

VIDEO ENCODING

FOR OTT

COGICOM



Clement Duval

**THIS COURSE FOCUSES ON THE H.264 AND H.265 CODECS
THAT REMAIN USED FOR THE MAJORITY OF OTT DELIVERY.
SAME CONCEPTS APPLY TO LATEST CODECS (VP9 AND AV1)**

COGICOM

OUR AGENDA

Encoding & Frame types

I frame P frame B frame Video Quality

Group of Pictures

Open/Closed Fixed/Variable Video Quality

Encoding & Display

Encoding order Display order

Making OTT Fragments

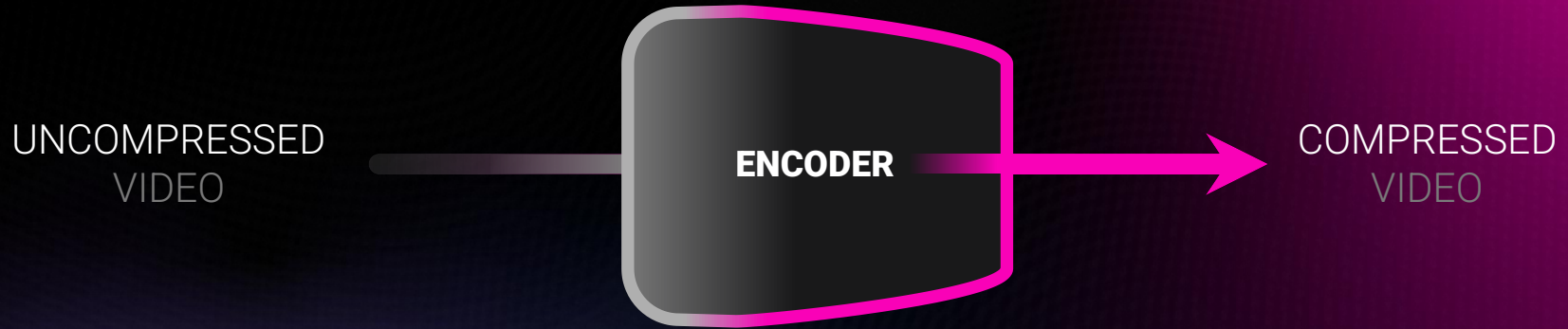
IDR frame GoP in fragments

VIDEO ENCODING

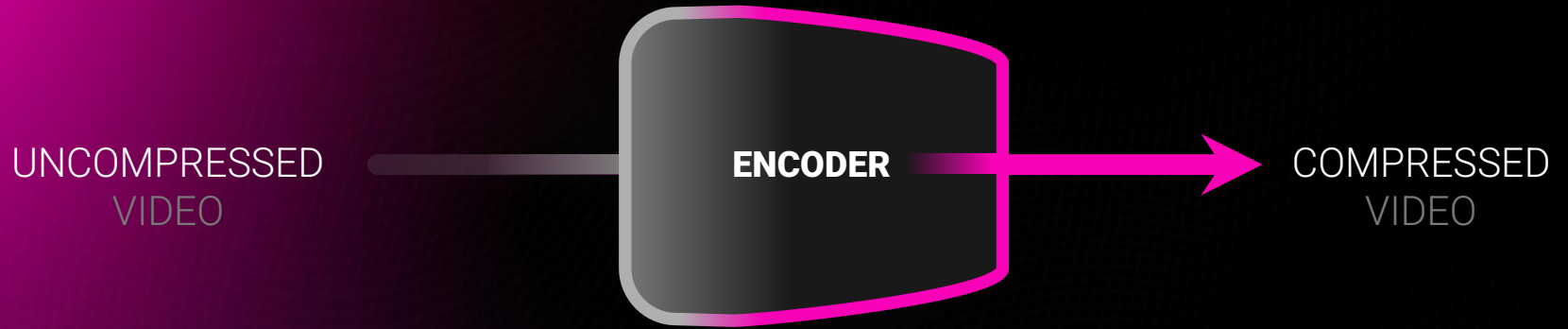
A LOSSY PROCESS

COGICOM

INTRO TO VIDEO ENCODING



INTRO TO VIDEO ENCODING



High bitrate
(ex. 620Mbps)

Bitrate / 100

Low bitrate
(ex. 6.2 Mbps)

COGICOM

VIDEO ENCODING & FRAME TYPES

INTRO TO VIDEO ENCODING

High bitrate
(ex. 620Mbps)



$\text{Bitrate} / 100$

Low bitrate
(ex. 6.2 Mbps)

$\approx \text{VIDEO QUALITY} \approx$



Video encoding is a
lossy process

COGICOM

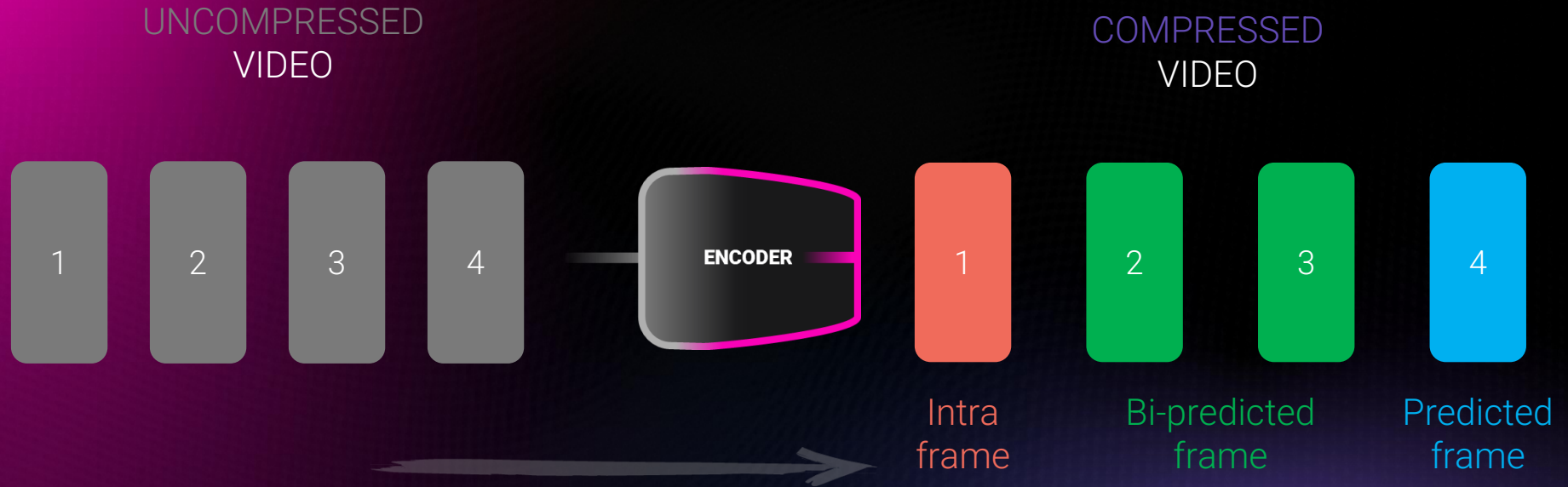
VIDEO ENCODING & FRAME TYPES

VIDEO COMPRESSION IS

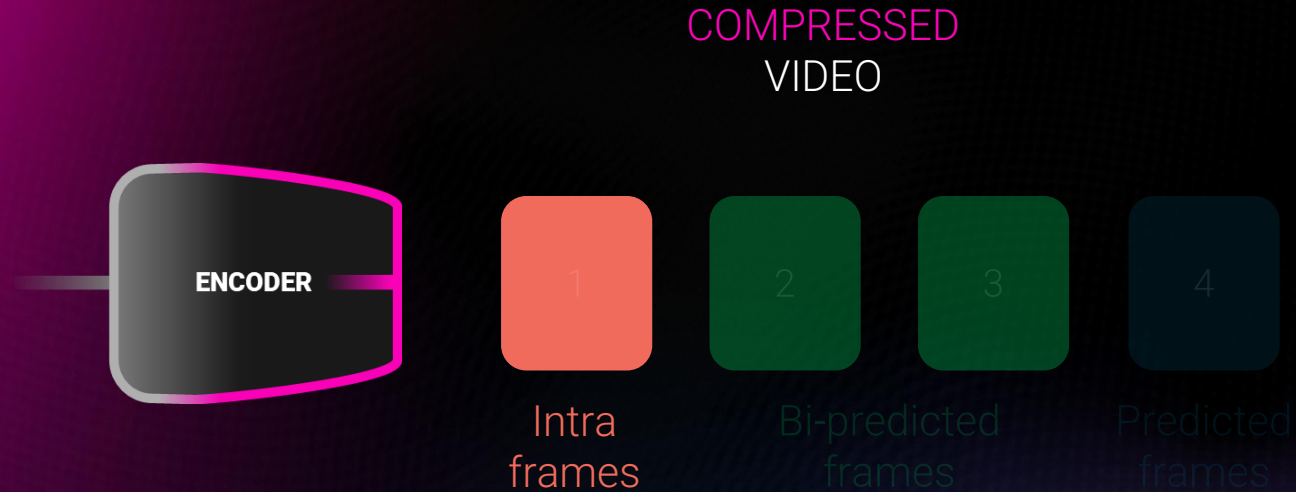
**A LOSSY PROCESS CONVERTING UNCOMPRESSED VIDEO
INTO A FORMAT THAT DRASTICALLY REDUCES BITRATE
WHILE MINIMIZING VISUAL QUALITY LOSS**

