How hard could it be?

Understanding network traffic at the picosecond level

Background

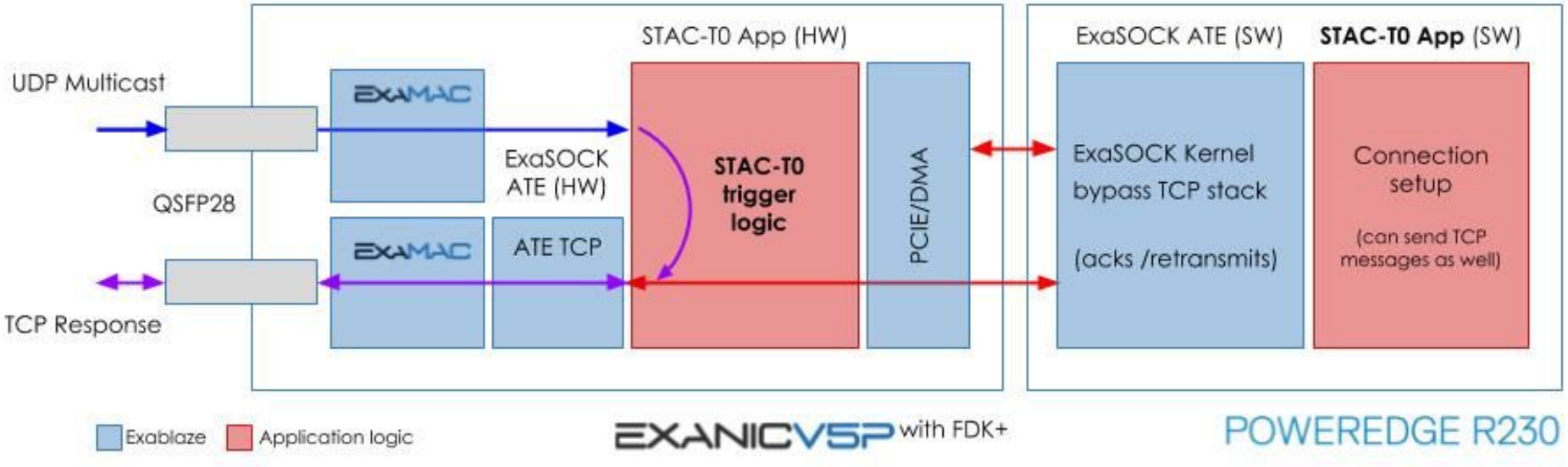
Introducing FDK-XP

-Everything from previous FDK

-Faster PCS/MAC

-Accelerated TCP Engine (ATE)

STAC-T0 (tick-to-trade)



STAC-T0 (tick-to-trade)

31ns*

Min. actionable latency

* Subject to final validation

STAC-T0.β1.*.*.ACTIONABLE.MIN)

STAC-T0 (tick-to-trade)

31ns*

Min. actionable latency

* Subject to final validation

STAC-T0.β1.*.*.ACTIONABLE.MIN)

STAC-T0 (tick-to-trade)

31ns*

Min. actionable latency

Why did it take so long?

* Subject to final validation

STAC-T0.β1.*.*.ACTIONABLE.MIN)

Possible reasons

Possible reasons

1. STAC can't measure things

Possible reasons

1. STAC can't measure things

2. It's harder than it looks

Possible reasons

1. ~~STAC can't measure things~~

2. **It's harder than it looks**

Enter the Picosecond

Problem:

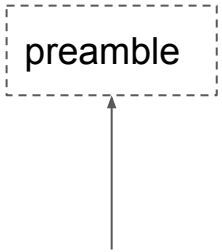
When did a field in my packet arrive?

In an ideal world...

An Ethernet fame:

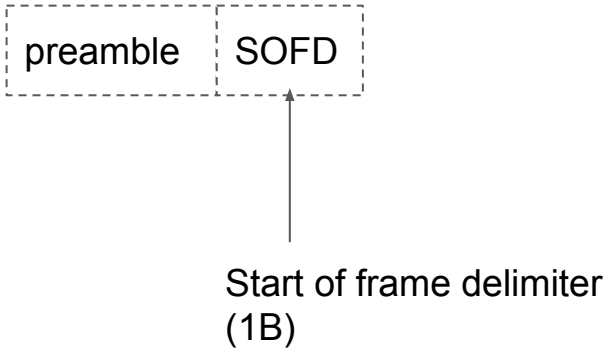
In an ideal world...

An Ethernet fame:



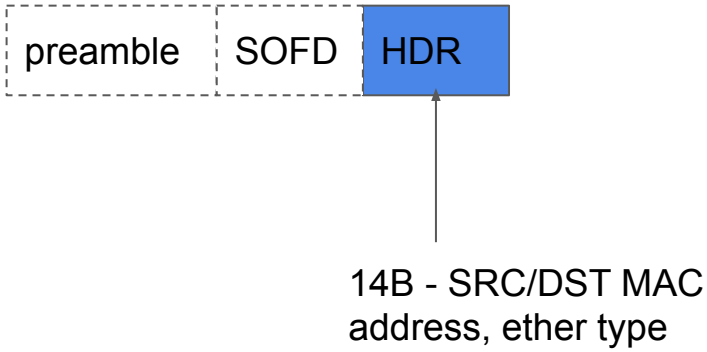
In an ideal world...

An Ethernet fame:



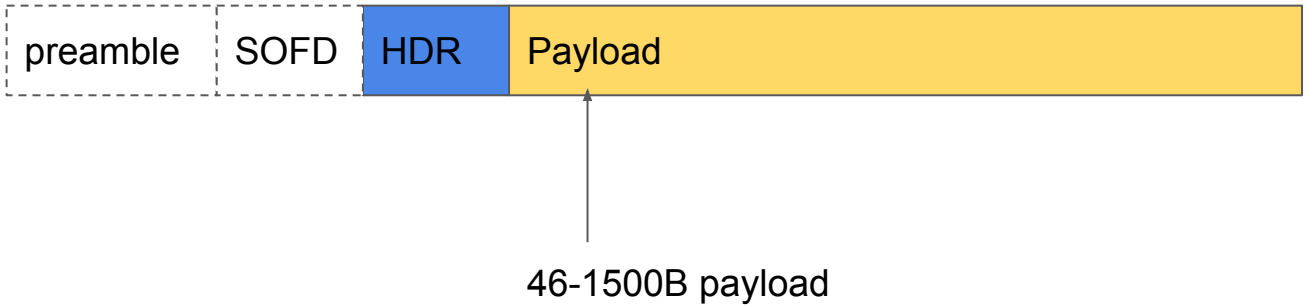
In an ideal world...

An Ethernet fame:



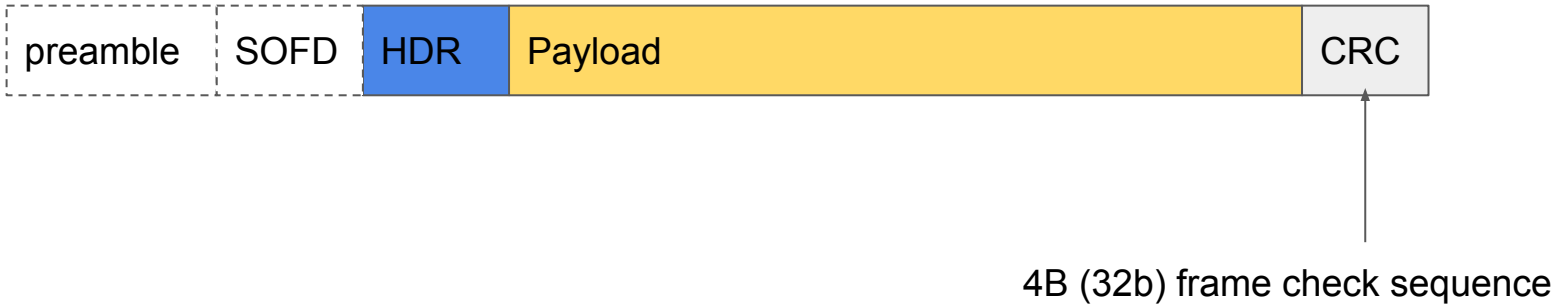
In an ideal world...

An Ethernet fame:



In an ideal world...

An Ethernet fame:

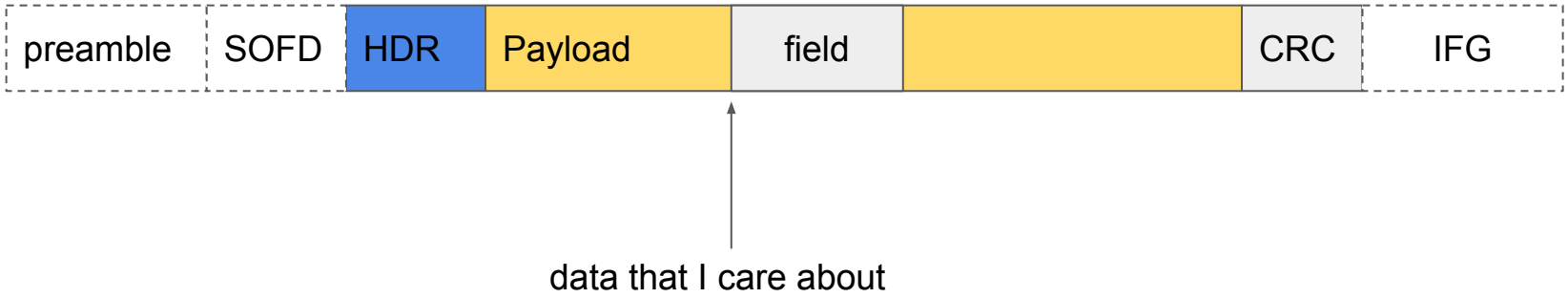


In an ideal world...

An Ethernet fame:

In an ideal world...

An Ethernet fame:

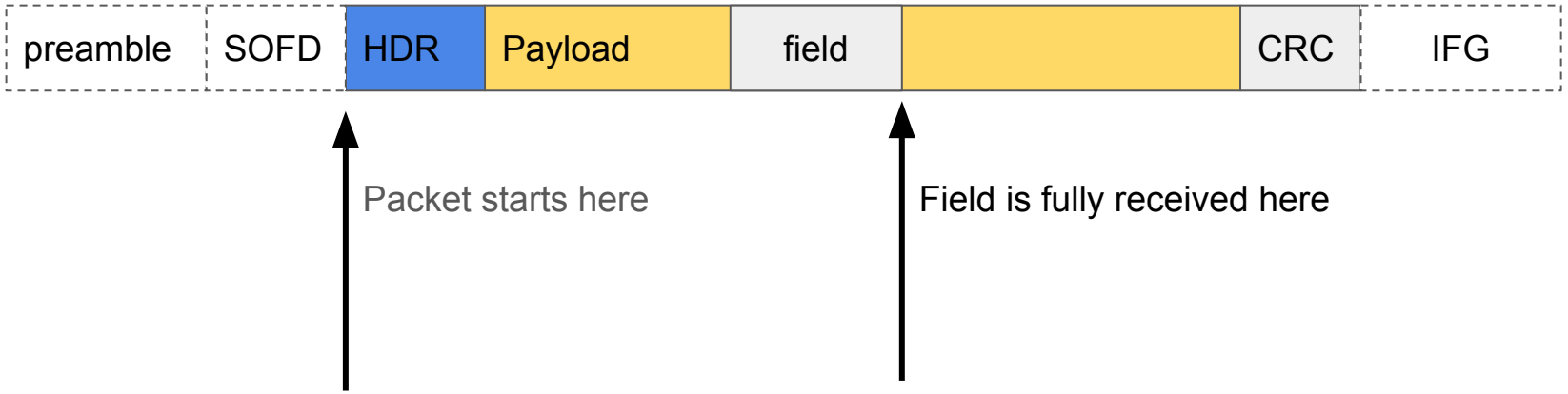


In an ideal world...

An Ethernet fame:

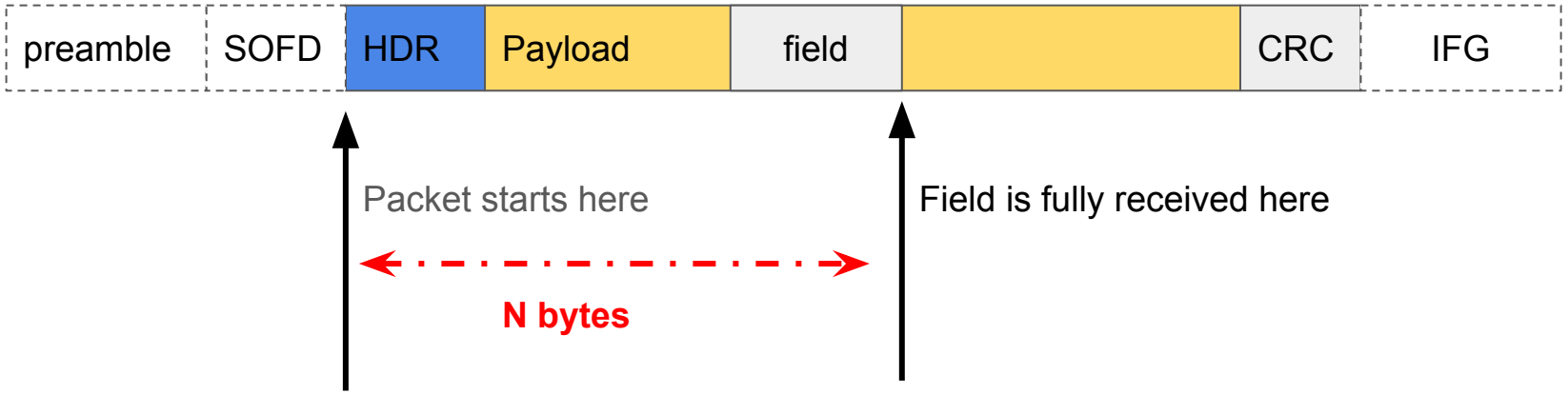
In an ideal world...

An Ethernet fame:



In an ideal world...

An Ethernet fame:



Ideal calculation

Delay = N

Ideal calculation

Ideal calculation

Line rate (giga-bits per second)

$$\text{Delay} = N \times 8 \times \frac{1}{10\ \text{Gb/s}}$$

Ideal calculation

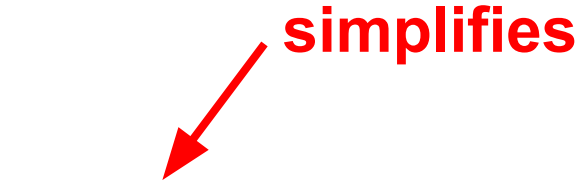
Convert to picoseconds (10^{-12})

$$\text{Delay} = N \times 8 \times \frac{1}{10\ \text{Gb/s}} \times 1/\text{ps}$$

Ideal calculation

$$\text{Delay} = N \times 8 \times 100$$

Ideal calculation



$$Delay = N \times 800$$

Ideal calculation - Example

Delay = 64B x 800 = 51,200ps

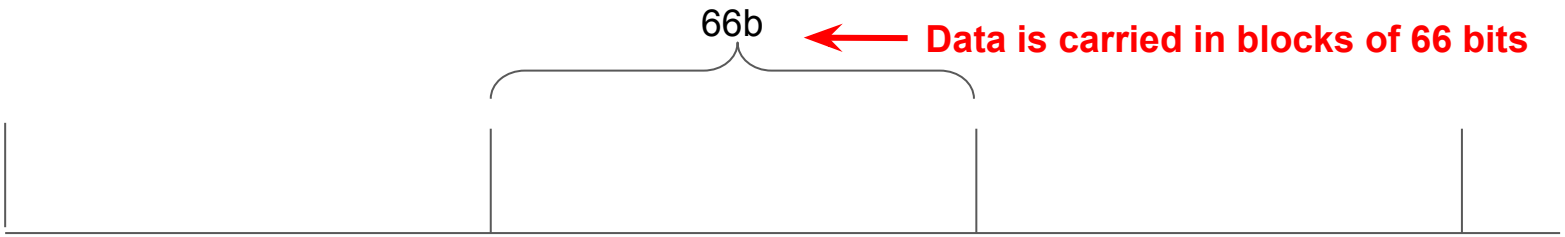
Finished?

Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s

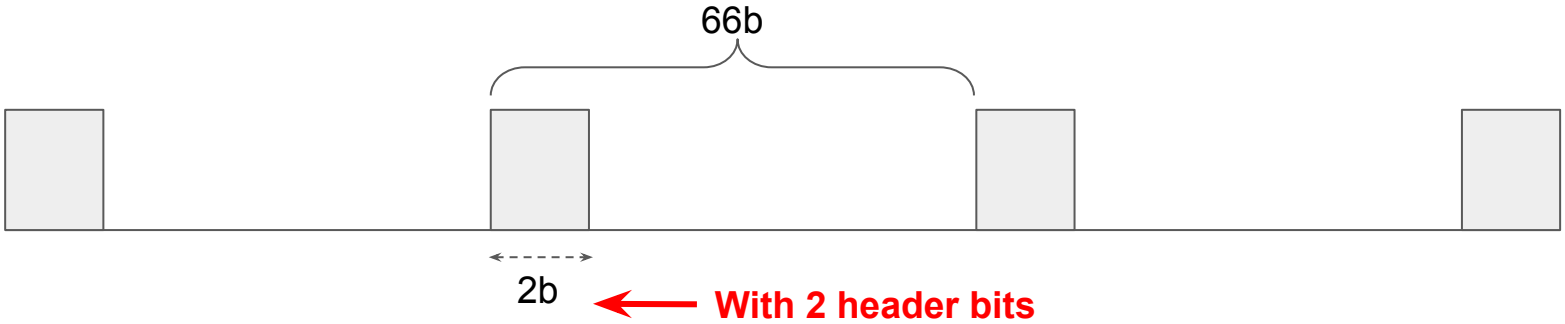
Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



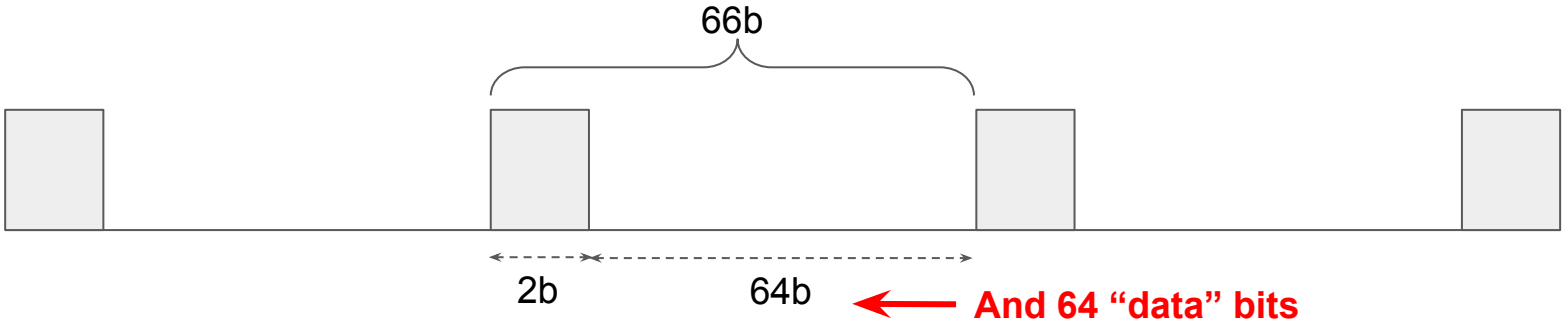
Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



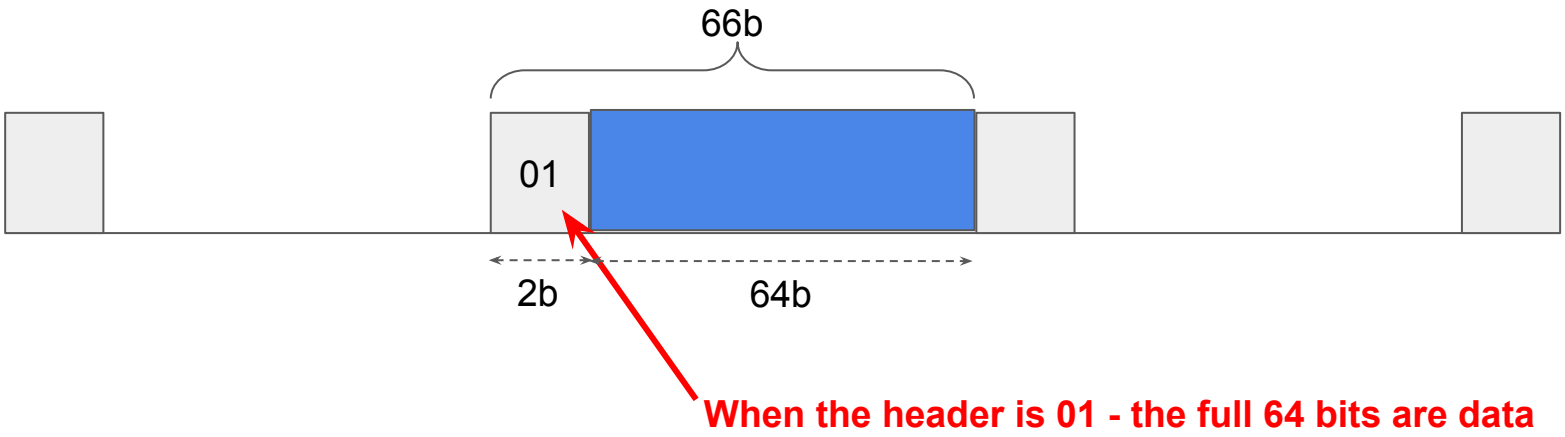
Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



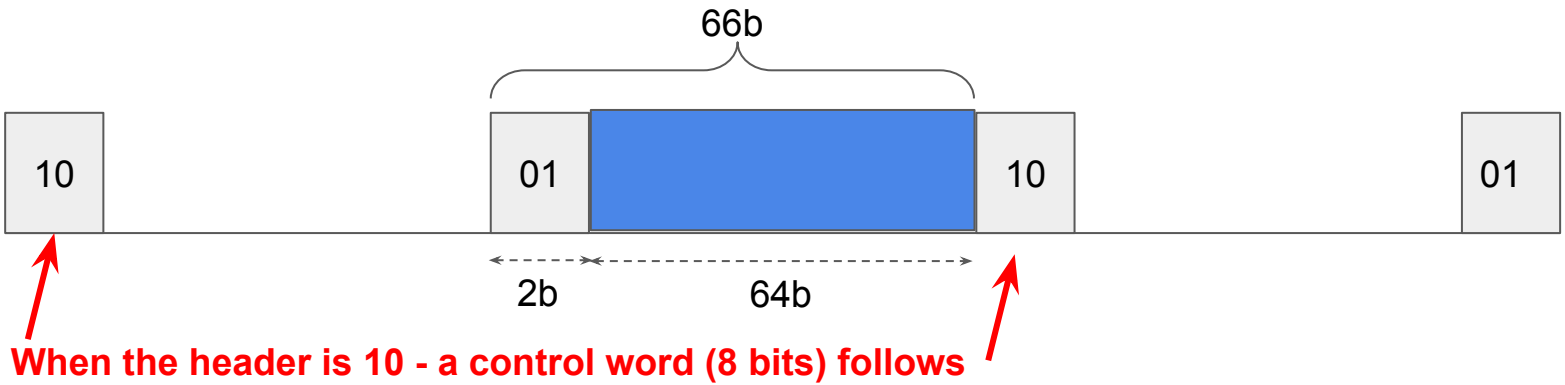
Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



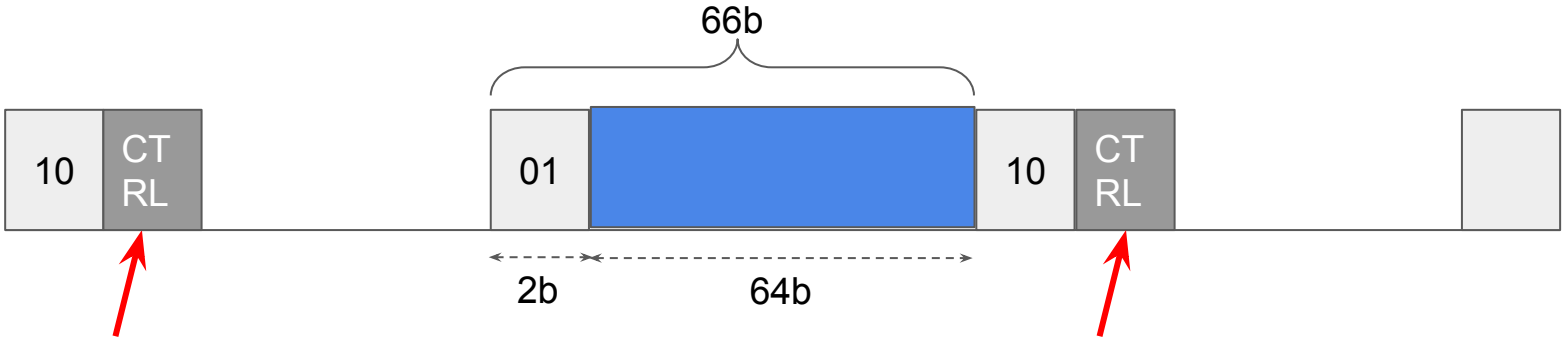
Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



Meanwhile in reality

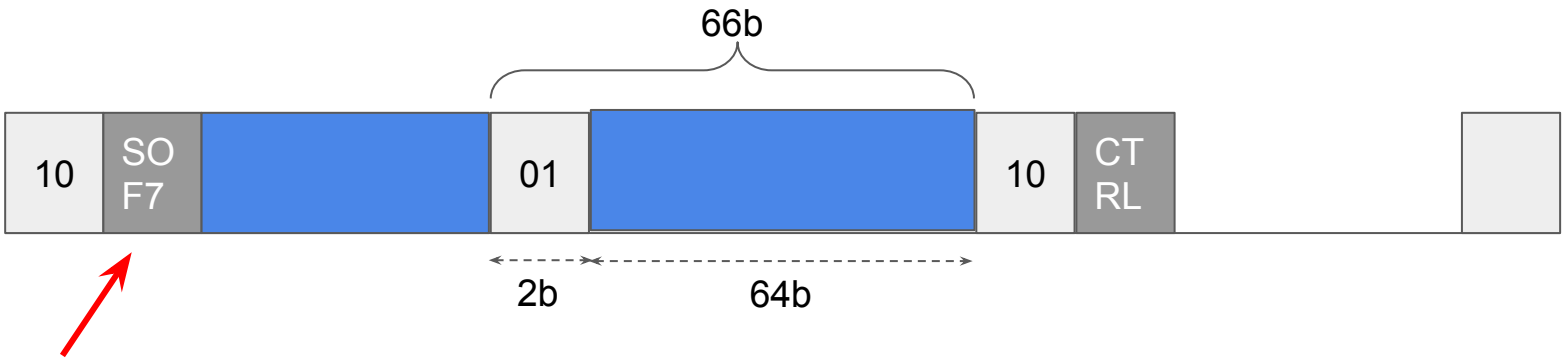
10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



The control word can have a number of values (256), but the most important ones for this discussion are ...

Meanwhile in reality

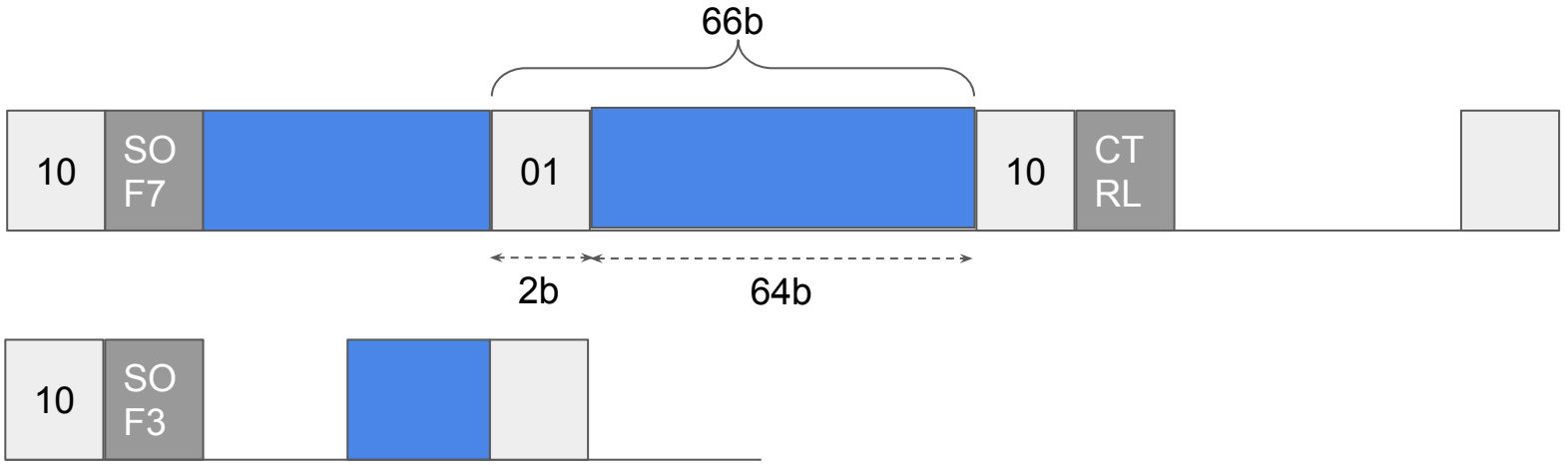
10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



Start of frame 7 (SOF7) - A frame has started there are 7B (56b) of data

Meanwhile in reality

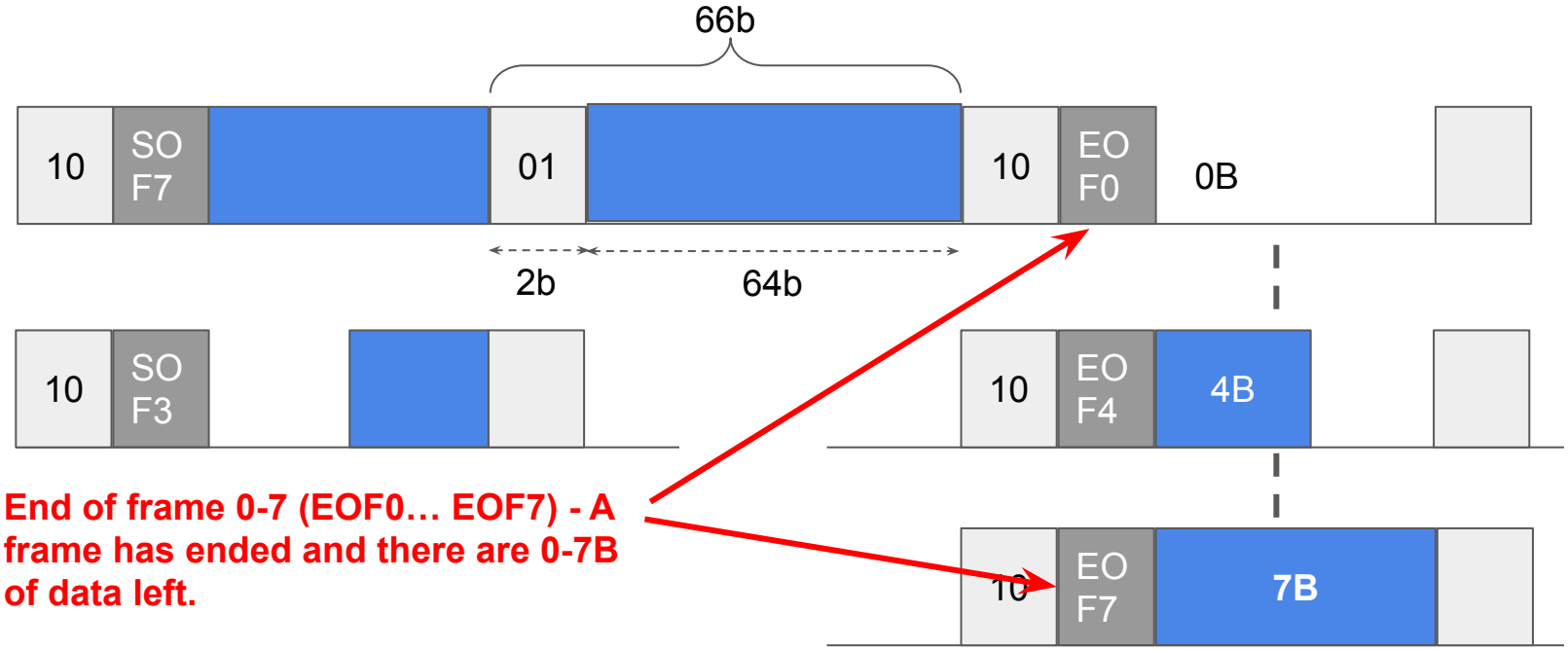
10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



And... start of frame 3 (SOF3) - A frame has started there are 4B of idle, then 3B of data (24bits)

Meanwhile in reality

10GbE is carried using 64b/66b encoding at 66/64 x 10 = 10.3125Gb/s



The making of an Ethernet frame

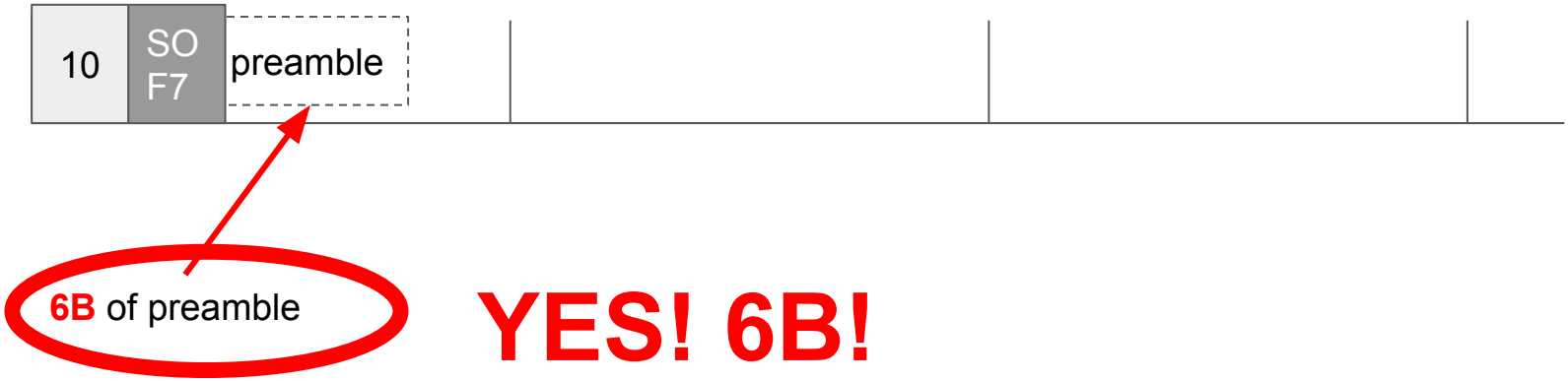
The Ethernet frame is then layered on top of the 64/66 encoding layer

The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...

