arXiv:2611.29921v1 [physics.ins-det] 23 Nov 2026

准备提交至 JINST

XTCP：一种用于通过 10 和 25 千兆以太网进行读出的开源 FPGATCP 端点

薛涛[a]，[b]，[1]

[a]粒子与辐射成像重点实验室（清华大学），教育部，中国北京清华园街 *1* 号[b]清华大学工程物理系，中国北京清华园街 *1* 号

电子邮箱：xuetaothu@tsinghua.edu.cn

摘要：XTCP 是一种在现场可编程门阵列（FPGA）上实现的开源传输控制协议（TCP），用于高速探测器读出。其目标是实现接近 TCP 有效载荷上限的持续单连接吞吐量。在同一块 Kintex UltraScale+KU5P 卡和 Linux 主机上分别进行的 10 千兆以太网传输中，XTCP 和 SiTCP-XG 向主机传输时的应用层吞吐量中位数均为 9.49 吉比特/秒。向 FPGA 传输时，XTCP 的中位数为 9.49 吉比特/秒，而所测试的 SiTCP-XG 配置为 5.76 吉比特/秒。XTCP 原型使用了更多查找表和片上存储器。在一次 96 吉比字节的回显测试中，测试跨越了多次序列号回绕；记录确认首次越过恢复边界的时刻，使得四次快速恢复得以实现，而此前的有符号比较会阻止这些恢复。在一个 25 千兆以太网比特流上，将 IP 最大传输单元从 1500 字节增至 9000 字节后，向主机传输的单向吞吐量中位数从 23.30 吉比特/秒升至 24.70 吉比特/秒，向 FPGA 传输的则从 22.09 吉比特/秒升至 24.08 吉比特/秒。在 9000 字节设置下，这些速率分别达到 TCP 有效载荷上限的 99.7% 和 97.1%。在另一次回显测试中，主机通过一个连接，在 3595 秒内验证了巨型帧原型返回的全部 8.77 太字节数据，平均每个方向的吞吐量为 19.51 吉比特/秒。

关键词：用于探测器读出的前端电子学；数据采集概念；数字电子电路

[1]通讯作者。

目录

1 引言 **1**

2 端点架构 2

2.1 发送和接收路径 22.2 数据通路宽度、时钟和片上存储 3

3 测量方法 **4**

3.1 实验设置与传输测试 43.2 应用层吞吐量与理论极限 53.3 持续回显与完整性验证 6

4 结果 **6**

4.1 单连接吞吐量与主机 CPU 利用率64.2 对主机中断调节的敏感性74.3 重传恢复与序列号回绕74.4 持续回显吞吐量9

5 与其他 FPGA 传输控制协议实现的比较10

5.1 在通用 10GbE 平台上的单连接测量105.2 资源利用率与已发表的吞吐量测量结果 10

6 结论11

1 引言

在连续探测器读出系统中，当主机暂时以低于探测器产生数据的速度接收数据时，前端必须对数据进行缓冲。CDEX-50 的无触发读出就是一个例子 [1]。以太网链路也被用于将像素探测器数据从现场可编程门阵列（FPGA）传输到主机 [2]。传输控制协议（TCP）通过标准套接字提供有序字节流。FPGA 实现必须跟踪对端已确认的字节数以及可用的接收缓冲区空间。

Uchida 展示了使用 SiTCP 硬件处理器实现千兆以太网读出的方法 [3]。SiTCP-XG 将这一方法扩展至 10 千兆以太网 [4]。针对 CMS 数据采集系统，Bauer等人实现了一个面向 Linux 接收端的单向 10 千兆以太网传输控制协议发送器 [5]。Sidler等人解决了另一项需求：现场可编程门阵列上可同时建立的传输控制协议连接数量 [6]。开源 Limago 协议栈将这一架构扩展至百吉比特以太网 [7]。

XTCP（极限TCP）是一个开源端点，用于在FPGA读出逻辑与标准主机套接字之间传输连续数据流。其可综合源代码被封装为

可配置 Vivado IP 核。三种构建配置分别支持 10 千兆以太网（10GbE）、使用标准帧的 25 千兆以太网（25GbE），以及支持巨型帧的 25GbE。每个端点服务一个连接。本研究评估应用层吞吐量，即单位时间内传输的 TCP 负载位数量。通过独立的单向传输表征发送与接收路径。在一个 25GbE 比特流上测试了两种 IP 最大传输单元（最大传输单元），以在硬件资源固定的情况下测量帧大小的影响。持续回显测试检查了数据完整性，测试范围涵盖 TCP 32 位序列空间的多次循环。

2 端点架构

2.1 发送与接收路径

RDX051 读出卡通过一个光通道，将 FPGA应用程序连接到主机套接字（见图 1）。应用程序生成或使用字节流。XTCP 在可编程逻辑中实现 TCP 连接状态、分组处理和以太网媒体访问控制（MAC）。BASE-R 物理编码子层（PCS）和物理介质连接（PMA）将 MAC 连接到串行收发器。主机向配置的 IPv4 地址和端口发起 TCP 连接。同一链路上的独立用户数据报协议（UDP）寄存器接口可用于访问配置寄存器和计数器，包括 FPGA应用程序已使用的字节数。

