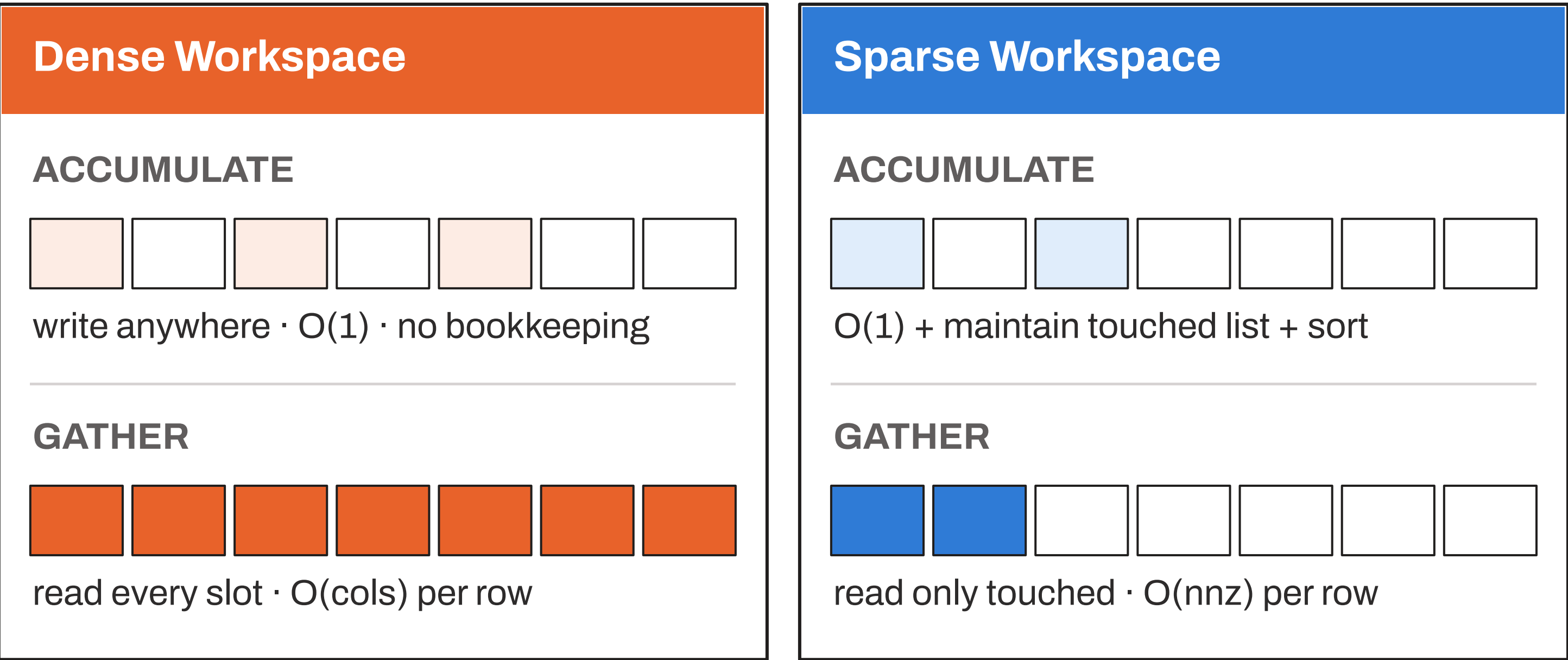


Data-Driven Kernel Selection for Sparse Tensors Without Autotuning

Linus Sheaff*, Shideh Hashemian, Michael O’Boyle & Jackson Woodruff School of Informatics, University of Edinburgh