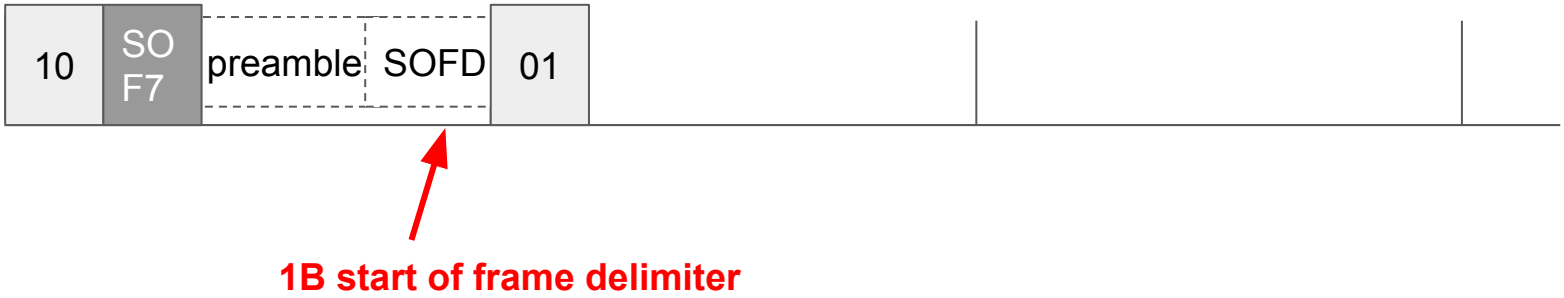
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



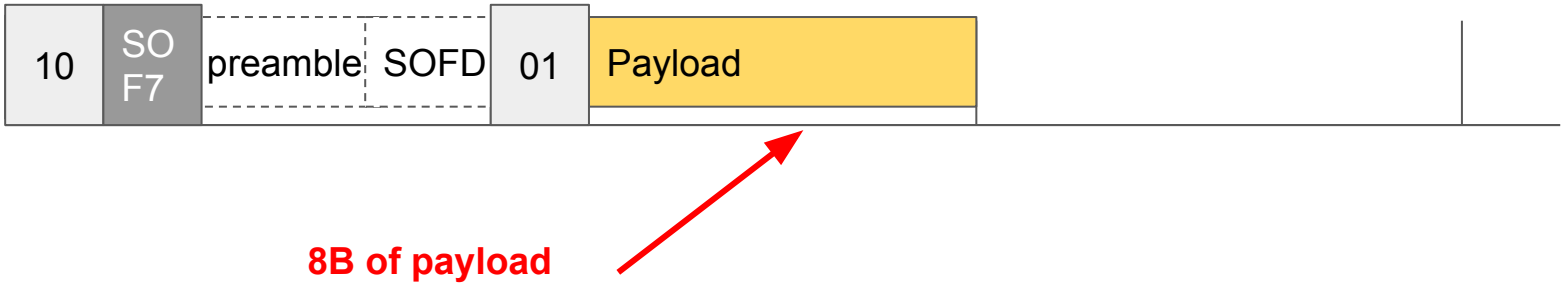
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For examle...



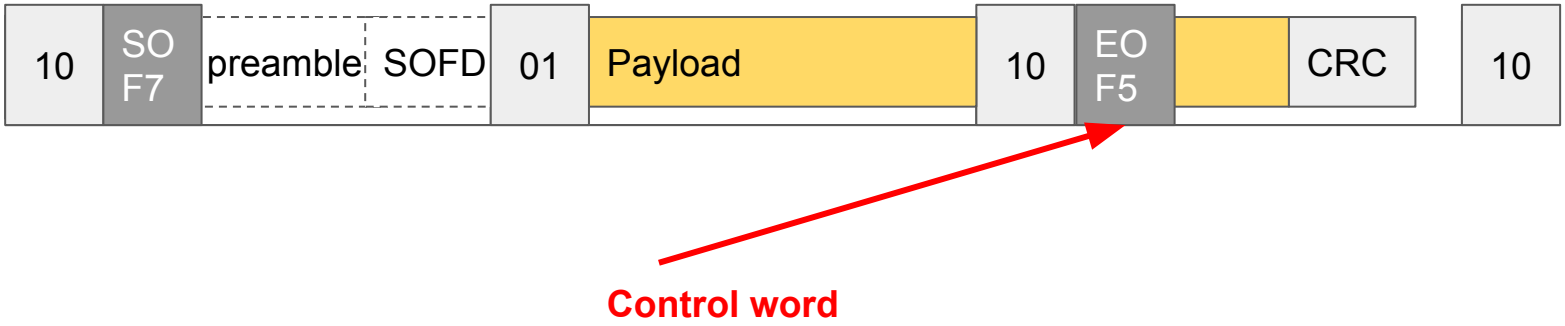
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



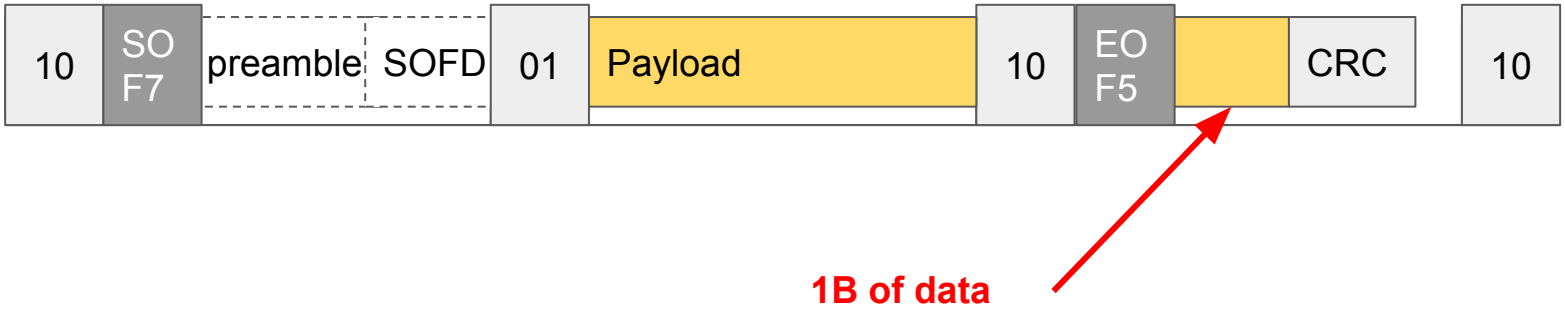
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



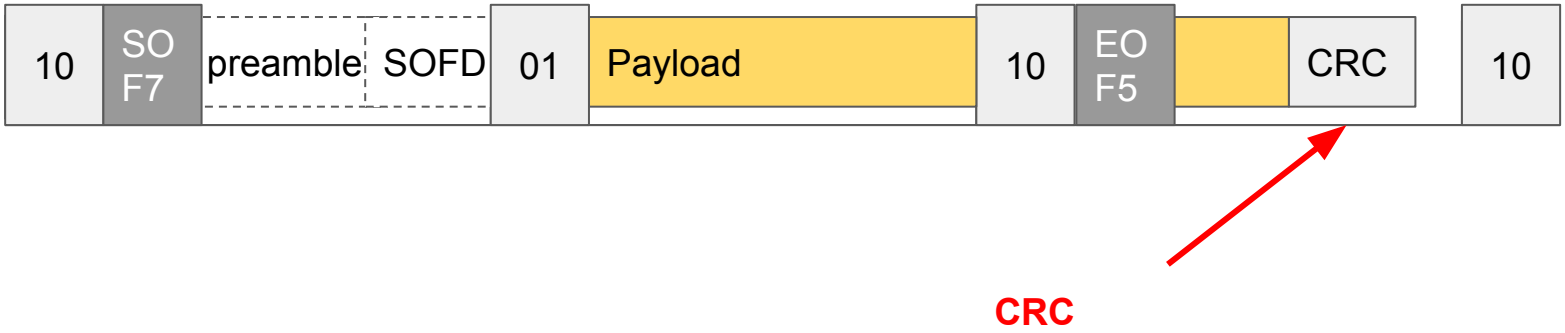
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



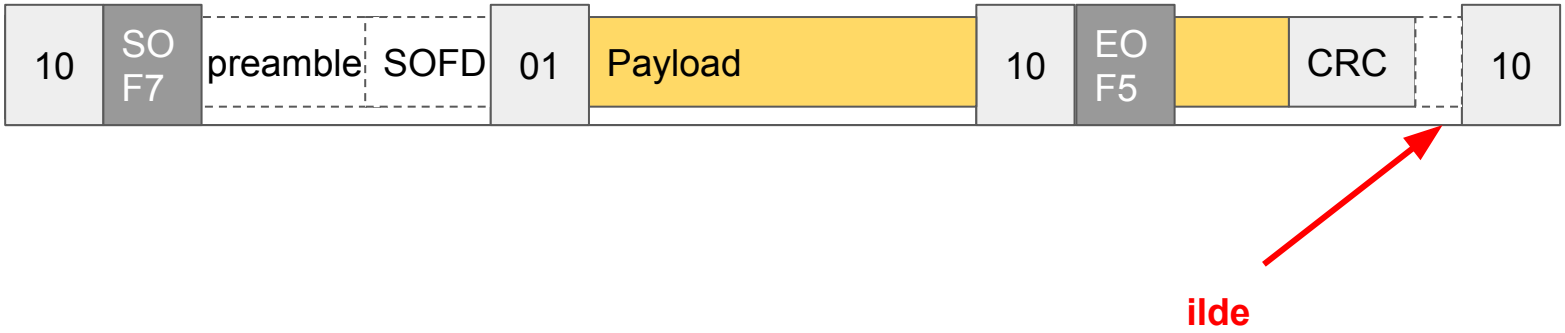
The making of an Ethernet frame

The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



The making of an Ethernet frame

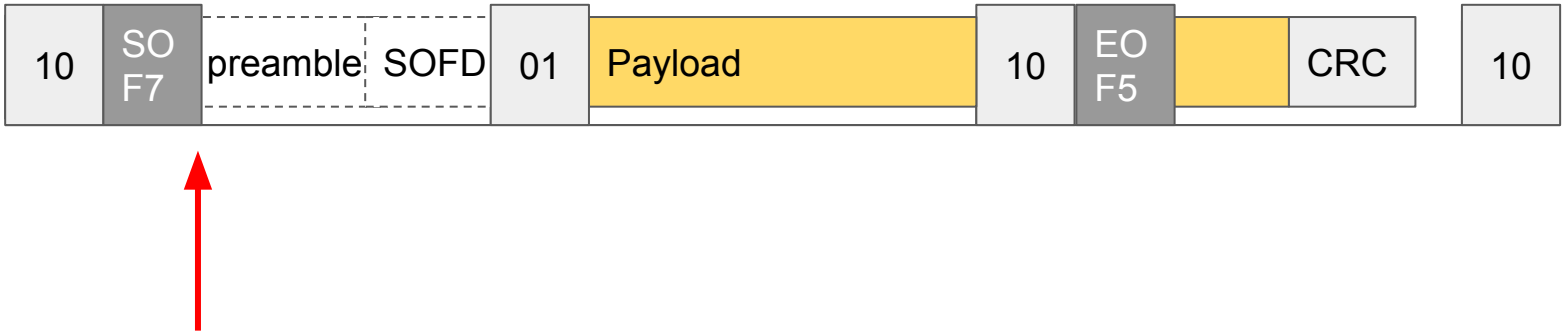
The Ethernet frame is then layered on top of the 64/66 encoding layer. For example...



This raises a lot of questions….

- **When does the frame start?**

When does a frame start?



When the 64/66 SOF7 signal is found?

When does a frame start?

Or when the SOFD finishes?

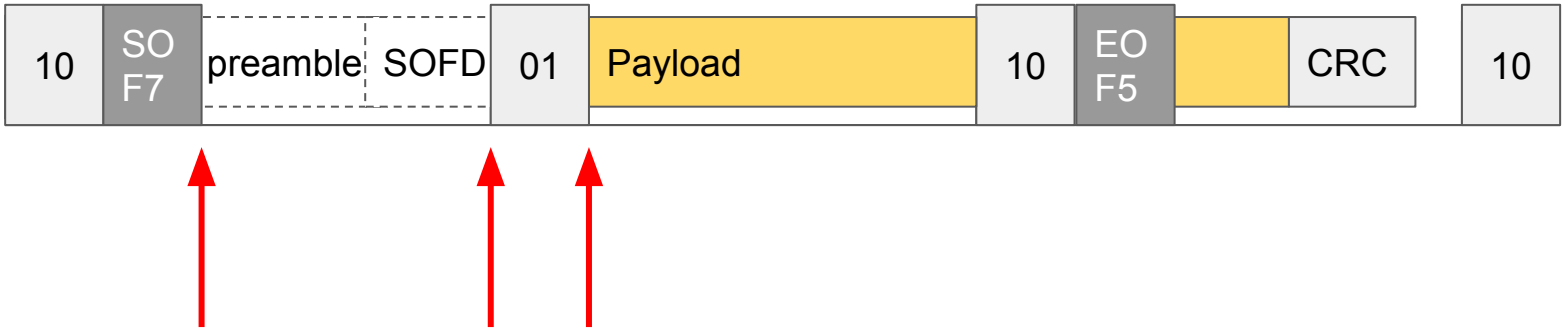
When does a frame start?

Or when the payload starts?

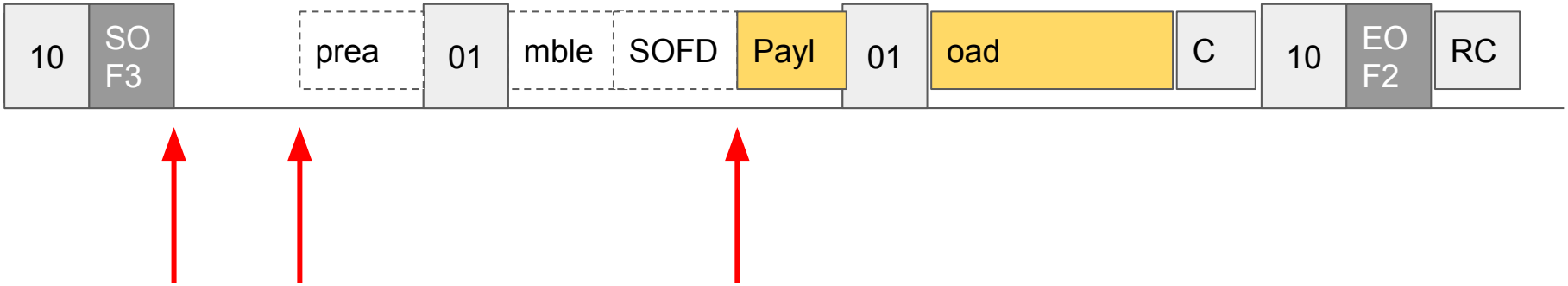
This raises a lot of questions….

