TECHNISCHE UNIVERSITÄT DRESDEN

FAKULTÄT ELEKTROTECHNIK UND INFORMATIONSTECHNIK

Fast-Start SerDes Links for Neuromorphic Hardware Systems

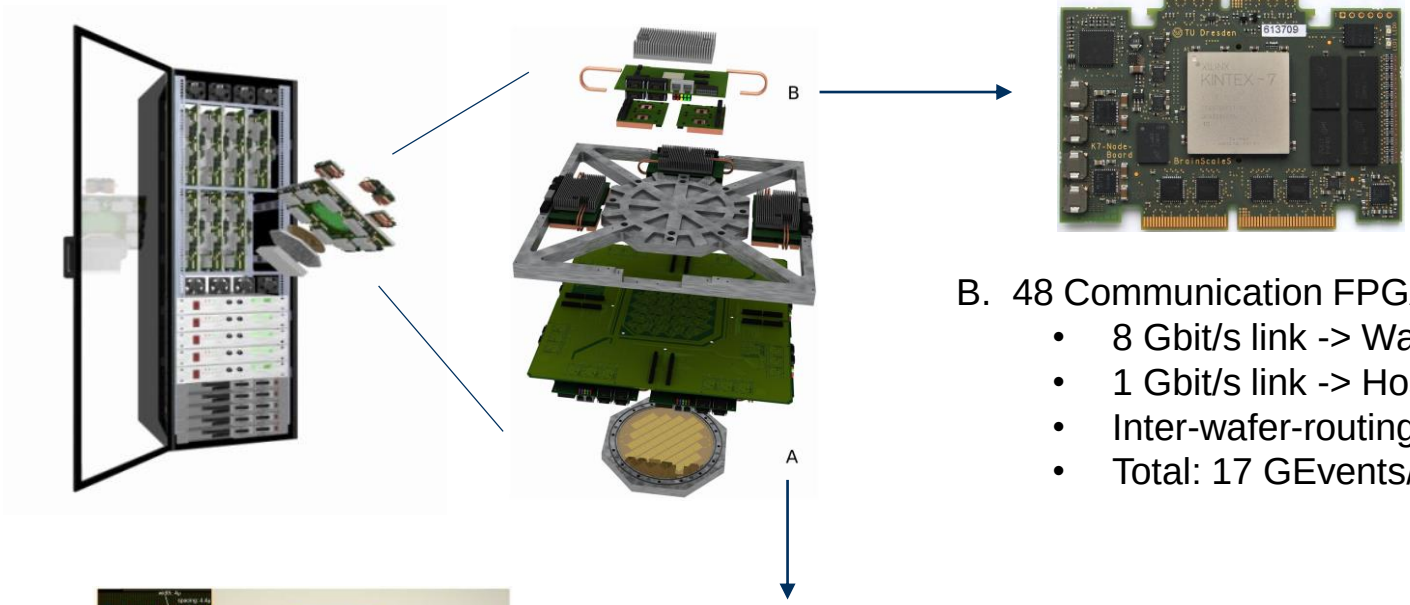
Presenter: Bitan Mallik
Institute: Principles of Electrical and Electronic Engineering
Chair: Highly-Parallel VLSI Systems and Neuro-Microelectronics

Erlangen, Thursday 11th Oct 2018

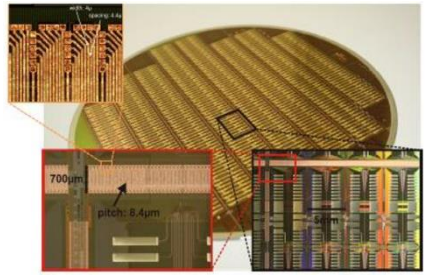
DRESDEN concept
Exzellenz aus Wissenschaft und Kultur

Motivation

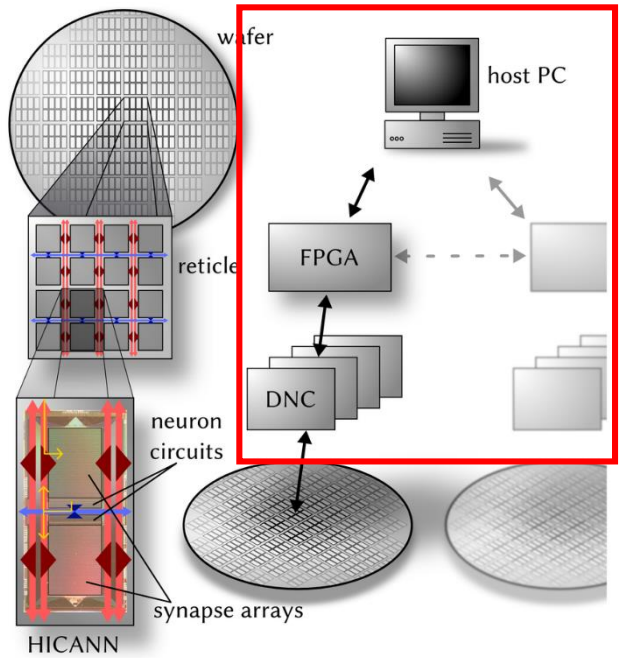
Neuromorphic Waferscale System: 1st Generation



