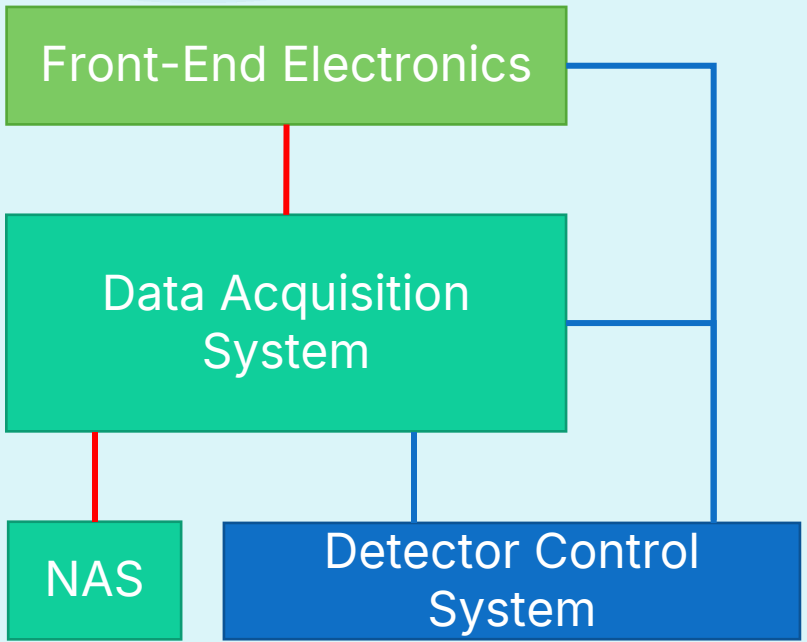
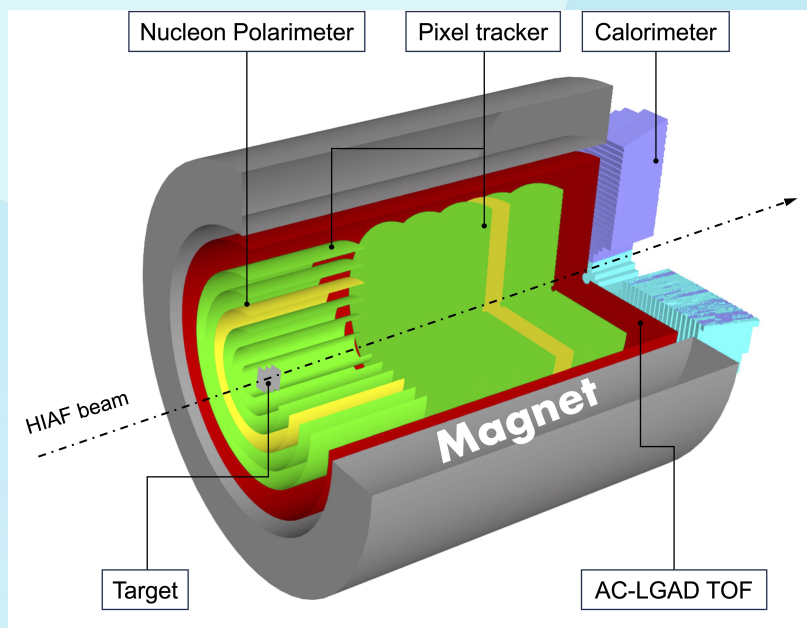