- When does the frame start (SOF)?
- **When is the SOF timestamped?**

When does a frame start? And when is it timestamped



When does a frame start? And when is it timestamped

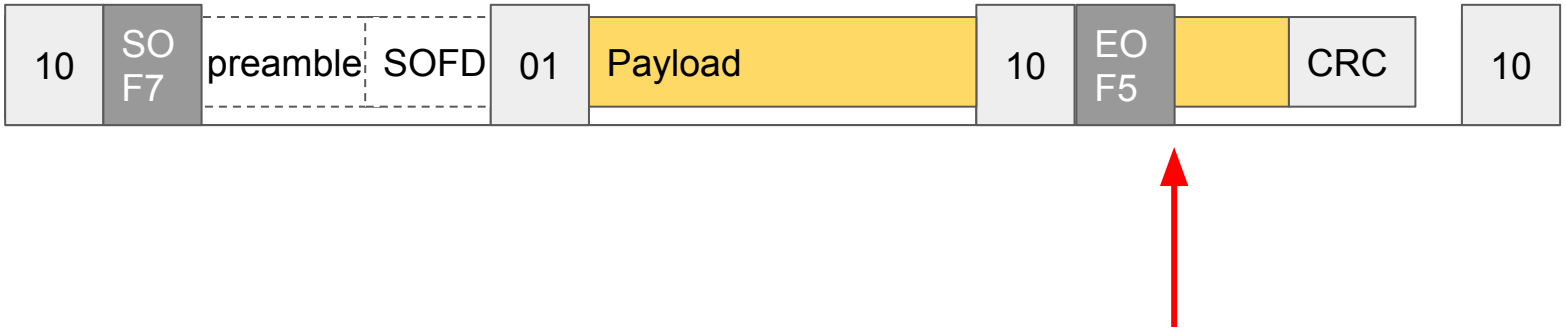


What about the SOF3 case?

This raises a lot of questions….

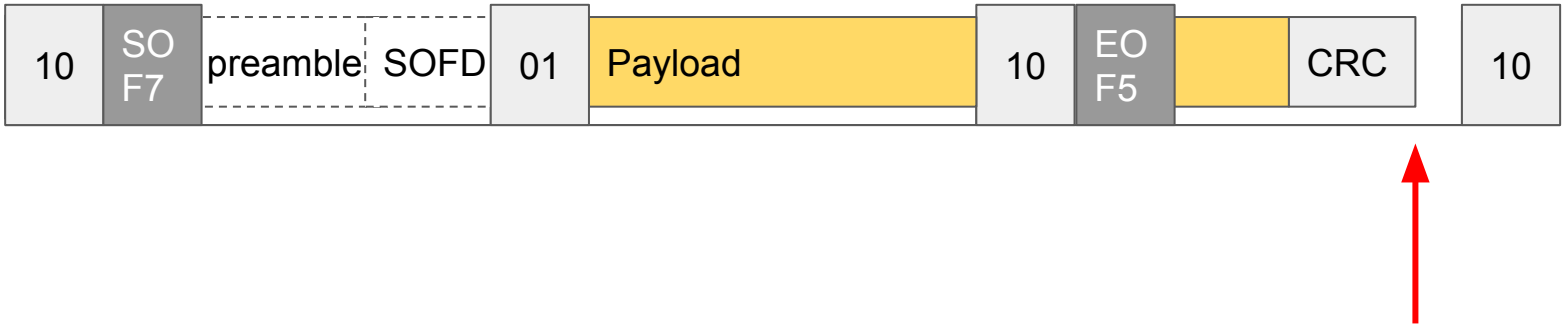
- When does the frame start? And when is it timestamped?
- **When does the frame end?**

When does a frame end?



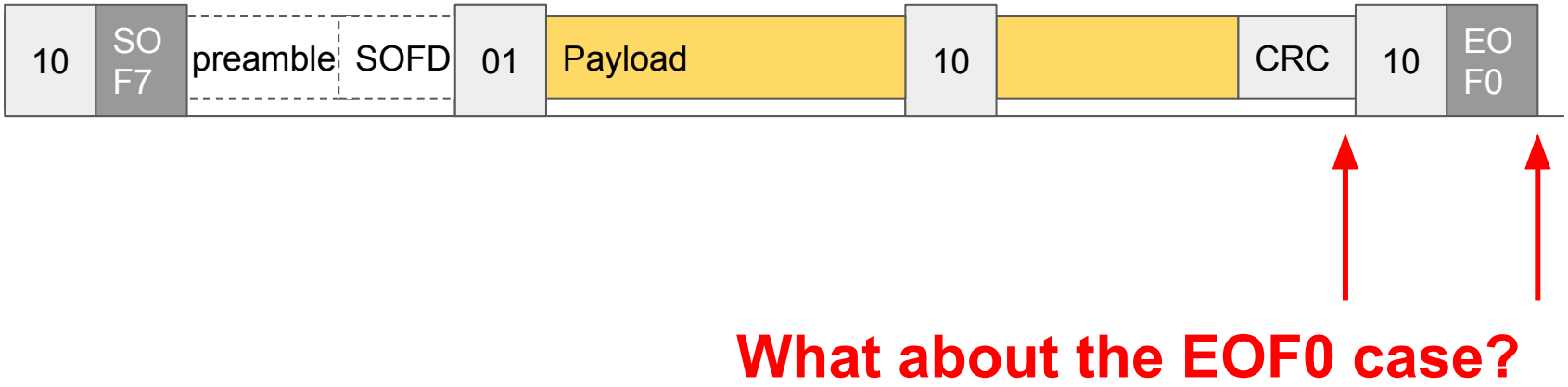
When the 64/66 EOF1-7 signal is found?

When does a frame end?



Or when the frame CRC arrives?

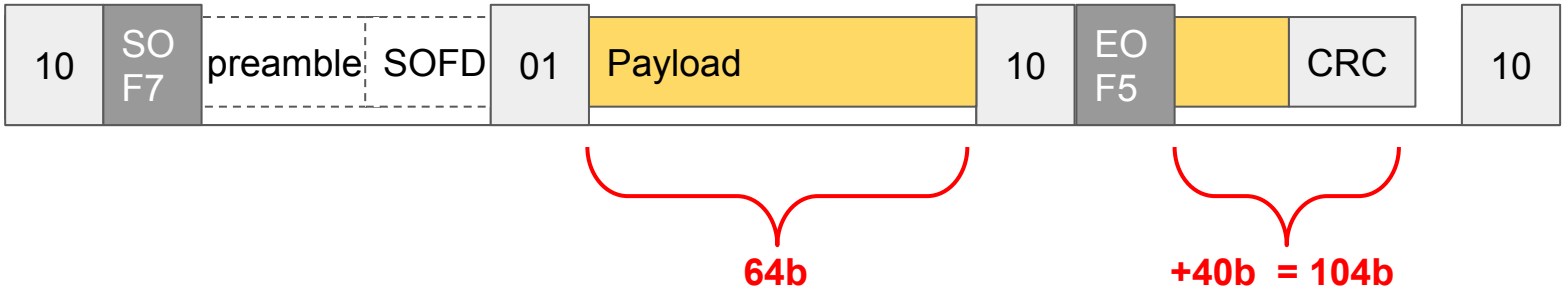
When does a frame end?



This raises a lot of questions….

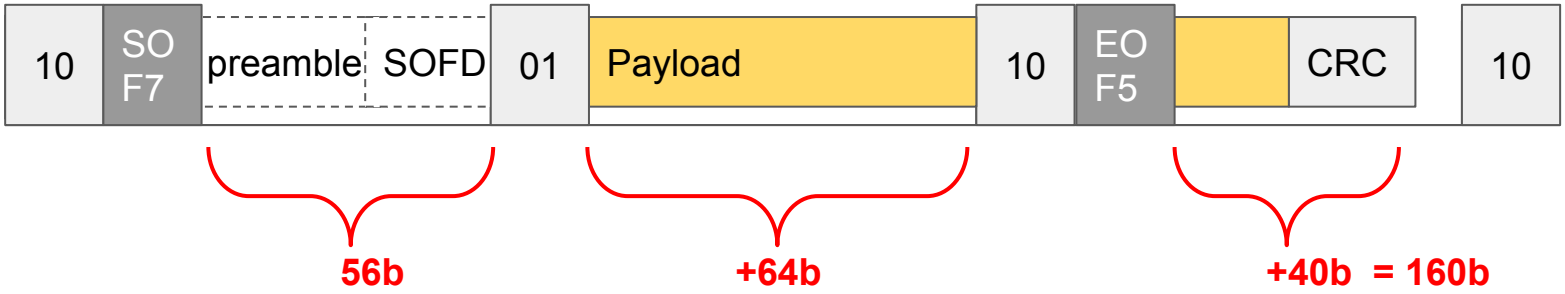
- When does the frame start? And when it it timestamped?
- When does the frame end?
- **How long is the frame? (in bits and in picoseconds)**

How long is the frame? (In bits / picoseconds)



Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

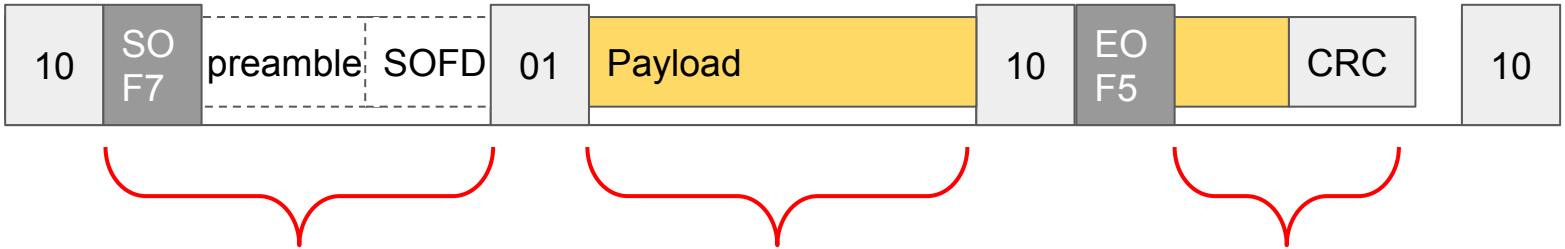
How long is the frame? (In bits / picoseconds)



Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)

How long is the frame? (In bits / picoseconds)

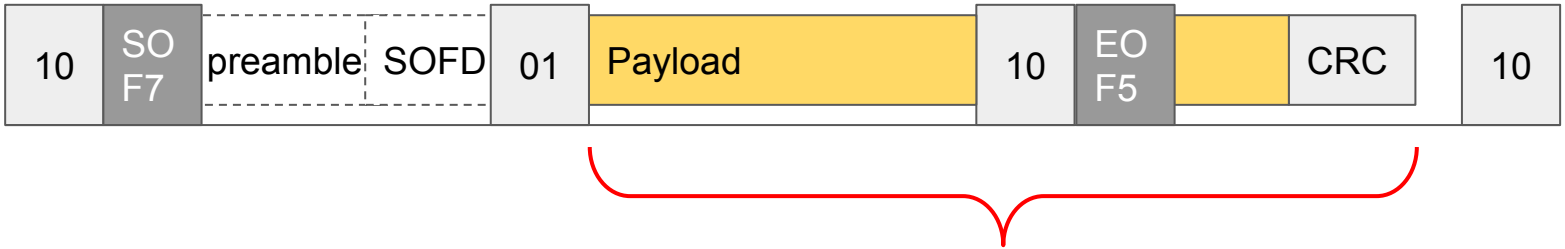


Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)

Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)

How long is the frame? (In bits / picoseconds)



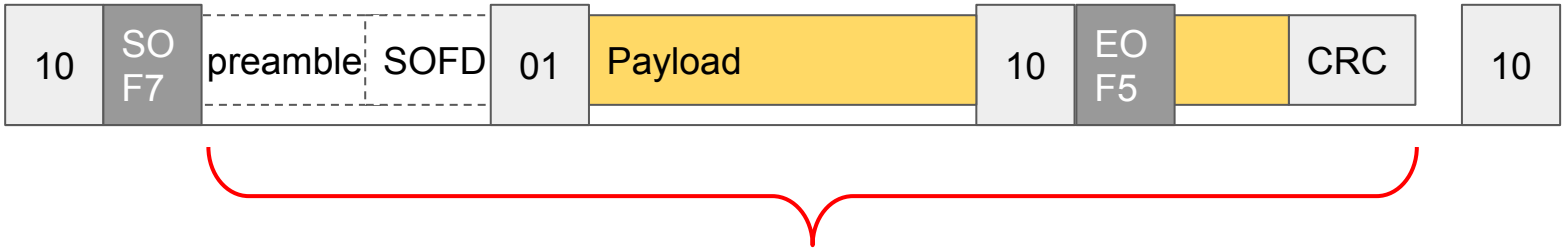
Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)

Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)

Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], ideal Ethernet view)

How long is the frame? (In bits / picoseconds)



56 +2 +64 +2 +8 +40 = 172b

Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

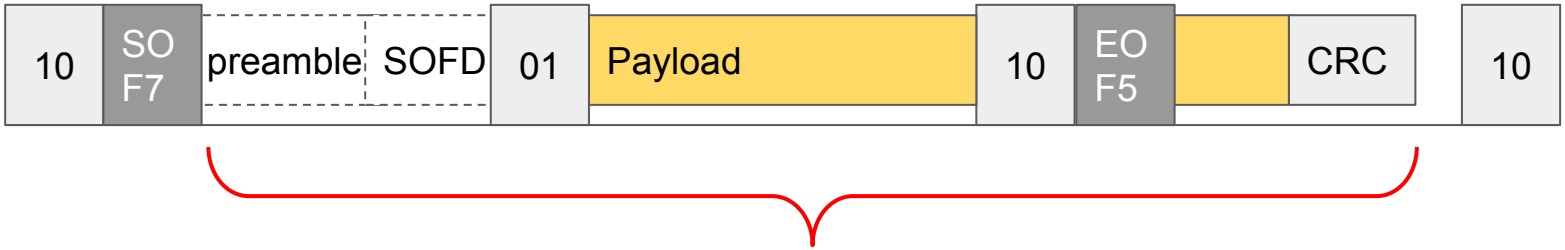
Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)

Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)

Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)

Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)

How long is the frame? (In bits / picoseconds)



56 +2 +64 +2 +8 +40 = 172b

Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)

Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)

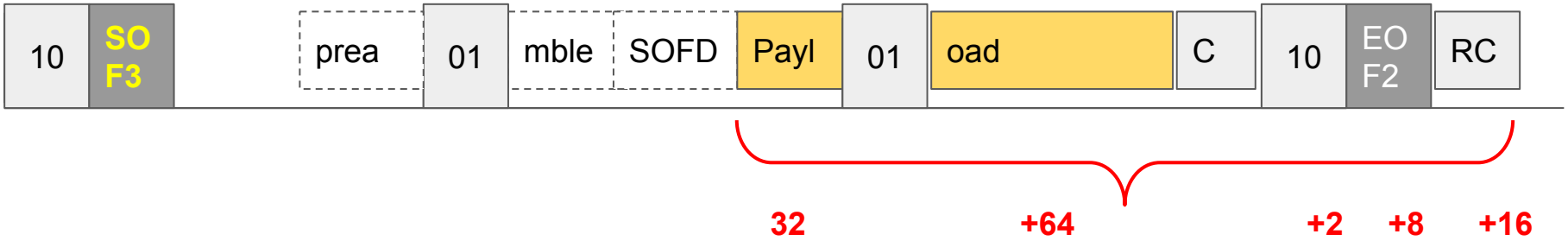
Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)

Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)

Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)

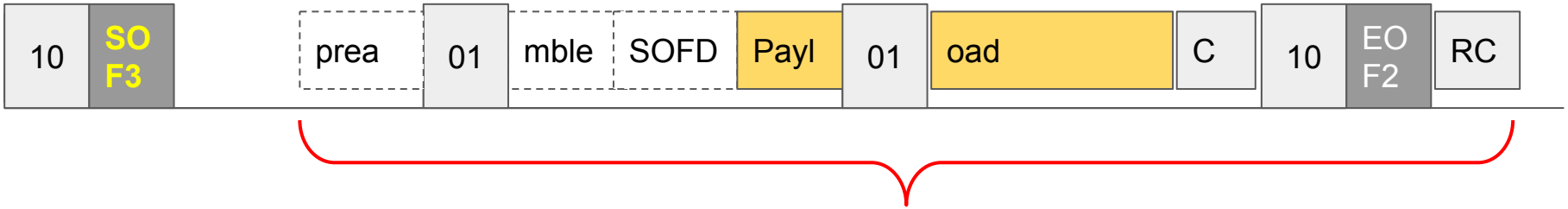
What about the SOF3 case?

