Leveraging Future Technology for Energy Efficient Databases
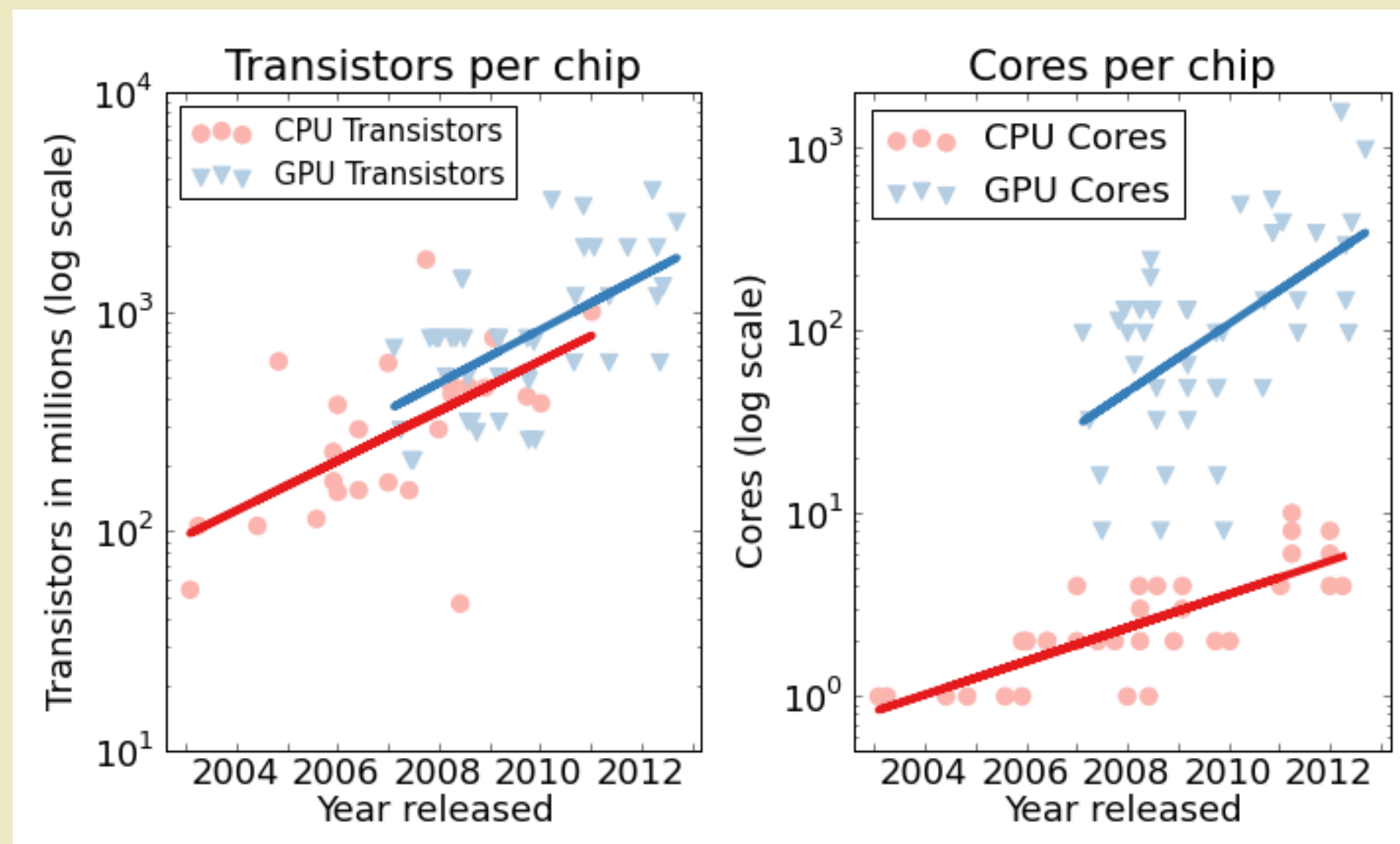
Jason Power, Yinan Li, Mark Hill, Jignesh Patel, David Wood
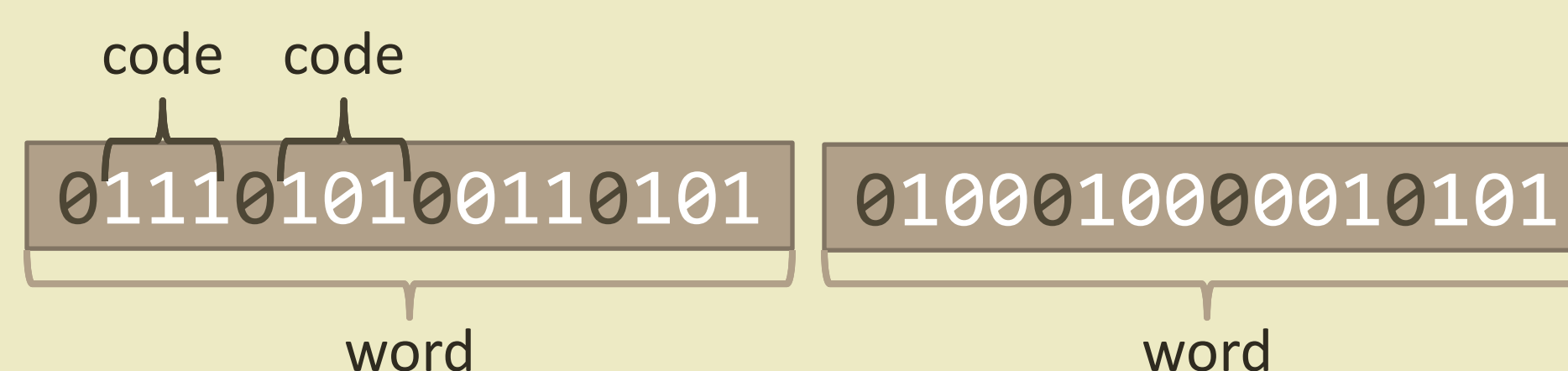
Future Technology Trends

