OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
EIT
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

Networks-on-Chip for heterogeneous 3D-Systems-On-Chip

Dissertation from Dr. Jan Moritz Joseph
Supervisor: Prof. Thilo Pionteck. Otto-von-Guericke-Universität Magdeburg, Institute for Information Technology und Communications, Germany

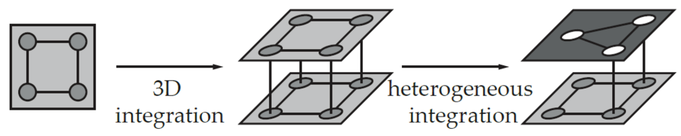
Introduction

Motivation:

Heterogeneous 3D integration: stacking of dies in disparate technologies

Need for **non-uniform routers** in 3D NoCs:

- **System level:** floorplan, topology, network synthesis
- **Architectural level:** router memory, router architectures, routing algorithms
- **Physical level:** design of links, power models
- **System simulation:** cover multiple abstraction levels, manifold design parameters



Modeling and Simulation

Motivation:

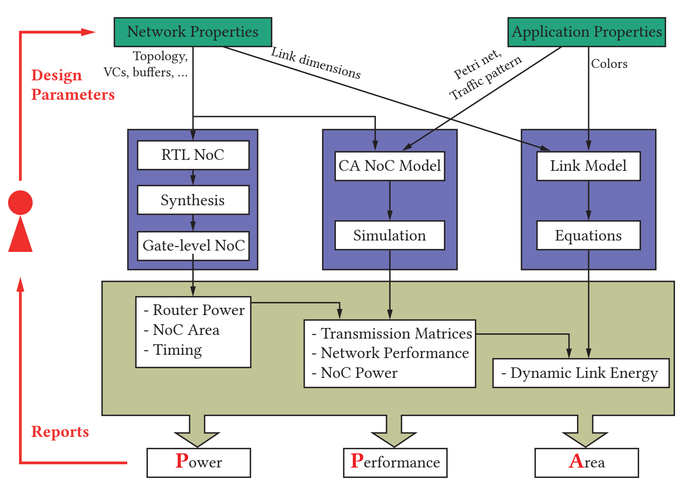
Lack of adequate models and simulation tools:

- **Need for full-stack simulation:** TLM benchmarks, cycle-accurate router models and bit-level accurate power estimations
- **Floorplan-awareness:** irregular 3D-topologies
- **Interfaces for novel design parameters:** per-VC and per-port buffer depths

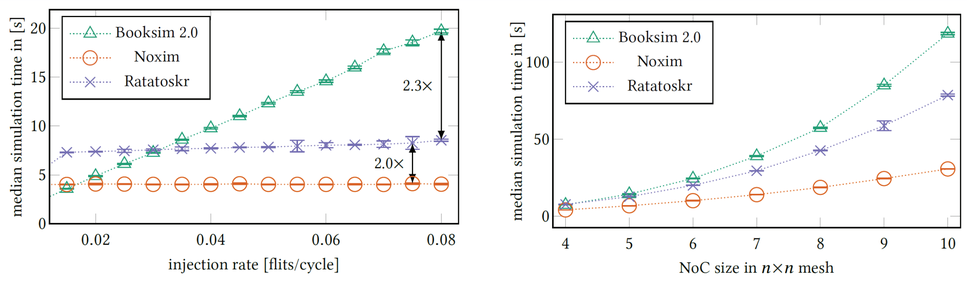
Methods and Tools:

Ratatoskr: design and simulation tool:

- **Heterogeneous 3D**
- Non-purely synchronous
- Any topology/floorplan
- Redistribution layers
- **Precise PPA analysis**
- Power estimation near bit-level accurate
- RTL NoC prototype
- Cycle-accurate NoC simulator
- TLM system benchmarks
- **Automated system generation**
- open-source:



Results:



- Dynamic link energy estimation **within 1% if bit-level simulations** at cycle-accurate simulation speed [PATMOS 2019, Integration 2019]
- Rigid **theroretical modelling** [ReCoSoC 2016, ReCoSoC 2017]
- **Rapid prototyping** and automated system design via interfaces [ReCoSoC 2018]
- **State-of-the-art performance** [arXiv 2020]

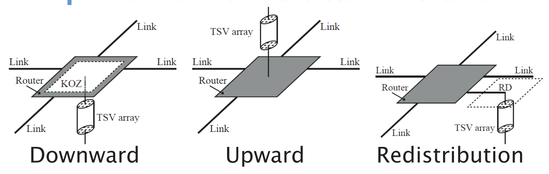
System-level optimization

Motivation:

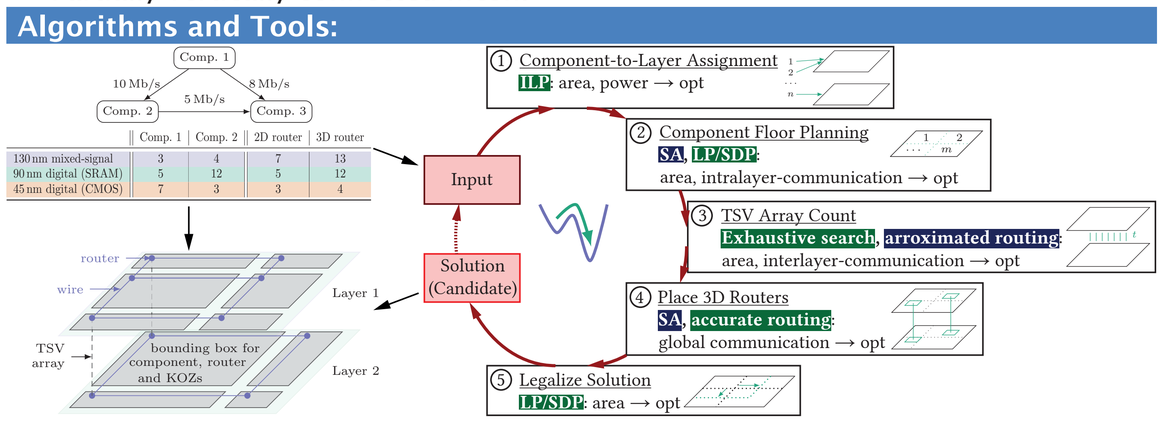
1. **PPA of routers and components vary per layer**
2. **Vertical interconnects with TSVs require area:** Keep-out-zones (KOZ) for Through silicon vias.
3. **Redistribution allows for more freedom in placement** of vertical links and routers.

Impact on system-level design:

- Integrated approach
- Models for KOZ and redistribution.
- Partially vertically connected mesh.



Algorithms and Tools:



Integrated mixed-approximate-optimal approach

Results:

- **12.4% - 12.9% better area using redistribution** from better placement over the same method without redistribution [ICCD 2019]
- **18.8% reduced whitespace due to better distribution of cores over layers** over conventional, manual placement w/o system-level consideration [ICCD 2019]
- **Up to 4.4% better communication** over conventional, manual placement w/o system-level consideration [ICCD 2019]

Architectural Optimization

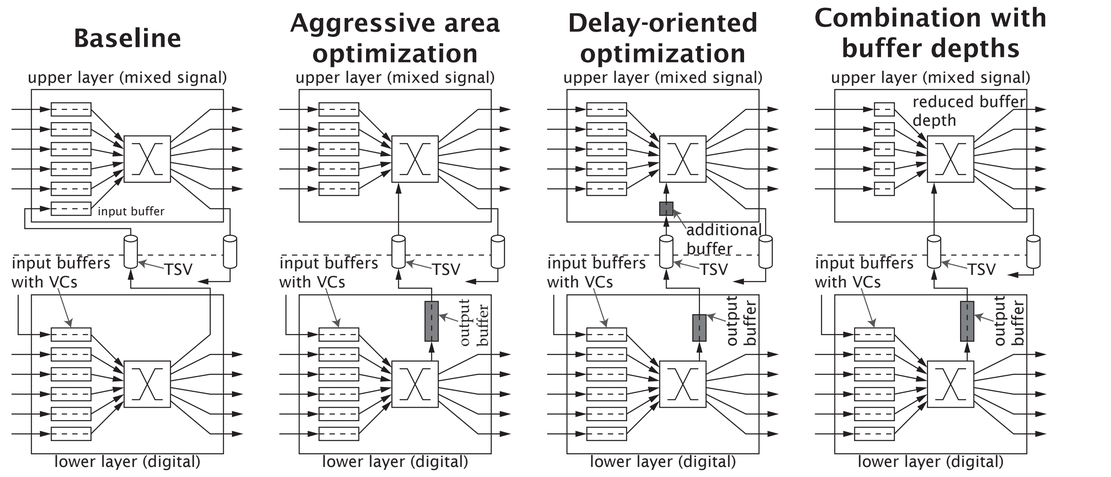
Optimization of Router Memory:

Motivation:

Different area and power costs of buffers between layers

- **Improve architectural parameters**: optimize *buffer depths*
- **New microarchitecture:** "spread" buffer over layer for better *buffer distributions*

Architecture:



Results:

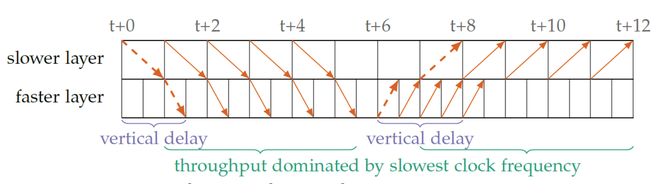
- **7.2% - 5.4% power savings** for aggressive and delay-oriented architecture vs. baseline in 130mn and 65mn nodes [MICPRO 2017]
- **Aggressive area optimization:** 9.6% area savings at 14% average performance loss [MICPRO 2017]
- **Delay-oriented optimization:** 8.3% area savings at 2.1% average performance loss [MICPRO 2017]
- **Combination with buffer depths:** 28% reduced area, 15% reduced power at 4.6% average performance loss [MICPRO 2017]