B. 48 Communication FPGAs
- 8 Gbit/s link -> Wafer
- 1 Gbit/s link -> Host PC
- Inter-wafer-routing
- Total: 17 GEvents/s



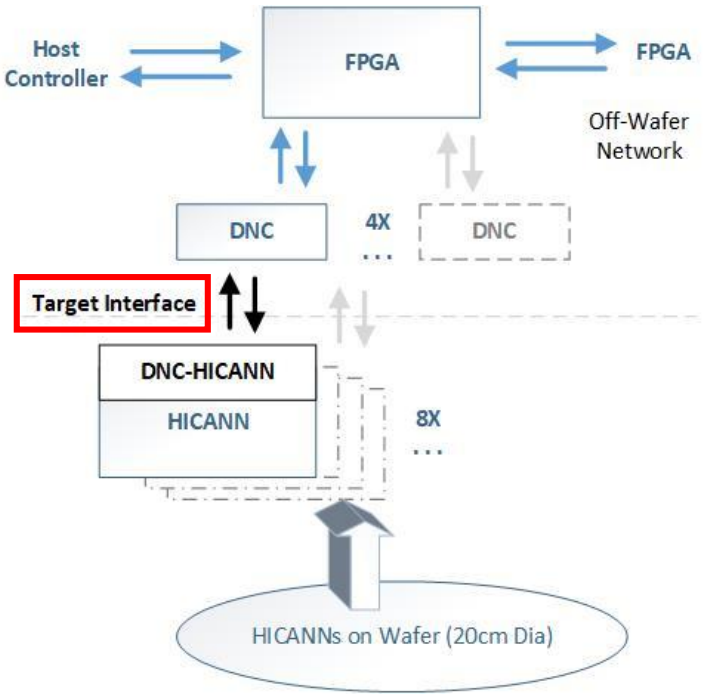
A. Wafer with 384 HICANNs
- 200000 neurons
- 50 million synapse
- Speed-up: 10^4



Off-Wafer Packet Routing Network

Introduction

Neuromorphic Waferscale System: 2nd Generation



Logical Off-Wafer Data Routing Network

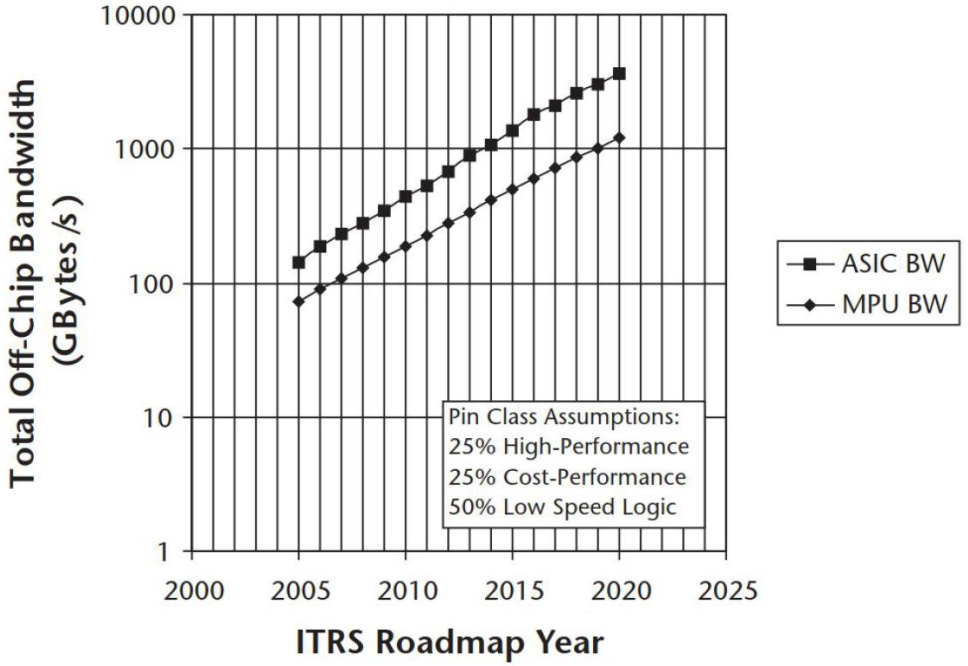
- Aim is to design a new **Serial link** for the next generation Neuromorphic Waferscale System.
- Target interface is **DNC-HICANN**

