

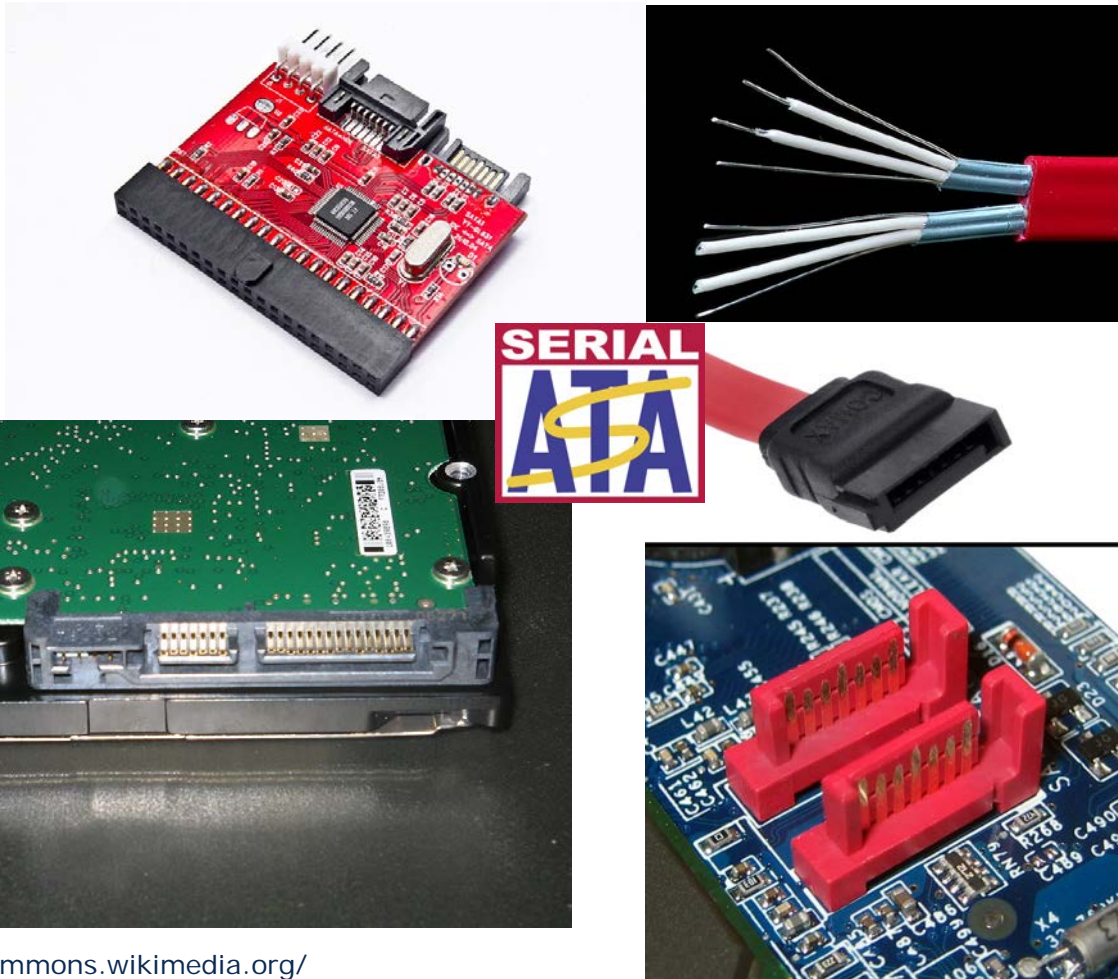


# Implementierung eines auf Streaming optimierten SATA-Host-Bus-Adapters (HBA)

Patrick Lehmann

Dresden, 11.04.2012





<http://commons.wikimedia.org/>

# Agenda

- 1 OSI-Referenzmodell
  - 1.1 Einordnung von Serial-ATA
  - 1.2 Einordnung von AT-Attachement
- 2 SATAController
  - 2.1 Electrical-Physical-Layer
  - 2.2 Logical-Physical-Layer
  - 2.3 Link-Layer
- 3 Vom Frame zum Paket / Kommando
  - 3.1 Implementierungsvarianten
  - 3.2 ATASstreamingController
- 4 Streaming-optimierter Datenzugriff
  - 4.1 Implementierungsvarianten
  - 4.2 StreamingDB
- 5 Ausblick

# 1 OSI-Referenzmodell

## 1.1 Einordnung von Serial-ATA

### Bereich des Standards:

- Neue elektrische Spezifikationen
  - Kabel, Stecker, Schirmung
  - Leitungskodierung
- Kompatibilitätsmodi
  - Transport von ATA-Registern
  - PIO- und DMA-Transfer
  - Native Übertragung von LBA48-Adressen
- Erweiterungen:
  - Geschwindigkeitsaushandlung
  - Neues Programmiermodell (AHCI)
  - Port-Multiplier (PM)
  - Native-Command-Queuing (NCQ)

[4] Serial ATA: High Speed Serialized AT Attachment – Rev. 1.0a

## 1.2 Einordnung von AT-Attachement

### ATA bleibt erhalten:

- Transport der Befehle durch Registertransfers
- Alle ATA Kommandos werden unterstützt
- Erweiterung um neue DMA-Befehle für NCQ

### Advanced Host Controller Interface (AHCI):

- Parallele Abarbeitung von bis zu 32 ATA Kommandos
- Erlaubt bis zu 15 Devices pro Host-Port (PM)
- Taskfile basiertes Programmiermodell
  - Im Hauptspeicher abgelegte Frames
  - autonomer Transport über DMA-Engine (First Party DMA)

[2] AT Attachement 8 – ATA/ATAPI Architecture Model

[3] AHCI – Serial-ATA – Advanced Host Controller Interface Specification 1.3

## OSI-Referenzmodell – angepasst auf SATA [5]

#	OSI-Layer	Modulname	
7	Application		
6	Presentation	StreamingDB	
5	Session	CommandLayer	
4	Transport	TransportLayer	
3	<i>Network</i>		
2	Link	LinkLayer	
1.2	Logical Physical	PhysicalLayer	
1.1	Electrical Physical	TransceiverLayer	

## OSI-Referenzmodell – angepasst auf SATA [5]

#	OSI-Layer	Modulname	
7	Application		
6	Presentation	StreamingDB	
5	Session	CommandLayer	
4	Transport	TransportLayer	
3	<i>Network</i>		
2	Link	LinkLayer	logische Verbindung - Übertragung von Frames, Sicherung per CRC
1.2	Logical Physical	PhysicalLayer	physische Verbindung - Verbindungsaufbau, Geschwindigkeitsaushandlung
1.1	Electrical Physical	TransceiverLayer	FPGA spezifische Schicht - Konfiguration des MGT für SATA - Clock-Netzwerk, DCMs, PLLs, OOB-Signaling



## OSI-Referenzmodell – angepasst auf SATA [5]

#	OSI-Layer	Modulname	
7	Application		
6	Presentation	StreamingDB	
5	Session	CommandLayer	ATA8 Befehlsschicht - Versand und Empfang von Paketen - DMA-Zugriff, Native-Command-Queuing (NCQ)
4	Transport	TransportLayer	Serial-ATA Kompatibilitätsschicht - Transport von ATA8 Paketen in Serial-ATA Frames - Übertragung von Status- und Error-Bitvektoren
3	<i>Network</i>		<i>entfällt bei SATA bis Rev. 2.0</i>
2	Link	LinkLayer	logische Verbindung - Übertragung von Frames, Sicherung per CRC
1.2	Logical Physical	PhysicalLayer	physische Verbindung - Verbindungsaufbau, Geschwindigkeitsaushandlung
1.1	Electrical Physical	TransceiverLayer	FPGA spezifische Schicht - Konfiguration des MGT für SATA - Clock-Netzwerk, DCMs, PLLs, OOB-Signaling