01 Kernel Selection Challenges

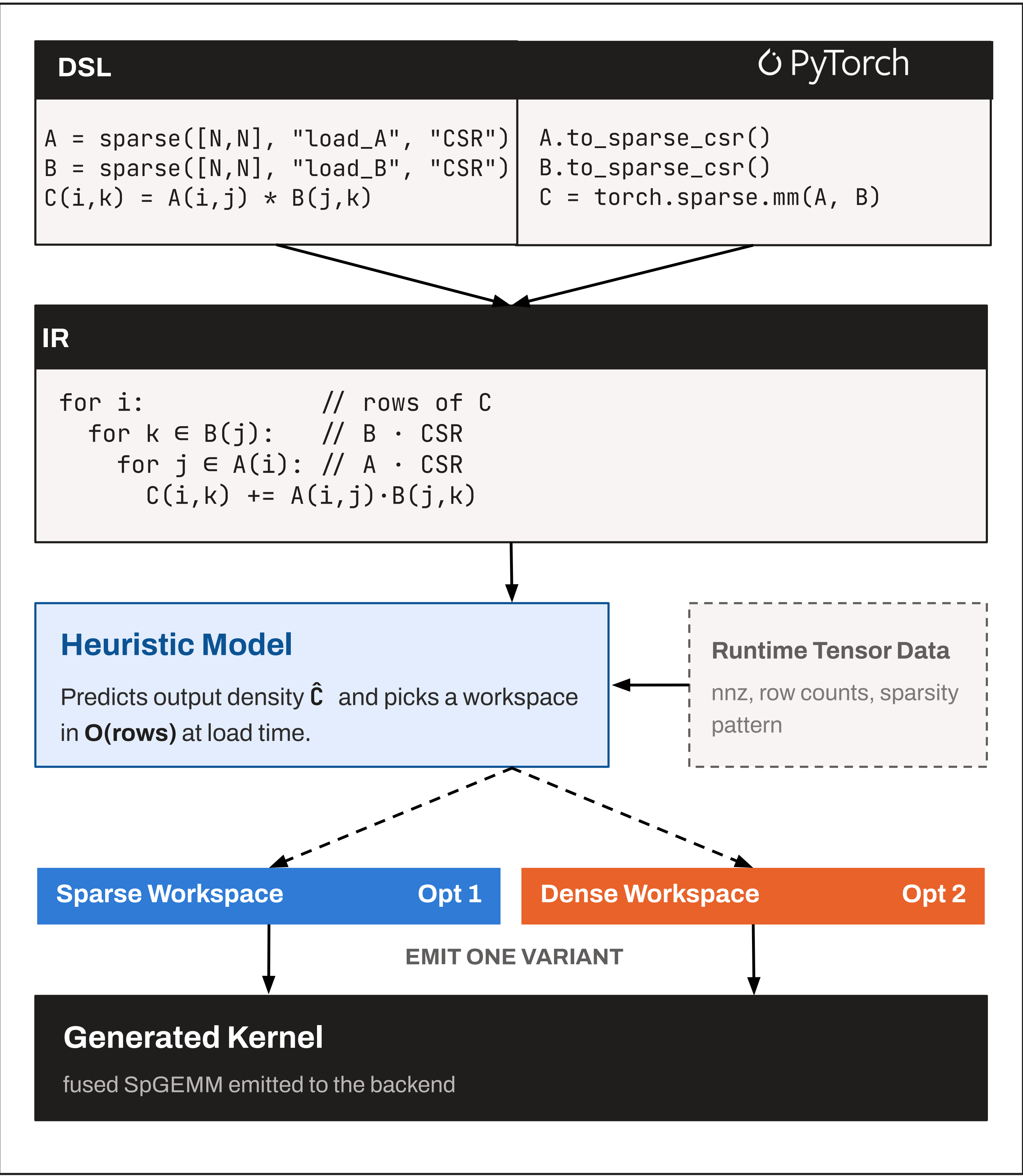
The fastest kernel depends on runtime tensor structure
As contributions arrive in arbitrary column order, results can't be emitted directly in sorted order so they must be accumulated in a workspace first.



Dense pays per column. Sparse pays per nonzero plus bookkeeping and a sort. The choice hinges on **predicted output structure**, estimated in $O(\text{rows})$ before executing.

