doi:10.11887/j.cn.201501007

http://journal. nudt. edu. cn

# 高密度存储服务器高速链路设计与仿真\*

沙超群1,任雪玉2,胡长军1,聂 华1,陈 进3,4

(1.北京科技大学 计算机与通信工程学院,北京 100083;2. 曙光信息产业有限公司,北京 100193; 3.哈尔滨工程大学 水声工程学院,黑龙江 哈尔滨 150001;4. 中国科学院 微电子研究所,北京 100029)

摘 要:针对高密度存储服务器中高速链路结构复杂、信号速率高、链路长度长等特点,在高速链路 PCI-E3.0和 SAS3.0设计过程中引入全面的信号完整性仿真。通过对高速印刷电路板设计中拓扑结构、材料类 型、印刷电路板结构等关键项目进行仿真设计,获得成本最低、性能达标的最优方案;通过全链路有源仿真, 预估系统性能,降低系统投产风险;通过系统实际信号测试,验证系统性能完全满足相应规范要求,仿真结果 有效可靠。

关键词:高速印刷电路板设计;信号完整性仿真;信号测试 中图分类号:TP336 文献标志码: A 文章编号:1001-2486(2015)01-039-08

# High-speed channel design and simulation of high density storage server

SHA Chaoqun<sup>1</sup>, REN Xueyu<sup>2</sup>, HU Changjun<sup>1</sup>, NIE Hua<sup>1</sup>, CHEN Jin<sup>3, 4</sup>

(1. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China;

2. Dawning Information Industry Corporation Limited, Beijing 100193, China;

3. College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin 150001, China;

4. Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China)

**Abstract:** Comprehensive signal integrity simulation was introduced in the design of PCI-E3.0 and SAS3.0 channels of high density storage server, which are characterized by complicated structures, high signal speed and long and dense traces. The optimized cost-effective design was obtained by simulating the key items of high-speed PCB design such as topology, material and high-speed via structure. The system performances were then evaluated by full channel active simulation to ensure the stability of the system and reduce the manufacture risks. The signal quality of storage server system totally meets the requirement of PCI-E3.0 and SAS3.0 specifications, verifying the effectiveness and reliability of signal integrity simulation.

Key words: high-speed printed circuit board design; signal integrity simulation; signal test

近年来,全球数据存储量呈现爆炸式增长,当前全球每48h所创造的数据量等同于从人类文明 诞生至2003年间产生数据量的总和<sup>[1]</sup>(其中,企 业与互联网数据以每年50%的速率在增长,据 Gartner预测,到2020年,全球数据量将达到 35ZB,等于80亿块4TB硬盘<sup>[2]</sup>)。数据存储理念 的变化,尤其是大数据、云计算等新的应用模式要 求存储服务器系统不断提升计算能力、存储容量、 出口带宽与存储密度,这对新一代存储服务器系 统设计,尤其是其中高速信号链路设计提出了新 的挑战。

1)链路结构复杂。单个系统中包含有 60~100

块 SAS3.0 硬盘,同时支持 4~8 组PCI-E3.0×16 标 准设备,有数百对高速传输线需要经过系统背板。

2)链路信号速率高。经过背板的高速链路 PCI-E3.  $0^{[3]}$ 和 SAS3.  $0^{[4]}$ 协议的速率分别达 8Gbps 和 12Gbps,传统高速印刷电路板(Printed Circuit Board, PCB)链路阻抗控制(±10%)已无 法满足要求,需要更加精细化的链路结构设计。

3)高速链路走线长。绝大部分的高速链路需要经过3次连接器,跨越4块电路板,最长的链路达1m,增加了链路损耗控制的难度。

针对系统高速链路的特点和要求,在系统设计 过程中引入全面的信号完整性(Signal Integrity,

<sup>\*</sup> 收稿日期:2014-06-12

**基金项目:**国家"863"高技术研究发展计划基金资助项目(2013 AA01 A209);北京市科委计划项目(D141100003414002) 作者简介:沙超群(1977—),男,江苏邳州人,博士研究生,E-mail;scq@ sugon.com; 任雪玉(通信作者),男,工程师,硕士,E-mail;renxy@ sugon.com