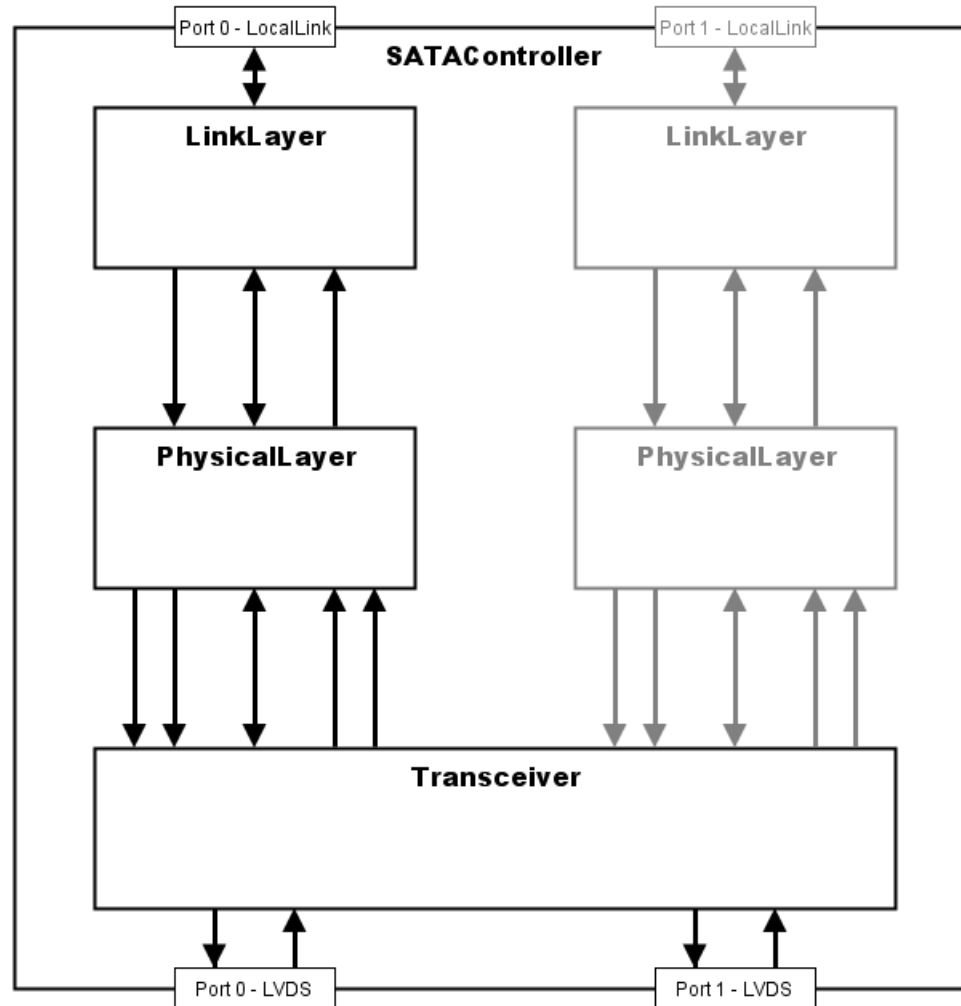
## OSI-Referenzmodell – angepasst auf SATA [5]

#	OSI-Layer	Modulname	
7	Application		Beispielanwendungen: - Genom Alignment, SHAP
6	Presentation	StreamingDB	Verwaltung einer Zuordnungstabelle - Zugriff auf Datenblöcke über eine Blocknummer
5	Session	CommandLayer	ATA8 Befehlsschicht - Versand und Empfang von Paketen - DMA-Zugriff, Native Command-Queuing (NCQ)
4	Transport	TransportLayer	Serial-ATA Kompatibilitätsschicht - Transport von ATA8 Paketen in Serial-ATA Frames - Übertragung von Status- und Error-Vektoren
3	Network		<i>entfällt bei SATA bis Rev. 2.0</i>
2	Link	LinkLayer	logische Verbindung - Übertragung von Frames, Sicherung per CRC
1.2	Logical Physical	PhysicalLayer	physische Verbindung - Verdrahtung, FFC, Geschwindigkeitssensoren
1.1	Electrical Physical	TransceiverLayer	FPGA spezifische Schicht - Konfiguration des MGT für SATA - Clock-Netzwerk, DCMs, PLLs, OOB-Signaling

