TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Informatik, Inst. für Tech. Informatik, Prof. für VLSI-Entwurfssysteme, Diagnostik und Architektur

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

Kolloquium zur Bachelorarbeit

Patrick Lehmann

Dresden, 20.06.2012

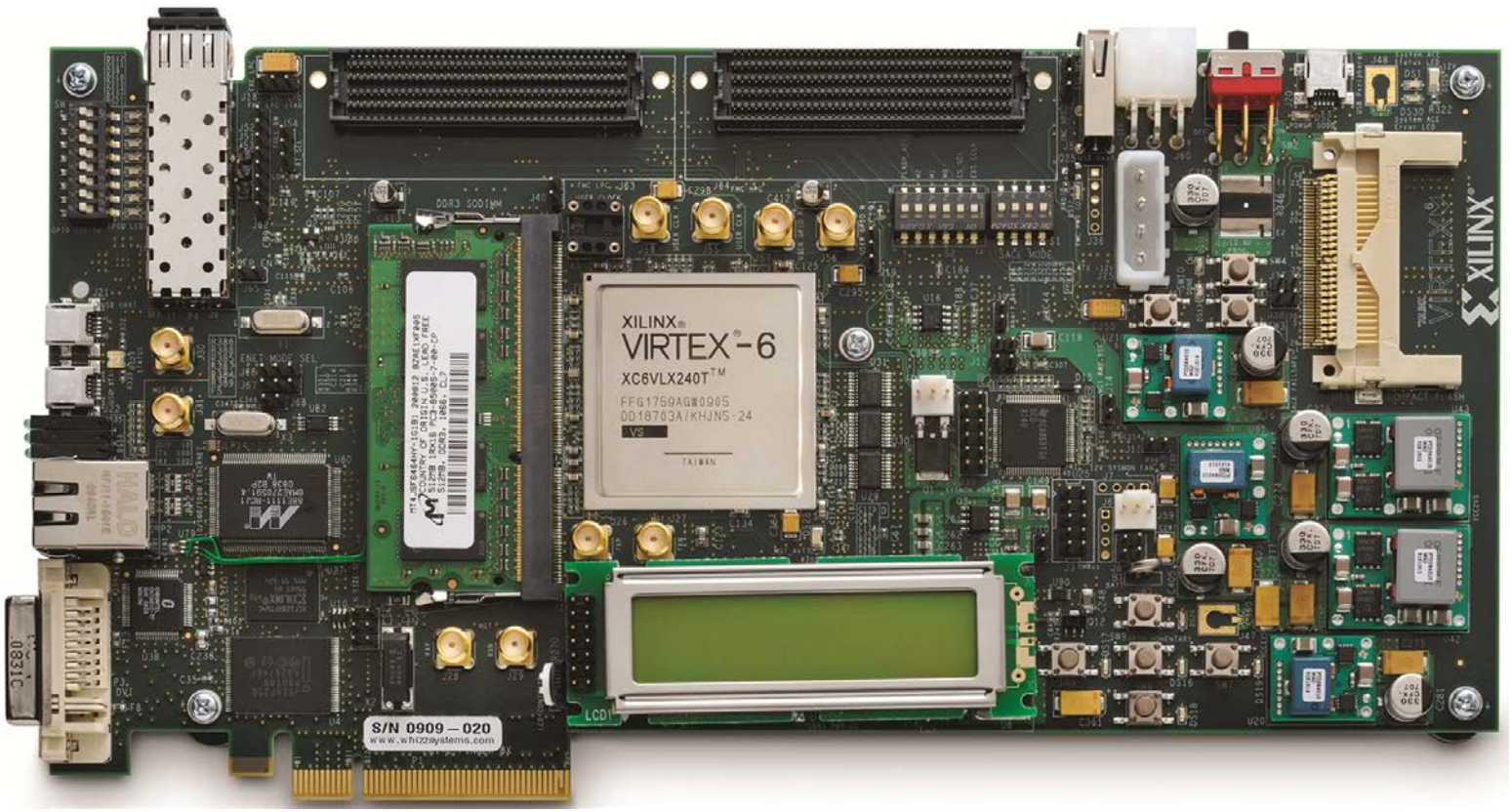
Implementierung eines auf Streaming optimierten SATA-Host-Bus-Adapters (HBA)

Versuchsaufbau aus: ML505 Developer Board, Samsung SP1, Xilinx Programmer und ATX Netzteil

Agenda

1 Einsatzszenario

1.1 FPGA-basierte Hardwarebeschleuniger



[1] Xilinx Press Kits, 08.12.2009 - Xilinx Virtex 6 FPGA ML605 Demonstration Board

1 Einsatzszenario

1.1 FPGA-basierte Hardwarebeschleuniger

Beispiel: Xilinx Virtex 6 (XC6VLX760)	
Recheneinheiten	118.560 Slices, 8.280 DSP Blöcke
Speicher intern	474.240 Flip-Flops 3.240 KiB Block RAM
Speicher extern	Typisch: 512 MiB – 4 GiB DDR3-SDRAM an 2 Kanälen Maximal: bis zu 8 Kanäle á 128 Bit Datenpfad
Datenanbindung	PCI Express 2.x/3.x mit bis zu 16 Lanes Gigabit-Ethernet SFP-Steckverbinder für InfiniBand, FibreChannel SATA

[1] Xilinx Virtex 6: Family Guide [DS150]

1 Einsatzszenario

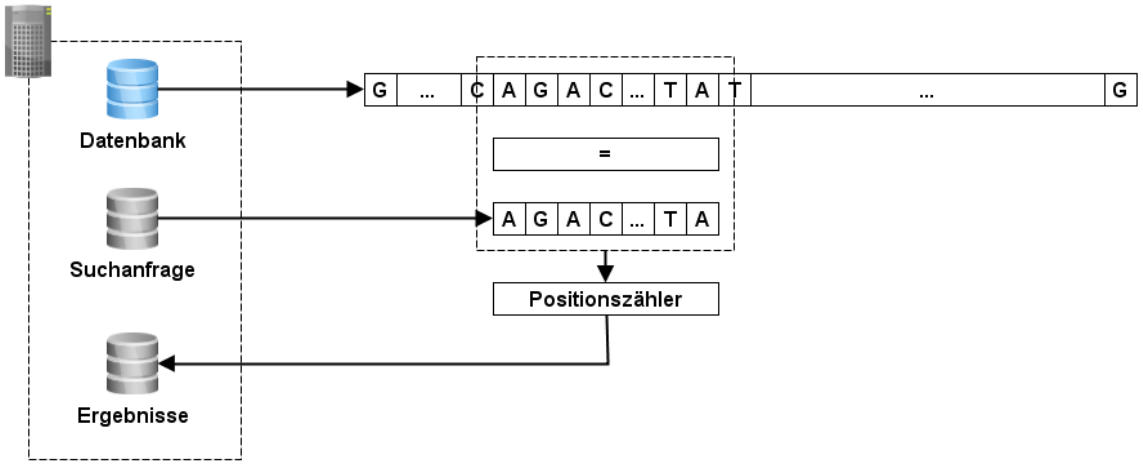
1.2 Short-Read-Mapping-Problem (SRMP)

Eingabe:

- Eine Datenbank von Genom-Sequenzen
- Suchanfrage bestehend aus einem Short-Read

Ausgabe:

