# FPGA Interconnect

**Nachiket Kapre**
nachiket@uwaterloo.ca

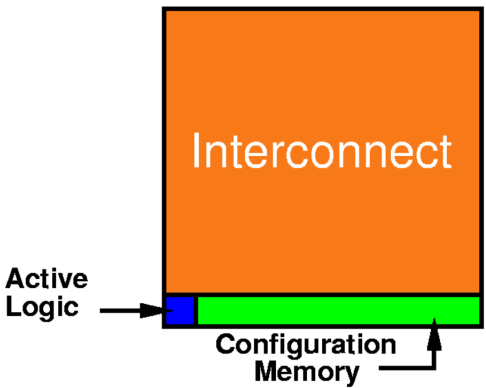
# Lecture Outline

- **Wiring and Switching** – Costs and tradeoffs
- **Putting it together** – sbox and cbox model

# Properties of Communication in Circuits

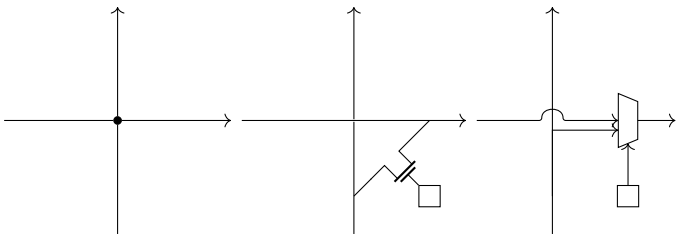
- Wiring in circuits connect gates and registers to each other
  - Persistent connection. Even if data doesn't change.
- Locality of connectivity.
  - Do not need all-to-all connectivity.
  - Input equivalence at LUT
  - Fewer long connections, More short connections.
- Wiring area and delay dominate overall circuit behavior

# Relative cost of Logic and Interconnect



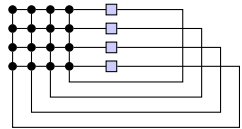

- Efficient to decompose large-input functions into lots of small LUTs + wires
  - Large LUT area scales as $2^k_{large}$
  - Alternative: $N \times 2^k_{small} + A_{interconnect}$
- Even with added expense of wires ($A_{interconnect}$), still better than $2^k_{large}$
- Interconnect area for *programmable logic* has two components:
  - Wires
  - Programmable Switches

# Directional Crosspoint



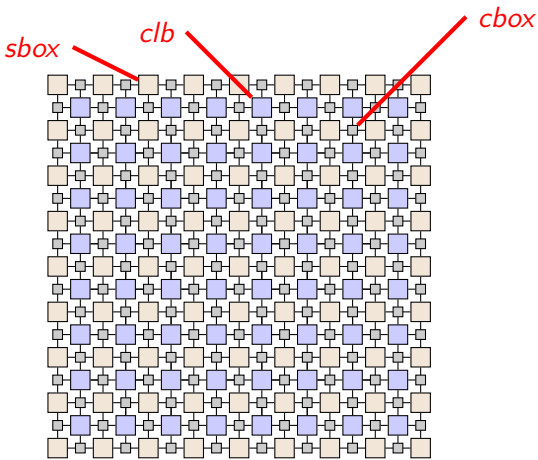
- A crosspoint for directional wires is one transistor + one SRAM cell
- Can also be visualized as a 2:1 mux driven by the two wires and controlled by the SRAM cell
- Hence, $A_{sw} = A_{Config} + A_{Mux}$

