

Latency considerations from an automotive camera system perspective

IEEE 802.3dm

July 15, 2024

Kirsten Matheus, BMW Group

Supporters

- Ahmad Chini (Broadcom)
- Claude Gauthier (NXP)
- Debajyoti Pal (Onsemi)
- Dongok Kim (Hyundai)
- Frank Wang (Realtek)
- Heiko Strohmeier (BOSCH)
- Hideki Goto (Toyota)
- Kevin So (Microchip)
- Scott Muma (Microchip)
- Steve Gorshe (Microchip)
- Steve Kang (Microchip)
- Yasuhiro Kotani (Denso)
- Yoshifumi Kaku (Denso)

Motivation

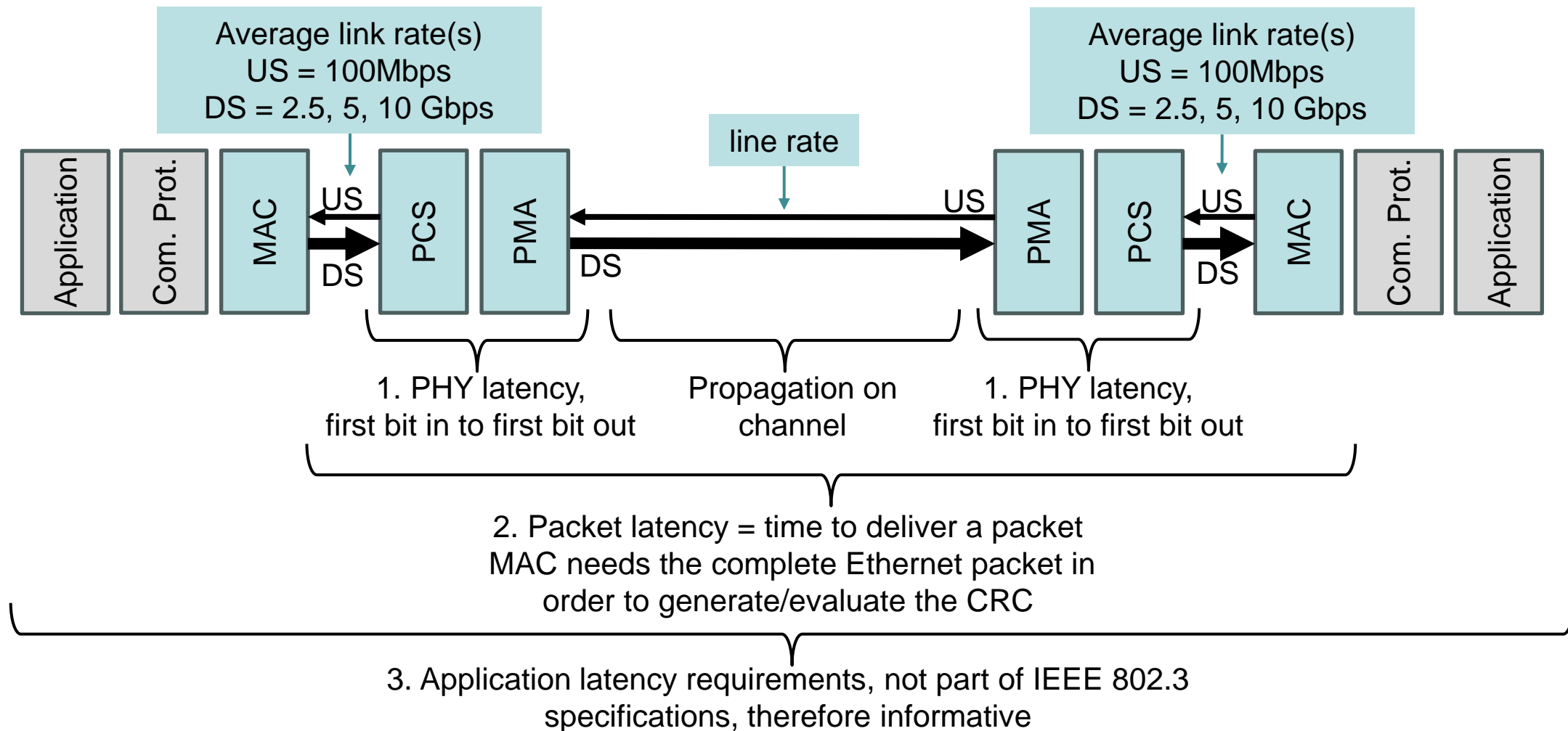
- Page 16 of https://www.ieee802.org/3/dm/public/0524/sedarat_3dm_02_202405.pdf gives the impression that a TDD system is at a significant disadvantage in respect to latency and FIFO when compared with an FDD or an FDD/CM system.
- This presentation looks at latency from an automotive camera system perspective in order to investigate this statement in detail.
- It provides concrete examples of latency and latency requirements in a camera system.
- It addresses the different parameters that impact the system latency of a transmission technology.
- It shows no latency disadvantage for TDD in case of automotive camera systems.

Content

- System overview
- Latency aspects to consider
 1. Latency requirements from camera application perspective (informative)
 2. Packet latency due to data rate and duplexing scheme
 3. Interrelation between packet latency and power consumption
 4. PHY latency
- Summary and conclusion