- Positionen und Längen der Übereinstimmungen

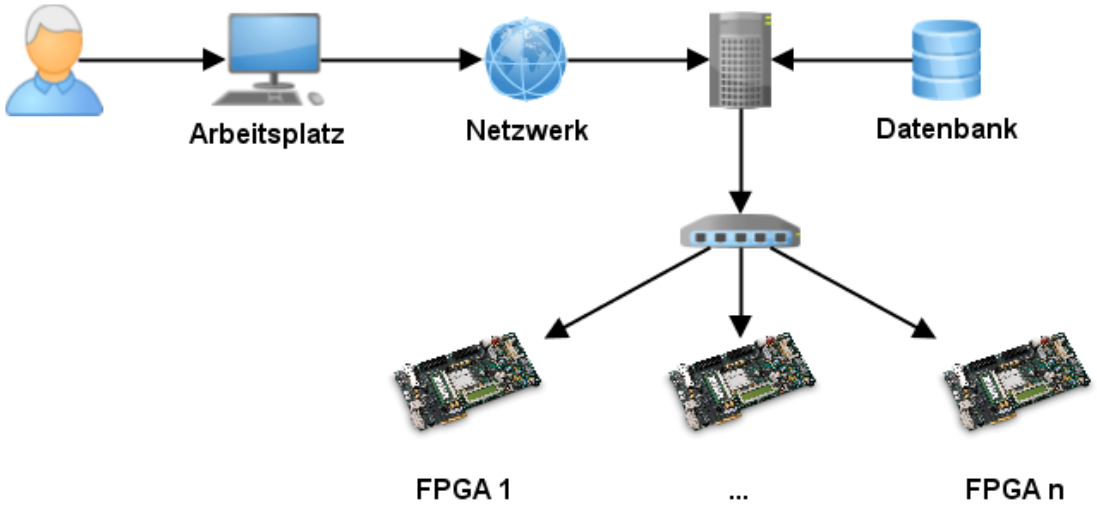


1 Einsatzszenario

1.2 SRMP - Gesamtarchitektur

Datentransport zwischen Server und FPGA:

- Datenbank (Konstantendaten)
- Suchanfragen
- Ergebnisse



2 Schichtenarchitektur

2.1 SATA- / ATA-Referenzarchitektur

#	OSI-Layer	Modulname	Aufgaben
7	Application		Short-Read-Mapper
6	Presentation	StreamingDB	Verwaltung einer Zuordnungstabelle
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*		*entfällt bei SATA bis Rev. 2.0*
2	Link	LinkLayer	logische Verbindung (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

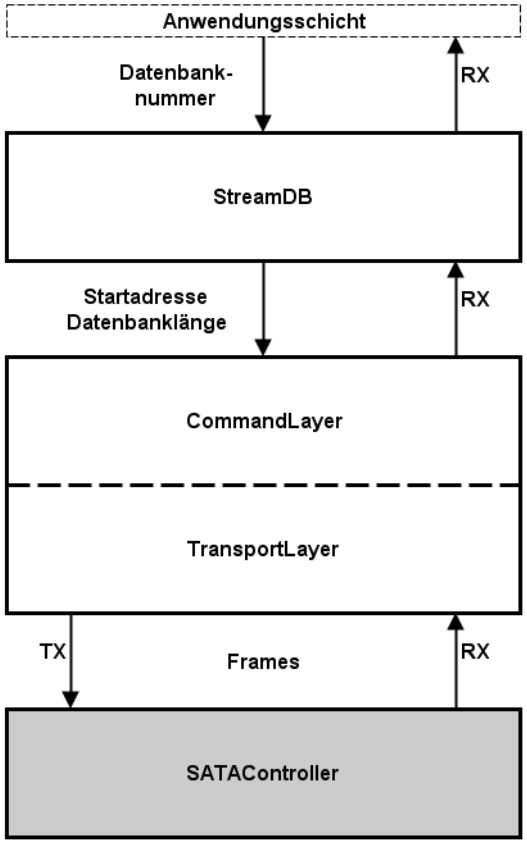
[2] AT Attachement – ATA/ATAPI Architecture Modell – ATA8-ATM
[4] ISO 7498-1:1994 – Open System Interconnection – Basic Reference Modell

2 Schichtenarchitektur

2.2 Anforderungskatalog

Anfrage über eine Datenbanknummer
↓
Ermittelt Datenbank-position und -länge
↓
Programmiert Register mit ATA8 Befehlen
↓
Erzeugt Frames
↓

Übersetzt Frames in Bitstrom



Datenbank

↑
(Überprüft Daten per CRC)
↑
Erneuert Register mit Statusmeldungen
↑
Zerlegt Frames
↑

Übersetzt Bitstrom in Frames

2 Schichtenarchitektur

2.2 Anforderungskatalog

Allgemein:

- Ein FPGA- und Hersteller-unabhängiges Interface
- Konfiguration des MGT für Serial-ATA
- Kapselung des FPGA-spezifischen Taktnetzwerkes
- Geschwindigkeitsumschaltung
- Erkennung von leeren Ports/neu angesteckten Geräten

Anwendungsschnittstelle:

- Lese- und Schreibzugriff auf Datenspeicher
- Abstraktion der Datenspeicher-spezifischen Adressierung
- Adressierung mehrerer Datenblöcke durch eine kompakte Datenbanknummer

2 Schichtenarchitektur

2.3 ATAStreamingController

TransportLayer Dienste:

- ATA Paket- und Sitzungsverfolgung
- Erzeugen und Auswerten von FISes
- Programmiermodel: ATA-Register
- Transparente Registertransfers

CommandLayer Dienste:

- Programmierung der ATA-Register mit ATA8-Befehlen
- Transfergenerierung
- Initialisierung des Gerätes / Lesen von Geräteparametern
- Adressumrechnung von Geräte-unabhängigen in Geräte-abhängige Adressen

[5] SATA Storage Technology, MindShare Press

2 Schichtenarchitektur

2.4 Abschätzung der Datentransferrate

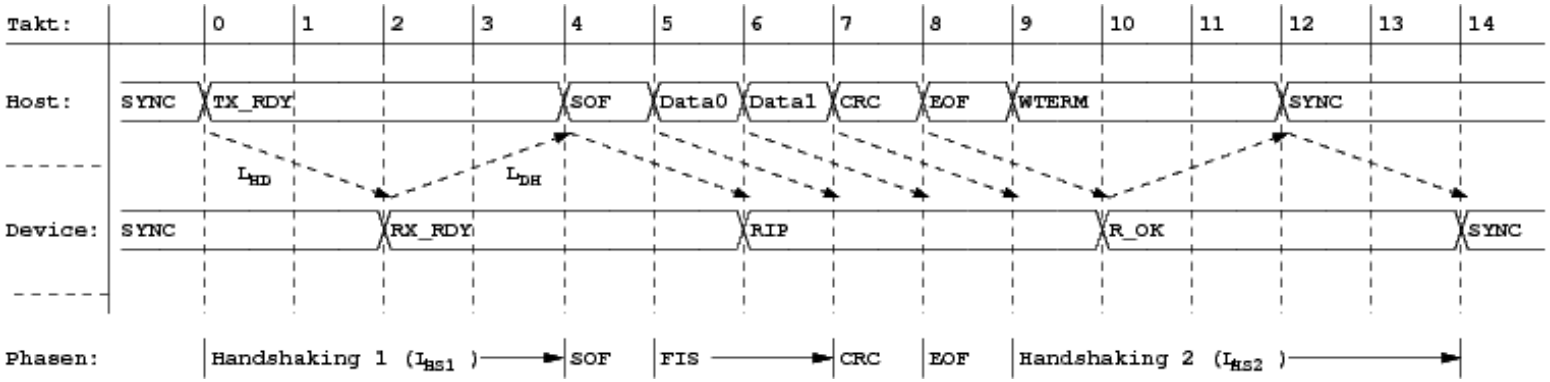
SATA Generation 1	Datenrate	„Verlust“
Symbolfrequenz	1,5 Gigaherz	
Kanalbitrate	1,5 Gigabit/s	
Kanaldekodierung (8B10B)	1,2 Gigabit/s (= 150 MB/s)	25,0 %
Basisumrechnung	143 Mebibyte/s	4,6 %

SATA Generation 2	Datenrate	„Verlust“
Symbolfrequenz	3,0 Gigaherz	
Kanalbitrate	3,0 Gigabit/s	
Kanaldekodierung (8B10B)	2,4 Gigabit/s (= 300 MB/s)	25,0 %
Basisumrechnung	286 Mebibyte/s	4,6 %

[3] Serial ATA: High Speed Serialized AT Attachment

2 Schichtenarchitektur

2.4 Abschätzung der Datentransferrate



$L_{HD} \approx 11\ \text{SATA} - \text{WORT} - \text{Takte}$

$T_{HD\_Register} = L_{HS1} + L_{SOF} + L_{DATA(5)} + L_{CRC} + L_{EOF} + L_{HS2} \approx 63\ Takte$