H-NS数据获取系统

陈凯（华中师大），杨俊峰（中科大），李敏（近物所）





01

读出需求

02

整体框架

03

研究现状

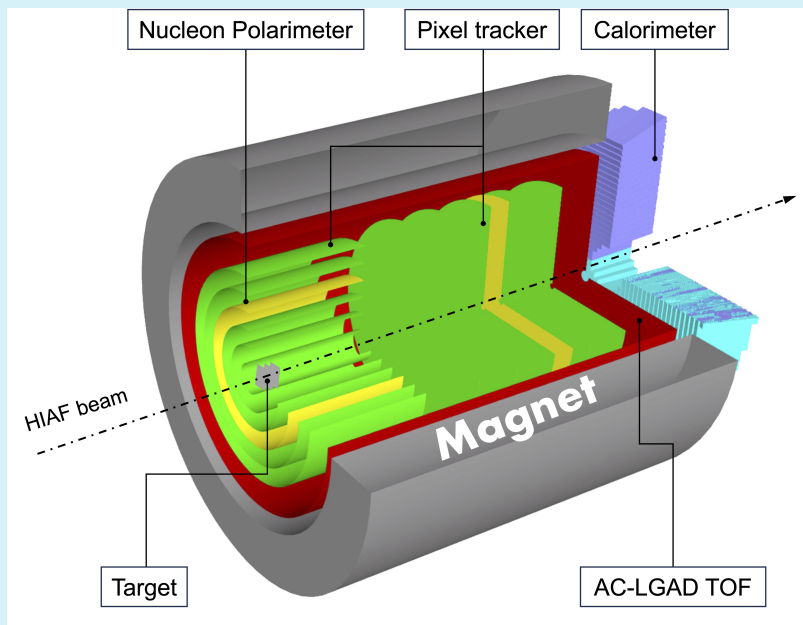
04

研究计划

01

读出需求

光纤数



MAPS

总面积：4.53 m² 芯片：11636
Stave：122 像素：5.67B
Ladder：570

ECAL内层

PbWO₄：~1280 (1225)
PMT读出板：80

LGAD

总面积：5.21 m²
芯片：13024

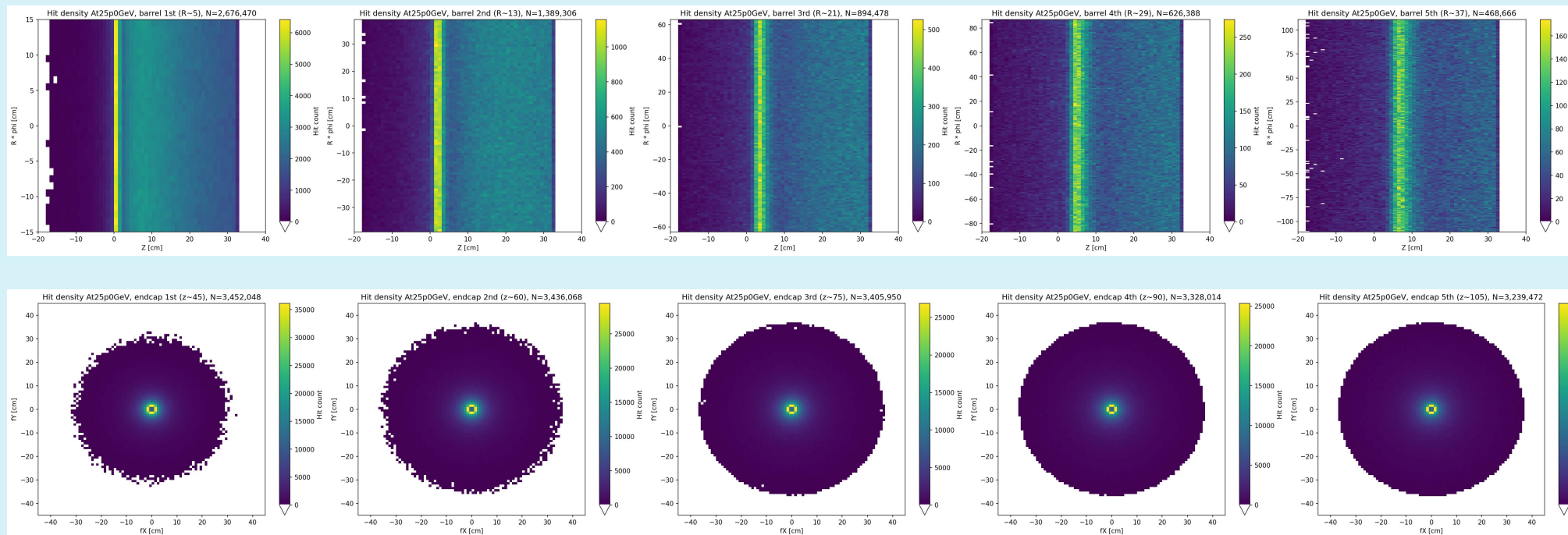
ECAL外层

双读出模块：232
每个模块层数：13
总通道：6032

MAPS-B	MAPS-E	LGAD-B	LGAD-E	
202		142	46	
ECAL-I	ECAL-O	TO-PMT	TO-LGAD	合计
80	95	13	49	627

数据率——MAPS

1MHz@25 GeV

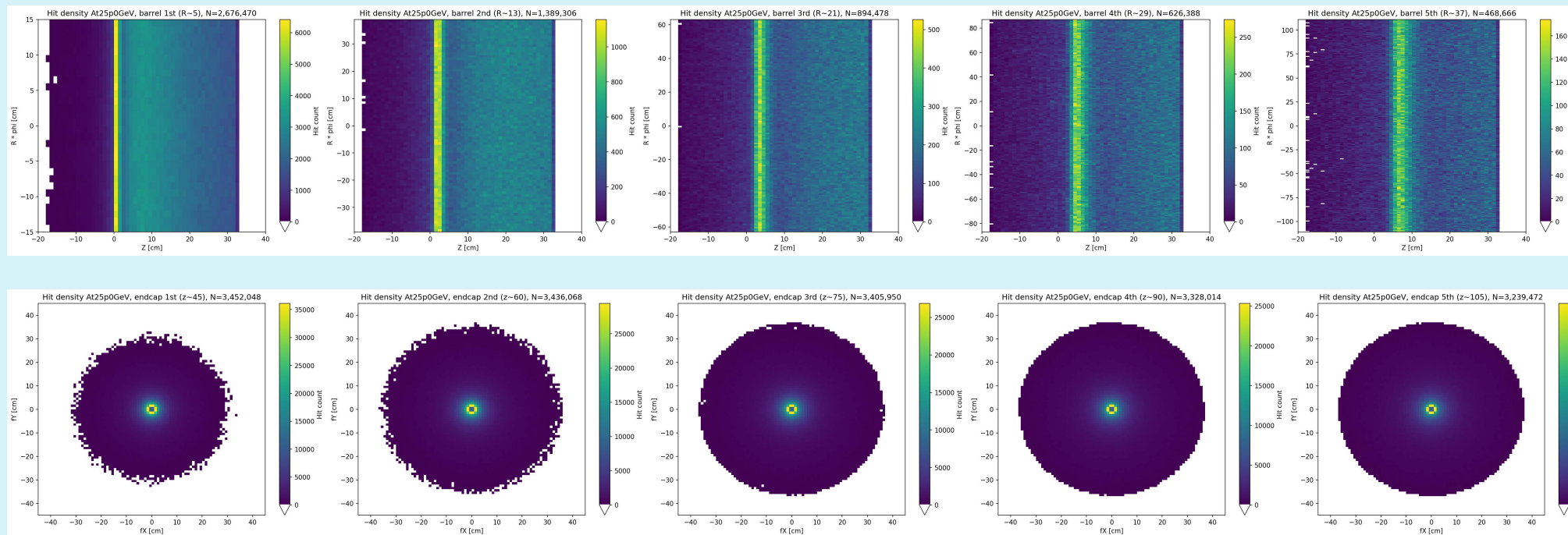


- 1MHz情况下总击中率22.94MHz
- 单击中4像素，占用2数据包（支持6像素）：272b
 - 去除空余2像素，可压缩至~208b
- 合计：272b x 22.94M = 6.384 Gb/s
 - 208b x 22.94M = 4.772 Gb/s

Barrel 1st	Barrel 2nd	Barrel 3rd	Barrel 4th	Barrel 5th
2.68	1.39	0.90	0.63	0.47
Endcap 1st	Endcap 2nd	Endcap 3rd	Endcap 4th	Endcap 5th
3.46	3.44	3.41	3.33	3.24

数据率——MAPS

1MHz@25 GeV

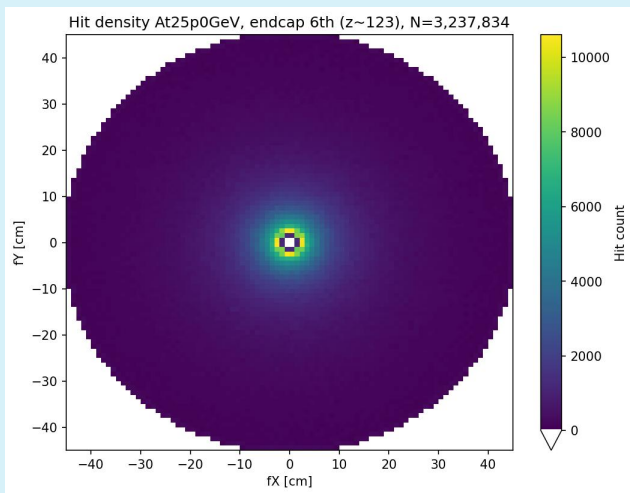
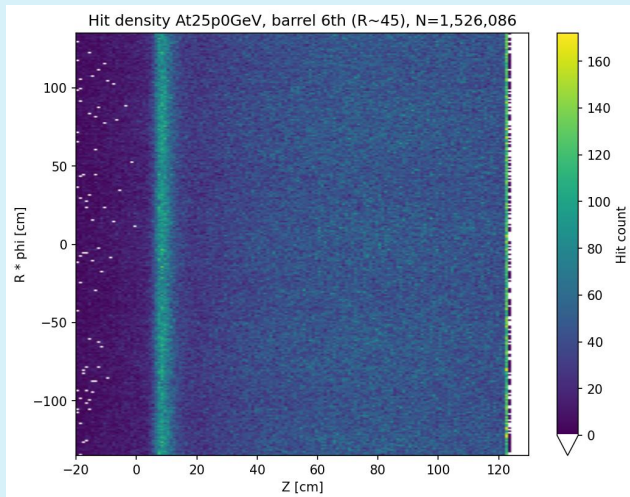


- 1 MHz情况下总击中率22.94 MHz
- 单击中4像素，占用2数据包（支持6像素）：272b
 - 去除空余2像素，可压缩至~208b
- 合计：272b x 22.94M = 6.384 Gb/s
 - 208b x 22.94M = 4.772 Gb/s

- 1GHz背景下，引起击中约30.34 MHz
- 数据率：272 x 30.34 = 8.26 Gb/s
- Fake hits: 5MHz, 272 x 5M = 1.36 Gb/s
- 合计：9.62 Gbps

数据率——LGAD

1MHz@25 GeV



- 1 MHz对应hit: 4.77 MHz (1.53+3.24)
- 128b/hit
 - 对应数据率: 0.61 Gb/s
- 1 GHz对应背景hit: 12.77 MHz
 - 对应数据率: 1.64 Gb/s

数据率——ECAL

ECAL Inner

- FEE: $8.2\text{B/hit} \times 32 \text{ channel} \times 2\% \text{ occupancy} = 5.248 \text{ B/event}$
 - 考虑overhead: 8B/event
- 合计: $8\text{B} \times 80 \text{ FEE} \times 1\text{MHz} = 5.12 \text{ Gbps}$
- 背景暂按3倍估算: 15.36 Gbps

ECAL Outer

- FEE: $5\text{B/hit} \times 64 \text{ ch} \times 4\% \text{ occupancy} = 12.8 \text{ B/event}$
 - 考虑overhead: 18B/event
- 合计: $18\text{B} \times 95 \text{ FEE} \times 1\text{MHz} = 13.68 \text{ Gbps}$
- 背景暂按3倍估算: 41.04 Gbps

读出要求

	Fiber	Event Data	Bkg Data	Data rate
MAPS	202	4.78	9.62	14.4
LGAD	188	0.61	1.64	2.25
ECAL-I	80	5.12	15.36	20.48
ECAL-O	95	13.68	41.04	54.72
T0-PMT	13	8	24	32
T0-LGAD	<49	8	24	32
	627	40.19 Gbps		155.85 Gbps
		24.19+16 kb/event		

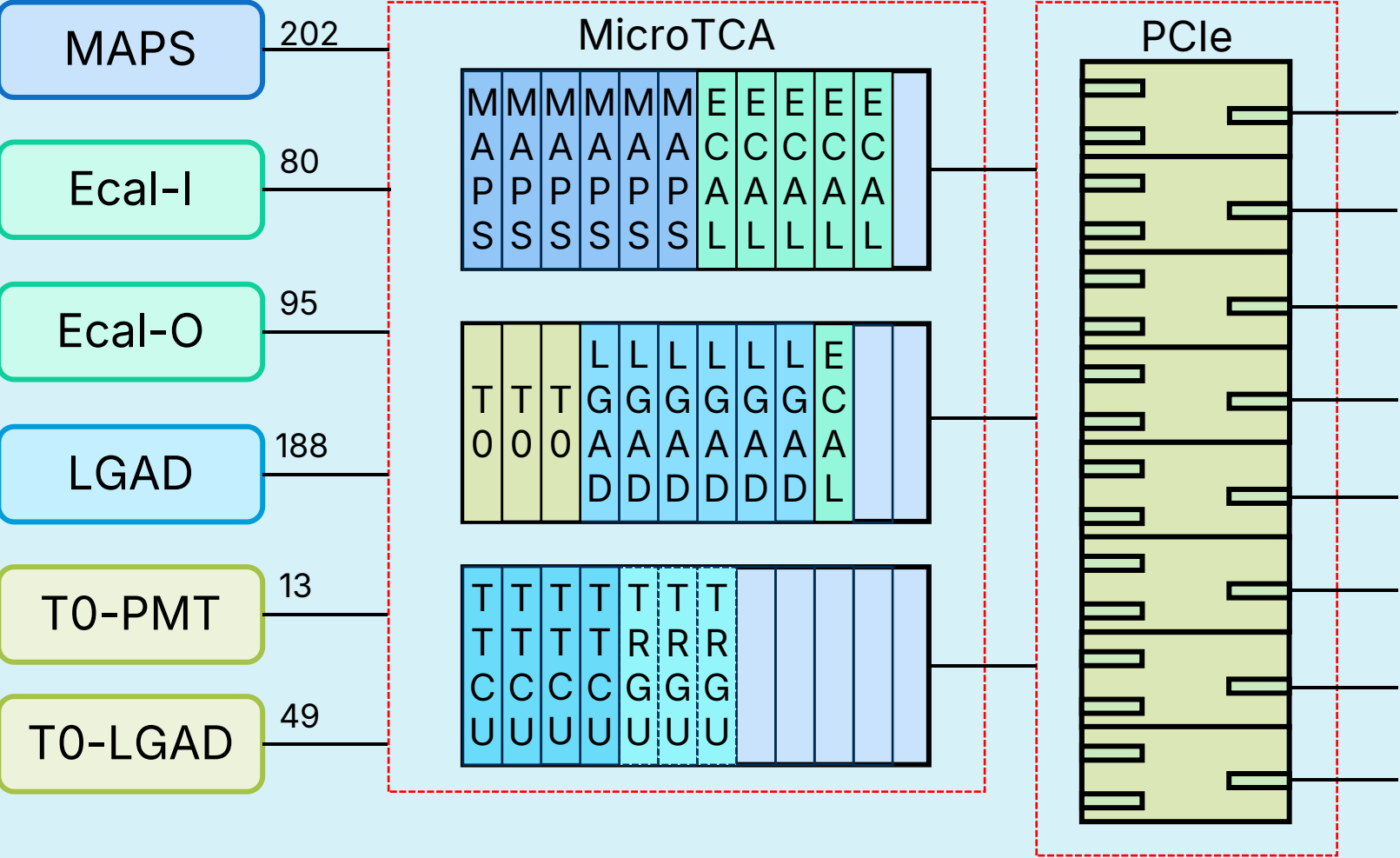
- 部分需求参数仍待确定
- Lambda相关事例约1%，即10 kHz，1PB存储看见按40.19 kb估算约230天
 - 数据在DAQ固件/软件应可继续压缩