- GPU cores scaling faster than CPU cores
- Integrated CPU-GPU chips
- Higher memory B/W (3D die-stacked DRAM)

Can we take advantage of these trends for database applications?

Background

- In-memory databases are common
  - Scan is a key operation for data warehousing
  - Use compressed data to reduce costs
- *Bitweaving*[1] leverages sub-word parallelism
  - "any size SIMD"
  - 5–25x speedup over naïve method on CPU
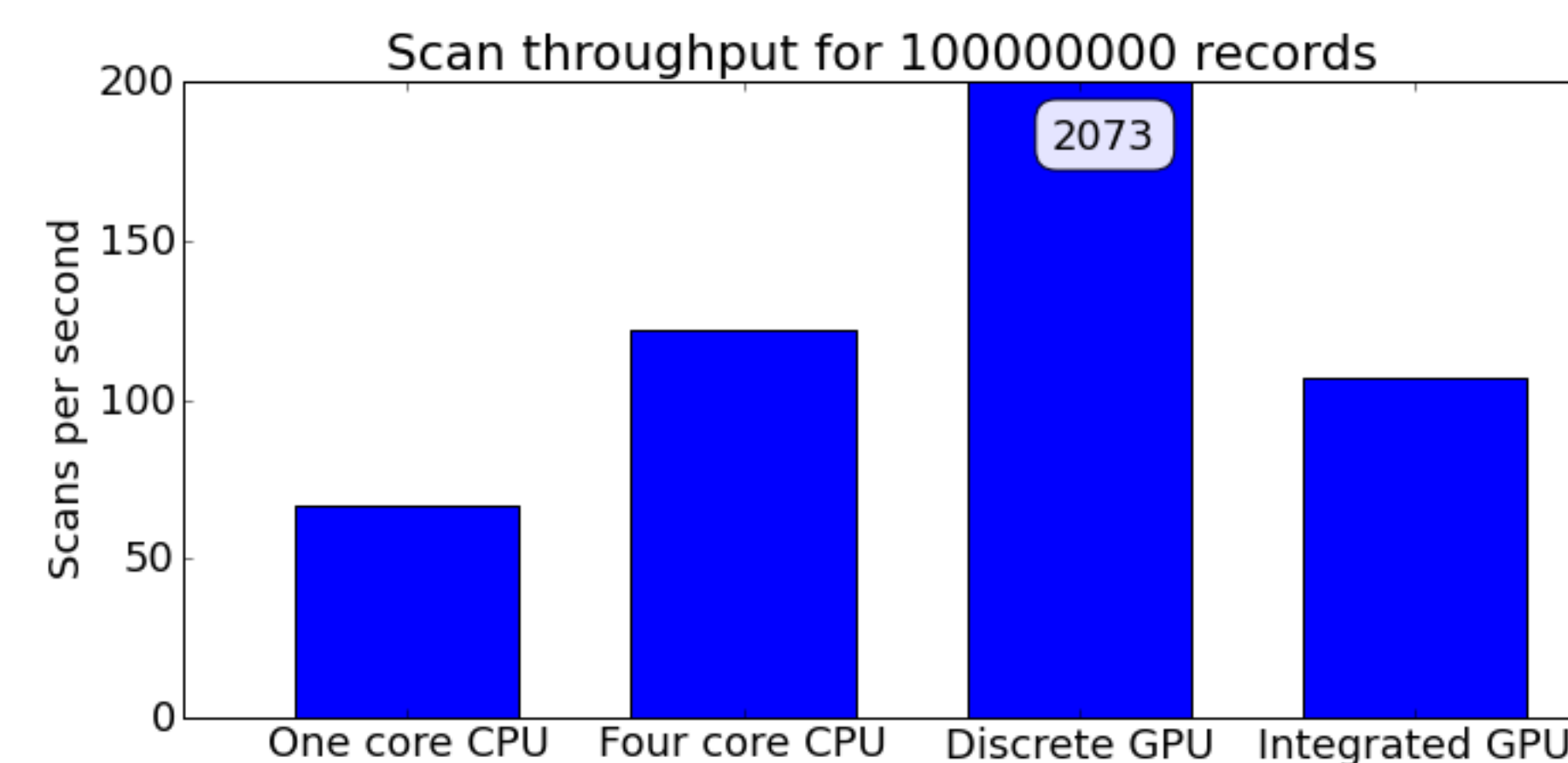- Pack many codes (records) per word

