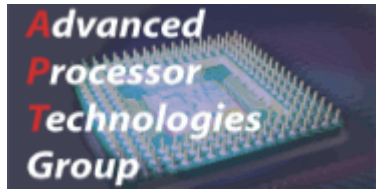



# A Low Latency Wormhole Router for Asynchronous On-chip Networks

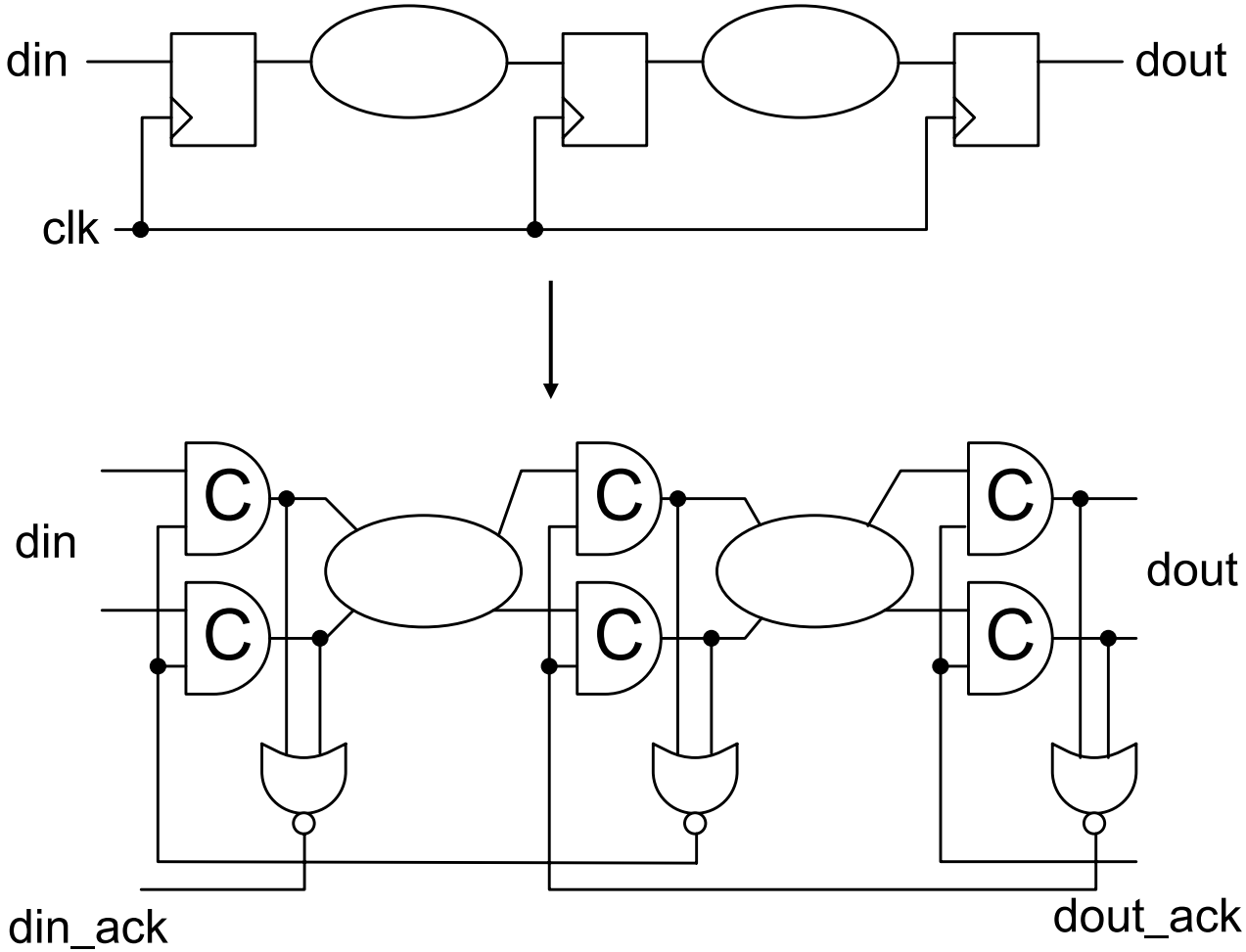
Wei Song and Doug Edwards

Advanced Processor Technologies Group (APT)
School of Computer Science
University of Manchester, UK

# Outline

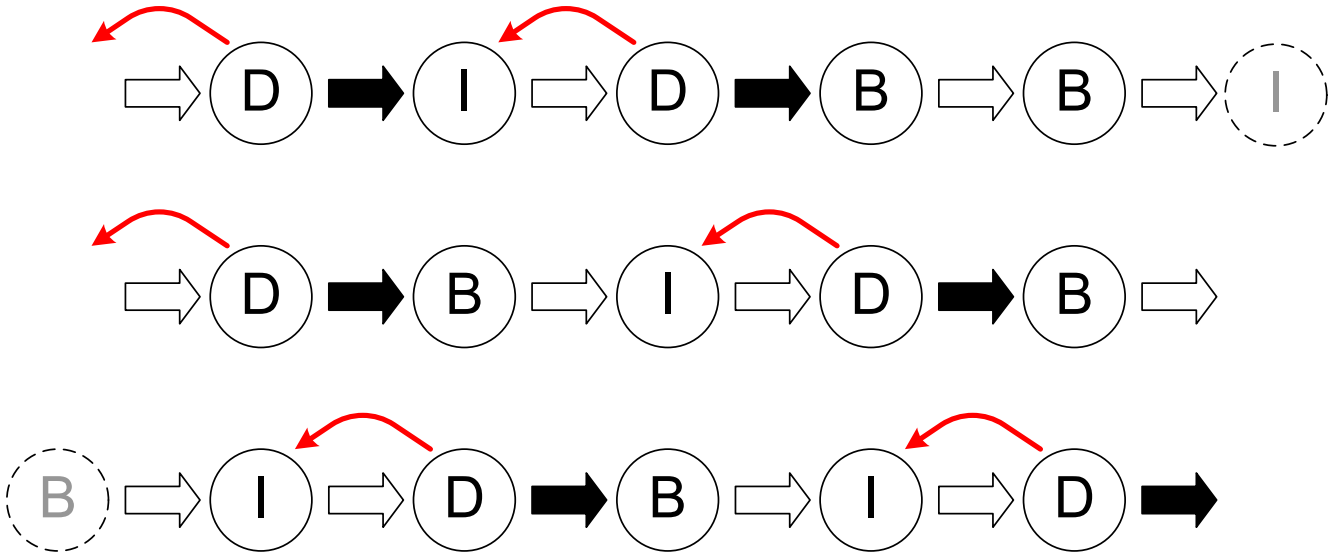
- Asynchronous On-chip Networks
  - Globally Asynchronous and Locally Synchronous (GALS)
  - Quasi Delay Insensitive (QDI) pipeline
  - *Target*: **general methods to improve speed**
- Solution
  - **Channel Slicing**
  - Using **Lookahead** pipeline on critical cycles
- Outcome
  - 32-bit wormhole router
  - **41.4%** latency reduction with **28.3%** area overhead