How long is the frame? (In bits / picoseconds)



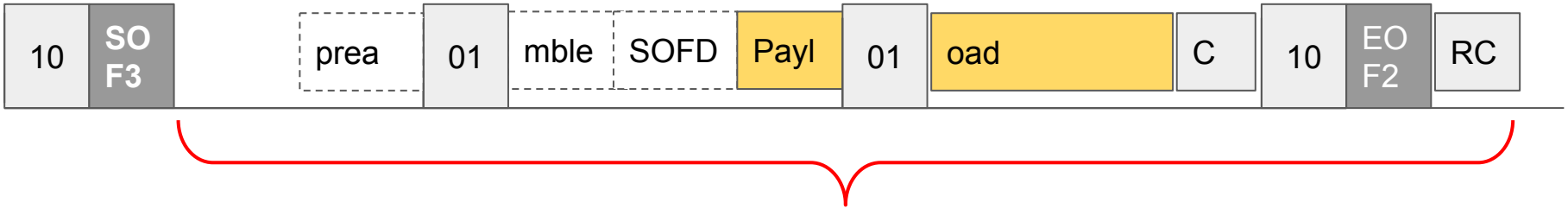
Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)
Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)
Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)
Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)
Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)
Is it 122b long @ 10.3125Gb/s = 11,830ps ? (PCS[SOF3], Ethernet view)

How long is the frame? (In bits / picoseconds)



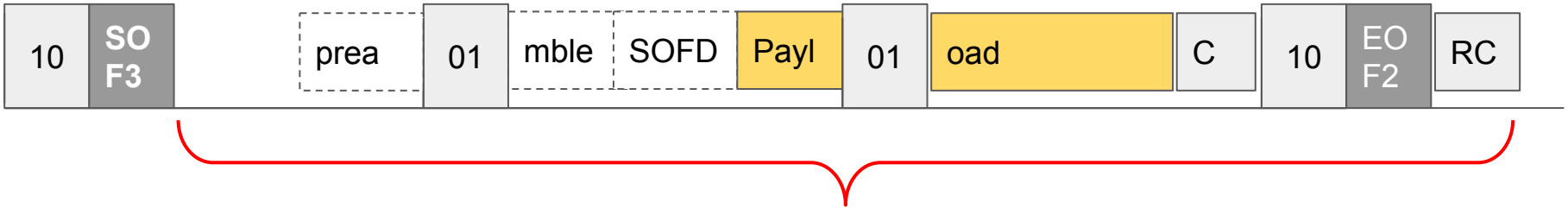
Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)
Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)
Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)
Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)
Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)
Is it 122b long @ 10.3125Gb/s = 11,830ps ? (PCS[SOF3], Ethernet view)
Is it 182b long @ 10.3125Gb/s = 17,648ps ? (PCS[SOF3], Ethernet view + preamble/SOFD)

How long is the frame? (In bytes / picoseconds)



Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)
Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)
Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)
Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)
Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)
Is it 122b long @ 10.3125Gb/s = 11,830ps ? (PCS[SOF3], Ethernet view)
Is it 182b long @ 10.3125Gb/s = 17,648ps ? (PCS[SOF3], Ethernet view + preamble/SOFD)
Is it 214b long @ 10.3125Gb/s = 20,752ps ? (PCS[SOF3/SOF7] Ethernet view)

How long is the frame? (In bytes / picoseconds)



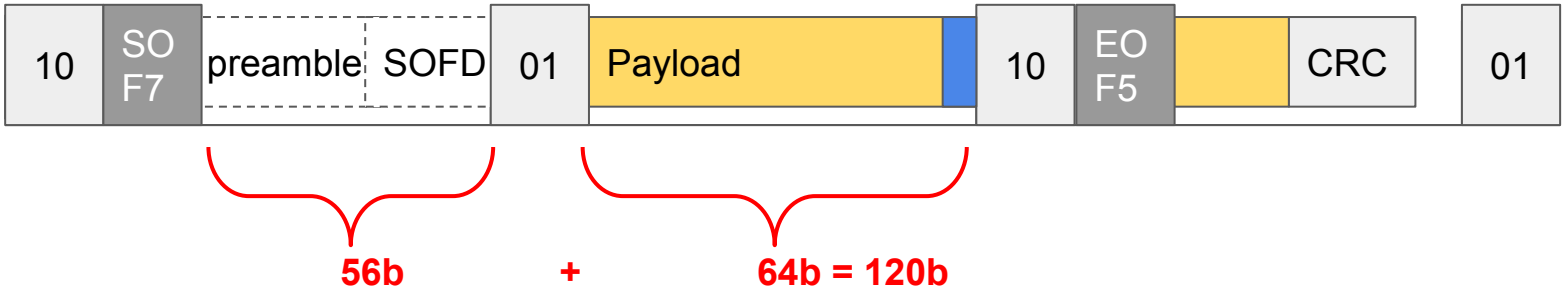
Is it 104b long @ 10Gb/s = 10,400ps ? (ideal Ethernet view)
Is it 160b long @ 10Gb/s = 16,000ps ? (ideal Ethernet view + preamble/SOFD)
Is it 168b long @ 10Gb/s = 16,800ps ? (ideal Ethernet view + XGMII preamble/SOFD)
Is it 114b long @ 10.3125Gb/s = 11,054ps ? (PCS[SOF7], Ethernet view)
Is it 172b long @ 10.3125Gb/s = 16,679ps ? (PCS[SOF7], Ethernet view + preamble/SOFD)
Is it 122b long @ 10.3125Gb/s = 11,830ps ? (PCS[SOF3], Ethernet view)
Is it 182b long @ 10.3125Gb/s = 17,648ps ? (PCS[SOF3], Ethernet view + preamble/SOFD)
Is it 214b long @ 10.3125Gb/s = 20,752ps ? (PCS[SOF3/SOF7] Ethernet view)

… and many more...

This raises a lot of questions….

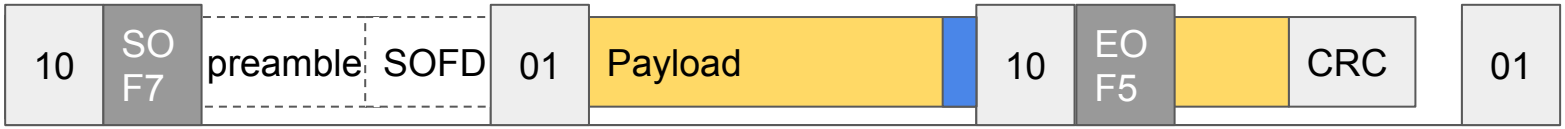
- When does the frame start? And when it it timestamped?
- When does the frame end?
- How long is the frame? (in bits and in picoseconds)
- **How far (ps) into the frame is an arbitrary offset?**

How far (ps) into the frame is an offset?



Is it 120b @ 10.000Gb/s = 12,000ps ?

How far (ps) into the frame is an offset?

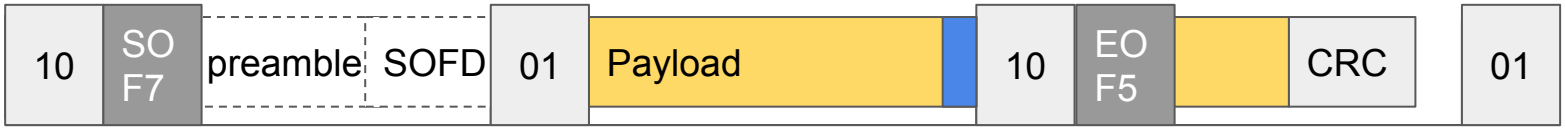


56b **+2** **+64b = 122b**

Is it 120b @ 10.000Gb/s = 12,000ps ?

Is it 122b @ 10.3125Gb/s = 11,830ps ?

How far (ps) into the frame is an offset?



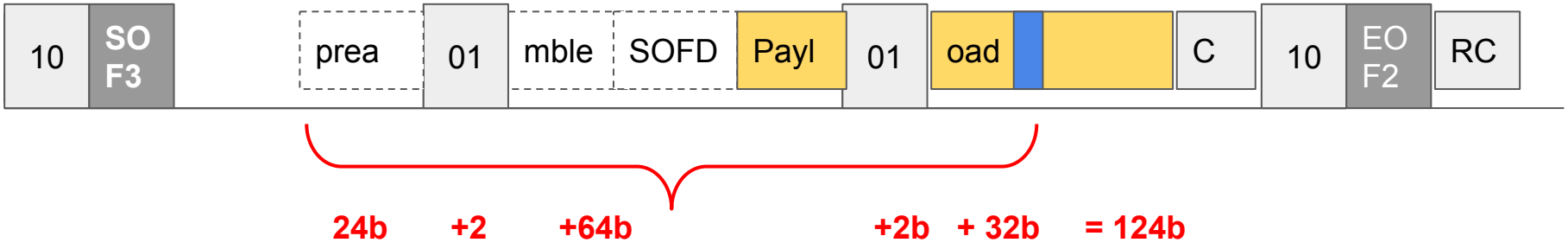
56b **+2** **+64b = 122b**

Is it 120b @ 10.000Gb/s = 12,000ps ?

Is it 122b @ 10.3125Gb/s = 11,830ps ?

Note: this is smaller than above!!!

How far (ps) into the frame is an offset?

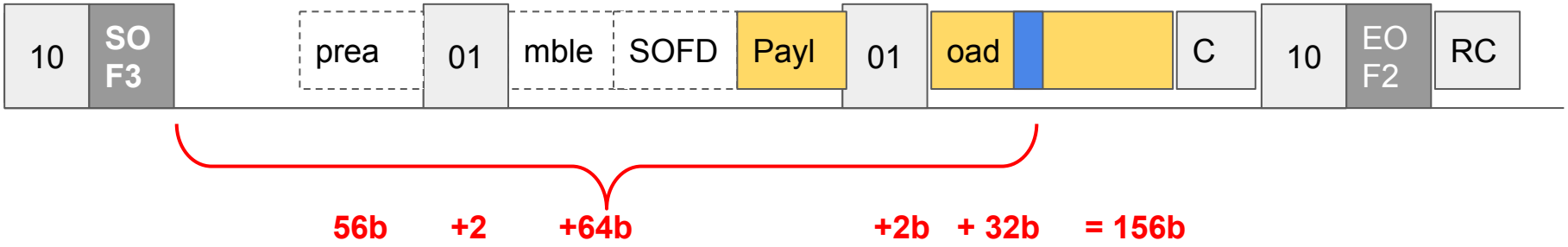


Is it 120b @ 10.000Gb/s = 12,000ps ?

Is it 122b @ 10.3125Gb/s = 11,830ps ?

Is it 124 @ 10.3125Gb/s = 12,024ps ?

How far (ps) into the frame is an offset?



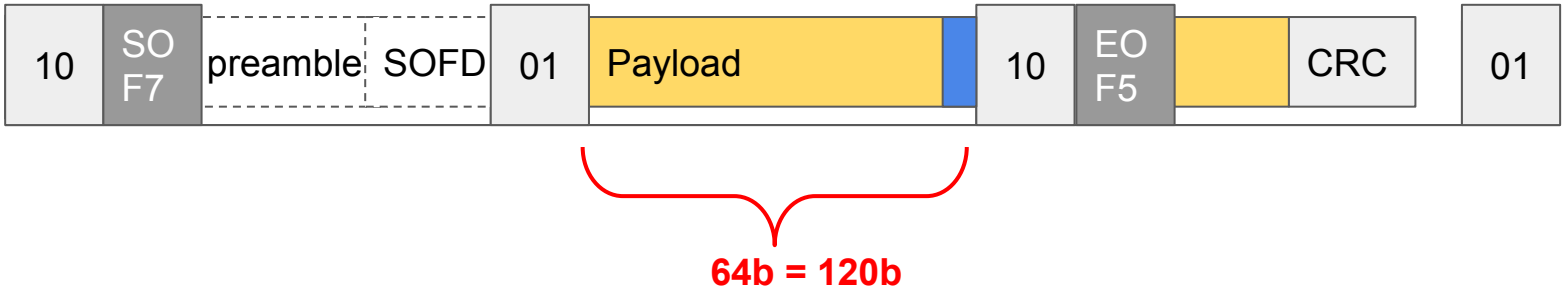
Is it 120b @ 10.000Gb/s = 12,000ps ?

Is it 122b @ 10.3125Gb/s = 11,830ps ?

Is it 124 @ 10.3125Gb/s = 12,024ps ?

Is it 156 @ 10.3125Gb/s = 15,127ps ?

How far (ps) into the frame is an offset?



Is it 120b @ 10.000Gb/s = 12,000ps ?

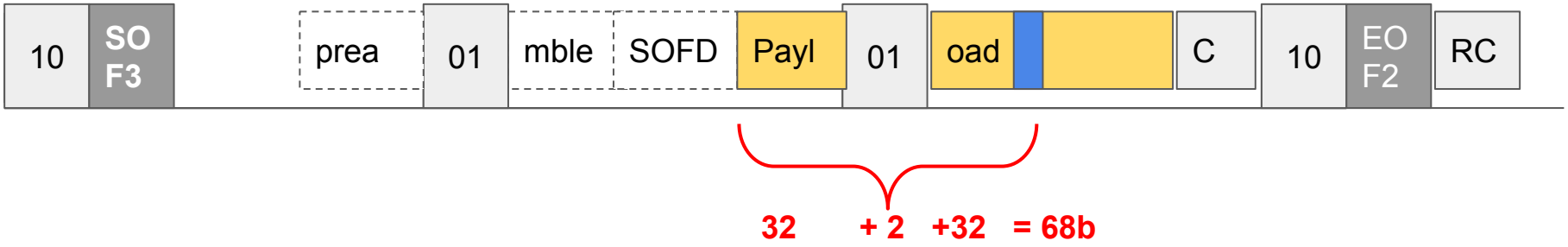
Is it 122b @ 10.3125Gb/s = 11,830ps ?

Is it 124 @ 10.3125Gb/s = 12,024ps ?

Is it 156 @ 10.3125Gb/s = 15,127ps ?

Is it 64b @ 10.3125Gb/s = 6,206ps ?

How far (ps) into the frame is an offset?



Is it 120b @ 10.000Gb/s = 12,000ps ?

Is it 122b @ 10.3125Gb/s = 11,830ps ?

Is it 124 @ 10.3125Gb/s = 12,024ps ?

Is it 156 @ 10.3125Gb/s = 15,127ps ?

Is it 64b @ 10.3125Gb/s = 6,206ps ?

Is it 68b @ 10.3125Gb/s = 6,594ps ?

Implications for uncertainty

1. Ethernet protocol has an **average** rate of 10Gb/s at layer 2, but PCS effects are visible at individual packet sizes. Thus PCS encodings must be taken into account and 10.3125Ghz must be used.

Implications for uncertainty

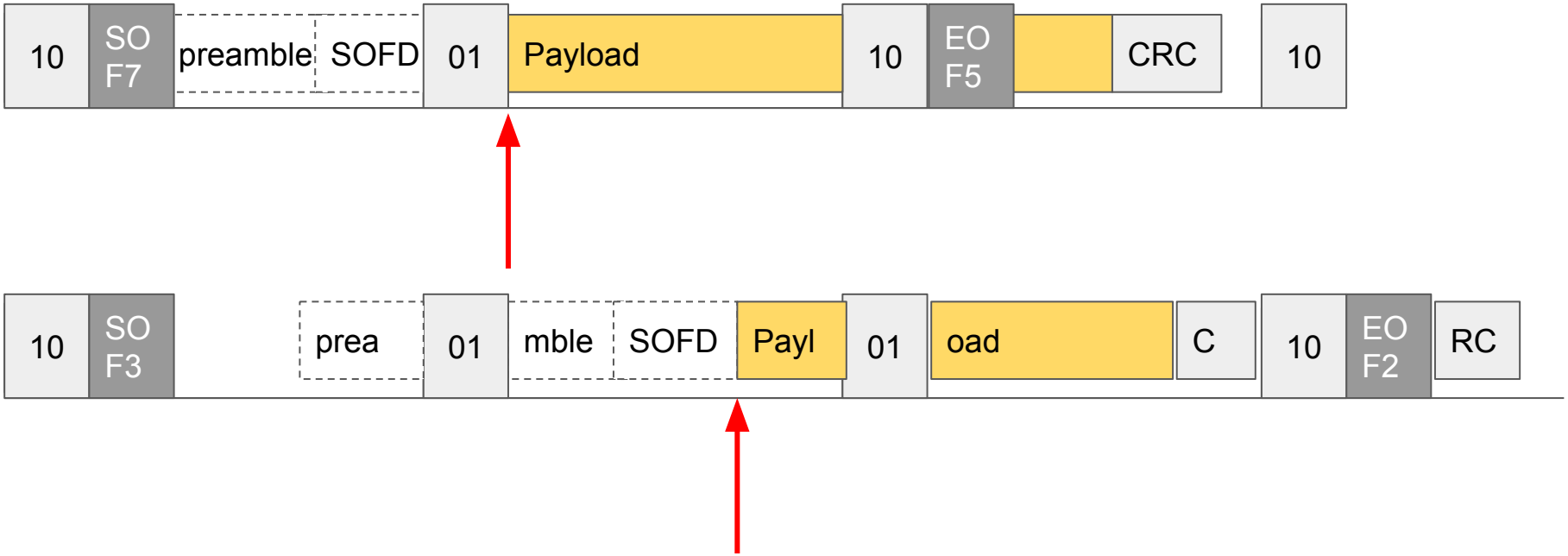
1. Ethernet protocol has an **average** rate of 10Gb/s at layer 2, but PCS effects are visible at individual packet sizes. Thus PCS encodings must be taken into account and 10.3125Ghz must be used.
2. Timestamps at PCS SOF3/7 and Ethernet layer SOFD have different absolute offsets*. Since both SOF3 and SOF7 may appear, these need to be accounted for.