02

整体框架



数据流



	μTCA	PCle	服务器	网卡
MAPS	6	2	1	1
LGAD	6	2	1	1
ECAL	6	3	3	3
T0	3	2	2	2
TRG	3	1	1	1
TTC	4			
合计	28	10	8	8

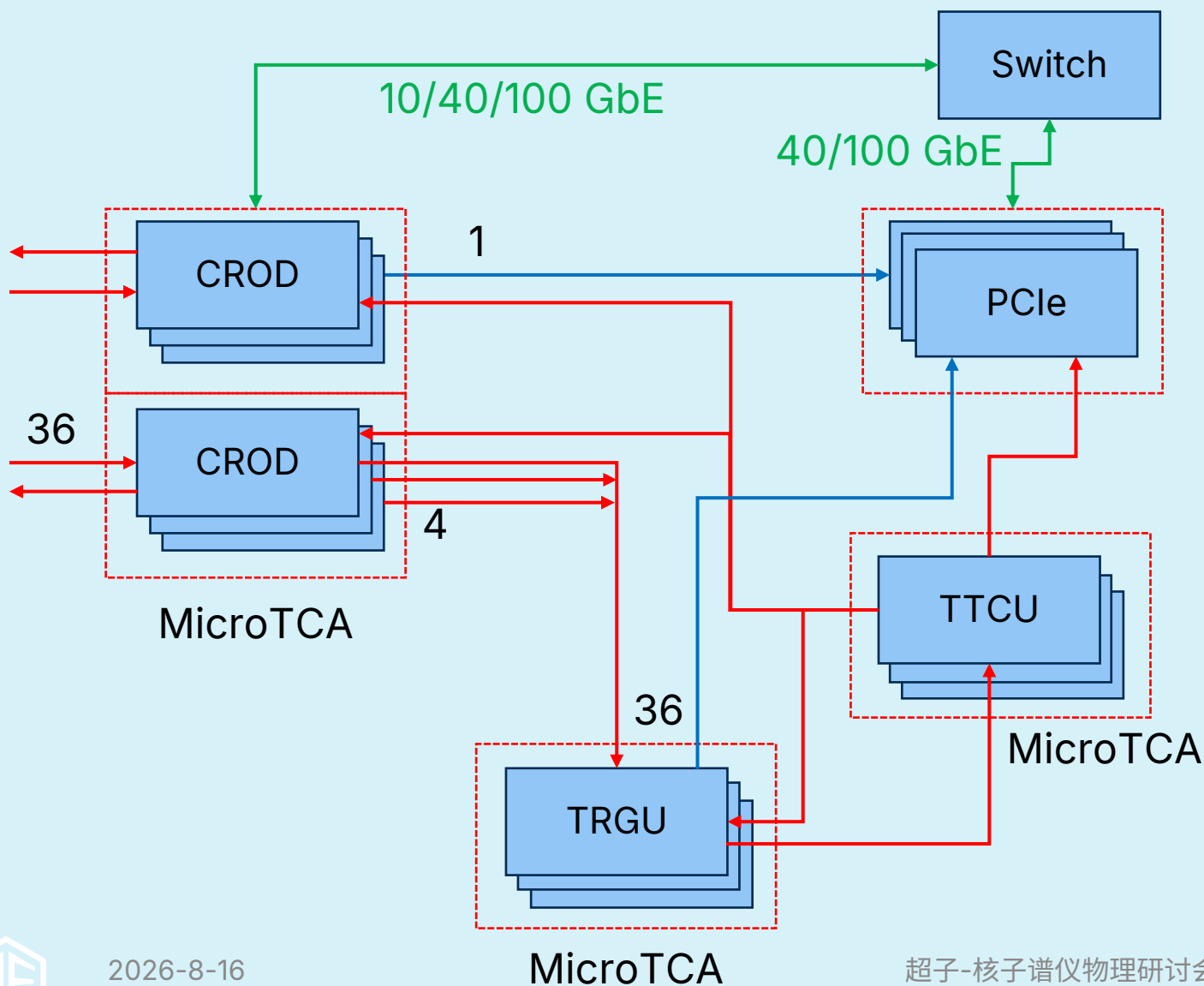
Data Handler

兼容触发、无触发模式

MicroTCA接收光纤>36

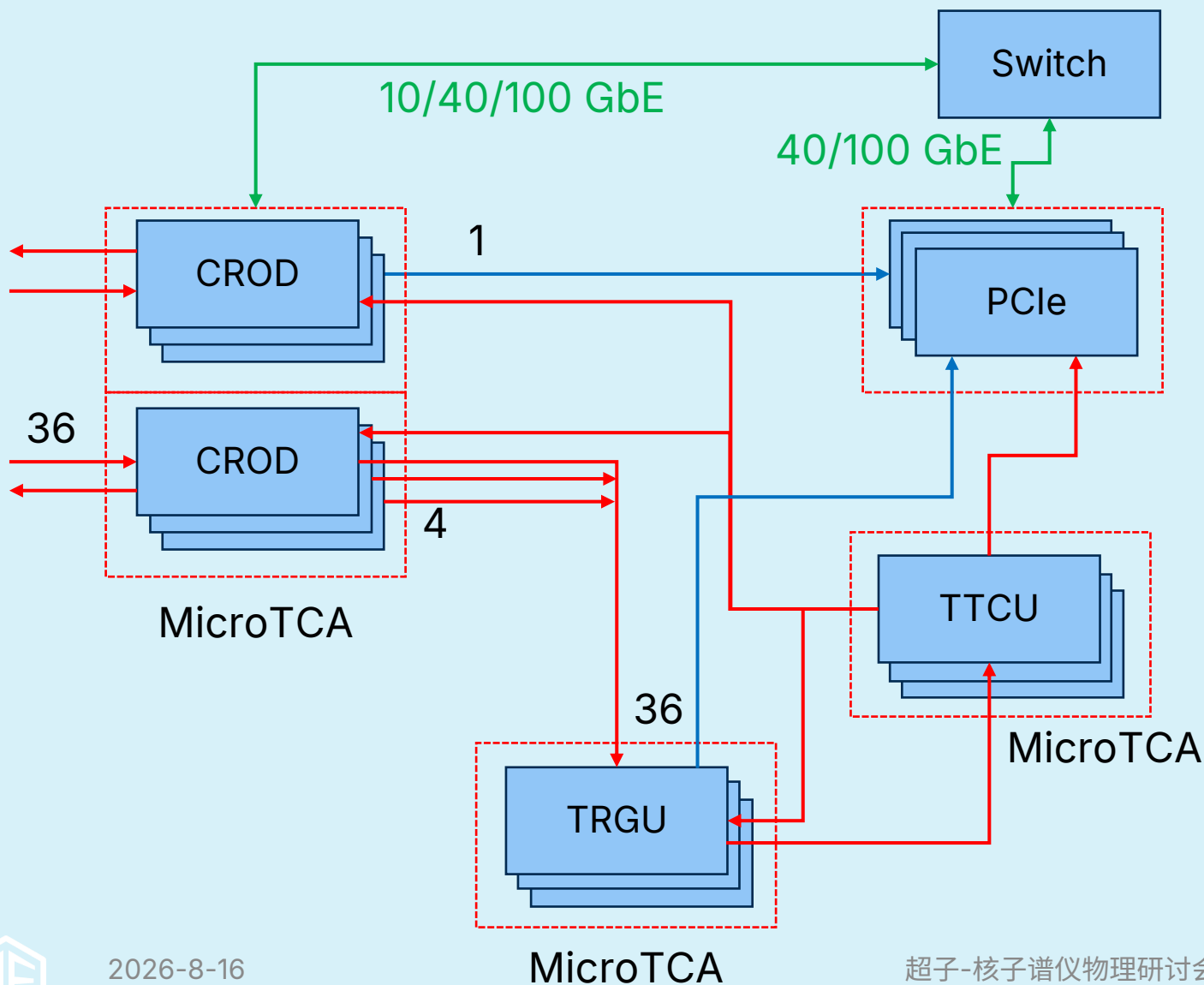


整体结构——TRGU



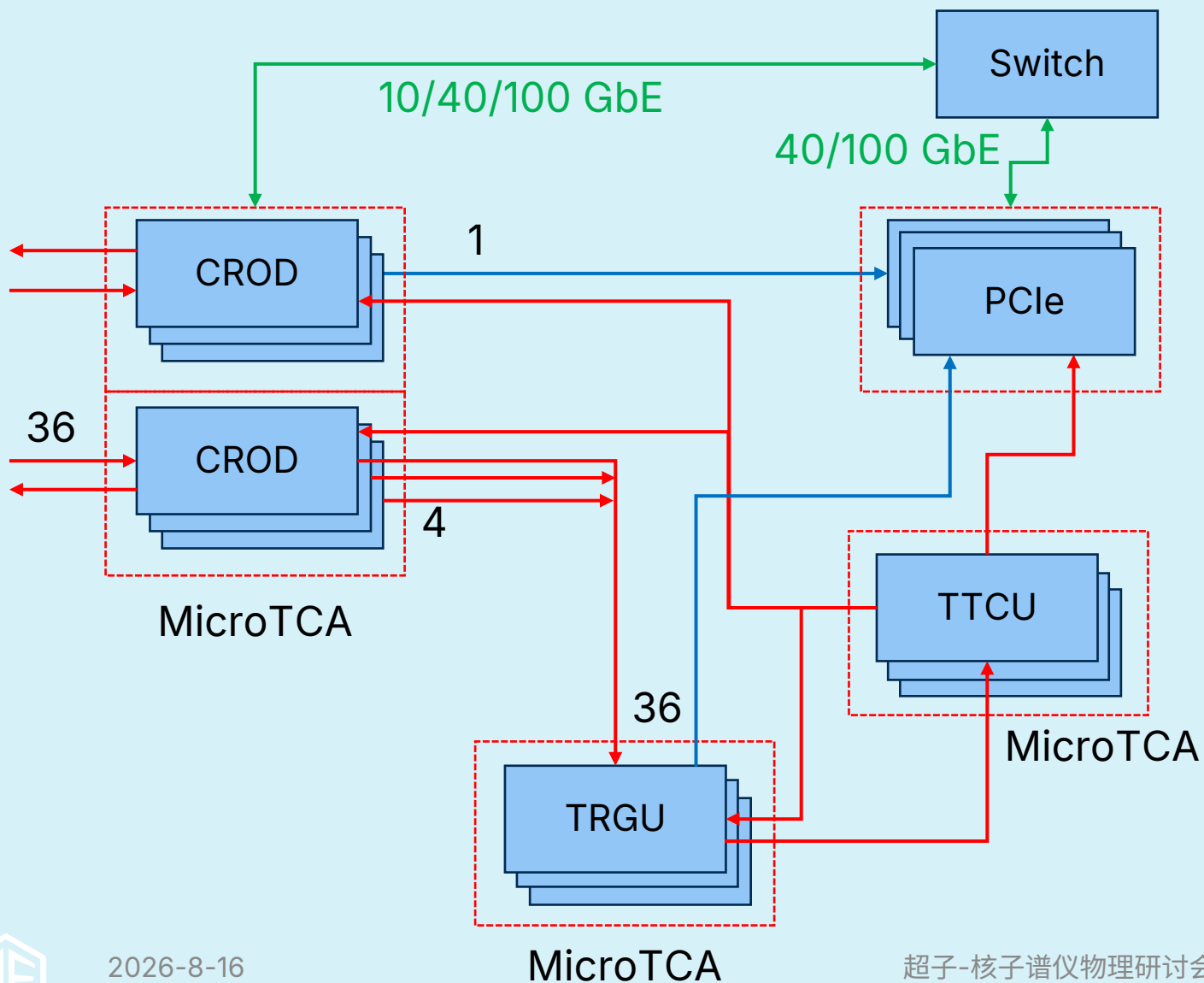
- 触发系统
 - 当背景事例引起数据过高时，将集成一个相对简单的触发系统
 - 硬件与DAQ、时钟板等共用（模块化设计）
 - 触发数据源及算法待讨论
- 数据来源
 - 在CROD实现，即直接对接FEE
 - 或从部分CROD接收必要的信号
 - 或接收外部其他信号
- 输出
 - 触发信号将送往TTC进行分发
 - 潜在可能性部分数据直接进入PCle
- 延时
 - 固定且尽可能小

整体结构——TTCU



- TTC模块——时钟分发