# QDI pipeline



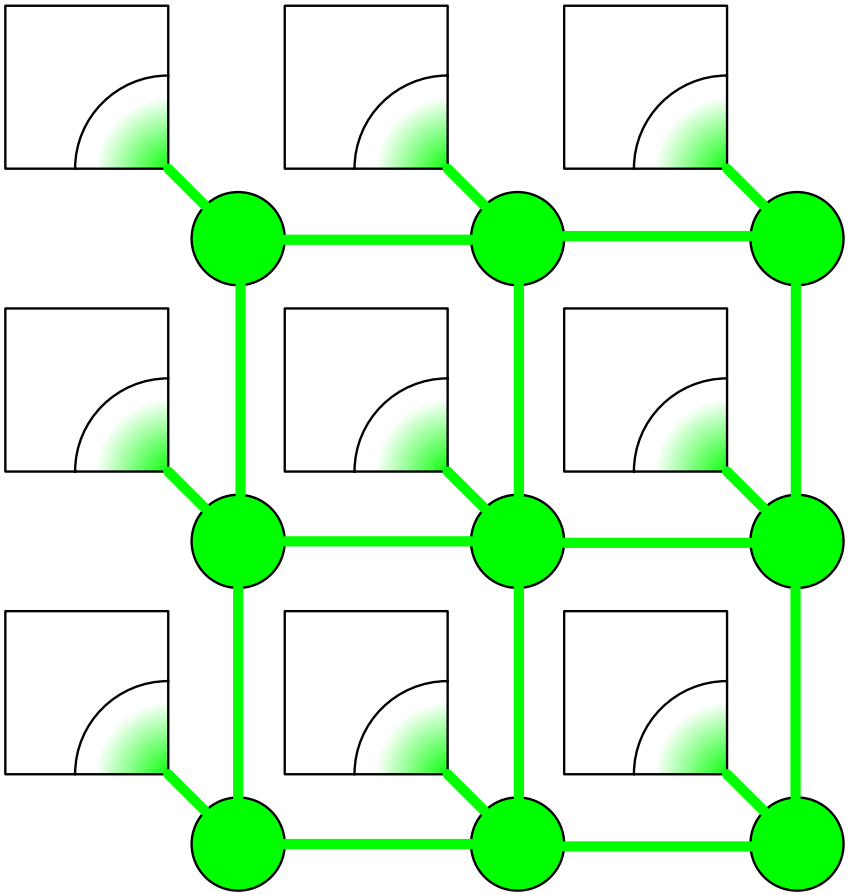

- Low Power
  - No clock tree
- Tolerance to Process Variation
  - Using delay insensitive handshakes

# Asynchronous Data Flow



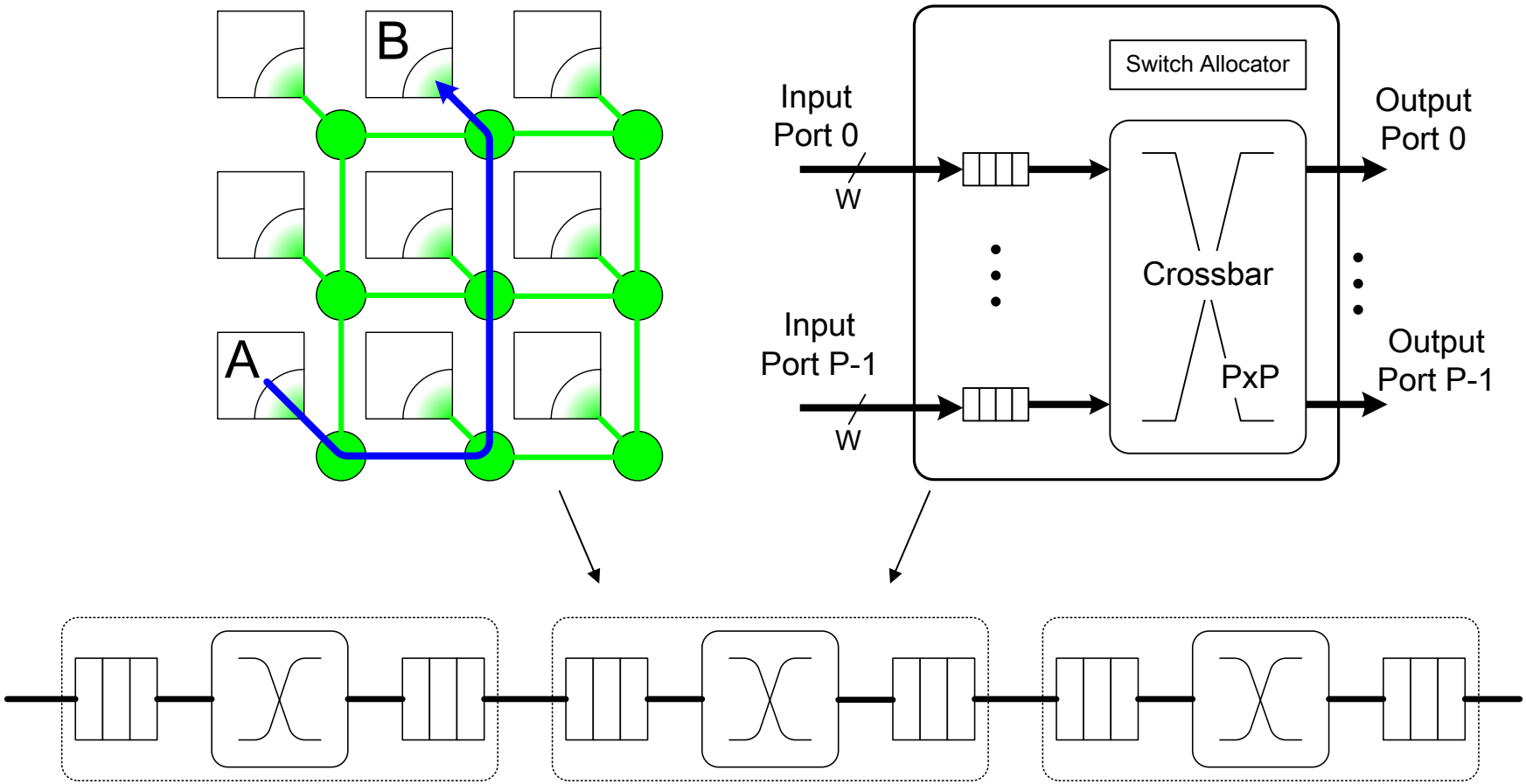

- One-hot coding
  - 01 0
  - 10 1
  - 00 idle, bubble
- Bubble propagation
- Critical cycle