SI) 仿真<sup>[5]</sup>, 通过仿真的方法指导高速链路设计, 提高系统高速信号质量。首先基于高速链路拓扑结构特点, 通过 SI 仿真分析以及 3D 电磁场仿真完成材料选型与过孔结构优化, 并最终得到全链路的无源模型(仿真工具为 ANSYS HFSS 15.0.7)。同时根据PCI-E3.0/SAS 3.0 标准规范, 设计PCI-E3.0 和 SAS3.0 有源收发器模型, 完成相应链路的有源仿真(仿真工具为 ANSYS Designer 15.0.7), 评估系统预期性能。最后使用误码仪和高速示波器, 对PCI-E和 SAS 链路的实际性能进行测试和验证。

# 1 系统 SI 仿真设计

#### 1.1 高速链路拓扑设计

系统中PCI-E与 SAS 链路的拓扑结构较为类 似,如图1所示,PCI-E链路从主板出发,经过背板 接到 I/O 扩展模块上,模块上有标准PCI-E SLOT, 用于外接标准PCI-E设备;SAS 链路从 SAS 标准卡 出发,经过背板接到硬盘底板上,硬盘底板可外接 标准硬盘或固态硬盘。链路整体均分为4个部分, 前3个部分为系统内板卡,第4部分为标准设备。





PCI-E最长链路中,主板部分长度约为230mm,背板部分长度约为330mm,I/O扩展板部分长度约为130mm,全链路总长约为690mm;SAS最长链路中,SAS子卡部分长度约为100mm,背板部分约为400mm,硬盘底板部分约为500mm,

全链路总长约为1m。这些链路3次经过连接器 (添加标准设备端),可以作为评价PCI-E和SAS 链路的最劣样本。根据PCI-E与SAS协议规范, 链路损耗需要分别控制在-10dB±2dB(4GHz) 与-18dB±2dB(6GHz)(其中的不确定度来自于 芯片封装性能与链路阻抗连续程度的差异)。

#### 1.2 材料选型

确定链路拓扑后,需要进行材料与连接器的 选型。材料选型原则为兼顾材料的链路损耗要求 及材料成本,避免过设计。因此,选择4种市场上 主流的 PCB 板材进行仿真,同时制作测试 PCB, 通过实测结果验证材料厂商所称的材料参数(仿 真工具为 ANSYS Designer,测试工具为矢量网络 分析仪)。





图 2 为 152mm 带状线的损耗仿真与实测结 果对比,表 1 为仿真拟合前后的材料参数与不同 材料的损耗对比。根据系统拓扑设计结果,PCI-E 链路的损耗在 4GHz 要求为 - 0.37 ± 0.07dB/in, SAS 损耗在 6GHz 要求为 - 0.45 ± 0.05dB/in。结 合仿真结果可知 N4000-13SI 为满足要求的最佳 性价比材料。

表 1	仿真拟合前后的材料参数与不同材料的损耗对比

-1 (117. 1 $-1$ 116.70/11170/11 106.55 <b>XX</b> V/1161164170/110/110/110/13/155 V/1 V/1110/110/110/110/110/110/110/110/110/1	Tab. 1	Insertion le	oss &	characteristic	parameters	of	different	materia
---	--------	--------------	-------	----------------	------------	----	-----------	---------

					I				
<u> 1</u> 4約米刊	材料手册参数			仿真拟合参数			损耗值(dB/in)		相对
材科关型	Dk	Df	铜箔	Dk	Df	$RMS(\mu m)$	4GHz	6GHz	成本
IT-180A <sup>[6]</sup>	4.7	0.018	STD	4.4	0.025	7	0.789	1.127	1
$TU-862HF^{[7]}$	4.5	0.015	RTF	4.4	0.018	5	0.520	0.730	1.3~1.5
N4000-13SI <sup>[8]</sup>	3.2	0.008	RTF	3.4	0.010	5	0.355	0.485	2.5~3
Megtron6 <sup>[9]</sup>	3.4	0.004	VLP	3.4	0.005	3	0.225	0.298	4 ~ 5

注: Dk(Dielectric Constant)表示材料的介电常数,Df(Dissipation Factor)表示材料的损耗因子,STD表示标准铜箔,RTF表示反转铜箔,VLP表示低损耗铜箔,RMS值表征铜箔的粗糙度。

