

1 Simulations

1.1 4x4 Crossbar

Bestimmen und begründen Sie wie die Werte zustande kommen

When observing the crossbar, we see certain values for minimum and maximum, as well as in between. For example the waiting time inside the arbiter queue is one of

$$t_{\text{wait}} \in \{0, 512, 1024, 1536\} \quad (1)$$

To get to a approximation of the avg values we then needed to weight these values by certain criteria:

- A number of Nodes is generating messages for a subset of nodes (could be from 1 to 4 in size) $\Rightarrow \binom{N}{n}$
- These subsets could be distributed in certain ways to the generating nodes. $\Rightarrow (N - n + 1)!$
-

$$Avg = \sum_{n=0}^N \frac{\binom{N}{n} (N - n + 1)! \cdot C_n}{\sum_{k=0}^N \binom{N}{k} (N - k + 1)!} \quad (2)$$

1.1.1 avg Input Queue Length

- Generated Packets
 - size: $s = 512 \text{ b}$
 - send interval: $t = 512 \text{ ns}$
- Connection to XBar
 - data rate: 1 Gbps

$$r_{\text{gen}} = \frac{512 \text{ b}}{512 \text{ ns}} = 1 \text{ Gbps} \quad (3)$$

$$C_n = 250 \text{ Mbps} \cdot n \quad (4)$$

$$r_{\text{q fill, eff}} = \sum_{n=1}^4 \frac{\binom{4}{n} (5 - n)! \cdot C_n}{141} \quad (5)$$

$$= 347.5 \text{ Mbps} \quad (6)$$

$$\Rightarrow l_{\text{q, avg}} = \frac{t_{\text{sim}} \cdot r_{\text{q fill, eff}}}{2 \cdot s} = 3394 \quad (7)$$