# Asynchronous On-chip Network

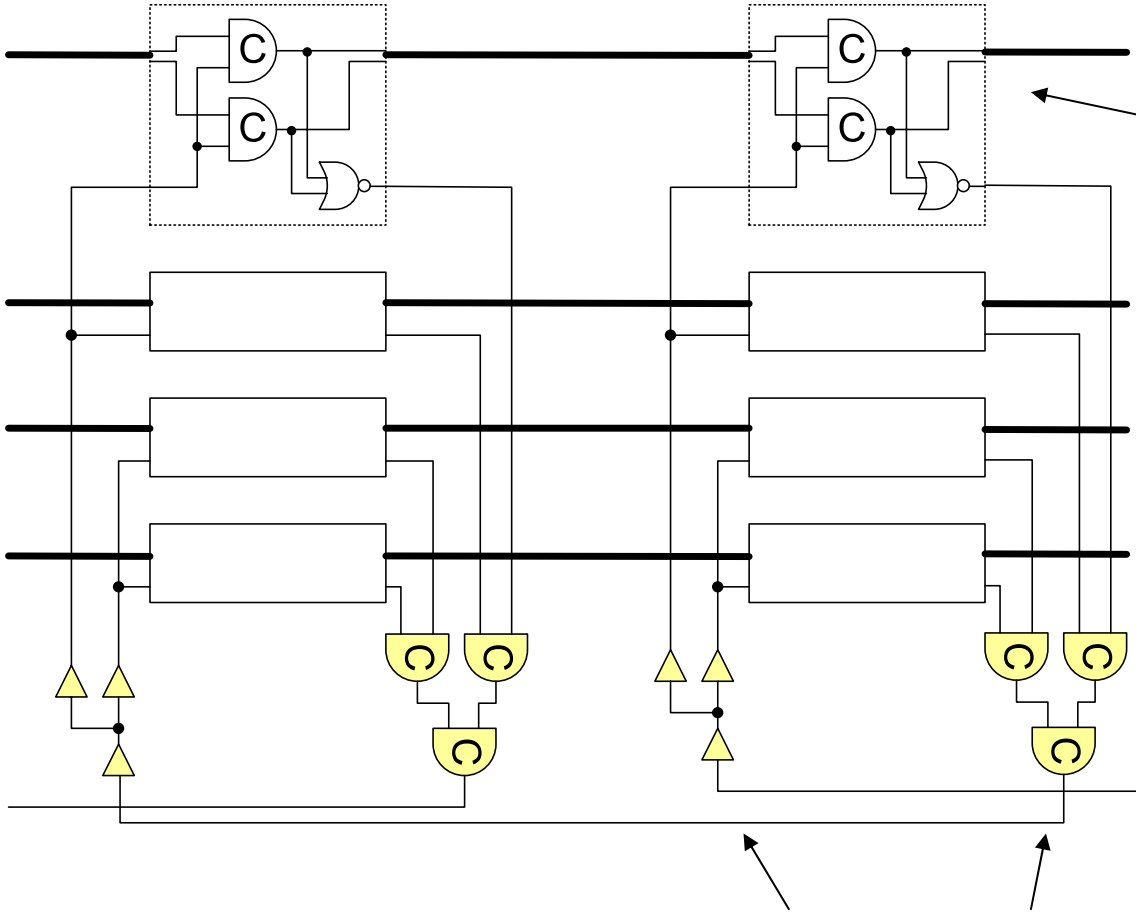


- NoC
  - Network-on-Chip
  - A scalable and distributed communication fabric
- GALS
  - Synchronous IP Blocks
  - Fully asynchronous routers