#### 1.3 PCB 结构精细化设计

高速 PCB 链路中的过孔、蛇形绕线、补偿线、 BGA 焊点、AC 耦合电容等结构影响链路的阻抗 连续性,其中,过孔是 PCB 设计中应用最广泛的 结构,也是造成链路阻抗不连续的关键因素之一, 需要进行精细化的设计来提高链路阻抗连续性, 减少链路对高速信号的反射。

对过孔的阻抗有较大影响的结构参数主要 有差分孔距、钻孔孔径、反焊盘半径和过孔根 长度,通过参数扫描仿真可以得到这些参数对 过孔特性的影响,见表2。一般工程实践中的 过孔优化步骤是建立3D过孔模型,通过扫描 不同的结构参数得到阻抗变化最小的结果,但 这种方法往往只能将过孔阻抗的偏差控制在 10%。为了进一步优化过孔阻抗特性,在此基 础上设计了3种特殊的过孔结构,图3为4种 结构(额外添加标准结构作为对照)的结构透 视图。

	Tab. 2 Effect of university via structural parameters						
나기 소粉		<u>기</u>	皆参数值增大(↑	)时特征参数的3	变化		
过扎参数 —	Lo	Cg	Lm	Cm	阻抗	特性	
差分孔距	-	-	$\uparrow$	$\downarrow$	<b>↑</b>	感性	
钻孔孔径	$\downarrow$	$\uparrow$	-	$\uparrow$	$\downarrow$	容性	
反焊盘半径	-	$\downarrow$	-	-	<b>↑</b>	感性	
过孔根长度	-	-	-	-	$\downarrow$	容性、谐振	



• • • •	0000	• • • •	• • • •
(a)标准结构	(b) 纺锤形结构	(c)梯形结构	(d)雨滴形结构
(a) Standard	(b) Spindle	(c) Trapezoid	(d) Raindrop
	图 3 4 种	过孔结构	
	Fig. 3 Four	via structure	

图 3(a)为标准结构;图 3(b)为纺锤形结构, 该设计思想为通过减小外层 PCB 反焊盘大小,降 低过孔两端的阻抗,通过增大内层 PCB 的反焊 盘,提升过孔中段的阻抗;图 3(c)为梯形结构,其 设计思路为通过增加一小块入线层参考面来使得 阻抗更加连续;图 3(d)为雨滴形结构,设计思路 为通过增加入线线宽,降低过孔入线端阻抗。对 这4种结构分别进行仿真与测试验证,仿真与实 测的 TDR 阻抗结果对比如图 4 所示(仿真工具为 ANSYS HFSS,测试工具为矢量网络分析仪)。

由图4可知,纺锤形与梯形结构的阻抗更加 连续(走线阻抗约为102.5Ω),阻抗波动偏差约 为5%,优于经过仿真优化的标准结构(约为 6%~7%)。考虑到纺锤形结构的布线工作量要 大于梯形,在原型系统设计时选取设计工作量相 对较小的梯形方案。

# 1.4 PCI-E3.0/SAS3.0 全链路仿真

原型系统的 PCB 设计初步完成后,通过对



PCI-E与 SAS 全链路进行 3D 电磁场仿真(基于 ANSYS HFSS),得到全链路的无源仿真模型(S 参 数)。仿真过程中,对于较长的链路,需要将其分 割为多个部分,分段仿真,最后将各个部分的 S 参 数级连起来,得到全链路的 S 参数(仿真工具为 ANSYS Designer)。图 5 分别为PCI-E与 SAS 中最 长链路的插入损耗曲线。





其中, PCI-E 全链路在 4GHz 时的损耗为 -9.17dB, SAS 全链路在 6GHz 时的损耗为 -17.74dB,结合相应的频域参数模板可知, PCI-E的仿真结果满足相应规范的要求, SAS 的仿 真结果在低频时低于模板,在基频点(6GHz)以上 满足模板要求。