$T_{Transfer} = L_{HD\_Register} + N_{Frame} \cdot L_{Data(2048)} + L_{DH\_Register}$

$$T_{Request} = L_{DATA(2048)} + (N_{Frame} - 1) \cdot L_{DATA(2048)} + (N_{Transfer} - 1) \cdot (L_{HD\_Register} + L_{DH\_Register} + L_{TransferPause})$$

3 Ergebnisse

3.1 FPGA Ressourcenauslastung

	SATAC + ATASC	ISCID.2009.124[2]	FCCM.2012.45[2]
LUTs	ca. 460 (2.080) [3]	ca. 2.610	ca. 1.760
Slices	ca. 510 (1.100) [3]	unbekannt	ca. 750
Register	ca. 1.170 (1.180) [3]	ca. 2.230	ca. 850
Gesamtverbrauch[1]	**7 % (15 %)**	**19 %**	**10 %**

[1] Gesamtverbrauch = max(LUTs, Slices, Register)
[2] inklusive Schreibunterstützung
[3] Werte des Zwischenvortrags (April 2012)

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

FCCM.2012.45
Groundhog – A Serial ATA Host Bus Adapter (HBA) for FPGAs
Louis Woods, Ken Eguro
IEEE 20th International Symposium on Field-Programmable Custom Computing Machines, 2012

3 Ergebnisse

3.1 FPGA Ressourcenauslastung

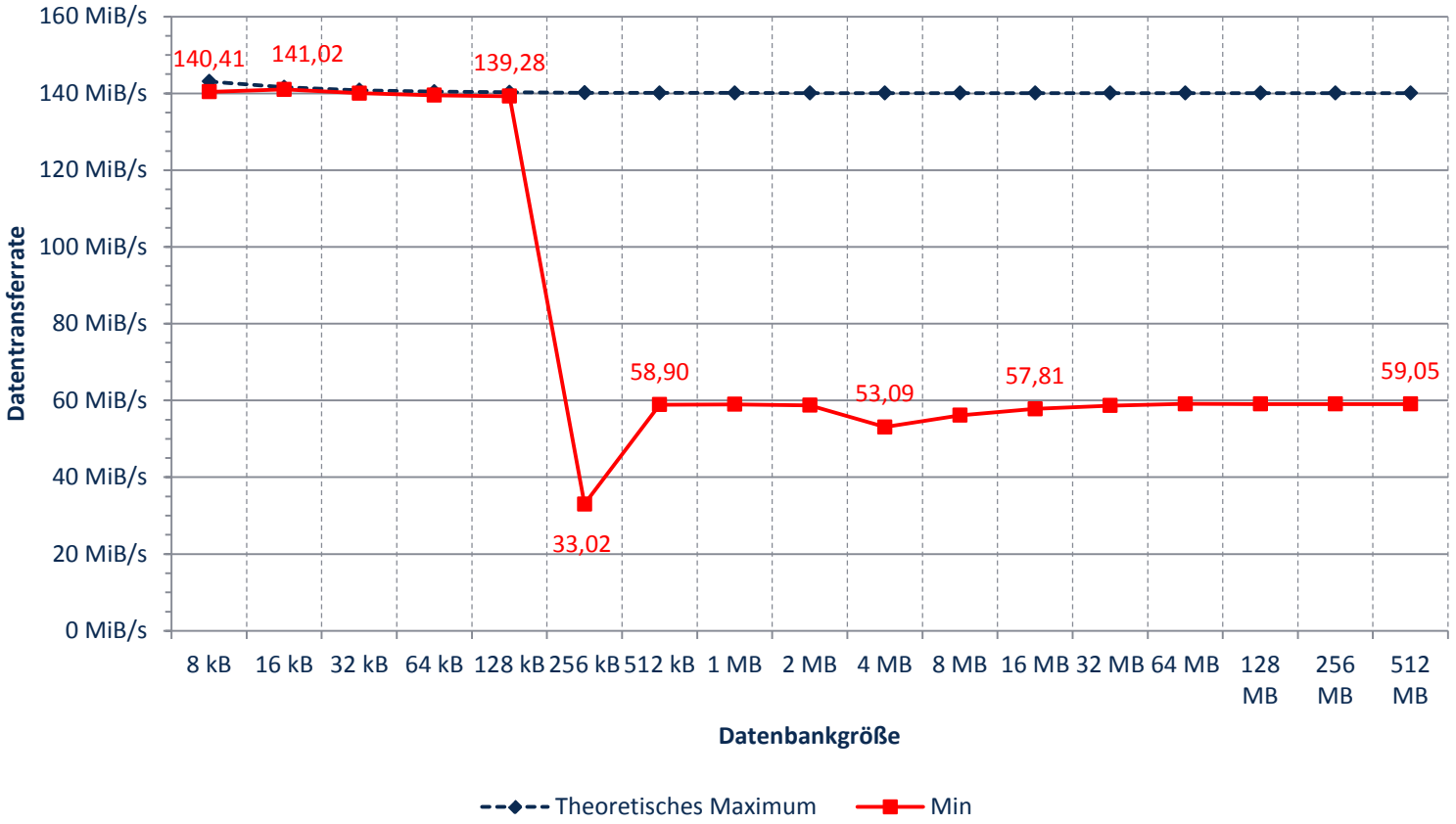
VHDL Modul	Slices	Reg.	LUTs	BlockRAM	BUFGs	DCMs
Gesamt[1]	7.200	28.800	28.800	60	32	12
SATAController	272	597	192	4	3	1
ATAStreamingCont.	238	569	267	4 (9) [2]	-	-
StreamDB	27	60	37	2	-	-
Summe	**537**	**1.226**	**496**	**10 (15) [2]**	**3**	**1**

[1] Xilinx Virtex 5 (XC5VLX50T)