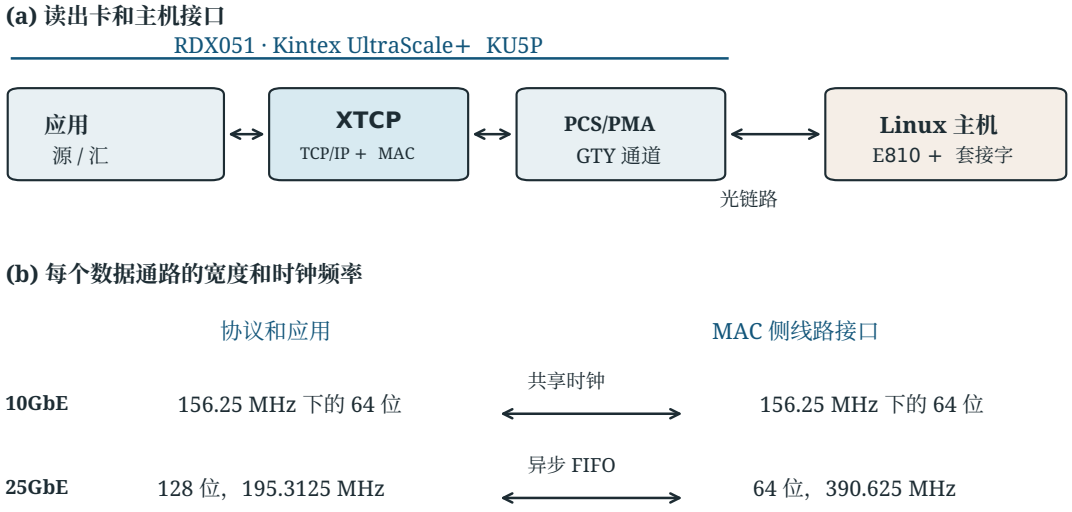


图1。RDX051 上 v1.0.0 测量的读出路径，以及两种链路速率对应的时钟频率。箭头表示双向接口。在 25 千兆以太网中，独立的发送和接收异步先进先出缓冲区（FIFO）连接两个时钟域。在两种链路速率下，差分收发器参考时钟均为 156.25 MHz。

发送接口接收一个数据字及其有效字节数，因此传输可以在部分字处结束。端点通过置位近满信号向源施加反压。它会将已接收的字节存储在循环缓冲区中，直到对端确认这些字节。数据包构建器

根据通告接收窗口、可用发送信用额度以及协商确定的最大报文段长度（MSS）选择字节范围。它构建首部、组装有效载荷并计算校验和。两个帧缓冲区使得一个帧可以在前一个帧传输期间进行组装。经过验证的累积确认（ACK）会释放相应的字节以供重新使用。在此之前，发送端可以重传这些字节，同时接收新的应用程序数据。

接收时，输入队列会缓存帧以供解析。端点会检查帧长度和分组长度、校验和、连接标识符以及 TCP序列号，然后再将数据交付给应用。端点使用字节地址和逐字节写使能，将连续的有效载荷字节写入应用的循环缓冲区。独立的重排序缓冲区可以保留在间隙之后接收的报文段。应用在消耗数据时会推进其读指针。通告接收窗口取决于应用缓冲区中的可用空间、可配置的限制值以及协商得到的窗口缩放。

图2 展示了确认数据与释放接收缓冲区空间之间的区别。填补间隙可以使累积确认前移，而应用读取指针保持不变。反过来，消耗数据可以增大通告窗口，而下一个预期序列号保持不变 [8]。

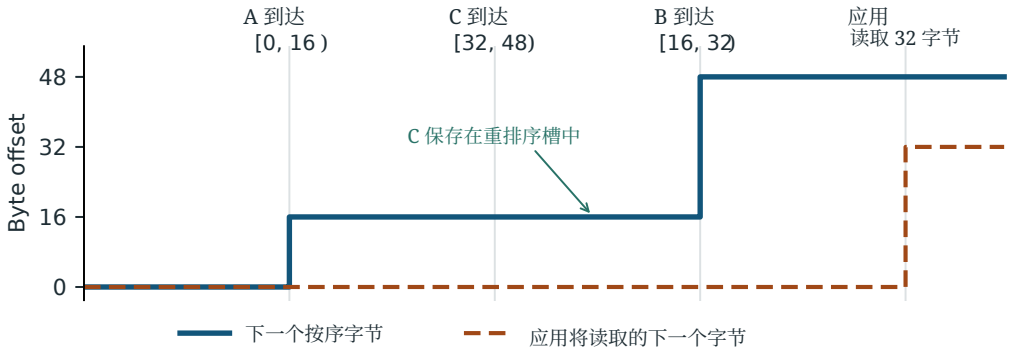


图 2。示意接收序列，字节偏移量相对于第一个有效载荷字节计算。分段按 A、C、B 的顺序到达。B 填补间隙且端点提交 C 后，累积确认前移至偏移量 48，确认截至偏移量 47 的字节。随后，应用消耗 32 字节。水平间距表示事件顺序。图中省略了确认传输和内部存储器周期。

2.2 数据通路宽度、时钟与片上存储

10GbE 实现以 156.25 MHz 的频率，每个时钟周期传输 8 字节。两种 25GbE 配置均采用 195.3125 MHz 的 128 位应用接口，以及 390.625 MHz 的 64 位 MAC 侧接口。在计入协议开销之前，每种配置均可传输 25 吉比特/秒。数据通路宽度加倍后，协议逻辑的时钟周期为 5.12 ns，而线路侧数据通路的周期为 2.56 ns。字节使能可在宽度转换过程中保留部分字。该设计采用 *verilog-ethernet* [9], width 提供的以太网成帧和循环冗余校验 (CRC) 逻辑