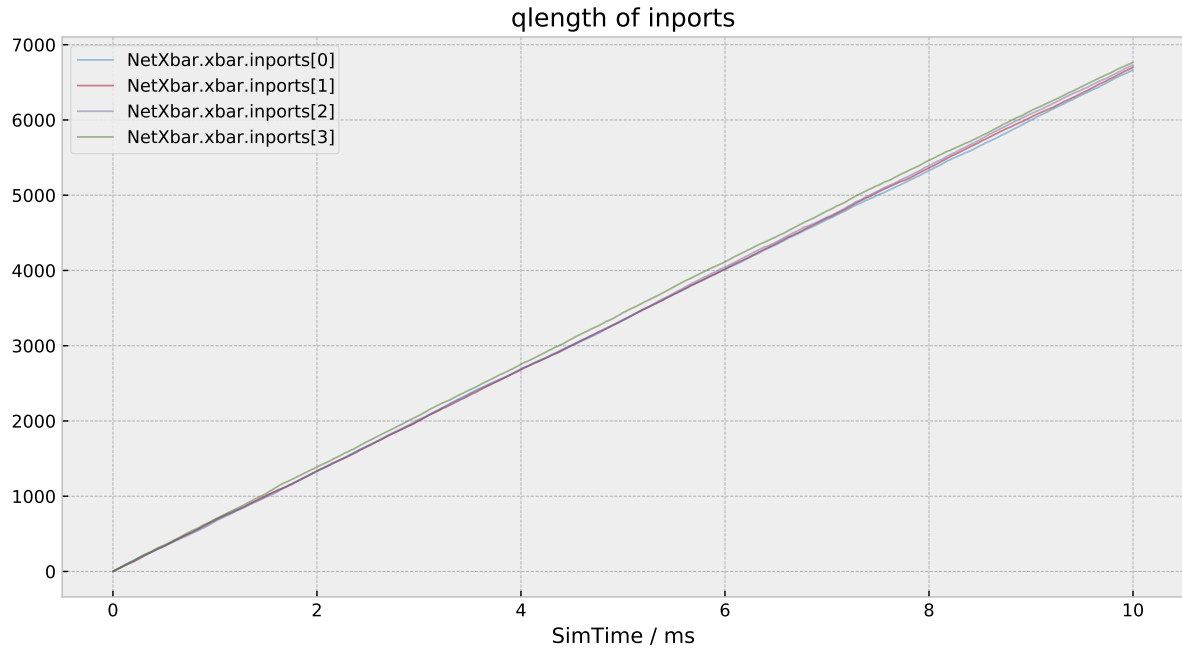


Abbildung 1: Pre Optimization

1.1.2 avg End-to-End Latency

- (no delays inside buffers, because the generated data rate, even for all 4 apps, is lower than a single data rate channels maximum throughput)
- minimum
 - App \rightarrow C \rightarrow Inport \rightarrow C \rightarrow Outport \rightarrow C \rightarrow App
 - delay for packet: $t_{delay} = 512ns$ (per `DatarateChannel C`)

$$\Rightarrow t_{e2e,min} = 1.536 \mu s \quad (8)$$

- maximum
at end of simulation, inport buffer full, all in to one out

$$l_{inportq,max} = 7031 \quad (9)$$