[2] Fehlerhafte Erkennung einer 34-Bit FIFO während der Synthese

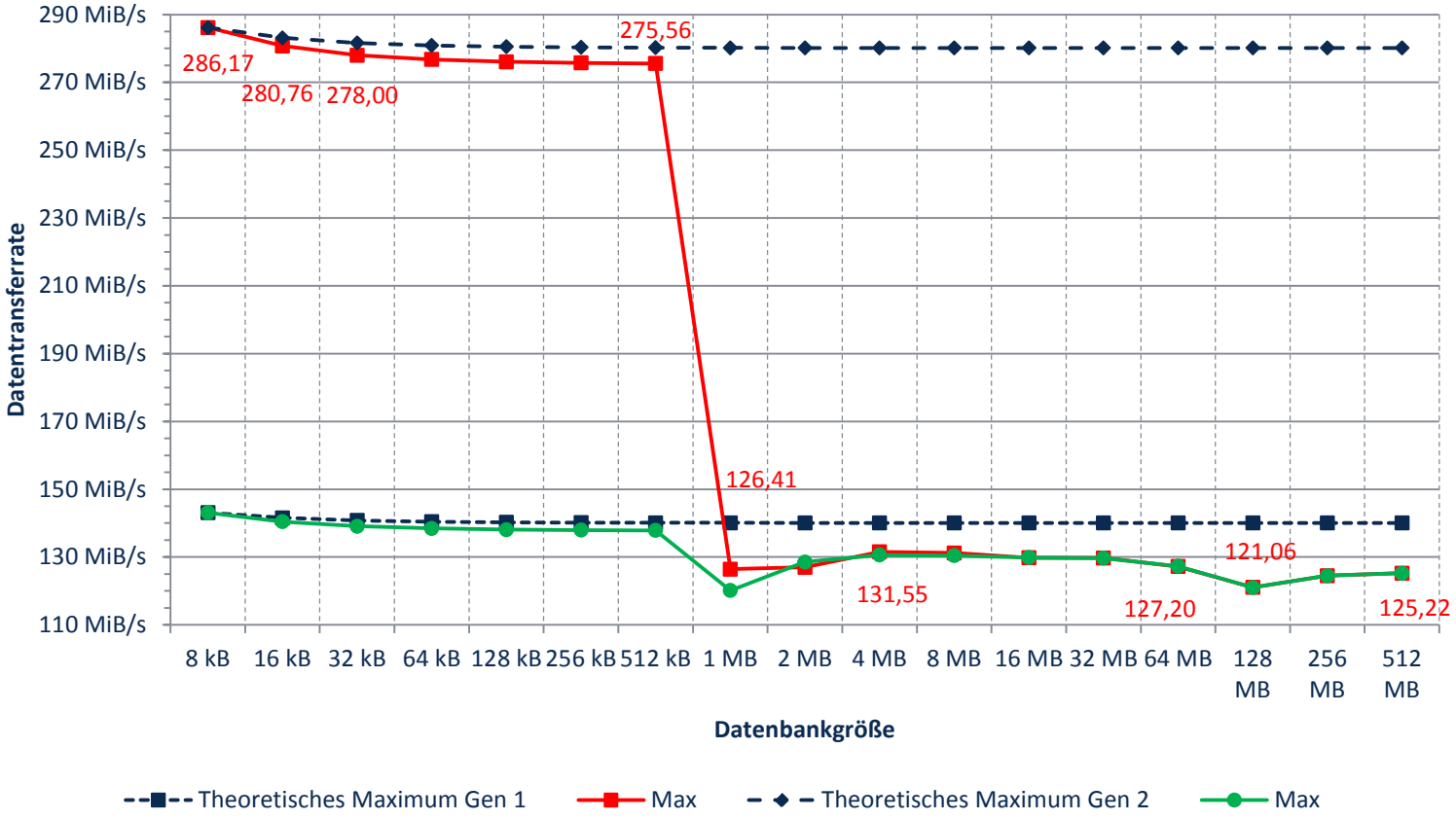
3 Ergebnisse

3.2 Messergebnisse – Samsung SP1 (160 GB)



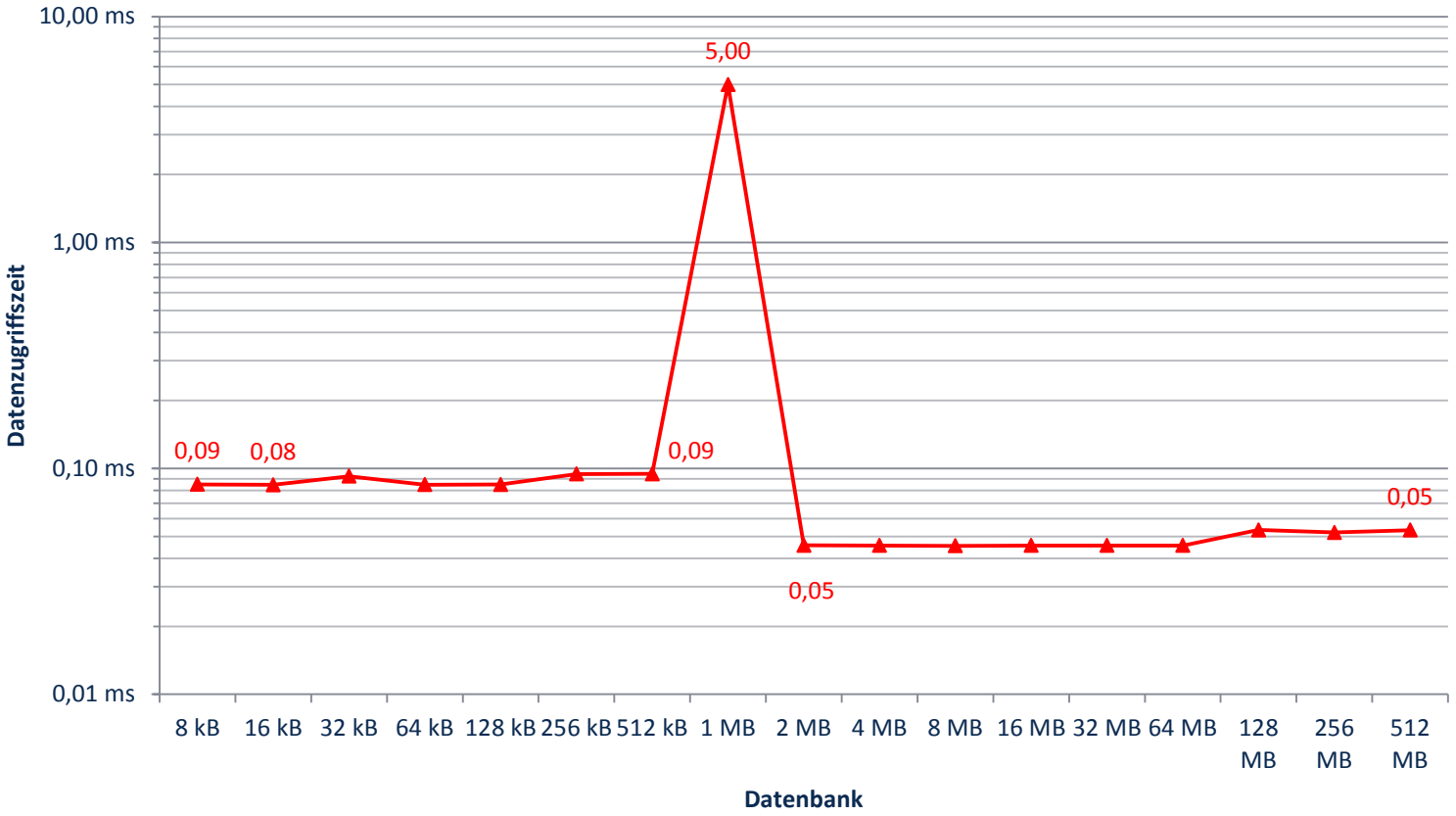
3 Ergebnisse

3.2 Messergebnisse – Western Digital (500 GB)



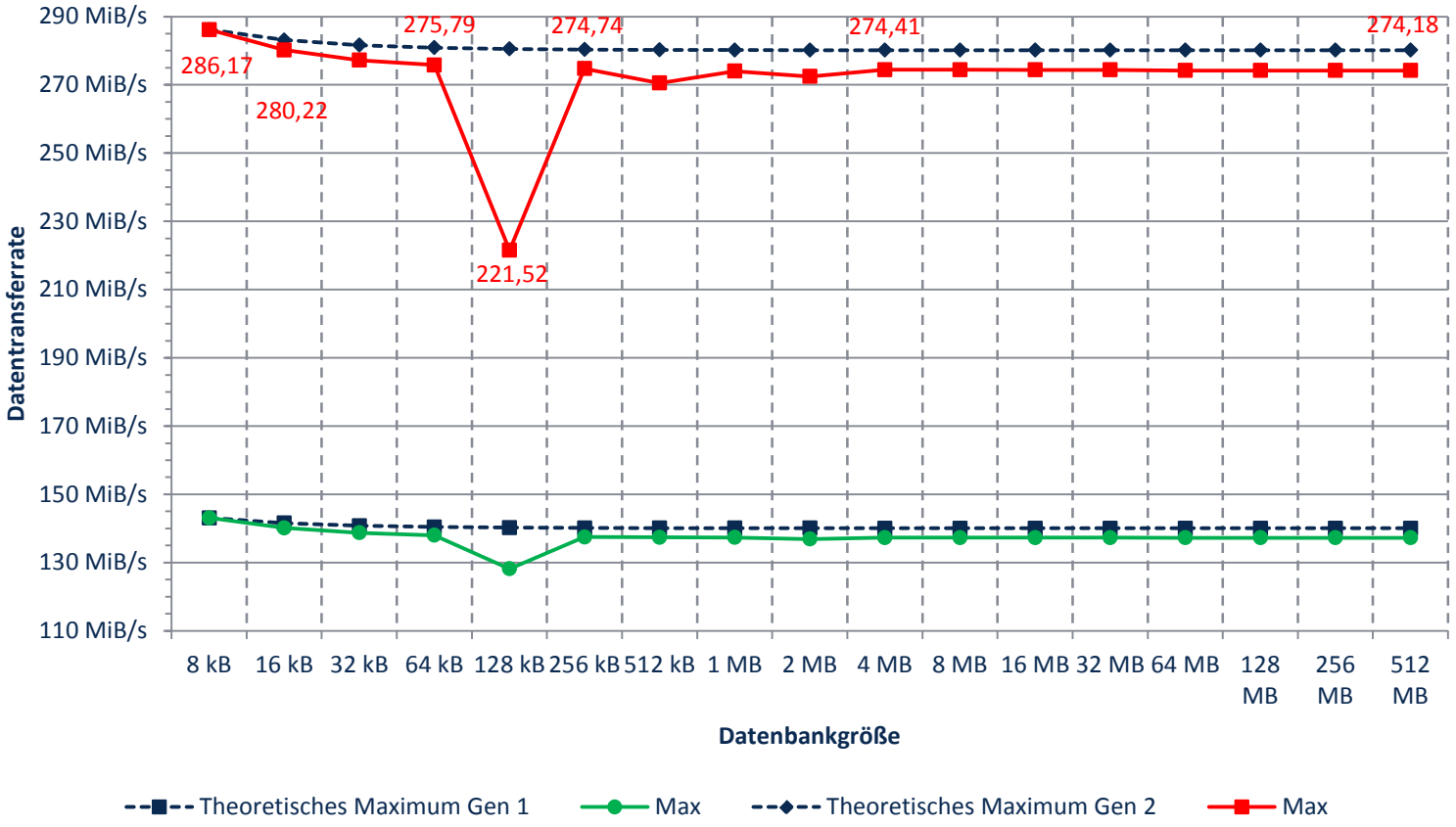
3 Ergebnisse

3.2 Messergebnisse – Western Digital (500 GB)



3 Messergebnisse

3.2 Messergebnisse – OCZ Vertex3 (120GB, SSD)



4 Zusammenfassung

Zusammenfassung

SATAController:

- Ein stabieler Link wird hergestellt und gehalten
- SATA Generation 1 und 2 werden unterstützt

ATAStreamingController:

- Kompatibel zu Festplatten und SSDs
- Datentransferraten von bis zu 285 MiB/s

StreamDB:

- Testdatenbanken bis 512 MiB werden übertragen
- Die CRC16-Prüfung war erfolgreich

4 Zusammenfassung

Ausblick

SATAController:

- Fehlerhafte Frames wiederholt senden
- Test auf dem Virtex 6
- Portierung auf eine Altera Plattform (Stratix2GX)

ATAStreamingController:

- Implementierung des Schreib-Datenpfades
- Anbindung an den Java-Bytecode-Prozessor SHAP

Literaturverzeichnis

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

- Xilinx Datasheets: DS150
- Xilinx Press Kits: 08.12.2009, Xilinx Virtex 6 FPGA ML605 Demonstration Board