System overview

US = Upstream, low data rate direction
DS = Downstream, high data rate direction



Ethernet system latency

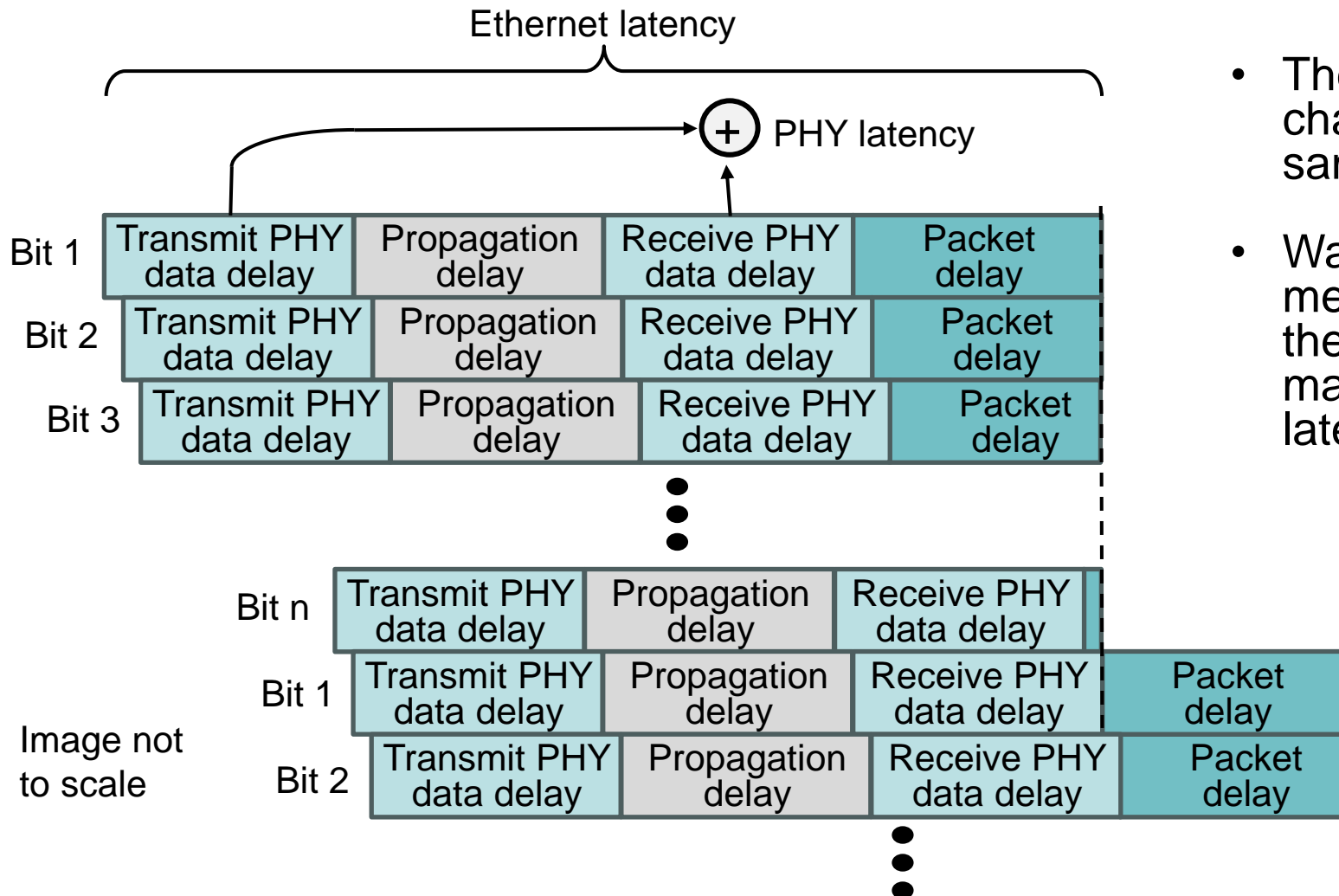


Image not to scale

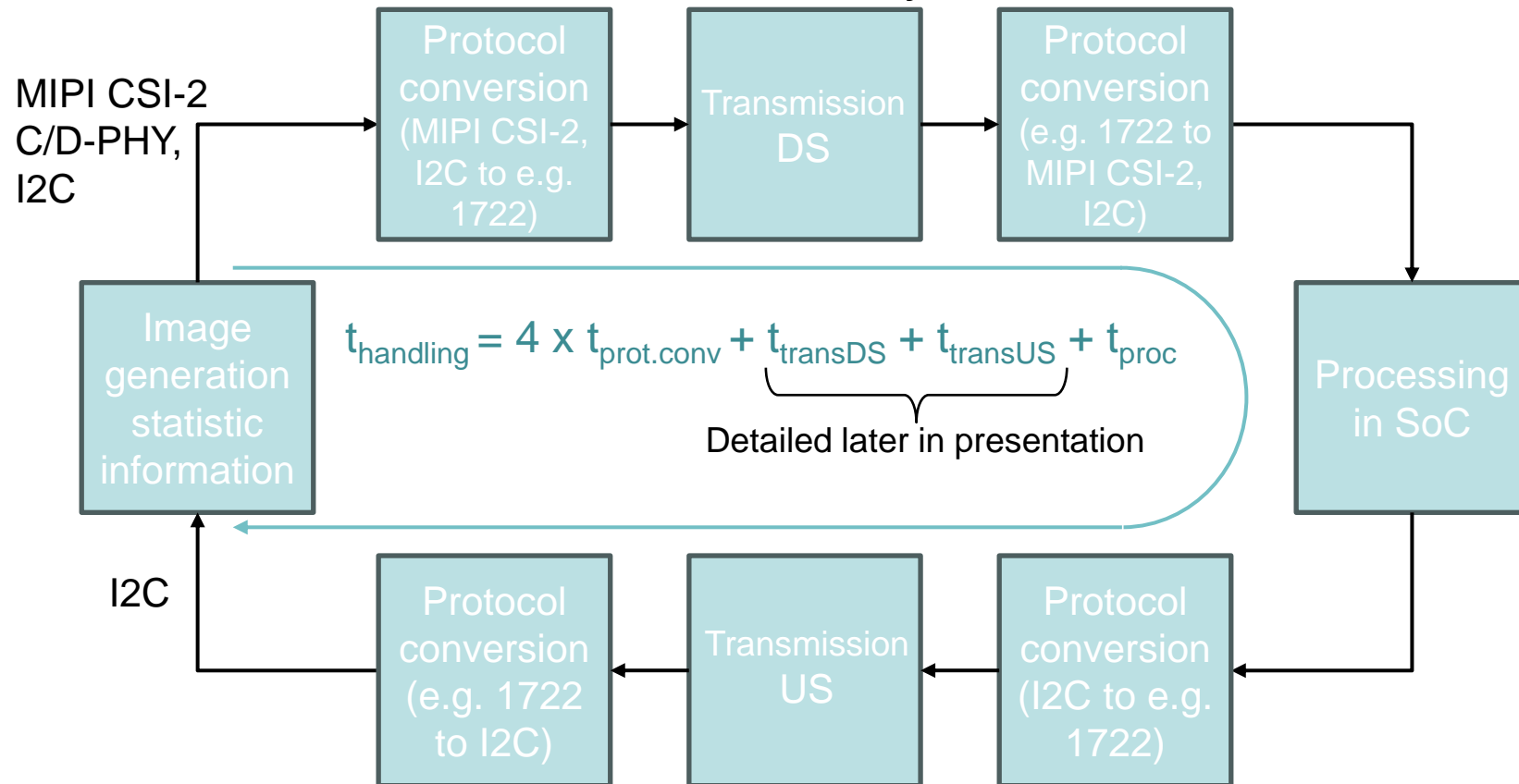
- The propagation delay on the channel is small (<90ns) and the same for all duplexing options.
- Wait times caused by the duplexing method are typically attributed to the transmit PHY delay. However, it may also be attributed to the packet latency or considered separately.

Content

- System overview
- Latency aspects to consider
 1. Latency requirements from camera application perspective (informative)
 2. Packet latency due to data rate and duplexing scheme
 3. Interrelation between latency and power consumption
 4. PHY latency
- Summary and conclusion

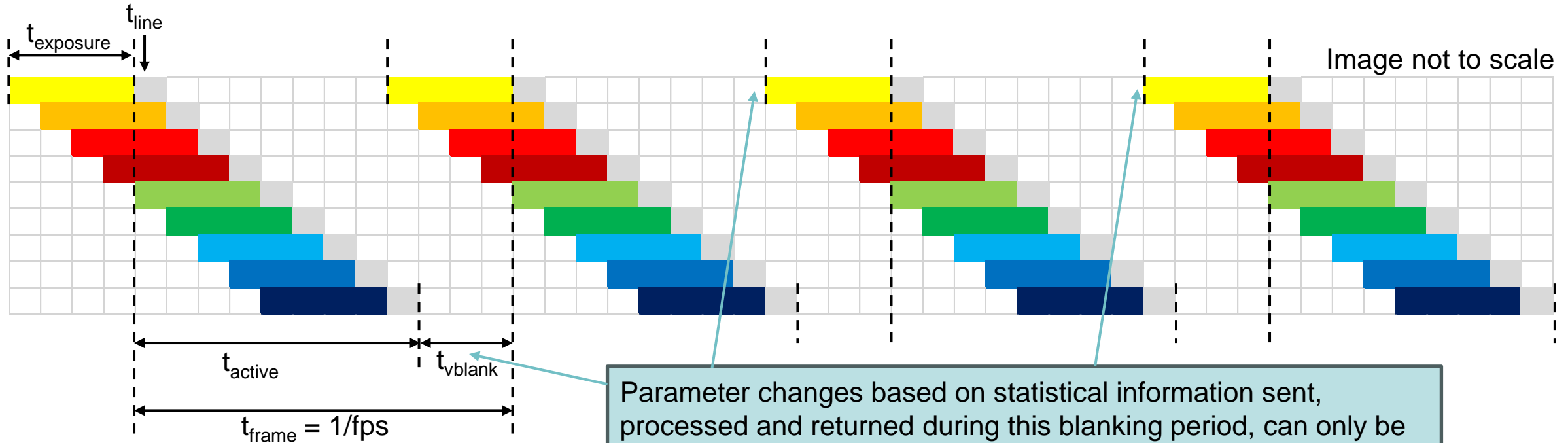
Camera application latencies

Automotive camera communication typically distinguishes between uni-directional video data and bi-directional control data communication. US latencies are relevant for control data only.



*) Global shutter takes the complete image at once, stores it and then line-wise transfers it. More costly and therefore less common.

Rolling*) shutter imager timing example



Parameter changes based on statistical information sent, processed and returned during this blanking period, can only be applied to the image capture after next. If handling takes longer, the new parameters get applied to yet one image later.