$$r_{dequeue,min} = 250 \text{ Mbps} \quad (10)$$

$$t_{in \text{ queue, max}} = \frac{t_{q,inport} \cdot s}{r_{dequeue}} = 3.599 \text{ ms} \quad (11)$$

$$(12)$$

- on avg

$$t_{e2e, avg} = t_{in \text{ queue, max}} / 2 = 1.8 \text{ ms} \quad (13)$$

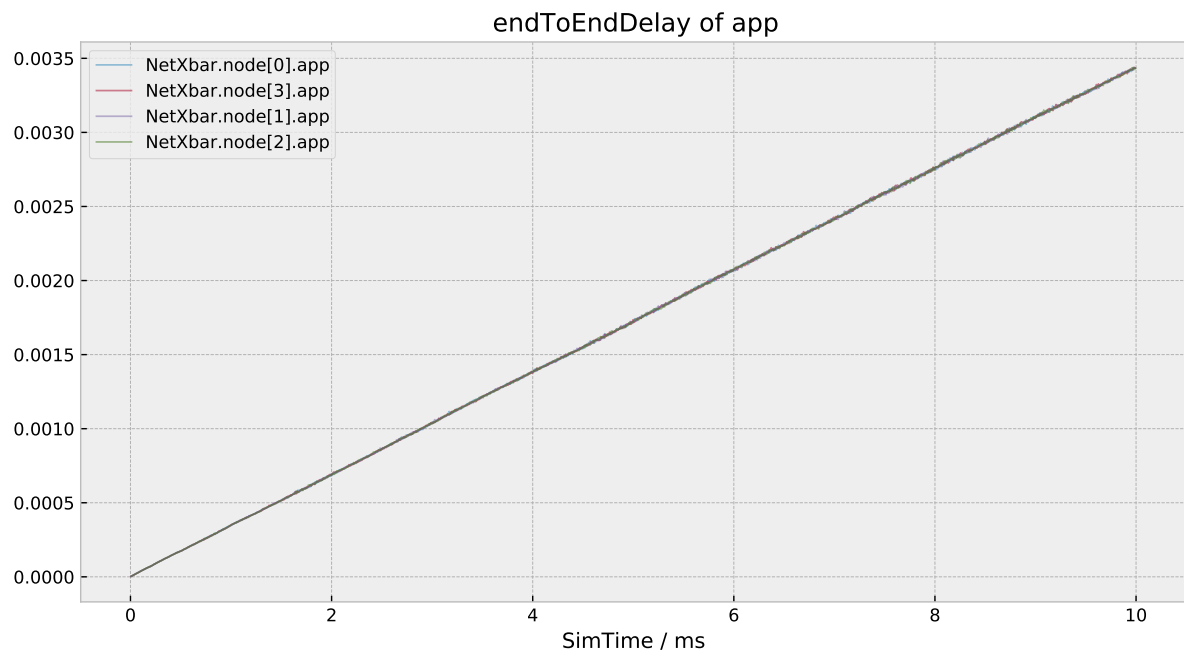


Abbildung 2: Pre Optimization e2e latency

1.1.3 avg Arbiter Request Queue Length

- cases

$$C_n = n - 1 \quad 1 \leq n \leq 4 = N \quad (14)$$

- on avg (analogously to t_{e2e})

$$\Rightarrow l_{arbq,avg} = \sum_{n=1}^4 \frac{\binom{4}{n} (5-n)! \cdot C_n}{141} = 0.39 \quad (15)$$

