Ambit

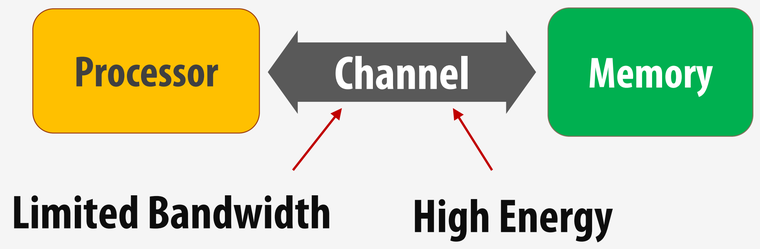
In-memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology

Vivek Seshadri Donghyuk Lee Thomas Mullins Hasan Hassan Amirali Boroumand,
Jeremie Kim Michael A. Kozuch Onur Mutlu Phillip B. Gibbons Todd C. Mowry
Carnegie Mellon University Microsoft Research India NVIDIA Research Intel ETH Zürich

Applications for Bulk Bitwise Operations

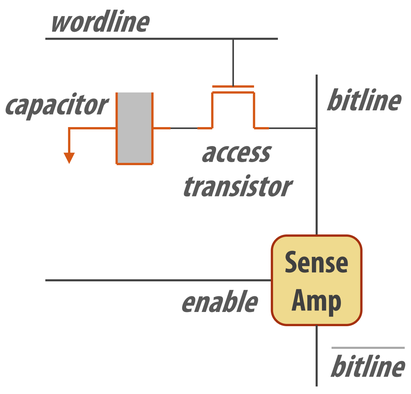
[1] Li and Patel, BitWeaving, SIGMOD 2013 [2] Goodwin+, BitFunnel, SIGIR 2017

Memory Channel Bottleneck

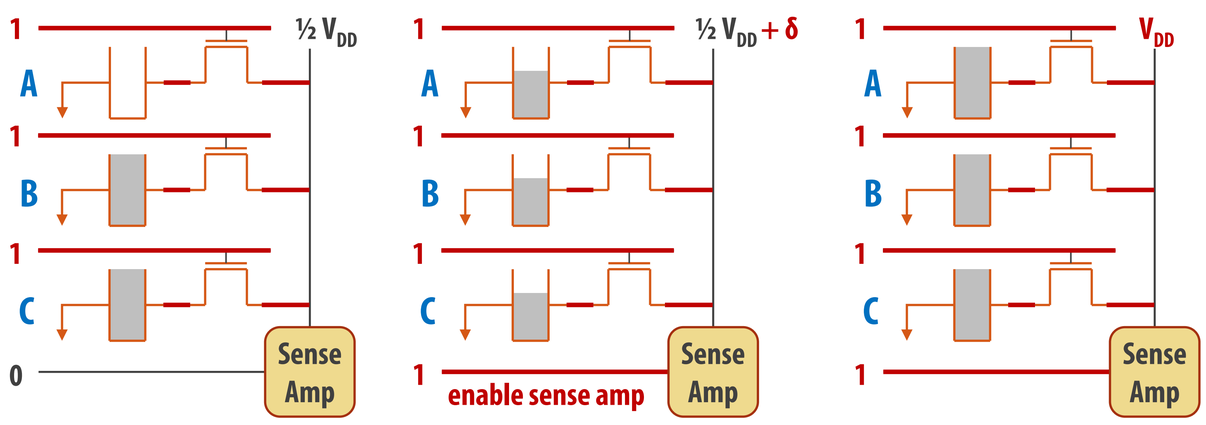


Our Approach: Perform bitwise operations completely inside DRAM

DRAM Cell + Sense Amp



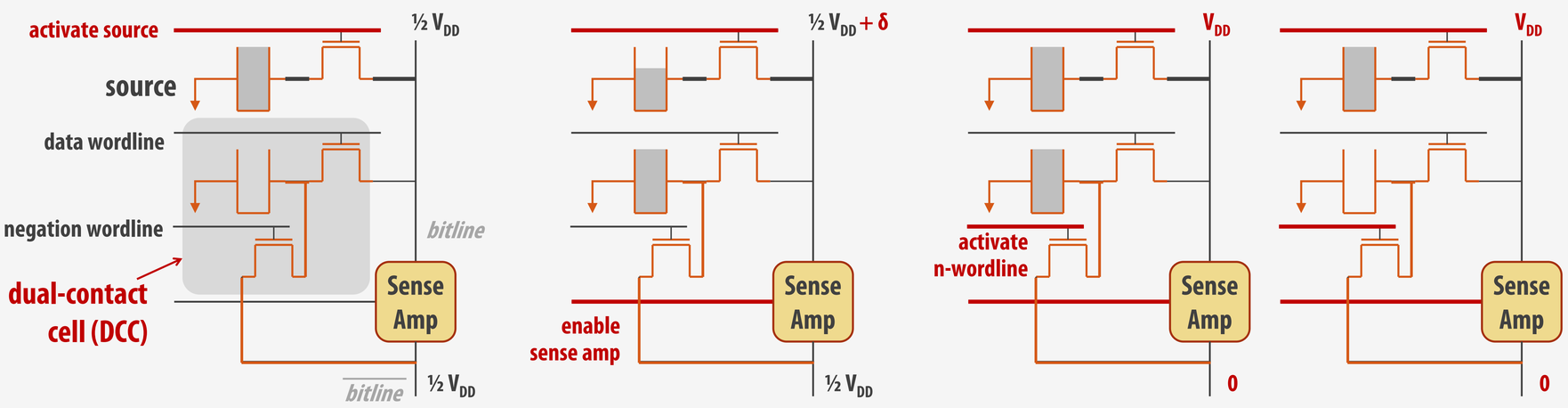
Ambit AND/OR – Triple-Row Activation (TRA)



$$\text{Output} = AB + BC + CA$$
$$C\,(A \text{ or } B) \;+\; \sim C\,(A \text{ and } B)$$