Example:

Frame rate = 30 fps $\rightarrow t_{\text{frame}} = 33.3 \text{ ms}$

Active image = 29 ms (design decision $< t_{\text{frame}}$) means $(33.3-29)/29 = \sim 15\%$ blanking overhead

$t_{\text{line}} = t_{\text{active}}/\# \text{ active lines (imager capability)} = 29 \text{ ms}/2160 = 13.4 \text{ us} \ll \text{any other time in the system}$

$t_{\text{exposure}} = \text{variable (but } < t_{\text{frame}} \text{)} ! 10 \text{ ms is a typical upper value for 8Mpx}$

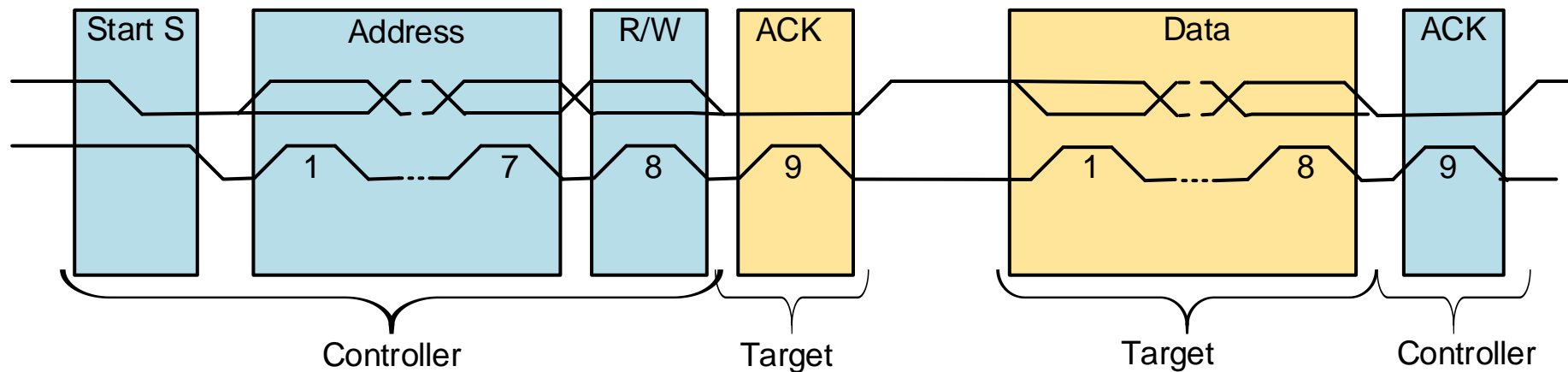
$$t_{\text{vblank}} = 33.3 \text{ ms} - 29 \text{ ms} = 4.33 \text{ ms} > t_{\text{handling1}}, t_{\text{handling2}} < 37.66 \text{ ms} = 4.33 \text{ ms} + 33.3 \text{ ms}$$

\rightarrow Ethernet latencies of 10us in the DS and of 100us in the US are sufficiently small for t_{handling} .

Implications of I2C

In automotive camera applications, a common protocol for control data is I2C, which follows a controller/target communication concept. The controller typically sits in the SoC, the camera is the target. The basic I2C format requests acknowledgements received for every byte sent within a certain time. This time determines the latency requirements for the I2C control traffic. However, I2C timing is typically not critical, because

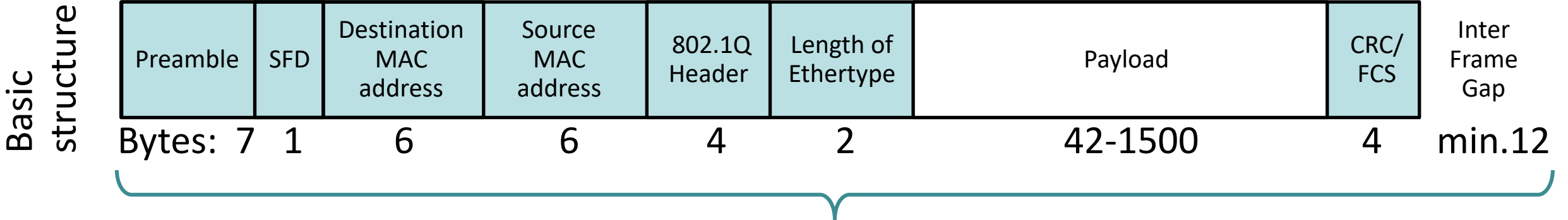
- a) the typical rate for I2C in camera applications is 100 kbps, 400 kbps, or 1 Mbps
- b) I2C allows for clock stretching to accommodate delays in ACK responses
- c) a number of I2C commands may be sent in a bulk, as may the ACKs.



Content

- System overview
- **Latency aspects to consider**
 1. Latency requirements from camera application perspective (informative)
 2. **Packet latency due to data rate and duplexing scheme**
 3. Interrelation between latency and power consumption
 4. PHY and coding latency
- Summary and conclusion

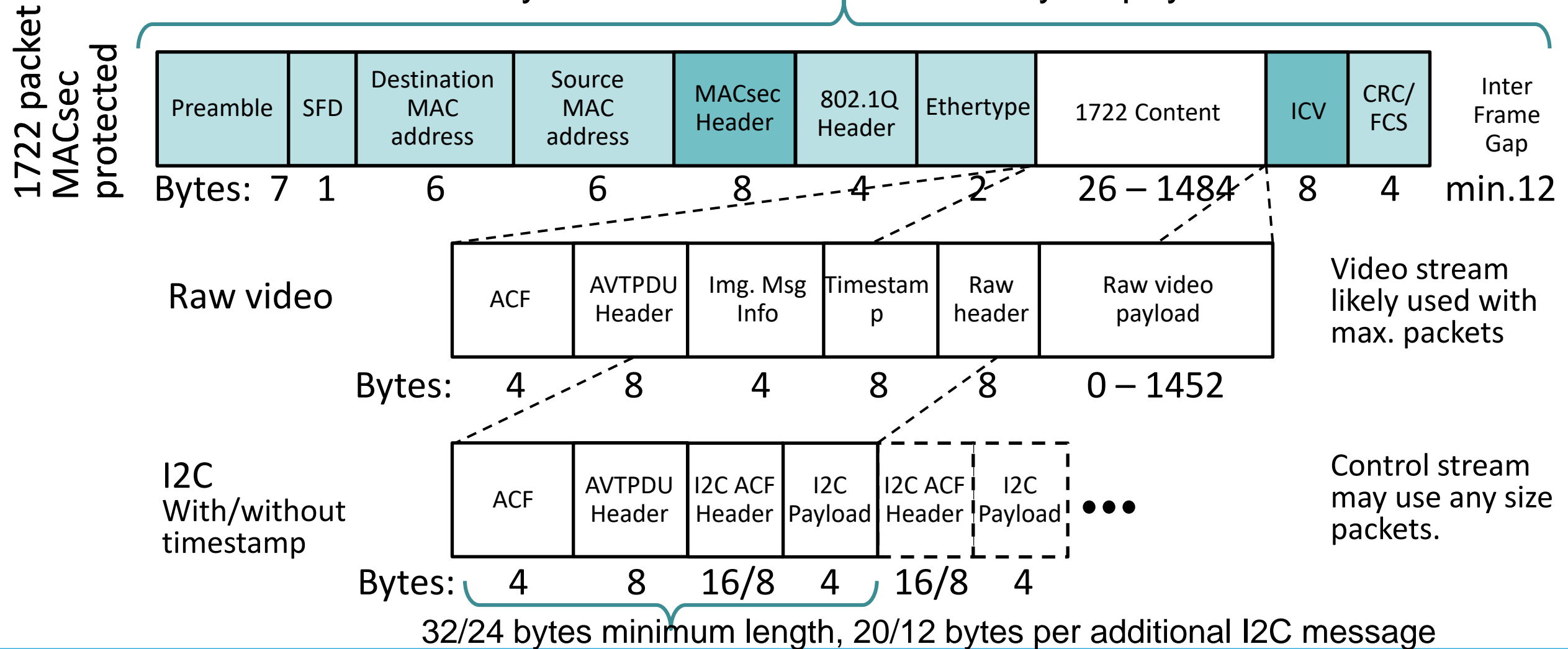
Ethernet frame and packet



42 bytes overhead and 42-1500 bytes payload
= 84 bytes min to 1542 bytes max.

Example frames for automotive camera use

58 bytes overhead and 26-1484 bytes payload



Packet latency per Eth. frame depending on link rate

100 Mbps

$$84 \times 8 / (100 \times 10^6) \text{ s} = 6.72 \text{ us to } 1542 \times 8 / (100 \times 10^6) \text{ s} = 123.36 \text{ us}$$

1 Gbps

$$84 \times 8 / (1000 \times 10^6) \text{ s} = 0.672 \text{ us to } 1542 \times 8 / (1000 \times 10^6) \text{ s} = 12.336 \text{ us}$$

2.5 Gbps

$$84 \times 8 / (2500 \times 10^6) \text{ s} = 0.268 \text{ us to } 1542 \times 8 / (2500 \times 10^6) \text{ s} = 4.9344 \text{ us}$$

5 Gbps

$$84 \times 8 / (5000 \times 10^6) \text{ s} = 0.134 \text{ us to } 1542 \times 8 / (5000 \times 10^6) \text{ s} = 2.4672 \text{ us}$$