转换和异步 FIFO 来自 *verilog-axis* [10], , PCS/PMA 则来自 AMD 10G/25G 以太网子系统 5.0 版 [11]。

表1 列出了在 Vivado 2025.2 中完成布局布线后，Kintex UltraScale+ KU5P 现场可编程门阵列（`xcku5p-ffvb676-2-i`）上三种 XTCP v1.0.0 配置所使用的资源。端点统计包含 MAC 和帧处理逻辑。示例应用程序有自己的接收缓冲区。所有配置都为未确认数据的发送缓冲区分配 256 KiB。10GbE 设计为输入帧缓冲区分配 64 KiB，为应用接收缓冲区分配 256 KiB。25GbE 设计中的每个缓冲区均为 512 KiB。所有 TCP 缓冲都在芯片上。以 24.70 吉比特/秒的读出速率计算，一个初始为空的 256 KiB 发送缓冲区可容纳约 85 μ 秒的输入数据。等待确认的字节数以及接近满载的阈值，会降低主机服务中断期间可用的容量。

表 1。XTCP v1.0.0 布线后的 FPGA 资源利用率。端点各列统计查找表（LUT）、触发器（FF）、块 RAM 原语、UltraRAM 块和数字信号处理（DSP）块。电路板各列包含应用、PCS/PMA 和仪器模块。时序分析报告的最差建立裕量为 WNS。

	端点						电路板设计	
配置	LUTs	触发器	RAMB36	RAMB18	URAM	DSP	查找表	最差负时序裕量（纳秒）
10GbE 标准	14156	8 807	82	0	4	1	17 095	0.043
25GbE 标准	23 215	11272	198	2	4	1	26 740	0.127
25GbE 巨型帧	23 385	11227	216	2	32	1	26 851	0.032

25GbE 巨型帧设计中大部分新增内存用于接收端重排序。两种 25GbE 设计均提供 64 个重排序槽。为容纳 8960 字节的 TCP 负载，标准帧设计中的每个槽占用 2 KiB，巨型帧设计中的每个槽占用 16 KiB。槽容量增至八倍，使 UltraRAM 数量从 4 个增加到 32 个。更大的帧缓冲区和时钟域跨越 FIFO 需要额外使用 18 个 RAMB36 块。端点逻辑增加了 170 个查找表。支持巨型帧的实现也接受 1500 字节的最大传输单元，因此可以使用单一 FPGA 配置测试这两种帧大小。

3 测量方法

3.1 实验设置和传输测试

将三种 XTCP v1.0.0 配置依次编程到同一块 RDX051 网卡中。通过一条光通道，将该网卡直接连接到 Linux 主机中的 Intel E810 网络接口控制器（网卡）（Core Ultra 9 285 HX，内核 6.17.0-14-generic）。所有 TCP 连接均由主机发起。链路使用 IPv4，未启用 VLAN 标记。在端口处确认了协商后的链路速率以及已禁用的前向纠错（FEC）设置。每种链路速率均使用额定支持该速率的光缆。在使用 25GbE 光缆进行的初步 10GbE 测试中，主机记录到 63 个接收 CRC 错误，FPGA 记录到 17 个帧校验序列错误。使用对应的 10GbE 和 25GbE 光缆时，吞吐量测量期间这两项计数均未增加。

测试了四种设置。标准帧 10GbE 和 25GbE 设计使用 1500 B 的 IP MTU。支持巨型帧的 25GbE 设计分别在 1500 B 和 9000 B 下进行了测试。在每种设置下，11

项回显测试涵盖了从 1 到 65 539 B 的有效载荷长度，包括 7 和 8、15 至 17、1460 和 1461，以及 8960 和 8961 B。每个方向传输 1 048 593 B，以检查应用字节计数器。这些测试总计 52 项功能检查，涵盖了部分字以及跨越多个传输控制协议段的传输。

对于每种设置，主机在不同连接上运行三个工作负载。每个工作负载重复三次，共得到 36 次计时传输。对于 FPGA 到主机传输，现场可编程门阵列生成确定性字节模式，C 客户端则验证接收到的每个字节。对于主机到 FPGA 传输，客户端发送该模式，并等待已确认字节数达到请求值。FPGA应用程序字节计数器为传输长度提供了独立检查。每次单向传输包含 4 吉比字节 + 17 字节。对于回显，主机通过一个连接同时进行发送和接收。现场可编程门阵列从应用缓冲区读取接收到的字节，并通过传输控制协议将其返回。主机验证所有返回的字节，并等待未完成的发送字节得到确认。每次回显传输每个方向包含 1 吉比字节 + 17 字节。额外的 17 字节用于测试两种接口宽度下的末尾部分字传输。

客户端使用单调时钟测量经过时间，从连接建立完成后开始，到有效载荷传输完成时结束。对于主机发起的传输，该时间间隔包括等待最终确认的时间。通过 UDP 接口读取 FPGA应用程序字节计数器。主机对发送和接收均使用默认的套接字缓冲区大小及自适应中断调节。客户端 CPU 利用率按用户态和系统态 CPU 时间之和除以经过时间计算，其中 100% 对应一个逻辑核心。操作系统 CPU 计数器的变化给出了主机上繁忙逻辑核心的平均数量，其中包括中断处理和其他主机活动。

