ESE534: Computer Organization

Day 19: April 5, 2010

Interconnect 4: Switching

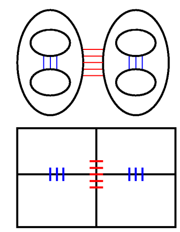
Previously

- 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
- Exploiting Multiple Metal Layer

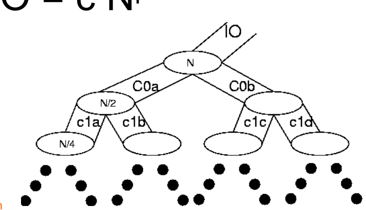
Switching:

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

Observation

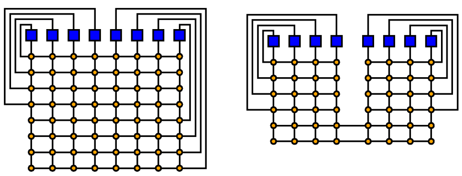
- Locality that saved us wiring, also saves us switching

$$IO = c\,N^p$$



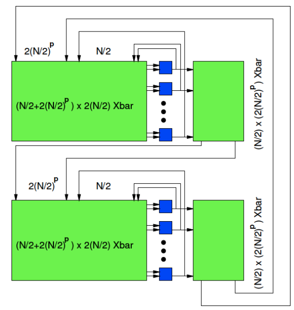
Consider

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



Preclass 1

- Crosspoints?
- Symbolic ratio?
- Ratio $N=2^{15}$, $p=2/3$?

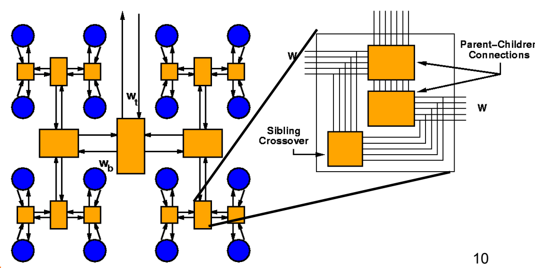


What next

- When something works once?
- …we try it again…

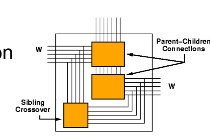
Recurse

- Repeat at each level ➔ form a tree



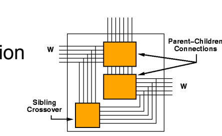
Result

- 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 $O(N^{2p})$
      - $2\ W_{top}\times W_{bot}+(W_{bot})^2$
      - $2\ (N^p\times(N/2)^p)+(N/2)^{2p}$
      - ~~$N^{2p}(1/2^p + 1/2^{2p})$~~ Correction from lecture
      - $N^{2p}(2/2^p + 1/2^{2p})$



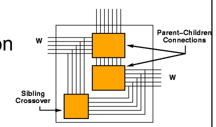
Result

- 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 $O(N^{2p})$
      - $N^{2p}(2/2^p + 1/2^{2p})$
    - switches at one level down is
      - $2\times(N/2)^{2p}(2/2^p + 1/2^{2p})$
      - $2\times(1/2^p)^2 \times(N^{2p}(2/2^p + 1/2^{2p}))$
      - $2\times(1/2^p)^2 \times$ previous level



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 $O(N^{2p})$
      - $N^{2p}(1/2^{p} + 1/2^{2p})$
    - 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

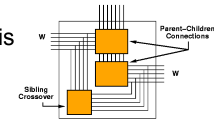


Now believe this is correct. Just got confused In lecture.

Example: p=2/3
$2^{(1-2\times2/3)}=2^{-1/3}=0.7937$

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

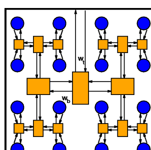
- Good news
  - asymptotically optimal
  - Even without switches, area $O(N^{2p})$
    - so adding $O(N^{2p})$ switches not change

Bad News

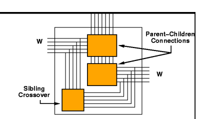
- Switches area >> wire crossing area
  - Consider 8λ wire pitch $\Rightarrow$ crossing $64\ \lambda^2$
  - Typical (passive) switch $\Rightarrow$ $2500\ \lambda^2$
  - Passive only: 40× area difference
    - worse once rebuffer or latch signals.
- …and switches limited to substrate
  - whereas can use additional metal layers for wiring area

Additional Structure

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



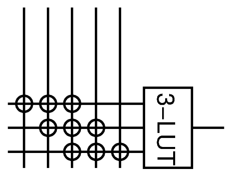
Can we do better?