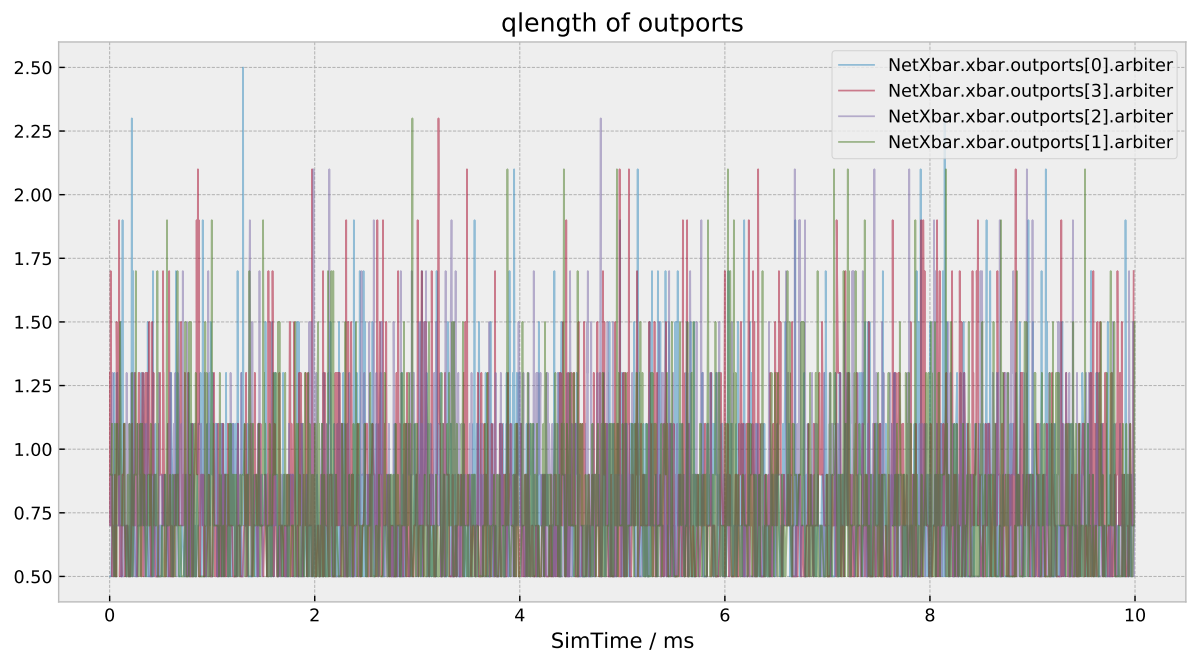


Abbildung 3: Pre Optimization Queue Length

1.1.4 avg Arbiter Request Queue Time

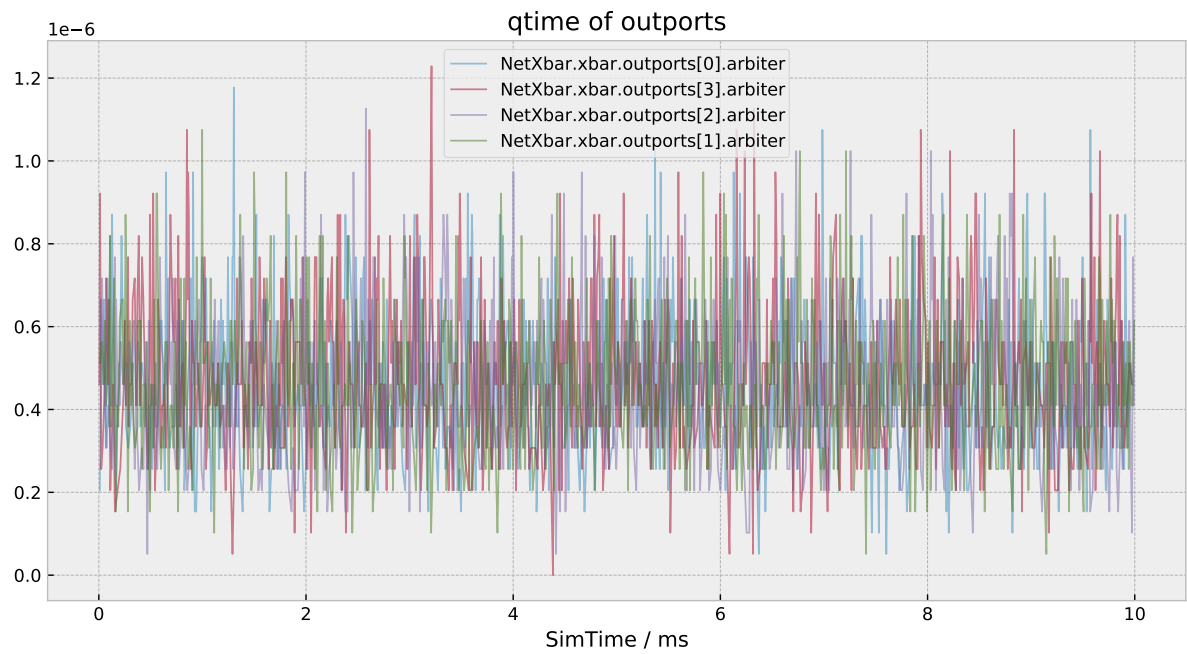


Abbildung 4: Pre Optimization Arbiter Queue Times

- cases

$$C_n = 512 \text{ ns} * n \quad (16)$$

- on avg

$$\Rightarrow t_{arbq,avg} = \sum_{n=1}^4 \frac{\binom{4}{n} (5-n)! \cdot C_n}{141} \quad (17)$$