**ATAStreamingController**

**SATAController mit FIFO-Interface**

## 2 SATAController

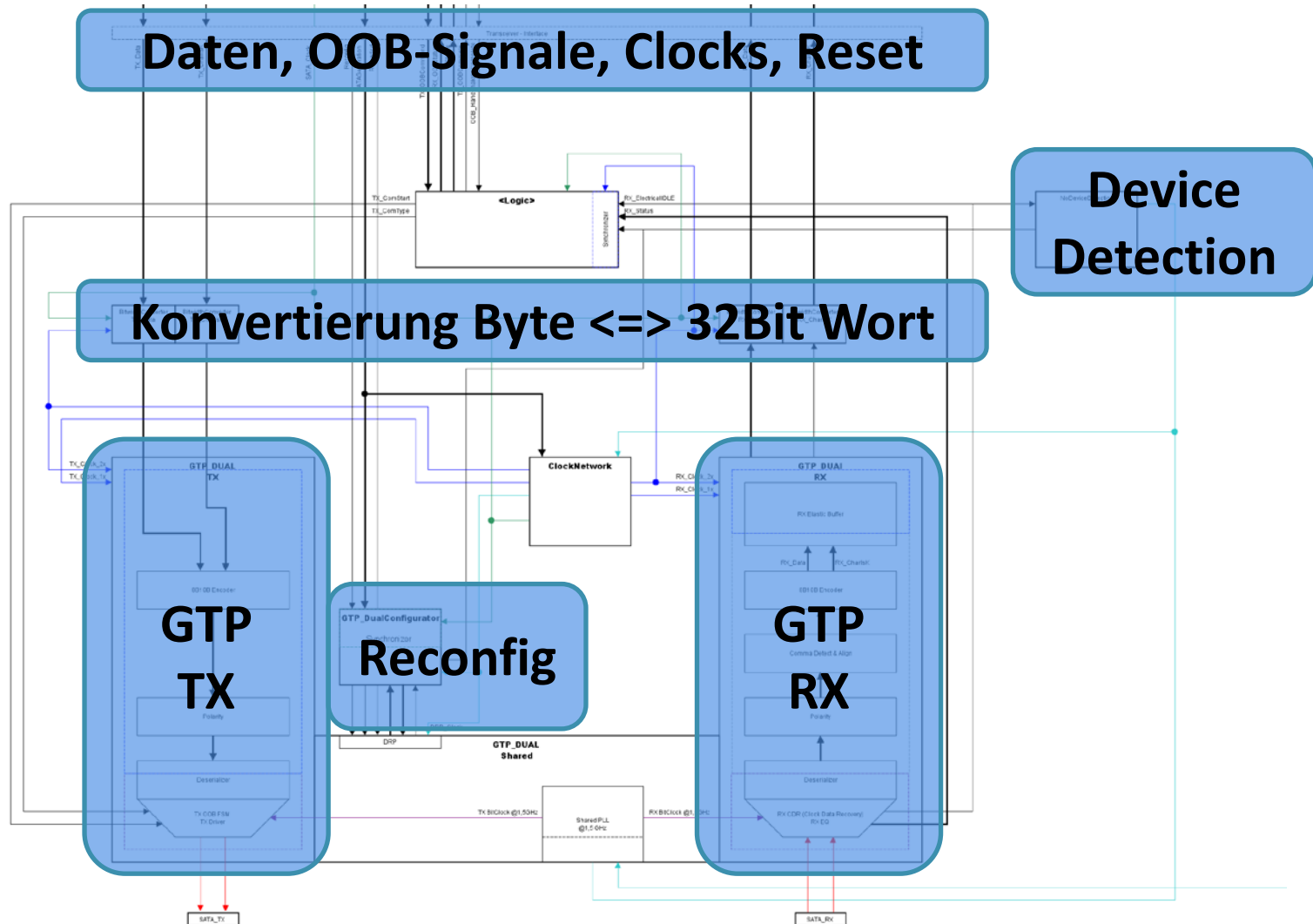


## 2.1 Electrical Physical Layer

### Angebotene Dienste:

- Ein FPGA und Hersteller unabhängiges Interface
- Konfiguration des MGT für Serial-ATA
- Kapselung des FPGA spezifischen Taktnetzwerkes
- Geschwindigkeitsumschaltung
  - Virtex5: Reprogrammierung zur Laufzeit
  - Virtex6: Ändern von ClockDividern zur Laufzeit  
Reset des MMCM
- Erkennung von leeren Ports/neu angesteckten Geräten

[3] PHY Interface for PCI Express 3.0 Architecture – PIPE 2.00

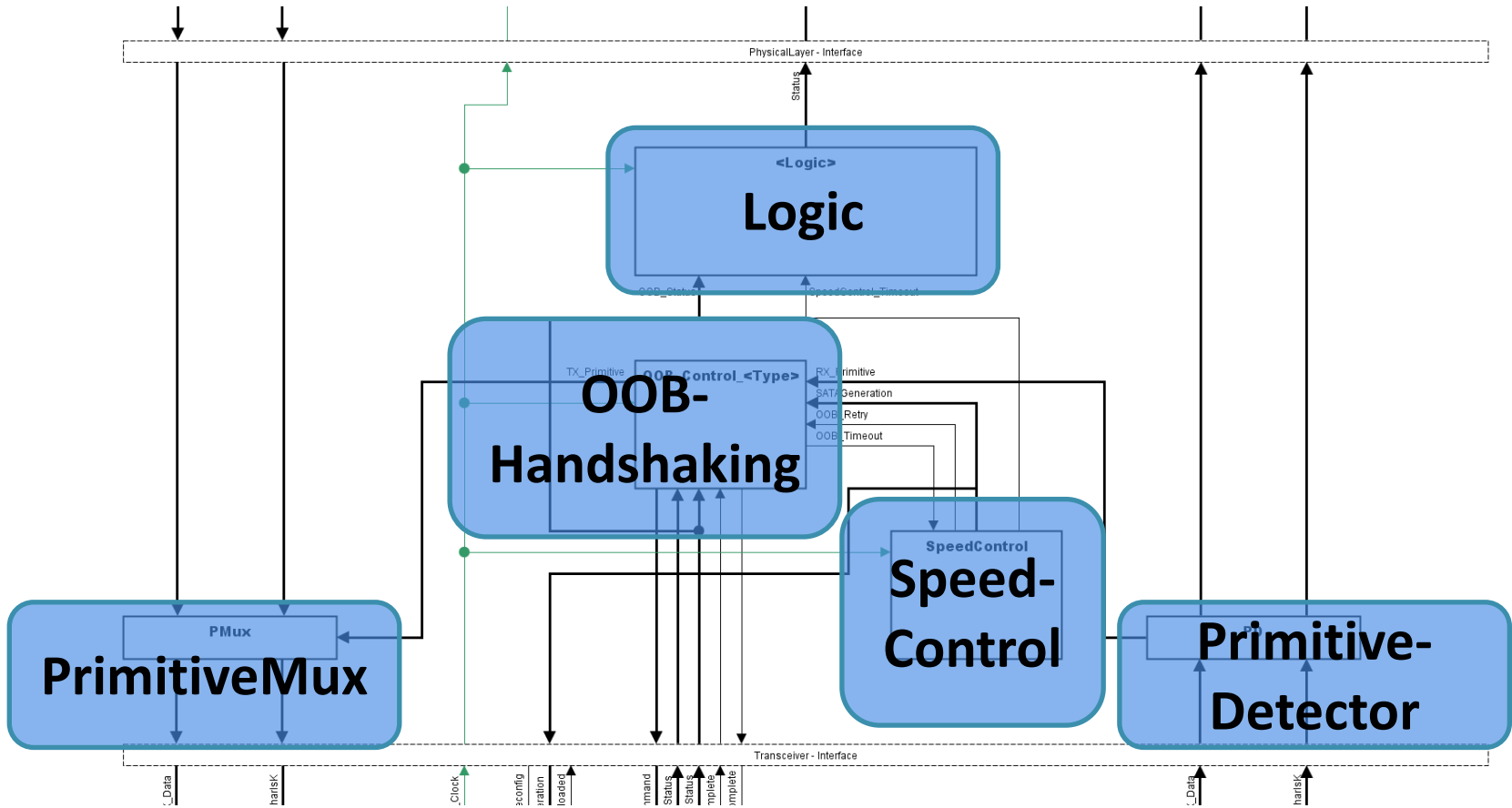


## 2.2 Logical Physical Layer

### Angebotene Dienste:

- Physischer Verbindungsaufbau
- Geschwindigkeitsaushandlung
- Erkennung von Leitungsfehlern (z.B. 8B10B Fehler)
- Senden und Empfangen von Hardresets

[6] SATA Storage Technology, MindShare Press





## 2.3 Link Layer

### Angebotene Dienste:

- Logische Verbindung herstellen und halten
- Verpacken von Daten in Frames
- CRC-Berechnung und Prüfung am Paketende
- Scrambling
- SATA-Primitive in Datenstrom einfügen/entfernen
- Flusskontrolle durch  $HOLD_p$  Primitive
- Richtungs-aushandlung zwischen Host und Device
- Bestätigungs-/Fehlermeldung über  $R\_OK_p$ ,  $R\_ERROR_p$  Primitive
- Frames wiederholt senden
- Sende- und Empfangs-FIFOs (ca. 8kB)