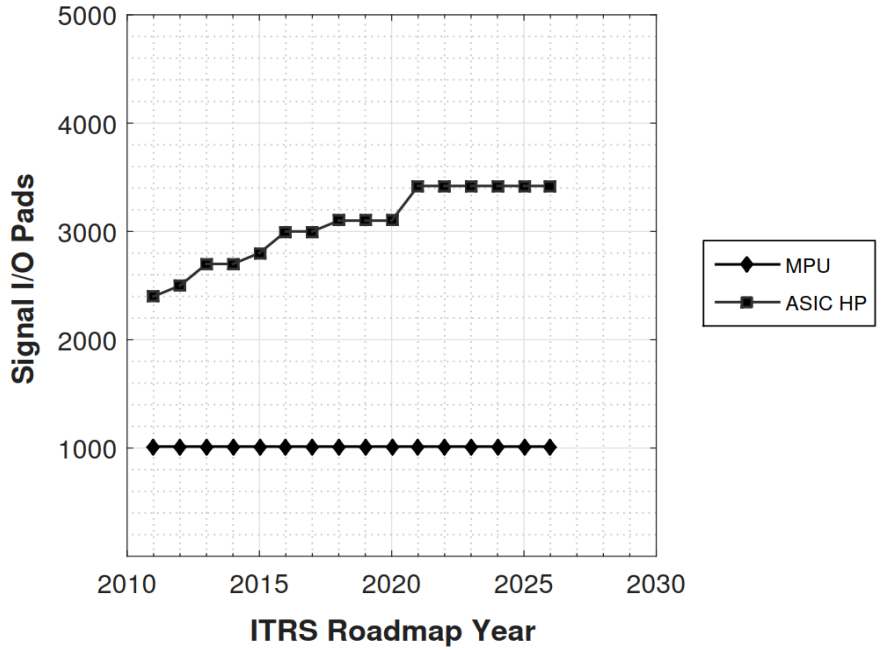
New Serial Link Properties:

- **Asynchronous** operation to minimize power and maximize total throughput.
- Activate **Burst Mode** communication to achieve low Latency and low Energy/bit.
- Switch to **TSMC® 65 nm CMOS process** to get higher integration density.

Current Trends

Why Asynchronous Operation?

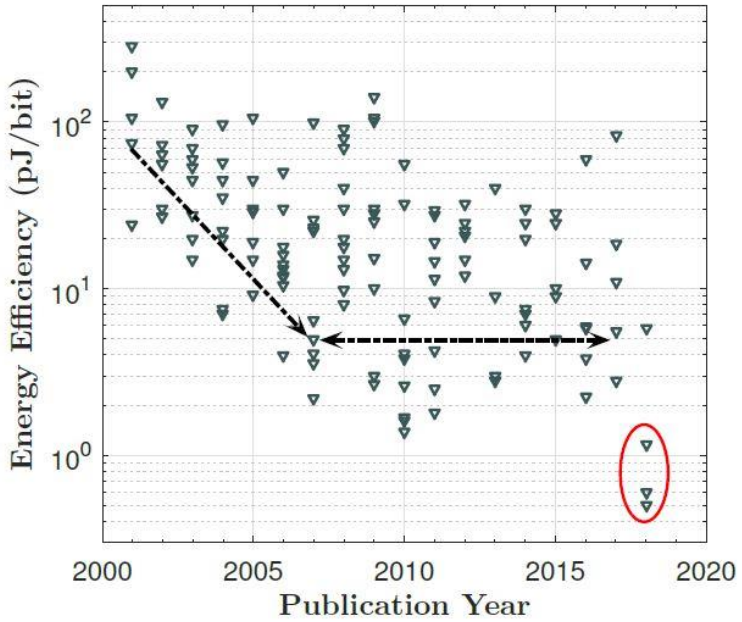




- Serial link operates more efficiently in asynchronous mode –> low power
- Install double the number of links in the same hardware space –> high bandwidth