Our recommendations….

Our recommendations….

- When does the frame start? *At the start of the payload (DST MAC)*

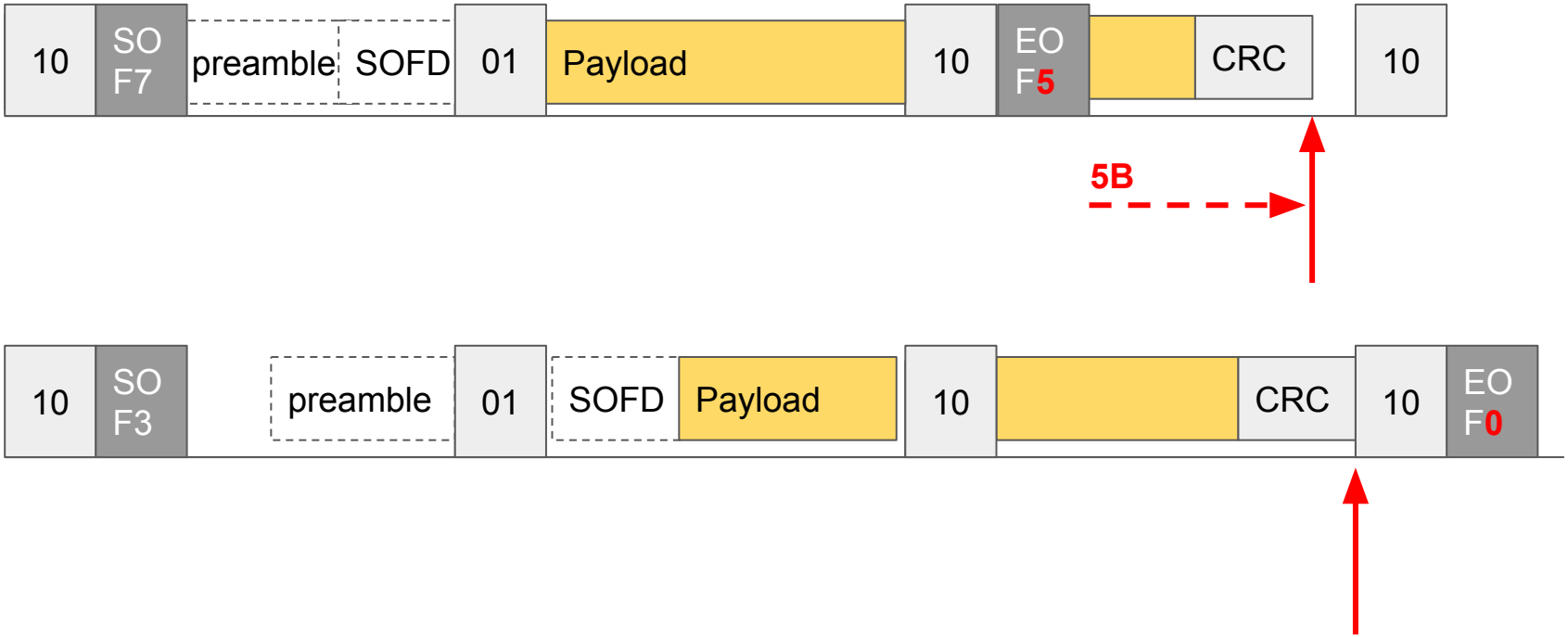
When does a frame start? *At the start of the payload (DST MAC)*



Our recommendations….

- When does the frame start? *At the start of the payload*
- When does the frame end? *At the end of the CRC*

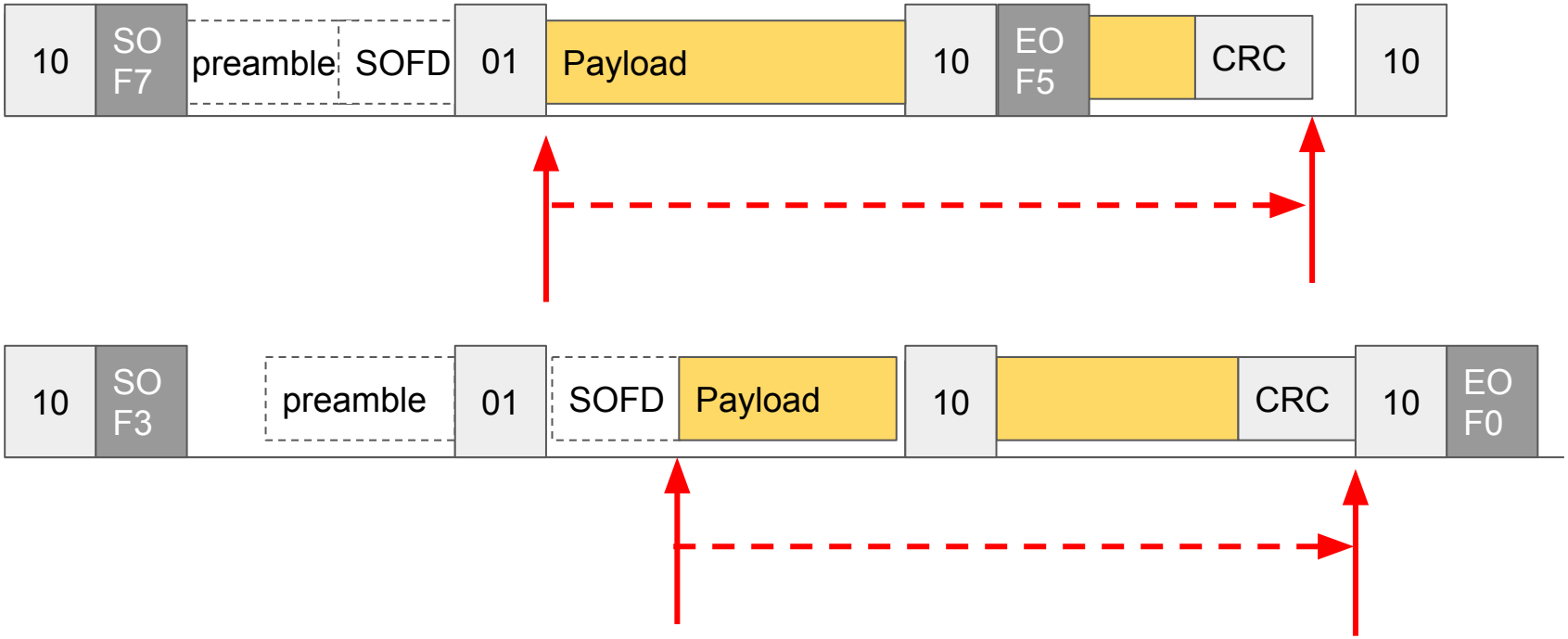
When does a frame end? *At the end of the CRC*



Our recommendations….

- When does the frame start? *At the start of the payload*
- When does the frame end? *At the end of the CRC*
- How long is the frame? *(CRC - payload) @ 10.3125G*

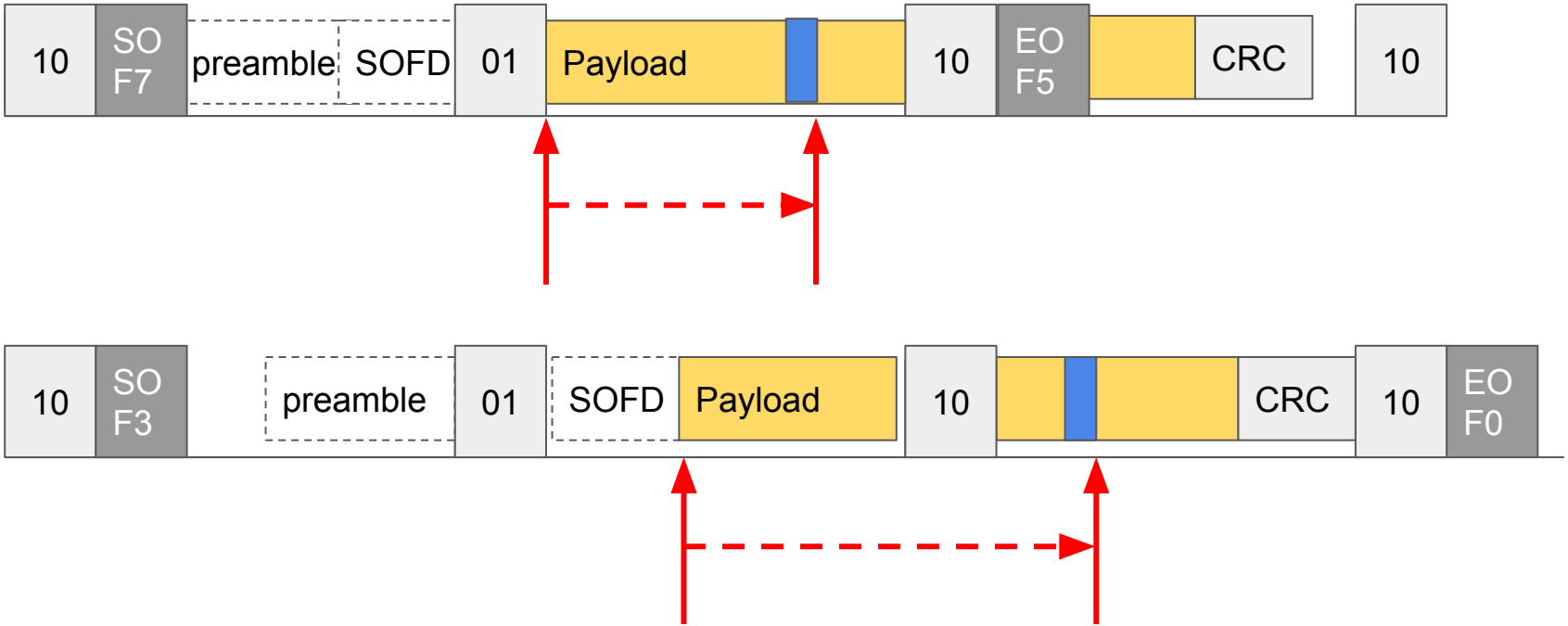
How long is the frame? *(CRC - payload) @ 10.3125Gbs*



Our recommendations….

- When does the frame start? At the start of the payload
- When does the frame end? At the end of the CRC
- How long is the frame? (CRC - payload) @ 10.3125G
- How far (ps) is an offset? *(bit offset - payload) @ 10.3125G*

How far is an offset?? *(bit offset - payload) @ 10.3125G*



Our recommendations!

- When does the frame start? **At the start of the payload**
- When does the frame end? **At the end of the CRC**
- How long is the frame? **(CRC - payload) @ 10.3125G**
- How far (ps) is an offset? **(bit offset - payload) @ 10.3125G**

Worked example

1. Message is **503**B long, excluding FCS (4B)

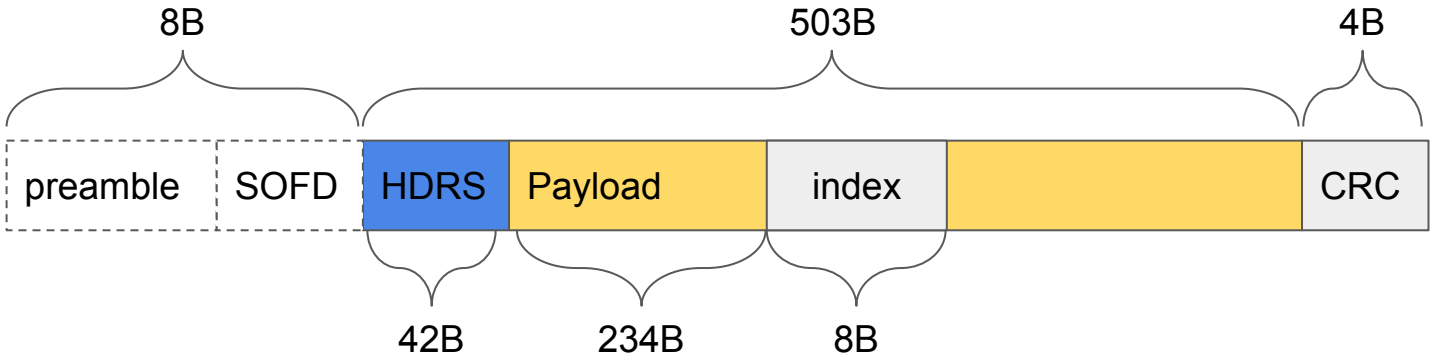
Worked example

1. Message is 503B long, excluding FCS (4B)
2. The field is 8B long and is offset is at **234**B from the IP/UDP headers.

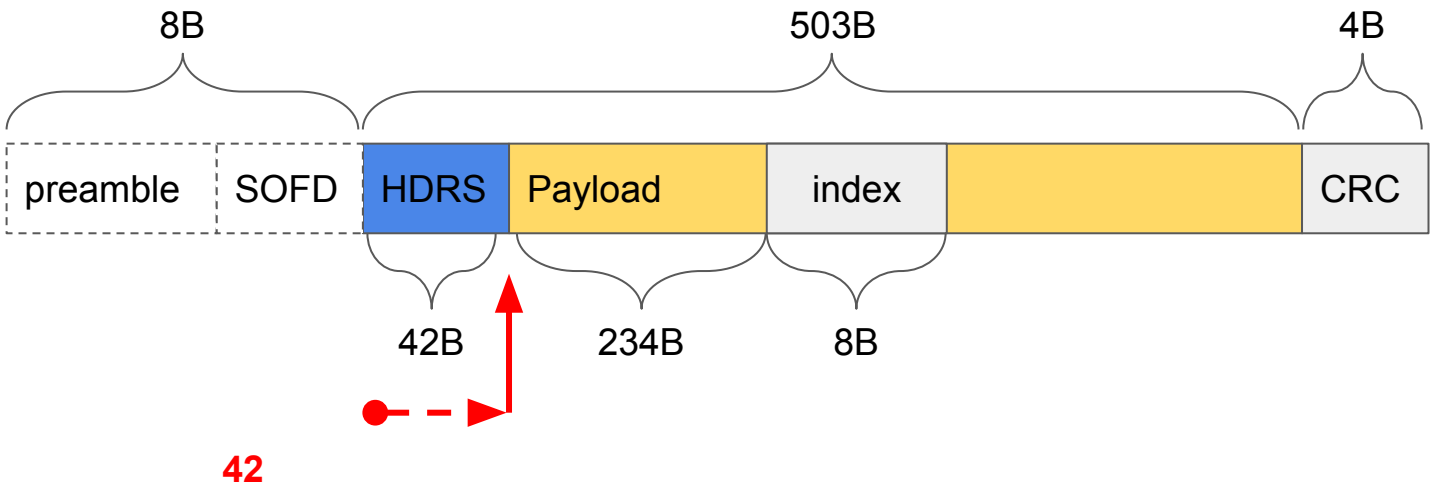
Worked example

1. Message is 503B long, excluding FCS (4B)
2. The field is 8B long and is offset is at 234B from the IP/UDP headers.
3. The Ethernet + UDP + IP headers are **42**B from the “start of frame” (not including preamble + SOFD)