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

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

[3] http://www.serialata.org/

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

[4] http://www.iso.org/

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

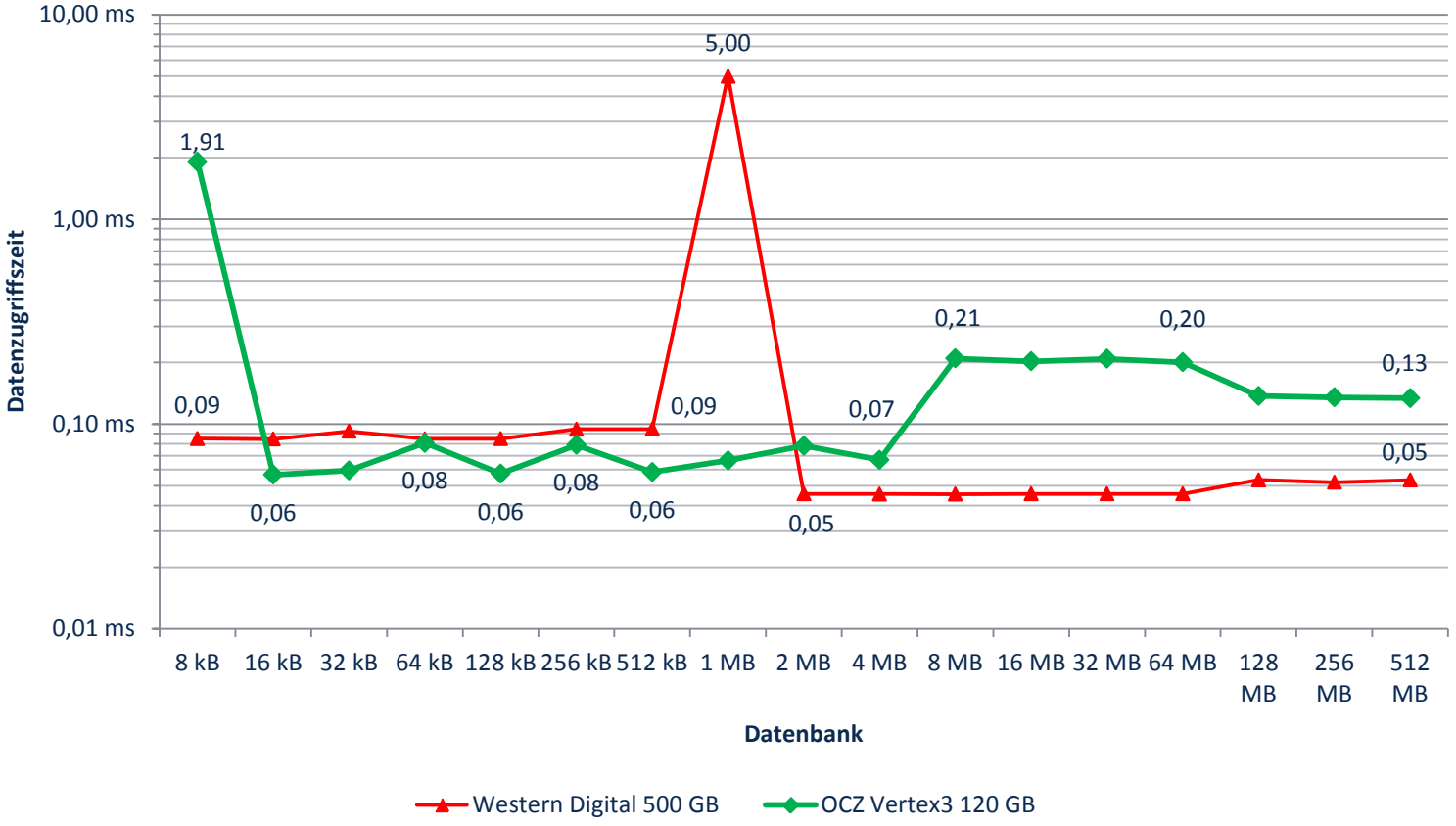
[5] SATA Storage Technology

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

Anhang

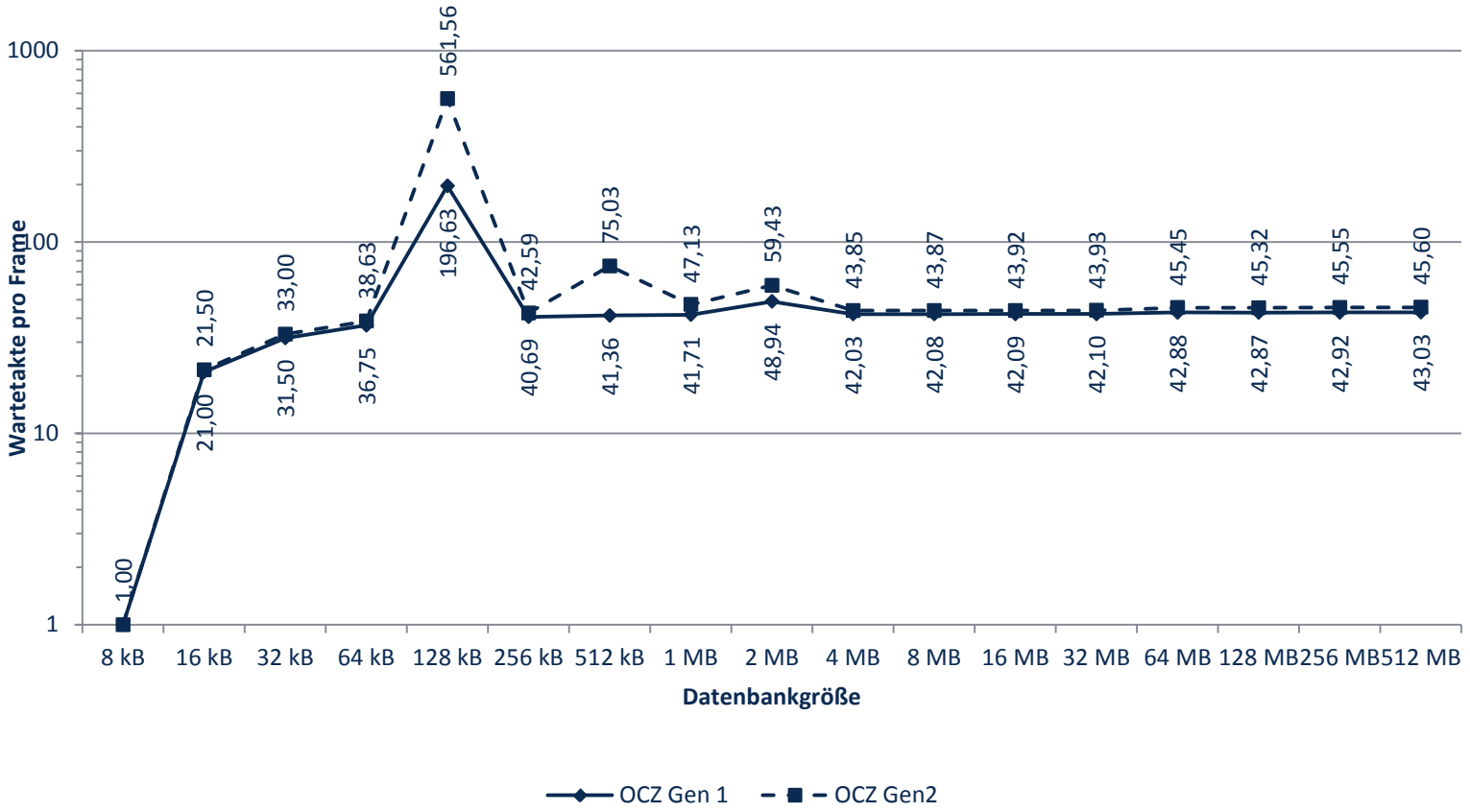
3 Ergebnisse

3.2 Messergebnisse – Western Digital (500 GB)



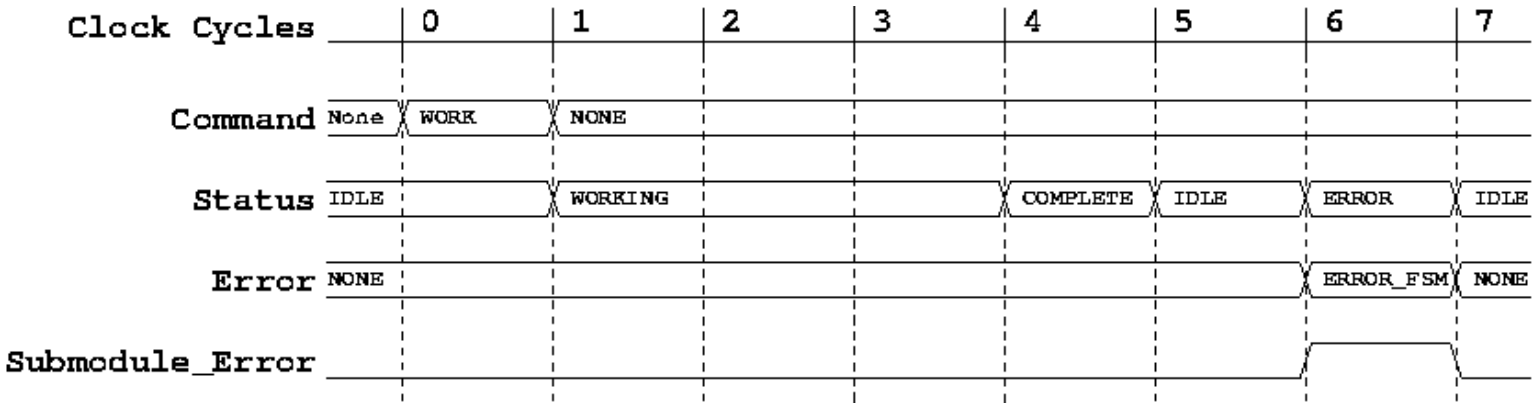
Anhang

Wartetakte – OCZ Vertex 3 (120 GB, SSD)



