CS184a: Computer Architecture (Structure and Organization)

Day 16: February 14, 2005

Interconnect 4: Switching

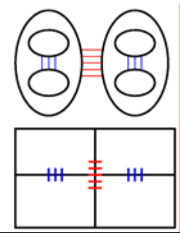
Previously

- Used Rent's Rule characterization to understand wire growth

$$IO = c\ N^p$$

- Top bisections will be $\Omega(N^p)$
- 2D wiring area

$$\Omega(N^p)\times\Omega(N^p) = \Omega(N^{2p})$$



We Know

- How we avoid $O(N^2)$ wire growth for "typical" designs
- How to characterize locality
- How we might exploit that locality to reduce wire growth
- Wire growth implied by a characterized design

Today

- Switching
  - Implications
  - Options

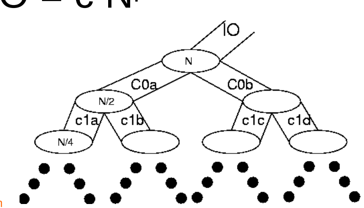
Switching:

How can we use the locality captured by Rent's Rule to reduce switching requirements? (How much?)

Observation

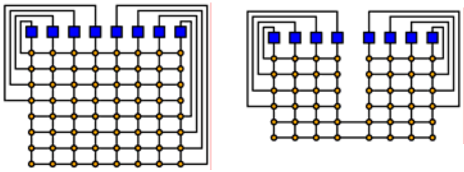
- Locality that saved us wiring, also saves us switching

$$IO = c\ N^p$$



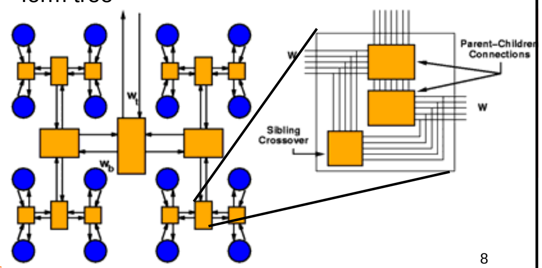
Consider

- Crossbar case to exploit wiring:
  - split into two halves, connect with limited wires
  - N/2 x N/2 crossbar each half
  - N/2 x $(N/2)^p$ connect to bisection wires
  - $2(N^2/4) + 2(N/2)^{p+1} < N^2$



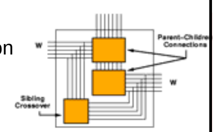
Recurse

- Repeat at each level
  - form tree



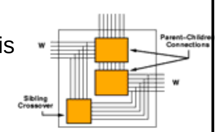
Result

- If use crossbar at each tree node
  - $O(N^{2p})$ wiring area
    - for p>0.5, direct from bisection
  - $O(N^{2p})$ switches
    - top switch box is $O(N^{2p})$
    - switches at one level down is
      - $2\times(1/2^p)^2 \times$ previous level
      - $(2/2^{2p})=2^{(1-2p)}$
      - coefficient < 1 for p>0.5



Result

- If use crossbar at each tree node
  - $O(N^{2p})$ switches
    - top switch box is $O(N^{2p})$
    - switches at one level down is
      - $2^{(1-2p)} \times$ previous level
    - Total switches:
      - $N^{2p} \times (1+2^{(1-2p)} + 2^{2(1-2p)} + 2^{3(1-2p)} + \ldots)$
      - get geometric series; sums to O(1)
      - $N^{2p} \times (1/1-2^{(1-2p)})$
      - $= 2^{(2p-1)}/(2^{(2p-1)}-1) \times N^{2p}$



Good News

- Good news
  - asymptotically optimal
  - Even without switches, area $O(N^{2p})$
    - so adding $O(N^{2p})$ switches not change

Bad News

- Switches area >> wire crossing area
  - Consider 6λ wire pitch $\Rightarrow$ crossing 36 λ^2
  - Typical (passive) switch $\Rightarrow$ 2500 λ^2
  - Passive only: 70x area difference
    - worse once rebuffer or latch signals.
- …and switches limited to substrate
  - whereas can use additional metal layers for wiring area

Additional Structure

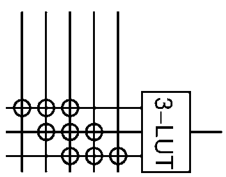
- This motivates us to look beyond crossbars
  - can depopulate crossbars on up-down connection without loss of functionality?

Can we do better?