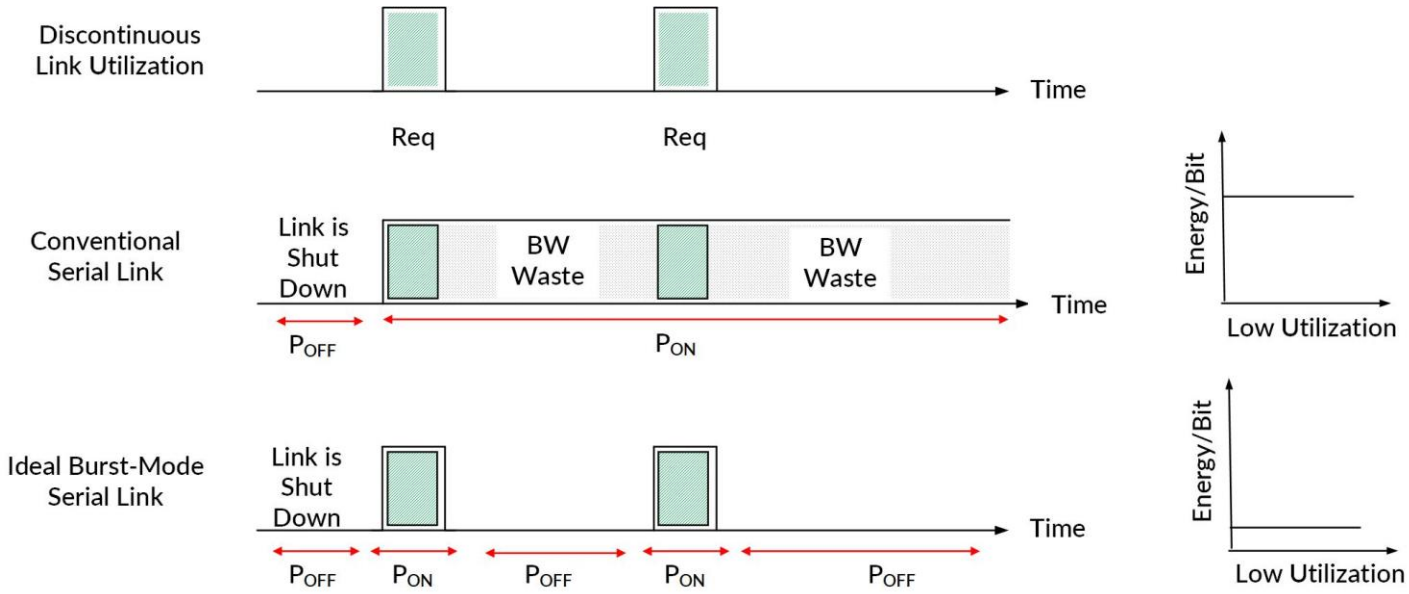
Current Trends

Why Burst Mode Operation?



- Serial links with Burst Mode operation get up to **≤1 pJ/bit** energy efficiency.

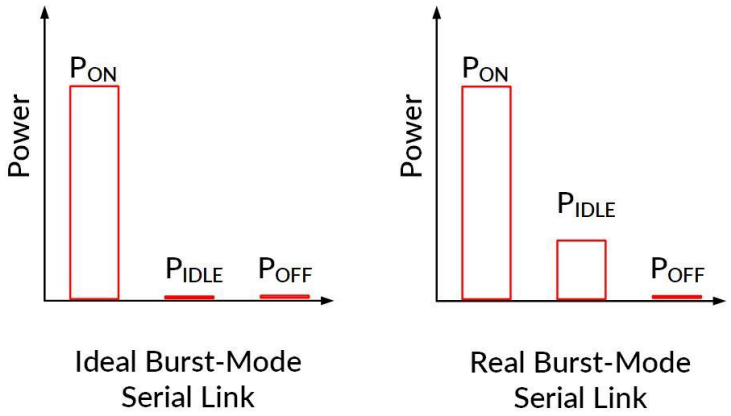
Burst Mode Communication Principle

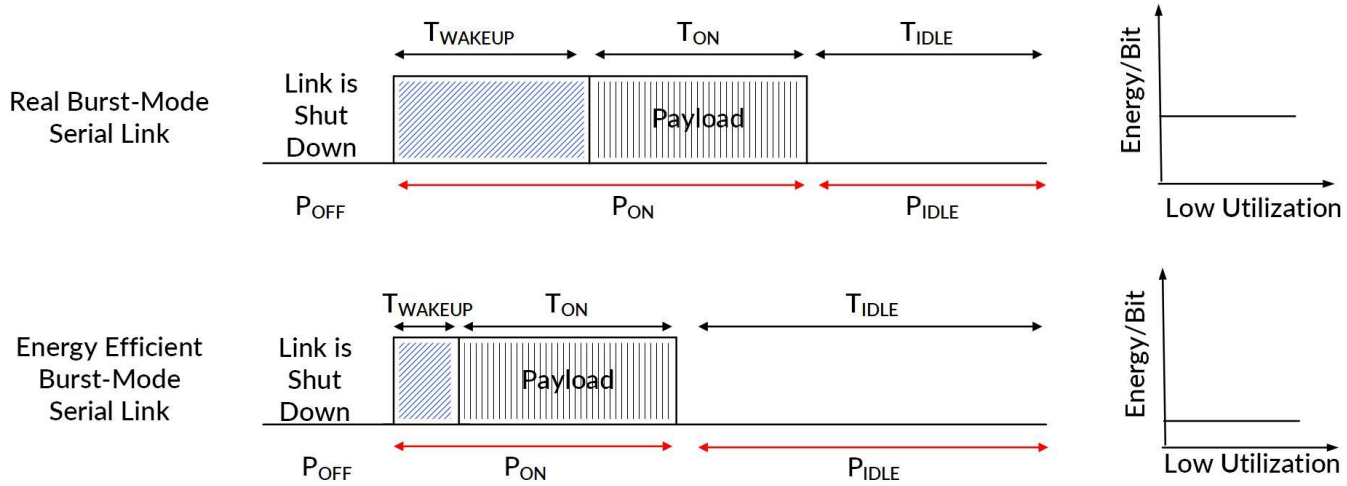


- Payload is transferred in burst of packets
- Only power-on during payload shipment otherwise power-off
- Low energy efficiency for low link utilization (proportional)