 - 分发时钟到DAQ与触发各模块
 - CROD进一步分发时钟到前端电子学
 - 全链路相位不确定性（复位/重新上电等情况下）小于10ps
 - 环路具备一定的延时校准能力
- TTC模块——控制
 - 从TRGU接收触发信号并统一产生系统级触发命令
 - 其余实时控制命令
 - 接收BUSY等状态信号

整体结构——CROD与PCIe



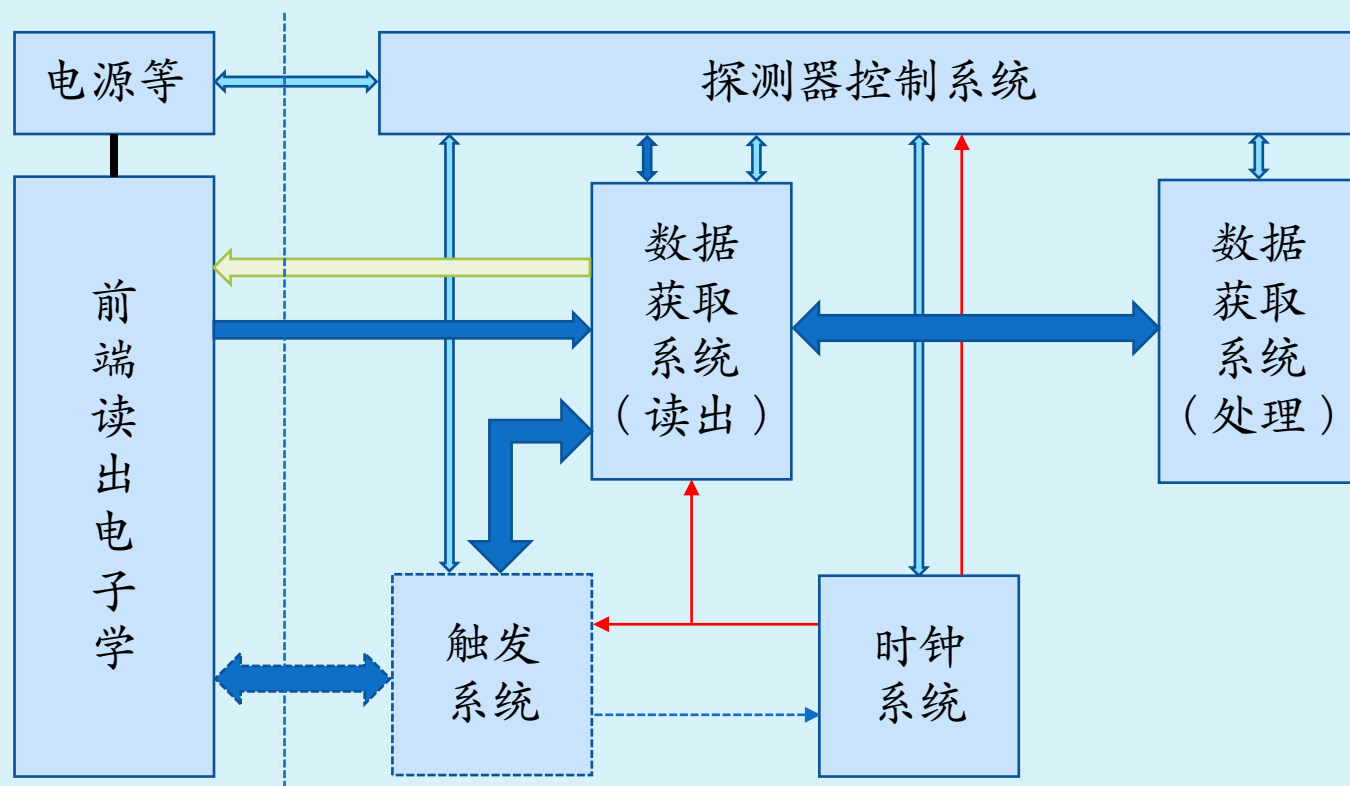
• CROD

- 单模块对接36个前端模块
- 时钟与控制命令分发
- 高速数据流持续接收
- 进行一定的简单数据处理
- 具备经10~100GbE直接对接交换机的可能性

• PCIe

- 进一步汇总数据
- 经NIC进入交换机

与探测器控制系统（DCS）的集成

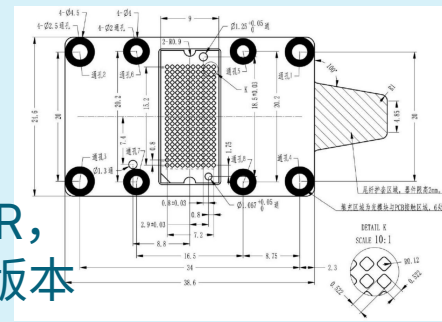
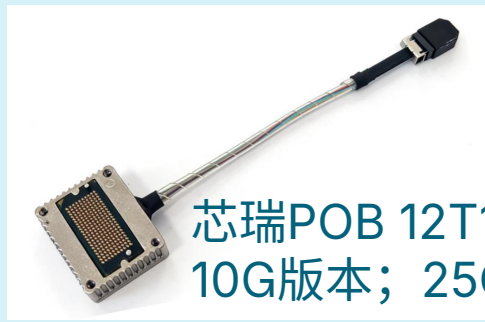
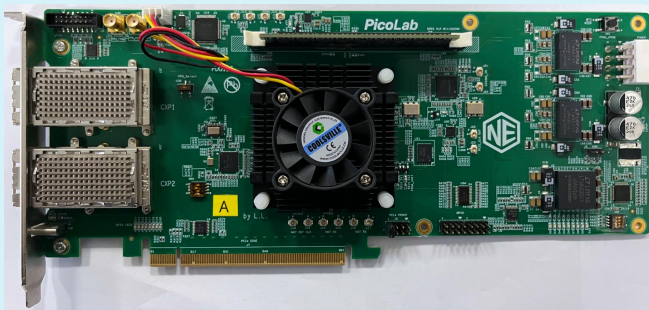
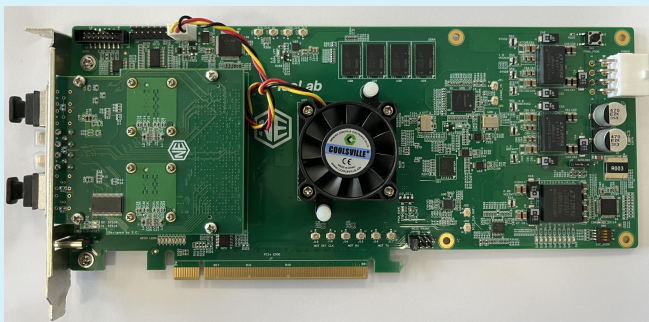