### 2 有源仿真

在 PCB 设计完成后,需要评估链路设计是否 满足相应规范要求。结合前文提取的全链路模 型,依据PCI-E3.0和 SAS3.0协议规范设计了通 用芯片的有源仿真模型,对系统PCI-E与 SAS 链 路的信号质量进行有源仿真。保证在仿真结果满 足规范要求的前提下进行投产,降低系统投产风 险。有源仿真基于 ANSYS Designer 仿真工具,为 了能够适应不同厂商的芯片,在仿真时优化设计 了收发器有源模型,模型的建立依据相应的PCI-E 与 SAS 的协议规范。

#### 2.1 PCI-E3.0 有源仿真

PCI-E3.0通用芯片模型中包含均衡模块<sup>[10]</sup>:①在TX端设置有一个3Tap(Tap表示移位寄存器的阶数)的FIR预加重模块;②在RX端设置有CTLE模块和1TapDFE模块,CTLE模块包含2个极点和1个零点,高频补偿为6~12dB; DFE采用LMS自适应算法<sup>[11]</sup>。在仿真过程中,为使仿真环境更接近真实测试环境,在链路的末端添加一段损耗约为-4dB(4GHz)的链路,用于模拟PCI-E标准卡设备,图6为使用PRBS31编码得到的RX端眼图仿真结果(8Gbps)。

图 6(a)中 TX/RX 均衡均未开启;图 6(b)中 TX 端均衡开启;图 6(c)中 TX/RX 均衡均开启; 图 6(d)为(a)~(c)条件下 BER 浴盆曲线对比。



Fig. 6 Simulated eyes of PCI-E channel (8Gbps)

仿真统计结果见表 3,为了能够更好地模拟真实 芯片的环境,在发送端加入 0.04UI 的占空比失真 (Duty Cycle Distortion, DCD),加入了 2ps 的随机 抖动(Random Jitter, RJ)。协议标准定义的误码 率眼高和眼宽分别为 35mV 和 37.5ps,添加抖动 噪声的情况下仿真结果能够满足规范要求。

表 <b>う</b>	PUI-L谜	路眼图仍身	1余件与3	<b>治米(</b> 80	зbps)	
Eye simu	lation con	ditions and	results of	PCI-E	channel (	(8Gbps)

TX Pre-emphasis		emphasis	CTLE(JD)	1 Tap	DCD(III)	DI(DMC)	Eye Height	Eye Width
Item	Pre(dB)	Post(dB)	- CILE(dB)	DFE	DCD(UI)	KJ(KMS)	(mV)	(%UI)
a	-	-	-	-	0.04	$2 \mathrm{ps}$	-	-
b	3.5	3.5	-	-	0.04	$2 \mathrm{ps}$	65	$53(66 \mathrm{ps})$
с	3.5	3.5	6	Adaptive	0.04	$2 \mathrm{ps}$	86	$69(86 \mathrm{ps})$

表 4 SAS 链路眼图仿真条件与结果(12Gbps)

Tab. 4 Eye simulation conditions and result of SAS channel (12Gbps)

Item	TX Pre-emphasis	CTLE(dB)	5 Tap DFE	DCD(UI)	RJ(RMS)	Eye Height (mV)	Eye Width (%UI)
а	-	_	-	0.04	$2 \mathrm{ps}$	_	-
b	Optimized(12dB)	-	-	0.04	$2 \mathrm{ps}$	36	55(45 ps)
с	Optimized(12dB)	4	Adaptive	0.04	$2 \mathrm{ps}$	36	$47(39 \mathrm{ps})$

### 2.2 SAS3.0 有源仿真

与PCI-E3.0相比,SAS3.0的链路速率更高, 为得到更加可靠的仿真效果,同时与市场中主流 的SAS3.0芯片性能相吻合,在模型中设计了更 加复杂的均衡模块:①在TX端设置一个4Tap FIR预加重模块,支持1Tap Pre-Cursor和2Tap Post-Cursor;②在RX端设置CTLE和5Tap DFE 模块,CTLE模块包含2个极点和1个零点,高频 补偿为0~11dB;DFE采用LMS自适应算法。

Tab. 3

为让 SAS3.0 的预加重参数更好地匹配链路 特性,依据 SAS 全链路的冲激响应曲线计算预加 重模块的均衡器系数。预加重模块补偿前后的全 链路冲激响应曲线的对比如图 7 所示(为使对比 结果更加明显,图中对预加重模块引入的 1UI 时 延进行了补偿)。