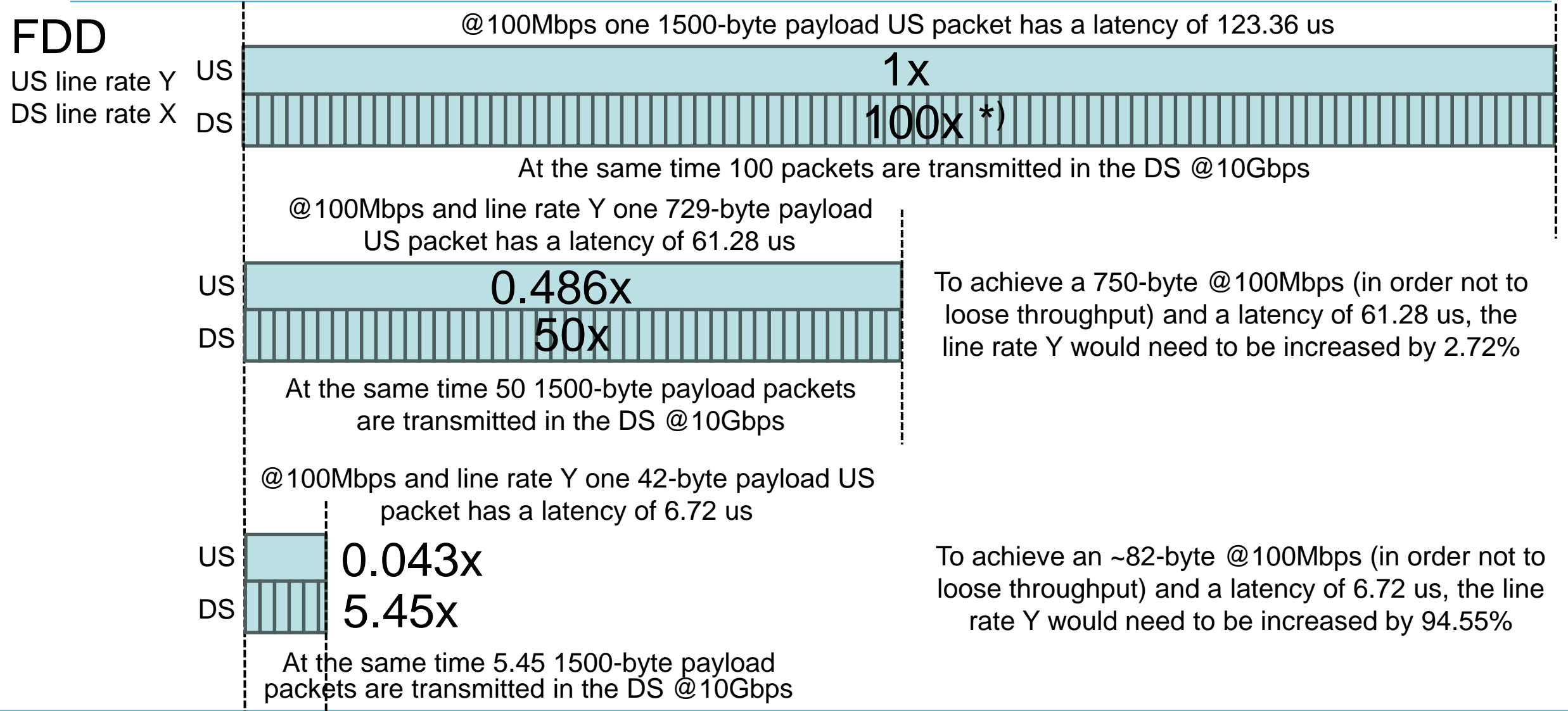
10 Gbps

$$84 \times 8 / (10000 \times 10^6) \text{ s} = 0.0672 \text{ us to } 1542 \times 8 / (10000 \times 10^6) \text{ s} = 1.2336 \text{ us}$$

For lower link rates, the packet latency may be significantly larger than single digit us or even tens of us.
For higher link rates, it is in the single us range or even significantly smaller.

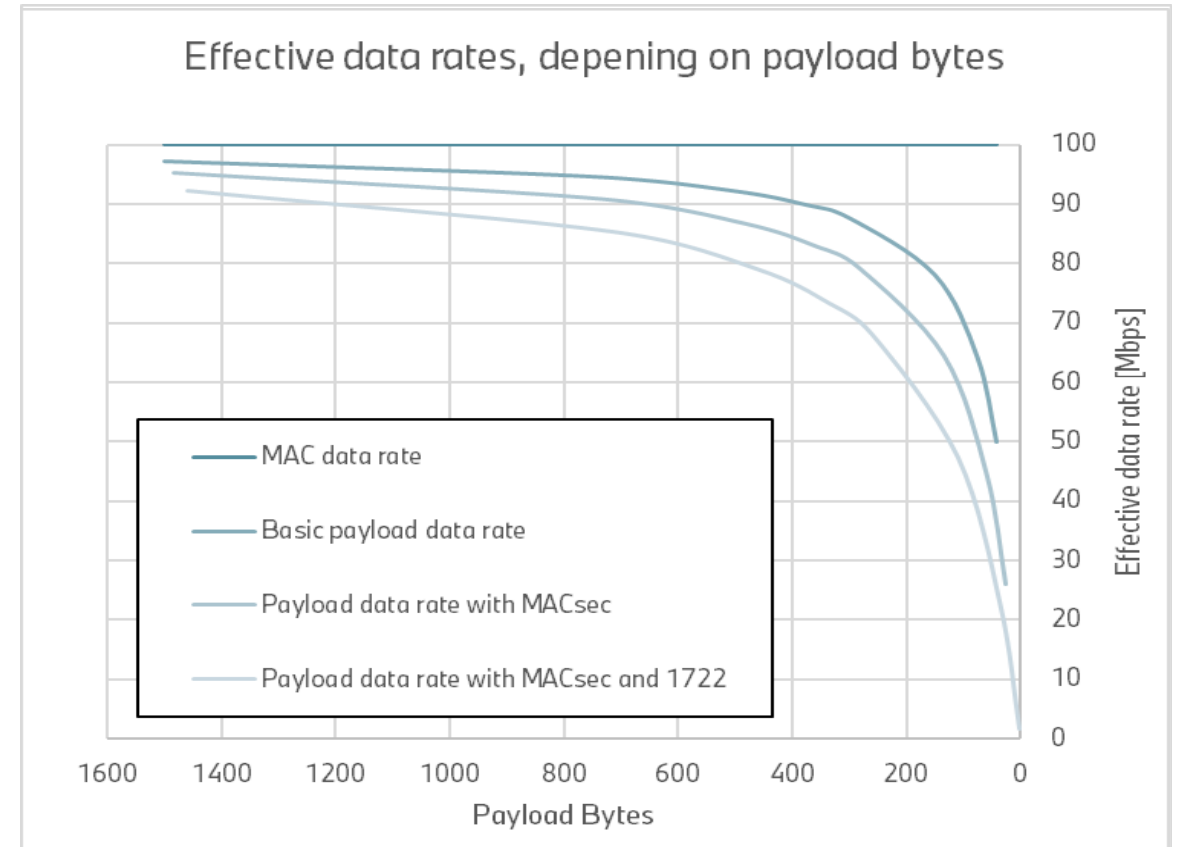
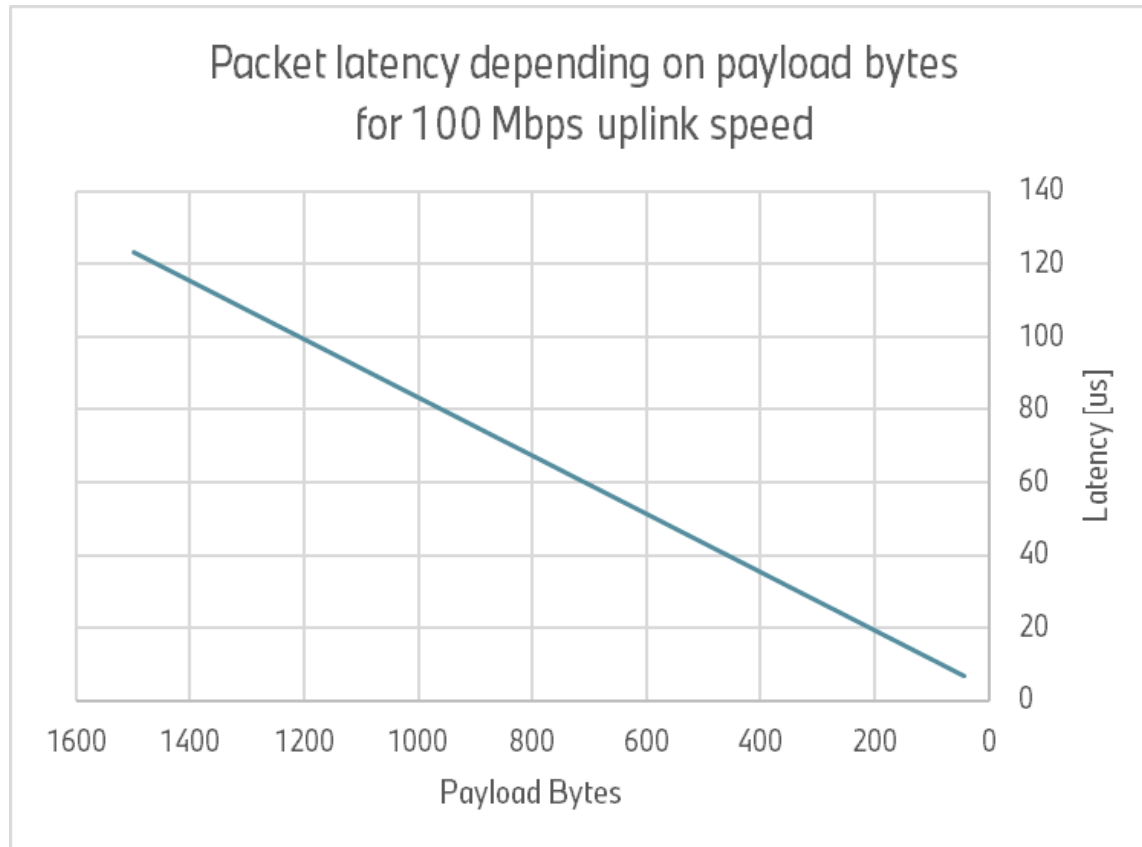
Packet latency for FDD and its reduction (1)