- Each CPU operation acts on many codes

[1] Li. Patel. BitWeaving: Fast Scans for Main Memory Data Processing. SIGMOD '13.
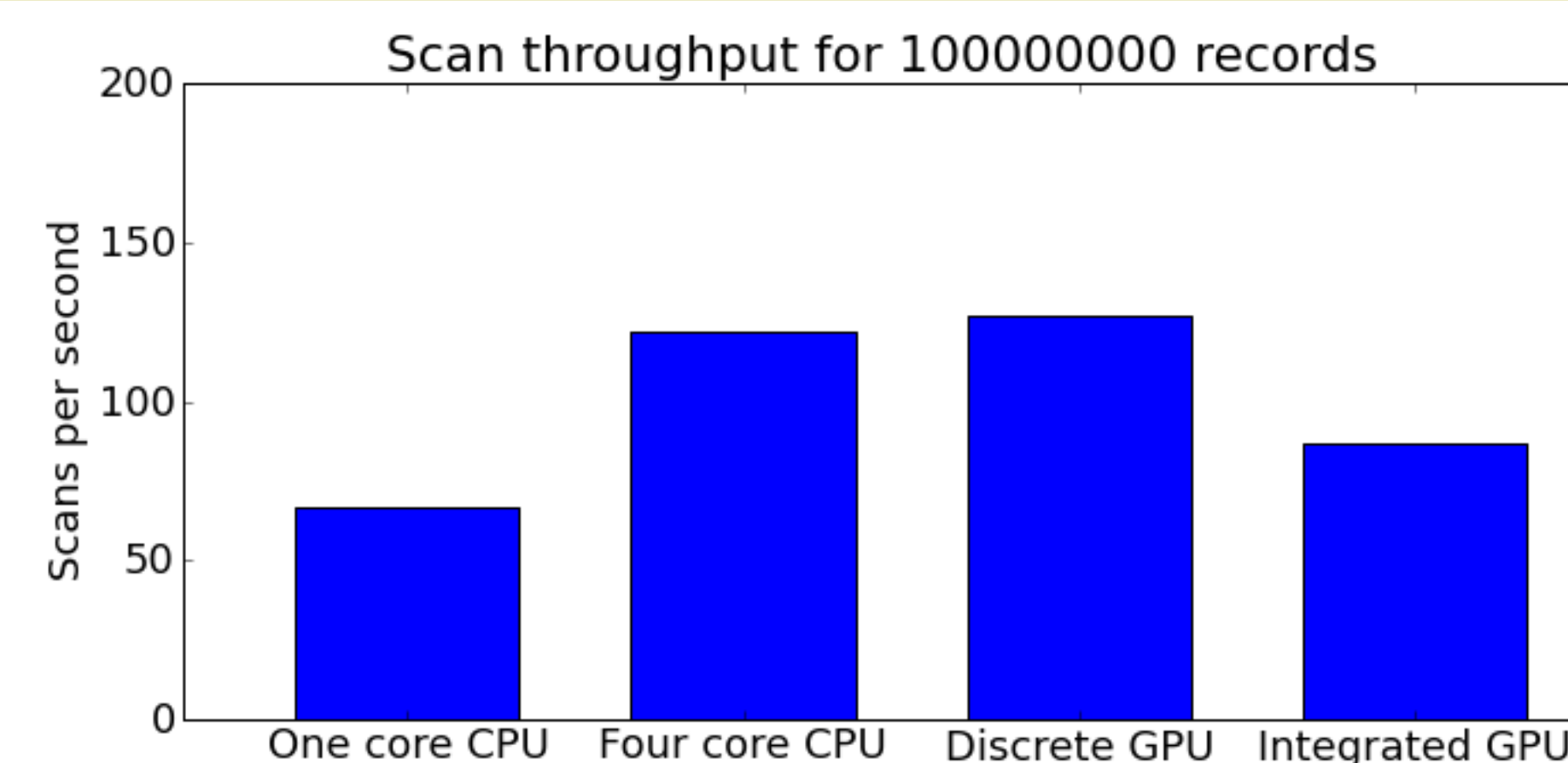
GPU Performance