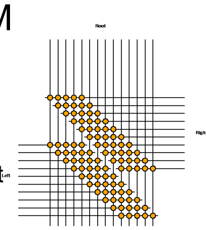
- Crossbar too powerful?
  - Does the specific down channel matter?
- What do we want to do?
  - Connect to *any* channel on lower level
  - Choose a subset of wires from upper level
    - order not important

N choose K

- Exploit freedom to depopulate switchbox
- Can do with:
  - $K \times (N-K+1)$ switches
  - Vs. $K \times N$
  - Save ~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 → $2^p/(2^p-1) \approx 2.7$
- Similary, Left-Right
  - order not important ⇒ reduces switches

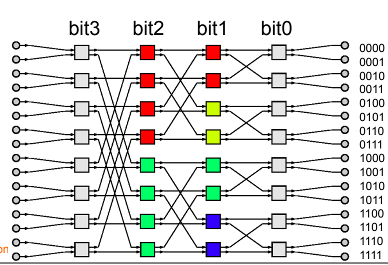
Multistage Switching

Multistage Switching

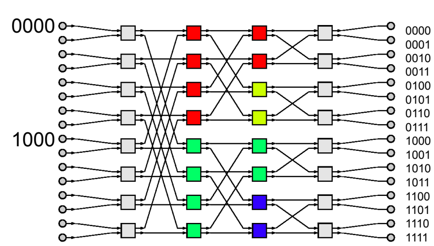
- We can route any **permutation** with fewer switches than a crossbar
- If we allow switching in stages
  - Trade increase in switches in path
  - For decrease in total switches

Butterfly

- Log stages
- Resolve one bit per stage

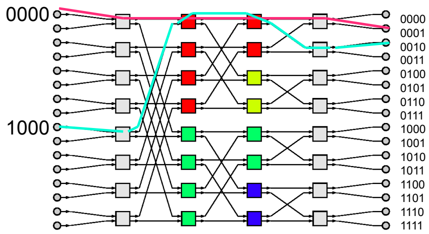


What can a Butterfly Route?



- 0000 → 0001
- 1000 → 0010

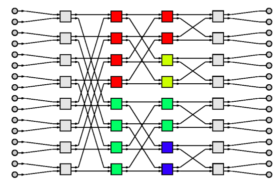
What can a Butterfly Route?



- 0000 → 0001
- 1000 → 0010

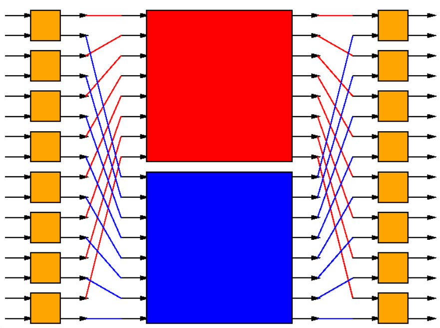
Butterfly Routing

- **Cannot** route all permutations
  - Get internal blocking

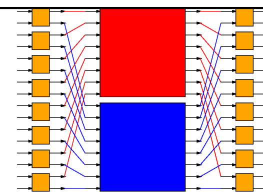


- What required for non-blocking network?

Decomposition



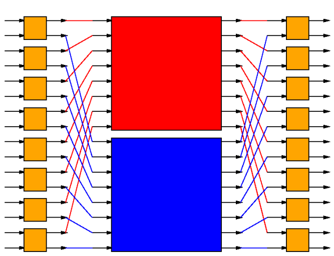
Decomposed Routing



- Pick a link to route.
- Route to destination over red network
- At destination,
  - What can we say about the link which shares the final stage switch with this one?
  - What can we do with link?
- Route that link
  - What constraint does this impose?
  - So what do we do?