- Designated rows **t1, t2, t3** for TRA
- **Copy** data to designated rows
- **Reserved addresses** for reserved rows
  - e.g., ACTIVATE reserved address 0
  - simultaneous activation of t1, t2, t3
- SPICE simulations → **TRA highly reliable**

Ambit NOT – Dual-Contact Cell



Ambit Energy Savings

Operation	Energy ↓
not	59.5X
and/or	43.9X
nand/nor	35.1X
xor/xnor	25.1X
mean	35.0X

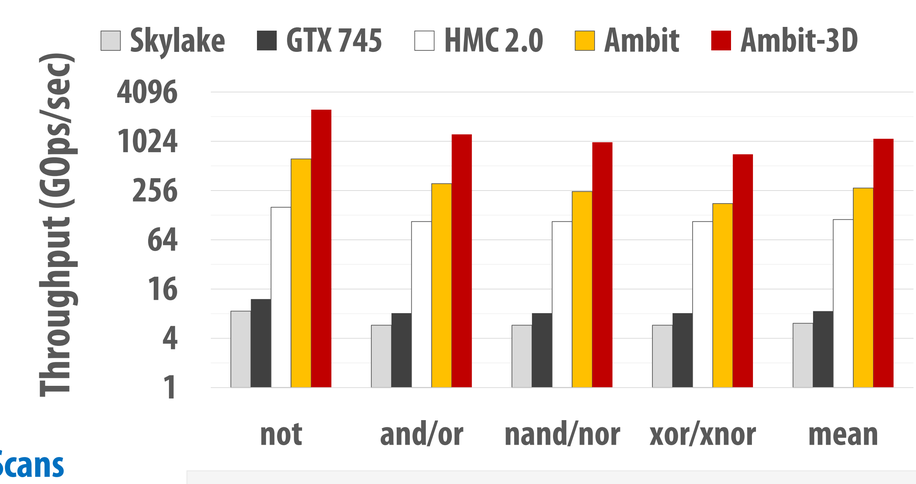
DRAM Implementation

- Reserve designated rows in each subarray
- Use RowClone to perform data copy
- Use separate small row decoder for designated/DCC rows
  - Overlap back-to-back activations
  - Lower row decoder complexity

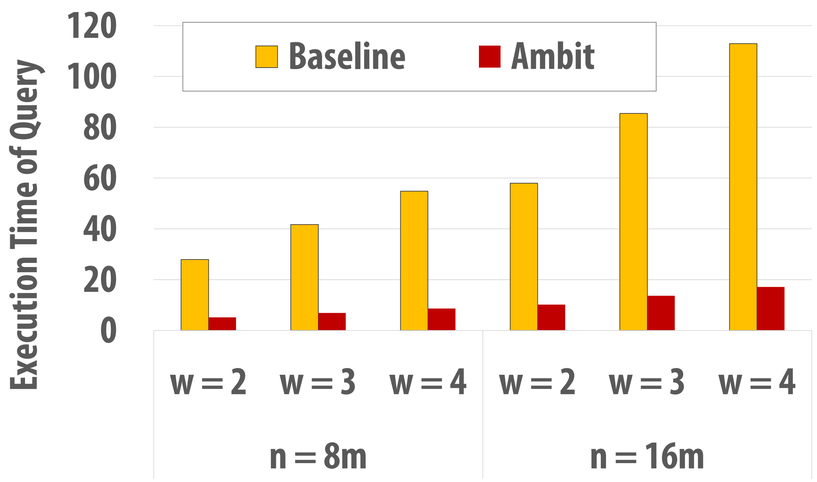
Integration with System

- Option 1: **PCIe device**
- Option 2: **System memory bus**
  - Same DRAM bus interface
  - Avoid CPU-device communication
  - CPU Instructions to trigger Ambit
  - Existing cache coherence support

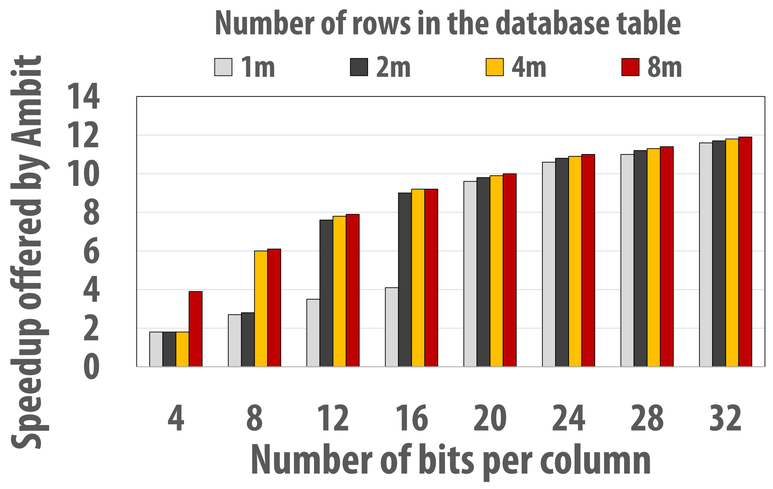
Ambit Throughput Improvement



Database Bitmap Indices



BitWeaving – Accelerating Database Scans



Other Applications

- Set operations: Ambit better than red-black trees
- BitFunnel: accelerating web search document filtering
- Cryptography, genome analysis, masked initializations