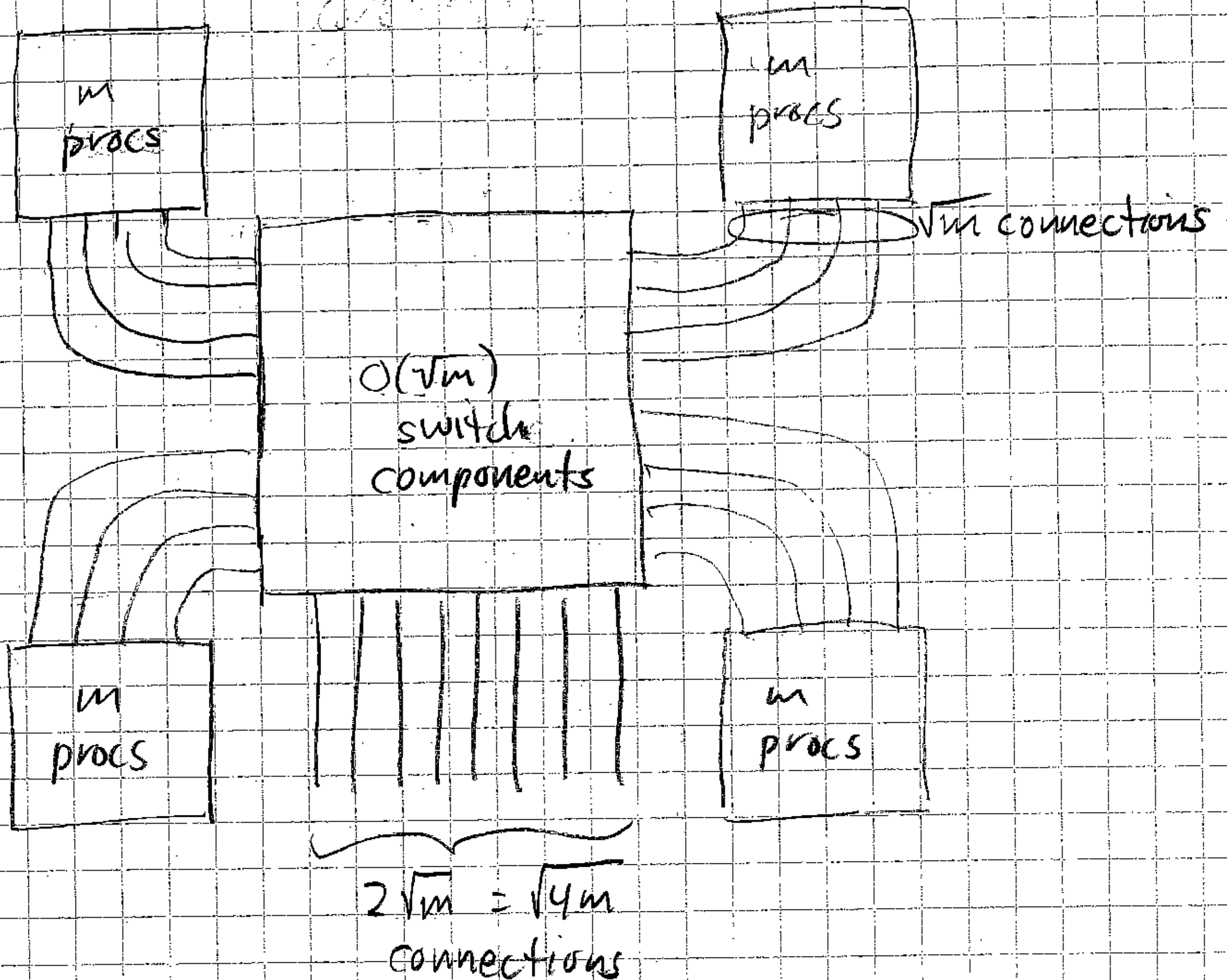
Area-Universal Networks

Idea: Make #wires leaving region proportional to perimeter of region (like 2D mesh or TOM), but small diameter.

Fat-Tree (slide 2)