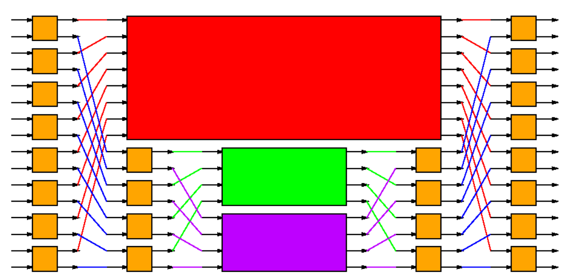
Decomposition

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



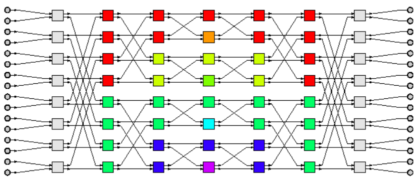
Recurse

If it works once, try it again…



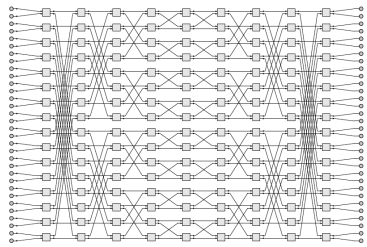
Result: Beneš Network

- $2\log_2(N)-1$ 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



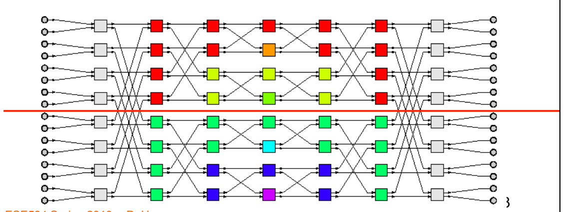
Preclass 2

- Switches in Benes 32?
- Ratio to 32x32 Crossbar?



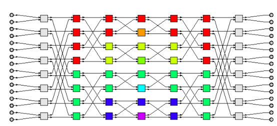
Beneš Network Wiring

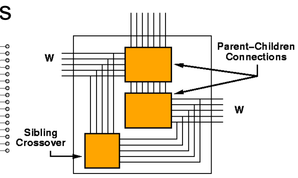
- Bisection: N
- Wiring → $O(N^2)$ area (fixed wire layers)



Beneš Switching

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



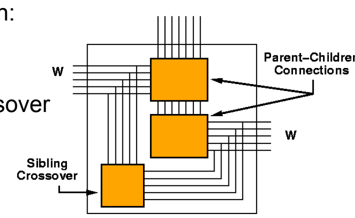


Beneš Switching

- 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…

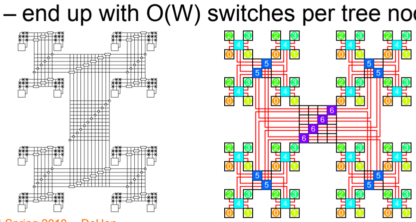
- Believe do not need Beneš on the up paths
- Single switch on up path
- Beneš for crossover
- Switches in path:
  - log(N) up
  - \+ log(N) down
  - \+ 2log(N) crossover
  - = 4 log(N)
  - = O(log (N))



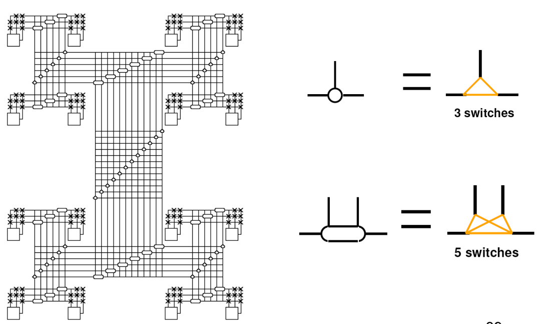
Linear Switch Population

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

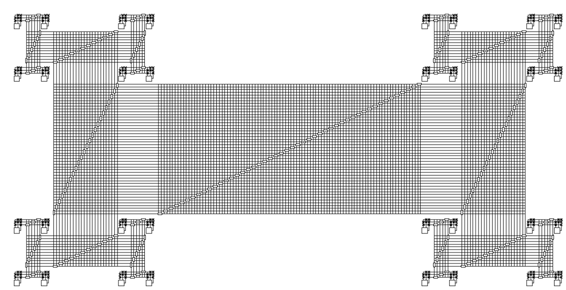


Linear Switch (p=0.5)



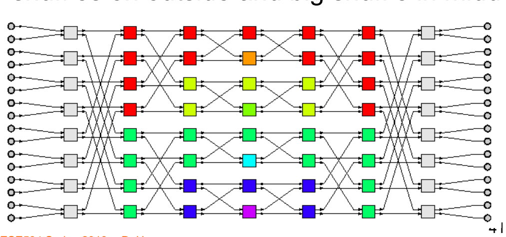
Linear Population and Beneš

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



Beneš Compare

- Can permute stage switches so local shuffles on outside and big shuffle in middle