将最高点定为采样原点,原冲击响应曲线 (虚线)在原点、Pre-Cursor、Main-Cursor、Post-Cursor 对应的采样点电压依次为*x*<sub>-1</sub>,*x*<sub>0</sub>,*x*<sub>1</sub>,*x*<sub>2</sub>,构 建采样输入矩阵 X 与期望输出向量 y:

$$\boldsymbol{X} = \begin{bmatrix} x_0 & x_{-1} & 0 & 0 \\ x_1 & x_0 & x_{-1} & 0 \\ x_2 & x_1 & x_0 & x_{-1} \\ 0 & x_2 & x_1 & x_0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (1)$$

得到原始的预加重均衡器系数 c

$$\boldsymbol{c} = \boldsymbol{X}^{-1} \cdot \boldsymbol{y} = \begin{bmatrix} c_{-1} & c_0 & c_1 & c_2 \end{bmatrix}$$
(2)

得到归一化的预加重均衡器系数

$$\boldsymbol{c}' = \frac{\boldsymbol{c}}{\sum |c_i|} \tag{3}$$

c'的值为(-0.14,0.57,-0.27,-0.02)。对 比补偿前后的冲激响应曲线,预加重使得响应曲 线更加接近理想的冲激响应。图 8 为使用 PRBS31 编码得到的眼图仿真结果(12Gbps)。

图 8(a)中 TX/RX 均衡均未开启;图 8(b)中 TX 端均衡开启;图 8(c)中 TX/RX 均衡均开启; 图 8(d)为(a)~(c)条件下 BER 浴盆曲线对比。 眼图仿真结果统计见表 4,SAS3.0 协议中并未给 出 12Gbps 接收端的眼图模板,业界一般要求接收 端的眼宽大于 20.8ps。仿真结果表明链路在添 加抖动的情况下能够满足要求。







# 3 测试验证

### 3.1 PCI-E/SAS 全链路无源性能

在 PCB 生产完成后,为了评估系统链路真实的损耗特性,同时对比验证仿真结果,需要对 PCI-E/SAS 全链路进行无源测试。图 9 为PCI-E/SAS 链路的损耗仿真与实测结果的对比(链路损耗测试基于矢量网络分析仪),可以看出,在 20GHz 范围内,仿真结果与测试结果能够达到较好的匹配。

### 3.2 PCI-E3.0 测试结果

使用误码仪作为标准 TX 输出端,接入到系统PCI-E链路中,在PCI-E SLOT 接入标准PCI-E 3.0测试夹具,使用高速示波器得到PCI-E3.0的眼图测试结果(8Gbps),如图 10 所示。



图 10(a)~(c)的测试分别对应图 6(a)~ (c),测试中设置的预加重、均衡条件与仿真条件 一致,图 10(d)为链路接收端误码率仿真与实测 对比(无 RX 端均衡),3 种眼图的测试结果细节 见表 5。由于测试时的接收端均衡为高速示波器 模拟,并没有实际的物理设备,因此添加物理预加 重后的眼图结果作为评价链路的指标。实测结果 满足PCI-E3.0 的 35mV 眼高和 37.5ps 眼宽要求, 且实测结果与仿真结果(括号中的数值)接近,实 测结果略小于仿真结果。

## 3.3 SAS3.0 测试结果

使用误码仪作为标准 TX 输出端,接入到系统 SAS 链路中,在 SAS 硬盘接口处接入标准 SAS3.0 测试夹具,使用高速示波器得到 SAS 的眼 图测试结果(12Gbps),如图 11 所示。





图 11(a)~(c)的测试分别对应图 8(a)~ (c),测试中所设的预加重、均衡条件与仿真条件接 近。由于误码仪的预加重模块只能实现 3 Tap,实 测环境中的预加重效果有所降低,图 11(d)为链路 接收端误码率仿真与实测对比(无 RX 端均衡),3 种眼图的测试结果细节如表 6 所示。添加物理的 预加重后 SAS 链路的接收端眼图张开,与业内的 评价指标相比,接收端的眼宽指标满足 20.8ps 的 要求。由于实际设备的预加重性能较低(实际设备 为 3 Tap,仿真模型为 4 Tap),信号在添加 TX 预加重 情况下的眼高眼宽小于仿真值。在添加预加重的 情况下进行误码率测试,系统的误码率能够达到 10<sup>-15</sup>,满足存储系统的性能要求。