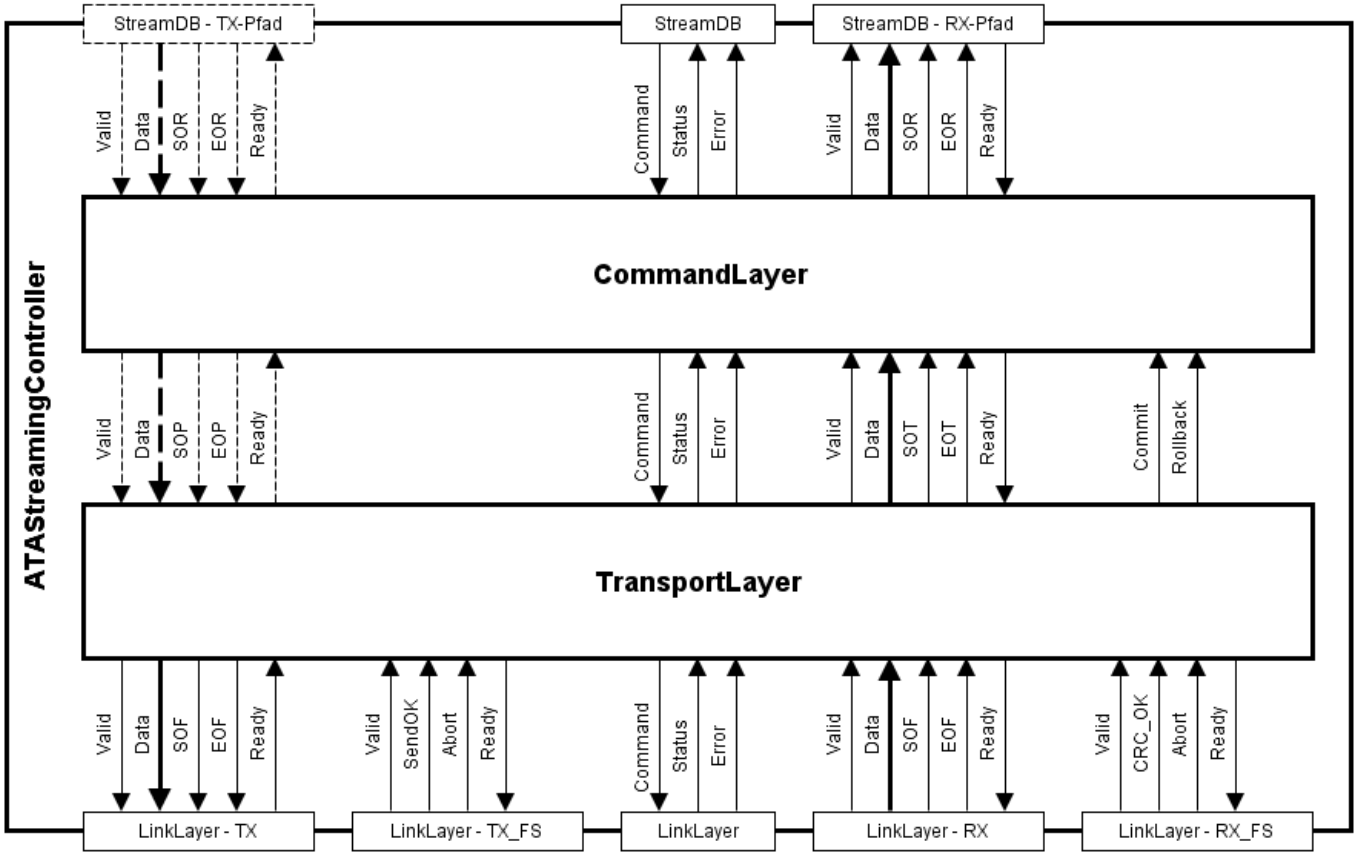
Anhang

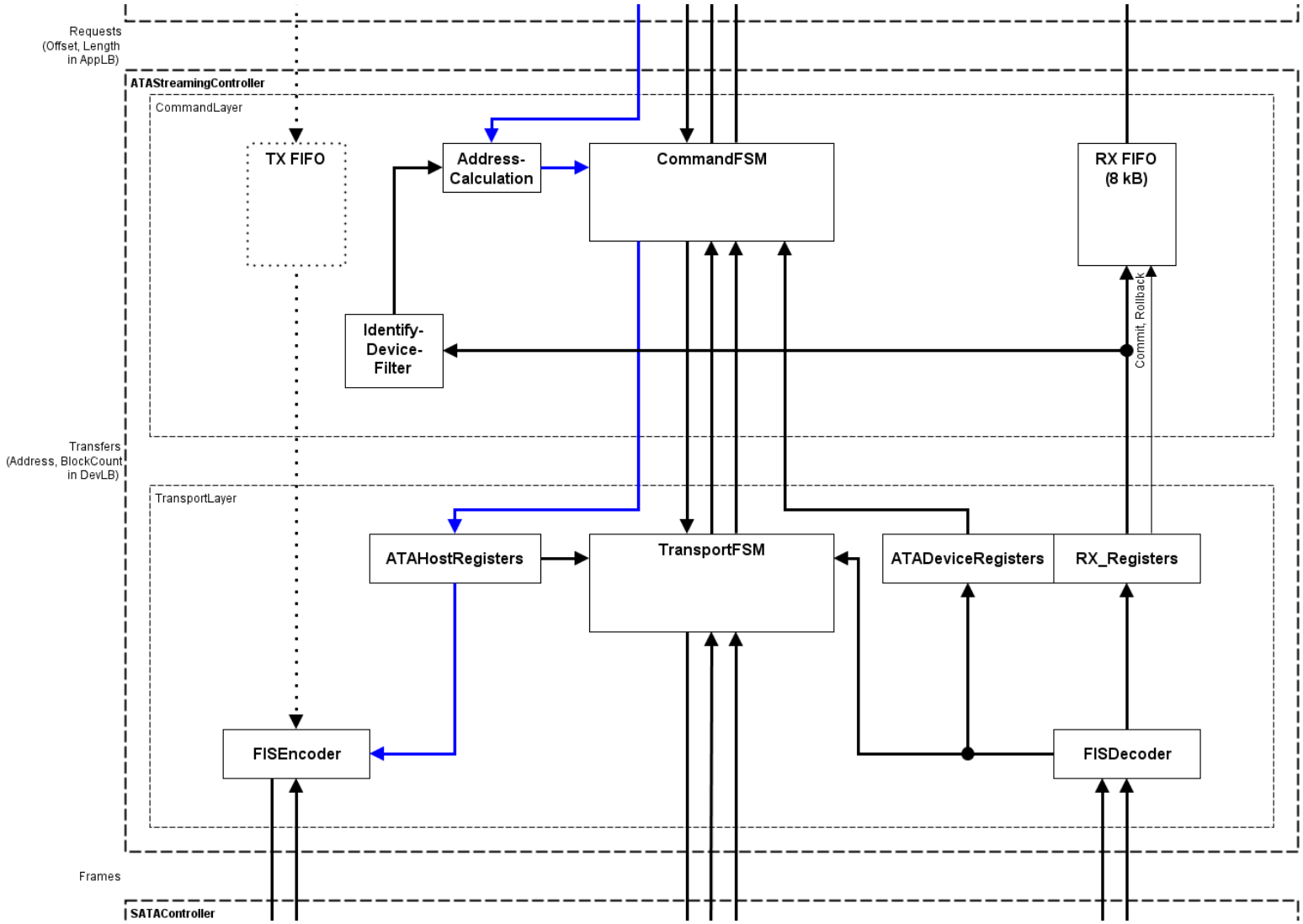
Command-Status-Error - Interface



Anhang

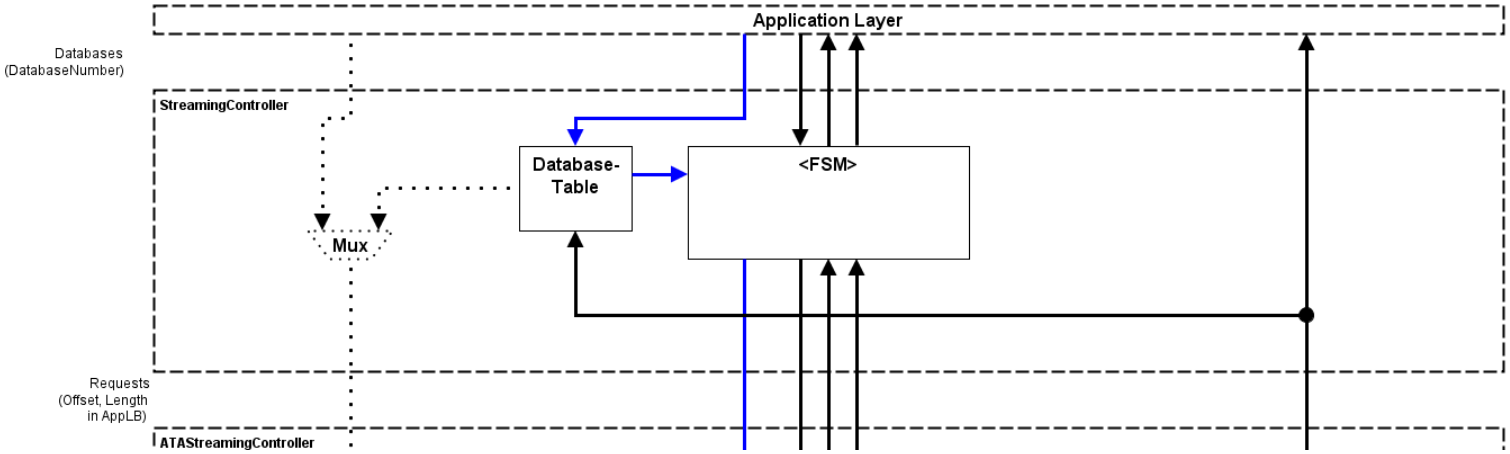
ATAStreamingController




Requests
(Offset, Length
in AppLB)
ATAStreamingController
CommandLayer
TX FIFO
Address-
Calculation
CommandFSM
RX FIFO
(8 kB)
Identify-
Device-
Filter
Commit, Rollback
Transfers
(Address, BlockCount
in DevLB)
TransportLayer
ATAHostRegisters
TransportFSM
ATADeviceRegisters
RX_Registers
FISEncoder
FISDecoder
Frames
SATAController