*) 50/25 packets in case of 5 and 2.5 Gbps respectively



Packet latency for FDD and its reduction (2)

The FDD packet latency can be reduced by shortening the UL Ethernet packets. Shortening the Ethernet packets reduces the available payload bytes (throughput) over-proportionally as more overhead data is being transmitted. If this was to be counter-measured, it would require increasing the (line) rate.



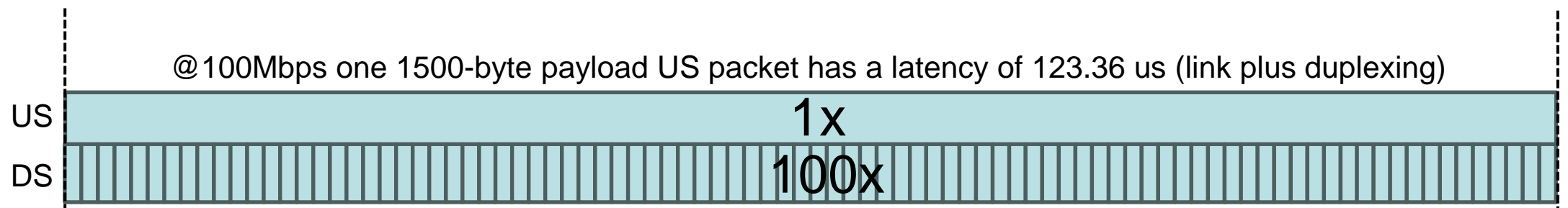
**) assuming the same processing blocks/overhead in the FDD and TDD DS.

Packet and duplexing latency

*) assuming a propagation delay of ~90ns plus an additional header of 250 bytes = ~0.2us for switching header between US and DS

FDD

US line rate Y
DS line rate X

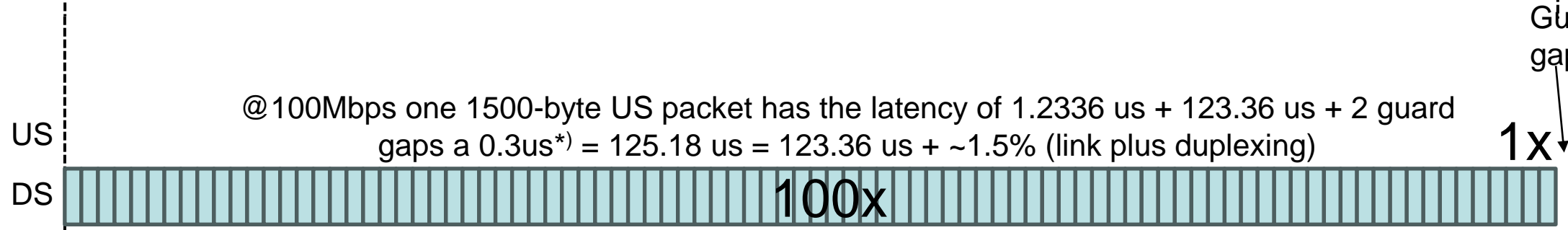


@100Mbps one 1500-byte payload US packet has a latency of 123.36 us (link plus duplexing)

At the same time 100 packets are transmitted in the DS @10Gbps

TDD

DS=US
line rate X**)

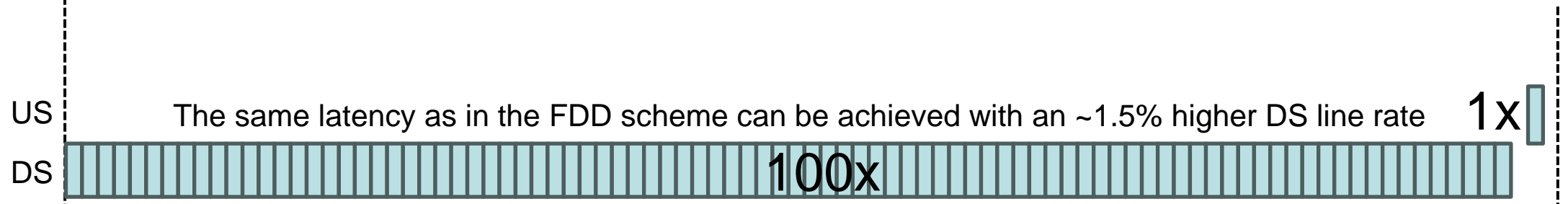


@100Mbps one 1500-byte US packet has the latency of 1.2336 us + 123.36 us + 2 guard gaps a 0.3us*) = 125.18 us = 123.36 us + ~1.5% (link plus duplexing)

At the same line rate 100 packets are transmitted

TDD

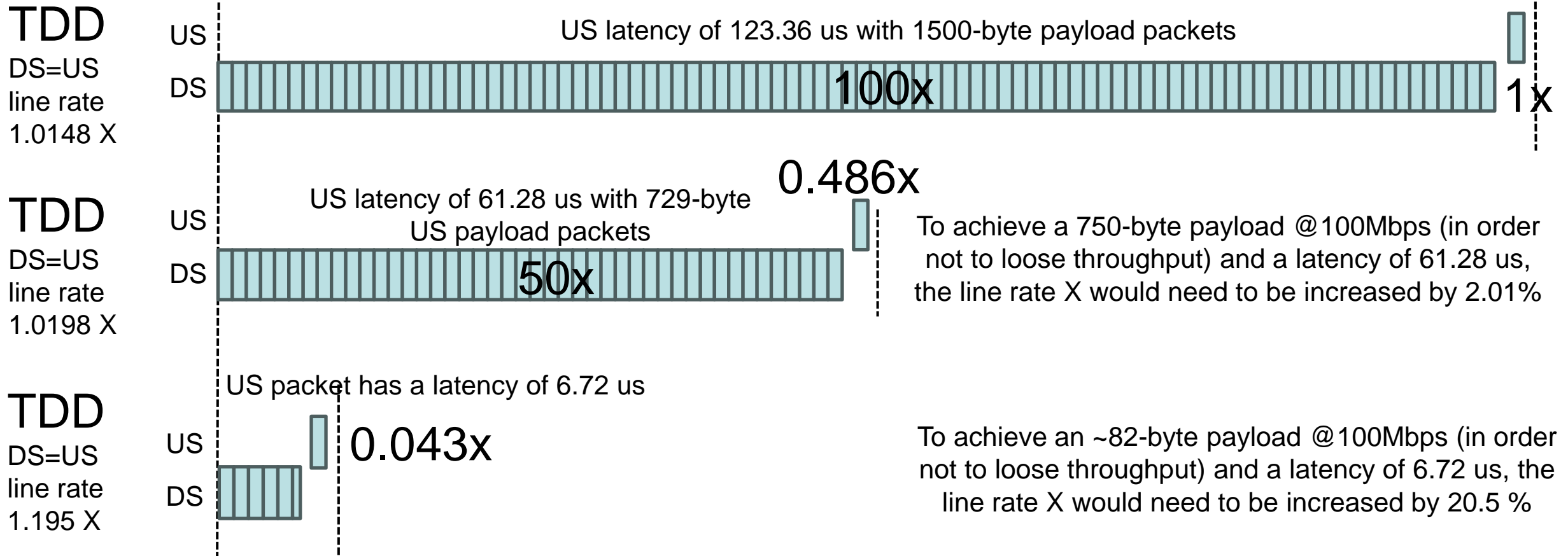
DS=US
line rate
1.0148 X



The same latency as in the FDD scheme can be achieved with an ~1.5% higher DS line rate

Shortening the latency in a TDD system

The link and duplexing latency can be also be reduced by shortening the US Ethernet packets and sending fewer packets in the DS burst. Using the same reduced payload length as for the FDD example, the line rate further needs to increase to achieve the same throughput because of the added guard gaps and switching headers.



