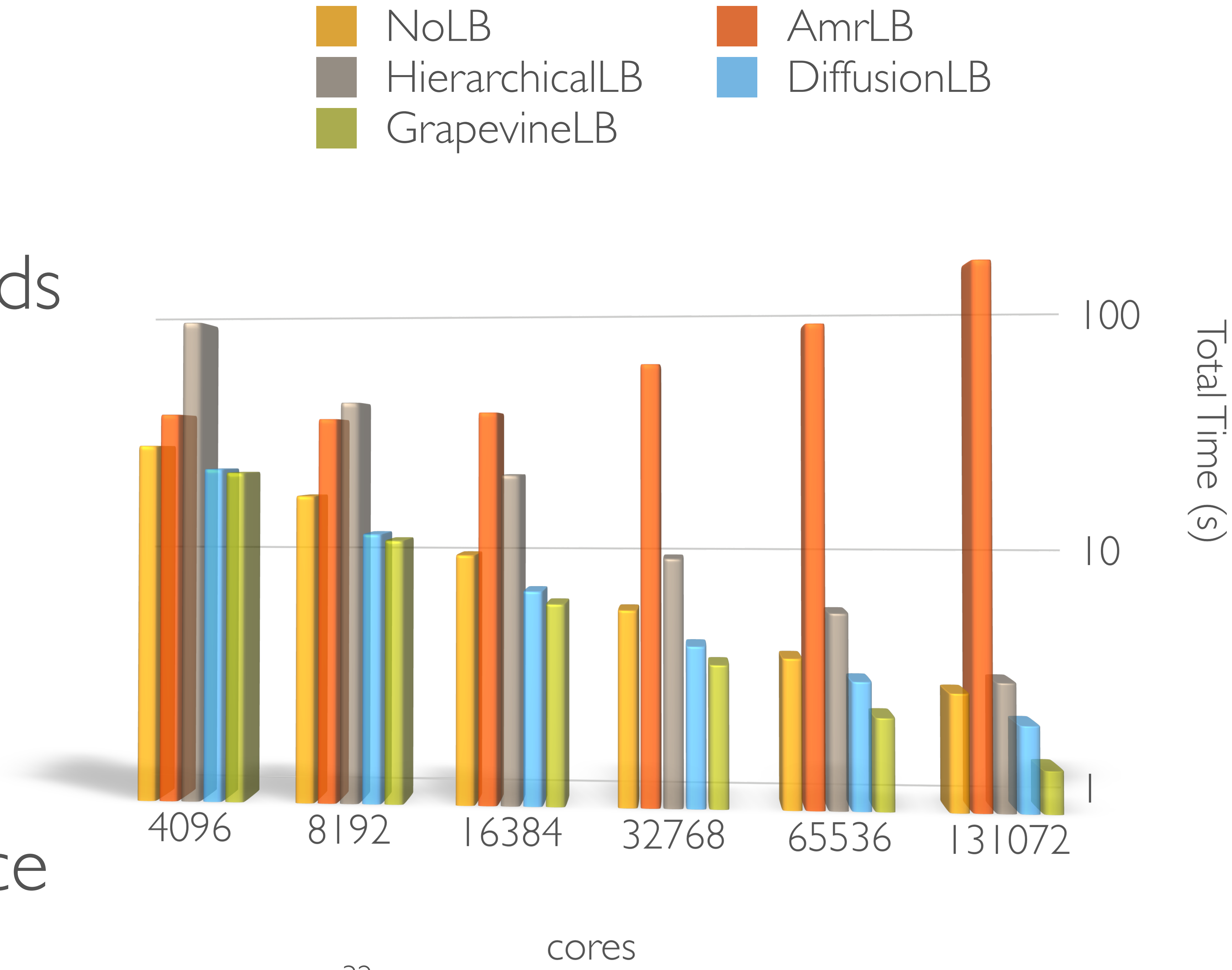


AMR- TOTAL TIME

Centralized, Hierarchical- overhead exceeds benefit

DiffusionLB- marginal benefit

GrapevineLB- best performance



SUMMARY

Simple strategy

Good quality with less overhead

Tunable

ACKNOWLEDGEMENTS

PPL group

Funded by:

DOE grant DOE DE-SC0001845 and by NSF ITR

-HECURA-0833188

THANK YOU!