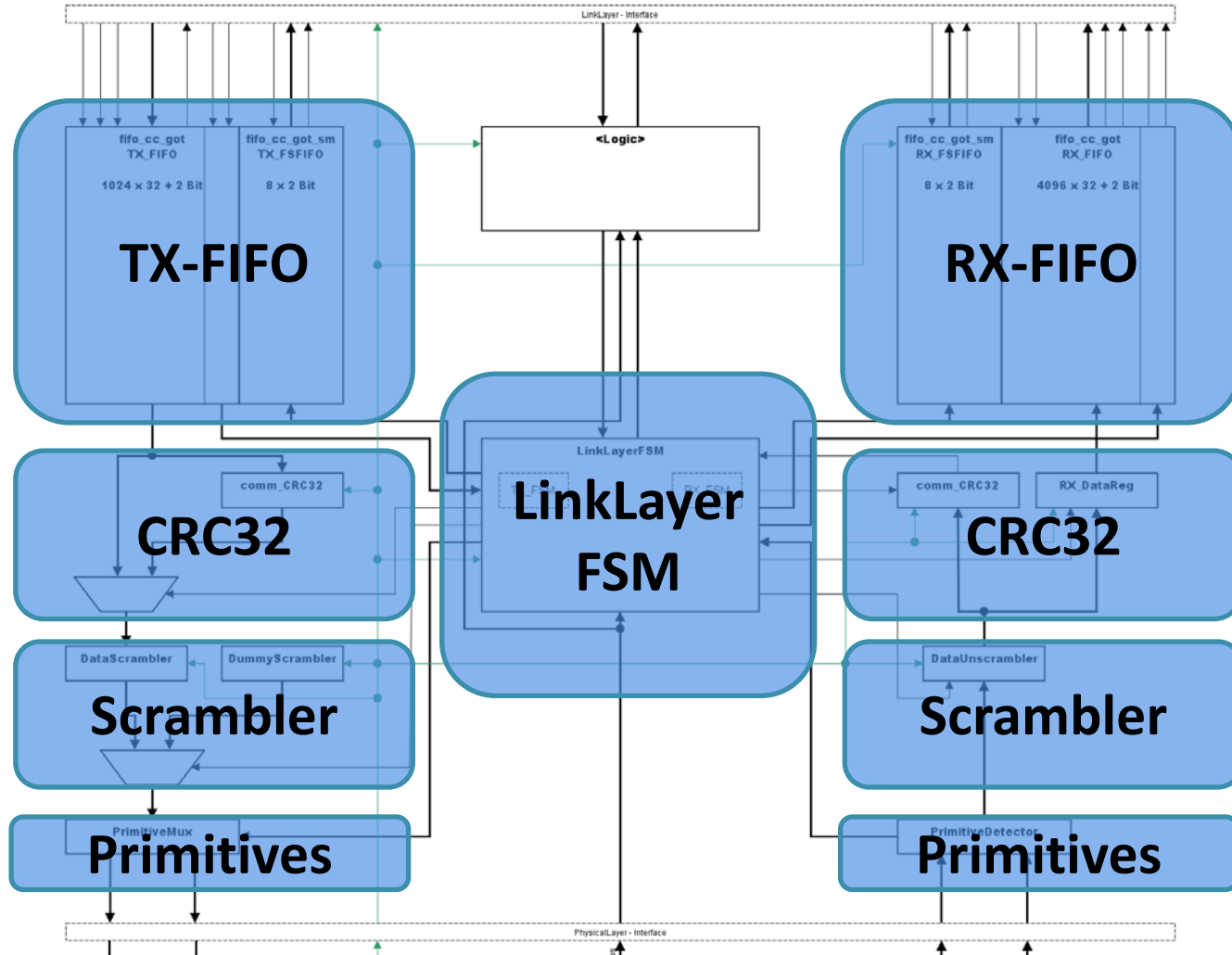
[6] SATA Storage Technology, MindShare Press

## 2.3 Link Layer

### LinkLayer Interface:

- Command, Status, Error
- 32-Bit FIFO Interface für TX und RX
  - \*\_Ready, \*\_Valid
  - \*\_Data, \*\_SOF, \*\_EOF
- 2-Bit FIFO Interface für Frame-Status
  - \*\_Ready, Valid
  - \*\_SendOK, \*\_CRC\_OK, \*\_Abort
- Diverse Generics zur Parametrierung (Auswahl)
  - Anzahl der Ports
  - Porttyp (Host / Device)
  - Timeout-Zeiten

[1] Xilinx LocalLink Interface Specification (SP006)

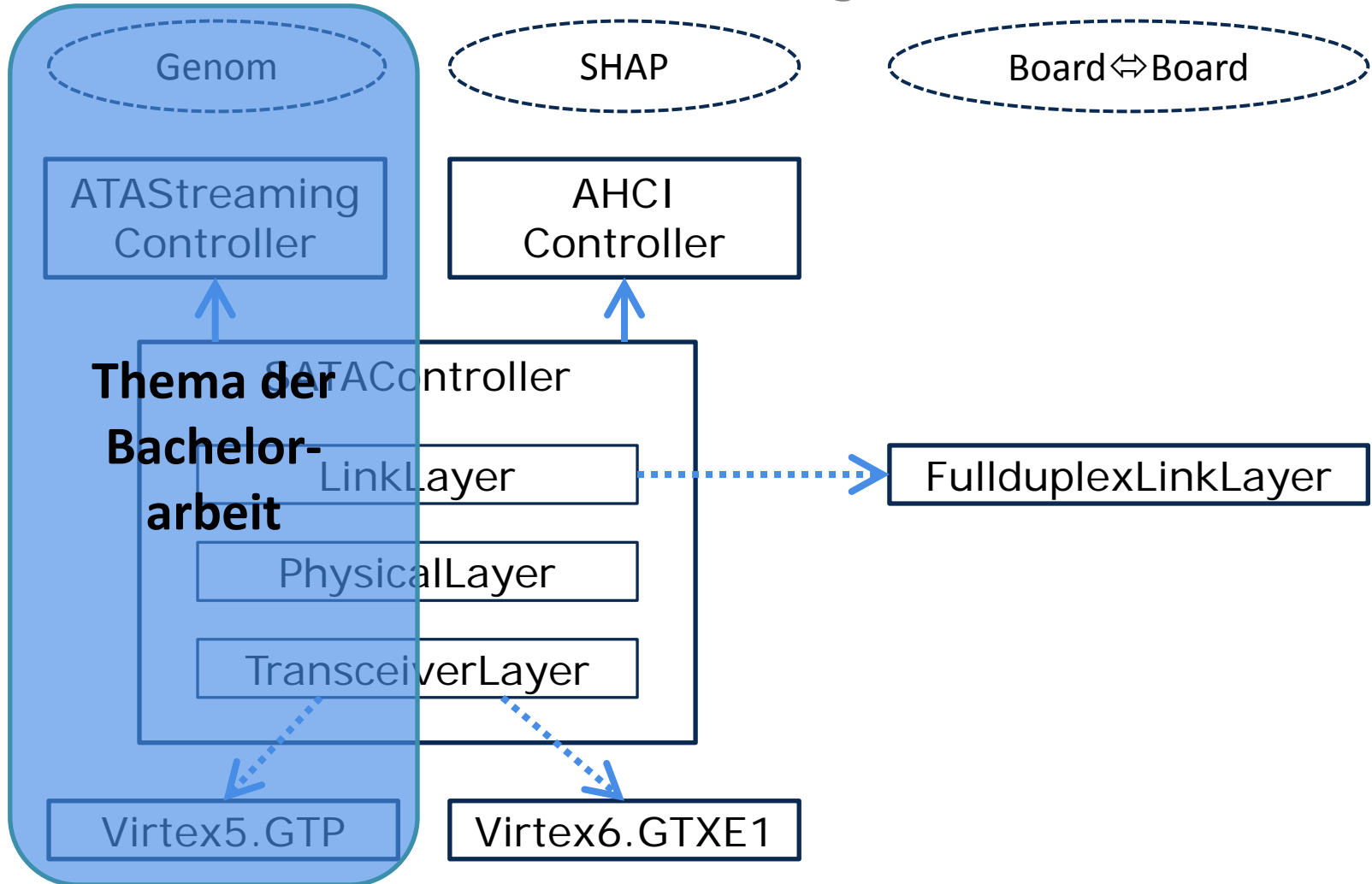


## FPGA Ressourcenverbrauch

	<b>Virtex5 – LX50T</b>		<b>Virtex6 – LX240T</b>	
	<b>1 Port</b>	<b>2 Ports</b>	<b>1 Port</b>	<b>2 Ports</b>
LUTs	ca. 1250	ca. 2600	ca. 1450	ca. 2060
Slices	ca. 620	ca. 1270	ca. 660	ca. 910
Register	ca. 600	ca. 1030	ca. 660	ca. 660
BlockRAM	4	8	4	8
MGTs	1	1	1	2
BUFG / BUFR	3 / -	5 / -	2 / 1	4 / 1
DCM bzw. MMCM	1	1	1	2
Gesamtverbrauch <sup>1</sup>	8,5 %	17,6 %	1,8 %	2,4 %