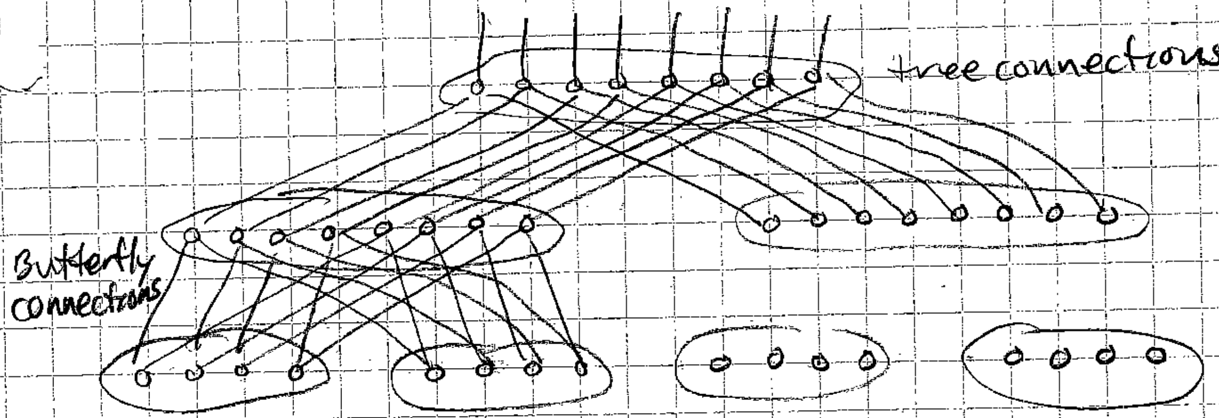
What is area $A(N)$ of N-leaf fat-tree?
Embed in $TOM(\sqrt{N}) \Rightarrow A(N) = O(N \lg^2 N)$
or $S(N) = \sqrt{A(N)}$

$$S(N) = 2S(N/4) + \Theta(\sqrt{N})$$
$$= \Theta(\sqrt{N} \lg N)$$
$$\therefore A(N) = \Theta(N \lg^2 N)$$

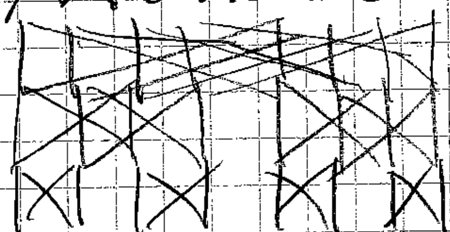
« Know master theorem for final exam! »

How many switches?
$$H(N) = 4H(N/4) + \Theta(\sqrt{N})$$
$$= \Theta(N).$$

Butterfly fat-tree (Slide 3)

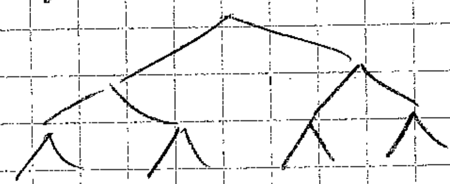


Only $\bowtie$ connections



butterfly
(or Benes)

Only tree connections

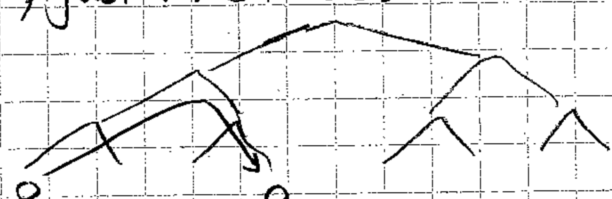


CBT

Alternate: area-universal network
Other ratios: variable growth of connections
- scalable to available technology
- no math law governs fatness.

Routing on fat-trees

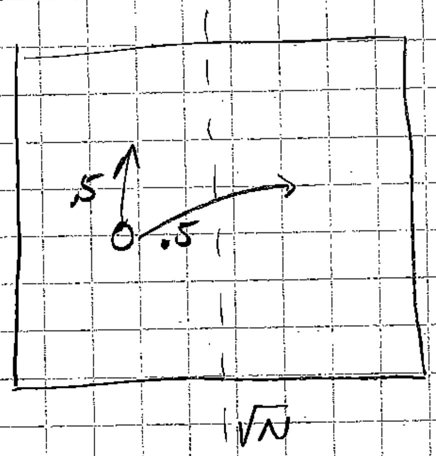
Message goes from source to dest via least common ancestor, just like tree:



Like phone network with exchanges.

6.896
5/5/04
L22.3

Issue: some routing problems harder than others.
Example: random communication is hard



$E[\#\text{ crossing bisection}] = N/2$
Bandwidth of bisection $= \sqrt{N}$
Time $\geq \frac{N}{2} \div \sqrt{N} = \Omega(\sqrt{N})$.

Load factor $\lambda = \max_{\{\text{cuts}\}} \frac{\#\text{msgs crossing cut}}{\text{bandwidth of cut}}$