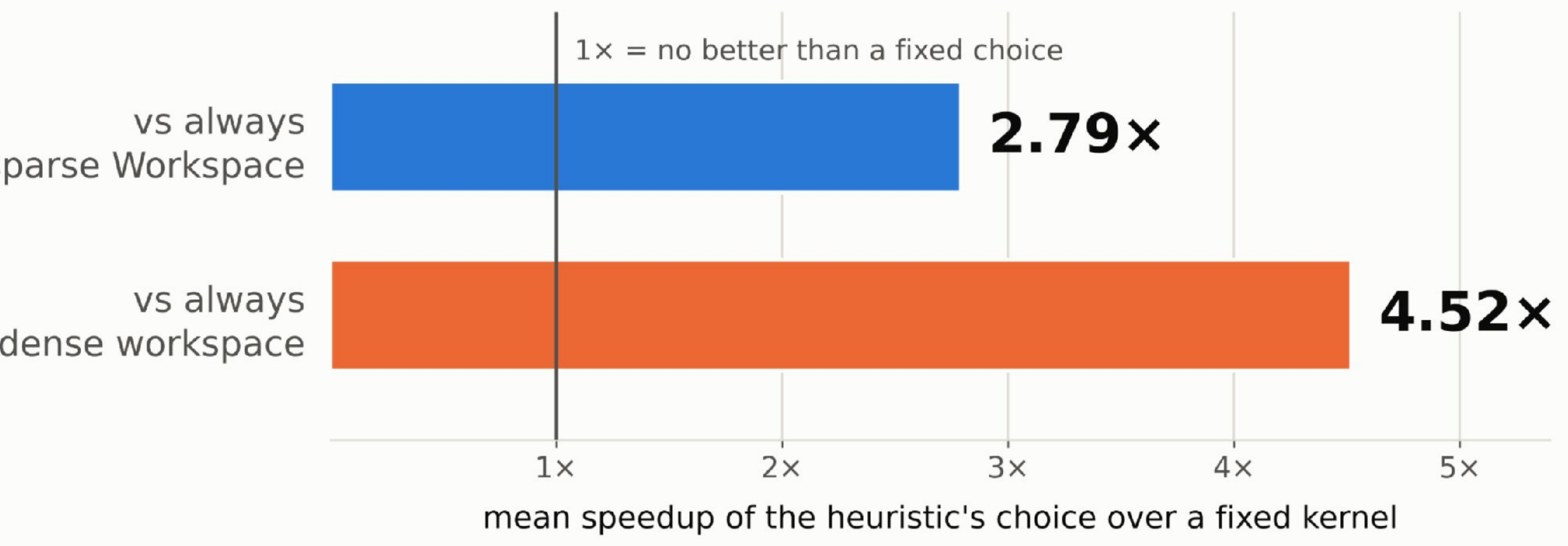
03 Selecting the Fastest Kernel Variant at Runtime

We estimate the fastest kernel variant with simple, low-overhead signals.
No autotuning, no recompilation.

