Presentation Outline

- **Motivation**
  - Parallel links limitations
  - Novel high-speed serial links

- **Link Architectures**
  - "Register-Pipelined" and "Wave-pipelined" parallel links
  - Single gate-delay serial link

- **Comparative study: parallel vs. serial**
  - Analytical models
  - Scalability
  - 65nm case study
Parallel link limitations

- Parallel links limitations
  - Constructed of multiple (N) wires and repeaters
  - Incur high leakage power
  - Occupy large chip area (routing difficulty)
  - Present a significant capacitive load
  - Buses have often low utilization and most of the time just leak (line drivers and repeaters)
Bit-Serial Interconnect

- Fewer lines, fewer line drivers and fewer repeaters
- Reduced leakage power
- Reduced chip area
- Better routability

**BUT**

- Should work \( N \) times faster!
Serial Link

- Standard serial links are very slow
- Hope lies in *novel serial links*
  - Data cycle of a few gate-delays (inverter FO4 delay)

- This work considers one of the fastest serial links
  - With single gate-delay data cycle ($d_4$)
Our target

- To show that novel serial link outperforms the parallel one for:
  - Long ranges
  - Advanced technology nodes
Method

● Choose
  – *Parallel* link implementation representatives
  – *Serial* link implementation representatives

● Compare the parallel and serial link approaches in terms of:
  – Area
  – Power
  – Latency
  – Technology scaling
"Register-Pipelined" Parallel Link

- Fully synchronous
- Interconnect as combinational logic between registers
- Source synchronous or global clock

High cost for high bit rates!
"Wave-Pipelined" Parallel Link

Bit rate is limited by relative skew of the link wires
Crosstalk Mitigation and Power Reduction

- Shielding / Spacing
- Staggered repeaters
- Interleaved bi-directional lines
- Asynchronous signaling
- Data encoding
- Data pattern recognition with special worst-case handling

- This work analyzes the two extremes of shielding:
  - Unshielded wires (a)
  - Fully-shielded wires (b)
Single Gate-Delay Serial Link

- Transition signaling instead of sampling
  - Two-phase NRZ Level Encoded Dual Rail (LED) asynchronous protocol, a.k.a. data-strobe (DS)
- Acknowledge per word instead of per bit
- Wave-pipelining over channel
- Differential encoding (DS-DE, IEEE1355-95)
- Low-latency synchronizers

R. Dobkin, et al., High Rate Wave-Pipelined Asynchronous On-Chip Bit-Serial Data Link, ASYNC07
R. Dobkin, Parallel vs. Serial On-Chip Communication, SLIP08
Analytical Models

Parallel and Serial Link Bit Rates

- Please refer to the paper for details on the exact analytical models employed in the work
Parallel Link Bit Rate Limitations (1)

A. Fastest available clock
   - Ring oscillator limitation: $8 \cdot d_4$
   - Fast processors: $11 \cdot d_4$ (e.g. CELL)
   - Standard SoC/ASIC: $100-400 \cdot d_4$

B. Synchronization Latency
   - May take several clocks in case of asynchronous clock relation

C. Clock uncertainty
   - Extended critical path
Parallel Link Bit Rate Limitations (2)

D. Delay Uncertainty
- The skew and jitter of the clock
- Repeater delay variations
- Wire delay variations
  - mostly metal thickness variations
- Via variations
- Cross-Coupling (Crosstalk)
- Geometry
  - Outcome of routing congestion and multi-layer structure
Parallel Link Minimal Clock Cycle (1)

\[ T_{CLK} > 2 \cdot (\delta_{MAX} - \delta_{MIN}) + 4 \cdot \Delta_{CLK} + T_{SU} + T_{H} \]

Latest data clocking

Clock Uncertainty

Maximal Data Delay, \( \delta_{MAX} \)

Minimal Data Delay, \( \delta_{MIN} \)

Earliest data clocking


R. Dobkin, Parallel vs. Serial On-Chip Communication, SLIP08
Impact of Process Variations in Repeaters on Multi-Wire Delay Uncertainty

- **Variation types**
  - Random variations
    - closely placed devices
  - "Systematic" variations
    - location on the die

- **Relative skew ($\delta_{\text{MAX}} - \delta_{\text{MIN}}$)**
  - Repeaters in the same stage are highly correlated
  - Random variations are averaged out thanks to large repeater sizing
  - Systematic inter-stage variations are averaged out along the link

! Relative skew among the lines due to variations in repeaters is small

! Multi-wire delay uncertainty is dominated by **Cross-Coupling**
Parallel Link Minimal Clock Cycle (2)

- Minimal clock cycle:
  \[ T_{CLK} > 2 \cdot \Phi(L) + 4 \cdot \Delta_{CLK} + T_{SU} + T_H \]

- System clock limitation:
  \[ T_{CLK}^{PAR} = \max \{ 2 \cdot \Phi(L) + 4 \cdot \Delta_{CLK} + T_{SU} + T_H, T_{SYSTEM-CLOCK} \} \]

- Register-pipelined link:
  \[ T_{CLK}^{PAR} = T_{SYSTEM-CLOCK} \]
  - Distance between successive pipeline stages is affected by Delay Uncertainty