# Limits of crossbar



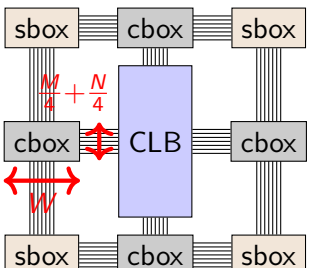
- Over-provisions connectivity
- $O(N)$ wires in the bisection
  - $N$ wires in horizontal and vertical cuts
  - 2D layout $O(N^2)$ wiring area
- $O(N^2)$ switches
- What dominates? (1) wires or (2) switches
  - $\lambda$ is a unit of length defined by the manufacturing technology *i.e.* 32 nm, 18 nm, 7 nm.
  - Wire pitch (spacing between wires) $\approx 10\lambda$
    - Wire area for $N$ wires is $(N \times 10\lambda)^2 = N^2 \times 100\lambda^2$
  - Switch area (2:1 mux+SRAM cell) is $1000\lambda^2$
    - Switch area for $N^2$ switches is $N^2 \times 1000\lambda^2$
  - Switch area ($N^2 \times 1000\lambda^2$) >> Wire area ($N^2 \times 100\lambda^2$)

# Interconnect of Modern FPGAs



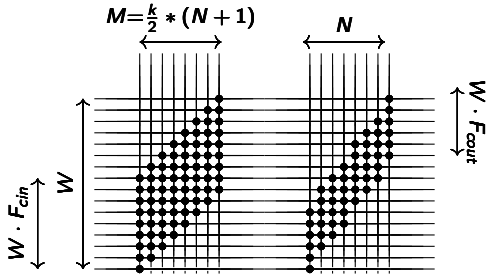

- Island-style FPGAs used by Xilinx/Altera
  - Each island is a LUT cluster (from last lecture)
  - Also called CLB (Configurable Logic Block)
- **Switch-box** (sbox): allows horizontal and vertical traffic to turn
- **Connection-box** (cbox): allows wires to enter/exit CLB

# Interconnect of Modern FPGAs



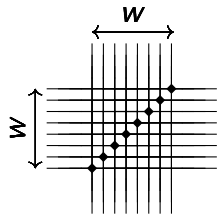

- Parameters of interest
  - $W$ is channel width (number of wires in horizontal/vertical tracks)
  - Recall, $M=\frac{k}{2} * (N+1)$, number of inputs to CLB
- **Connectionbox** parameters
  - $F_{cin}$ is the fraction of $W$ tracks connected to $M$ CLB inputs
    - Recall, $M=\frac{k}{2} * (N+1)$
  - $F_{cout}$ is the fraction of $W$ tracks connected to $N$ LUT outputs
- **Switchbox** parameters
  - $F_s$ is the number of switches connected to each wire at input of sbox

# Connection-box Architecture (cbox)



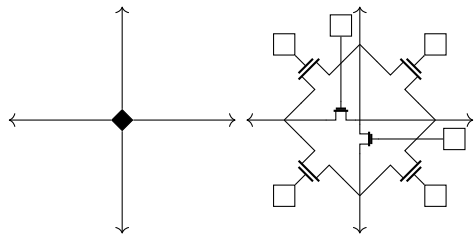

- Switch Area calculation
  - Num. switches = $M * W * F_{cin} + N * W * F_{cout}$
  - Cbox Area = $(W*M*F_{cin}+W*N*F_{cout})*1000\lambda^2$
  - At 10%, $W * (M + N) * 100\lambda^2$
- Wiring cost is a function of $W * M$ dictated by $10\lambda$ pitch
  - $(W * 10\lambda) * ((M + N) * 10\lambda)$
  - $W * (M + N) * 100\lambda^2$

# Switch-box Architecture (sbox)



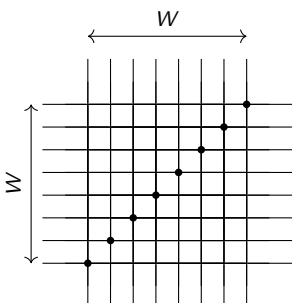

- Switch Area
  - Switch count $= W \times F_s \times 2$
  - Switch area $= W * F_s * 2 * 1000\lambda^2 = W * 6000\lambda^2$
- Wiring cost is a function of $O(W^2)$ dictated by $10\lambda$ pitch
  - $(W * 10\lambda) * (W * 10\lambda)$
  - $W^2 * 100\lambda^2$
- wire dominated at $W > 60$