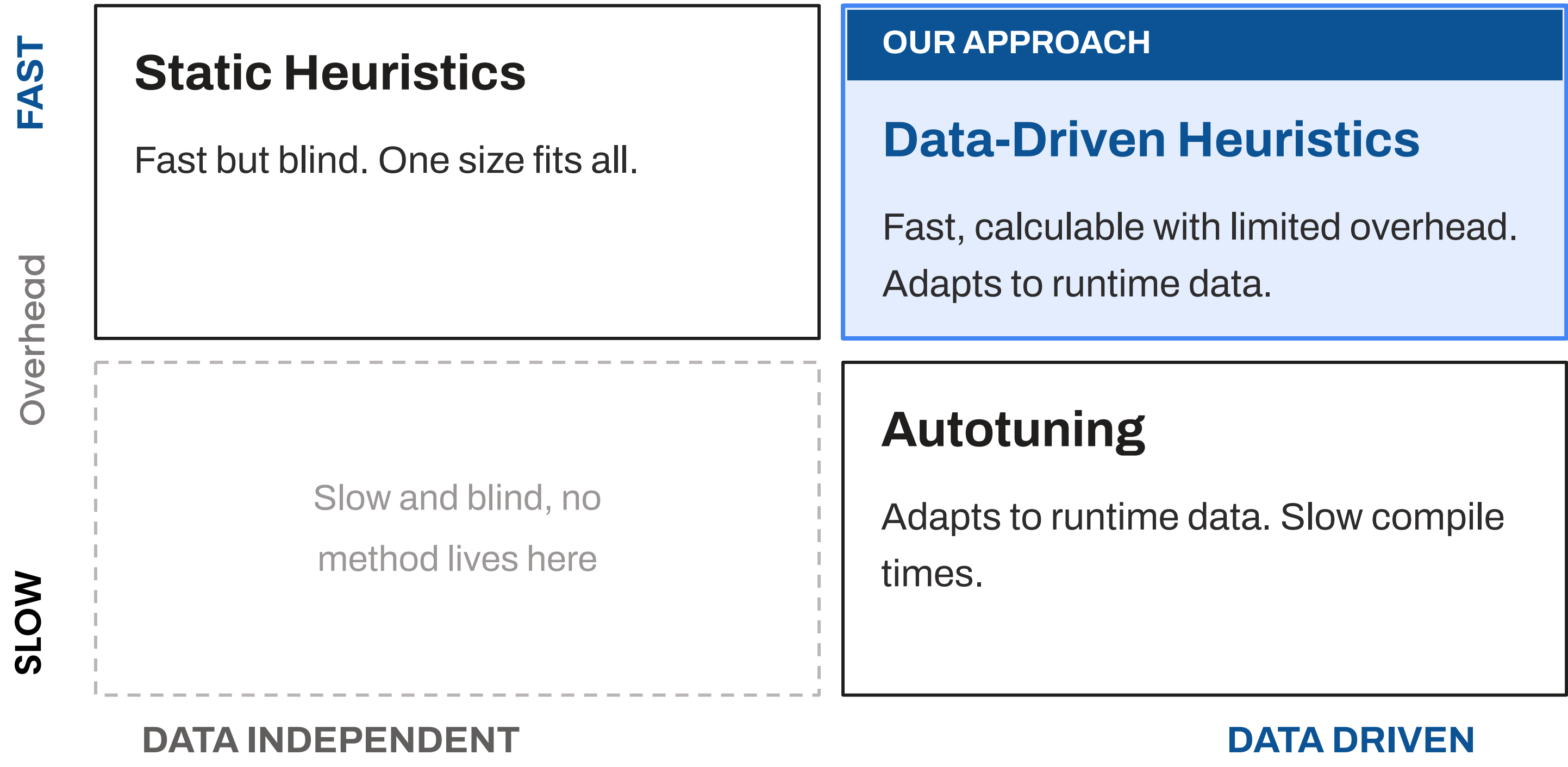


06 Heuristic Impact

Choosing the fastest kernel variant yields substantial speedups over naively applying one.