COGICOM

INTRO TO VIDEO ENCODING



INTRO TO VIDEO ENCODING

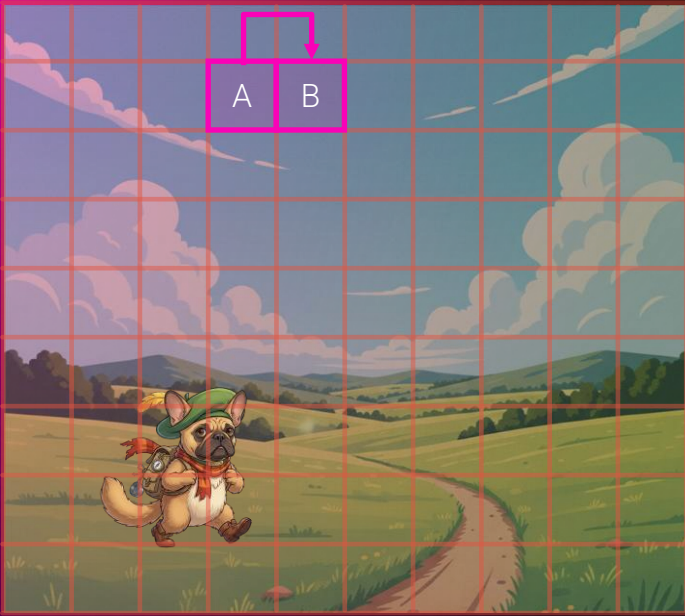


SPATIAL PREDICTION

THE INTRA CODED FRAME

COGICOM

INTRA PREDICTION: INTRA (KEY) FRAMES



Divide pictures into blocks

Neighboring blocks are likely to be similar

Encodes the residual ($A - B$)

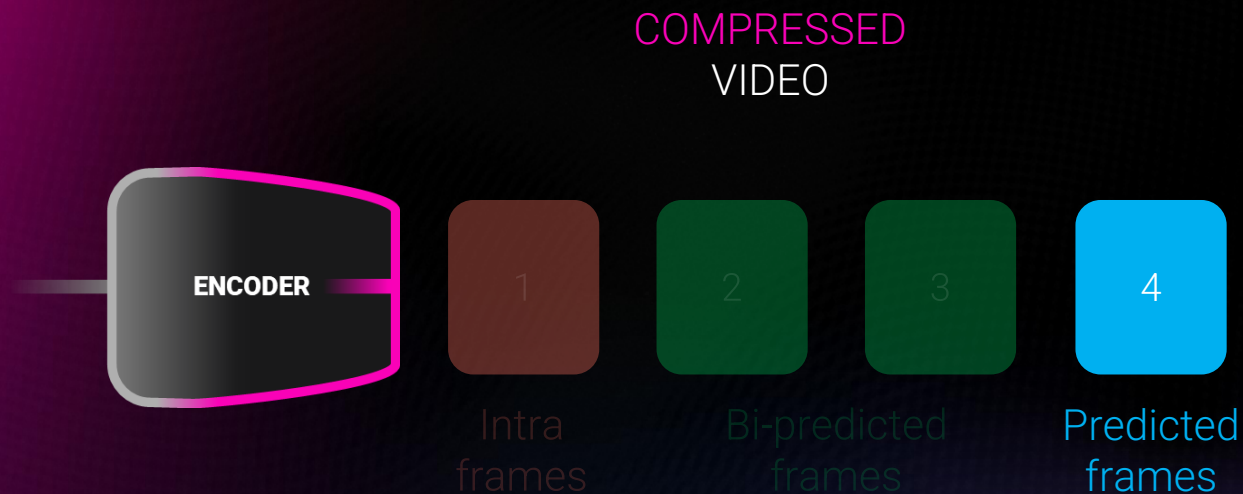
Intra-frame coding

TEMPORAL PREDICTION

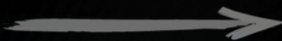
THE PREDICTED & BI-PREDICTED FRAMES

COGICOM

TEMPORAL PREDICTION



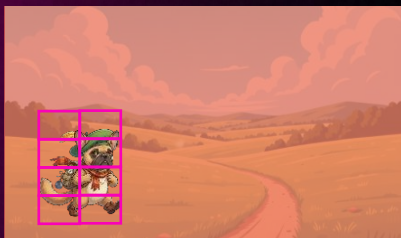
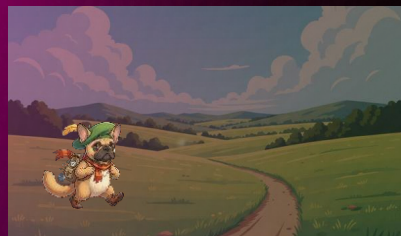
TEMPORAL PREDICTION: ENCODING 4 PICTURES