Current Trends

Activation of Burst Mode Operation in Serial Link: Practical Scenario





Target Figure of Merits of the Link

- Very low T_{WAKEUP} for low Latency and low BMEE
- P_{IDLE} << P_{ON} (P_{IDLE} ~ P_{OFF}) for low BMEE
- P_{ON} as low as possible for low BMEE
- Minimum energy loss during wakeup (low E_{WAKEUP})

$$\text{BMEE} = \frac{T_{ON} * P_{ON} + T_{WAKEUP} * P_{ON} + E_{WAKEUP} + T_{IDLE} * P_{IDLE}}{Number\ of\ payload\ bits}$$

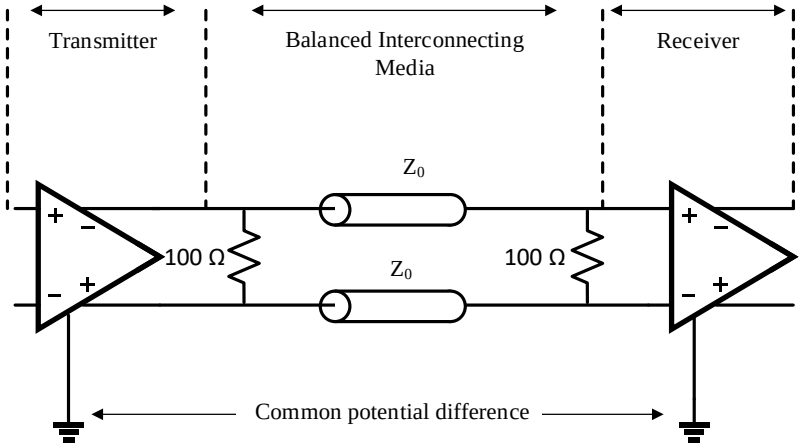
Figure of Merits

Target Figure of Merits of the Link

- **Power switching** to activate Burst Mode operation
- Very fast start typically **T_{WAKEUP} in few nano-seconds**
- Architecture with **LVDS IO** standard to get low P_{ON}
- Design to achieve very low P_{IDLE} (**P_{IDLE} << P_{ON}**)
- Avoid unbound leakages during switching (low E_{WAKEUP})