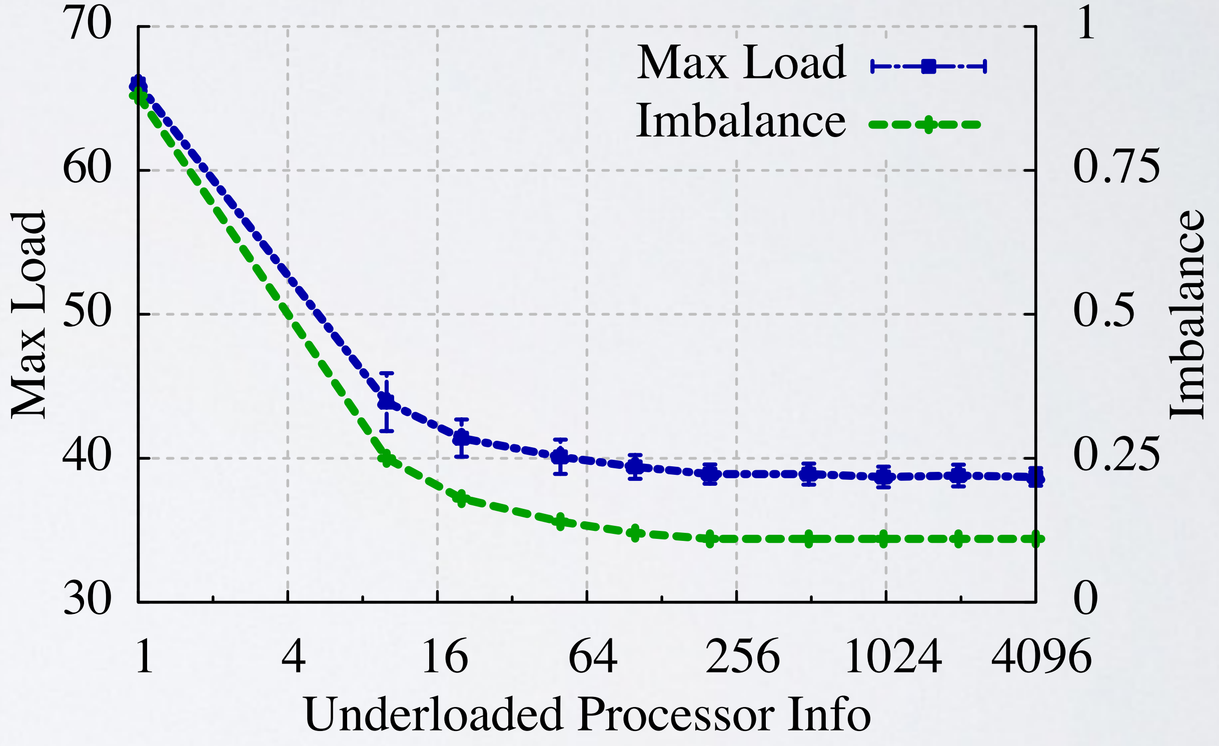
LEANMD: TOTAL TIME

	4096	8192	16384	32768	65536	131072
NoLB	519.19	263.30	131.56	67.18	41.49	27.20
DiffuseLB	342.15	170.41	99.67	58.47	34.91	24.29
GreedyLB	336.34	184.09	112.23	90.19	99.51	105.35
HierarchLB	325.00	163.65	84.62	44.56	33.49	22.43
GrapevineLB	305.20	152.21	79.94	43.88	31.3	21.53

