Xilinx 5G Telco Accelerator Cards

Xilinx Wired and Wireless Group
Xilinx is expanding our 5G product offering with the addition of Telco Accelerator Cards

The first card in our Telco Accelerator series is called “T1”

T1 is a two-in-one 75W “plug and play” card that accelerates Fronthaul and L1 Functions and is already in limited sampling

The “Virtualization of 5G” provides a great opportunity to expand our existing footprint in 2nd and 3rd phases of 5G deployments
Radio Access Network

Towers

Radio Unit (RU)

Edge Cloud

T1

Distributed Unit AKA “Baseband Unit”

Core Network

Centralized Unit (CU)

Edge/Core Routers
What is the Virtualization of 5G?

Traditional Model (LTE)

- OEM’s create Proprietary equipment for Wireless Deployments
- 5G Operators are moving away from this model
  - “Vendor Lock” creates a lack of competition
  - Difficult for Operators to deploy Software Services
    - VR, Gaming, Automotive etc.

Open-RAN Model (5G)

- Virtual BBU implemented in a standard server form factor
  - Similar to what we saw with “Open Compute” 10 years ago
  - Utilize Open Interfaces for Multi-Vendor Compatibility
  - New players can drive competition and innovation
  - Software Services can now be deployed all the way to the Edge!
Open RAN Concept is Growing in Popularity… Fast

- In 2019 and prior, most Xilinx 5G customer demand came from traditional OEMs

- In 2020, demand dramatically shifted towards Open and Virtualized RAN architectures
  - O-RAN Radio Unit (O-RU)
  - O-RAN Distributed Units (O-DU)

- Xilinx Telco Accelerator Cards address the O-DU portion of the Open vRAN Market

Projected Market Split
Traditional RAN vs. Open vRAN

Source: Data extracted from ABI Research (www.abiresearch.com)
5G Virtualization
Traditional Base Band Unit

Traditional Baseband Unit (BBU) from an OEM

Let’s look at what’s inside…
What’s Inside a Traditional Base Band Unit?

Traditional BBU

Chips inside a traditional BBU

**General Purpose Processor**
L2/L3 Protocol Layer Processing
Chips Used ➔ x86 or ARM

**Fronthaul FPGA**
Terminates CPRI traffic to/from Radio Unit
Chips Used ➔ Mid-sized FPGA’s like Kintex or Zynq

**Layer 1 Baseband ASIC or FPGA**
Low-PHY and High-PHY Functions
Chips Used ➔ First Deployments are FPGA, then ASIC’s in 2nd Gen
Introducing T1 – Fronthaul and L1 Offload FPGA Card

**PCle Form Factor Card**
FHHL 75W Card
NEBS Compliant

**Fronthaul Termination**
2x SFP28 Cages for Radio links
16 Layers @ 100MHz OBW

**L1 Lookaside Acceleration**
High throughput FEC at low power
16 Layers @ 100MHz OBW

Same Chips – But in a O-RAN Compliant PCle Card!
Frees up GPP’s for Software at the Edge
O-RAN Virtual BBU in a Commodity Server

**Standard Server**
Can be ruggedized or not, based on environment
Available from Dell, SuperMicro, HPE etc.

**General Purpose Processor**
Upper-Layer Protocol Layer Processing (Open RAN etc.)
Can be x86 or ARM

**Fronthaul and L1 Offload in PCIe Cards**
Processors cannot handle high-volume 5G traffic alone
L1 and Fronthaul functions managed in PCIe Cards
New Open RAN Virtual Base Band Unit (vBBU)

New vBBU Distributed Unit
Open, Disaggregated, FPGA Accelerated!
T1 Deeper Dive
T1 Offloads Difficult Functions from the CPU

O-RAN Stack

- **O-DU Split 6**
  - MAC
  - Channel Encoding
  - Channel Decoding
  - Rate Matching
  - Rate De-Matching
  - Scrambling
  - Descrambling
  - Modulation
  - Demodulation
  - Layer Mapping
  - IDFT (optional)
  - Equalization
  - Precoding
  - Channel Estimation
  - HARQ
  - HARQ

- **O-DU Split 7.2**
  - MAC
  - Fronthaul eCPRI
  - PTP Timing
  - Fronthaul eCPRI
  - PTP Timing

- **O-RU Split 8**
  - Precoding
  - Resource Element Mapping
  - Beamforming
  - IFFT
  - Cyclic Prefix Inversion
  - Digital and Analog RF
  - Resource Element Demapping
  - Port Reduction
  - FFT
  - Cyclic Prefix Removal
  - Analog and Digital RF

**Red Functions**: Map well to FPGA

**Grey Functions**: Map well to CPU

**Total Card Power**: < 75W
Demonstrated T1 Performance using FlexRAN on Dell R740

### Stand-Alone Server (No T1)

**Server**

<table>
<thead>
<tr>
<th>Component</th>
<th>Process</th>
<th>Split 6</th>
<th>MAC</th>
<th>Process</th>
<th>Split 6</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>Rate Matching</td>
<td></td>
<td>MAC</td>
<td>Rate Matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel Encoding</td>
<td>Scrambling</td>
<td></td>
<td></td>
<td>Channel Decoding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scrambling</td>
<td></td>
<td></td>
<td></td>
<td>Rate De-Matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modulation</td>
<td></td>
<td></td>
<td></td>
<td>Descrambling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layer Mapping</td>
<td></td>
<td></td>
<td></td>
<td>Demodulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fronthaul eCPRI</td>
<td></td>
<td></td>
<td></td>
<td>IDFT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTP Timing</td>
<td></td>
<td></td>
<td></td>
<td>Equalization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**L1 Performance**