3.2 应用层吞吐量与理论极限

单个方向的应用层吞吐量为 $8N/T$ 比特/秒，其中 N 是在 T 秒的时间间隔内交付给接收应用的 TCP 有效载荷字节数。每个字节流位置只计数一次。协议头和重传不会增加有效载荷字节数。对于回显，N 是经主机验证后返回的字节数。因此，报告的回显吞吐量按每个方向计算。吞吐量以十进制单位表示。以吉比字节表示的传输大小采用 1 吉比字节 $= 2^{30}$ 字节。

对于由满尺寸分段组成的单向数据流，TCP 有效载荷理论最大吞吐量为

$$R_{\max} = R_{\mathrm{MAC}} \frac{M - 40\,\mathrm{B}}{M + 38\,\mathrm{B}}, \tag{3.1}$$

其中 R_{MAC} 为标称以太网 MAC 速率，M 为 IP MTU。在这些测试中，IPv4 和 TCP 首部各占 20 字节。分母包括 14 字节的以太网头、4 字节的帧校验序列、前导码和帧起始定界符占用的 8 个字节时间，以及帧间隙占用的 12 个字节时间 [12]。该模型假设反向信道能够承载 ACK 流量。在全部 36 个连接中，均未协商 TCP 时间戳。对于 1500 和 9000 字节的 MTU，Linux 报告的发送 MSS 值分别为 1460 和 8960 字节。MTU 为 1500 字节时，式（3.1）在 10GbE 下给出 9.493 吉比特/秒，在 25GbE 下给出 23.732 吉比特/秒。MTU 为 9000 字节时，25GbE 的上限为 24.784 吉比特/秒。回显期间，ACK 与数据段共享各链路方向，且应用将发送与接收耦合在一起。

3.3 持续回显与完整性验证

持续测试使用了三个原型比特流，每种配置各一个，测试时间早于 v1.0.0 构建版本。发送端根据字节的位置和一个种子生成有效载荷字节。接收端会检查同一连接上返回的每个字节。测试安排持续一小时，并留出约 5 秒让未完成的数据排空并关闭连接。每次运行都产生 359 个完整且互不重叠的 10 秒时间间隔。总体平均值包括传输结束前返回的所有字节。对于 10GbE 原型，在区间边界处对经验证的接收字节数进行了插值。对于两个 25GbE 原型，每个区间的吞吐量取该区间的发送速率和经验证的接收速率中的较小值。第五百分位数取 359 个区间速率按升序排列后的第 18 个值。

10GbE 原型使用了 Intel X710 主机适配器。两个 25GbE 原型均使用 E810 主机，并启用了自适应中断调节；FPGA 接收窗口上限设为 480 KiB，主机接收和发送线程使用相同的一对逻辑核心。所有持续运行均采用相同的字节验证标准。固定长度的 v1.0.0 测量在三种配置下均使用 E810 主机。补充数据通过比特流哈希标识每个已编程映像。

4 结果

4.1 单连接吞吐量与主机 CPU 利用率

表2 给出了各 v1.0.0 配置的吞吐量中位数。图3 显示了每种设置下三次试验的最小值和最大值。对于 10GbE 标准帧，两个单向吞吐量中位数与理论 TCP 有效载荷上限的差距均不超过 0.1%。对于标准帧 25GbE 设计，FPGA 到主机和主机到 FPGA 的中位数分别达到该上限的 98.8% 和 93.0%。在这些单向测试中，两个端点均未记录到 TCP 重传。

表 2。XTCP v1.0.0 各配置的应用层吞吐量。测得的速率为每种工作负载下三个连接的中位数。所有速率均以吉比特/秒为单位。理论极限对应于满尺寸分段的单向传输。

配置	最大传输单元（字节）	现场可编程门阵列到主机	主机到现场可编程门阵列	每个方向的回显
10GbE 标准	1500	9.493	9.487	9.486	9.427
25GbE 标准	1500	23.732	23.443	22.068	12.155
25GbE 巨型帧	1500	23.732	23.296	22.093	12.115
25GbE 巨型帧	9000	24.784	24.702	24.077	16.419

在同一支持巨型帧的比特流上，将 MTU 从 1500 字节增至 9000 字节后，FPGA 到主机方向的中位吞吐量从 23.296 吉比特/秒升至 24.702 吉比特/秒，主机到 FPGA 方向则从 22.093 吉比特/秒升至 24.077 吉比特/秒。在较大的 MTU 下，这些数值分别相当于理论 TCP 有效载荷上限的 99.67% 和 97.15%。在有效载荷吞吐量不变的情况下，将 MSS 从 1460 字节增至 8960 字节，可使每秒满尺寸报文段的数量减少 6.14 倍。在 FPGA 到主机测试中，客户端 CPU 利用率（包括字节校验）从单个逻辑核心的 67.1% 降至 48.5%。主机整体的平均 CPU 利用率从 1.17 个繁忙核心降至 0.70 个繁忙核心。在反向传输中，客户端 CPU 利用率从 44.0% 降至 38.2%，相应的主机整体数值为 0.57 和 0.58 个繁忙核心。