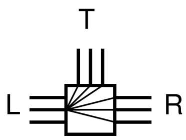


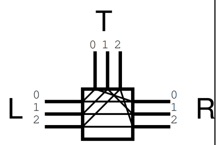
Linear Consequences: Good News

- Linear Switches
  – O(log(N)) switches in path
  – $O(N^{2p})$ wire area
  – O(N) switches

  – More practical than Beneš crossover case

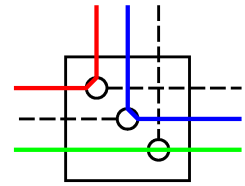
Preclass 3





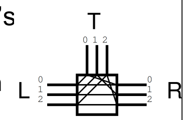
- Route Sets:
  - T→(L & R), L→ (T & R), R→(T & L)
  - T→L, R→T, L→R
  - T→L, L→R, R→L

Mapping Ratio

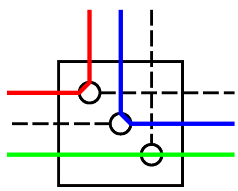


- Mapping ratio says
  - if I have W channels
    - may only be able to use W/MR wires
      - for a particular design's connection pattern
    - to accommodate any design
      - forall channels

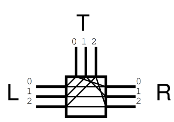
        physical wires ≥ MR × logical



Mapping Ratio



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

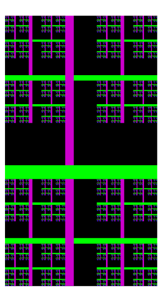


Linear Consequences: Bad News

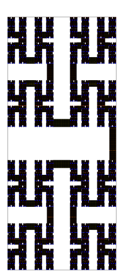
- Lacks guarantee can use all wires
  - as shown, at least mapping ratio > 1
  - likely cases where even **constant** not suffice
    - expect no worse than logarithmic
- Finding Routes is harder
  - no longer linear time, deterministic
  - **open** as to exactly how hard

Area Comparison

Both:
p=0.67
N=1024

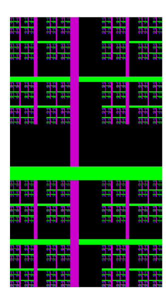


M-choose-N perfect map

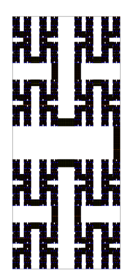


Linear MR=2

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 MR greater than 1.5

Multilayer Metal

Day 7

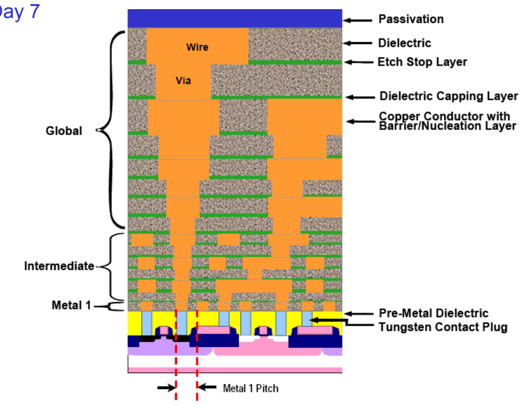
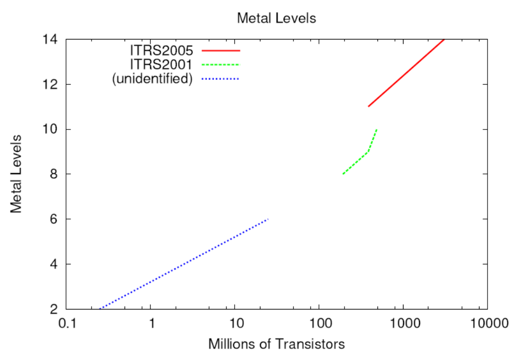


Figure 70 Cross-section of Hierarchical Scaling—MPU Device

Day 7

Wire Layers = More Wiring

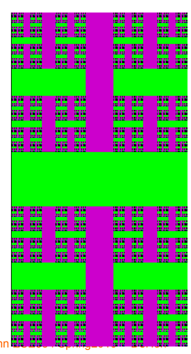


Opportunity

- Multiple Layers of metal allow us to
  - Increase effective pitch
  - Potentially route in 3D volume