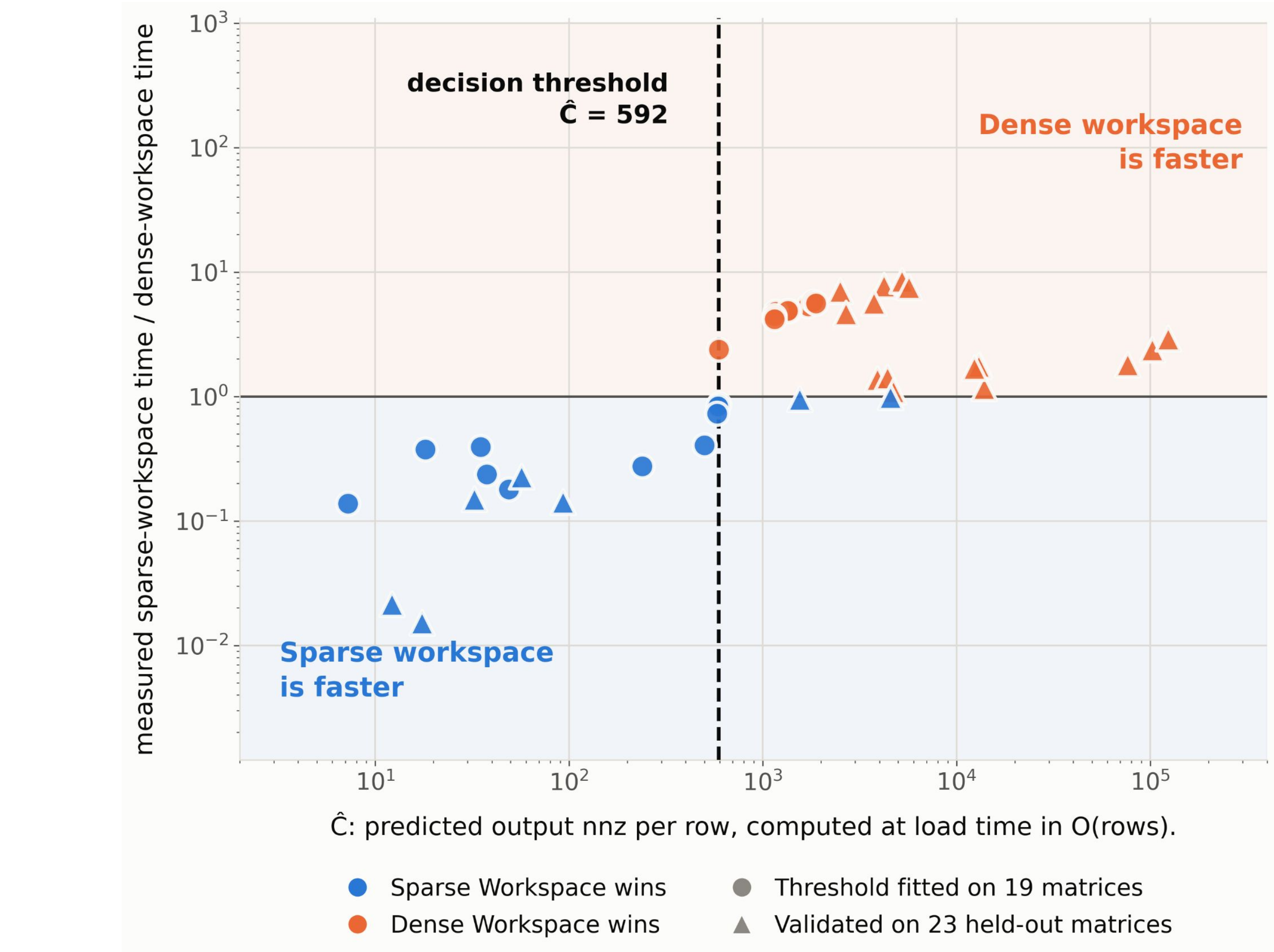


02 Data-Driven Heuristics are Fast & Adaptive



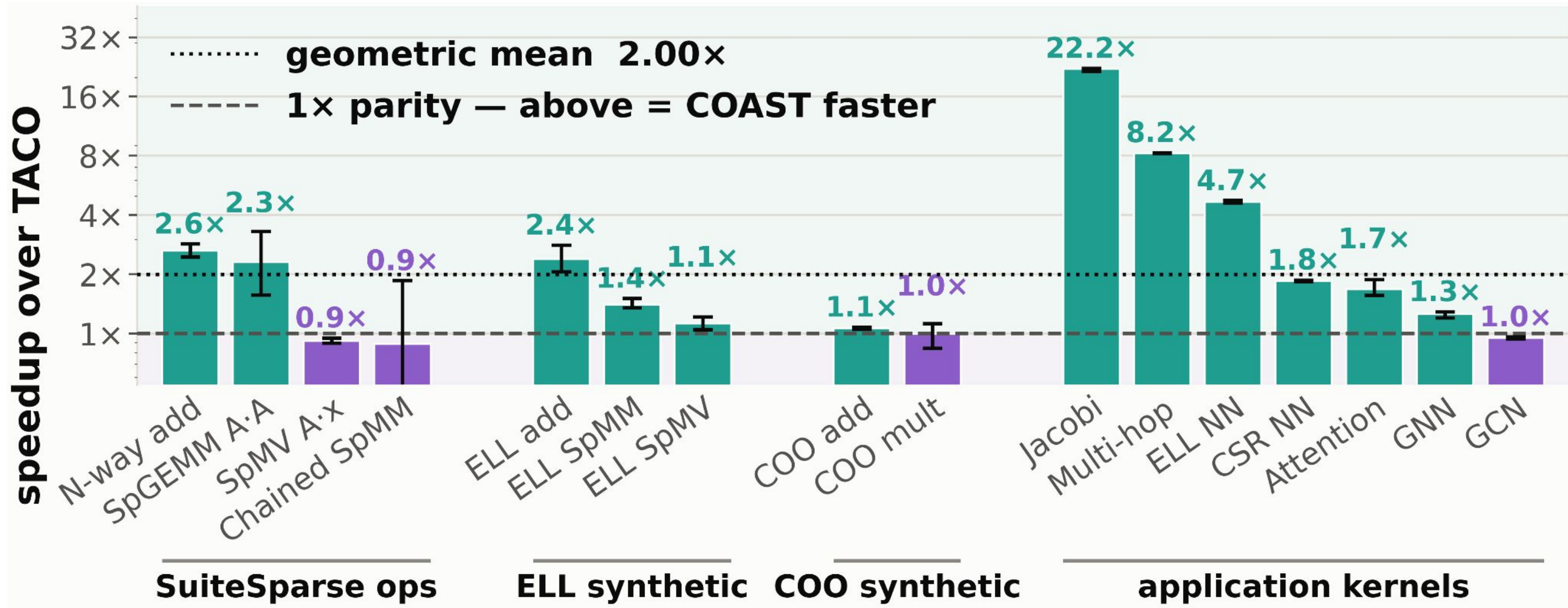
04 Fit the Heuristic Model

Our model is only 1.01× slower than an oracle that always chooses the correct program.



05 Compiler Performance

We compare our tool against TACO across a range of benchmarking kernels. We demonstrate a geometric mean **speedup of 2.00×**



07 Further Work

Dataflow analysis that predicts intermediate sparsity, thus extending near-oracle kernel selection across the whole expression DAG.