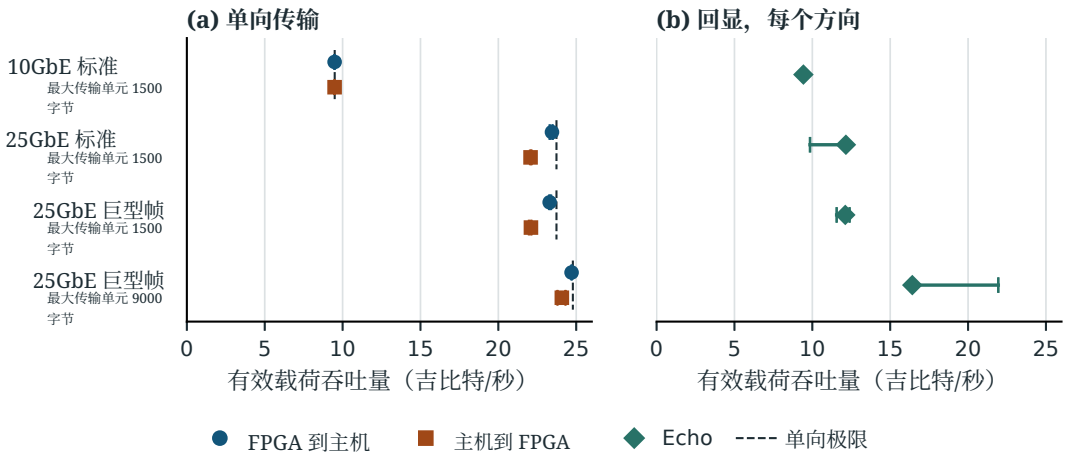


图 3。每次通过一个 TCP 连接进行固定长度传输。符号表示中位数，水平线表示三次试验中的最小值和最大值。(a) 中的虚线标示了根据公式（3.1）得出的理论极限。(b) 中的每个数据点均按每个方向各计一次回显的有效载荷。

在回显测试中，FPGA 应用程序会返回接收到的数据，同时主机的发送和接收线程并发运行。在 10GbE 下，每个方向的回显中位吞吐量为 9.427 吉比特/秒。对于支持巨型帧的 25GbE 比特流，MTU 为 1500 字节时的中位吞吐量为 12.115 吉比特/秒，MTU 为 9000 字节时则为 16.419 吉比特/秒。在 9000 字节 MTU 下，三次运行的吞吐量介于 16.346 和 21.942 吉比特/秒之间，两个端点均未记录到 TCP 重传。使用标准帧25GbE比特流并将 MTU 设为 1500 字节时，中位吞吐量为 12.155 吉比特/秒，最低为 9.852 吉比特/秒。

4.2 对主机中断调节的敏感性

在对标准帧 25GbE 原型进行的一项独立测试中，每次单向传输的数据量为 16 吉比字节。在两个 FPGA 编程周期中，每个方向分别测量了三条连接在自适应接收中断调节下的表现，以及三条连接在固定接收间隔 8 μ秒下的表现。两种设置使用相同的主机源端口，发送中断调节则保持自适应。在主机到 FPGA 的传输中，两个周期的中位吞吐量在自适应设置下分别为 23.023 和 23.022 吉比特/秒，在 吉比特/秒、间隔为 8 μ秒时分别为 23.222 和 23.407。在反向传输中，自适应设置下的中位数分别为 23.406 和 23.325 吉比特/秒，而间隔为 吉比特/秒时分别为 23.496 和 23.478 8 μ秒。两个周期中，固定间隔在两个方向上的中位数均较高。在主机到 FPGA 的传输方向，网卡还会处理返回的确认。比特流、传输长度和计时流程均不同于表2中的设置。

4.3 重传恢复与序列号回绕

为避免对同一笔未完成数据反复进行快速重传，10GbE 发送端会检查累积确认是否已越过保存的恢复边界 [13]。32 位 TCP 序列号每 4 GiB 回绕一次。确认一旦前进超过 2 GiB，有符号比较就可能将其判定为落在一个已经越过的边界之前。在 9.49 吉比特/秒的速率下，这个半周期为 1.81 秒。修订后的

发送端记录严格超过该边界的第一个已验证的 ACK，并在设置新边界时清除此标志（图 4(a)）。

FPGA 计数器每隔约 0.10 秒采样一次，在两次 96 GiB 回显运行中标出了 15 个事件（图 4（b））。在较早的连接中，全部五次快速恢复都发生在允许半周期内，而全部四次超时都发生在受阻半周期内。每次超时之前都有三个符合条件的重复 ACK。过时的比较逻辑阻止了快速重传，导致发送端只能等待其计时器。修订后的发送方进行了六次快速恢复，且没有发生超时。在这六个序列位置，较早的比较逻辑本可允许两次进入恢复，但阻止了另外四次。