COGICOM

TEMPORAL PREDICTION

TEMPORAL PREDICTION: ENCODING 4 PICTURES



INTRA

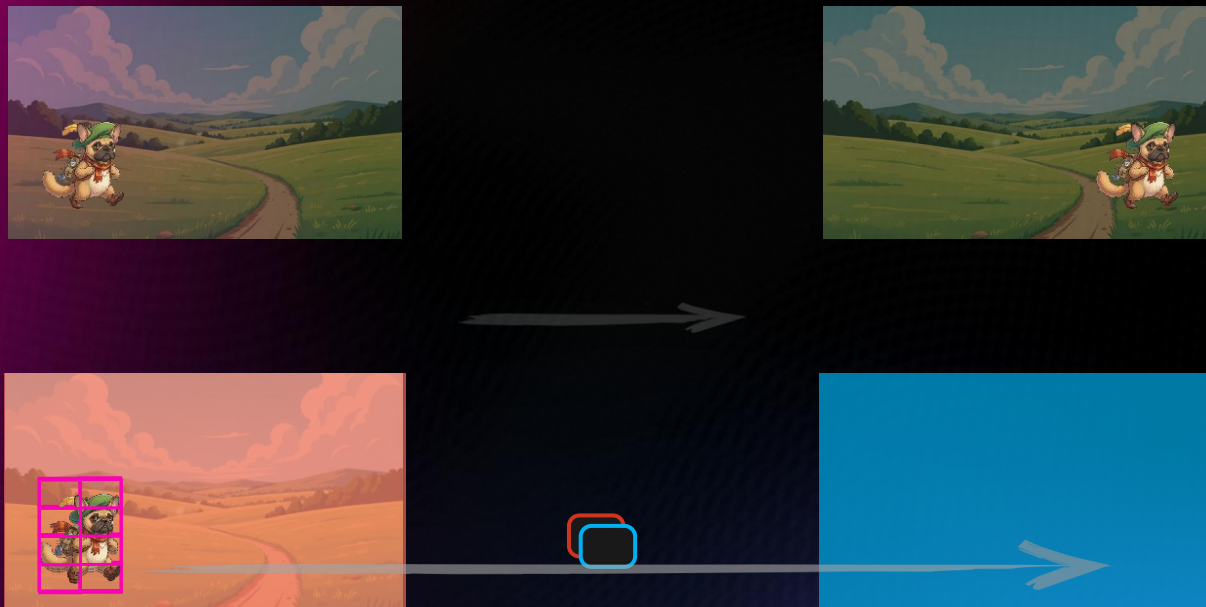
COGICOM

Copying the
landscape

PREDICTED

TEMPORAL PREDICTION

TEMPORAL PREDICTION: ENCODING 4 PICTURES



INTRA

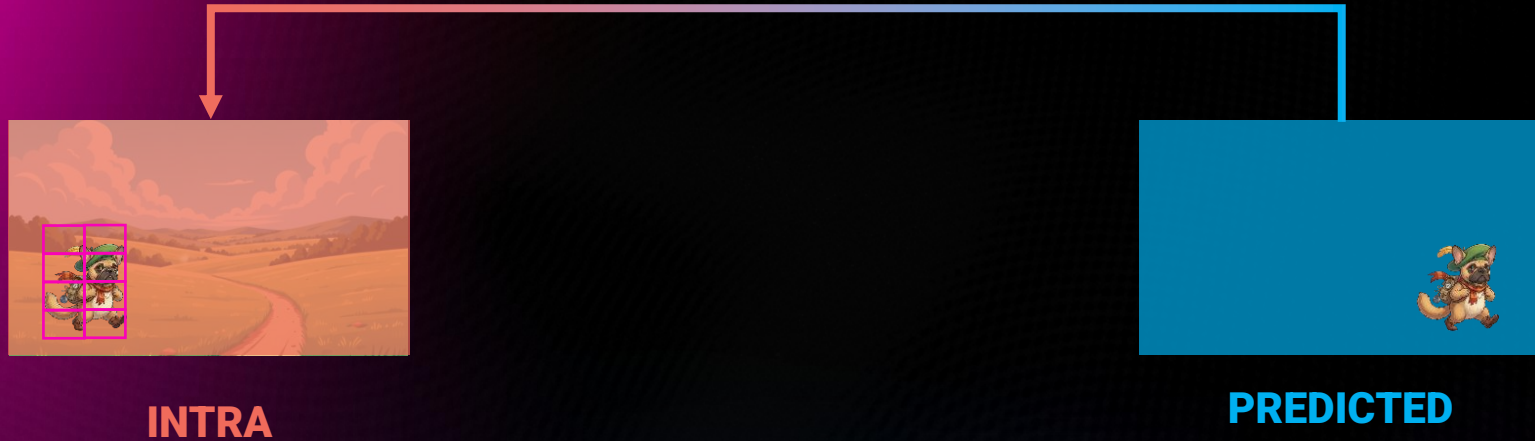
COGICOM

Copying/translating
the moving object

PREDICTED

TEMPORAL PREDICTION

TEMPORAL PREDICTION: ENCODING 4 PICTURES

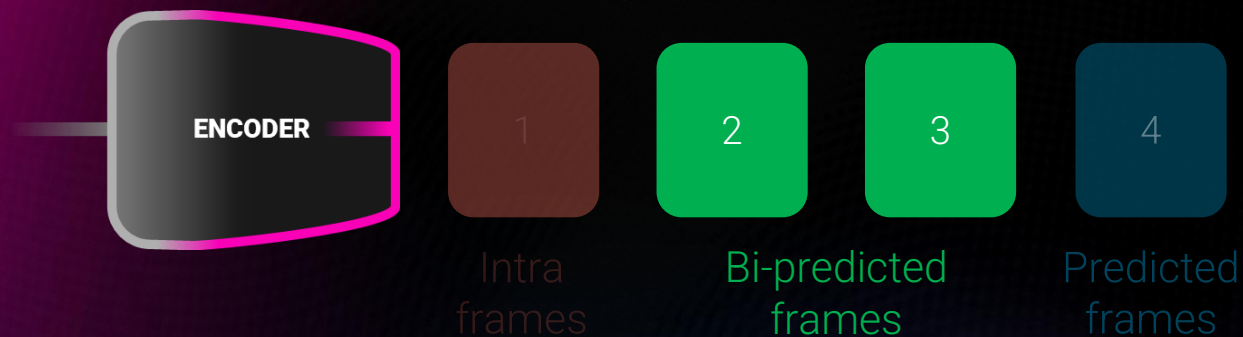


The predicted frame uses the past I frame
as a reference to be encoded & decoded

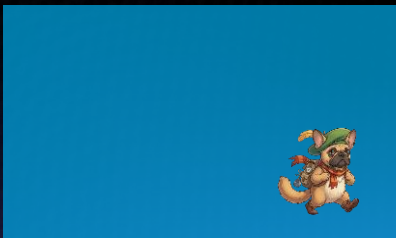
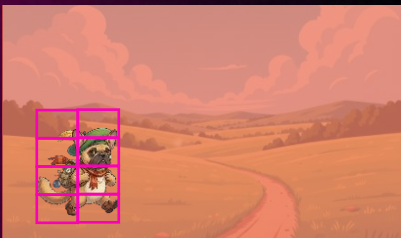
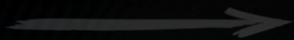
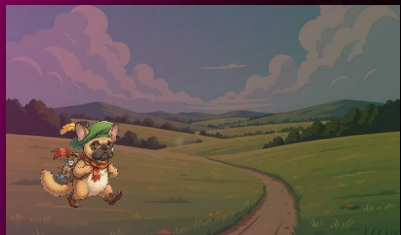
**PREDICTED FRAMES USE A
SINGLE (PAST) REFERENCE FRAME
TO BE ENCODED & DECODED**

COGICOM

THE BI-PREDICTED FRAME



TEMPORAL PREDICTION: ENCODING 4 PICTURES

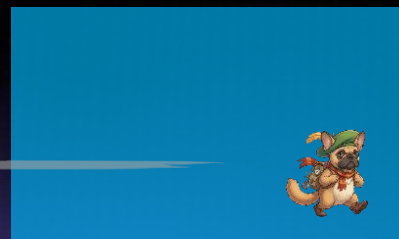
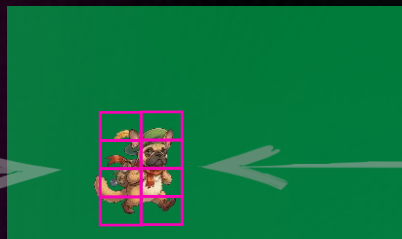
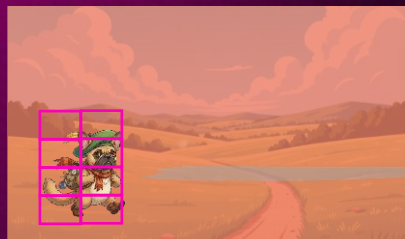


COGICOM¹

4

TEMPORAL PREDICTION

TEMPORAL PREDICTION : ENCODING 4 PICTURES



¹
COGICOM

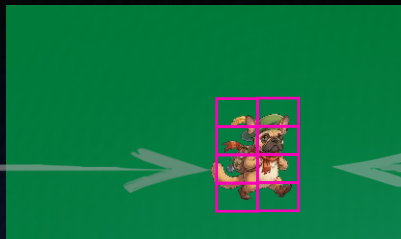
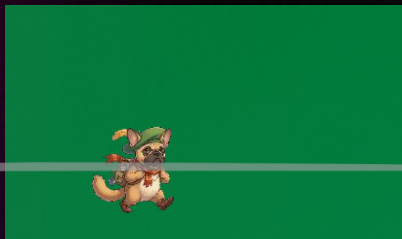
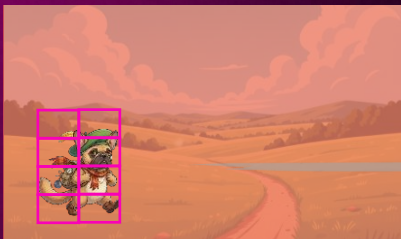
2

3

4

TEMPORAL PREDICTION

TEMPORAL PREDICTION: ENCODING 4 PICTURES



¹
COGICOM

2

3

4

TEMPORAL PREDICTION

BI-PREDICTED FRAMES USE
TWO REFERENCE FRAMES
TO BE ENCODED & DECODED

COGICOM

SUMMARY

SPATIAL & TEMPORAL PREDICTION

COGICOM

PICTURE TYPES: SUMMARY

I Frame

P Frame

B Frame

Reduces spatial
redundancy

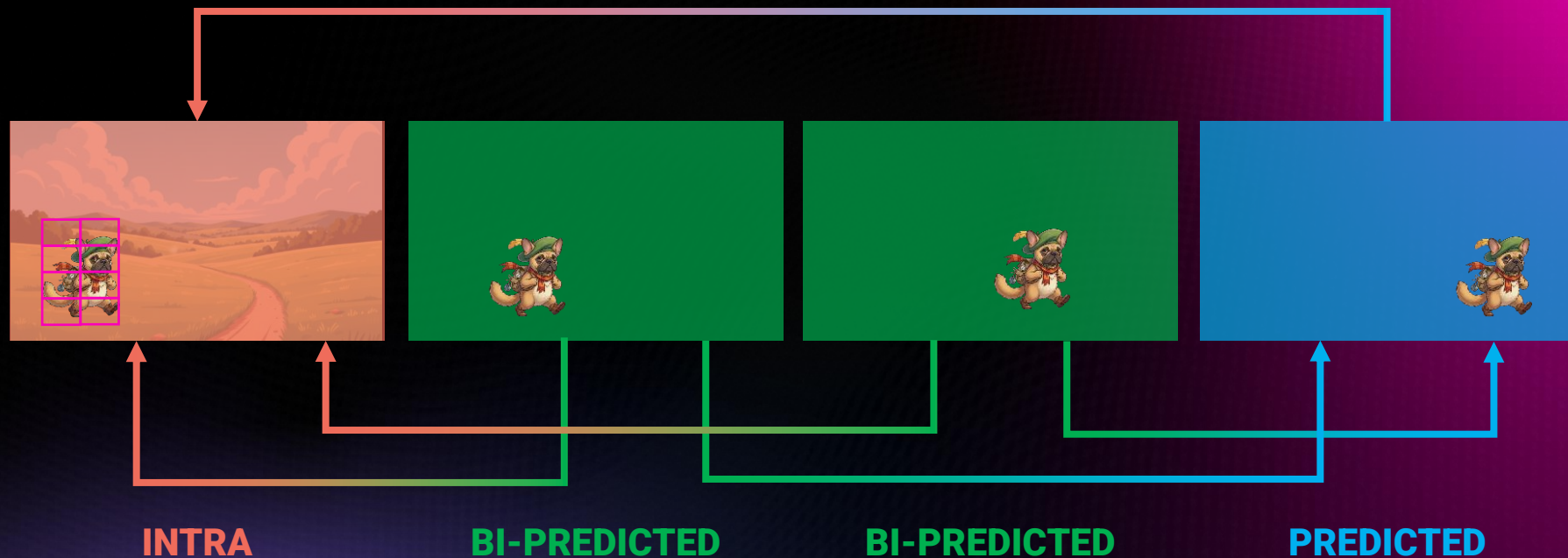
Reduces temporal redundancy

Encoded and decoded
alone

Uses 1 temporal reference

Uses 2 temporal references

TEMPORAL PREDICTION & REFERENCES



COGICOM

TEMPORAL PREDICTION

ENCODING EFFICIENCY

& FRAME TYPES

COGICOM

I FRAMES ARE THE LEAST EFFICIENT FRAMES

- **THEY ONLY USE SPATIAL PREDICTION**

B FRAMES ARE THE MOST EFFICIENT FRAMES

- **THEY REDUCE TEMPORAL REDUNDANCY WITH 2 REFERENCES**

ENCODING EFFICIENCY

Not at
scale!