GPUs have huge potential

- GPUs can provide massive speedup (17x)
- Even an integrated GPU shows speedup
- However, this has optimistic assumptions:
  - All data is resident on GPU
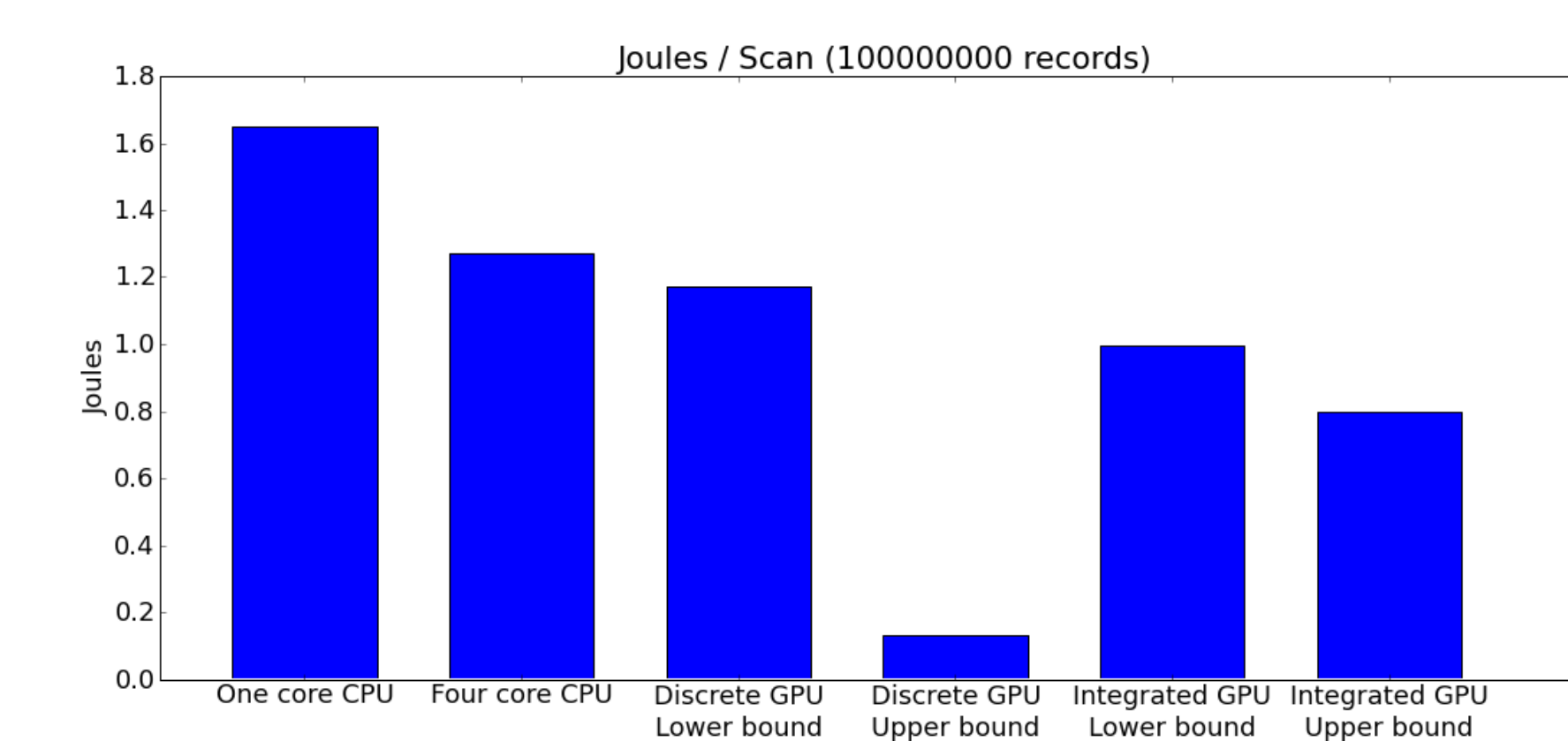 - No kernel launch overheads and the scan results are accumulated on GPU