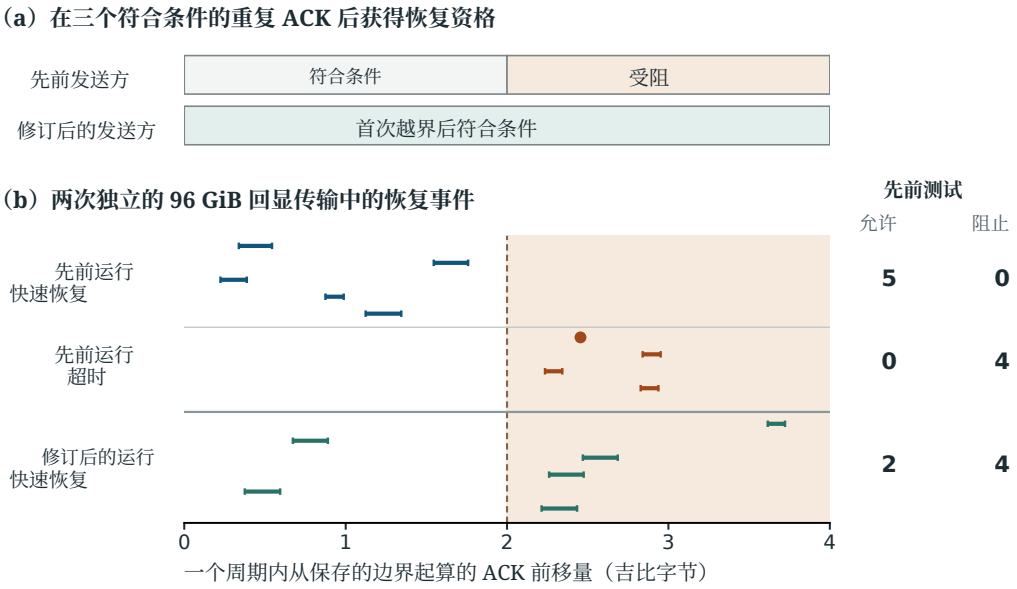


图 4。跨越 4 GiB 序列周期的快速恢复。(a) 首次 ACK 越界后的边界测试。(b) 两次独立、非配对运行中的所有事件，按各组内的时间顺序排列。柱线界定了相对于保存的边界、以 4 GiB 为模的 ACK 前移量。圆点表示上下界重合。阴影和计数将先前的比较方法应用于两次运行。

在分别进行的持续回显运行中，较早版本和修订版设计分别记录了 121 次和 5 次超时。在 359 个 10 秒时间间隔中，吞吐率低于 9 吉比特 / 秒的时间间隔从 113 个降至 5 个。在一次单独的 1 GiB 传输中，修订版设计记录了 1 次超时，并以 4.478 吉比特/秒的速率返回了所有字节。此次传输覆盖的序列空间不到一半，因此不能仅凭完成时间来判断是否发生了序列号回绕。

10GbE 原型的受控测试涵盖了丢包、重复、乱序、零通告窗口和重新连接。在一次 512 MiB 的传输过程中，接收端故障注入导致 180 个数据包被丢弃。现场可编程门阵列记录到一次重传超时，并重放了 180 个发送描述符。主机在 1.458 s 内验证了整个有效载荷。25GbE 设计针对接收窗口重新开放、报文段长度边界和持续数据完整性进行了测试。对于巨型帧设计，仿真涵盖了未对齐的有效载荷长度，并独立验证了 54 个长度处于边界值的测试帧的以太网 CRC、IPv4 校验和及 TCP 校验和。

4.4 持续回显吞吐量

在三次持续测试中，现场可编程门阵列均返回了完整的传输数据流，且主机验证了每个字节（见表3）。25GbE 巨型帧原型通过一个连接，在 3595.002 s 内返回了 8 768 350 064 640 字节，两个方向各自经历了超过 2000 个完整的 TCP 序列号循环。其平均回显吞吐量为每个方向 19.512 Gbit/s，标准帧原型则为 13.283 Gbit/s 。两次测试使用了相同的主机和接收窗口上限，但最大传输单元和现场可编程门阵列缓冲区分配均不同。

表 3。三种工程原型上的连续回显，每行一个连接。返回数据以十进制 TB 为单位，采用每 TB 10^{12} B。最右侧三列给出每个方向的吞吐量，单位为吉比特/秒。百分位数和最小值采用完整的 10 秒时间窗口计算。

原型	耗时（秒）	返回量（太字节）	Mean	第 5 百分位数	最小值
10GbE 标准	3595.002	4.246	9.450	9.457	8.503
25GbE 标准	3595.001	5.969	13.283	11.395	11.306
25GbE 巨型帧	3595.002	8.768	19.512	16.447	16.439

图 5 给出了回显吞吐量随时间的变化。

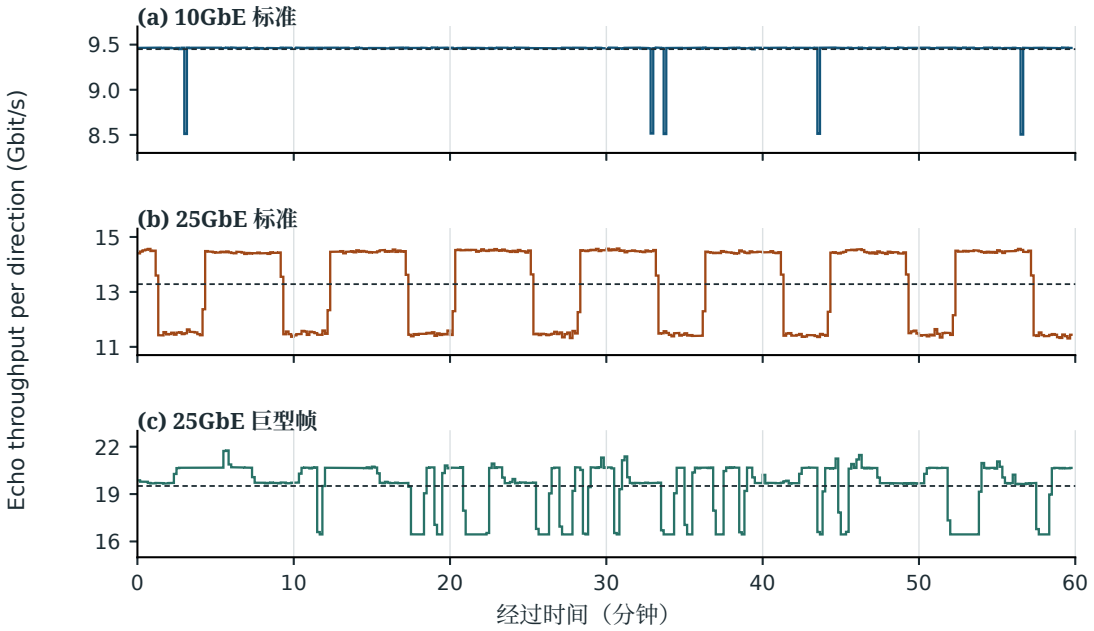


图 5。每个原型在连续 359 个 10 秒时间间隔内测得的每个方向的回显吞吐量。虚线表示包括最后一个不完整时间间隔在内的全程平均值。面板 (a) 使用经验证的接收字节数。面板 (b) 和 (c) 使用每个时间间隔内发送字节数与经验证的接收字节数中的较小值。垂直刻度不同，以便解析每种配置内的变化。