# Data-path Abstraction



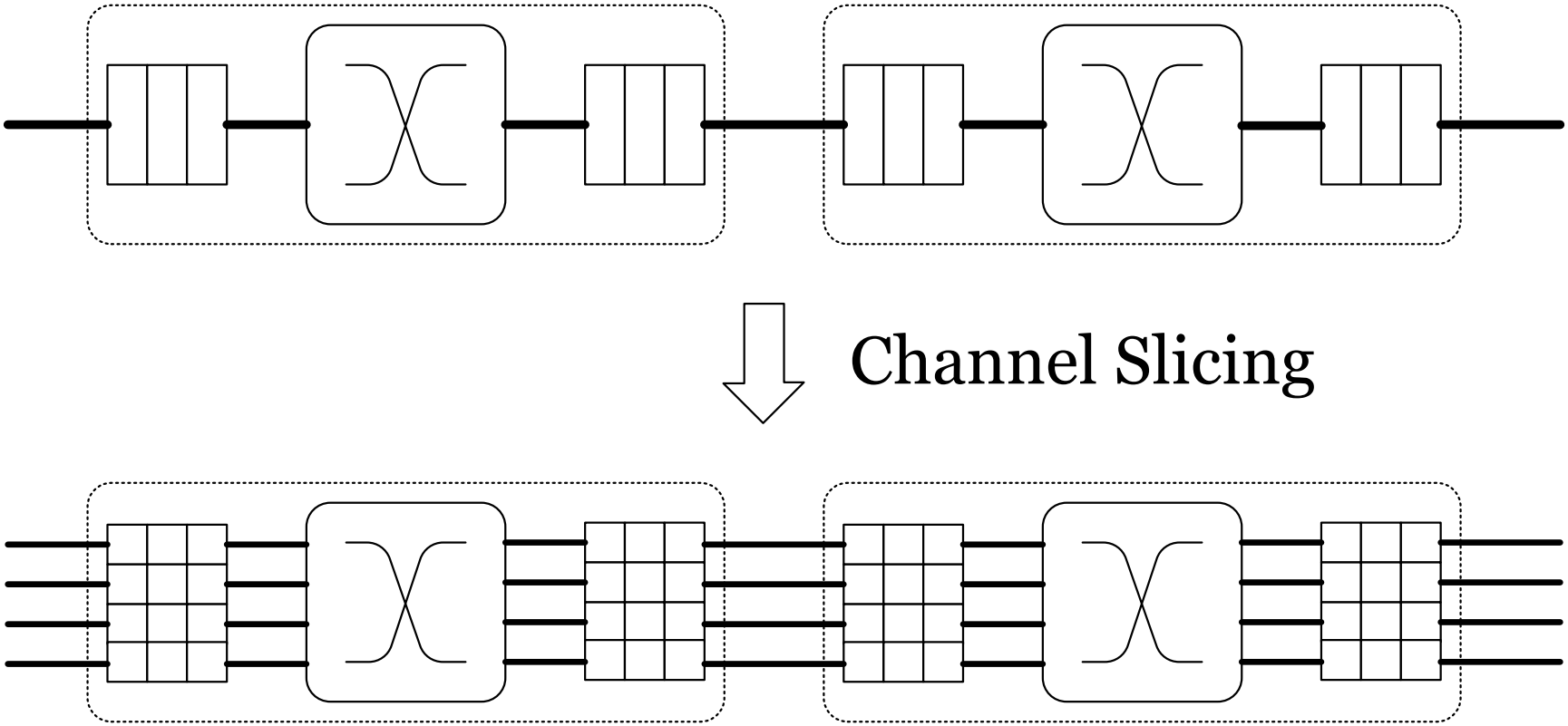
# Synchronized Pipeline Style



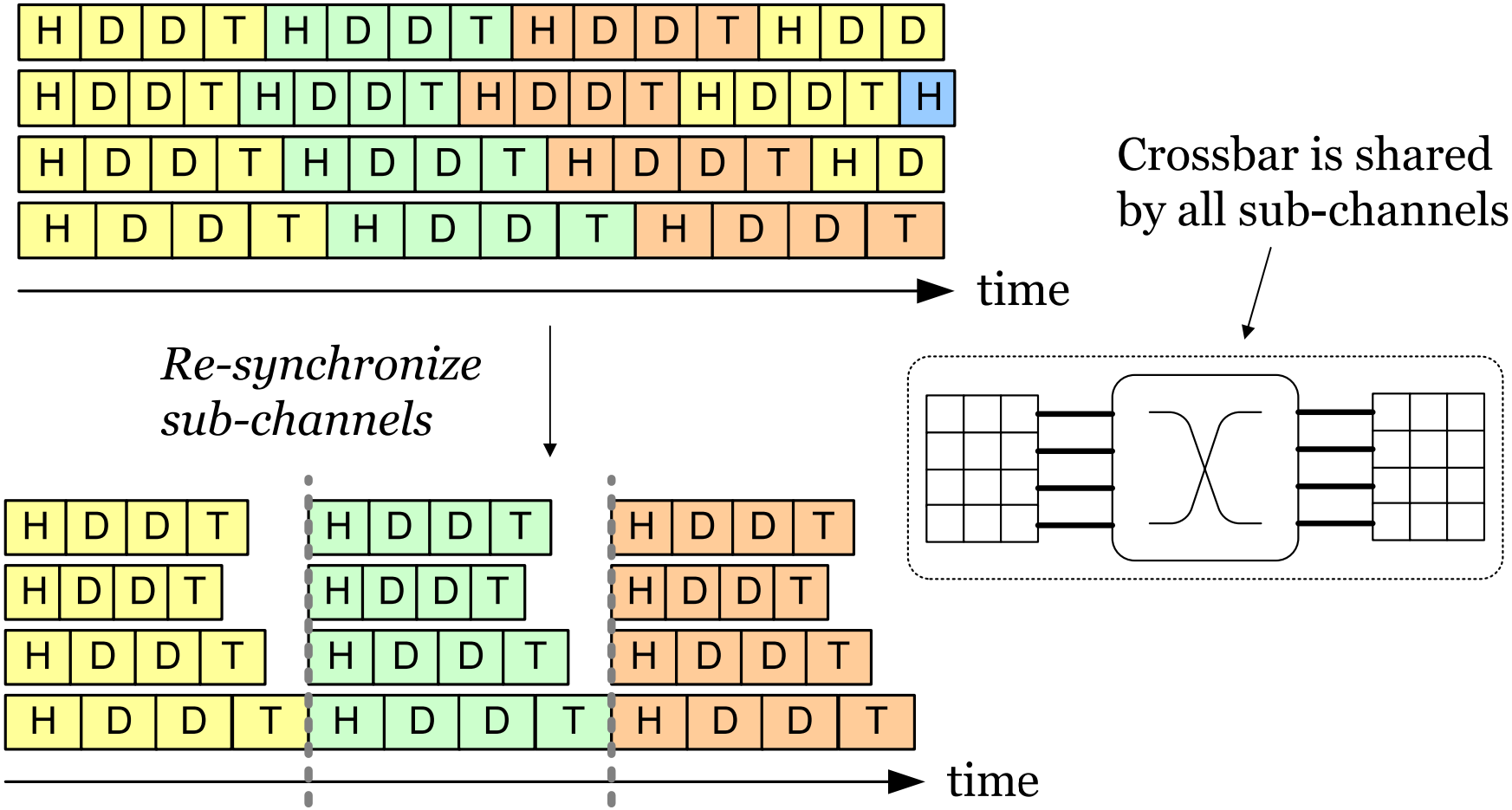

1-bit sub-channel

Extra latency overhead

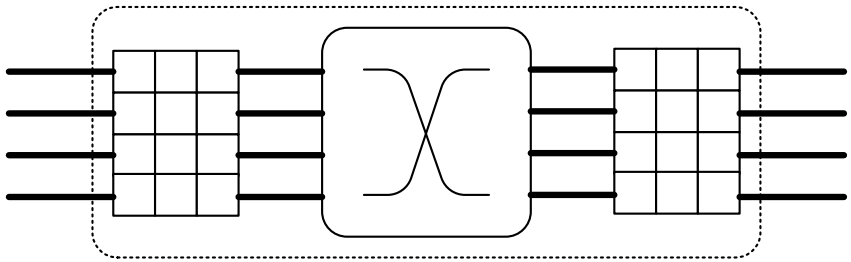
# Using Independent Sub-channels



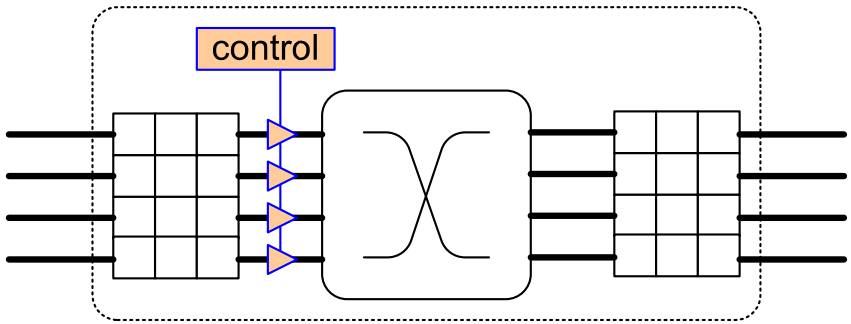
# Problem in Flow Control



# Solution: Re-synchronization

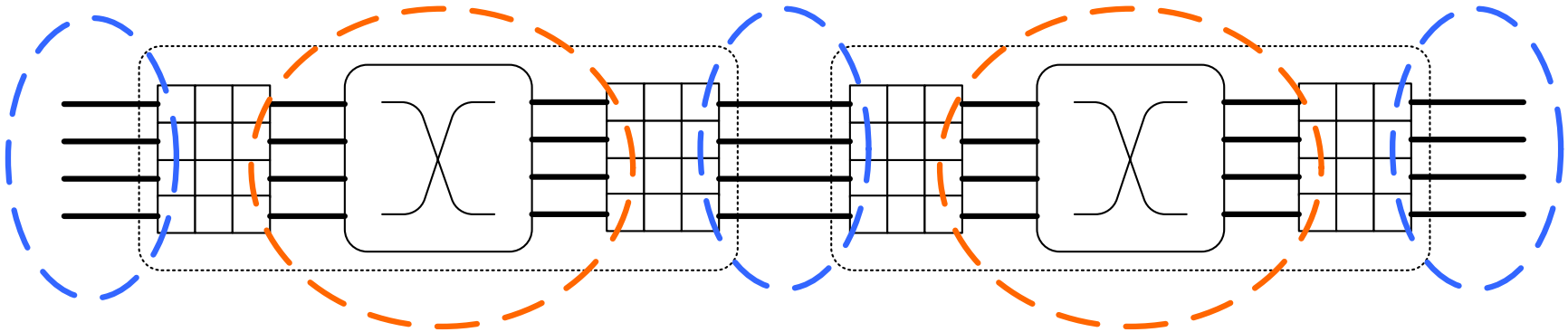