One picture uncompressed HD

3 M Bytes

I frame

500kBytes

P frame

150kBytes

B frame

20kBytes

CODING EFFICIENCY

Low

High

COGICOM

ENCODING EFFICIENCY

I FRAMES ARE THE LEAST EFFICIENT FRAMES

- **THEY ONLY USE SPATIAL PREDICTION**

B FRAMES ARE THE MOST EFFICIENT FRAMES

- **THEY REDUCE TEMPORAL REDUNDANCY WITH 2 REFERENCES**

Quiz 1

COGICOM

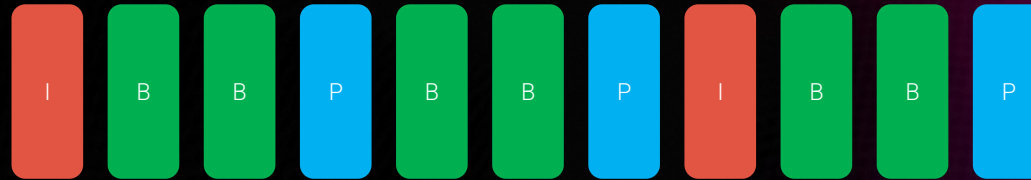
STREAM STRUCTURE

NUMBER OF CONSECUTIVE B FRAMES

COGICOM

REAL-WORLD ENCODED STREAM

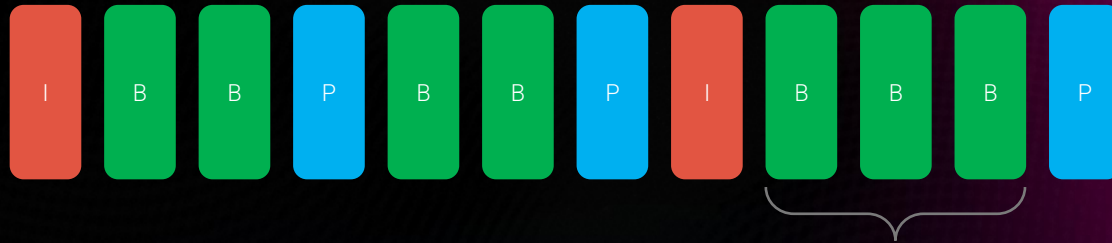
Number of consecutive B frames = 2



In the real world, encoded stream is defined by
the number of consecutive B frames used

REAL-WORLD ENCODED STREAM

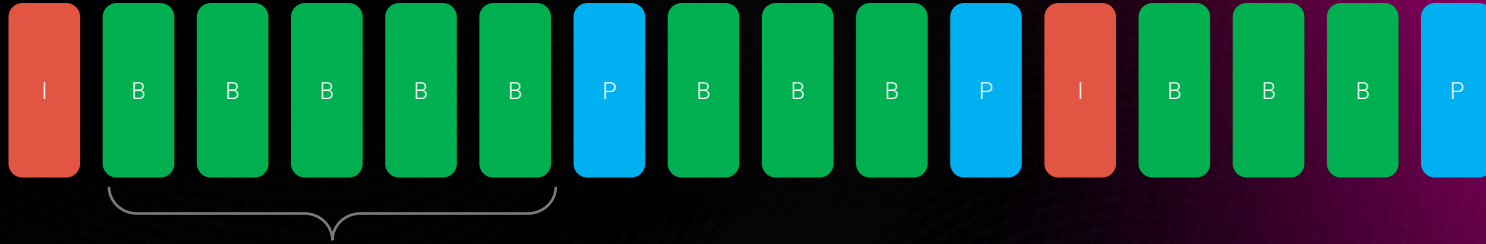
Maximum number of consecutive B frames = 3



In the real world, stream complexity is defined by
the maximum number of consecutive B frames

REAL-WORLD ENCODED STREAM

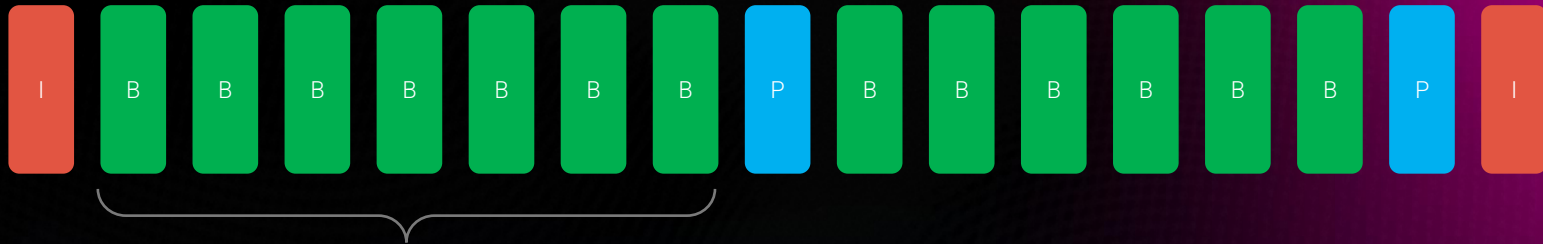
Maximum number of consecutive B frames = 5



In the real world, stream complexity is defined by
the maximum number of consecutive B frames

REAL-WORLD ENCODED STREAM

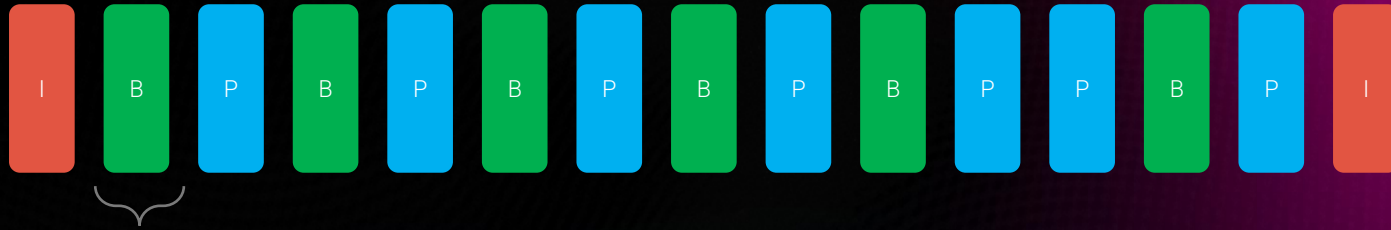
Maximum number of consecutive B frames = 7



In the real world, stream complexity is defined by
the maximum number of consecutive B frames

REAL-WORLD ENCODED STREAM

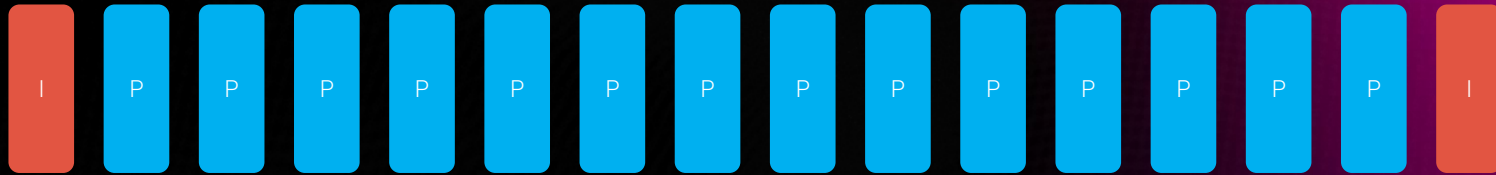
Maximum number of consecutive B frames = 1



In the real world, stream complexity is defined by
the maximum number of consecutive B frames

REAL-WORLD ENCODED STREAM

Maximum number of consecutive B frames = 0



In the real world, stream complexity is defined by
the maximum number of consecutive B frames

**ENCODING QUALITY IS DEFINED BY THE
MAXIMUM NUMBER OF CONSECUTIVE B FRAMES
WHICH IS THE NUMBER OF B FRAMES
BETWEEN TWO ANCHOR FRAMES**