- 前端电子学与DCS间的双向数据流的转发（经CROD进入DCS网络）
- CROD、PCIe、TTCU、TRGU涉及的服务器、机箱的监控
- DAQ（数据处理部分）的监控

03

研究现状

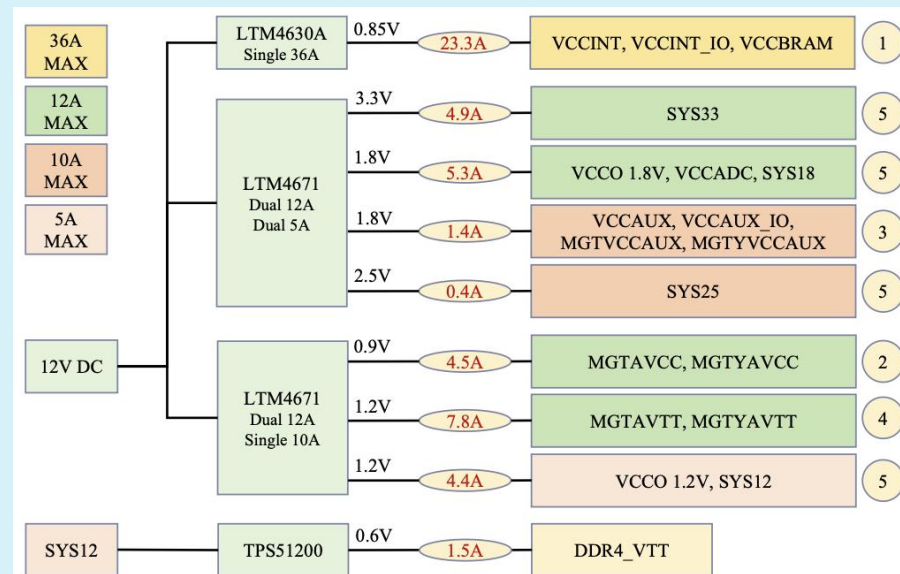
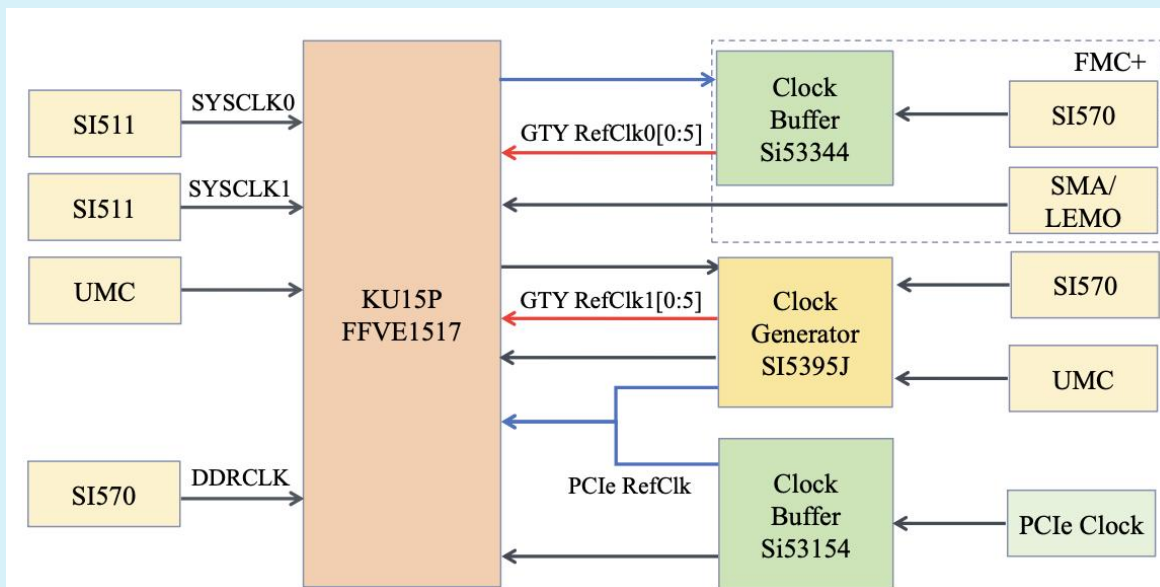
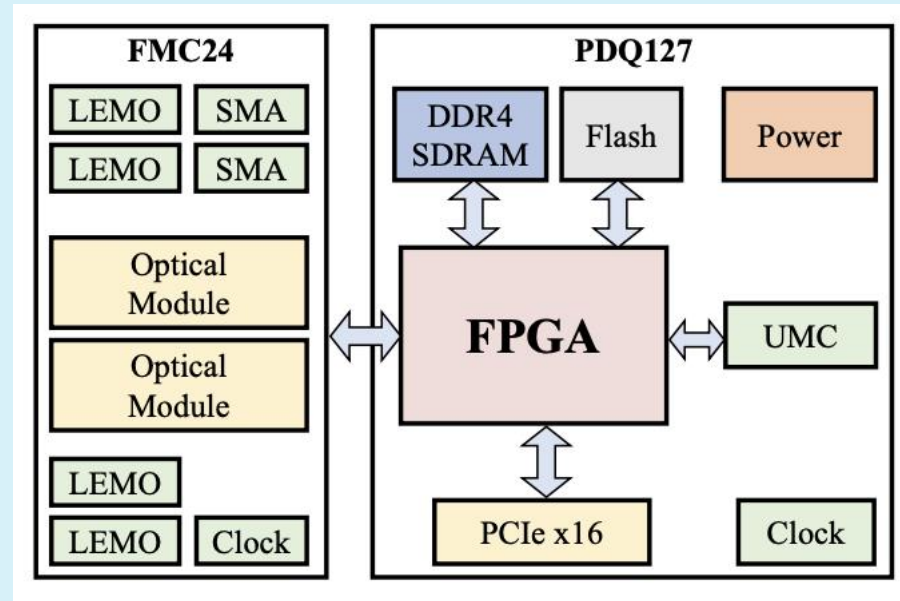
PCIe硬件



- PDQ127: 量产x20 (含3x PDQ125)
 - STCF & H-NS0
- PDQ126: 量产x4
 - NvDEx

	FPGA	Fiber links	PCIe	Throughput	Transceiver
PDQ126	KU15P	24	Gen3/4	1/2x14.76 GB/s	CXP
PDQ127	KU15P	24	Gen3/4	1/2 x14.76 GB/s	POB

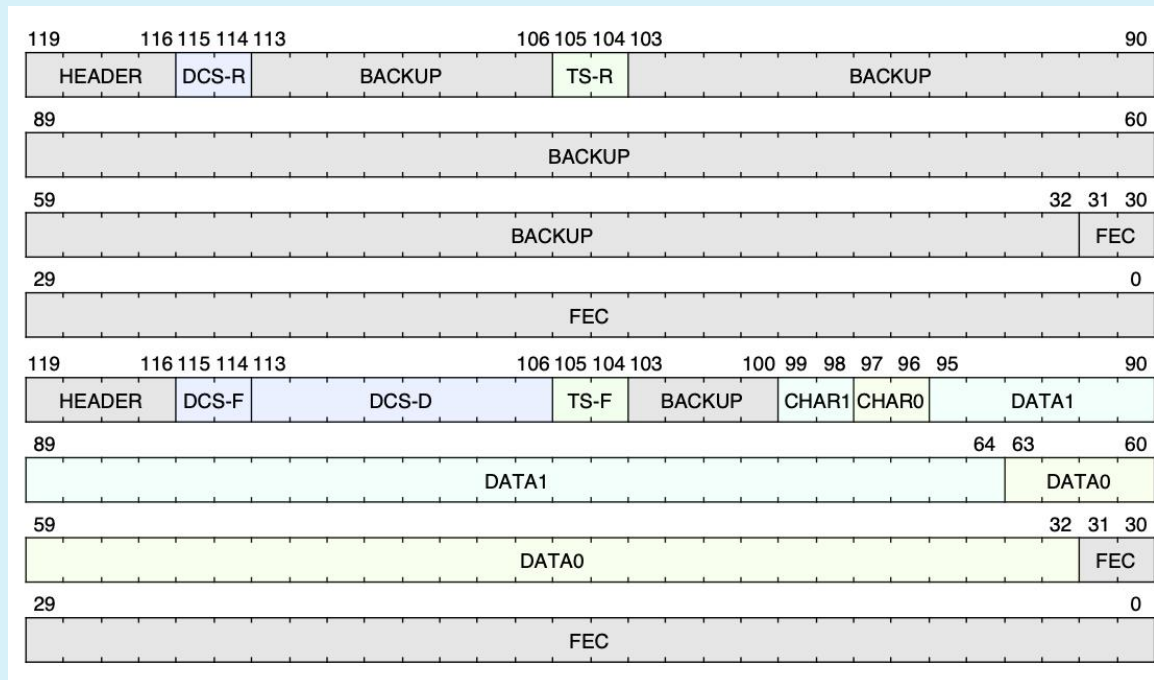
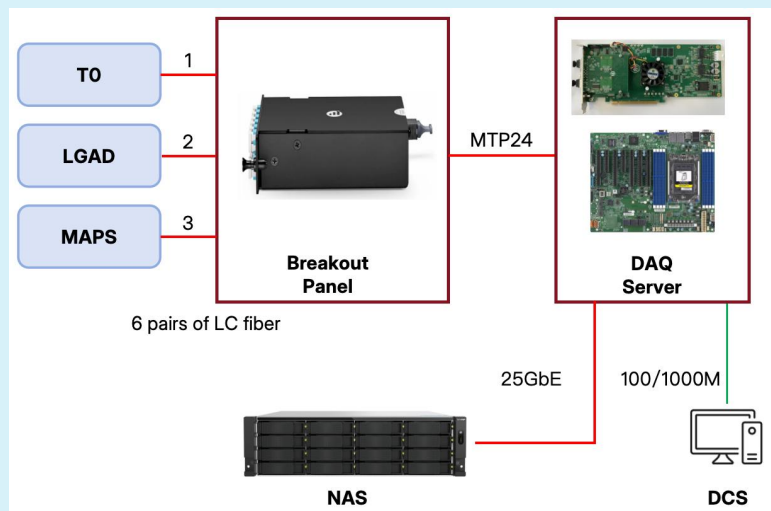
PCIe硬件——时钟/电源管理/光口