A sub-channel controller for each sub-channel




- Re-synchronize once per frame
- Algorithm:
  1. Wait for head flit
  2. Routing
  3. *Data transmission (**parallel**)*
  4. Tail detected
  5. Go to 1

# Critical Cycle Analysis



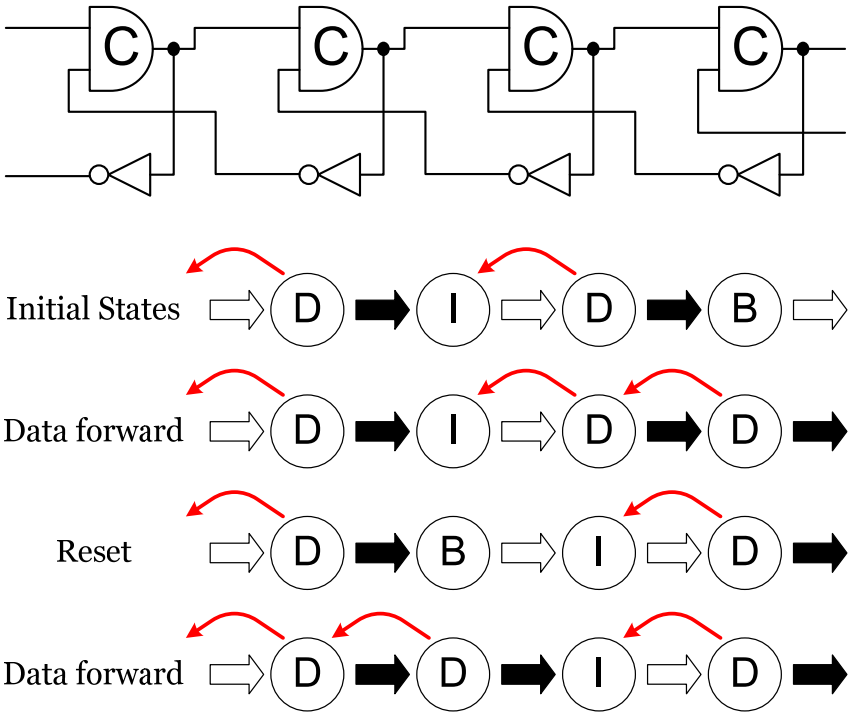
- Long interconnect
  - Buffer insertion
  - More pipeline stages
  - Wave-pipeline
- Crossbar
  - High fan-out
  - Routing control
  - Inside the router
  - **Critical cycle**

