Coding for memory systems

WNCG, UT Austin

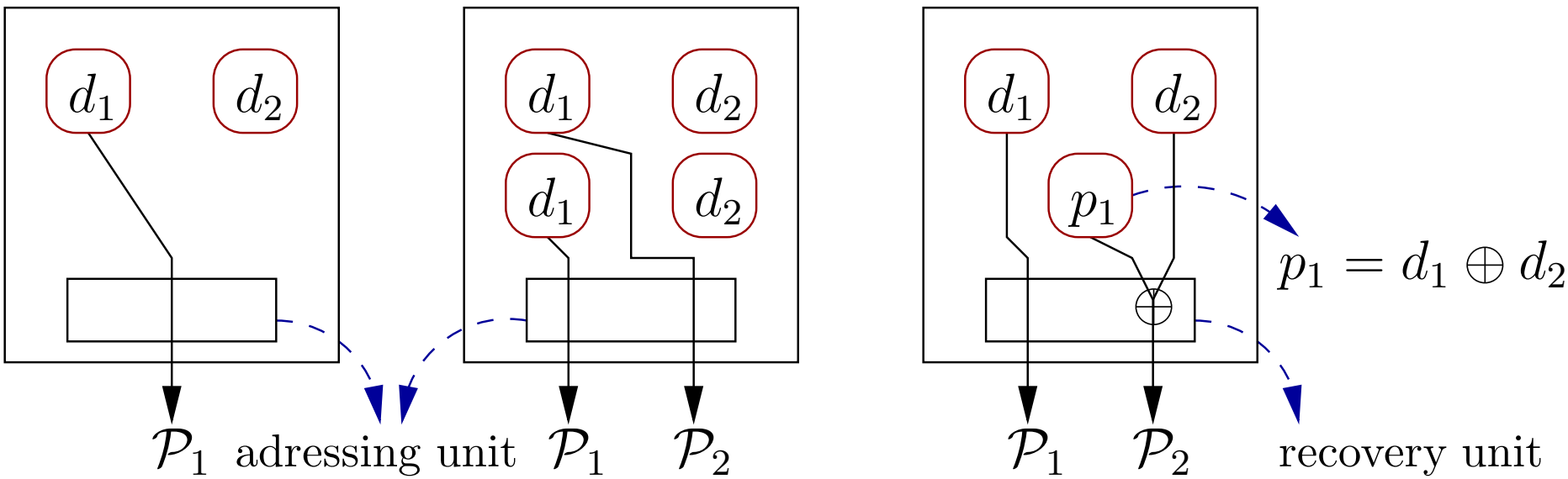
December 7, 2012

Coding for memory systems

- Idea: Add redundancy to the data stored in memory bank in order to reduce latency, and increase memory access speed.
- Basic mechanism: Erasure codes allow to add redundancy such that when a memory bank is serving to a processor, another processor can be served simultaneously by re-constructing the data of the busy memory bank (i.e., considering the busy bank as erasure).

Simplest example

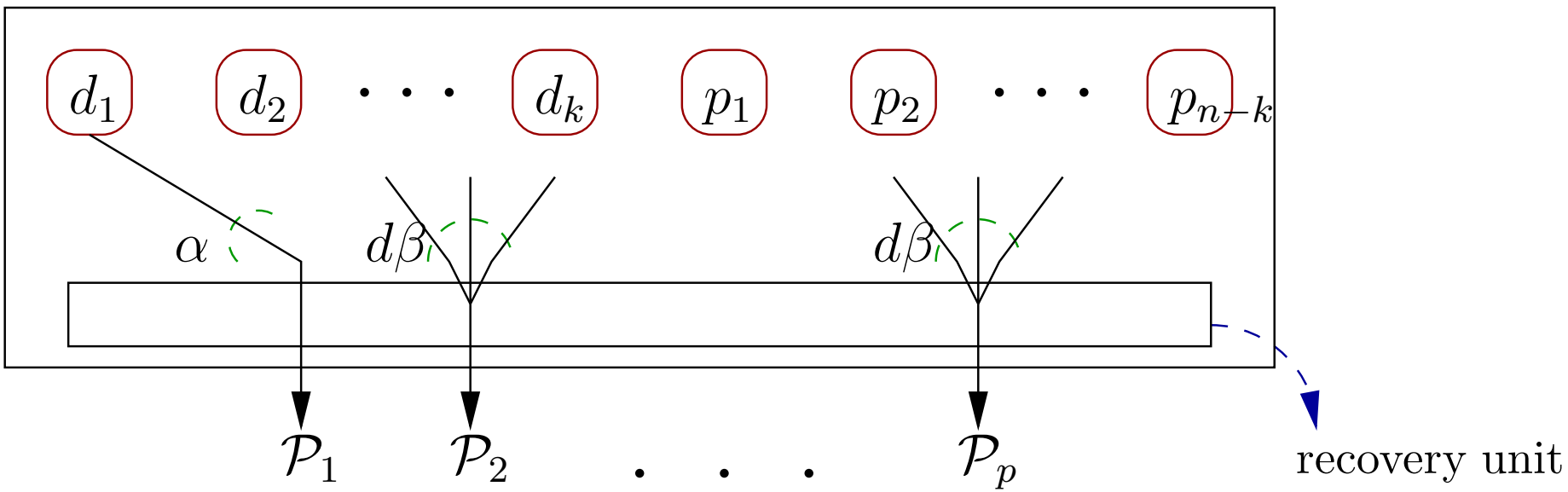
Scenario: A processor is accessing to one of the two memory banks. We want to double the capacity, and add another processor. But, this may create bank conflicts. Algorithmic Memory (by Memoir Systems) uses coding to resolve conflicts.