Day 18

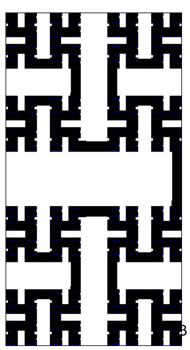
Larger "Cartoon"



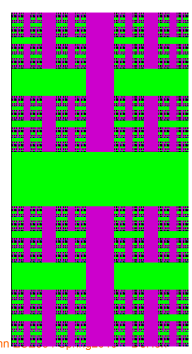
1024 LUT Network

P=0.67

LUT Area 3%



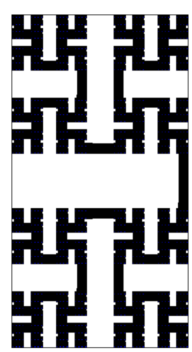
Challenge



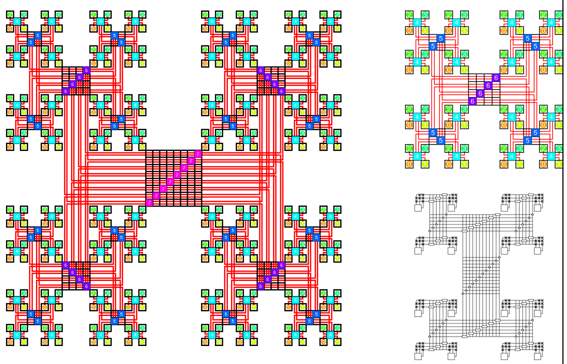
Spread or balance wires

Homogenize switches

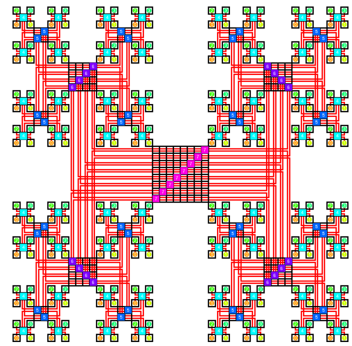
Avoid via saturation



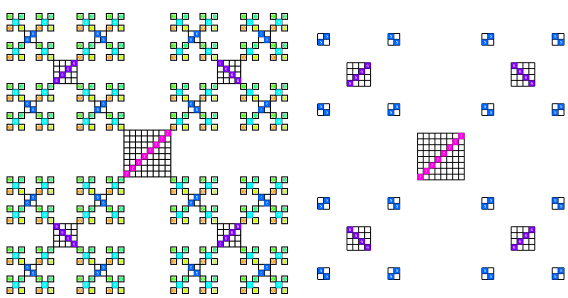
Linear Population Tree (P=0.5)



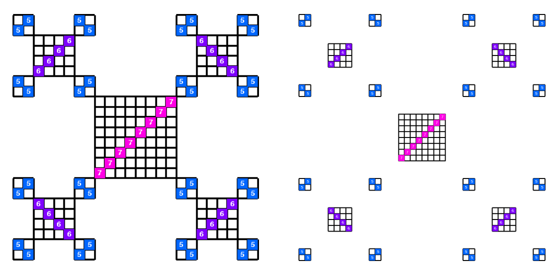
Linear Population Tree (P=0.5)