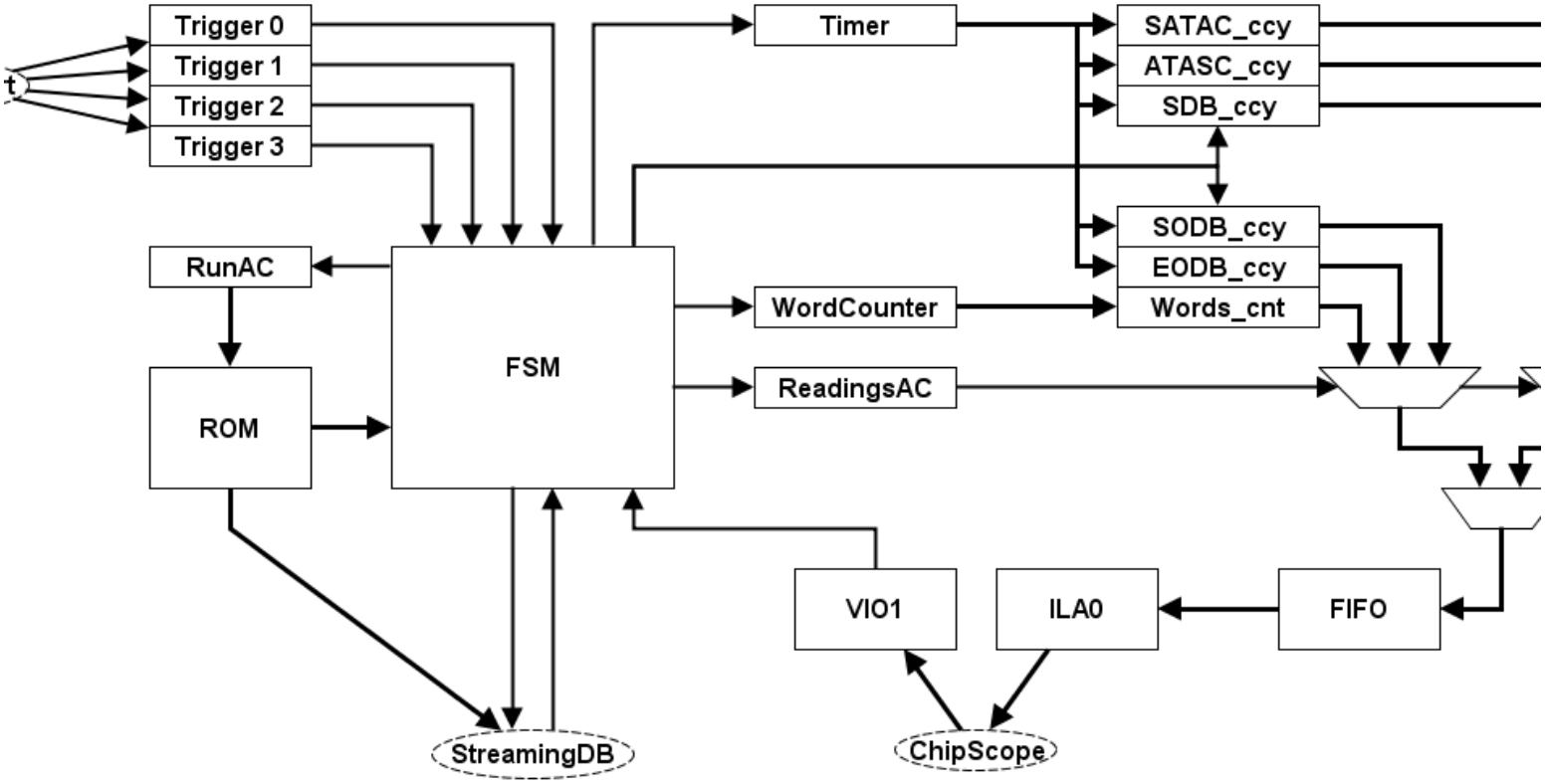
Anhang

StreamDB



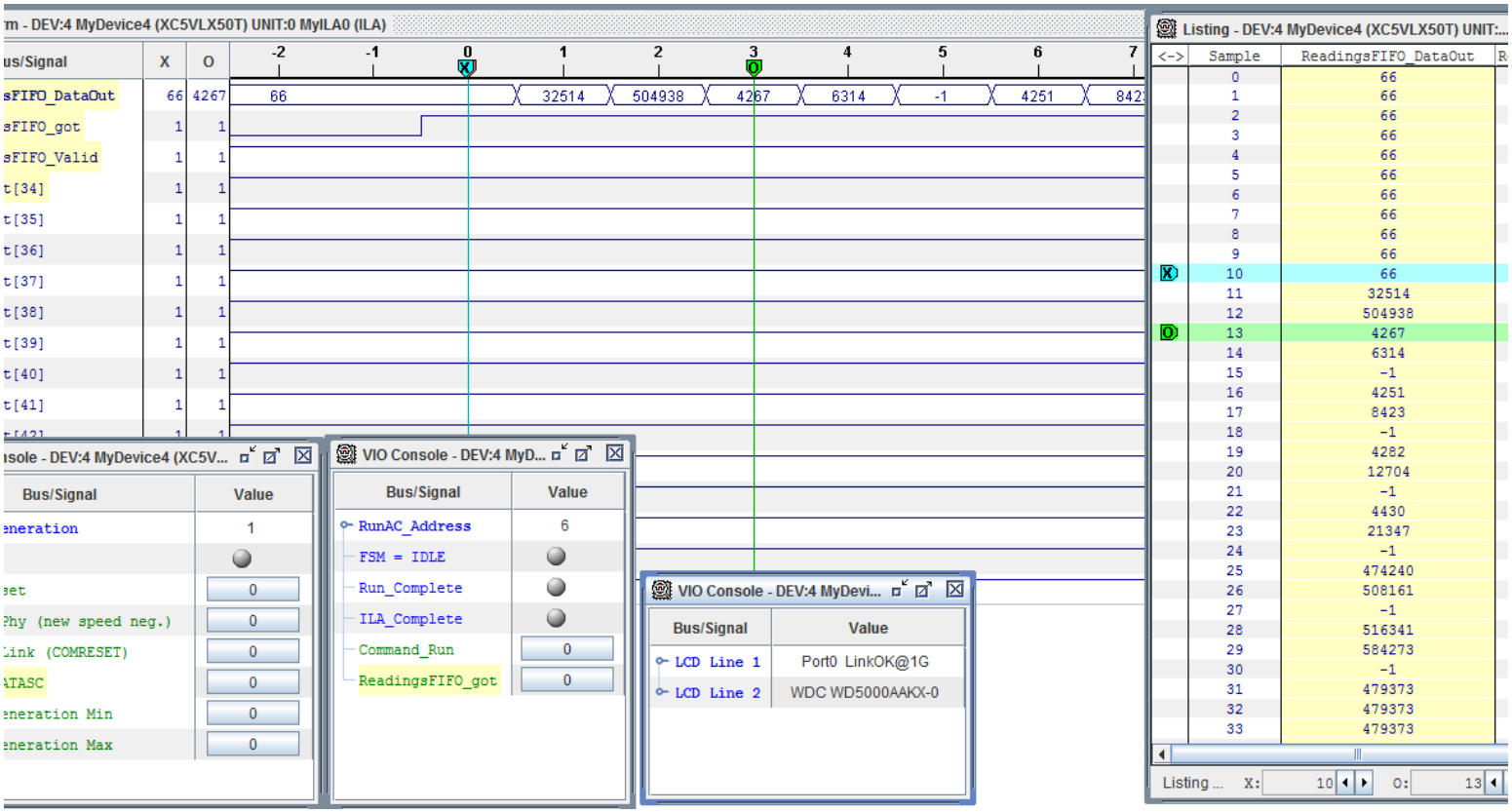
Anhang

Messwerterfassung - ATASCMeasurement



Anhang

Messwerterfassung - ChipScope Ansicht



Anhang

Aufbau der MasterTable

MasterTable (MT)

- Erster 8 KiB 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 KiB

⇒ Festplatten bis zu 32 TiB Gesamtkapazität

⇒ 1016 Datenbanken von 8 KiB bis 32 TiB

Anhang

Aufbau der MasterTable

[Bytes]	4		8		12	
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						

Anhang

Host Device Register FIS

	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
DW0	Features (7:0)								Command (7:0)								C 1	R	R	Reserved					FIS Type (27h)							
DW1	Device (7:0) R	L 1	R	D	R	R	R	R	LBA (23:0)																							
DW2	Features (15:8)								LBA (47:24)																							
DW3	Control (7:0) H	R	R	R	R	S 0	N	Z 0	Reserved								Count (15:0)															
DW4	Reserved								Reserved								Reserved								Reserved							

[2] AT Attachement 8 – ATA Serial Transport (ATA8-AST)

Anhang

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

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

Anhang

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

Anhang

GPGPU-basierte Hardwarebeschleuniger

GPGPU-Grafikkarte (GeForce GTX 690):

- Recheneinheiten je GPU:
  - 1536 Stream-Prozessoren
  - 128 Textureinheiten
  - 32 Rasterendstufen
- Speicher je GPU:
  - 256 Bit Speicherinterface
  - 2 GB GDDR5-SDRAM
  - Theoretische Speicherbandbreite: 192 GiB/s
- Anbindung:
  - PCI Express 3.0 mit 16 Lanes

[X] http://www.nVIDIA.com