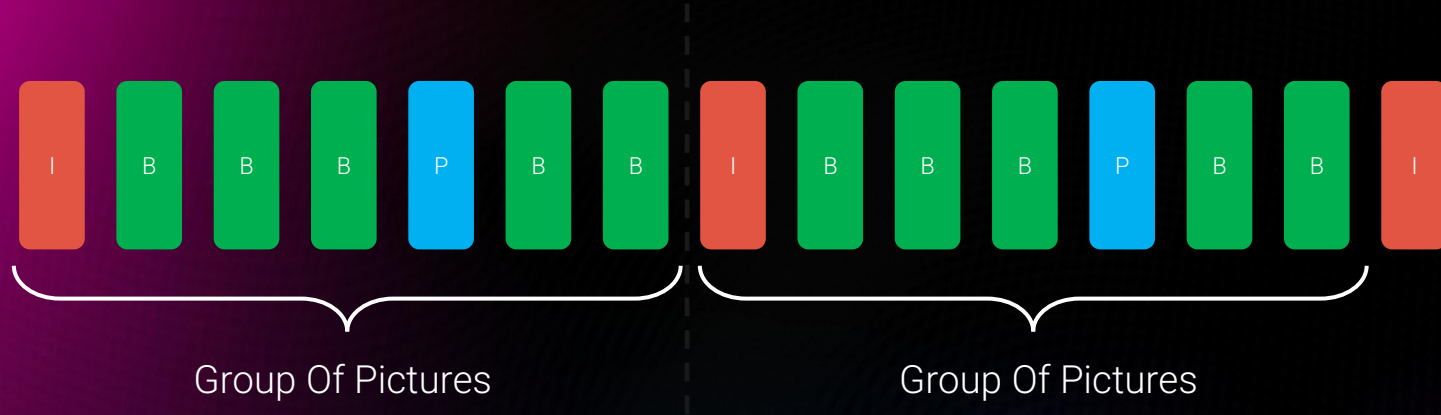
COGICOM

GROUP OF PICTURES

RESETTING THE TEMPORAL PREDICTION PROCESS

COGICOM

INJECTING I FRAMES PERIODICALLY: THE GOP



GoP is a time interval between two I frames

Expressed in seconds or in number of frames

THE NEED FOR GROUP OF PICTURES

Channel Zapping

Packet loss

Scene changes

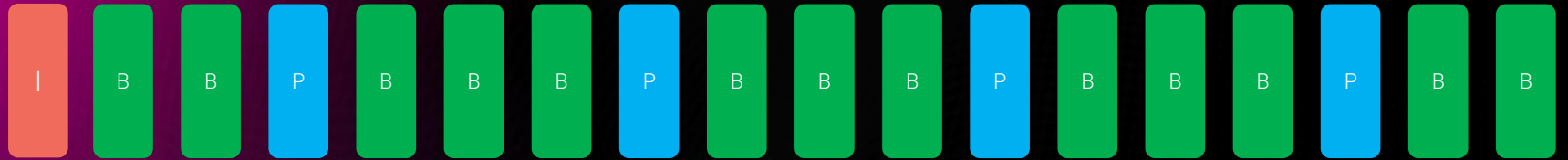
THE NEED FOR GROUP OF PICTURES

Channel Zapping

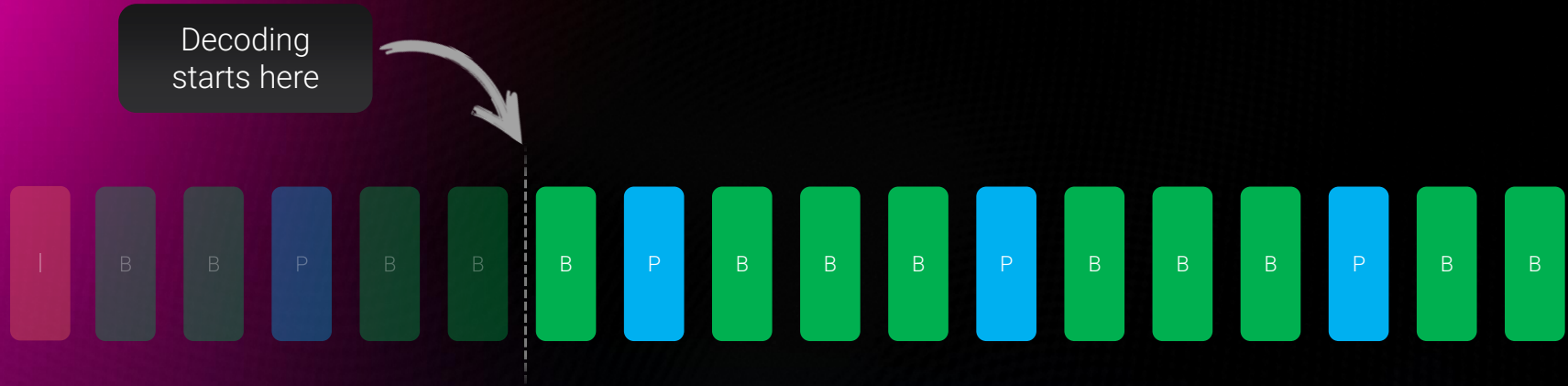
COGICOM

GROUP OF PICTURES

INTER ONLY & CHANNEL ZAPPING

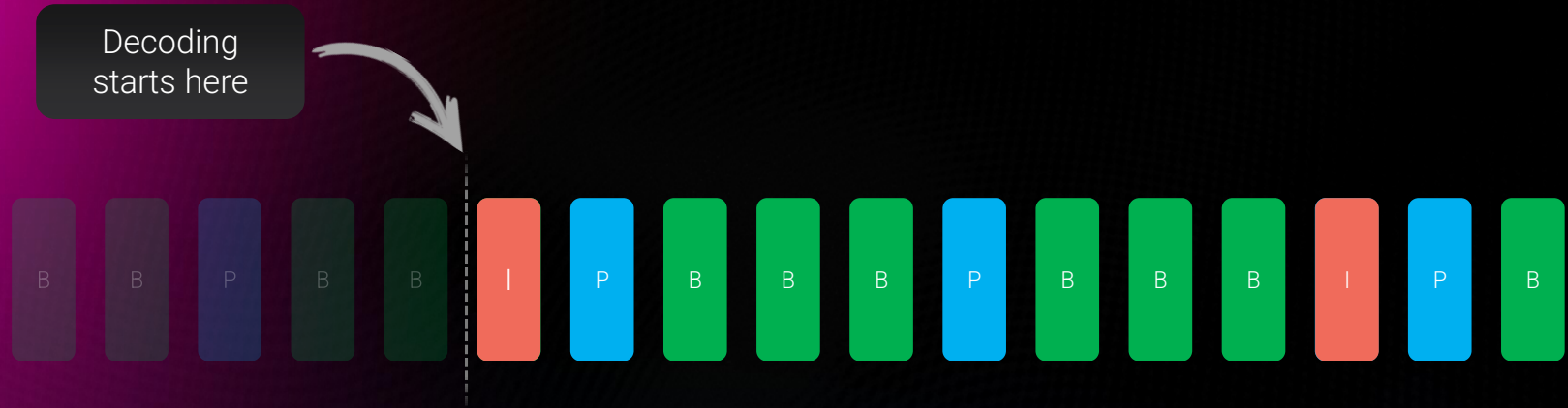


INTER ONLY & CHANNEL ZAPPING



If using inter frames only, a live decoder cannot start decoding in the middle of the live transmission

SOLUTION: INSERT I FRAMES AT REGULAR INTERVAL



Inserting I frames regularly allows for decoding in the middle of a live stream.

THE NEED FOR GROUP OF PICTURES

Channel Zapping

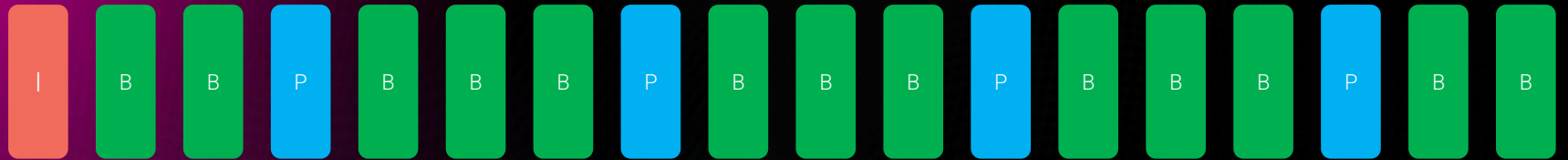
Packet loss

Scene changes

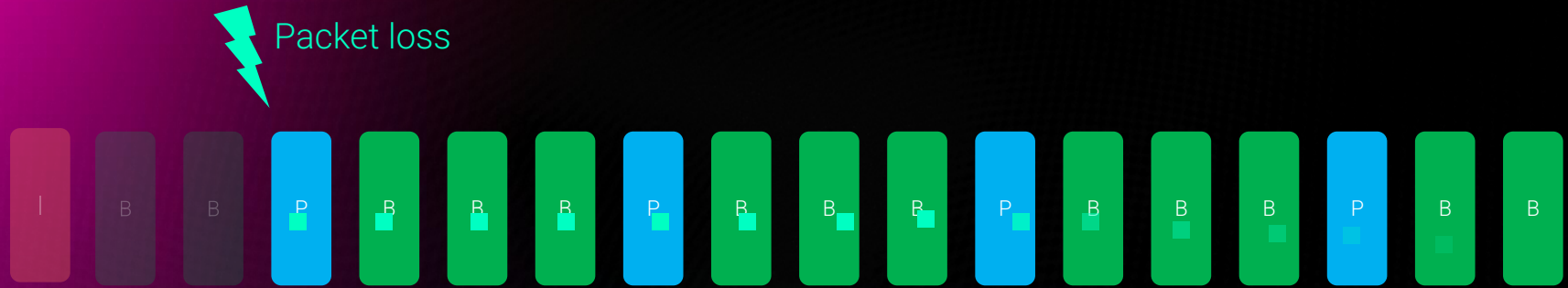
THE NEED FOR GROUP OF PICTURES

Packet loss

INTER ONLY & PACKET LOSS

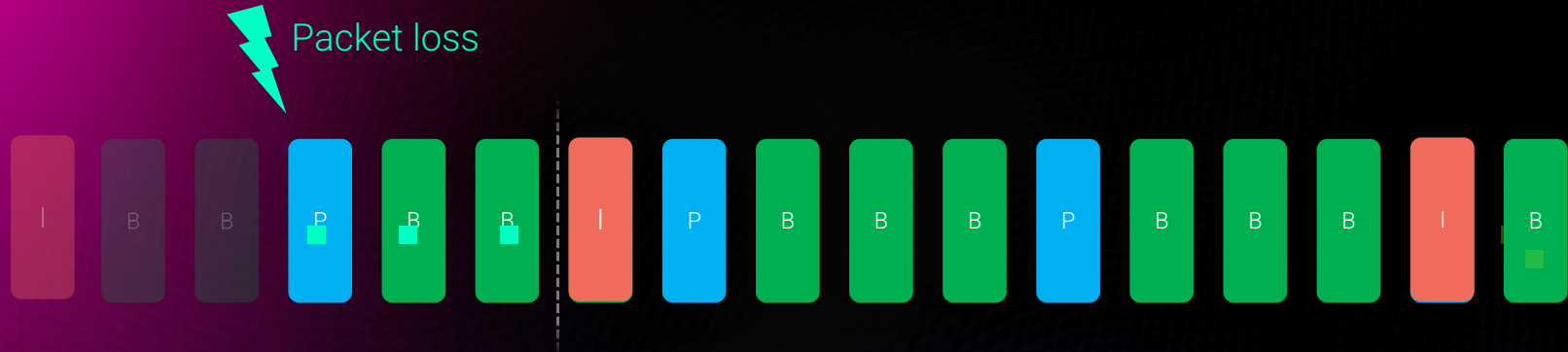


INTER ONLY & PACKET LOSS



With inter frames only, video data lost during transmission may propagate indefinitely

INTER ONLY & PACKET LOSS



Inserting I frames regularly limits error propagation throughout the live stream.

THE NEED FOR GROUP OF PICTURES

Channel Zapping

Packet loss

Scene changes

THE NEED FOR GROUP OF PICTURES

Scene changes

Coming up!