Relative comparison of latency reductions and rate increases for the FDD and TDD examples

**) Below the minimum payload size*

US latency [us]	Stable FDD US rate				Stable payload throughput			
	Payload bytes	Throughput reduction	FDD US rate increase	TDD DS=US rate increase	Payload bytes	Throughput reduction	FDD US rate increase	TDD DS=US rate increase
123.36	1500	0%	0%	1.49%	1500	0%	0%	1.49%
61.68	729	2.8%	0%	1.98%	750	0%	2.72%	2.01%
41.12	472	5.6%	0%	2.47%	500	0%	5.45%	2.52%
30.84	343.5	8.4%	0%	2.96%	375	0%	8.17%	3.05%
24.672	266.4	11.2%	0%	3.46%	300	0%	10.89%	3.57%
12.336	112.2	25.2%	0%	5.96%	150	0%	24.51%	6.21%
6.168	35.1 ^{*)}	53.2%	0%	11.09%	75	0%	51.75%	11.62%
3.454	1.176 ^{*)}	97.2%	0%	19.5%	42	0%	94.55%	20.5%

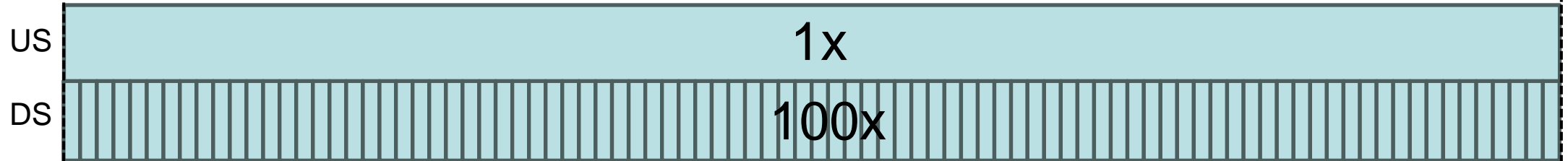
Because the overhead is transmitted slower in the US of an FDD than of a TDD system, compensating for it causes relatively higher rate increases. The situation becomes worse if other overheads (MACsec, 1722, ...) are considered. Naturally, the absolute FDD US transmission rate stays smaller than the one for a TDD US.

Principle comparison with FDX/EEE

@100Mbps one 1500-byte payload US packet has a latency of 123.36 us (packet plus duplexing)

FDD

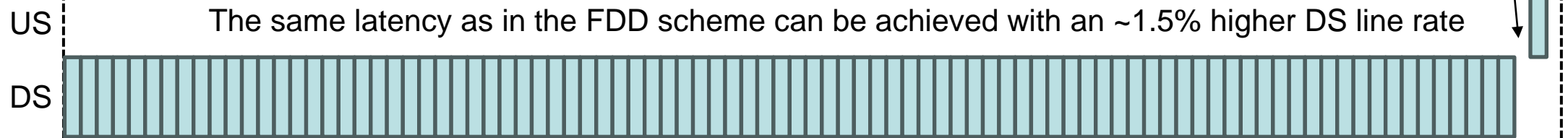
US line rate Y
DS line rate X



At the same time 100 packets are transmitted in the DS @10Gbps

TDD

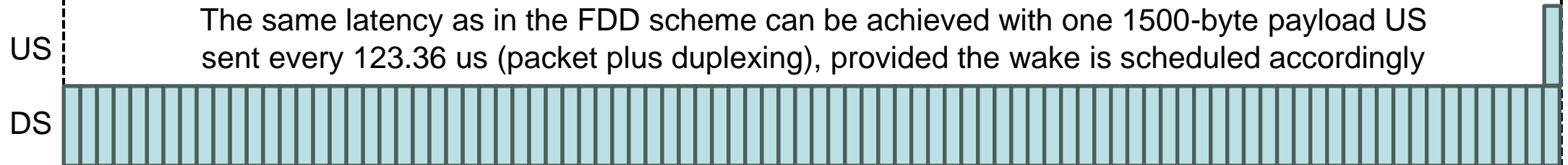
DS=US
line rate
1.0148 X



At the same time 100 packets are transmitted in the DS @10Gbps

FDX/EEE

DS=US
line rate X



At the same time 100 packets are transmitted in the DS @10Gbps

Shortening the latency in a FDX/EEE system

The goal of EEE in a full duplex (FDX) system is to save power in the US direction, which per 802.3dm objectives needs to be available at an average 100 Mbps link rate (i.e. 25 – 100 times slower than DS).

The most power is saved when a 1500-byte payload US packet is sent every 123.36us. Shortening the packet reduces the latency but also the power that can be saved.

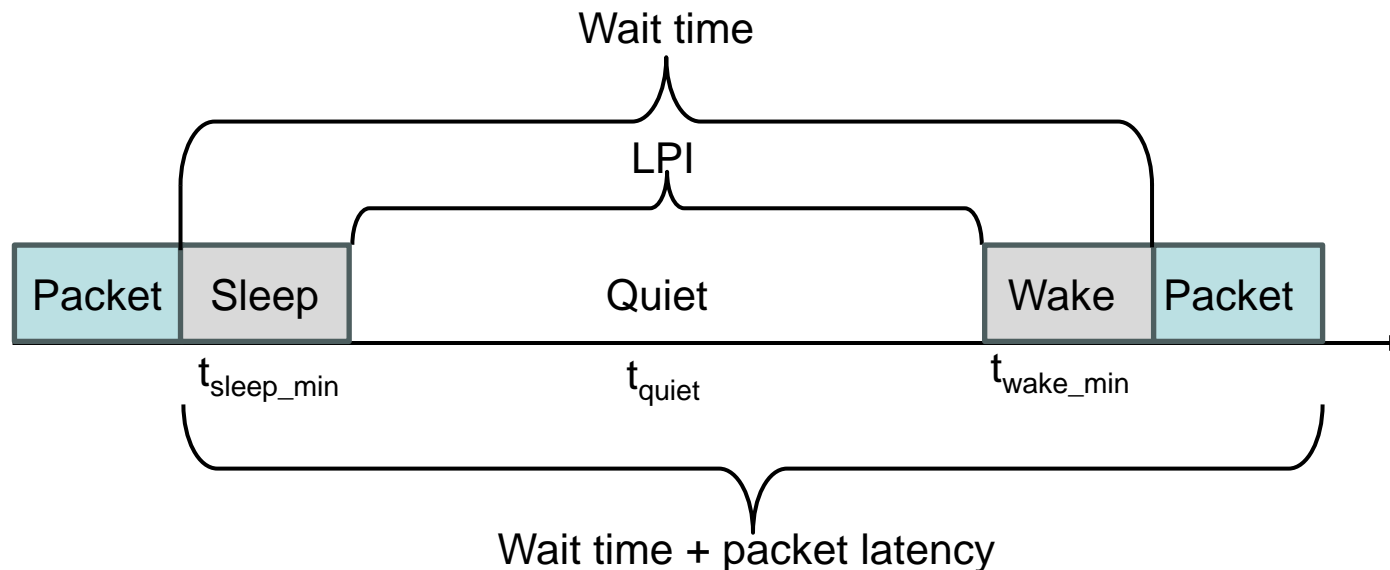
The assumption is that for EEE to effectively save power the link should be quiet at least 50% of the energy saving period. This determines the minimum US packet length.

Other than in a TDD system, EEE as is, is not scheduled. Normally, the link is activated, when content is available. It would need to be organized on higher layers, that the respective amount of data is aggregated before being transmitted.

Wait time and packet length in 802.3ch/EEE

*)Table 78-2 **)Table 78-4

	$t_{\text{sleep_min}}$ *)	$t_{\text{wake_min}}$ **)	Min. wait time for $t_{\text{quiet}} \geq t_{\text{sleep_min}} + t_{\text{wake_min}}$	No. payload bytes for 100Mbps link rate	Through- put	Wait time + packet latency
2.5GBASE-T1	10.24 us	25.6 us	71.68 us	~891 bytes	95,5%	74.67 us
5GBASE-T1	5.12 us	12.8 us	35.84 us	~415 bytes	90,8%	36.57 us
10GBASE-T1	2.56 us	6.4 us	17.92 us	~184 bytes	81,44%	18.1 us



Note in addition: 802.3ch works in interleaving and RS-FEC blocks of $L \times 3600$ bits = $L \times 450$ bytes. This might lead to additional quantification impacts (or wasted bandwidth for idle transmission).