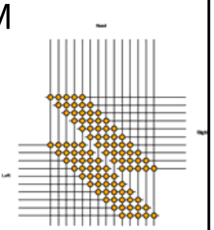
- Crossbar too powerful?
  - Does the specific down channel matter?
- What do we want to do?
  - Connect to *any* channel on lower level
  - Choose a subset of wires from upper level
    - order not important

N choose K

- Exploit freedom to depopulate switchbox
- Can do with:
  - $K \times (N-K+1)$ swtiches
  - Vs. $K \times N$
  - Save $\sim K^2$



N-choose-M



- Up-down connections
  - only require concentration
    - choose M things out of N
  - *i.e.* order of subset irrelevant
- Consequent:
  - can save a constant factor ~ $2^p/(2^p-1)$
    - $(N/2)^p \times N^p$ vs $(N^p - (N/2)^p+1)(N/2)^p$
    - $P=2/3 \rightarrow 2^p/(2^p-1) \approx 2.7$
- Similary, Left-Right
  - order not important $\Rightarrow$ reduces switches

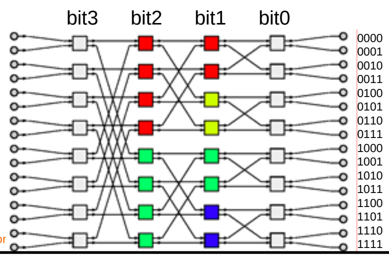
Multistage Switching

Multistage Switching

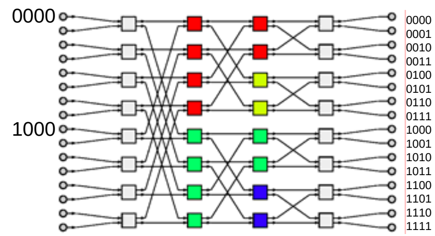
- Can route any **permutation** w/ less switches than a crossbar
- If we allow switching in stages
  - Trade increase in switches in path
  - For decrease in total switches

Butterfly

- Log stages
- Resolve one bit per stage



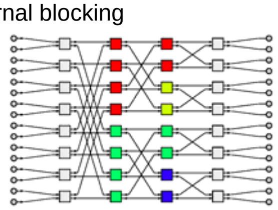
What can a Butterfly Route?



- 0000 → 0001
- 1000 → 0010

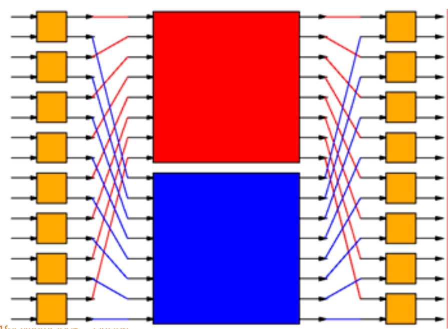
Butterfly Routing

- Cannot route all permutations
  - Get internal blocking



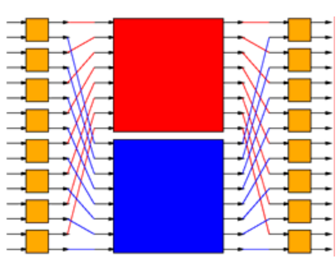
- What required for non-blocking network?

Decomposition



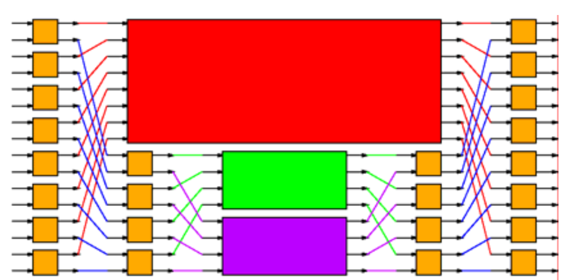
Decomposition

- Switches: $N/2 \times 2 \times 4 + (N/2)^2 < N^2$



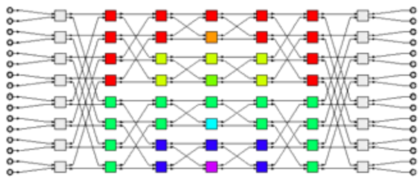
Recurse

If it works once, try it again…



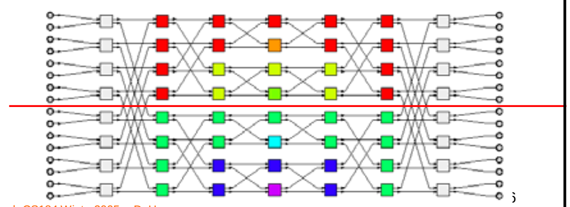
Result: Beneš Network

- $2\log_2(N)-1$ stages (switches in path)
- Made of N/2 2×2 switchpoints
  - (4 switches)
- $4N\times\log_2(N)$ total switches
- Compute route in O(N log(N)) time