TSMC® 65 nm Deep N-well Process

LVDS Link Architecture

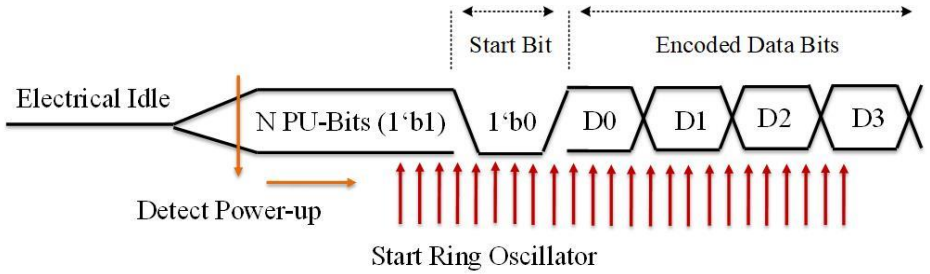


Components

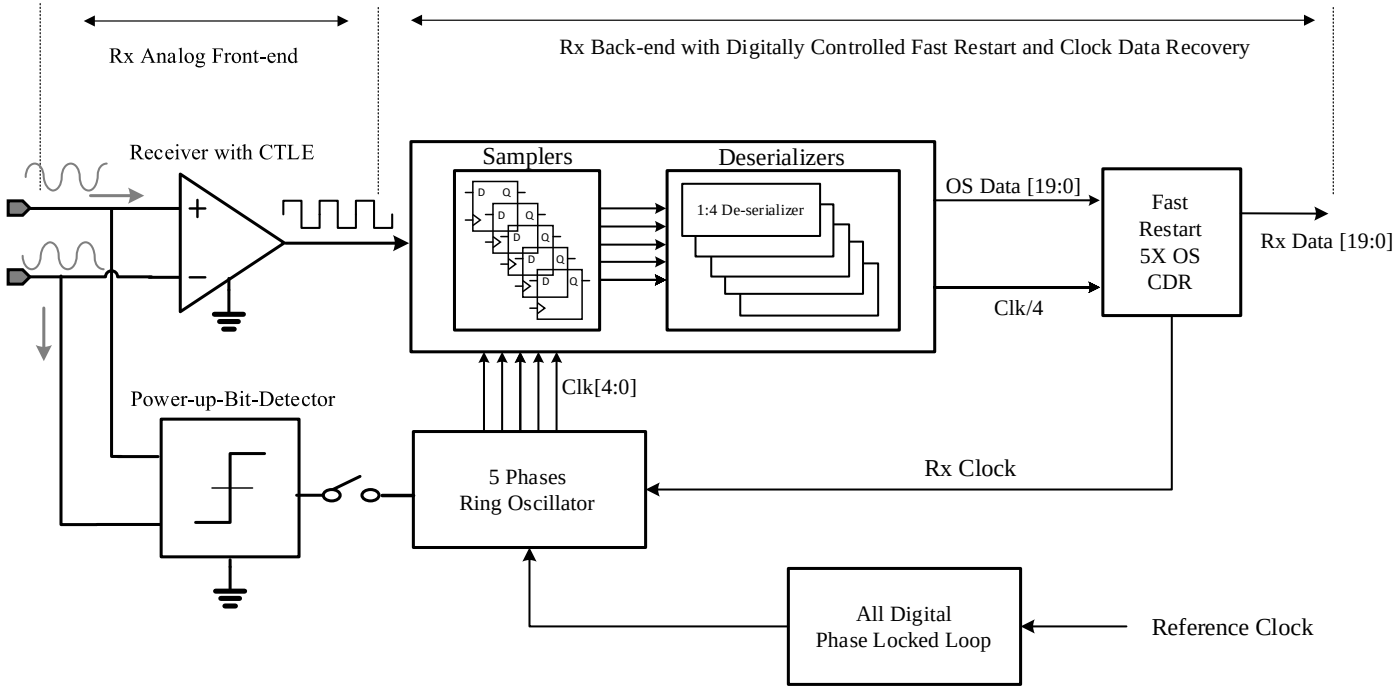
- Fast-Start Transmitter (Tx)
- Fast-Start Receiver (Rx)
- Transmission Line (e.g. backplane, PCB connector, micro-strip line, FR4, twin-axial cable etc.)

Receiver Architecture

Concept of the Fast-Start Receiver (Rx)



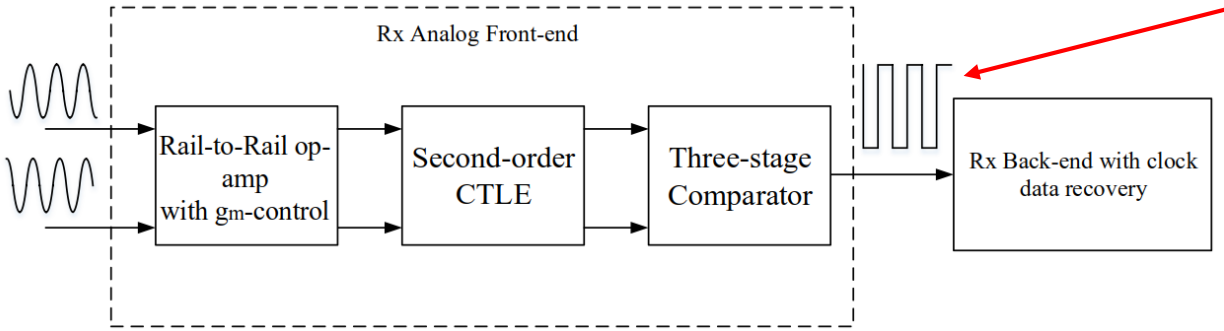
The behavior is verified in Verilog!