<sup>1</sup> Gesamtverbrauch = max(LUTs, Slices, Register)

# SATAController - Anwendungsfälle



## 3 Vom Frame zum Paket / Kommando

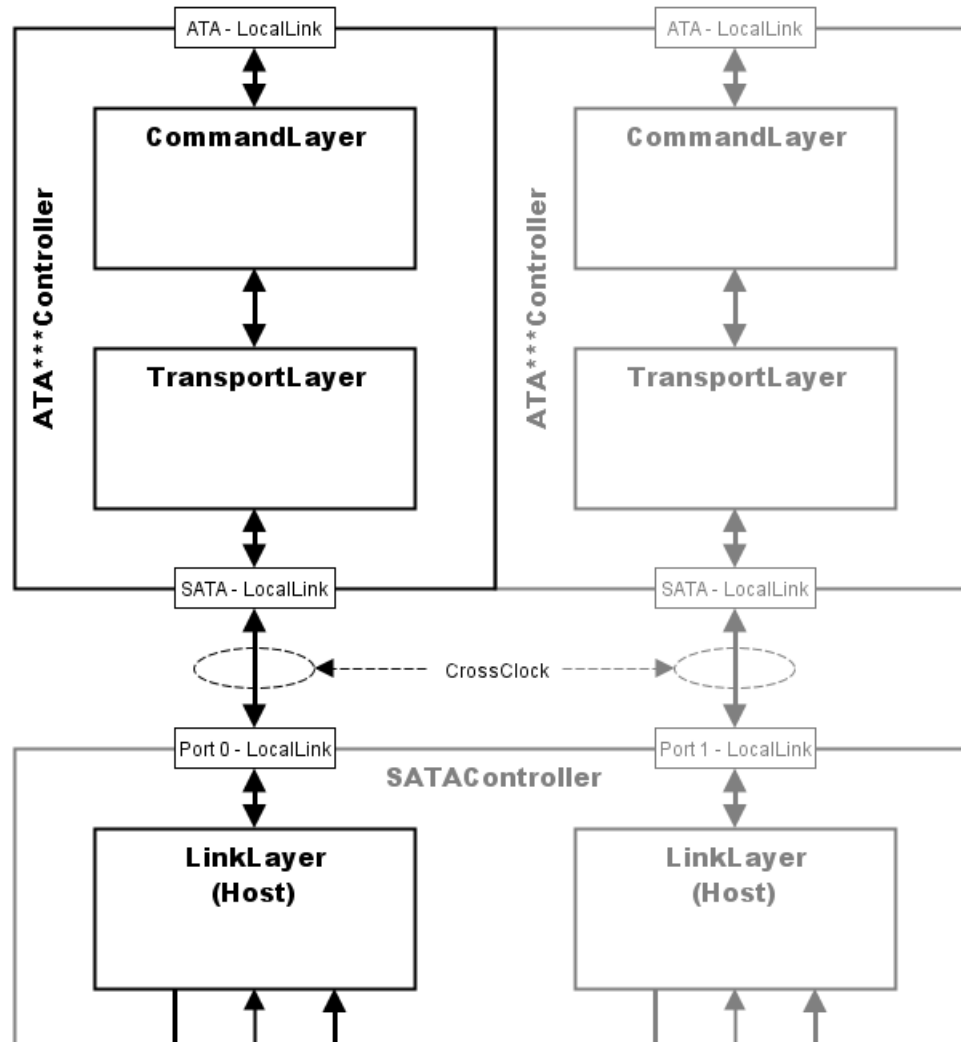
## Implementierungsvarianten

### AHCI Controller:

- Hauptspeicher (z.B. SRAM)
- DMA-Engine
- Blockweise Datentransfers
- Zusätzlicher Overhead durch DMA-Setup-FISes
- + NCQ ist realisierbar (bis zu 32 ausstehende Operationen)
- + Vollduplex

### ATAStreamingController:

- BlockRAM für FIFOs (ca. 8 kB)
- Halbduplex
- + DMA Bursttransfers (bis zu 32 MB/Burst)
- + FIFO Interface





## Unterstützte ATA8 Kommandos

Command Category	Command Name
Non-Data	FLUSH_CACHE_EXT
PIO-Data-In	IDENTIFY_DEVICE
DMA-In	READ_DMA_EXT <i>READ_FPDMA_QUEUED</i>
DMA-Out	WRITE_DMA_EXT <i>WRITE_FPDMA_QUEUED</i>

### Kompatibilitätseinschränkung:

- DMA-Transfers
- LBA Adressierung im 48-Bit Modus

[2] AT Attachernet 8 – ATA/ATAPI Command Set

## ATAStreamingController

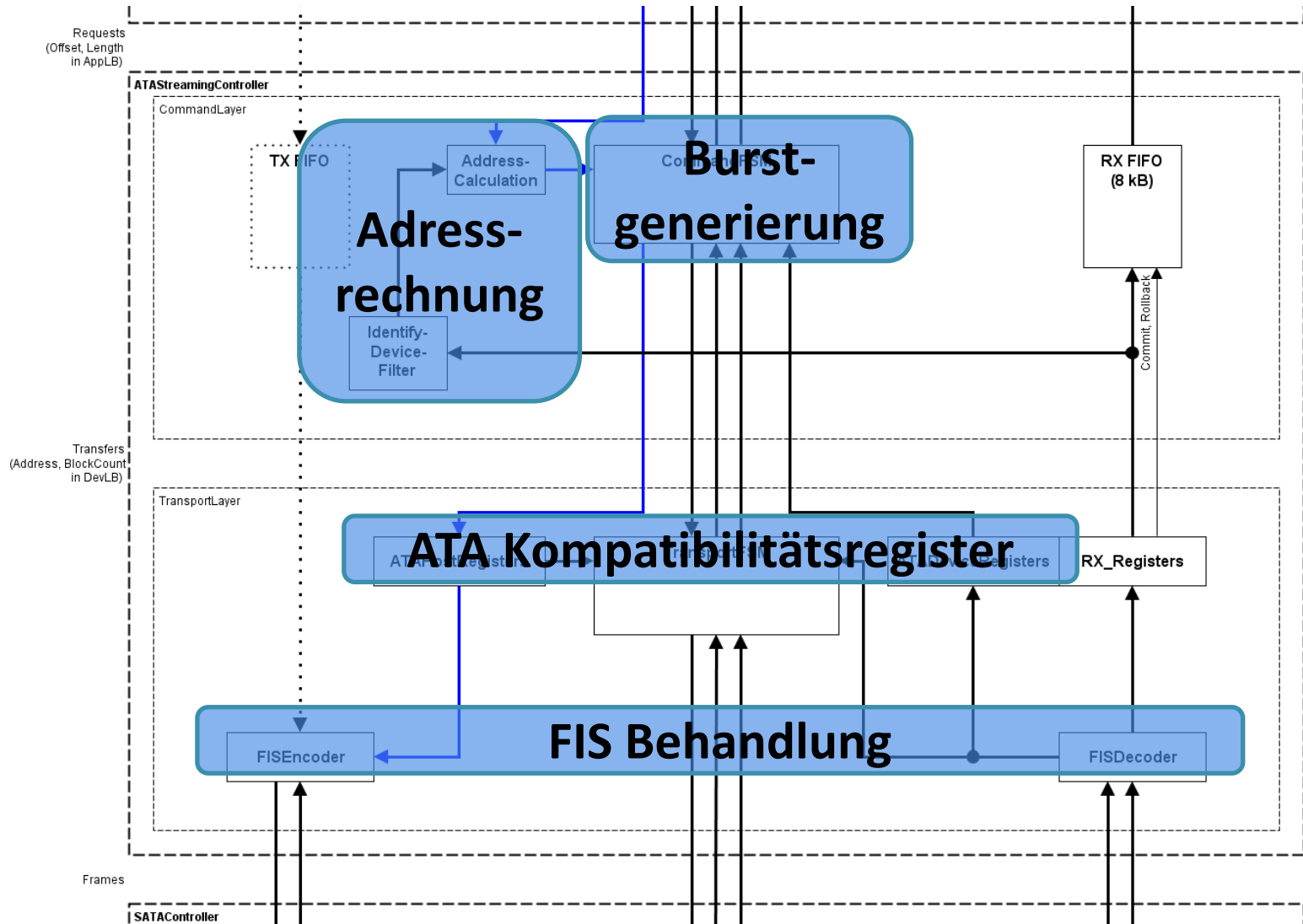
### TransportLayer Dienste:

- ATA Paket- und Sitzungsverfolgung
- Transparente Register Transfers
- Extraktion von Daten FISes

### CommandLayer Dienste:

- ATA8 Kommando Generierung
- Auswertung des IDENTIFY\_DEVICE Antwortpaketes
- Adressumrechnung:
  - Umrechnung von n AppLB in m DevLB
  - 1 AppLB := 8 kB
  - 1 DevLB umfasst je nach Device 512 Bytes, 4 kBytes, ...

[6] SATA Storage Technology, MindShare Press



## FPGA Ressourcenverbrauch

	Virtex5	
	SATAC + ATASC	ISCID.2009.124
LUTs	ca. 2080	ca. 2610
Slices	ca. 1100	ca. 1390 <sup>2</sup>
Register	ca. 1180	ca. 2230
Gesamtverbrauch <sup>1</sup>	15 % <sup>3</sup>	19 %

<sup>1</sup> Gesamtverbrauch = max(LUTs, Slices, Register)

<sup>2</sup> Slices mit einer LUT/Slice-Rate von 1,88 abgeschätzt

<sup>3</sup> kleiner als 1 SHAP Kern

### ISCID.2009.124

Implementing a Serial ATA Controller base on FPGA

Wei Wu, Hai-bing Su, Qin-zhang Wu

Second International Symposium on Computational Intelligence and Design, 2009

## 4 Streaming-optimierter Datenzugriff

## Implementierungsvarianten

### „echtes“ Dateisystem:

- Partitionstabelle
- Viele Indirektionsstufen
- Ungeeignet für FSMs => Einsatz eines Softcores
- Teilweise mit Patenten geschützt
- Ungenutzte Dateisystemfunktionen (z.B. Ordner, Löschen)
- + Einfaches Zugriffsmodell im Rechner über Treibermodell

### „einfache“ Zuordnungstabelle:

- LowLevel Schreibzugriff auf RAW-Device im Rechner
- + Master Table im BlockRAM speicherbar
- + Durch einfache FSMs zu realisieren
- + Zugriff über Datenbanknummer

# Datenzuordnungstabelle

## Master Table (MT)

- Erster 8 kB Block auf der Festplatte
- Metadaten (Version, Datum, LBU Größe)
- Anzahl Datenbankeinträge
- Datenbank Einträge (je 8 Byte)
  - Daten Startadresse in LBU    32 Bit
  - Daten Länge in LBU            16 Bit
  - CRC16-IBM Checksumme       16 Bit    (zu Testzwecken)

## Beispiel:

1 Logical Block Unit (LBU) = 8 kB

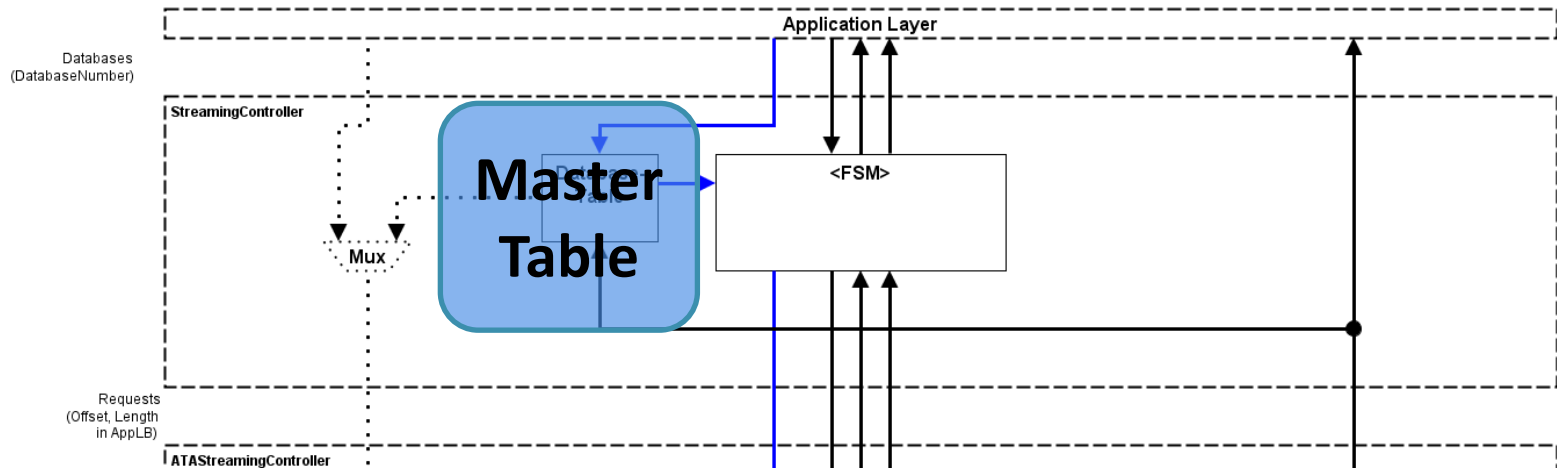
⇒ Festplatten bis zu 32 TB Gesamtkapazität

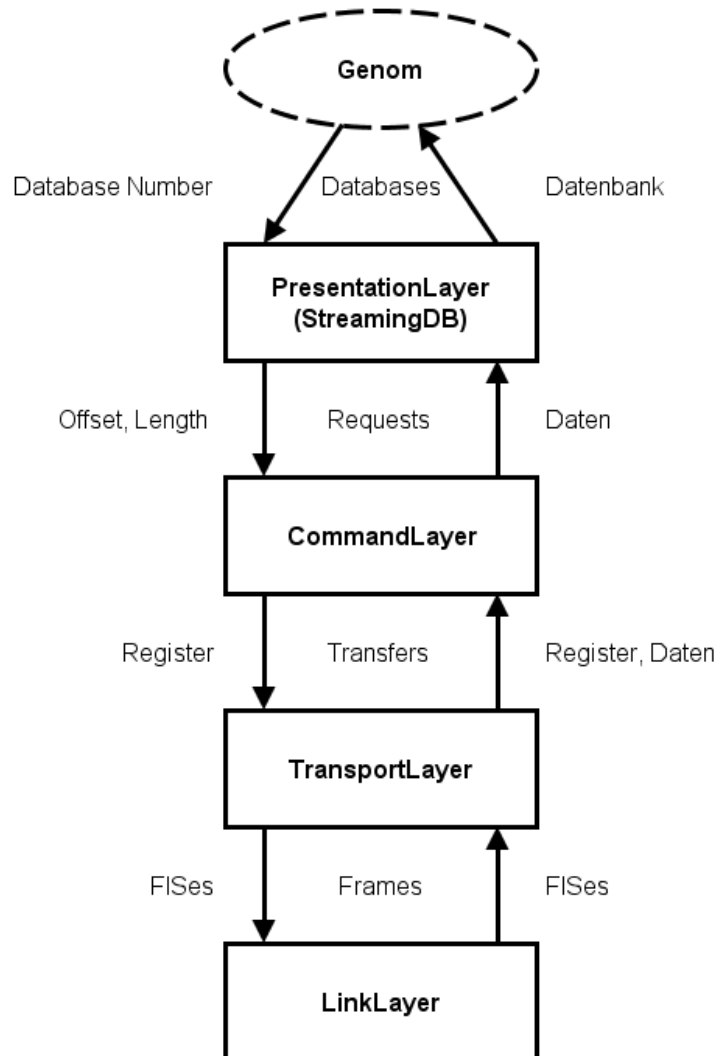
⇒ Ca. 1000 Datenbanken von 8 kB bis 32 TB