COGICOM

GROUP OF PICTURES

GROUP OF PICTURES IS THE INTERVAL BETWEEN TWO I FRAMES

RESET THE TEMPORAL PREDICTION SO THAT DECODERS CAN:

- **DECODE A LIVE STREAM AS IT IS RUNNING**
- **RECOVER FROM A LIVE STREAM IMPAIRMENT**
- **MAKE SCENE CHANGE ENCODING MORE EFFICIENT**

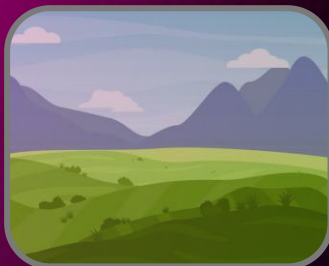
VARIABLE vs FIXED GOP

OPTIMIZING THE I FRAME PLACEMENT

COGICOM

UNDERSTANDING SCENE CHANGES

SCENE A



SCENE B



≠

Very low temporal redundancy

COGICOM

FIXED vs VARIABLE GOP

UNDERSTANDING SCENE CHANGES

SCENE A

SCENE B

LET'S USE **TEMPORAL PREDICTION**
TO ENCODE THIS SCENE CHANGE

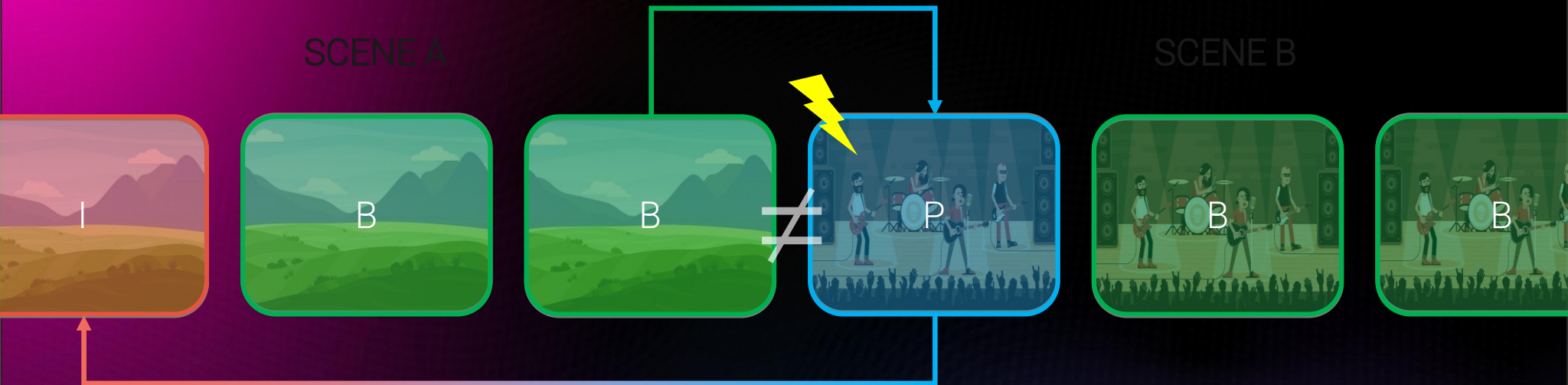
≠

Very low temporal redundancy

COGICOM

FIXED vs VARIABLE GOP

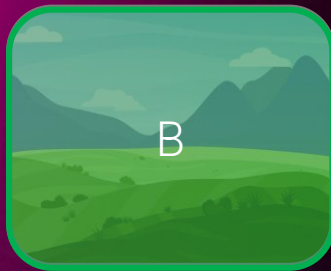
SCENE CHANGE & TEMPORAL PREDICTION



Using a P or B around the scene change would
yield **poor encoding efficiency**

OPTIMIZING ENCODE AT SCENE CHANGE WITH I FRAME

SCENE A



SCENE B



Inserting an **I frame** at scene change
provides better visual result

Also called:

- *Dynamic GoP*
- *Scene cut detection*
- *Adaptive GoP*



WHEN USING **VARIABLE GOP,**
THE ENCODER DECIDES THE **BEST MOMENT TO INSERT AN I FRAME**

- **BETTER ALLOCATE BITS ACCORDING TO THE CONTENT**
- **INCREASES VIDEO QUALITY**
- **INCREASES ENCODING/DECODING COMPLEXITY**

COGICOM

VARIABLE GOP IS DEFINED BY A **DURATION SETTING**
THIS SETTING BECOMES THE **MAX GOP SIZE**, AT WHICH
THE ENCODER **FORCES THE INSERTION OF AN I FRAME**

COGICOM

VARIABLE GOP DURATION & SCENE CHANGE

5 fps

GoP size = 1 sec
variable GoP = yes



COGICOM

FIXED vs VARIABLE GOP

FIXED GOP DURATION

5 fps

GoP size = 1 sec
variable GoP = no



COGICOM

FIXED vs VARIABLE GOP

FIXED GOP

**FORCES I FRAME INSERTION AT REGULAR TIME INTERVAL
REGARDLESS OF THE CONTENT**

COGICOM

GOP TYPES: INTEROP & VIDEO QUALITY

FIXED GOP

The encoder inserts I frames at fixed time interval

*Optimized for
interoperability*

VARIABLE GOP

The encoder chooses the best moment to insert I frames

*Optimized for
Video Quality*

GOP TYPES: INTEROP & VIDEO QUALITY

FIXED GOP

*Optimized for
interoperability*

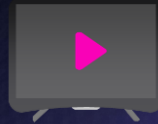
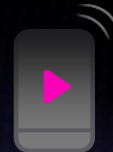
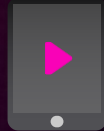
VARIABLE GOP

*Optimized for
Video Quality*

FIXED GOP USED ON MOST OTT CONFIGURATION

FIXED GOP

Optimized for
interoperability



COGICOM

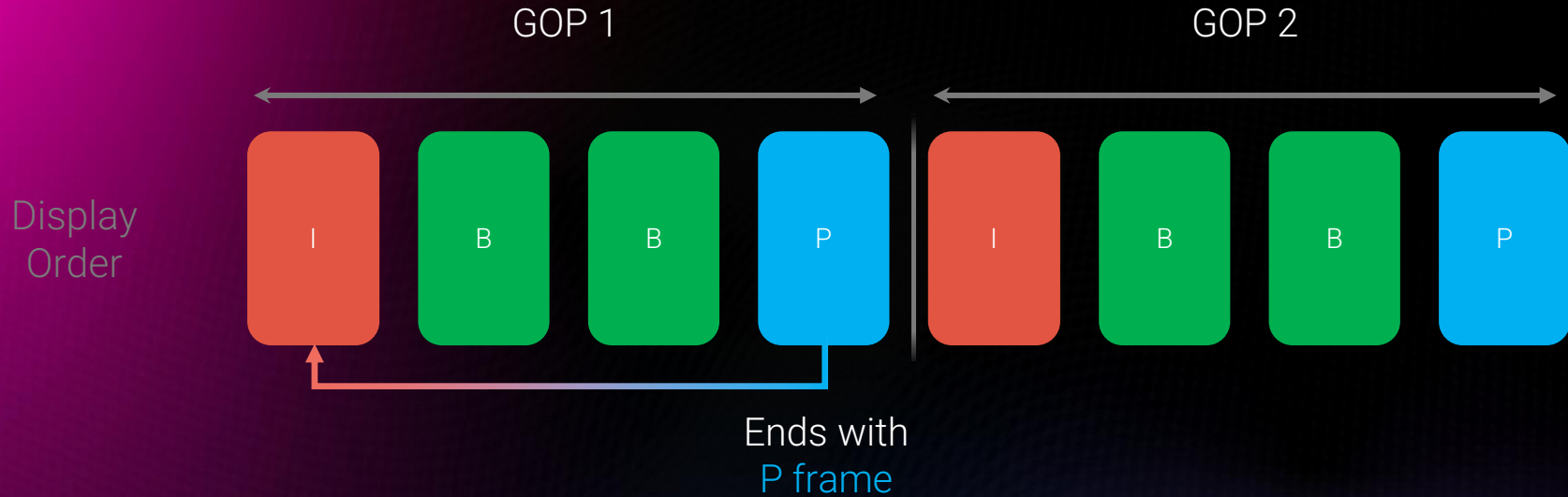
FIXED vs VARIABLE GOP

CLOSED vs OPEN GOP

INDEPENDENT OR NOT

COGICOM

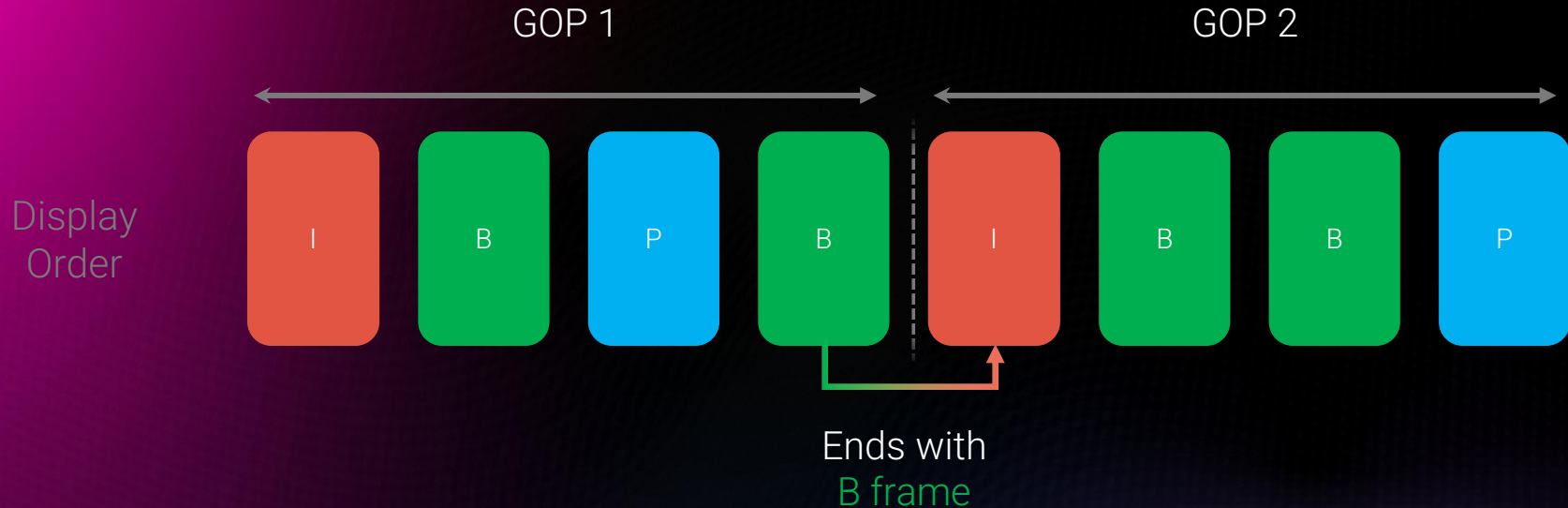
CLOSED GOP: INDEPENDENT



GoP are independent from each other

Closed GoP end with a P frame

OPEN GOP: DEPENDENT



GoP are dependent from each other

Open GoP end with a B frame

OPEN/CLOSED GOP: INTEROP & VIDEO QUALITY

CLOSED GOP

GoPs are **independent**

End with a **P** frame

Optimized for
interoperability

COGICOM

OPEN GOP

GoPs are **dependent**

End with a **B** frame

Optimized for
Video Quality

CLOSED vs OPEN GOP

OPEN/CLOSED GOP: INTEROP & VIDEO QUALITY

CLOSED GOP

Optimized for
interoperability

OPEN GOP

Optimized for
Video Quality

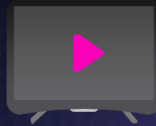
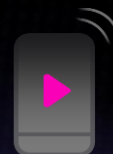
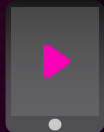
COGICOM