10GbE 测试中有几个低吞吐量区间，使整体平均值低于各区间速率的第五百分位数。对于标准帧 25GbE 原型，区间吞吐量反复在约 11.4 至 14.5 吉比特/秒之间波动。对于巨型帧原型，359 个区间中有 71 个

区间低于 18 吉比特/秒，其中包括一段持续 120 秒的连续时段。该次测试中，两个端点均未记录到 TCP 重传。低吞吐量区间的持续时间与连续输入流的缓冲有关。要确定其原因，需要对两个端点的队列占用和服务进行时间分辨测量。

5 与其他 FPGA 传输控制协议实现的比较

5.1 在通用 10GbE 平台上的单连接测量

SiTCP-XG [4] 和 10GbE XTCP 原型（表 3）使用了相同的 KU5P 卡、光链路和配备 Intel X710 网卡的 Linux 主机（MTU 为 1500 字节，MSS 为 1460 字节）。对于每个端点和方向（表 4），同一客户端分别在两组、每组三次的设置下，对六次 32 吉比字节传输进行计时。对于 FPGA 到主机传输，计时在主机验证最后一个字节时结束；对于主机到 FPGA 传输，计时在确认 FPGA 应用程序的字节计数时结束。表 2 报告了在 E810 主机上进行的较短传输。

表 4。在通用 10GbE 平台上进行的六次传输的单连接中位数。客户端 CPU 时间包括每吉比字节有效载荷对应的用户态和系统态时间。

	吞吐量（吉比特/秒）		进程 CPU 时间（秒/吉比字节）	
端点	现场可编程门阵列→主机	主机→现场可编程门阵列	现场可编程门阵列→主机	主机→现场可编程门阵列
SiTCP-XG	9.491	5.758	0.322	0.165
XTCP 原型	9.493	9.488	0.318	0.064

两个端点均以相近的客户端 CPU 开销向主机提供了 9.49 吉比特/秒的吞吐量。在反向传输方向上，SiTCP-XG（128 KiB 发送缓冲区和 32 KiB 应用接收缓冲区）与 XTCP（256 KiB 应用接收缓冲区）公布的最终窗口中位数分别为 32 752 B 和 262 124 B。对于 SiTCP-XG，Linux 将发送端繁忙时间的 99.998–100% 判定为受接收窗口限制；对于 XTCP，该比例为 99.758–99.962%。即使在 9.488 吉比特/秒的速率下，这一时间比例仍接近 1。1.65 倍的吞吐量差异适用于这些配置。若要单独确定窗口大小的贡献，需要独立改变窗口大小。

一次单独的 SiTCP-XG 回显测试在 1.425 秒后因有效载荷不匹配而停止。X710 接收 CRC 错误计数增加了 19。计数器无法确定这两类错误中的任何一类的原因，并且此次运行无法提供持续吞吐量估计。

5.2 资源利用率与已发布的吞吐量测量结果

表 5 列出了每个端点或 TCP卸载引擎（TOE）所统计的模块。两个 10GbE 端点的统计均包含 MAC。完整的 SiTCP-XG 和 XTCP 原型设计（包括应用、PCS/PMA 和调试逻辑）分别使用了 8423 和 17023 个查找表。其应用程序接收缓冲区分别占用 8 个和 64 个块 RAM 瓦片。XTCP 还分配了片上存储器，用于发送保留、输入帧和段重排序。

Design Gateway 的单连接TOE25G-IP具有128位用户端口，最低时钟频率为195.3125 MHz。其 KU5P 示例不包含 MAC，并为发送数据分配64 KiB、为发送数据包分配16 KiB、为接收数据分配64 KiB [14]。其 PC演示报告显示，在采用8960字节有效载荷和主机端模式校验时，从现场可编程门阵列到主机的速率为22.63 Gbit/s；采用现场可编程门阵列端校验时，从主机到现场可编程门阵列的速率为21.24 Gbit/s。每个数值对应一次68.72 吉字节

表5。所列模块的 FPGA 资源利用率。一个块RAM块代表一个 RAMB36原语或两个 RAMB18原语。前四行给出了本研究所用 KU5P 的布线后计数。其余各行给出了引用来源中报告的计数。短横线表示未报告的 UltraRAM 计数，n/a 表示器件不含 UltraRAM。

设计与模块	FPGA	LUTs	FF、BRAM 块、URAM		
SiTCP-XG，端点	KU5P	5406	9954	70.5	0
XTCP 10GbE 原型，端点 KU5P		13 802	8787	82	4
XTCP 25GbE标准，端点	KU5P	23 215	11272	199	4
XTCP 25GbE巨型帧，端点	KU5P	23 385	11 227	217	32
TOE25G-IP，TOE [14]	KU5P	3904	3869	20	2
Sidler等人，TCP/IP协议栈 [6]	VX690T	19 026	20611	279	n/a
Limago，TCP卸载引擎 [7]	VU9P	36 469	36 229	247.5	–