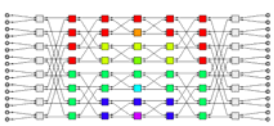
Beneš Network Wiring

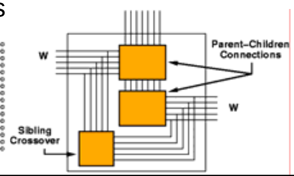
- Bisection: N
- Wiring → $O(N^2)$ area (fixed wire layers)



Beneš Switching

- Beneš reduced switches
  - N^2 to N(log(N))
  - using multistage network
- Replace crossbars in tree with Beneš switching networks



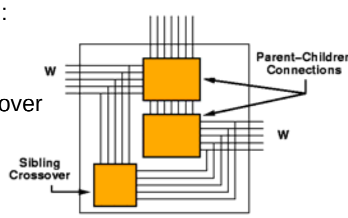


Beneš Switching

- Implication of Beneš Switching
  - still require $O(W^2)$ wiring per tree node
    - or a total of $O(N^{2p})$ wiring
  - now O(W log(W)) switches per tree node
    - converges to O(N) total switches!
  - $O(\log^2(N))$ switches in path across network
    - strictly speaking, dominated by wire delay $\sim O(N^p)$
    - but constants make of little practical interest except for very large networks ☹

Better yet…

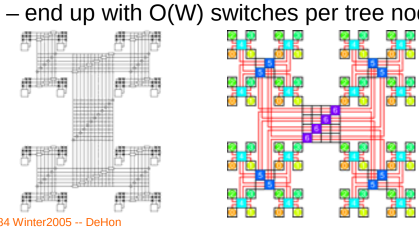
- Believe do not need Beneš on the up paths
- Single switch on up path
- Beneš for crossover
- Switches in path:
  - log(N) up
  - + log(N) down
  - + 2log(N) crossover
  - = 4 log(N)
  - = O(log (N))



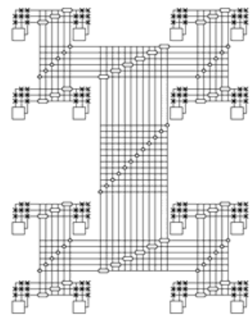
Linear Switch Population

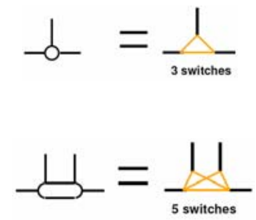
Linear Switch Population

- Can further reduce switches
  - connect each lower channel to O(1) channels in each tree node
  - end up with O(W) switches per tree node



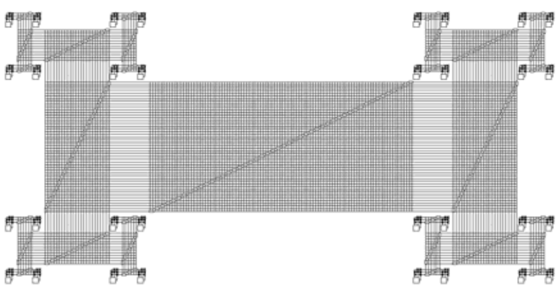
Linear Switch (p=0.5)





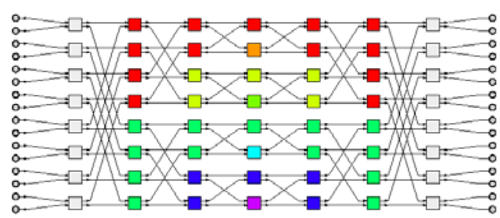
Linear Population and Beneš

- Top-level crossover of p=1 is Beneš switching



Beneš Compare

- Can shuffle stages so local shuffles on outside and big shuffle in middle



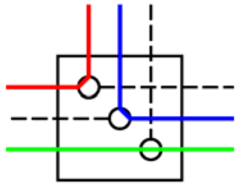
Linear Consequences: Good News

- Linear Switches
  - O(log(N)) switches in path
  - O(N^{2p}) wire area
  - O(N) switches
  - More practical than Beneš crossover case

Linear Consequences: Bad News

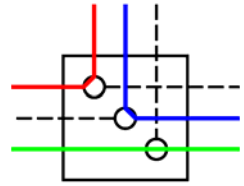
- Lacks guarantee can use all wires
  - as shown, at least mapping ratio > 1
  - likely cases where even **constant** not suffice
    - expect no worse than logarithmic
- Finding Routes is harder
  - no longer linear time, deterministic
  - **open** as to exactly how hard