A Functional Model of the Fast-Start Receiver

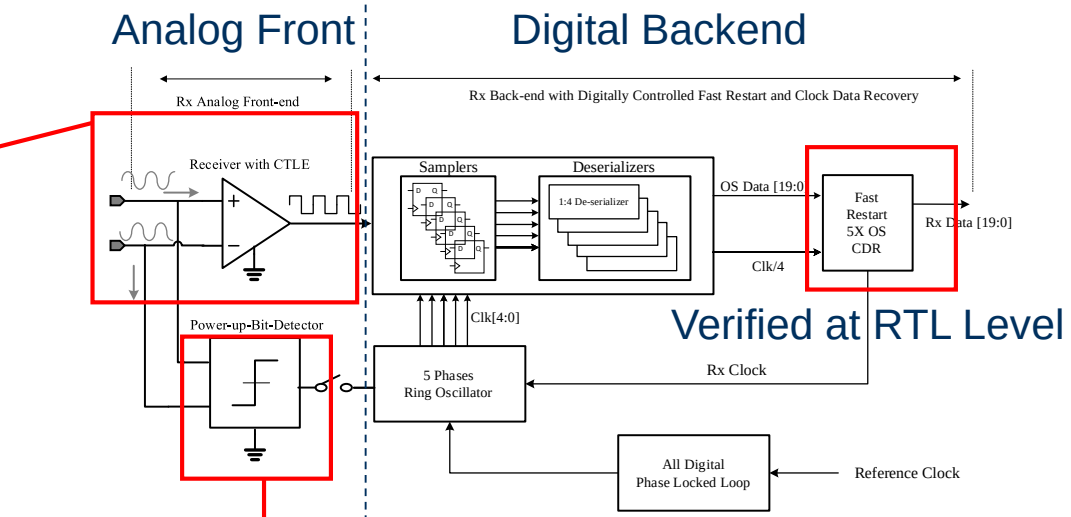
Fast-Start Rx Design

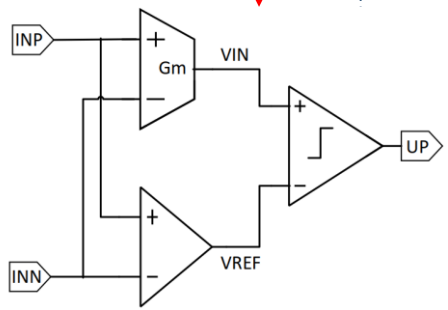
Fast-Start Rx Block Level Design



- Typical T_{WAKEUP} < 10 ns (worst case ~ 40 ns)
- P_{ON} ~ 10.6 mW and P_{IDLE} ~ 150 µW (98% reduced)

Verified at the Transistor Level!





Almost instant switch on (< 1 ns)

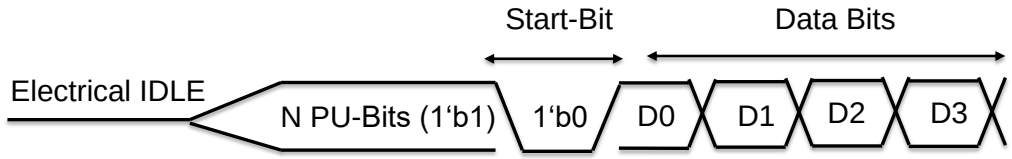
- **OS CDR 24 UI**
- Deserializer 4 UI
- Sampler 1 UI
- RO (in progress)

(in progress..)

Transmitter Architecture

Concept of the Fast-Start Transmitter (Tx)

- A compatible receiver expects the target data frame at input
- After wake up, the transmitter sends a continuous stream of Power Up Bits (1'b1), until the receiver is powered on.
- V_{OCM} during Idle can be maintained by a bypass circuit