CLOSED vs OPEN GOP

CLOSED GOP USED ON MOST OTT CONFIGURATION

CLOSED GOP

Optimized for
interoperability



COGICOM

CLOSED vs OPEN GOP

CLOSED GOP & I FRAME BEATING



Closed GOP may lead to
I frame flickering

COGICOM

CLOSED vs OPEN GOP

GOP DURATION

& VIDEO QUALITY

COGICOM

GOP DURATION & INTRA FRAME INSERTION



Channel Zapping

Impact of packet loss

Scene change encoding

GOP DURATION & INTRA FRAME INSERTION



Channel Zapping

Impact of packet loss

Scene change encoding

COGICOM



Decreases Video Quality
if inserted too frequently..

But ... why?

GOP DURATION & VIDEO QUALITY

I FRAMES ARE THE **LEAST EFFICIENT FRAMES**
THEY ARE BIG IN SIZE

COGICOM

0.5 SEC GOP & VIDEO QUALITY

Video bitrate = 1Mbps

GoP = 0.5 second

25 frames per second

I frame = 60kB (480kb)

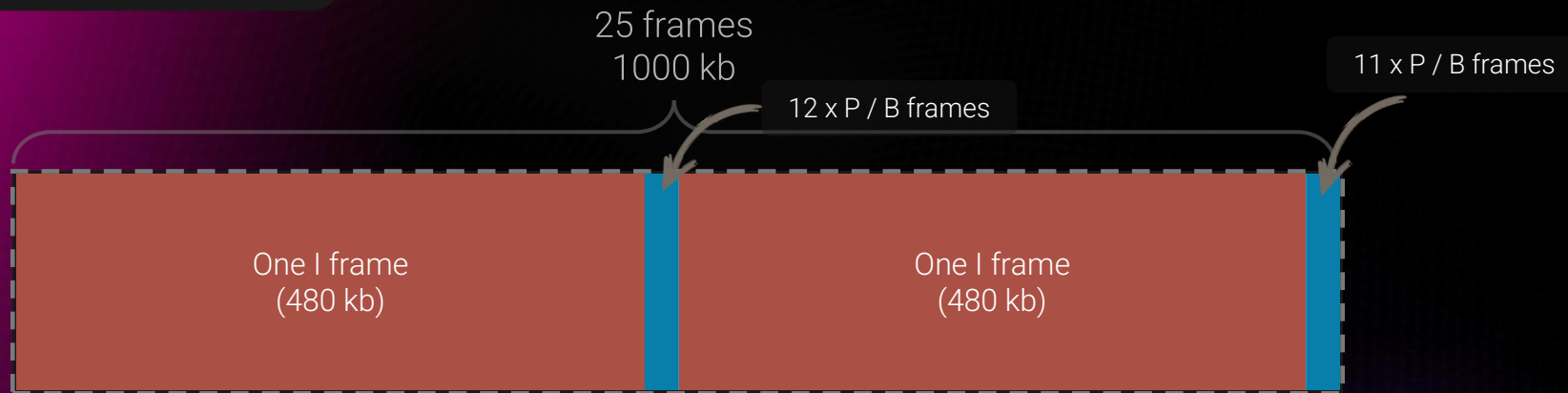
0.5 SEC GOP & VIDEO QUALITY

Video bitrate = 1Mbps

GoP = 0.5 second

25 frames per second

I frame = 60kB (480kb)



$40 \text{ kb} / 23 \text{ frames} = 1.74 \text{ kb}$
217 Bytes per Inter frame

COGICOM

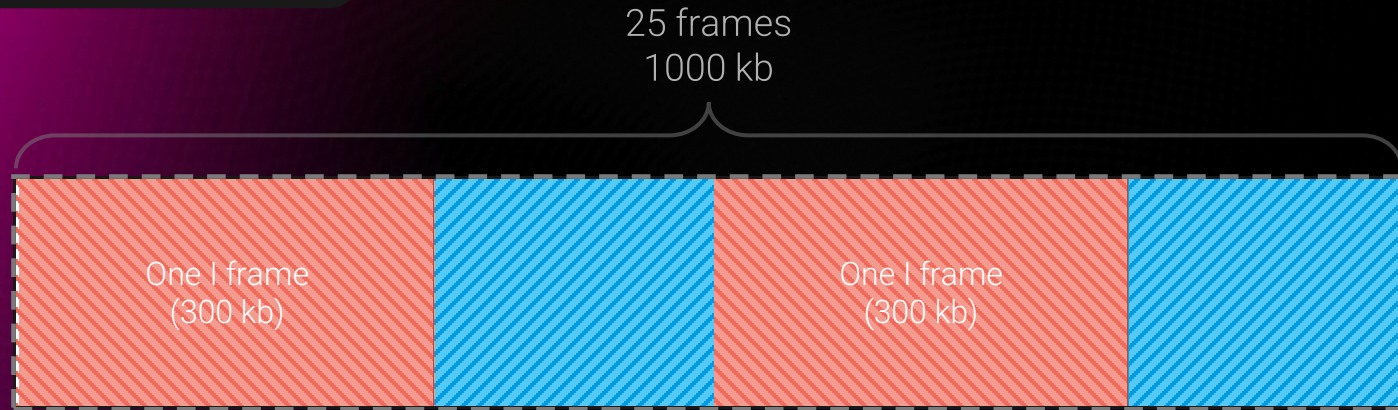
GOP DURATION & VIDEO QUALITY

DECREASING SIZE OF THE I FRAME?

Video bitrate = 1Mbps

GoP = 0.5 second

25 frames per second



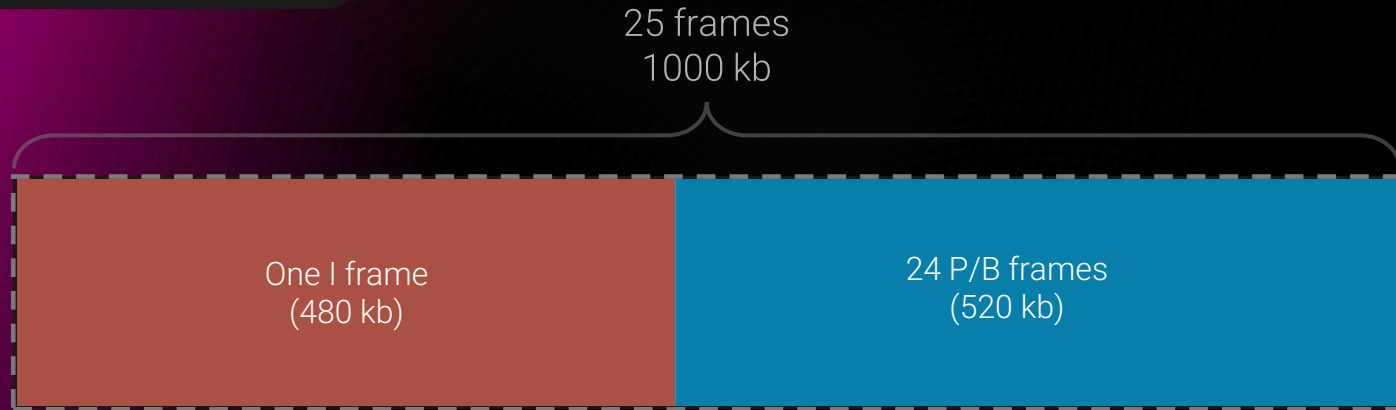
Reducing I frame size **reduces video quality of the entire sequence** inside the GoP