# Diamond Crosspoint



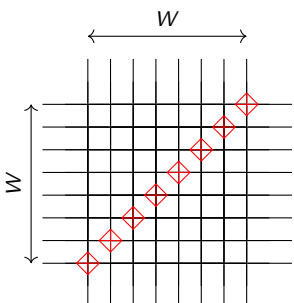
- A diamond switchbox works for bidirectional wires
- Six possible links (fanout is also allowed)
- Need a new SRAM cell for each link
- Hence, $A = 6 * A_{sw} = 6 * (A_{Config} + A_{Mux})$

# Understanding diamond crosspoints (Xilinx XC4000s)



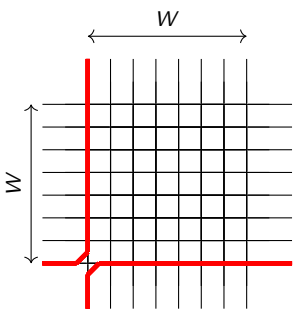

- Each diamond crosspoint (switch) is actually 6 transistors + 6 SRAM cells (one per wire)
- If we have $W$ channels, we need one diamond crosspoint per wire
- Naive crossbar will have $W^2$ switches, but we only have $W$ switches
- Same crosspoint can route two independent routes!

# Understanding diamond crosspoints (Xilinx XC4000s)



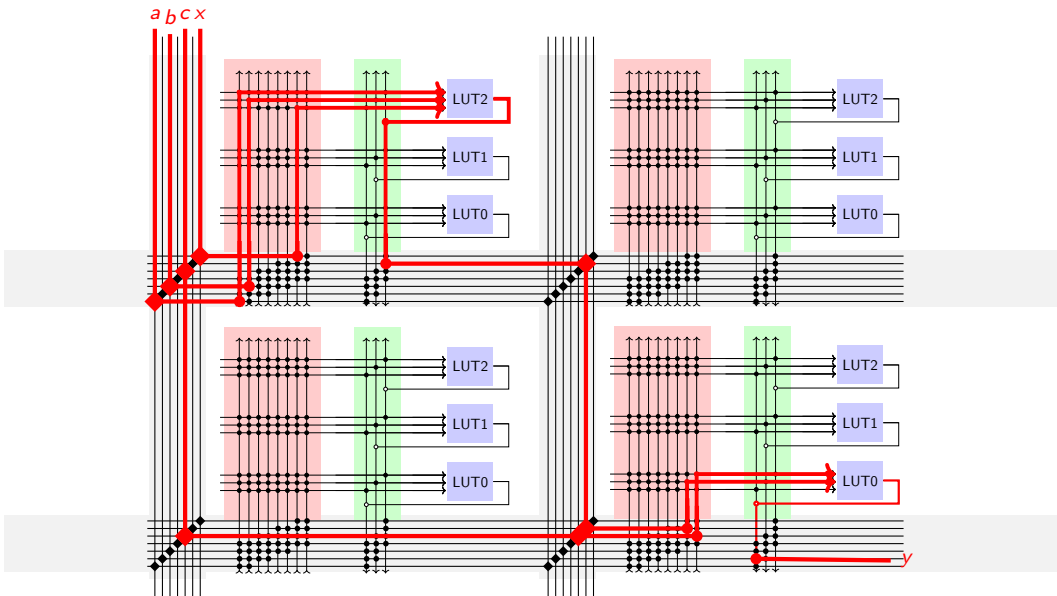

- Each diamond crosspoint (switch) is actually 6 transistors + 6 SRAM cells (one per wire)
- If we have $W$ channels, we need one diamond crosspoint per wire
- Naive crossbar will have $W^2$ switches, but we only have $W$ switches
- Same crosspoint can route two independent routes!