Optimization of Throughput and Latency:

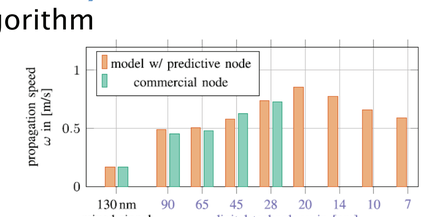
Motivation:

Different clock frequency and throughput between layers

- Co-Design of router architecture and routing algorithm



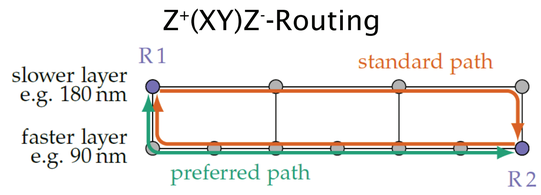
Throughput limitations



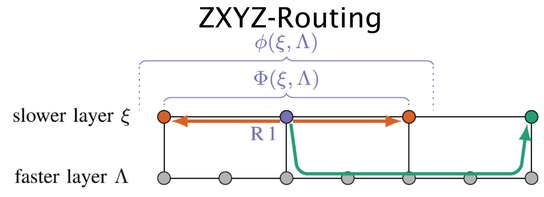
Transmission speed differences

Routing:

Z⁺(XY)Z⁻-Routing

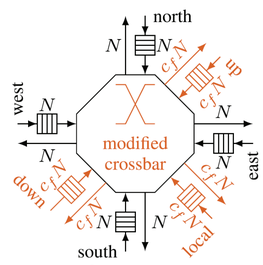


ZXYZ-Routing

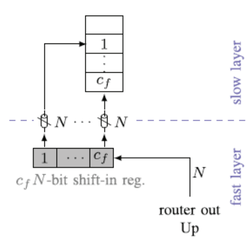


Architecture:

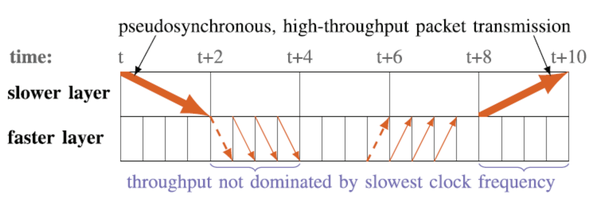
Pseudosynchronous high-vertical throughput router



Modified crossbar



High-throughput vertical links

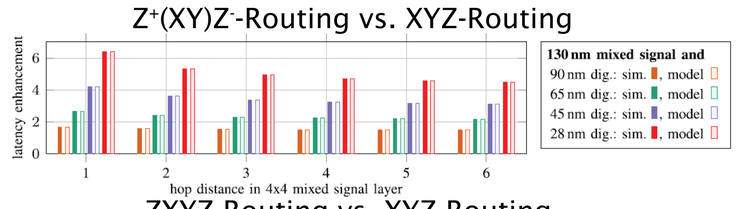


Exemplification of improved throughput

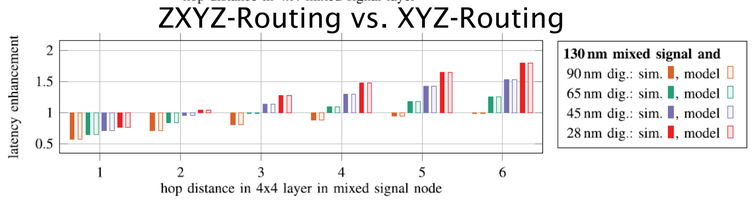
Results:

- **1.5x to 6.5x better latency** from Z⁺(XY)Z⁻-Routing
- **0.54x to 1.79x better latency** from ZXYZ-Routing
- **4x throughput increase at 10.6% area costs** from high-throughput router (45nm mixed signal node)

Z⁺(XY)Z⁻-Routing vs. XYZ-Routing



ZXYZ-Routing vs. XYZ-Routing



- **3D Vision SoC Case Study [Zárandy 2011]** (15nm digital, 45nm mixed signal): **2x throughput** increase, **2.26x average latency** speedup, **41.4% dynamic power** savings at **2.1% area** increase [IEEE Access 2019]

Conclusion

- **Asymmetric 3D NoCs:** novel design paradigm for NoCs in heterogeneous 3D SoCs
- **Complete tool-stack** for system design and simulation
- **Algorithms for system-level optimization**
- **Novel router microarchitectures optimizing buffers**
- Latency and throughput limitations can only be tackled via a **co-design of routing algorithms and router architectures**

Get the PDF!