Content

- System overview
- Latency aspects to consider
 1. Latency requirements from camera application perspective (informative)
 2. PHY and coding latency
 3. **Interrelation between latency and power consumption**
 4. Packet latency due to data rate and duplexing scheme
- Summary and conclusion

Impact for reduced DS data rates (1)

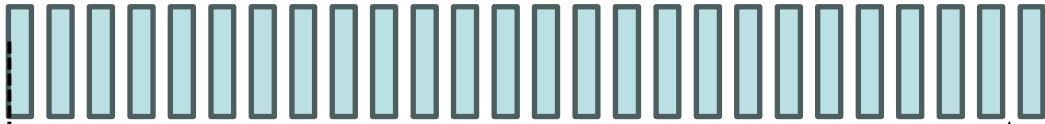
In a camera system, power saving is essential. This applies to the low-speed US data, but also to the high-speed DS traffic.

In case a significantly lower DS throughput is needed than the available link rate would support, it would be desirable to save power also DS.

When applying EEE for the DS the same assumption applies as in the FDX/EEE US case: The link should be quiet 50% of the power saving time. Also there is a minimum quiet time, which needs to be observed.

Example: DS link rate 10 Gbps, needed 60%

Option 1: DS packets with 1500-byte payload are evenly distributed, packet duration = 1.2336 us



Option 2: DS packets with 1500-byte payload are sent in minimum bulk length, i.e. $\text{gap} = 2 \times (t_{\text{sleep_min}} + t_{\text{wake_min}})$



Option 3: DS packets with 1500-byte payload are sent with 100 Mbps wait time equivalent



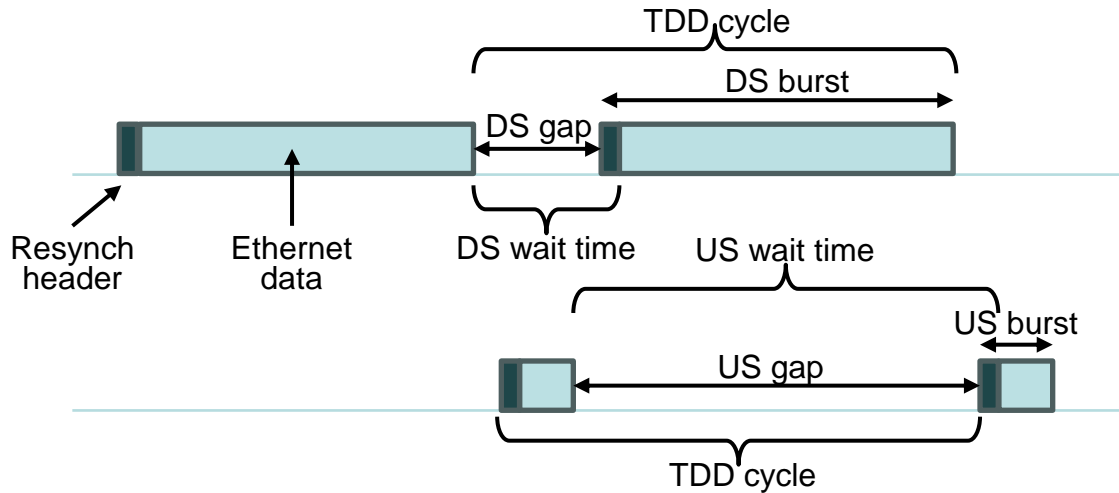
	Option 1	Option 2	Option 3
Wait time	0.74 us	17.92 us	122.13 us
No. packets in bulk	1	21.8	148.5
Latency	1.97 us	19.15 us	123.36 us

LPI not possible
Minimum power saving
minimum extra buffering
minimum extra latency

Maximum power saving
maximum extra buffering
maximum extra latency

In case an EEE-like power saving in the DS is enabled, the realization is virtually independent of duplexing scheme. However, for TDD the mechanisms are inherent. There are no wake and sleep signals in the gap and therefore there is no min. quiet time and more flexibility to realize also smaller gap times and latencies.

Example ASA-MLE wait times and packet latencies



DS link rate	US link rate	Line rate	DS gap	US gap	Resynch header	Wait time plus packet latency DS	Wait time plus packet latency US
2.5 Gbps	100 Mbps	4 Gbps	0.88 us	3.28 us	0.192 us	7.08 us (1542 bytes)*)	3.95 us (150 bytes)
5 Gbps	100 Mbps	8 Gbps	0.64 us	2.56 us	0.192 us	4.44 us (1542 bytes)*)	2.99 us (150 bytes)
10 Gbps	100 Mbps	12 Gbps	0.72 us	26.32 us	0.192 us	2.15 us (1542 bytes)	26.67 us (200 bytes)
10 Gbps	1 Gbps	16 Gbps	0.64 us	2.56 us	0.192 us	2.07 us (1542 bytes)	2.87 us (150 bytes)

*) Transmission spread over two TDD cycles

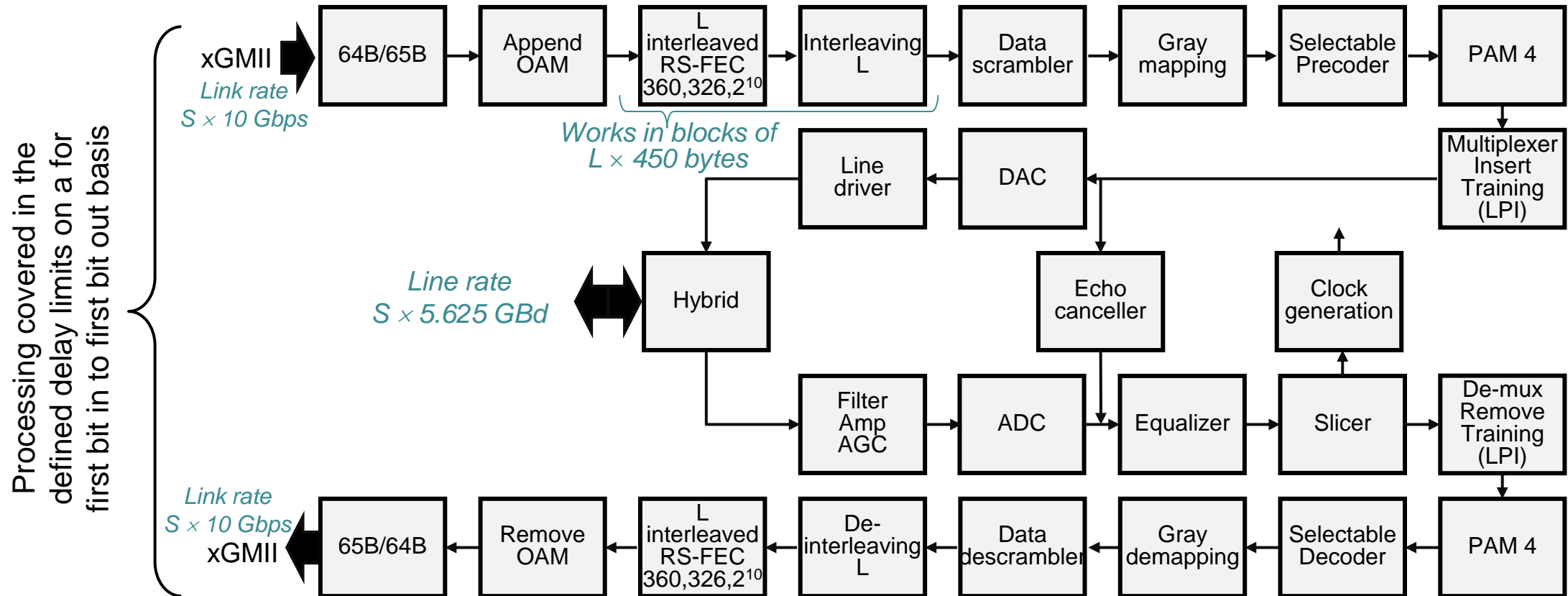
Content

- System overview
- **Latency aspects to consider**
 1. Latency requirements from camera application perspective (informative)
 2. Packet latency due to data rate and duplexing scheme
 3. Interrelation between latency and power consumption
 4. **PHY latency**
- Summary and conclusion

Link rate	S
2.5 Gbps	0.25
5 Gbps	0.5
10 Gbps	1

PHY example for 802.3ch

Clause 149.10 „The sum of the transmit and receive data delays for an implementation of the PHY shall not exceed the limits shown in Table 149–20. Transmit data delay is measured from the input of a given unit of data at the XGMII to the presentation of the same unit of data by the PHY to the MDI. Receive data delay is measured from the input of a given unit of data at the MDI to the presentation of the same unit of data by the PHY to the XGMII.“ The propagation delay on the channel is not included.