<table>
<thead>
<tr>
<th>Component</th>
<th>Throughput</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoder</td>
<td>0.718 Gbps</td>
<td>45 us</td>
</tr>
<tr>
<td>Decoder</td>
<td>0.183 Gbps</td>
<td>62.7 us</td>
</tr>
</tbody>
</table>

**Fronthaul**

- FH: 4T4R @100 MHz OBW
- **2 Sectors** w/ Redundant Port: NIC-Dependent, 24 Cores
- **4 Sectors**: NIC-Dependent, 64 Cores

### Server With T1 Card

**Server** + **T1 Card**

<table>
<thead>
<tr>
<th>Component</th>
<th>Process</th>
<th>Split 6</th>
<th>MAC</th>
<th>Process</th>
<th>Split 6</th>
<th>MAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>Rate Matching</td>
<td></td>
<td>MAC</td>
<td>Rate Matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel Encoding</td>
<td>Scrambling</td>
<td></td>
<td></td>
<td>Channel Decoding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate Matching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scrambling</td>
<td></td>
<td></td>
<td></td>
<td>Rate De-Matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modulation</td>
<td></td>
<td></td>
<td></td>
<td>Descrambling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layer Mapping</td>
<td></td>
<td></td>
<td></td>
<td>Demodulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fronthaul eCPRI</td>
<td></td>
<td></td>
<td></td>
<td>IDFT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTP Timing</td>
<td></td>
<td></td>
<td></td>
<td>Equalization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**L1 Performance**

<table>
<thead>
<tr>
<th>Component</th>
<th>Throughput</th>
<th>Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoder</td>
<td>17.7 Gbps</td>
<td>14.15 us</td>
</tr>
<tr>
<td>Decoder</td>
<td>7.8 Gbps</td>
<td>16.21 us</td>
</tr>
</tbody>
</table>

**Fronthaul**

- FH: 4T4R @100 MHz OBW
- **2 Sectors** w/ Redundant Port: NIC-Dependent, 24 Cores
- **4 Sectors**: NIC-Dependent, 64 Cores

**T1 Provides L1 Performance**

- 42x Higher Encoder Throughput
- 24x Higher Decoder Throughput
- 3.2x Lower Encoder Latency
- 3.8x Lower Decoder Latency

**T1 Provides Real 5G Fronthaul**

- Sub ns PTP Timestamping
- ORAN Layer Mapping in NIC
- HW Redundancy and Fallback

---

*Traditional NIC’s provided none of this*
Available Reference Designs

L1 Reference Design

- **L1 Software**
  - BBDev API
  - QDMA Driver
  - L1 Stack

- **T1 Card**
  - PCIe Gen3
  - QDMA
  - CB CRC Attach / Detach
  - Rate Matching / De-matching
  - LDPC Encode
  - LDPC Decode
  - HARQ Engine

Fronthaul Reference Design

- **Fronthaul Software**
  - C/ U/ S/ M-Plane Software
  - DPDK API’s and IQ Streaming I/F
  - Linux TCP Stack and QDMA Driver

- **T1 Card**
  - PCIe Gen3
  - QDMA
  - S/M Plane Queues
  - O-RAN Framer
  - Synchronization
  - eCPRI Framer
  - Packet Parser and Interconnect
  - 25G + PTP

**Fully Operational Reference Designs**
Removes adoption barriers for companies that are not FPGA savvy

**FlexRAN Software Stack**
Layer-1 → BBDev standard API’s
Fronthaul → DPDK Drivers

**Standard QDMA Interface**
Same interface used by Alveo

**FPGA IP and Integration is already done!**
No need for RTL team or additional 3rd parties
T1 Deployment Scenarios

2 Sector with Redundancy

High Reliability Deployment
One SFP28 for Traffic
Second SFP28 for Fallback
Full ORAN Classification on Card

Oversubscribed with FHGW

Highest RU / DU Ratio
Requires external FHGW for oversubscription
Moves DU closer to Core
Can merge DU and CU

4 Sector Direct to RU’s

High RU / DU ratio w/o FHGW
Scalable (more cards = more radios)
Uses Radio Daisy Chaining

Highly Scalable

More T1 Cards = More Radios
Fronthaul and L1 scale together
Frees up XEON’s for Operator Services
Review of T1

Replaces two incumbent cards with a single T1
Fronthaul and L1 on a single 75W card

<table>
<thead>
<tr>
<th>T1 (Sampling Now)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Form Factor</td>
<td>FHHL PCIe Card</td>
</tr>
<tr>
<td>Optimization</td>
<td>Hybrid (Fronthaul + L1)</td>
</tr>
<tr>
<td>FH Ports</td>
<td>2x SFP28 + 1588</td>
</tr>
<tr>
<td>FH BW</td>
<td>4 sectors of 4TRX @ 100HMz *</td>
</tr>
<tr>
<td>IEEE 1588</td>
<td>Yes – Stamp at PHY</td>
</tr>
<tr>
<td>L1 Encode</td>
<td>17.7 Gbps *</td>
</tr>
<tr>
<td>L1 Decode</td>
<td>7.8 Gbps *</td>
</tr>
</tbody>
</table>

T1 Performance Advantage

- 42x Higher Encoder Throughput
- 24x Higher Decoder Throughput
- 3.2x Lower Encoder Latency
- 3.8x Lower Decoder Latency
- Sub ns PTP Timestamping
- ORAN Layer Mapping in NIC
- HW Redundancy and Fallback

* Demonstrated live in T1 Ref Design Demo
Thank You
Xilinx Mission

Building the Adaptable, Intelligent World