传输 [15]。XTCP 巨型帧的中位数分别为24.70和24.08 Gbit/s。主机、传输长度和验证过程各不相同。该供应商分别介绍了资源示例和吞吐量演示。

在 MSS 为 536 字节时，Sidler等人测得，通过 25 个连接进入现场可编程门阵列的聚合应用吞吐量为 8.5 吉比特/秒，通过 500 个连接从现场可编程门阵列传出的吞吐量为 8.6吉比特/秒。其协议栈资源数量不包括网络接口和外部存储器接口 [6]。Limago TCP卸载引擎条目针对的是在禁用窗口缩放的情况下支持 10 000 个连接。其公布的单连接速率包括以太网、互联网协议和传输控制协议报头。其主机应用程序测量使用了最多 30 个并发连接 [7]。表2 中的 XTCP 速率统计的是单个连接上的有效载荷。

6 结论

开源 XTCP 端点能够在单个连接上维持较高的应用层吞吐量。在 10GbE 下，XTCP 和 SiTCP-XG 向主机传输的速率均达到 9.49 吉比特/秒。在所测试的配置中，XTCP 的主机到 FPGA 吞吐量是 SiTCP-XG 的 1.65 倍；两者在接收缓冲区大小和通告窗口方面存在差异。XTCP 原型使用了 13 802 个查找表和 4 个 UltraRAM 块，而 SiTCP-XG 使用了 5406 个查找表且未使用 UltraRAM。

在 25GbE、最大传输单元为 9000 字节的情况下，XTCP 在独立传输中向主机传输了 24.70 吉比特/秒，向现场可编程门阵列传输了 24.08 吉比特/秒。该端点占用了 23 385 个查找表、217 个块 RAM 瓦片和 32 个 UltraRAM 块。在持续回显测试中，巨型帧原型每个方向的平均速率为 19.51 吉比特/秒；按 10 秒时间间隔统计，第五百分位数为 16.45 吉比特/秒。对于连续数据源，这类时间间隔的持续时间决定了上游缓冲需要吸收输入速率与输出速率之间差值的时长。

数据和代码可用性

可综合的 XTCP 源代码、Vivado IP 软件包、RDX051 参考设计和主机测试工具可从 https://github.com/RFSoC47/XTCP 获取。补充材料包含各项吞吐量测量结果、布线后的资源计数，以及持续回显测试的时间序列。

致谢

作者感谢清华大学工程物理系电气车间的技术工程师 Y. Xue 和 J. Zhang，感谢他们精湛的焊接技艺以及在电路板组装方面给予的耐心协助；没有他们的帮助，这些测量便无法完成。

OpenAI Codex 协助整理实验台记录、编写绘图和验证脚本，以及撰写和编辑手稿。

参考文献

[1] H. Yang, T. Xue, Q. Pan, B. Liang, Y. Liu and J. Li, 《CDEX-50无触发读出电子系统的设计》, 《仪器学报》19 (2024) T10005。[2] H.-x. Li, J. Zhang, Y. Ding, S.-s. Cui, Z. Wang and X.-s. Jiang, 《面向像素探测器的基于10G以太网的数据采集系统的现场可编程门阵列实现》, 《辐射探测技术与方法》4 (2020) 31。[3] T. Uchida, 《面向千兆以太网的基于硬件的TCP处理器》, 《IEEE核科学汇刊》55 (2008) 1631。[4] Bee Beans Technologies, “SiTCP-XG库”。产品文档。[5] G. Bauer, T. Bawej, U. Behrens, J. Branson, O. Chaze, S. Cittolin 等, 《面向高能物理的现场可编程门阵列输出的10 Gbps TCP/IP数据流》, 《物理学杂志：会议系列》513 (2014) 012042。[6] D. Sidler, G. Alonso, M. Blott, K. Karras, K. Vissers and R. Carley, 《面向可重构硬件的可扩展10Gbps TCP/IP协议栈架构》, 载于《2015年第23届IEEE年度现场可编程定制计算机国际研讨会》, 第36–43页, 2015, DOI。[7] M. Ruiz, D. Sidler, G. Sutter, G. Alonso and S. López-Buedo, 《Limago：基于现场可编程门阵列的开源100 GbE TCP/IP协议栈》, 载于《2019年第29届现场可编程逻辑与应用国际会议（FPL）》, 第286–292页, 2019, DOI。[8] W. Eddy（编）, 《传输控制协议（TCP）》。RFC 9293, RFC编辑部, 2022。10.17487/RFC9293。[9] A. Forencich, “Verilog以太网”。https://github.com/alexforencich/verilog-ethernet。[10] A. Forencich, “Verilog AXI流组件”。https://github.com/alexforencich/verilog-axis。[11] AMD, “10G/25G高速以太网子系统产品指南（PG210）”。子系统概述, 访问于2026年11月22日。[12] IEEE, 《IEEE以太网标准》。IEEE Std 802.3-2022, 2022。10.1109/IEEESTD.2022.9844436。[13] T. Henderson, S. Floyd, A. Gurtov and Y. Nishida, 《TCP快速恢复算法的NewReno修改》。RFC 6582, RFC编辑部, 2012。10.17487/RFC6582。[14] Design Gateway, 《TOE25G IP核数据手册》。修订版1.5, 表1, 2023。[15] Design Gateway, 《TOE25G-IP带CPU演示说明》。修订版1.3, 图3-4和3-8, 2023。