$$= \quad (18)$$

1.1.5 avg Output Buffer Queue Length

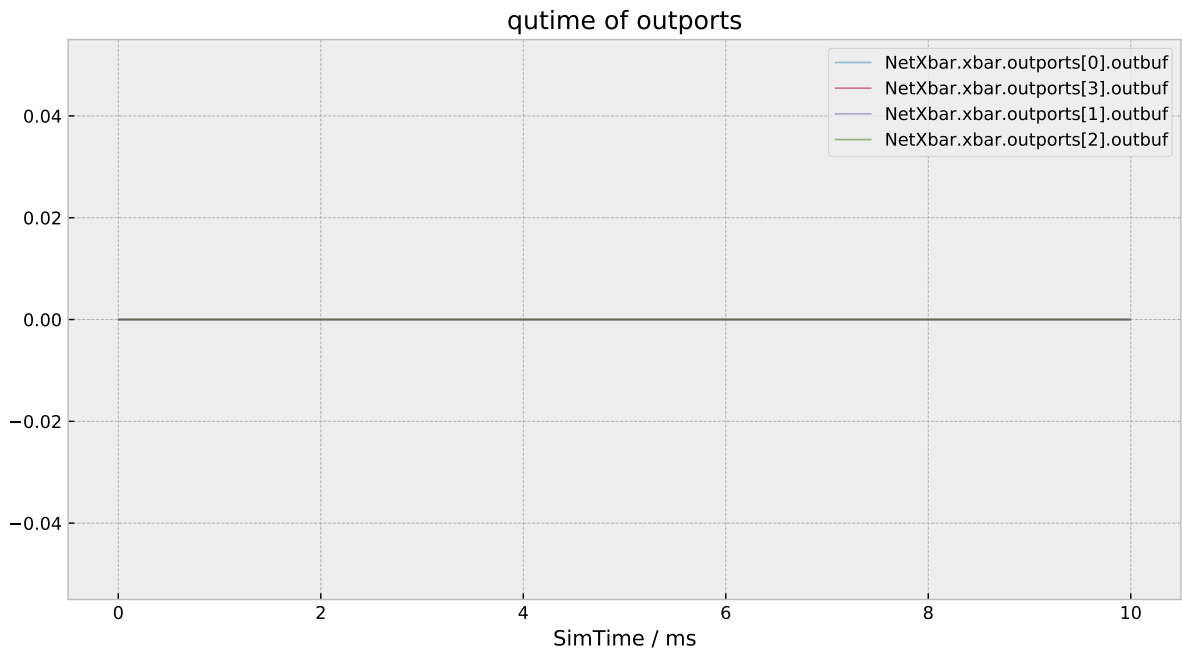


Abbildung 5: Pre Optimization Output buffer Queue Time

- Generated Packets
 - size: $s = 512 \text{ b}$
 - send interval: $t = \text{uniform}(1 \mu\text{s}, 10 \mu\text{s}) = 5.5 \mu\text{s}$
- Connection to and in XBar
 - data rate: 1 Gbps

$$r_{\text{generated}} = \frac{512 \text{ b}}{5.5 \mu\text{s}} = 93 \text{ Mbps} \quad (19)$$

$$\Rightarrow l_{q,avg} = 0 \quad (20)$$

1.1.6 avg Throughput

$$r_{cross} = 4 \cdot (r_{gen} - r_{q \text{ fill, eff}}) \quad (21)$$

$$= 4 \cdot (1000 \text{ Mbps} - 372 \text{ Mbps}) \quad (22)$$

$$= 2512 \text{ Mbps} \quad (23)$$

1.2 Throughput vs. Ports

# Nodes	$r_{q \text{ fill, eff}}$	$r_{q \text{ fill, sim}}$
2	600 Mbps	Mbps
4	347 Mbps	Mbps
8	167 Mbps	Mbps
16	90 Mbps	Mbps
32	48 Mbps	Mbps

1.3 Throughput vs. Injection Rate

- not in saturation until delay $\leq 832 \text{ ns}$
- saturation point $r_{sat} = \frac{512 \text{ b}}{832 \text{ ns}} = 615 \text{ Mbps}$
- The main reason will be the arbiter. It can only process packets at about 600 Mbps on avg., therefore bottlenecking the rest of the system

1.4 Throughput vs. Bandwidth

- Network in saturation until bandwidth $> 1600 \text{ Mbps}$
- throughput at saturation point $r_{sat} = 1 \text{ Gbps}$
- If the Arbiter can only work at $\approx 62\%$ of the bandwidth, we reach this saturation point if 62% of the bandwidth is 1 Gbps, therefore the required bandwidth is 1613 Gbps

2 Optimizations

2.1 Problems

In the previous exercise we determined the arbiter to be the limiting resource. We determined head of line blocking to be a limiting factor for performance. Since packets are sent to random nodes it is very likely for a packet having to wait because the route is currently occupied even though other packets in queue could be routed. Therefore virtual queues were implemented.

2.2 Solution

We can measure the performance gain in correlation to the average inport buffer length which is decreased by two orders of magnitude. Form around 3000 to around 15

2.3 Problems In Real Hardware

In a real hardware setting this would increase complexity. Either distinct buffers have to be used or some sort of control logic that holds pointers to the virtual queues. Spatial constraints have to be taken into account as well as scalability. With more possible routing decisions the number of virtual queues also rises.