Mapping Ratio



- Mapping ratio says
  - if I have W channels
    - may only be able to use W/MR wires
      - for a particular design's connection pattern
    - to accommodate any design
      - forall channels
        - physical wires $\geq$ MR $\times$ logical

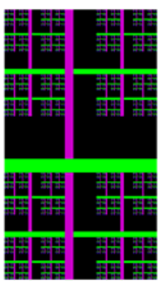
Mapping Ratio



- Example:
  - Shows MR=3/2
  - For Linear Population, 1:1 switchbox

Area Comparison

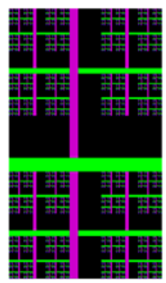
Both:
p=0.67
N=1024



M-choose-N perfect map

Linear MR=2

Area Comparison



M-choose-N perfect map

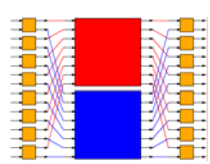
Linear MR=2

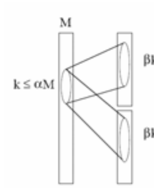
- Since
  - switch >> wire
- may be able to tolerate MR>1
- reduces switches
  - net area savings
- Empirical:
  - Never seen greater than 1.5

Expanders

Expander Theory

- (α,β)–expansion
  - Any group of size $k=\alpha N$ will expand connect to a group of size $\beta k=\beta\alpha N$ in each logical direction

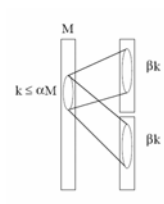




[Arora, Leighton, Maggs SIAM Journal of Comp. v25n3p600 1996]

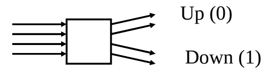
Expander Idea

- **IF** we can achieve expansion
  - Can guarantee non-blocking at each stage
- *i.e.*
  - Guarantee use less than αN
  - Guarantee connections to more stuff in next level
  - Since $\beta\alpha$N>αN available in next level
    - Guaranteed to be an available switch

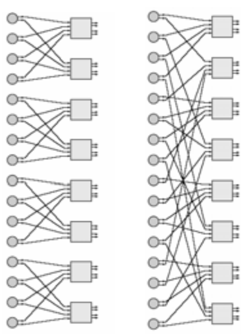


Dilated Switches

- Have multiple outputs per logical direction
  - **Dilation:** number of outputs per direction
  - *E.g.* radix 2 switch w/ 4 outputs
    - 2 per direction
    - Dilation 2



Dilated Switches allow Expansion



- On Right
  - Any pair of nodes connects to 3 switches
- Strictly speaking must have d>2 for expansion

Random Wiring

- Random, dilated wiring for butterfly can achieve

$$d > \beta + 1 + \frac{\beta + 1 + \ln 2\beta}{\ln(\frac{1}{2\alpha\beta})}$$

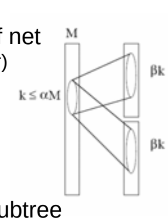
$$2d > 2\beta + 1 + \frac{2\beta + 1 + \ln 2\beta}{\ln(\frac{1}{2\alpha\beta})}$$

  - For tree…2→2^p (?)

[Upfal/STOC 1989, Leighton/Leiserson/Kravets MIT/LCS/RSS 8 1990]

Constraints

- Total load should not exceed α of net
  - L=mapping ratio (light loading factor)
  - αLW = number into each subtree
  - αLW ≥ W/2^p
    - L ≥ 1/($2^p\alpha$)



- Cannot expand past the size of subtree
  - $\beta\alpha$N≤N/2^p
  - $\beta\alpha$ ≤ 1/2^p

Extra Switches

- Extra switch factor: d×L
- Try:
  - β=2, α=1/10
  - d=8
  - dL≈40 (p=1)
- Try:
  - β=3/2, α=1/10 → dL≈30

Putting it Together

- Base, linear-population trees have O(N) switches
- Make larger by a factor of L (linear factor)
- Dilated version have a factor of d more switches
- Randomly wired expander
  - Can have O(N) switches
  - Guarantee routes
  - Constants < 100
  - Open: how tight can make it?

Big Ideas [MSB Ideas]

- In addition to wires, must have switches
  - Have significant area and delay
- Rent's Rule locality reduces
  - both wiring and switching requirements
- Naïve switches match wires at $O(N^{2p})$
  - switch area >> wire area
  - prevent benefit from multiple layers of metal

Big Ideas [MSB Ideas]

- Can achieve O(N) switches
  - plausibly O(N) area with sufficient metal layers
- Switchbox depopulation
  - save considerably on area (delay)
  - will waste wires
  - May still come out ahead (evidence to date)