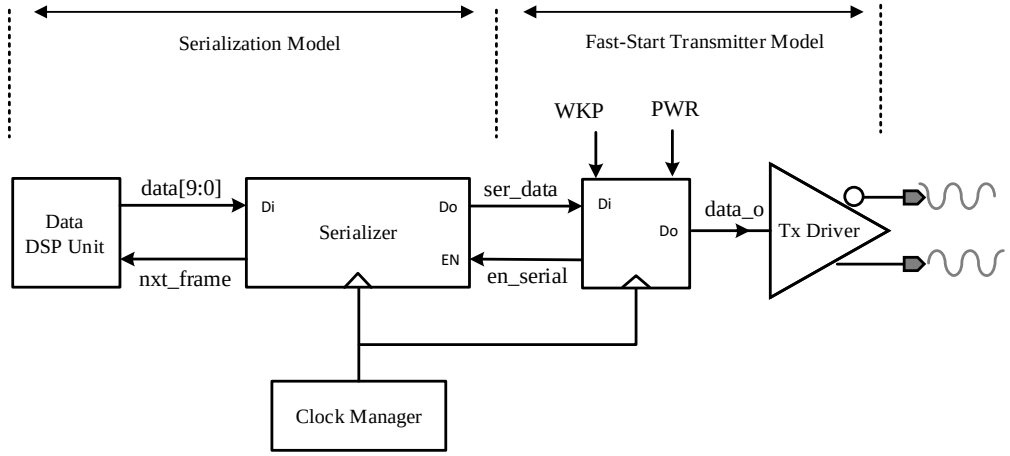


Control Scheme of the Fast-Start Transmitter

PWR	WKP	System State
0	?	Power Down
1	0	Idle
1	1	Start-up

Power-Up

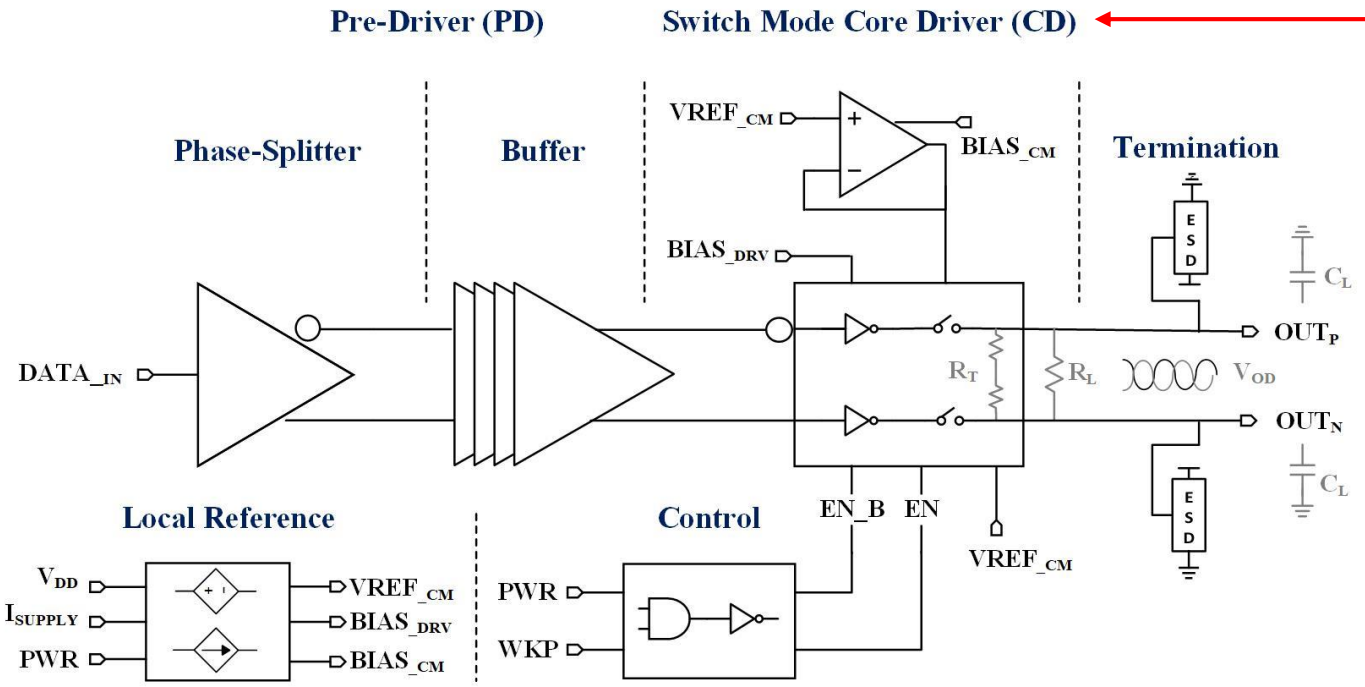
Wake-Up

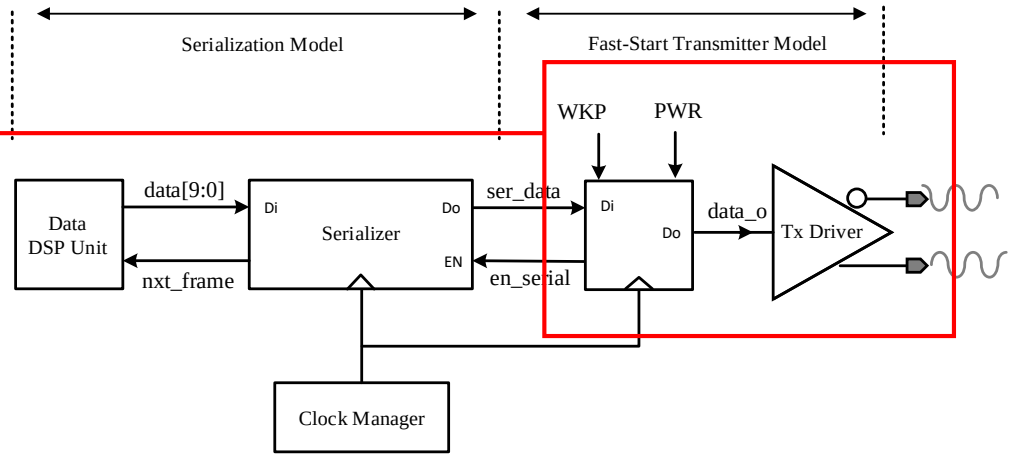


A Functional Model of the Fast Start Transmitter

Fast-Start Tx Design

Fast-Start Tx Block Level Design





- Typical T_{WAKEUP} < 1.5 ns (worst case ~ 4.14 ns)
- P_{ON} ~ 9.17 mW and P_{IDLE} ~ 209.5 µW (97% reduced)

Verified at the Transistor Level!

(in progress..)

Thank you for your attention!

Check the link: **www.humanbrainproject.eu/en/silicon-brains/**

Contact for querries: **mallik.bitan@gmail.com**

HPSN, TU Dresden is an official partner of the BrainScaleS since January 2011.
In 2013, 2016 and recently up to 2023 the project rights are revalidated.

Link: **https://tu-dresden.de/ing/elektrotechnik/iee/hpsn/forschung/forschungsprojekte**

DRESDEN concept