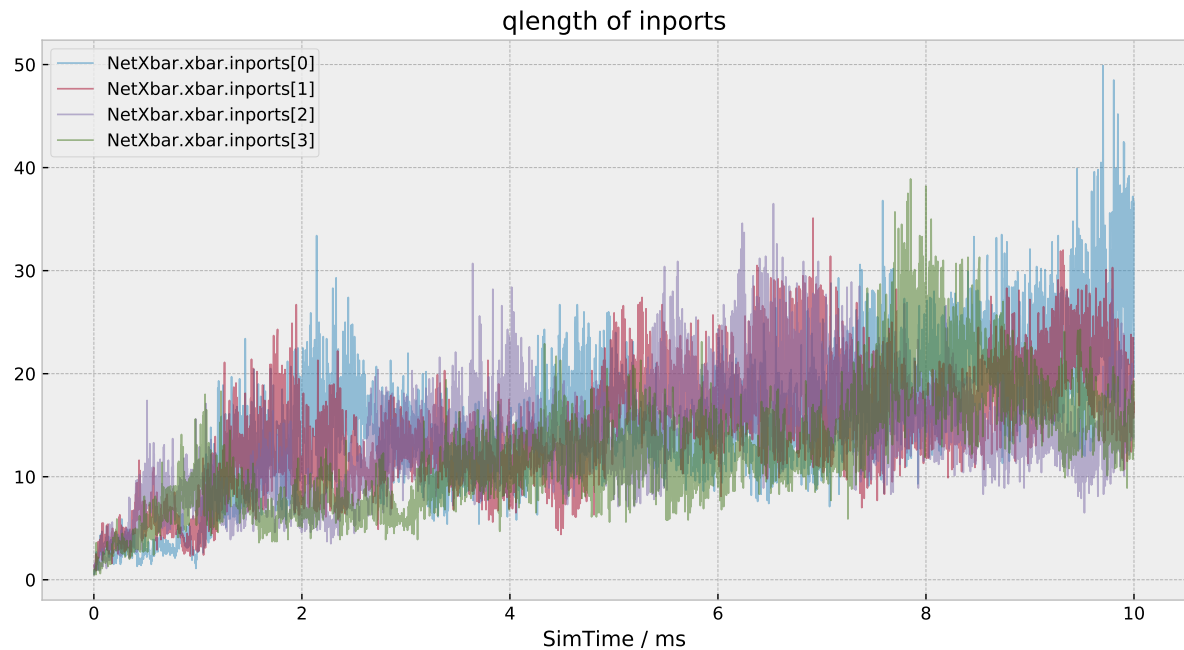


Abbildung 6: sim

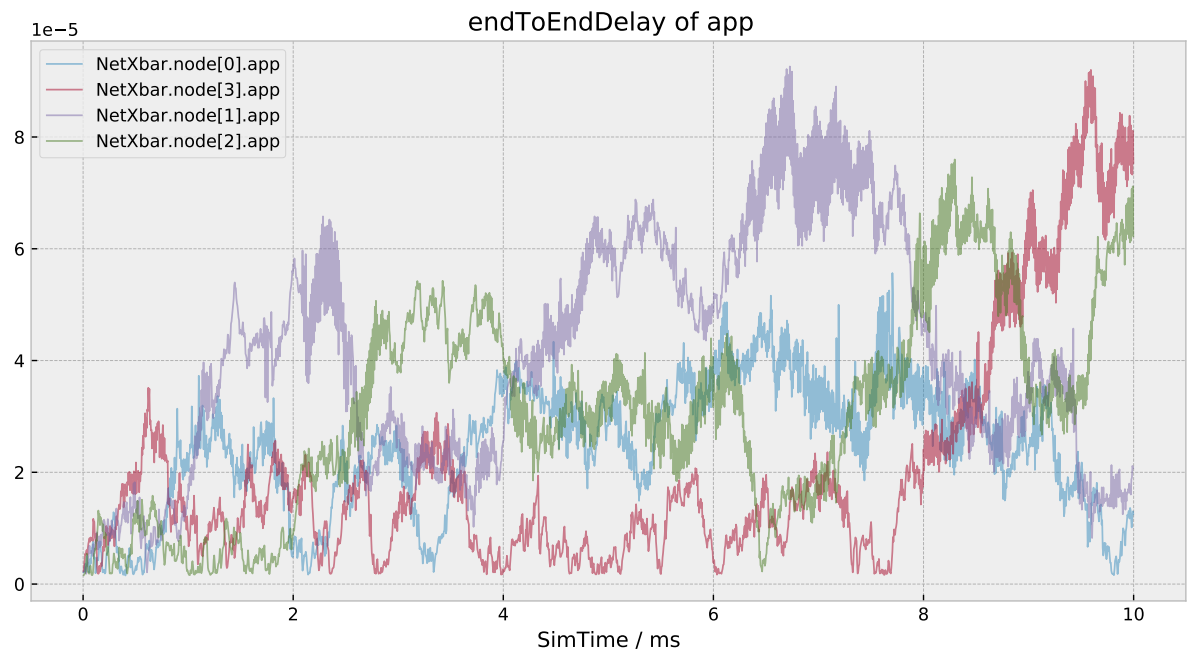