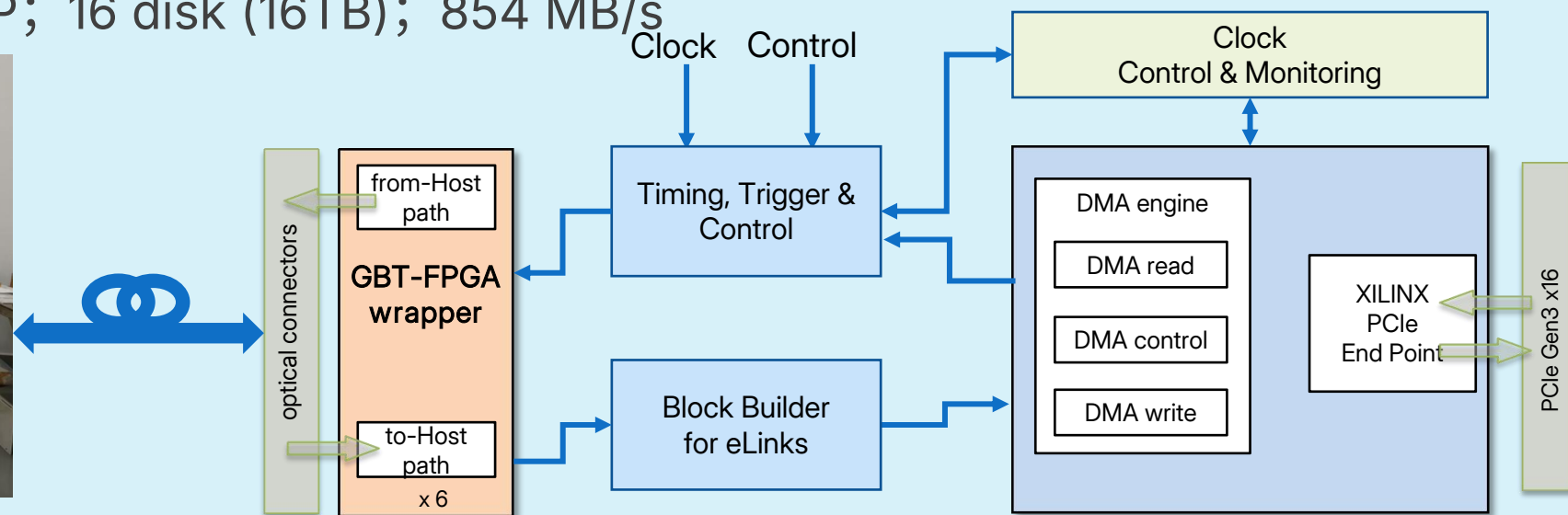
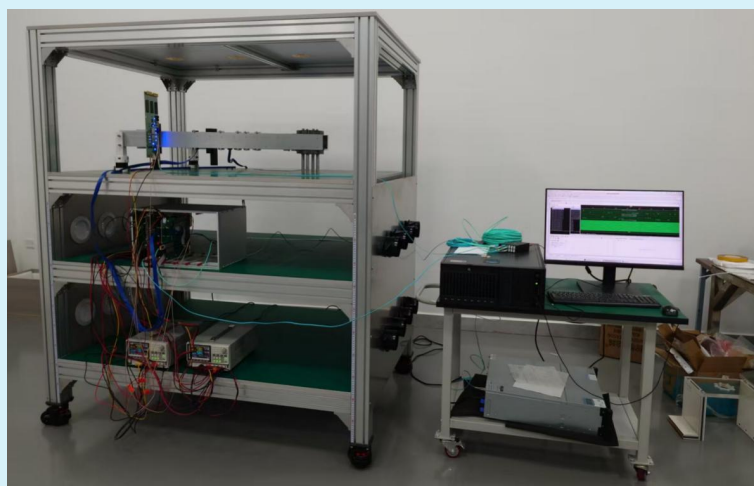
PCIe固件与软件

- PCIe DMA引擎：
 - Wupper: 多Descriptor; 双向高带宽DMA传输
- 光纤链路协议：
 - Supported: GBT; IpGBT; 10Gb Aurora; 8B10B (1.5G/2.5G/10G) ; 100Gb RDMA
 - To be supported: 10Gb Interlaken; 10Gb UDP
- 设计规范
 - 模块化
 - 标准AXI接口
 - BSP支持
- 固件/软件管理脚本化、Git管理
- 对接不同DAQ Framework

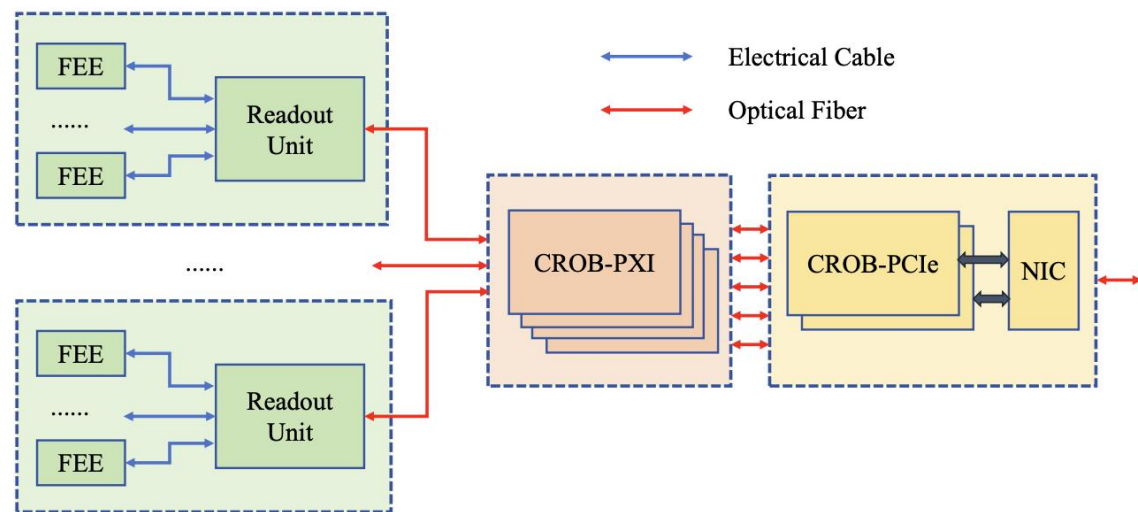
在H-NSO的应用



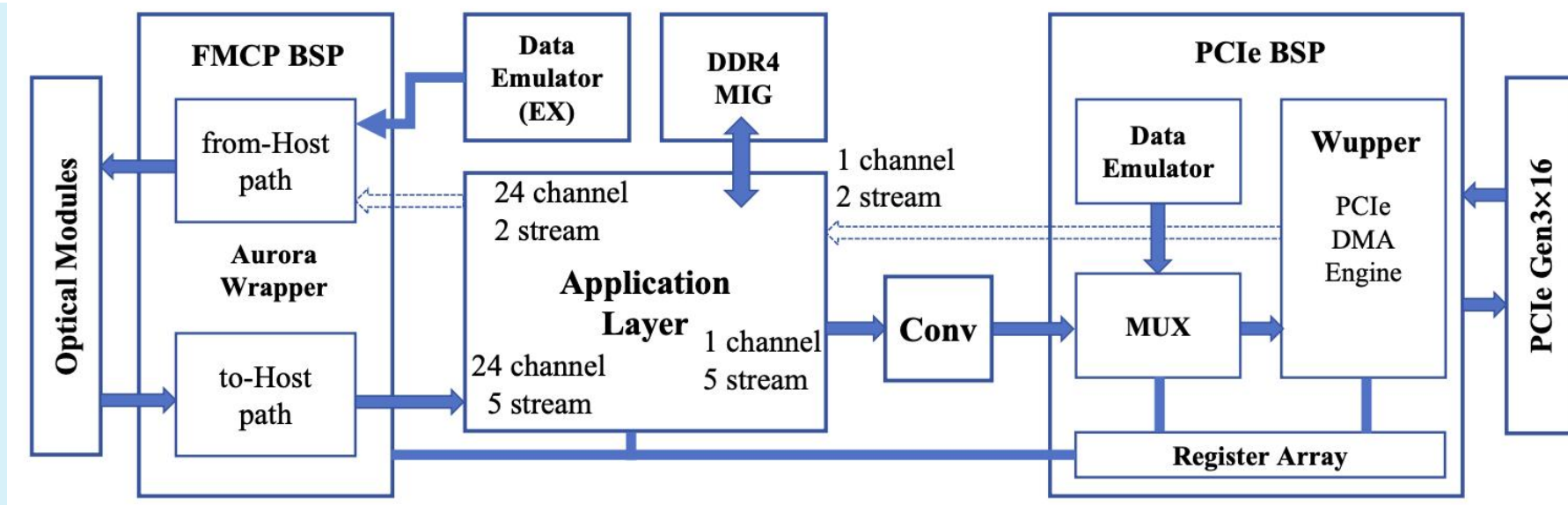
NAS: QNAP TS-h1677AXU-RP; 16 disk (16TB); 854 MB/s