# Datenzuordnungstabelle

[Bytes]	4	8	12	16		
0	Version [Major.Minor.Release.Build]		LastChange [Year.Month.Day.Hour.Minute.Second.Offset]			
16	HeaderBlockCount [AppLB]	HeaderSize [Bytes]	TableSize [Bytes]	DatabaseCount		
32	ChecksumPolynomial					
48						
64	DB[0].Offset	^.BlockCount	^.Checksum	DB[1].Offset	^.BlockCount	^.Checksum
80	DB[2].Offset	^.BlockCount	^.Checksum	DB[3].Offset	^.BlockCount	^.Checksum
96						
...	...					
8176	DB[1014].Offset	^.BlockCount	^.Checksum	DB[1015].Offset	^.BlockCount	^.Checksum
8192						







Virtex5: Test mit HDD



Virtex5: Test mit HDD



Virtex5: Test mit HDD



Virtex5: Test mit FPGA  
Virtex6: Simulation

## 5 Ausblick

## TODO Liste

### SATAController

- Frames wiederholt senden
- Scrambling der Sendedaten

### ATAStreamingController

- Fehler finden und beheben

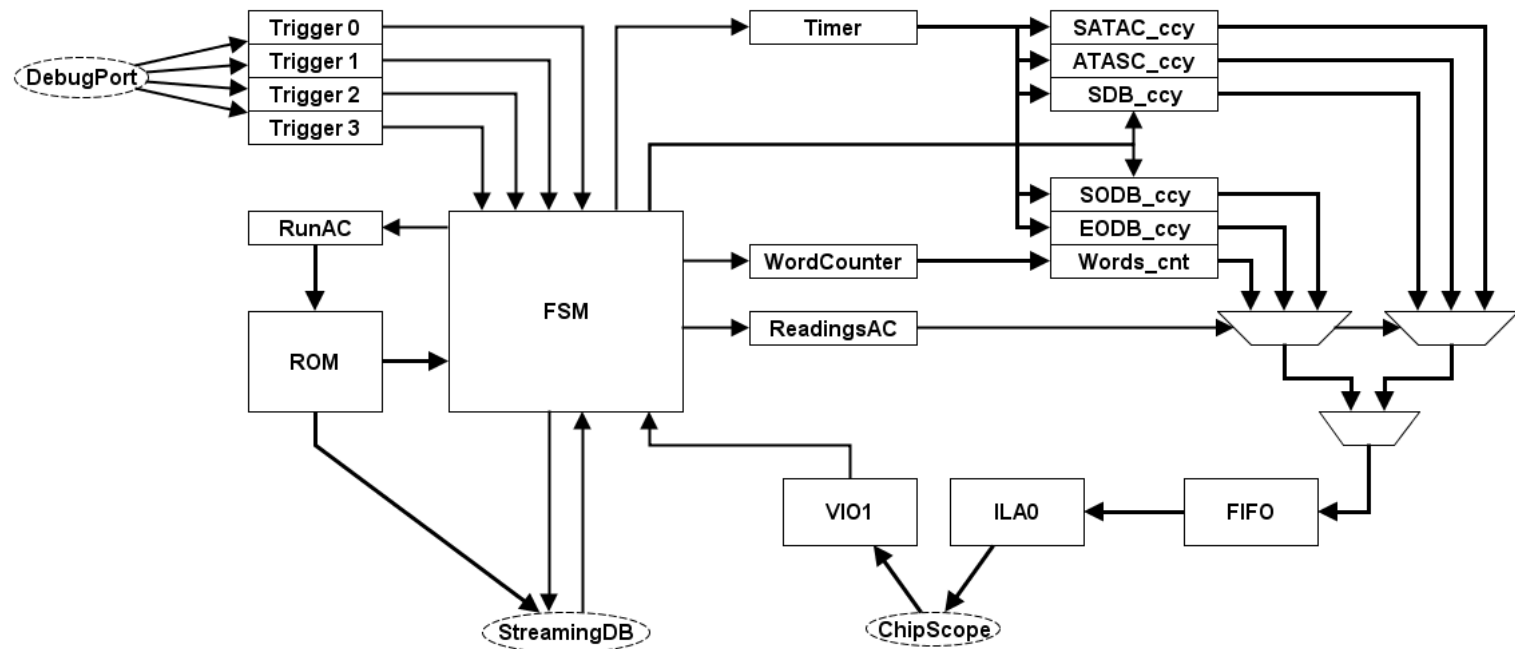
### StreamingController

- CRC16-IBM Checksumme auswerten

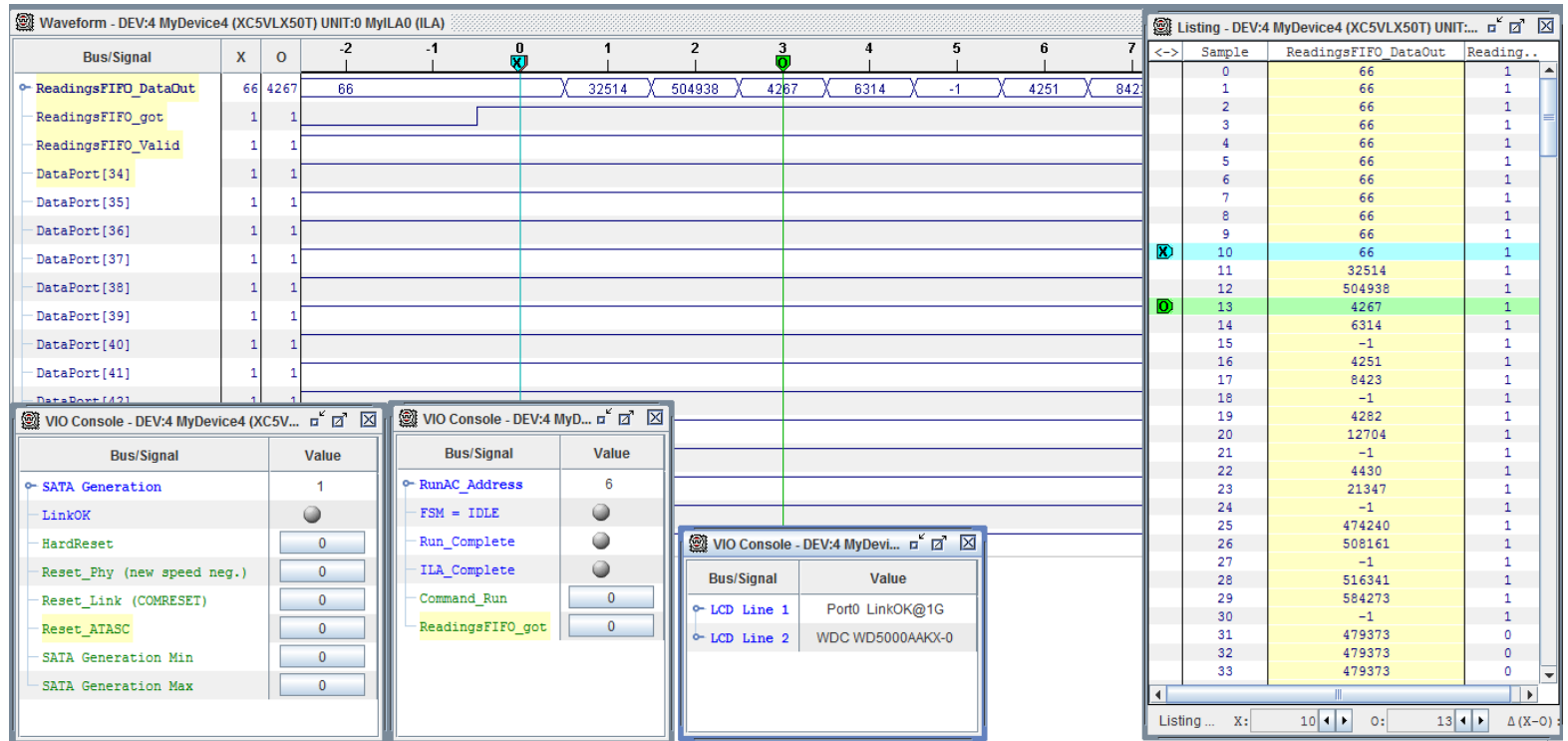
### Funktionstest des SATA-/ATA-Stacks

- Aufzeichnung von Messwerten

# Messwerverfassung



# ChipScope Ansicht



The screenshot displays the ChipScope interface with three main windows:

- Waveform - DEV:4 MyDevice4 (XC5VLX50T) UNIT:0 MyILA0 (ILA)**: Shows a digital waveform with time steps from -2 to 7. The 'ReadingsFIFO\_DataOut' signal is highlighted in yellow, showing values: 66, 32514, 504938, 4267, 6314, -1, 4251, 8423.
- Listing - DEV:4 MyDevice4 (XC5VLX50T) UNIT:0 MyILA0 (ILA)**: A table showing the sequence of data samples and their corresponding ReadingsFIFO\_DataOut values.
 

Sample	ReadingsFIFO_DataOut	Reading..
0	66	1
1	66	1
2	66	1
3	66	1
4	66	1
5	66	1
6	66	1
7	66	1
8	66	1
9	66	1
10	66	1
11	32514	1
12	504938	1
13	4267	1
14	6314	1
15	-1	1
16	4251	1
17	8423	1
18	-1	1
19	4282	1
20	12704	1
21	-1	1
22	4430	1
23	21347	1
24	-1	1
25	474240	1
26	508161	1
27	-1	1
28	516341	1
29	584273	1
30	-1	1
31	479373	0
32	479373	0
33	479373	0
- VIO Console - DEV:4 MyDevice4 (XC5VLX50T) UNIT:0 MyILA0 (ILA)**: Shows the state of various signals.
 

Bus/Signal	Value
SATA Generation	1
LinkOK	<input type="checkbox"/>
HardReset	<input type="text" value="0"/>
Reset_Phy (new speed neg.)	<input type="text" value="0"/>
Reset_Link (COMRESET)	<input type="text" value="0"/>
Reset_ATASC	<input type="text" value="0"/>
SATA Generation Min	<input type="text" value="0"/>
SATA Generation Max	<input type="text" value="0"/>
- VIO Console - DEV:4 MyDevice4 (XC5VLX50T) UNIT:0 MyILA0 (ILA)**: Shows the state of other signals.
 

Bus/Signal	Value
RunAC_Address	6
FSM = IDLE	<input type="checkbox"/>
Run_Complete	<input type="checkbox"/>
ILA_Complete	<input type="checkbox"/>
Command_Run	<input type="text" value="0"/>
ReadingsFIFO_got	<input type="text" value="0"/>
- VIO Console - DEV:4 MyDevice4 (XC5VLX50T) UNIT:0 MyILA0 (ILA)**: Shows the state of LCD lines.
 