Abbildung 7

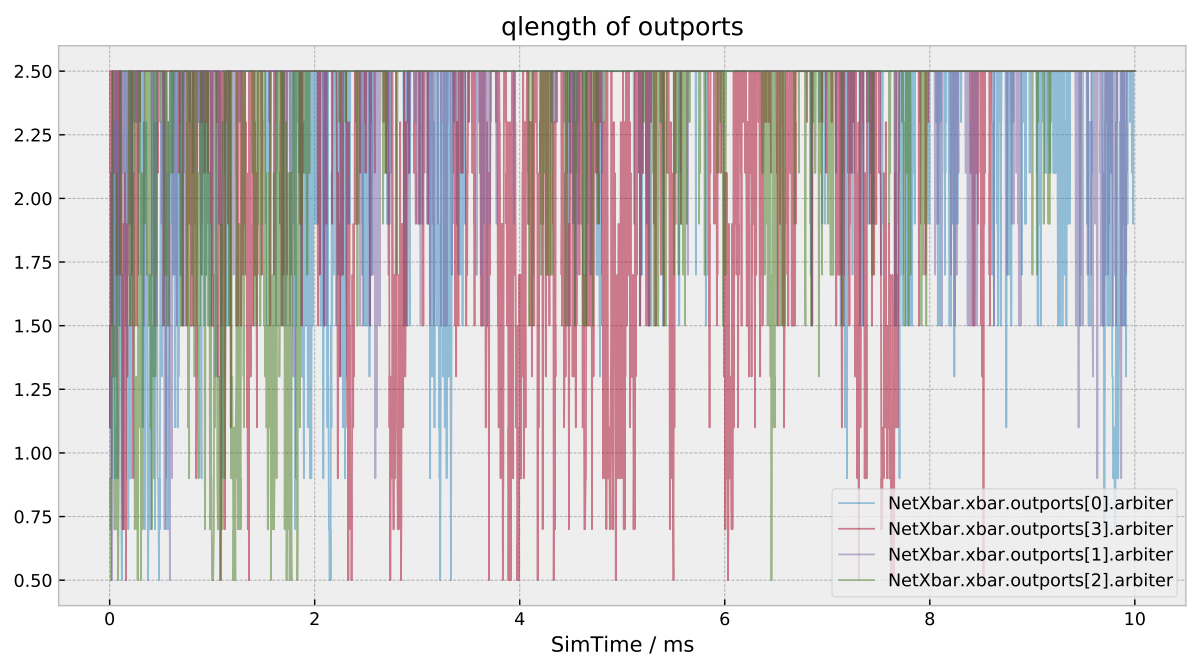


Abbildung 8