Worst case skew between two lines: Cross-coupling and wire variations

The rate is bounded by clock cycle rather than by the delay uncertainty

65nm example

Maximal Clock Generator Rate, 8 gate-delay cycle
Serial Link Bit Rate

- Skew due to transistor variations is neglected
  - much smaller than in parallel link
- Coupling factor is always known
  - LEDR encoding: there is only one transition per each transmitted bit
  - The skew is not affected by cross-coupling
    ➢ link delay is similar for all symbols
- Bit rate:

\[ B_{SER} = \frac{1}{d_4} \]
Scalability

- **Number of repeaters** (per millimeter) grows for more advanced technology nodes
- **Active area and leakage**: Minimal link length for serial link employment decreases with technology
- **Dynamic power**: Minimal link length for serial link employment decreases with technology
- **Interconnect area**: Serial link is always preferable

Number of repeaters grows with technology node scaling

Equal throughput Parallel and Serial links are assumed

Y.I. Ismail, et al., Repeater Insertion in RLC Lines for Minimum Propagation Delay, ISCAS99
R. Dobkin, Parallel vs. Serial On-Chip Communication, SLIP08
65nm Case Study
Goals and Set-up

- **Compare**
  - Wave-pipelined (shielded/unshielded) vs. Serial
  - Register-pipelined (shielded/unshielded) vs. Serial

- **In terms of:**
  - Area
  - Power
  - Latency
  - Length

- **All links deliver the same bandwidth**
  - $B_{SER}$ – the bandwidth of single serial link
Parallel Link Width for Equivalent Throughput

- Note impractical widths for:
  - Unshielded WP over 6mm
  - RP operating with clock cycle greater than $130 \cdot d_4$

Wave-Pipelined (WP) link width

- Maximal Width (128 Lines)
- Fully-Shielded ($8d_4$ Clock, $N=8$)
- Unshielded

Register-Pipelined (RP) link width

- Maximal Width (128 Lines)
- Equivalent Width
Wave-Pipelined Link vs. Serial Link: Active Area and Leakage Comparison

- Parallel is better
- Serial is better
- Same Area / Leakage
- Unshielded: Impractical
Wave-Pipelined Link vs. Serial Link: Total Area Comparison (Incl. Interconnect)

Serial is always better

Unshielded: Impractical

R. Dobkin, Parallel vs. Serial On-Chip Communication, SLIP08
Register-Pipelined Link vs. Serial Link: Active Area and Leakage Comparison

Serial is always better
Register-Pipelined Link vs. Serial Link: Total Area Comparison (Incl. Interconnect)

Serial is always better.
Wave-Pipelined Link vs. Serial Link: Dynamic Power Comparison

- Serial is better
- Too wide parallel link
- Fully-Shielded
- Unshielded: Impractical

Parallel is better

Impractical: Too wide parallel link

Serial is better >3mm
Wave-Pipeliined Link vs. Serial Link: Total Power Comparison

20% Utilization

Unshielded: Impractical

Fully-Shielded

Unshielded
Register-Pipelined Link vs. Serial Link: Dynamic Power Comparison

High Area penalty
Register-Pipelined Link vs. Serial Link: Total Power Comparison

20% Utilization

Total Power Ratio Parallel / Serial

Length [mm]

Unshielded, T=10d₄

Unshielded, T=130d₄

Fully-Shielded, T=10d₄

Fully-Shielded, T=130d₄

R. Dobkin, Parallel vs. Serial On-Chip Communication, SLIP08
## Test Case Summary

*Minimal length above which the serial link is preferred*

<table>
<thead>
<tr>
<th>Shielding</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully Shielded</td>
<td>Unshielded</td>
<td>Fully Shielded</td>
</tr>
<tr>
<td>Unshielded</td>
<td></td>
<td>Unshielded</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Length of parallel link</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>unlimited</td>
<td>up to 6mm</td>
<td>unlimited</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clock cycle of parallel link</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>$8d_4$</td>
<td>$8d_4$</td>
<td>$10d_4$ (fast)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$130d_4$ (slow)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$10d_4$ (fast)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$130d_4$ (slow)</td>
</tr>
</tbody>
</table>

To minimize the following: choose a serial link for links longer than:

<table>
<thead>
<tr>
<th>Area</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always</td>
<td>Always</td>
<td>Always</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 mm</td>
<td>4mm</td>
<td>3mm</td>
</tr>
<tr>
<td>3mm</td>
<td>3mm</td>
<td>1mm</td>
</tr>
<tr>
<td>3mm</td>
<td></td>
<td>3mm</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latency</th>
<th>Wave-Pipeline vs. Serial</th>
<th>Register-pipelined vs. Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 mm</td>
<td>Never*</td>
<td>4mm</td>
</tr>
<tr>
<td>4mm</td>
<td>12mm</td>
<td>2mm</td>
</tr>
<tr>
<td>9mm</td>
<td></td>
<td>3mm</td>
</tr>
</tbody>
</table>
Conclusions

- Novel high-speed *serial links* outperform *parallel links* for long range communication
- The *serial link* is more attractive for *shorter ranges* in *future technologies*
- Future large SoCs and NoCs should employ *serial links* to mitigate:
  - Area
  - Routing Congestion
  - Power
  - Latency
Thank You!

Questions?