Bus/Signal	Value
LCD Line 1	Port0 LinkOK@1G
LCD Line 2	WDC WD5000AAKX-0

# Literaturverzeichnis

[1] <http://www.xilinx.com/>

- Xilinx User Guides: UG190, UG191, UG196
- Xilinx Datasheets: DS202
- Xilinx Specifications: LocalLink Interface Specification (SP006)

[2] <http://www.T13.org/>

- AT Attachemnet 8 – ATA/ATAPI Architecture Model (ATA8-AAM; T13.1700-D)
- AT Attachemnet 8 – ATA/ATAPI Command Set (ATA8-ACS; T13.1699-D)
- AT Attachemnet 8 – ATA Serial Transport (ATA8-AST; T13.1697-D)

[3] <http://www.Intel.com/>

- AHCI – Serial-ATA – Adadvanced Host Controller Interface Specification 1.3
- PHY Interface for PCI Express 3.0 Architecture – PIPE 2.00

[4] <http://www.serialata.org/>

- Serial ATA: High Speed Serialized AT Attachment – Rev. 1.0a

[5] <http://www.iso.org/>

- Open System Interconnection - Basic Reference Model (ISO 7498-1:1994)

[6] SATA Storage Technology

- Donovan Anderson, MindShare Press, April 2007
- ISBN: 978-0-9770878-1-5



# Implementierung eines auf Streaming optimierten SATA-Host-Bus-Adapters (HBA)

Patrick Lehmann

Dresden, 11.04.2012





# Anhang

## 2.1 Electrical Physical Layer

### Transceiver Interface:

- Takt                      ClockIn\_150MHz, SATA\_Clock
- Reset                    Reset, ResetDone  
                              ClockNetwork\_Reset, \*\_ResetDone
- Controllpfade          Command, Status, TX\_Error, RX\_Error
- Datenpfade             TX\_Data, RX\_Data  
                              TX\_CharIsK, RX\_CharIsK
- OOB-Signaling         OOB\_Command, \*\_Status, \*\_Complete
- Sonstiges               SATAGeneration, HandshakingComplete
- Reconfig-Port         Reconfig, Complete, Reloaded, Lock, Locked
  
- LVDS-Signale:         TX\_n, TX\_p, RX\_n, RX\_p

[3] PHY Interface for PCI Express 3.0 Architecture – PIPE 2.00

## 2.1 Electrical Physical Layer

### Transceiver Submodule:

- ClockNetwork      Virtex5:    BUFG, DCM, „ClockMux“  
                         Virtex6:    BUFG, BUFR, MMCM  
                         Virtex7:    BUFG, BUFR, MMCM, „Quad-PLL“
- MGT                      Virtex5:    GTP\_DUAL, GTX  
                         Virtex6:    GTXE1  
                         Virtex7:    GTXE2
- BitWidthConverter    Virtex5:    8 ⇔ 32 Bit, inkl. Byte Alignment
- WordAligner            Virtex6:    16 Bit Alignment in 32 Bit-Worten
- Reconfigurator        Virtex5:    Reprogrammierung des FPGA
- Sonstige Logik        Transceiver Interface ⇔ spez. MGT Signale  
                         Reset-Netzwerk  
                         Crossclock-Pfade

## FPGA Ressourcenverbrauch

	<b>Virtex5 StreamingDB</b>
LUTs	ca. 2220
Slices	ca. 1180
Register	ca. 1260
Gesamtverbrauch <sup>1)</sup>	16 %

<sup>1)</sup> Gesamtverbrauch = max(LUTs, Slices, Register)