Bottlenecks in current hardware

- Using feasible methods in today's systems
  - GPU still has higher performance
  - Overheads dominate GPU execution
- For the discrete GPU:
  - Each scan ~50 μs
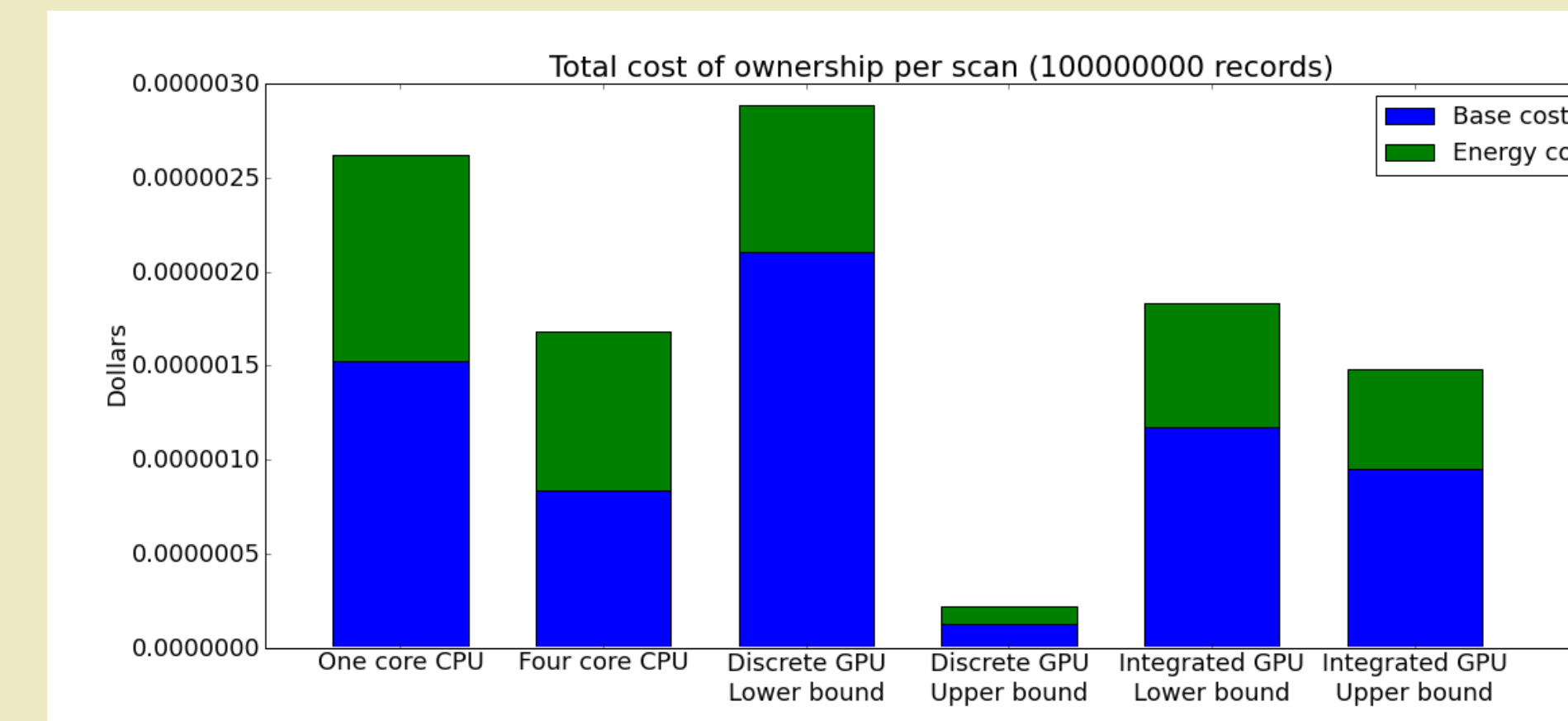 - Kernel launch &result copy ~5 ms
  - Overheads 100x more than useful work

Energy Efficiency

GPU can reduce total energy

- Current GPU hardware is energy efficient
- Bigger gap when bottlenecks are alleviated

Total cost of ownership for 3 years

- Need to reduce overheads to see TCO benefits on GPU platforms

Conclusions & Future Work

- Current GPUs can provide modest energy and performance benefits over CPUs for scan
- Future, tightly integrated GPUs will mitigate many overheads
- Next steps:
  - Integrate into more robust DB system
  - Run more realistic workloads (TPC-H)