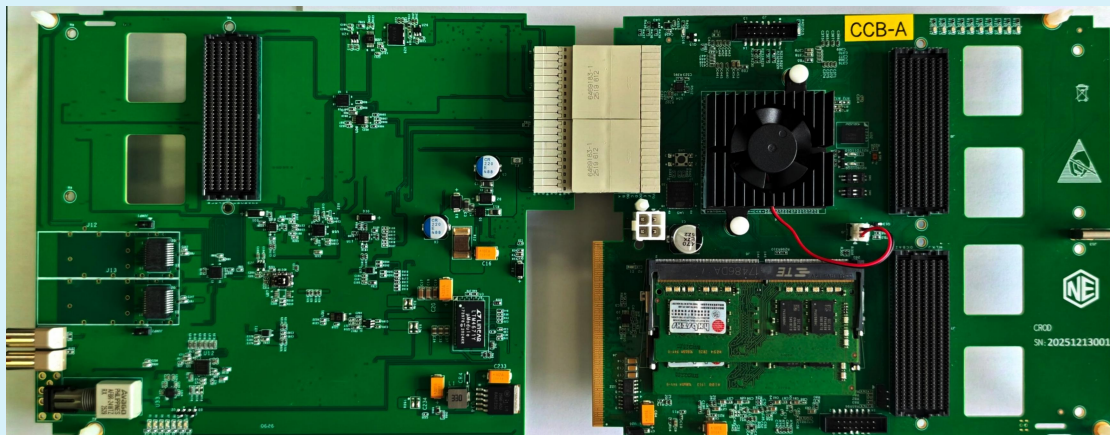
在STCF的应用



- ~180光纤: PXI至PCle
- ~10 PCle模块
- ~30.4 GB/s



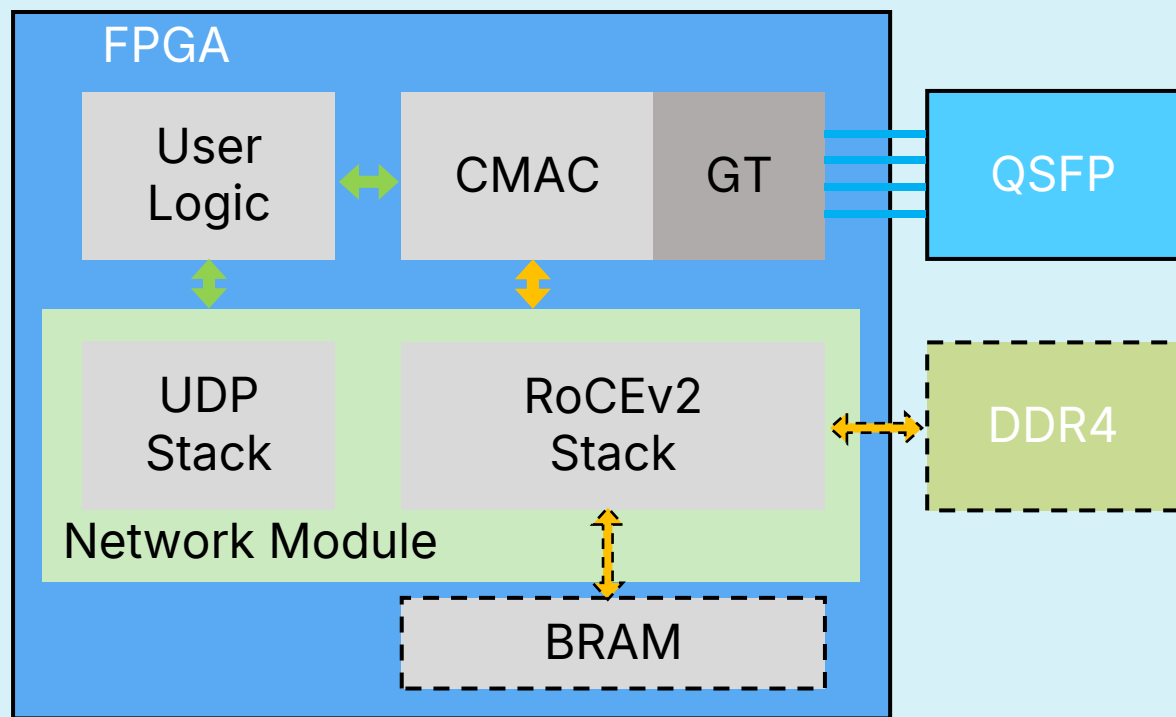
MicroTCA硬件



- 模块化设计：前插板GCM；后插板GRM；
- FMCP子卡（QSFP/SFP/POB），不同FMCP子卡用于不同目的
- V0：GCM性能符合预期；GRM测试进行之中
- 用于CROD（读出）；TRGU（触发）；TTCU（时钟及命令分发）
- 11~12卡/机箱

RDMA/UDP传输

- MicroTCA直连交换机，使用UDP（netFELIX亦计划采用）或RDMA。状态：
 - 基于相同FPGA的100Gb RDMA已应用于其他项目
 - 10Gb UDP比较成熟，可选方案多



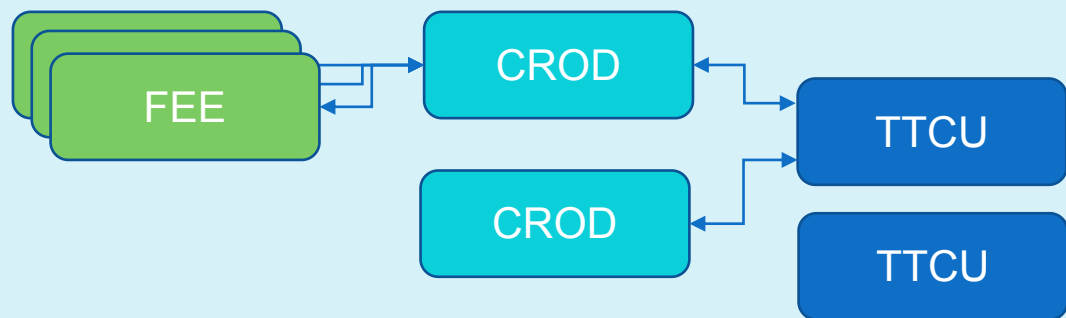
- ❑ 基于DavOS进行优化
 - 支持100Gbps的TCP、UDP、RoCEv2协议
 - 针对建链方式进行了优化
- ❑ 缓存模式
 - DDR4
 - 单内存条（4GB）可实现数十毫秒的缓存
 - 单个内存条难以支持双向同时传输
 - 逻辑复杂，读写延时较高
 - FPGA片内BRAM
 - 带宽高且配置灵活
 - AXIS协议传输，延时极低
 - 占用大量内部资源

RDMA传输性能初测

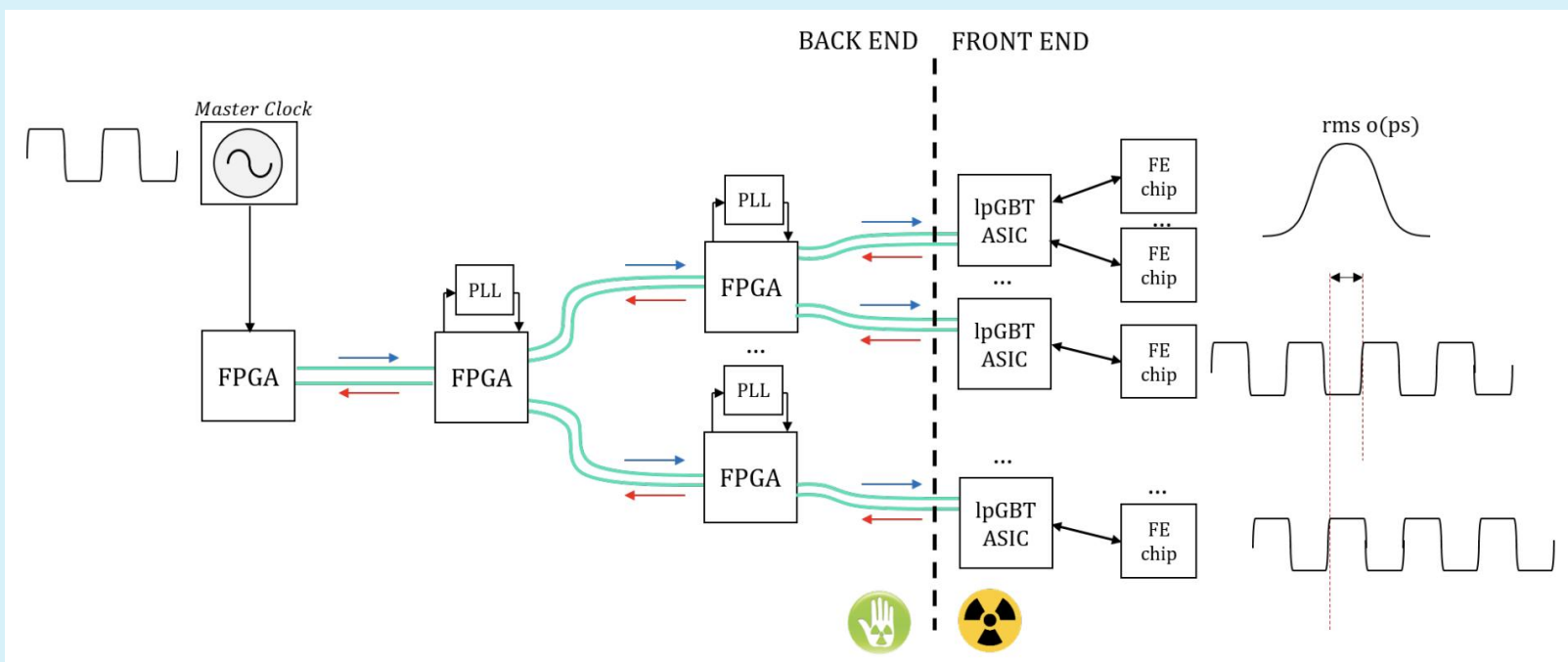
	DDR4缓存			BRAM缓存		
	FPGA to NIC	NIC to FPGA	双向	FPGA to NIC	NIC to FPGA	双向
Throughput	97.95 Gbps	97.86 Gbps	FPGA to NIC 68.39 Gbps NIC to FPGA 39.75 Gbps	97.92 Gbps	97.48 Gbps	FPGA to NIC 96.40 Gbps NIC to FPGA 95.24 Gbps

