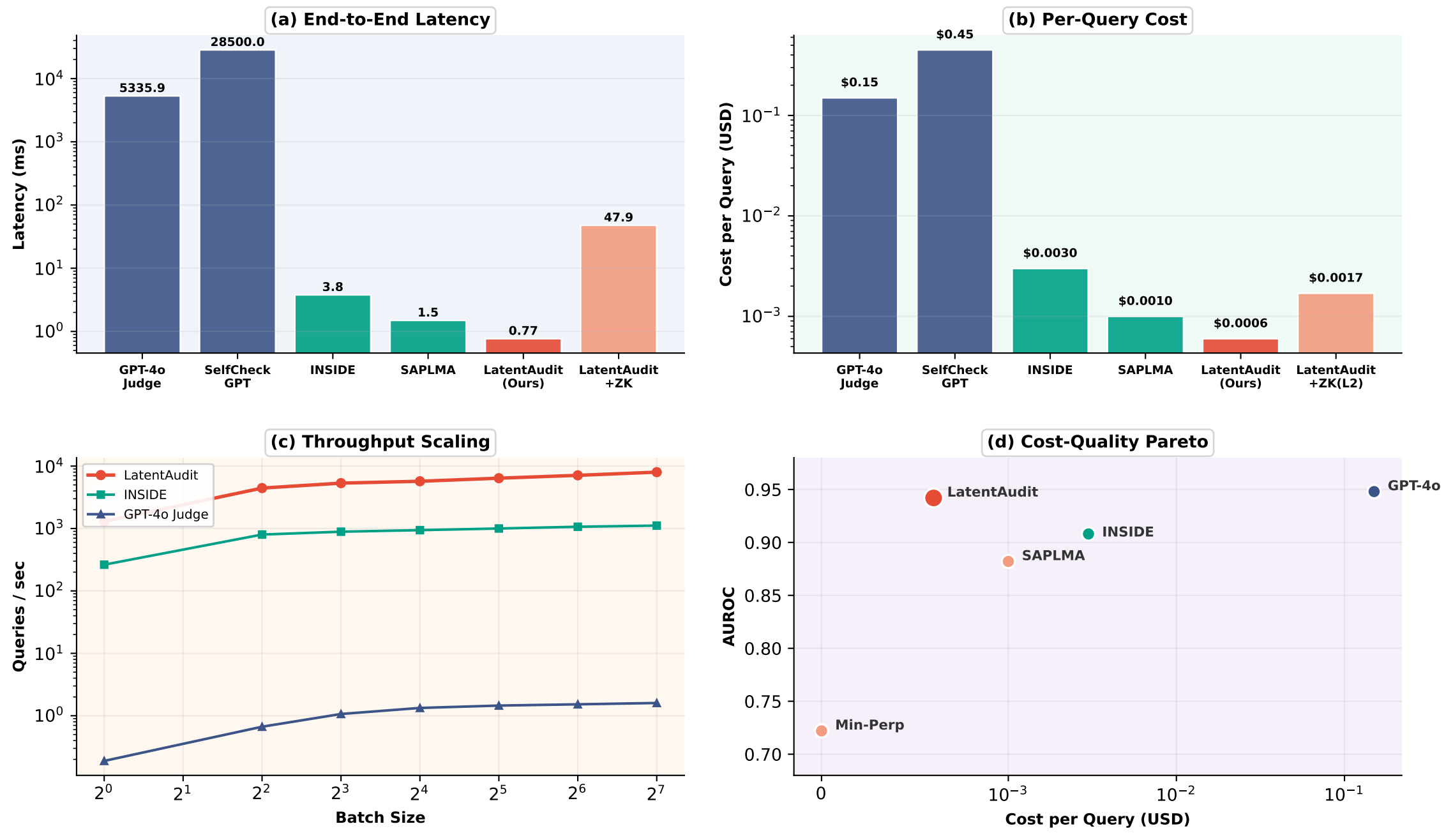
(a) End-to-End Latency
Latency (ms)
5335.9
28500.0
3.8
1.5
0.77
47.9
GPT-4o Judge
SelfCheck GPT
INSIDE
SAPLMA
LatentAudit (Ours)
LatentAudit +ZK
(b) Per-Query Cost
Cost per Query (USD)
$0.15
$0.45
$0.0030
$0.0010
$0.0006
$0.0017
GPT-4o Judge
SelfCheck GPT
INSIDE
SAPLMA
LatentAudit (Ours)
LatentAudit +ZK(L2)
(c) Throughput Scaling
Queries / sec
LatentAudit
INSIDE
GPT-4o Judge
Batch Size
(d) Cost-Quality Pareto
AUROC
LatentAudit
INSIDE
SAPLMA
GPT-4o
Min-Perp
Cost per Query (USD)