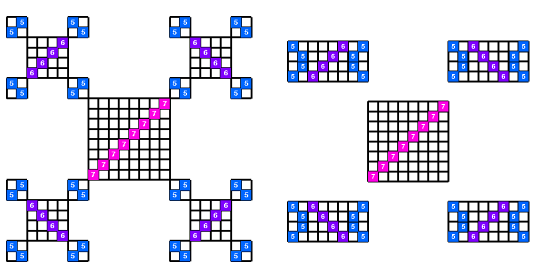
BFT Folding



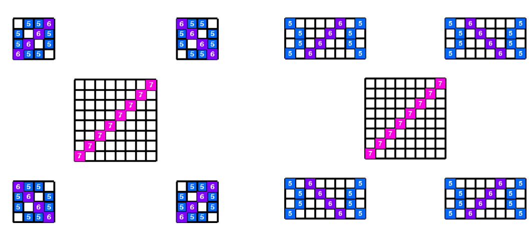
BFT Folding



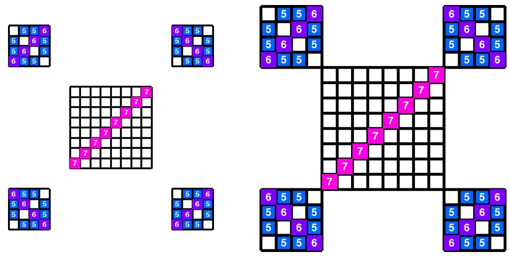
BFT Folding



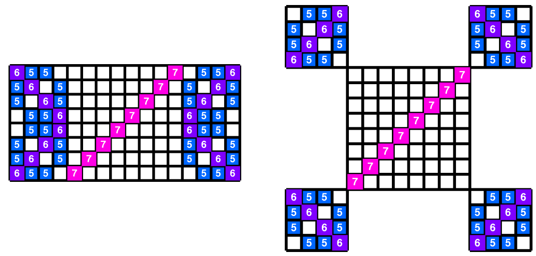
BFT Folding



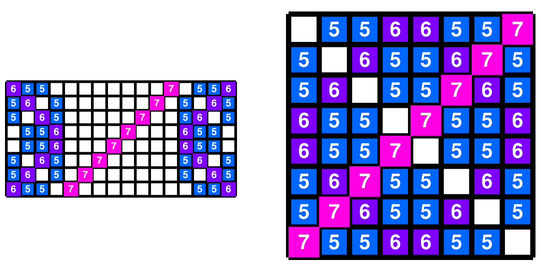
BFT Folding



BFT Folding

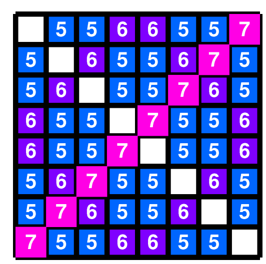


BFT Folding

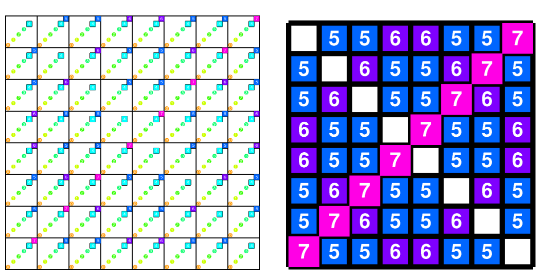


Invariants

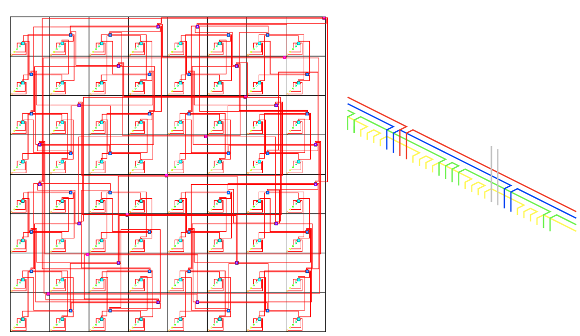
- Lower folds leave both diagonals free
- Current level consumes one, leaving other free



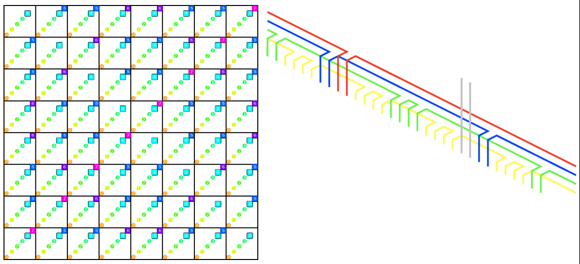
BFT Folding



Wiring

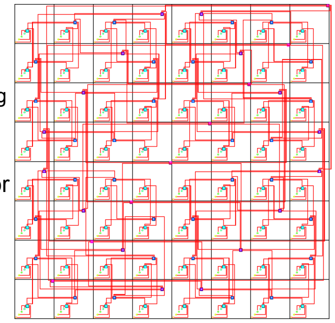


Avoid via Saturation



Compact, Multilayer Linear Population Tree Layout

- Can layout BFT
  - in O(N) 2D area
  - with O(log(N)) wiring layers
- Can be extended for p>0.5 as well
  - Wire layers grow as $O(N^{(p-0.5)})$



Admin

- HW6 Graded
- HW8 dues 4/12
  - Parts still need Wednesday's lecture
- Reading for Wednesday on web

Big Ideas [MSB Ideas]

- In addition to wires, must have switches
  - 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)