- 主机端网卡：MCX516A-CCATConnectX-5
- 单个DDR4 Module: Innodisk 16 GB DDR4 Mini-UDIMM VLP @ 2133MT/s
- 待测试：
 - 双DDR4 Module验证（实际NIC to FPGA方向需求不大）
 - 多模块经Switch到多计算节点的稳定性、拥塞测试

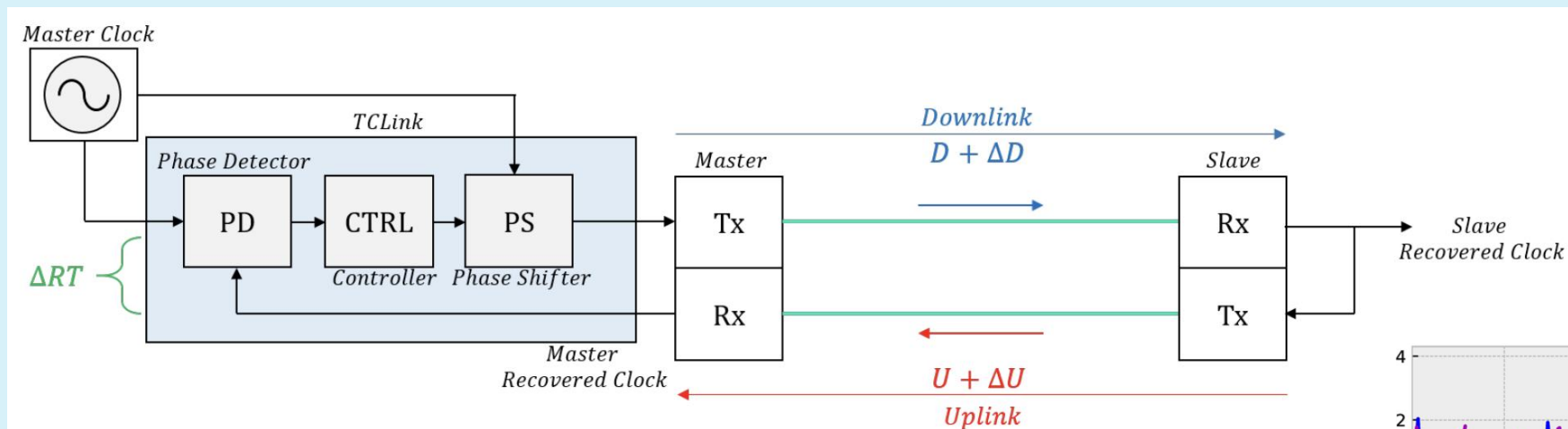
时钟分发——需求



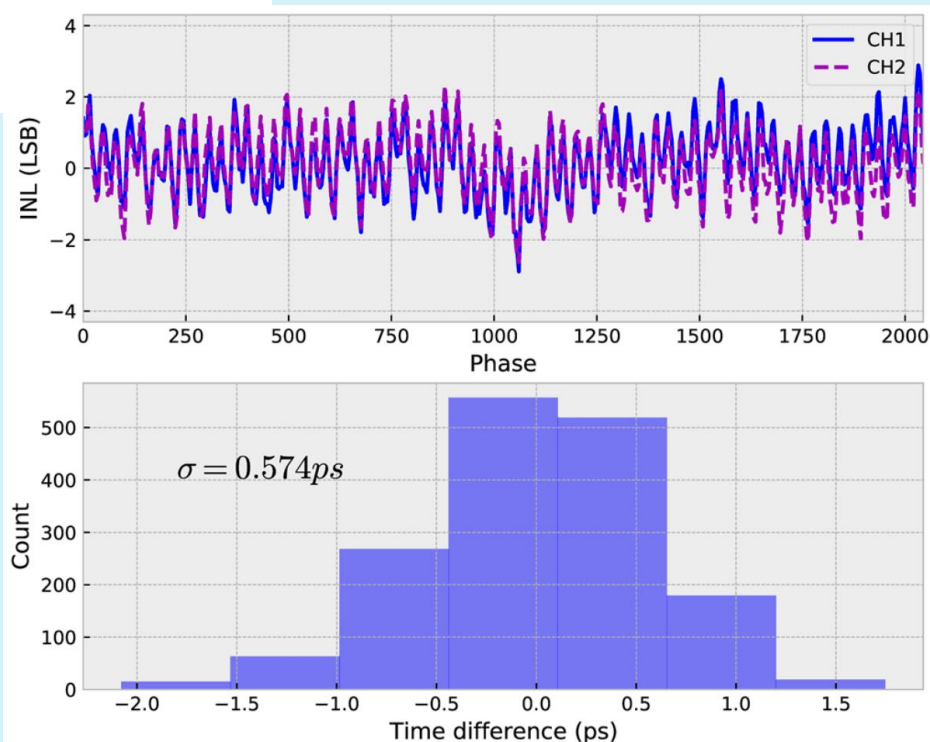
- TTCU间通过机箱背板、或板载电口传输时钟与命令
- TTCU与CROD间可通过背板、电口或光口传输时钟与命令
- CROD与FEE间通过光纤通信，嵌入时钟
- 传输协议本身需要固定延时
- FPGA高速收发器配置需优化：重上电；重插光线；复位，相位延时保持ps级稳定



时钟分发——相位稳定性、相位测量及调节



- FPGA MGT可实现RMS为皮秒级的相位可重复
- 全链路：TTCU-CROD-FEE，可重复性5~10ps
- 集成相位测量能力
 - 全数字DDMTD，ps/sub-ps精度
- 集成相位微调能力，补偿环境变化引起的漂移
 - ps级别步长



数据处理部分算力需求

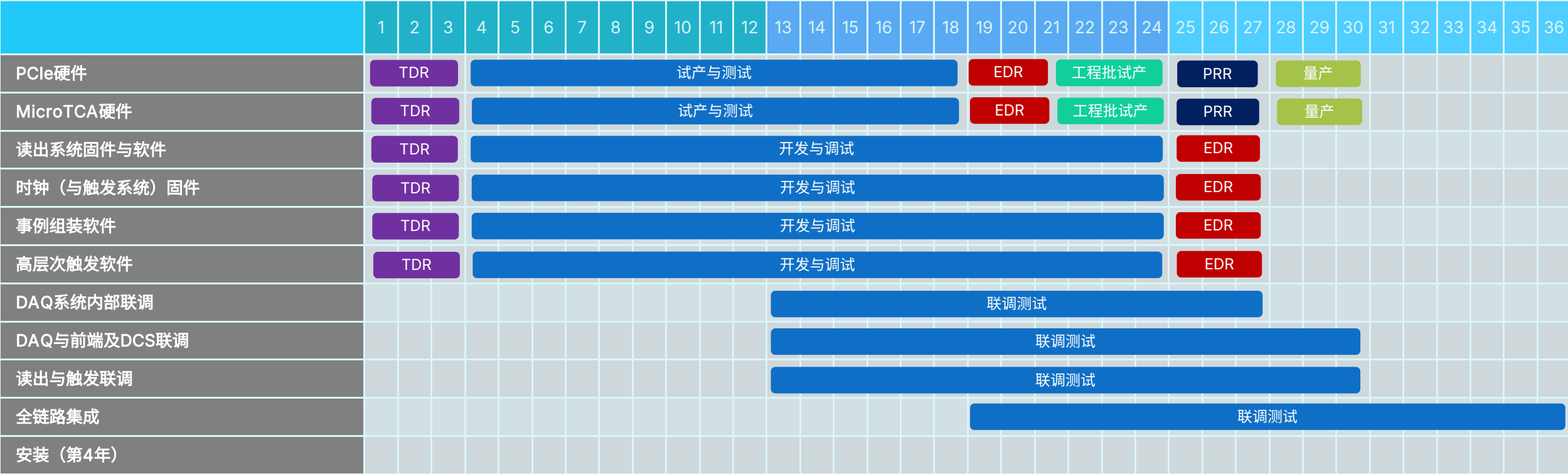
- STCF MDC二维径迹重建 (CPU+GPU) : 18.32ms/核/事例
 - H-NS: 1MHz L1事例, 约需18320 CPU核, 每个服务器192核, 约96个服务器
- LHCb Upgrade (Run3) : 32~40Tb/s (以32Tb/s计算)
 - Event Builder+HLT1 (含GPU节点) : 约170台服务器
 - HLT2 Farm: 1700~3700台服务器 (以1700计算)
 - H-NS: 按比例计算, 60台服务器约可处理~1Tb/s
- 综合以上三种计算方式, 采用GPU/FPGA加速方案, 按照60~80个服务器计算

04

研究计划



研究计划



- TDR

 技术设计报告/评审
- EDR

 工程设计评审
- PRR

 批产准备评审

第一年：
核心器件/设备的国产化替代
完成主要关键技术的初步验证



Thanks!