IEEE 802.3 defined PHY delay limits

Mode	Clause	Add. Info	Bit times	Pause quanta	Delay (tx/rx)
100 BASE-T1	96.10				360ns/960ns
1 GBASE-T1	97.10		7168	14	7.168us
2.5 GBASE-T1	149.10	L=1	10240	20	4.096 us
5 GBASE-T1	149.10	L=1	10240	20	2.048 us
5 GBASE-T1	149.10	L=2	13824	27	2.7648 us
10 GBASE-T1	149.10	L=1	10240	20	1.024 us
10 GBASE-T1	149.10	L=2	13824	27	1.3824 us
10 GBASE-T1	149.10	L=4	20480	40	2.048 us

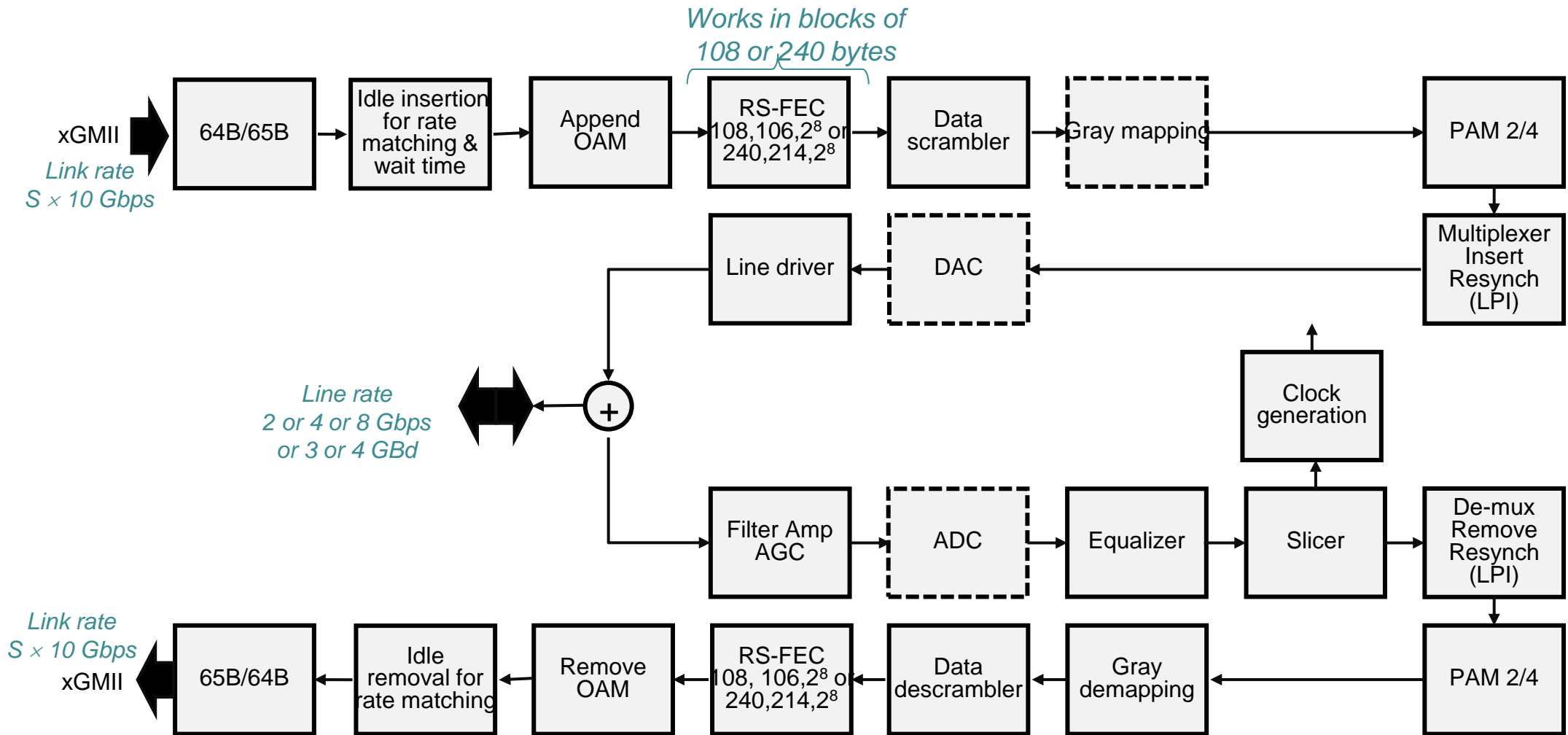
Working hypotheses: PHYs with similar processing steps (esp. in respect to FEC/interleaving) as 802.3ch can meet at least the same delay limits for that processing. PHYs with fewer/simpler steps can likely be even faster. (Wait times excluded)



Significantly smaller than t_{vblank}

Link rate	S
2.5 Gbps	0.25
5 Gbps	0.5
10 Gbps	1

PHY example for ASA-MLE



ASA-MLE does have less (complex) processing than 802.3ch. Same or smaller PHY latencies than 802.3ch expected.

PHY latency for an FDD system.

Without a specified FDD system only general statements are possible:

- The DS transceiver (camera side) differs significantly from the US transceiver (ECU side)
 - The DS transceiver (US receiver part) should be less complex than the US transceiver (DS receiver part) → Differentiation necessary.
- As transceivers transmit and receive at the same time, echoes overlay the receive signal. The closer the DS and US frequencies are the more disrupting. I.e., for DS 2.5Gbps with PAM4 (and e.g. US 100 Mbps) echoes are more disrupting than for 10 Gbps with PAM 4, which is more complex than in case of 10 Gbps with PAM 2. :
 - The high-speed (DS) receiver (ECU side) is more sensitive to echoes. An echo canceller (less complex than in case of FDX) significantly helps to improve SNR.
 - Inside the DS transceiver's low speed receiver the high-speed echoes are less disruptive. A light echo canceller helps also here to improve the SNR.
- The PHY transmit and receive latencies can expected to be in the same range as the respective IEEE 802.3 PHY latencies for the same speeds.

Comparison of latency values

	FDD	TDD (ASA-MLE)	FDX/EEE (802.3ch) 2.5/5/10 Gbps
Propagation delay	<90 ns	<90 ns	<90 ns
PHY processing latencies DS	Similar to ch (assuming similar processing)	At least as small as ch (no echo cancellation)	~4/2/1 us (w/o interleaving)
PHY processing latencies US	<1 us	At least as small as ch (no echo cancellation)	~4/2/1 us (w/o interleaving)
Packet latencies DS (including wait times), no power saving	~5/2.5/1.25 us (1542 bytes)	~ 7/4.4/2.1 us (1542 bytes)	~5/2.5/1.25 us (1542 bytes)
Packet latencies US (including wait times)	6.72 to 123.36 us (42 to 1542 bytes)	~3.95/2.99/26.83 us (150/150/200 bytes)	~74.67/36.57/18.1 us (891/415/184 bytes)

For all duplexing schemes, the Ethernet DS latency can be below ~10 us.

For all duplexing schemes, the Ethernet US latency can be below ~100 us.

Summary and conclusion

- In an FDD system, the Ethernet system US latency is dominated by the packet latency.
- In an TDD system, the Ethernet system US latency is dominated by the TDD wait time.
- Both can have the same length and be reduced by decreasing the Ethernet packet size. This either reduces the net throughput, or requires to increase the line rate.
- In an FDX/EEE system, the Ethernet system US latency is determined by the amount of power to be saved. In case of IEEE 802.3ch with EEE, EEE dominates the latency.
- The latency in the DS direction without extra power saving can be in the single digit μs range for all duplexing schemes.
- With power saving in the DS, the packets are best sent in a bulk. This almost levels all differences between the duplexing methods DS, with TDD being somewhat more flexible.
- A general statement that a TDD Ethernet system has a worse US (or DS) latency is incorrect. A TDD system will need a somewhat faster line rate for comparable PHY definition.
- In all cases, a typical automotive camera system functions with significantly larger latencies than the worst case asymmetric Ethernet latencies discussed in this presentation.
- Therefore, latency does not represent a crucial decision item between the options.

Thank You!