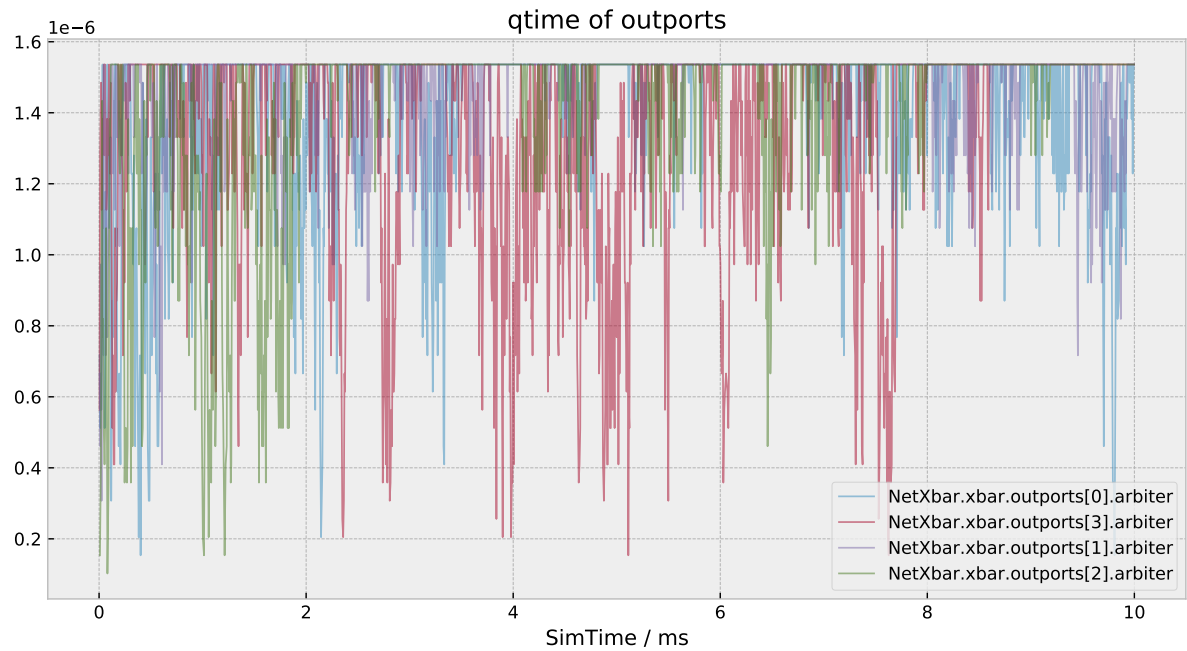


Abbildung 9

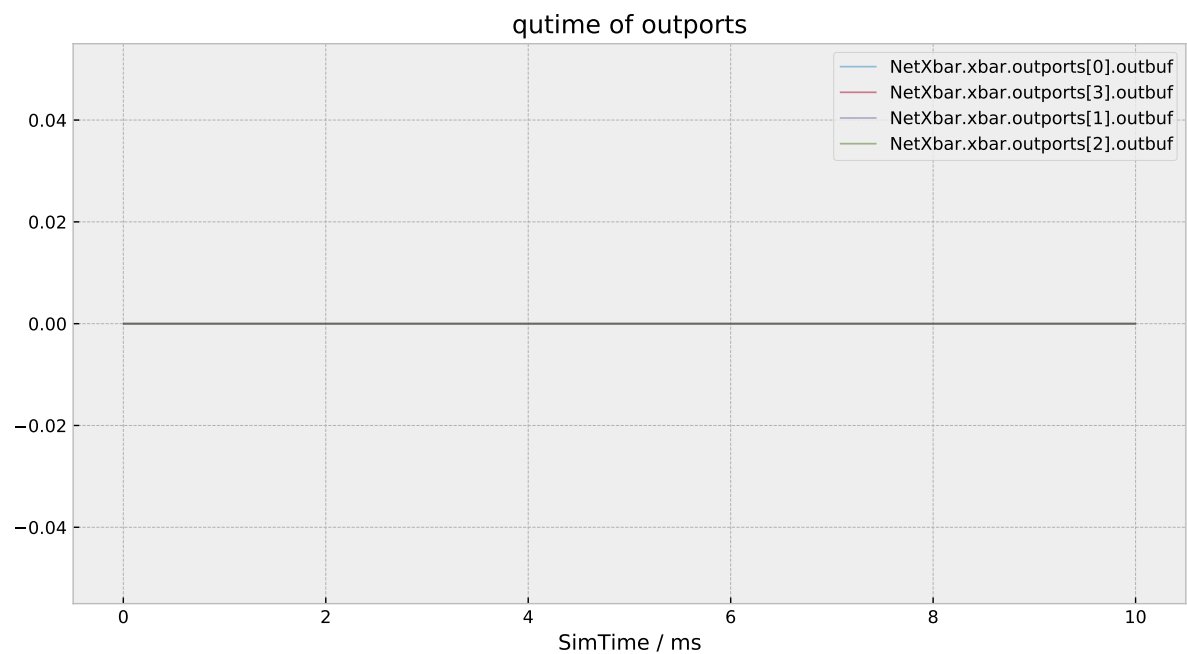


Abbildung 10