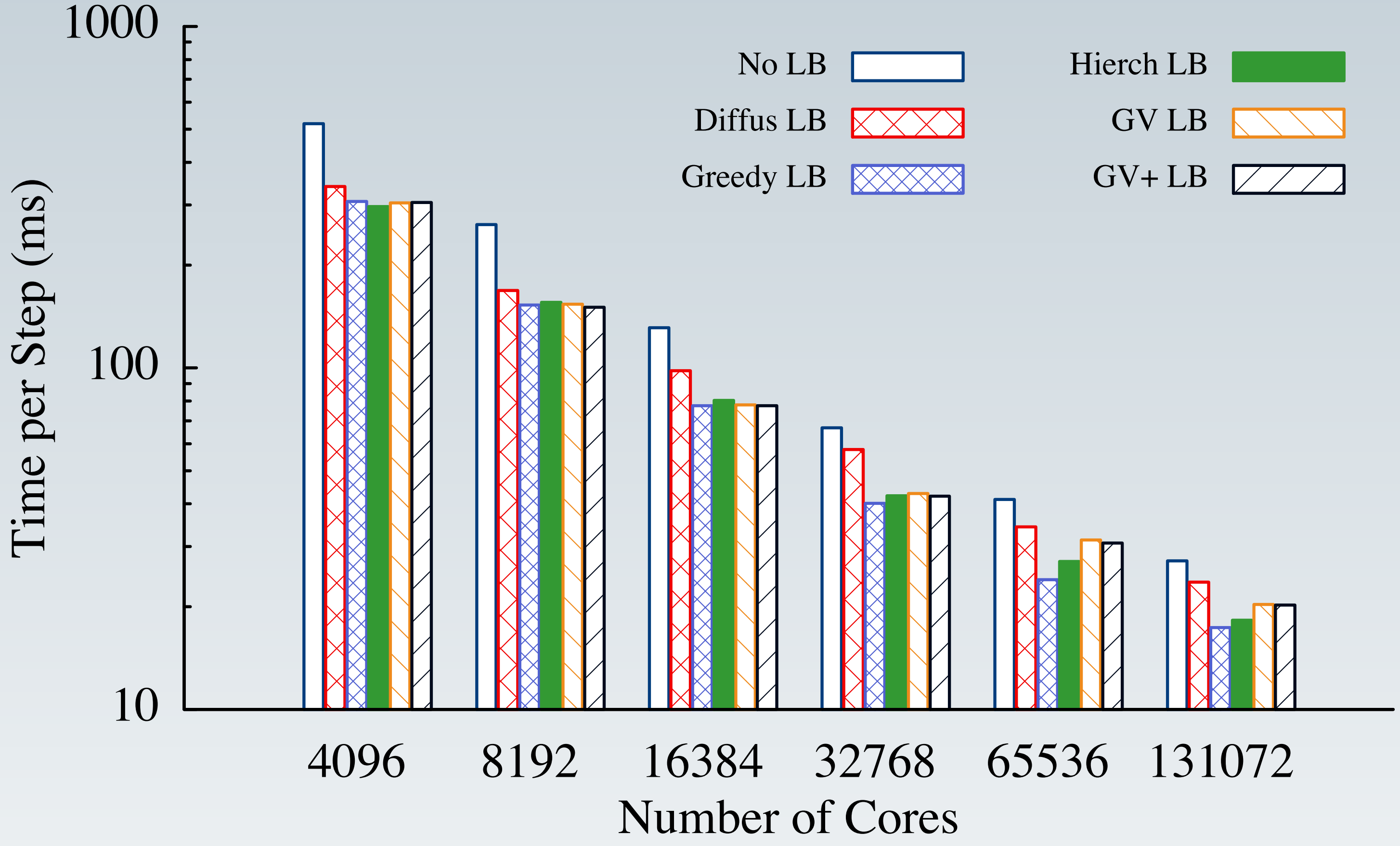
QUALITY OF LOAD BALANCE

- Quality metric

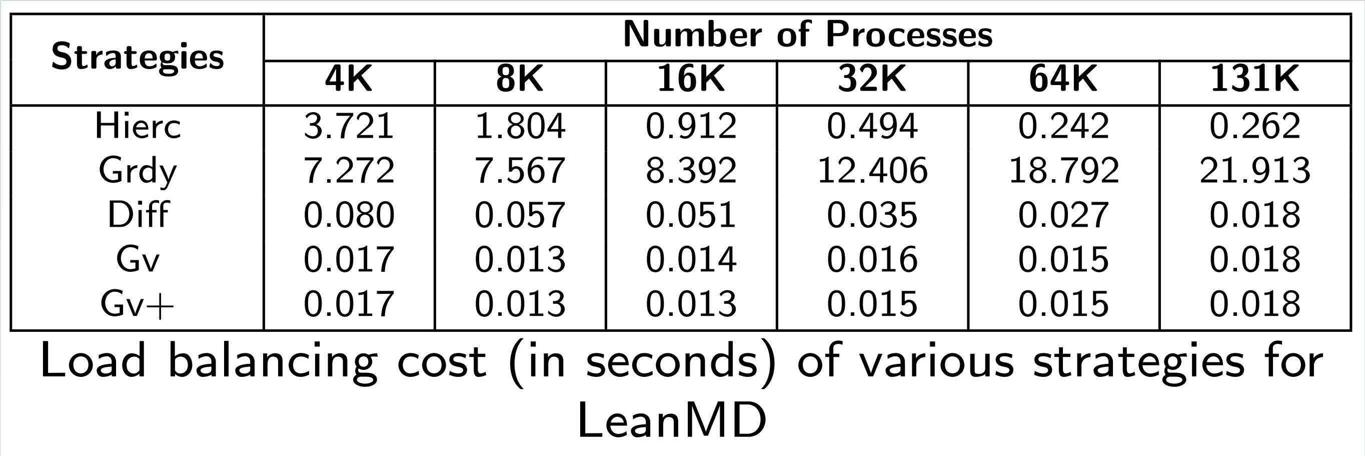
$I = L_{max}/L_{avg} - 1$



LEANMD-TIME PER STEP



LB OVERHEAD



Strategies	Number of Processes					
	4K	8K	16K	32K	64K	131K
Hierc	3.721	1.804	0.912	0.494	0.242	0.262
Grdy	7.272	7.567	8.392	12.406	18.792	21.913
Diff	0.080	0.057	0.051	0.035	0.027	0.018
Gv	0.017	0.013	0.014	0.016	0.015	0.018
Gv+	0.017	0.013	0.013	0.015	0.015	0.018

Load balancing cost (in seconds) of various strategies for LeanMD