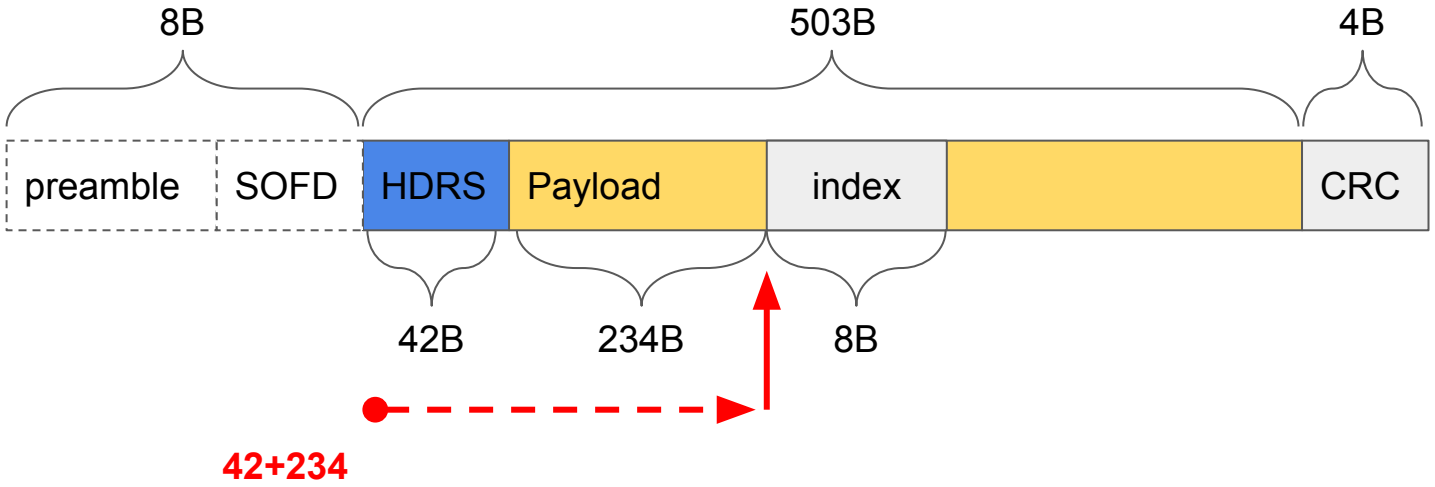
Ideal view of 507B Fame



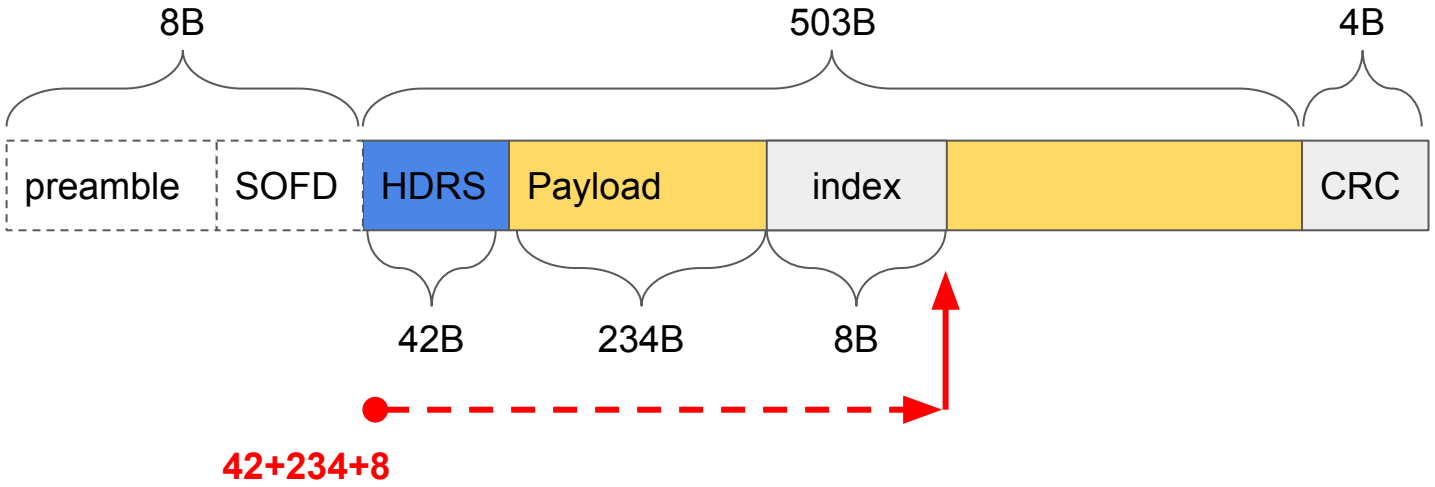
Ideal view of 507B Fame



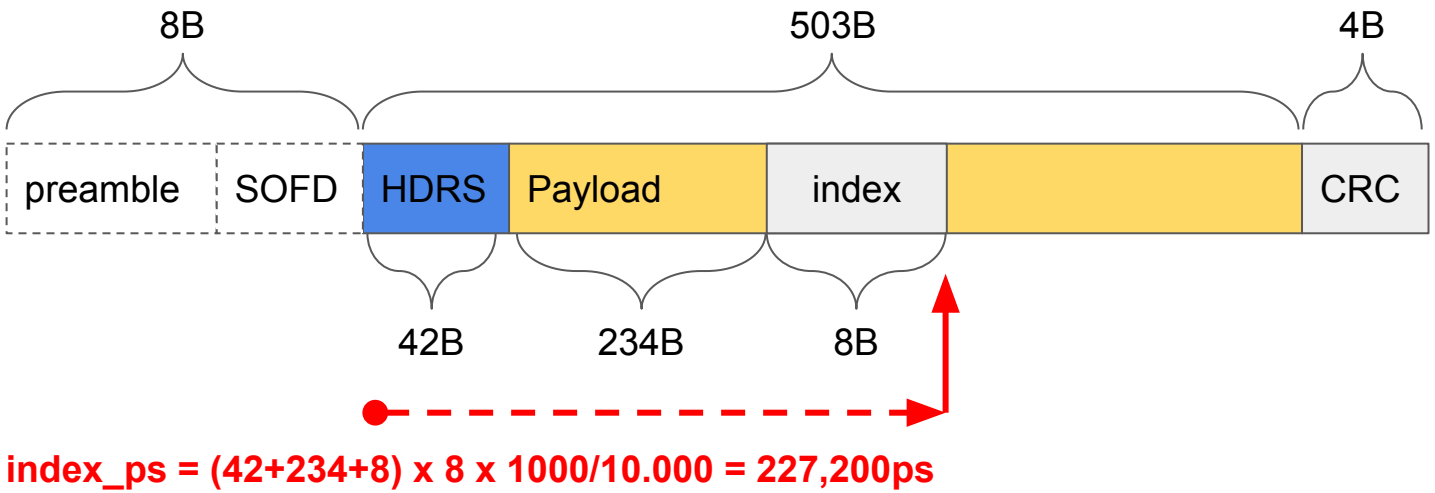
Ideal view of 507B Fame



Ideal view of 507B Fame

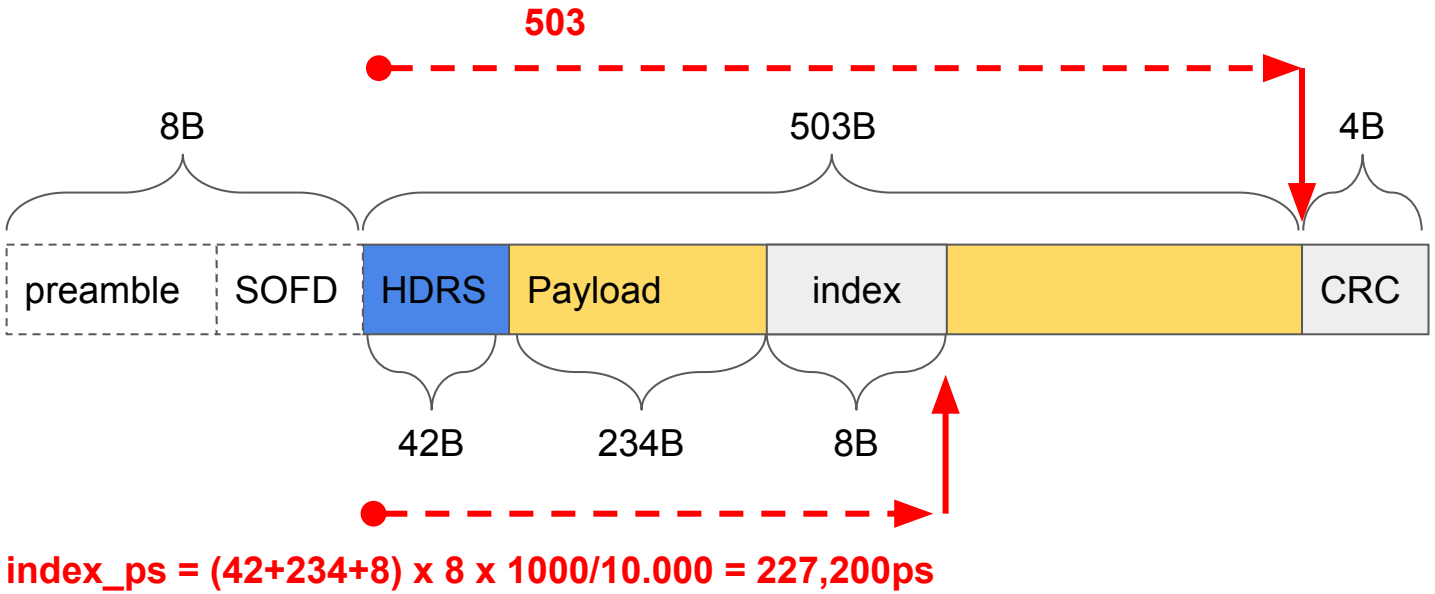


Ideal view of 507B Fame

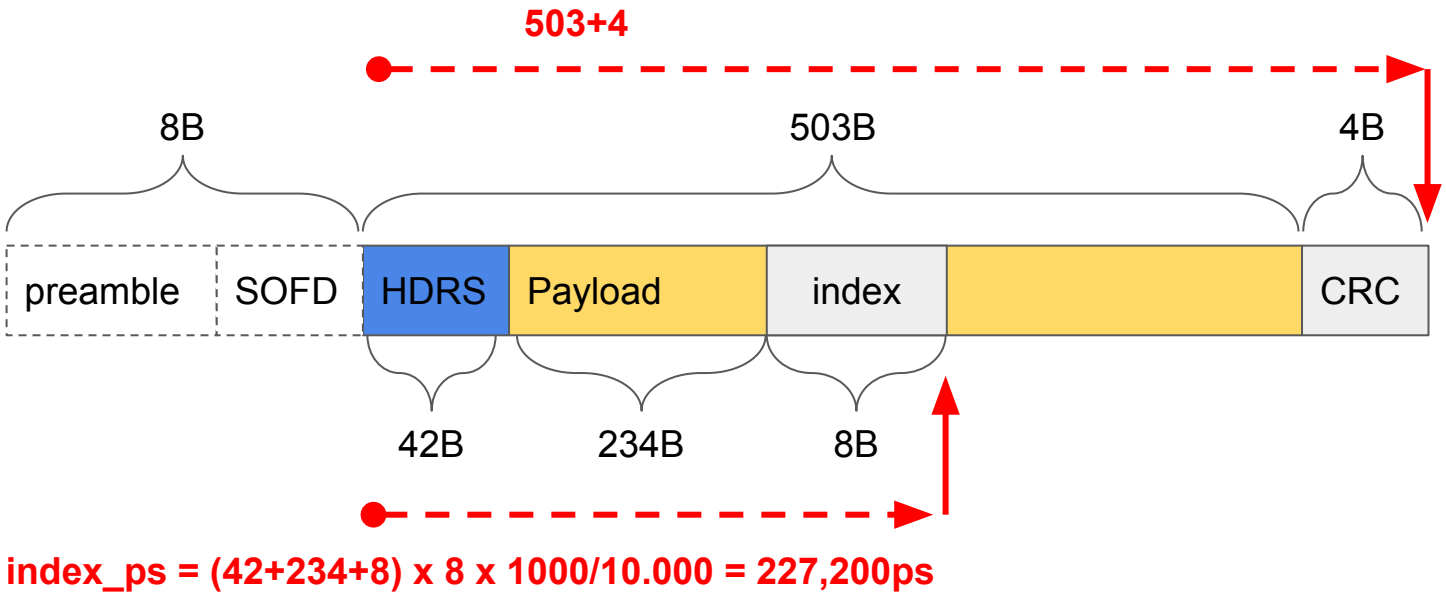


index_ps = (42+234+8) x 8 x 1000/10.000 = 227,200ps

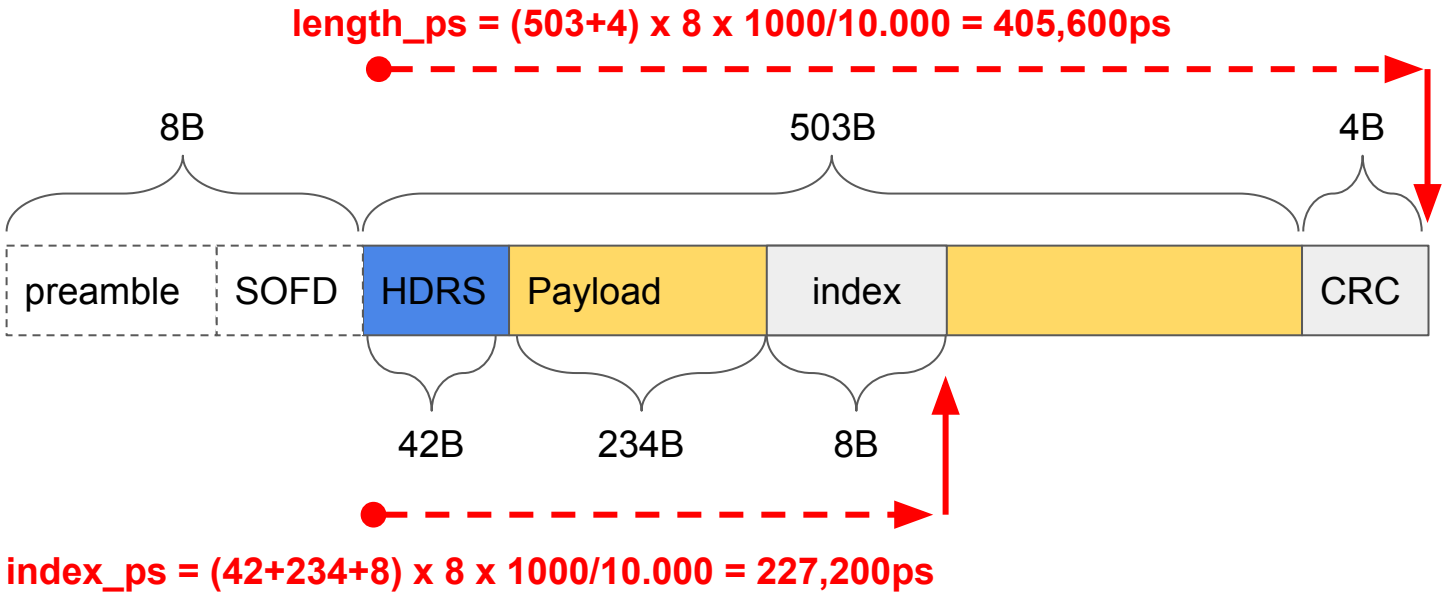
Ideal view of 507B Fame



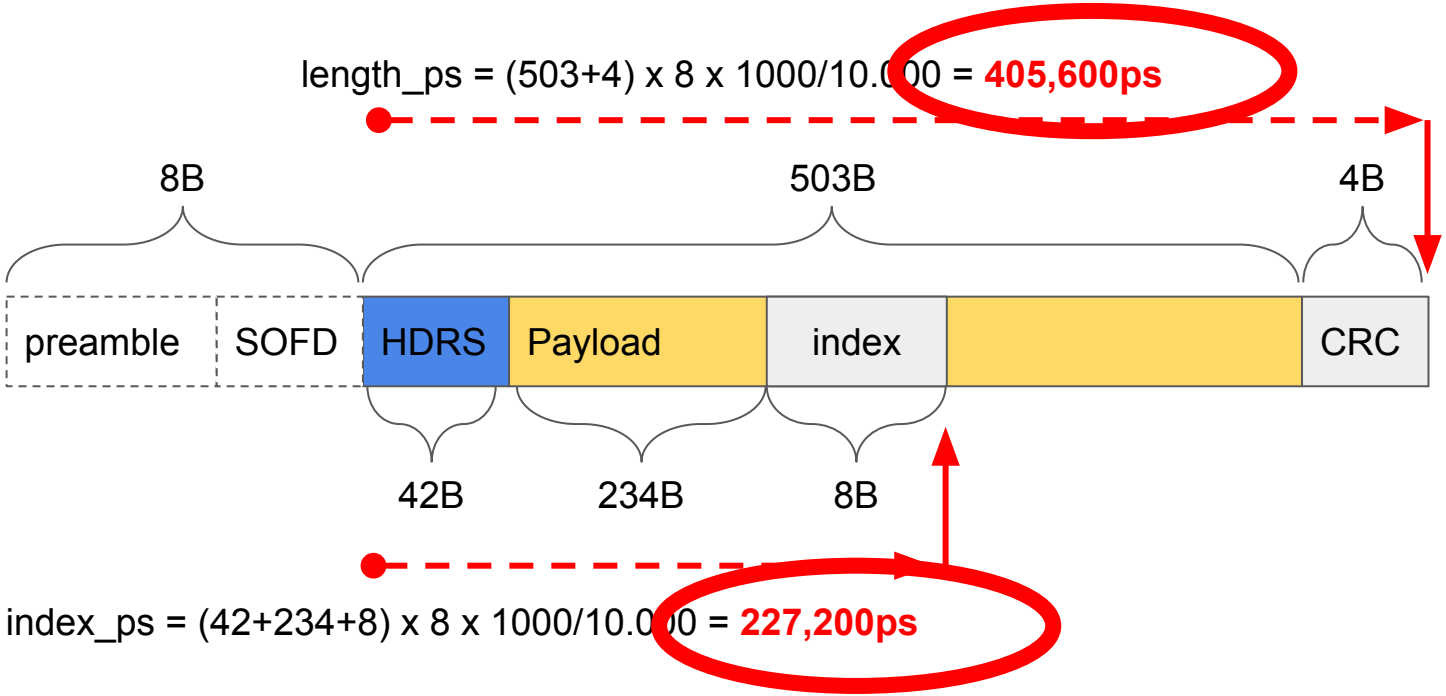
Ideal view of 507B Fame



Ideal view of 507B Fame



Ideal view of 507B Fame



Complications with PCS layer effects

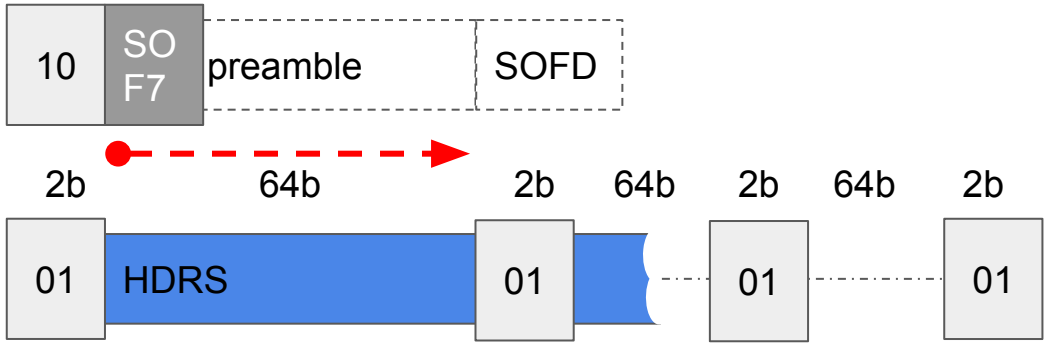
PCS **SOF7** view of a 507B fame:

Complications with PCS layer effects

PCS **SOF7** view of a 507B fame:

Complications with PCS layer effects

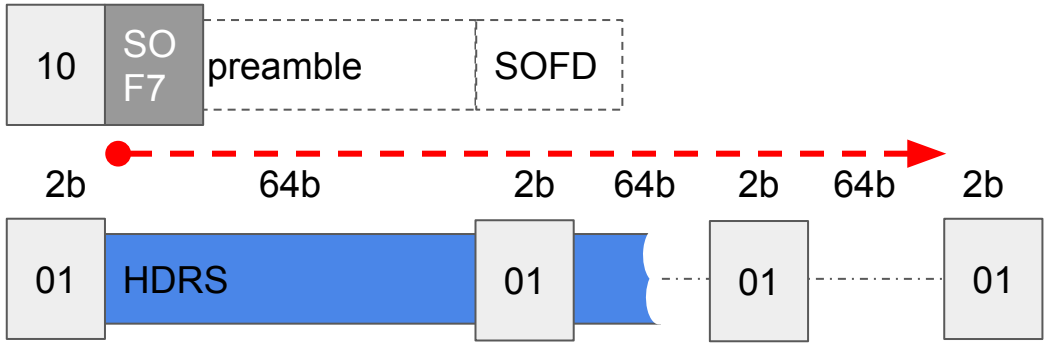
PCS **SOF7** view of a 507B fame:



64 +

Complications with PCS layer effects

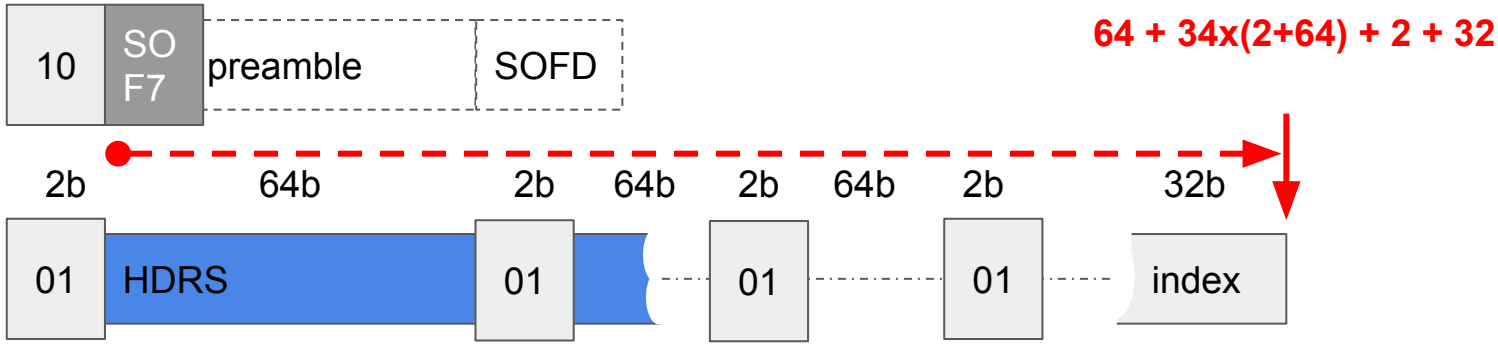
PCS **SOF7** view of a 507B fame:



64 + 34x(2+64)

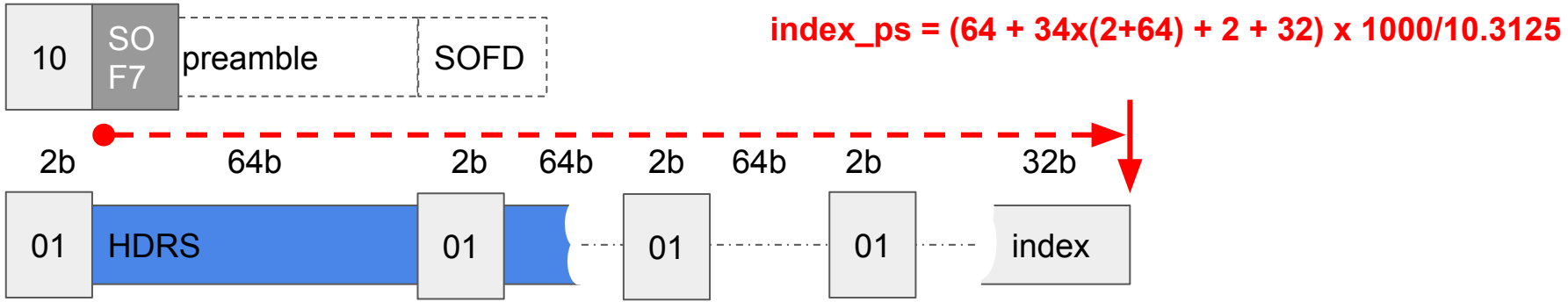
Complications with PCS layer effects

PCS **SOF7** view of a 507B fame:



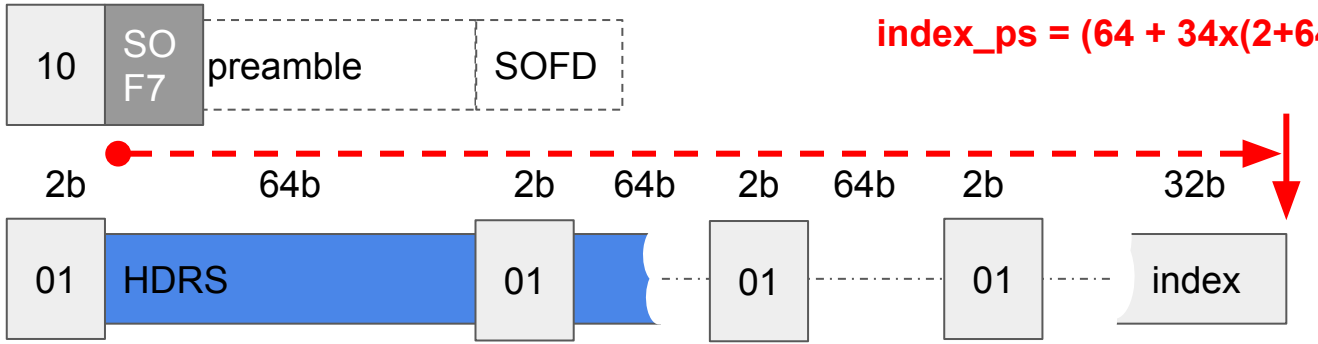
Complications with PCS layer effects

PCS **SOF7** view of a 507B fame:



Complications with PCS layer effects

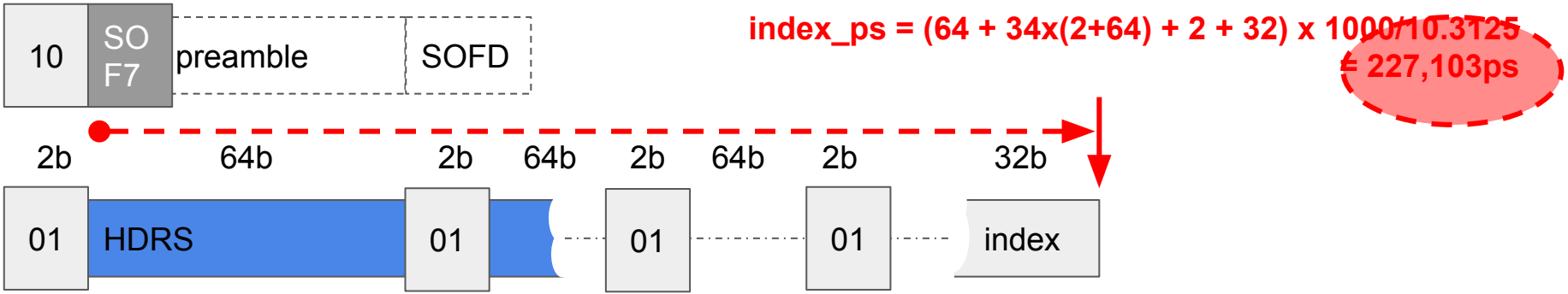
PCS **SOF7** view of a 507B fame:



index_ps = (64 + 34x(2+64) + 2 + 32) x 1000/10.3125
= 227,103ps

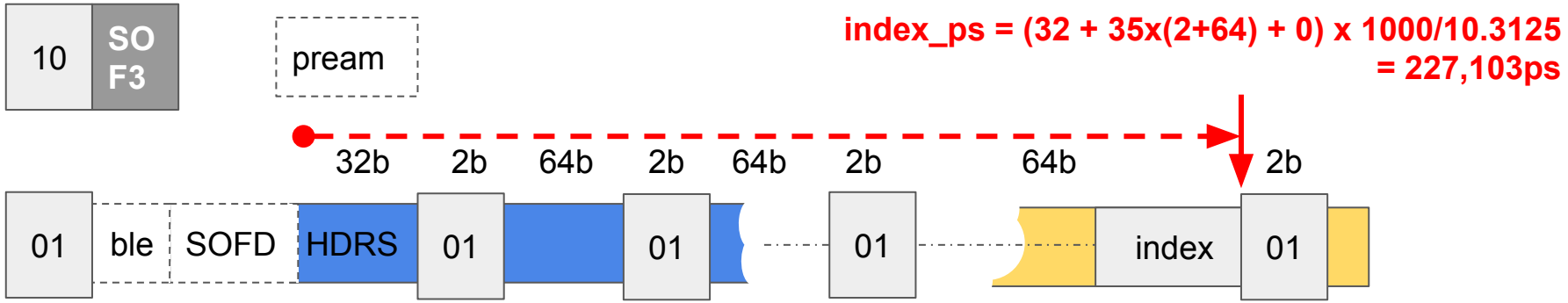
Complications with PCS layer effects

PCS **SOF7** view of a 507B fame:



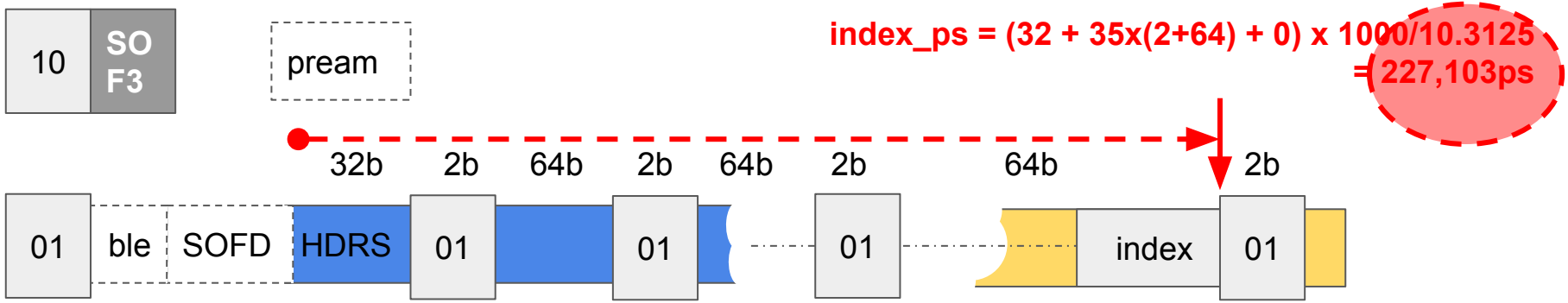
Complications with PCS layer effects

PCS **SOF3** view of a 507B fame:



Complications with PCS layer effects

PCS **SOF3** view of a 507B fame:



Results Summary for Index Offset

Message Type	507
Ideal time (ps) @ 10.000	227,200

Results Summary for Index Offset

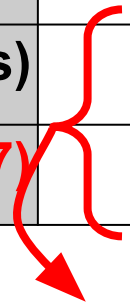