MANCHESTER 1824

The University of Manchester

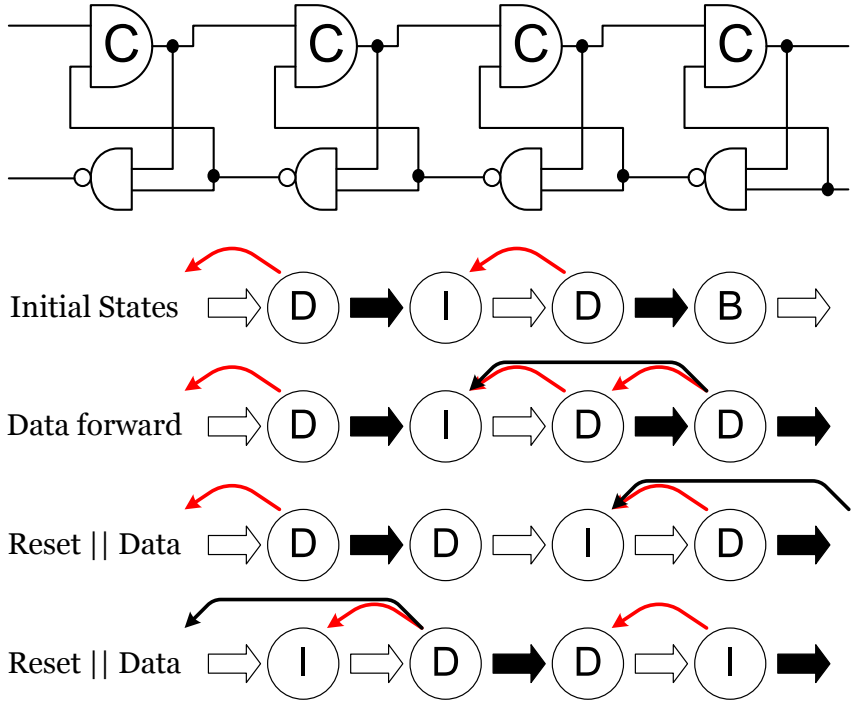
# Lookahead Pipeline

## Normal QDI pipeline



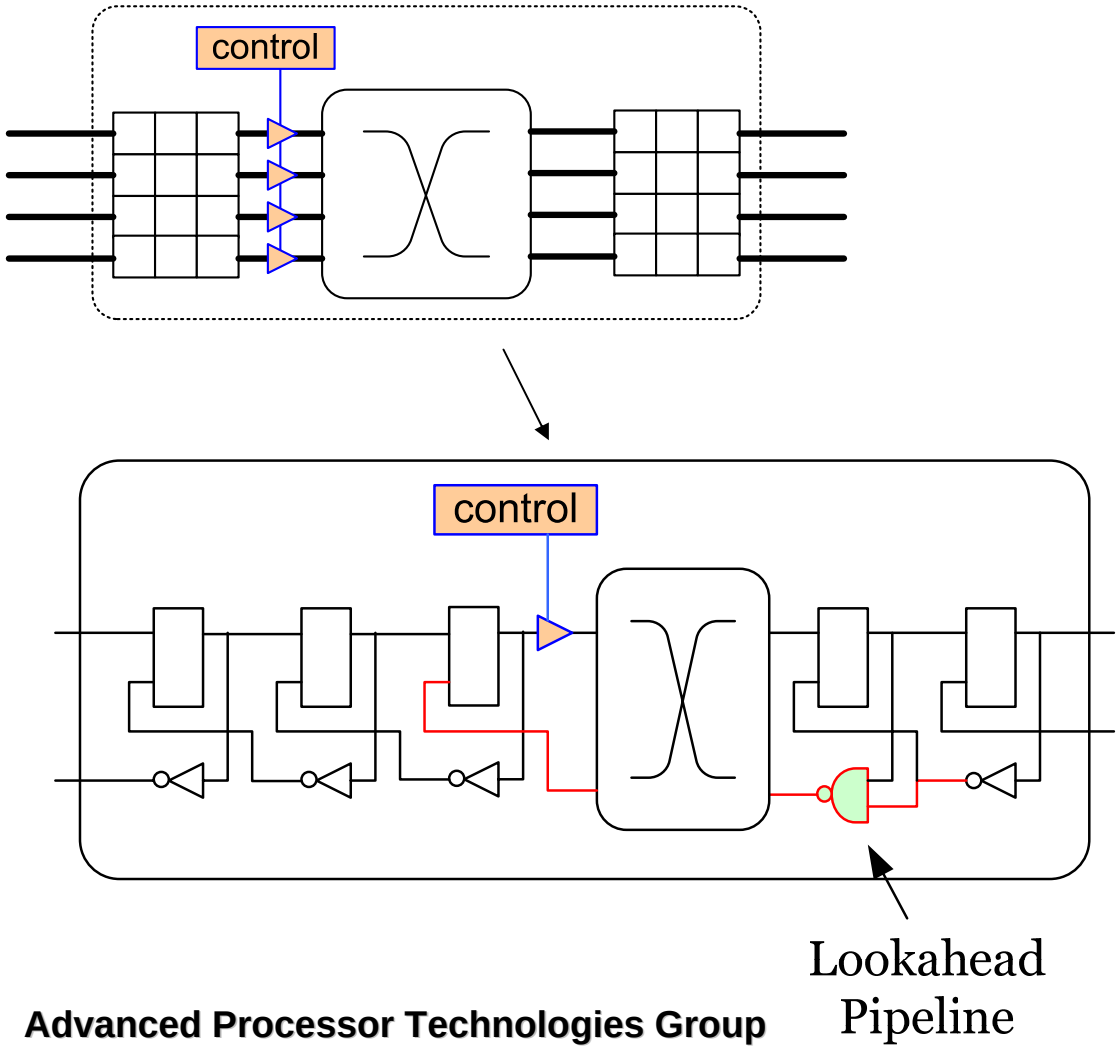

## Lookahead pipeline

[Montek Singh, 2007]




*1. Early acknowledge; 2. don't need an explicit bubble; 3. not strict QDI.*

# Using Lookahead in Router

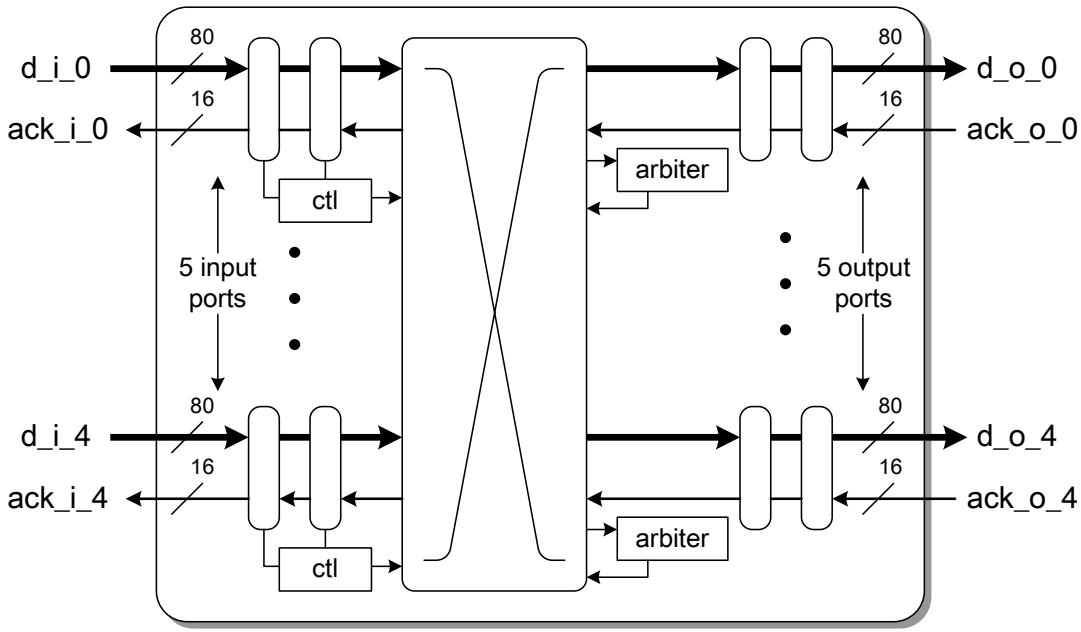



- Only utilized on the critical cycle.
- No significant area overhead.
- Timing assumptions are ensured.