表 5 PCI-E链路眼图测	则试结果细节
----------------	--------

Tab. 5	Details	of PCI-E	channel	eye	diagrams	(8Gbps)
--------	---------	----------	---------	-----	----------	---------

Equalization	Eye Height(mV)	Eye Width(ps)	$TJ(\mathrm{ps})$	RJ(ps)	Width@ BER $\times 10^{-12}$ (ps)
NO EQ	0	0	162.55	1.685 6	0
TX EQ	62.687 (65)	63.946 (66)	75.425	1.707 0	49.569
TX/RX EQ	101.13 (86)	97.832 (86)	40.410	1.210 4	90.015

#### 表 6 SAS 链路眼图测试结果细节

Tab. 6	Details	of SAS	5 channel	eye	diagrams	(12Gbps)
--------	---------	--------	-----------	-----	----------	----------

_						
	Equalization	Eye Height(mV)	Eye Width(ps)	$TJ(\mathrm{ps})$	RJ(ps)	Width@ BER $\times 10^{-12} (\mathrm{ps})$
	NO EQ	0	0	396.03	23.242	0
	TX EQ	30.441 (36)	37.043 (45)	56.046	2.130 8	27.282
	TX/RX EQ	39.237 (36)	46.945 (39)	46.646	1.526 2	36.683

#### 4 结论

本文基于 SI 仿真设计方法,对高密度存储服 务器系统中PCI-E3.0 与 SAS3.0 的链路进行设计 优化。在材料选型时,以最长链路的损耗作为指 标,对材料损耗进行控制,既满足了相关协议规范 的要求,又避免了材料的浪费;同时,对链路中不 连续结构,特别是过孔结构进行进一步优化,使阻 抗的偏差控制在 5% 之内,保证链路损耗在接近 规范极限的情况下,信号仍能满足协议要求。在 PCB 设计完成后,通过PCI-E3.0 与 SAS3.0 的有 源仿真,对系统性能进行初步的评价,确保仿真结 果满足协议规范后进行投产,大大降低系统的设 计风险。最后,对系统链路的无源与有源特性进 行测试验证对比,确认系统能够满足设计需求,同 时也验证 SI 仿真结果有效可靠。

# 参考文献(References)

[1] Schmidt E E. People Aren't Ready for the Technology Revolution [ EB/OL ]. [ 2014 - 06 - 12 ]. http:// readwrite.com/

- [2] Gantz J, Reinsel D. The digital universe decade-are you ready[R]. IDC\_925, 176 South StreetHopkinton, MA01748, United States, EMC Corporation, 2010.
- [3] PCI express base specification revision 3.0[S/OL]. PCI-SIG, 2010. http://www.pcisig.com/
- [4] Serial attached SCSI 3 (SAS 3) [S]. American National Standards Institute, Inc. 11 W. 42nd Street, New York, 2013.
- [5] Hall S H, Heck H L. Advanced signal integrity for high-speed digital design[M]. New York: John Wiley & Sons, 2009.
- [6] IT 180BS/IT 180TC specification [ R/OL ]. ITEQ Corporation. [2014 - 06 - 12]. http://www.iteq.com.tw/
- [7] TU 862HF specification [R/OL]. Taiwan Union Technology Corporation. [2014 – 06 – 12]. http://www.tuc.com.tw/
- [8] Nelco N4000 13/N4000 13 SI specification [R/OL]. Park Electrochemical Corporation. [2014 - 06 - 12]. http://www. parkelectro. com/
- [9] Megtron 6[R/OL]. Panasonic Corporation. [2014 06 12]. http://www.panasonic.com.c
- [10] Liu J, Ling X F. Equalization in high-speed communication systems[J]. IEEE Circuits and Systems Magazine, 2004,4 (2):4-17.
- [11] Lucky R W. The adaptive equalizer [J]. IEEE Signal Processing Magazine, 2006,23(3):104 – 107.