Note: Algorithmic Memory (by Memoir Systems) is using erasure codes (like Reed-Solomon codes). Different codes exist allowing efficiency for different tradeoffs.

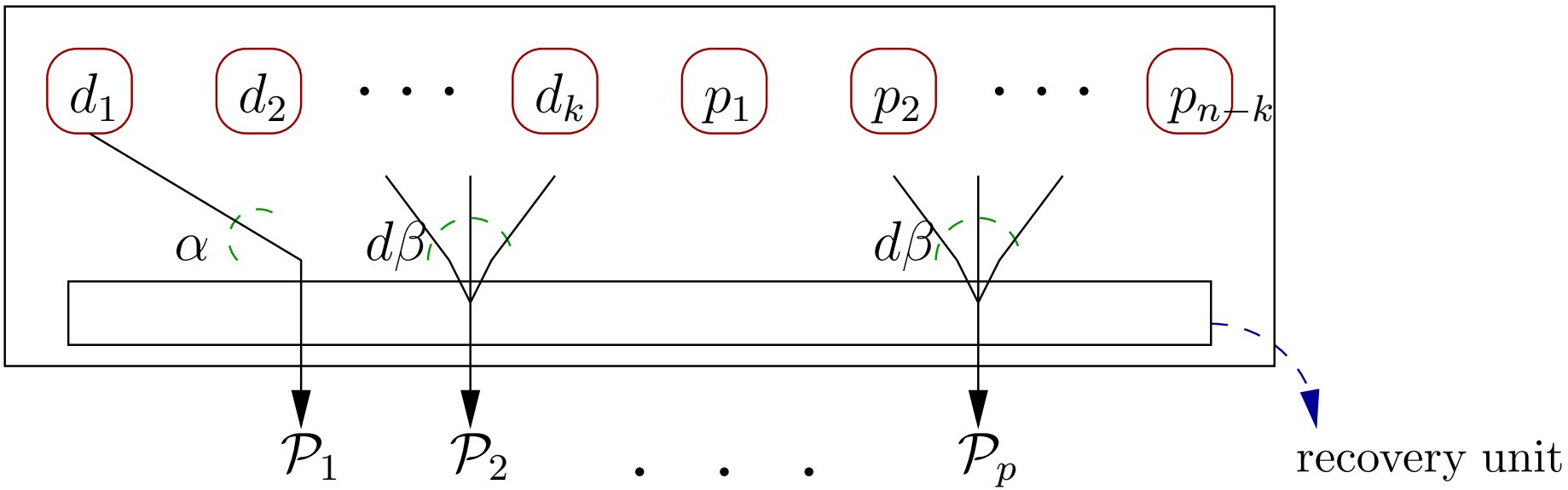
Regenerating codes

Regenerating codes allow for trading-off a) the complexity of the circuitry (or bandwidth) used for the recovery of busy bank and b) data bus bandwidth connecting banks to processors.



- Encoding k memory banks, d_1 to d_k, into n banks by adding $n-k$ parity banks p_1 to p_{n-k}

Regenerating codes



- α: Data bus bandwidth connecting banks to each processor
- d: Number of banks connected by recovery circuitry in the event that a busy memory bank needs to be connected to a processor
- β: Amount of data downloaded to the recovery circuitry

Code design for the worst case

- Consider all the processors are interested in the first memory bank, d_1.
- Serve $\mathcal{P}_1$ from bank storing d_1. Use recovery circuitry for the rest of the processors. Total nodes need to satisfy $n \geq 1 + (p-1)d$. As we seek to maximize *processors served per memory bank* p/n, we set $n = 1 + (p-1)d$ for this analysis.
- Given $n = 1 + (p-1)d$, the code trades-off the followings
  - Maximize *capacity of the system* k/n.
  - *Total recovery bandwidth per processor:* $\max\{\alpha, d\beta\}$.
  - *Data bus bandwidth from recovery unit to each processor:* α.

Extensions

- Allow for delay instead of worst case code design.
- Memory write operations, and update efficiency.