MANCHESTER 1824

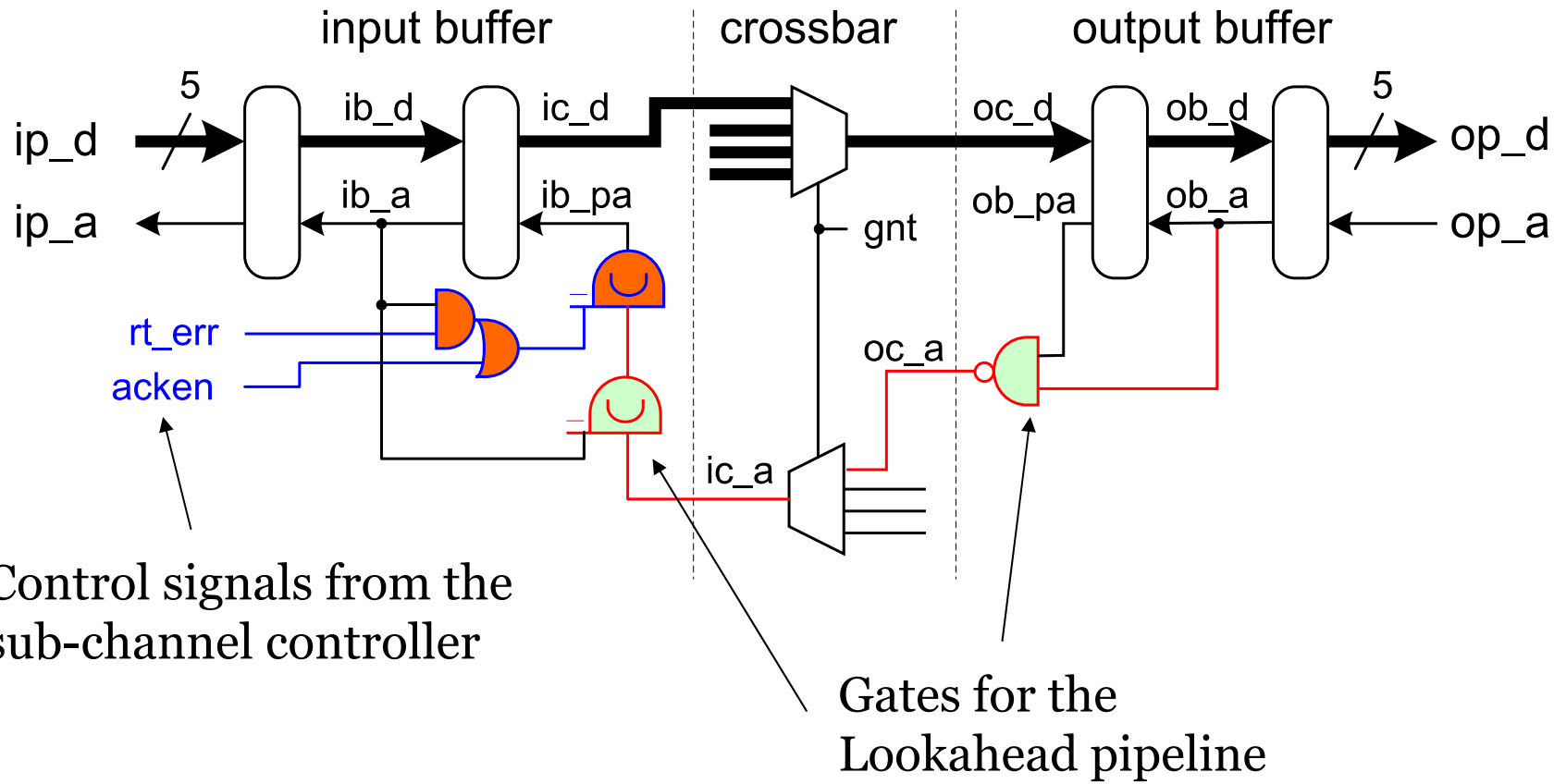
The University of Manchester

# A Wormhole Router Design

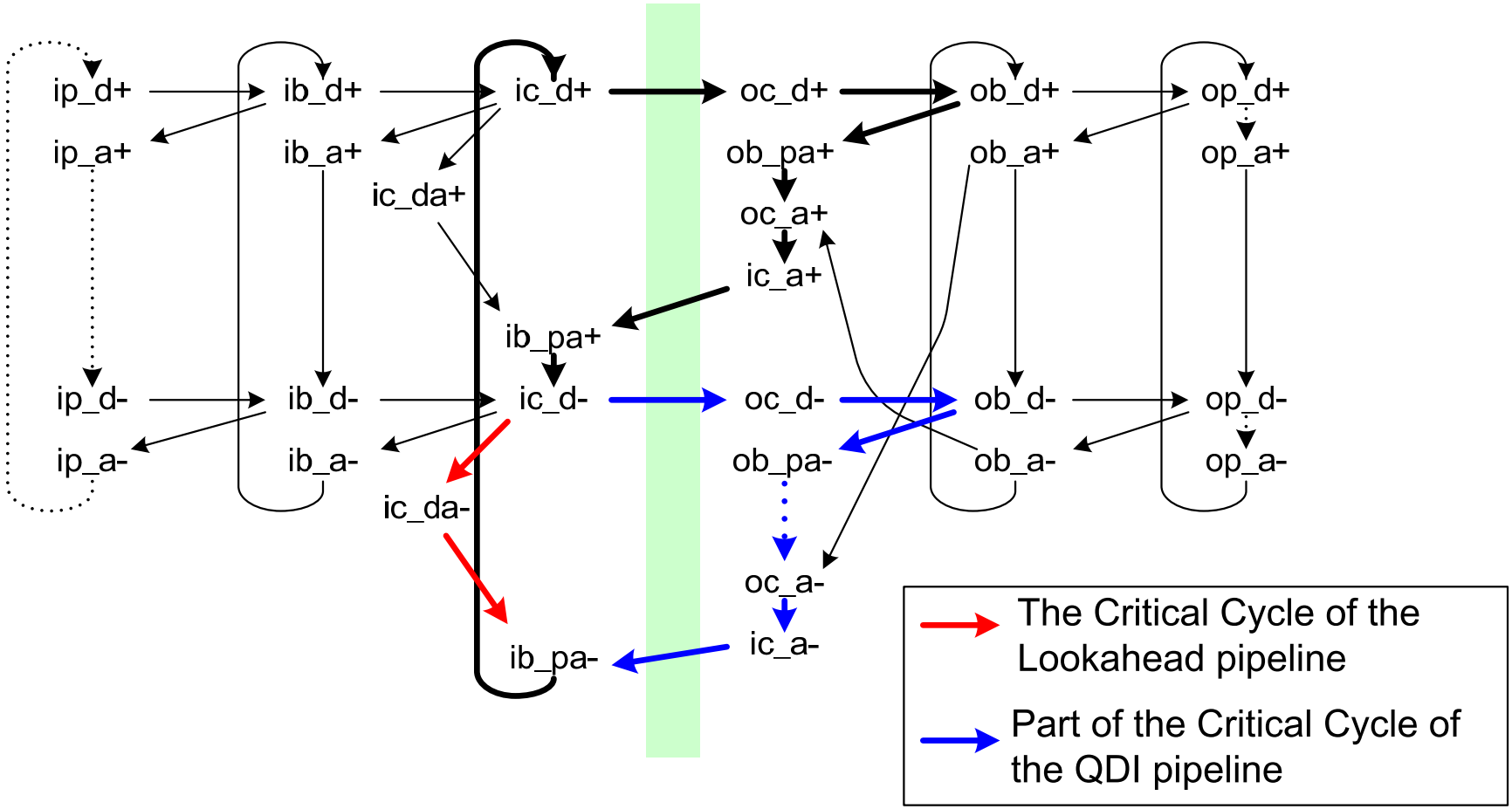



- 5-port router for the mesh topology
- 32-bit data-width
  - 16 1-of-4 sub-channels
- 2-stage input buffer
  - Control on the ack of the 2nd stage
- 2-stage output buffer
  - Make lookahead inside