Message Type	507
Ideal time (ps) @ 10.000	227,200
PCS SOF7 time @ 10.3125 (ps)	**227,103**

Results Summary for Index Offset

Message Type	507
Ideal time (ps) @ 10.000	227,200
PCS SOF7 time @ 10.3125 (ps)	**227,103**
PCS SOF3 time @ 10.3125 (ps)	**227,103**

Results Summary for Index Offset

Message Type	507
Ideal time (ps) @ 10.000	227,200
PCS SOF7 time @ 10.3125 (ps)	**227,103**
PCS SOF3 time @ 10.3125 (ps)	**227,103**
Uncertainty (SOF3/7)	**0**



Results Summary for Index Offset

Message Type	507	64	122
Ideal time (ps) @ 10.000	227,200	44,800	44,000
PCS SOF7 time @ 10.3125 (ps)	**227,103**	44,606	43,830
PCS SOF3 time @ 10.3125 (ps)	**227,103**	44,800	44,024
Uncertainty (SOF3/7)	**0**	**194**	**194**

Results Summary for Index Offset

Message Type	507	64	122
Ideal time (ps) @ 10.000	227,200	44,800	44,000
PCS SOF7 time @ 10.3125 (ps)	**227,103**	44,606	43,830
PCS SOF3 time @ 10.3125 (ps)	**227,103**	44,800	44,024
Uncertainty (SOF3/7)	**0**	**194**	**194**

Results Summary for Packet Length

Message Type	A	B	Response
Ideal time (ps) @ 10.000	405,600	54,400	97,600
PCS SOF7 time @ 10.3125 (ps)	**406,303**	**55,079**	**98,327**
PCS SOF3 time @ 10.3125 (ps)	**406,303**	**54,303**	**98,327**
Uncertainty (SOF3/7)	**0**	**-776**	**0**

Conclusions

1. It’s harder than it looks to do measurements at the picosecond scale.

Conclusions

1. It’s harder than it looks to do measurements at the picosecond scale.
2. Vendors need to specify where/when timestamps are taken to facilitate index offset/frame length calculations

Conclusions

1. It's harder than it looks to do measurements at the picosecond scale.
2. Vendors need to specify where/when timestamps are taken to facilitate index offset/frame length calculations
3. When taking into account PCS layer effects, some index offsets/frame lengths are 776ps longer/later than expected.

Questions?

(or tick the box)