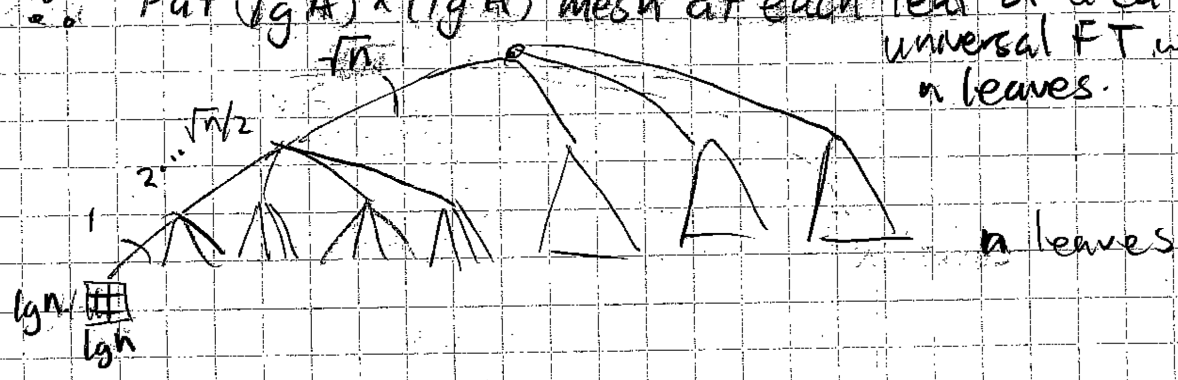
Lemma. Any set of messages with load factor λ can be routed on an N-leaf fat-tree in expected time $O(\lambda + \lg N)$.

Pf. [Leighton-Maggs-Ranade-Rao]. $\boxtimes$

Theorem. An N-leaf (N procs) fat-tree can simulate any area-N fat-tree in $O(\lg N)$ time.

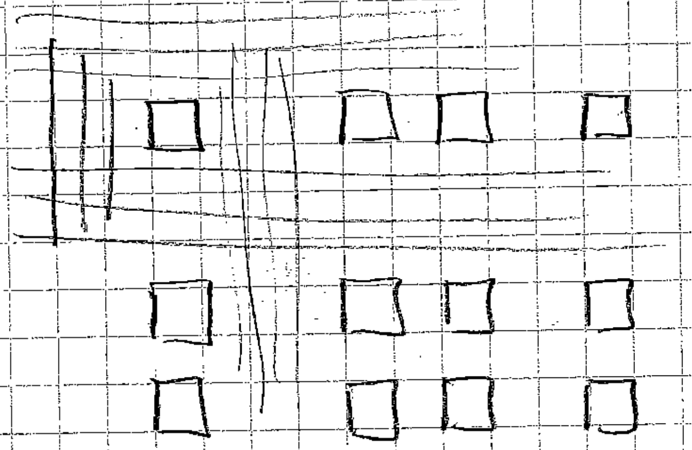
Proof. (Slides 4-5) $\lambda = O(1)$ $\boxtimes$

Tighter result: Area-$\Theta(A)$ network that can simulate any other area-A network in $O(\lg A)$ time.
Idea: Area-A area-univ FT has $N = A/\lg^2 A$ procs. <<sloppy asympt>>
$\therefore$ Put $(\lg A) \times (\lg A)$ mesh at each leaf of area-universal FT with n leaves.



$N = n \lg^2 n$

Area: 1 proc $\to (\lg n) \times (\lg n)$ procs expands each dimension of layout by <u>additive</u> $\sqrt{n} \lg n$.



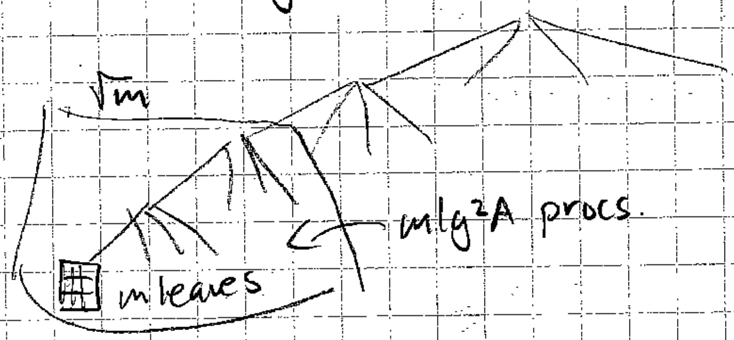
procs are aligned.

$\therefore$ Side length increases by const. factor.
$A = \Theta(n \lg^2 n)$. Procs are dense.

Simulation of network R

1. Divide R into $\sqrt{n} \times \sqrt{n}$ blocks of size $(\lg A) \times (\lg A)$
2. Simulate procs in (i,j) block of R by corresp. $(\lg A) \times (\lg A)$ leaf mesh of FT. Time $= O(\lg A)$
3. Communication among blocks:

- m-leaf subtree of FT has $\sqrt{m}$ external connections corresp. to $m \lg^2 A$-area region of R.



- Simulate wires of R with msgs.
  How many msgs out of $m \lg^2 A$ area? $O(\sqrt{m} \lg A)$

$\lambda = \dfrac{\sqrt{m} \lg A}{\sqrt{m}} = \lg A \Rightarrow$ Routing time $= O(\lg A)$ $\boxtimes$