# Data-path of a Sub-channel



# Latency Reduction Shown in STG



# Implementation and Simulation

- Verilog HDL netlists
  - Controller are generated from STGs using Petrify
  - Data-path are manually designed
- Implementation
  - Faraday Standard Cell Library using UMC 130nm technology
  - Synopsys DC + ICC + StarRC
- Simulation
  - Post-layout simulation with back-annotated latency from RC extraction
  - Typical corner (25 $^{o}$C, 1.2V)

# Speed Performance

| | CS + LH | ChSlice | Traditional |
|---|---|---|---|
| Cycle period | 1.7 ns | 2.2 ns | 2.9 ns |
| Router latency | 1.7 ns | 2.1 ns | 2.8 ns |
| Arbitration | 0.8 ns | 0.8 ns | 0.8 ns |

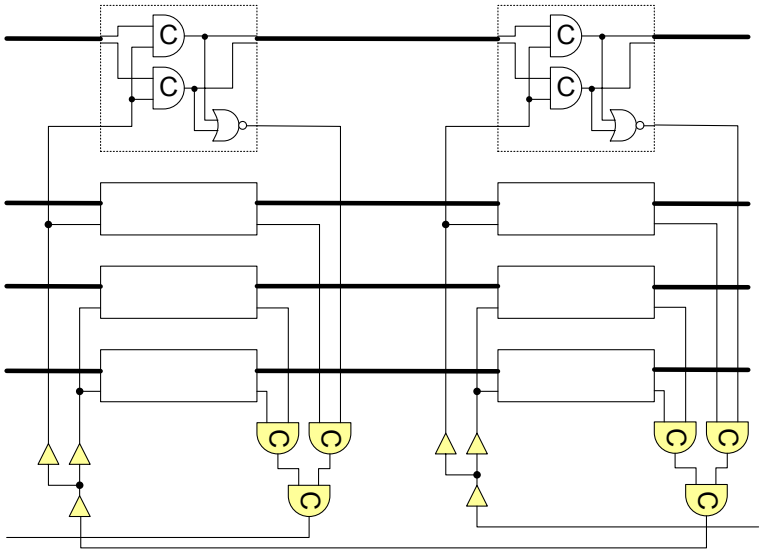
- Channel Slicing and Lookahead (CS+LH)
  - **590 MHz**, 41.4% cycle period reduction
- Channel Slicing only (ChSlice)
  - **450 MHz**, 24.1% cycle period reduction
- Traditional (without ChSlice or LH)
  - **345 MHz**

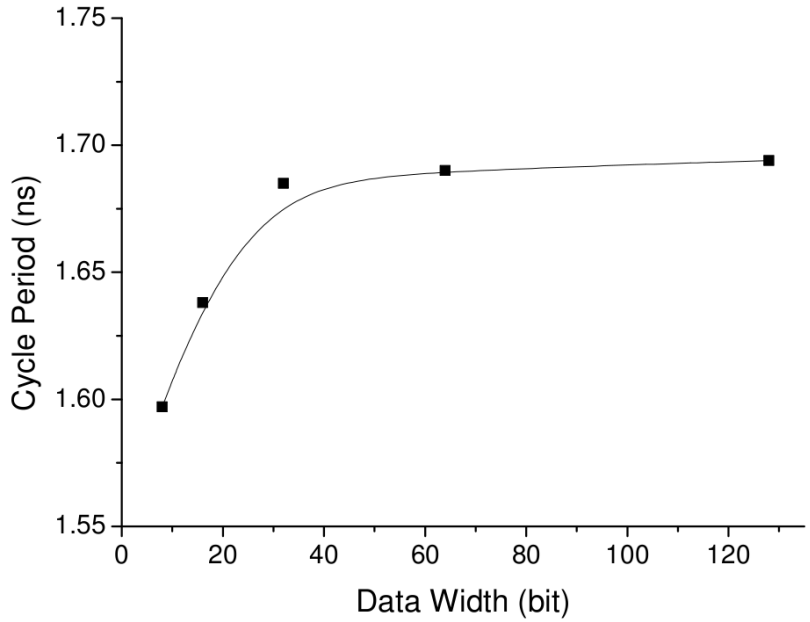
# Area Consumption

| | CS + LH | ChSlice | Traditional |
|---|---|---|---|
| Input Buffer | 6.2K | 5.8K | 4.3K |
| Output Buffer | 4.5K | 4.5K | 4.4K |
| Crossbar | 3.3K | 3.2K | 2.4K |
| Arbitration | 14.5K | 13.9K | 11.3K |

- Area in units of NAND2X1 Gate
- Channel Slicing 23.0% overhead
- Lookahead 5.3% overhead
- Total 28.3% overhead

# Data Width Effect




Cycle period increases when sub-channels are synchronized.

Cycle period is fixed when Channel Slicing is in use.

# Compare with Other Routers

| | Period | Tech | Special cell Library | Pipeline style |
|---|---|---|---|---|
| MANGO [2005] | 1.26 ns | 0.12 um | Unknown | Bundled-data |
| ANoC [2005] | 4 ns | 0.13 um | Yes | QDI |
| ASPIN [2008] | 0.88 ns | 90 nm | Custom | Bundled-data |
| QNoC [2009] | 4.8 ns | 0.18 um | Std cell | Bundled-date |
| **CS+LH [2010]** | 1.7 ns | 0.13 um | Std cell | Lookahead |

- Full standard cell design
- Delay insensitive, tolerance to process variation

# Conclusion

- QDI pipelines: low power and tolerant to process variation
- *Channel Slicing:* no C-element tree
- *Lookahead*: fast critical cycle.
- The wormhole router
  - 1.7 ns, 590MHz
  - 41.4% latency reduction with 28.3% area overhead

MANCHESTER 1824

The University of Manchester

# Thanks! Questions?

Contact info.

Wei Song

songw@cs.man.ac.uk