# Understanding diamond crosspoints (Xilinx XC4000s)




- Each diamond crosspoint (switch) is actually 6 transistors + 6 SRAM cells (one per wire)
- If we have $W$ channels, we need one diamond crosspoint per wire
- Naive crossbar will have $W^2$ switches, but we only have $W$ switches
- Same crosspoint can route two independent routes!

# Putting it together (LUTs + Cluster + Global)



# Relative Areas

- Logic Area (Assume $A_{Config} + A_{Mux} = 1000\lambda^2, A_{FF} = 0$)
  - $N * A_{kLUT} = N * (2^k * A_{Config} + 2^k * A_{Mux})$
  - $(2^k * N) * 1000\lambda^2$
- Interconnect Area (Assume $A_{Config} + A_{Mux} = 1000\lambda^2$)
  - Intra-Cluster Area
    - Output Area $= k * N * (N - k + 1) * A_{sw}$
    - Input Area $= k * N * M * A_{sw}$
    - $k^2 * N * (N - k + 1 + M) * 1000\lambda^2 \approx N^2 * \frac{k^2}{2} * 1000\lambda^2$
  - Cbox Area (Assume $F_{cin} = F_{cout} = 10\%$)
    - Switch Area $= M * W * F_{cin} * A_{sw} + N * W * F_{cout} * A_{sw}$
    - Wiring Area $= (W * 10\lambda) * ((M + N) * 10\lambda)$
    - $\max(W * (M + N) * 100\lambda^2, W * (M + N) * 100\lambda^2) \approx W * \frac{k}{2} * N * 100\lambda^2$
  - Sbox Area (Assume $F_s$=3)
    - Switch Area $= W * F_s * 2 * 1000\lambda^2$
    - Wiring Area $= (W * 10\lambda) * (W * 10\lambda)$
    - $\max(W * 6000\lambda^2, W^2 * 100\lambda^2) \approx W * 6000\lambda^2$

# Relative Areas

- Logic Area $= (2^k * N) * 1000\lambda^2$
- Interconnect Area
  - Intra-Cluster Area $= N^2 * \frac{k^2}{2} * 1000\lambda^2$
  - Cbox Area $= W * \frac{k}{2} * N * 100\lambda^2$
  - Sbox Area $= W * 6000\lambda^2$

# Relative Areas (Assume $k = 4$)

- Logic Area $= (16 * N) * 1000\lambda^2$
- Interconnect Area
  - Intra-Cluster Area $= 8 * N^2 * 1000\lambda^2$
  - Cbox Area $= 4 * W * N * 100\lambda^2$
  - Sbox Area $= W * 6000\lambda^2$

# Relative Areas (Assume $N = 8$)

- Logic Area $= 128 * 1000\lambda^2$
- Interconnect Area
  - Intra-Cluster Area $= 512 * 1000\lambda^2$
  - Cbox Area $= 32 * W * 100\lambda^2$
  - Sbox Area $= W * 6000\lambda^2$

# Relative Areas (Assume $W = 60$)

- Logic Area $= 128 * 1000\lambda^2$
- Interconnect Area
  - Intra-Cluster Area $= 512 * 1000\lambda^2$
  - Cbox Area $= 192 * 1000\lambda^2$
  - Sbox Area $= 360 * 1000\lambda^2$

# Relative Areas (Relative Ratios)

- Logic Area $= 128 * 1000\lambda^2$ – 11%
- Interconnect Area
  - Intra-Cluster Area $= 512 * 1000\lambda^2$ – 46%
  - Cbox Area $= 192 * 1000\lambda^2$ – 17%
  - Sbox Area $= 360 * 1000\lambda^2$ – 23%

# Class Wrapup

- Interconnect is ≈90% of the FPGA area
- Building *models* of hardware is helpful in deriving insight on operation
- Often OK to waste some LUTs, to use wires more efficiently
- Analytical model shows that for most realistic chips, switch area dominates wiring area for reasonable values of $W$