1 SEC GOP & VIDEO QUALITY

Video bitrate = 1Mbps

GoP = 1 second

25 frames per second

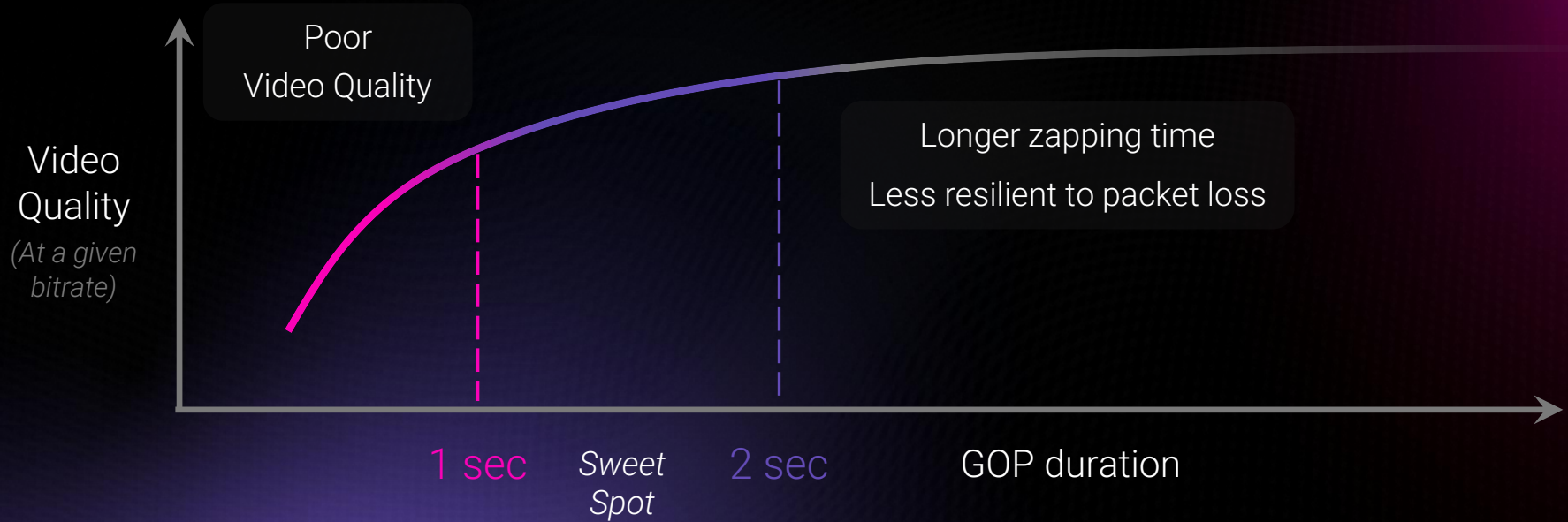


$520 \text{ kb} / 24 \text{ frames} = 21.67 \text{ kb}$
2780 Bytes per Inter frame

COGICOM

GOP DURATION & VIDEO QUALITY

GOP DURATION & VIDEO QUALITY



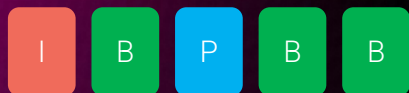
COGICOM

GOP DURATION & VIDEO QUALITY

GOP DURATION: TWO FAMILIES

“LONG GOP”

Temporal prediction



Contribution & Distribution

Low bitrate (< 30Mbps)

“I FRAME ONLY”

Spatial prediction only



High-quality professional format
(+ Offline editing)

High bitrate (≈ 100Mbps)

COGICOM

INTRA ONLY vs LONG GOP

GOP TYPES & VIDEO QUALITY

“LONG GOP”



Video Quality
(At a given bitrate)



“I FRAME ONLY”

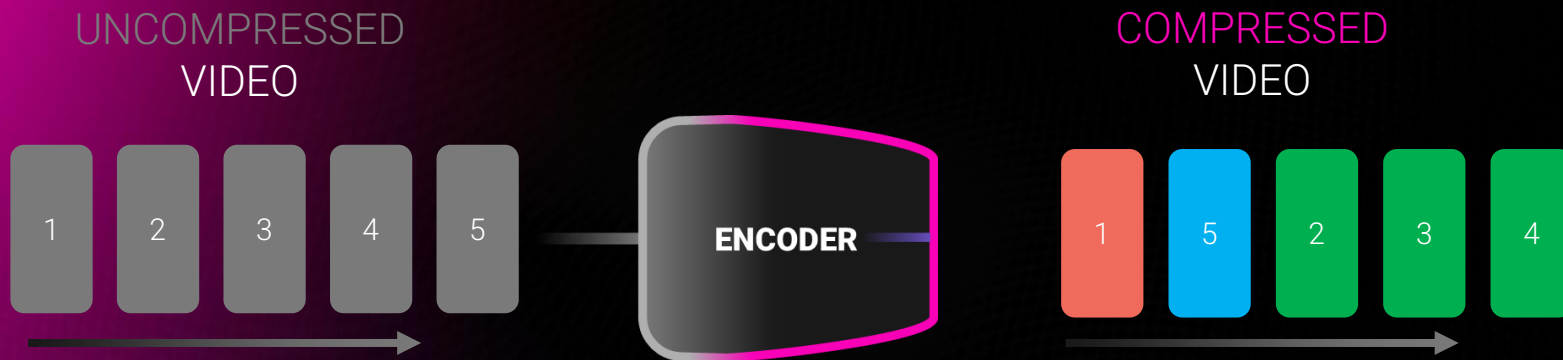
“I frame Only” requires a much higher bitrate
to have an equivalent video quality as “Long GOP”

ENCODING ORDER

HOW FRAMES ARE TRANSMITTED

COGICOM

ENCODING ORDER WITH 2 B FRAMES

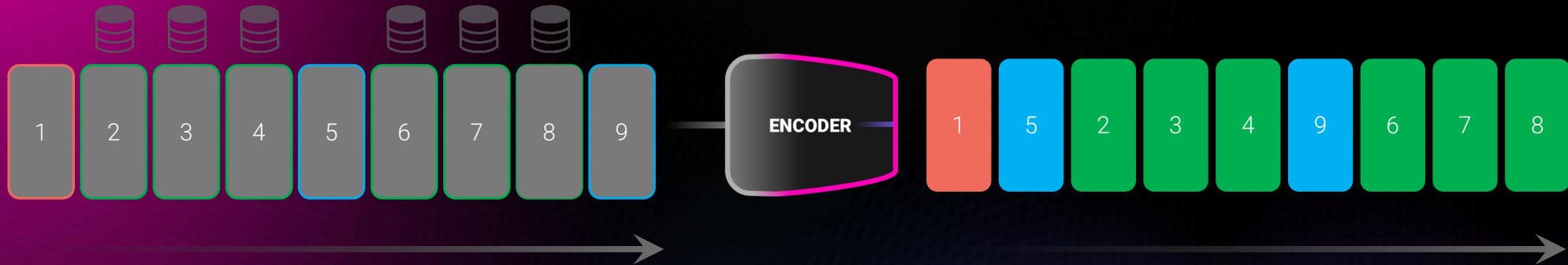


Encoder always encodes the references first

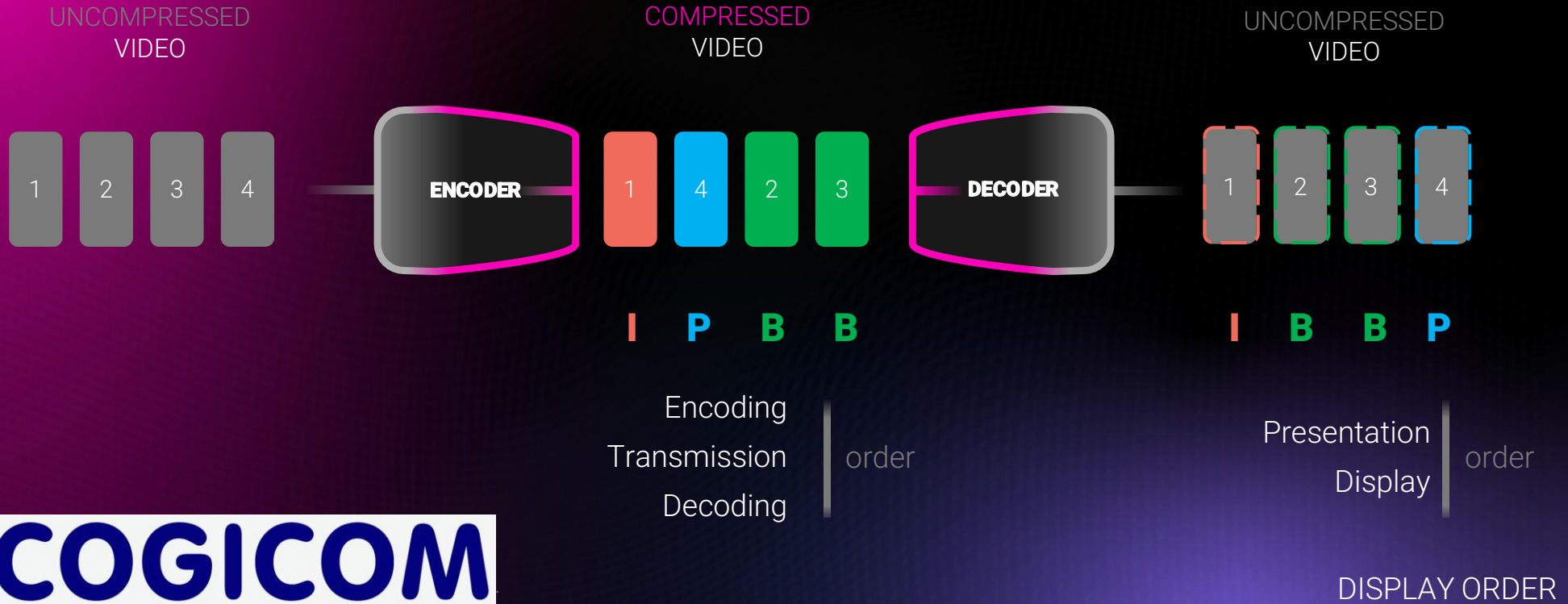
Encoding order $\neq$ source order, when B frames are used for encoding

ENCODING ORDER WITH 3 B FRAMES

Maximum number of consecutive B frames = 3



ENCODING vs DISPLAY ORDER



DISPLAY ORDER

UNCOMPRESSED
VIDEO

Presentation
Display

ENCODER

COMPRESSED
VIDEO

Encoding
Transmission
Decoding

DECODER

UNCOMPRESSED
VIDEO

Presentation
Display

COGICOM

DISPLAY ORDER

ENCODING ORDER IS

THE ORDER IN WHICH THE ENCODER GENERATES FRAMES

IT DIFFERS FROM THE DISPLAY ORDER WHEN B FAMES ARE USED

COGICOM

DISPLAY ORDER IS
THE ORDER IN WHICH UNCOMPRESSED PICTURES ARE
DISPLAYED OUT OF THE DECODER
USED TO CHARACTERIZE THE GOP STRUCTURE

COGICOM

Quiz 2

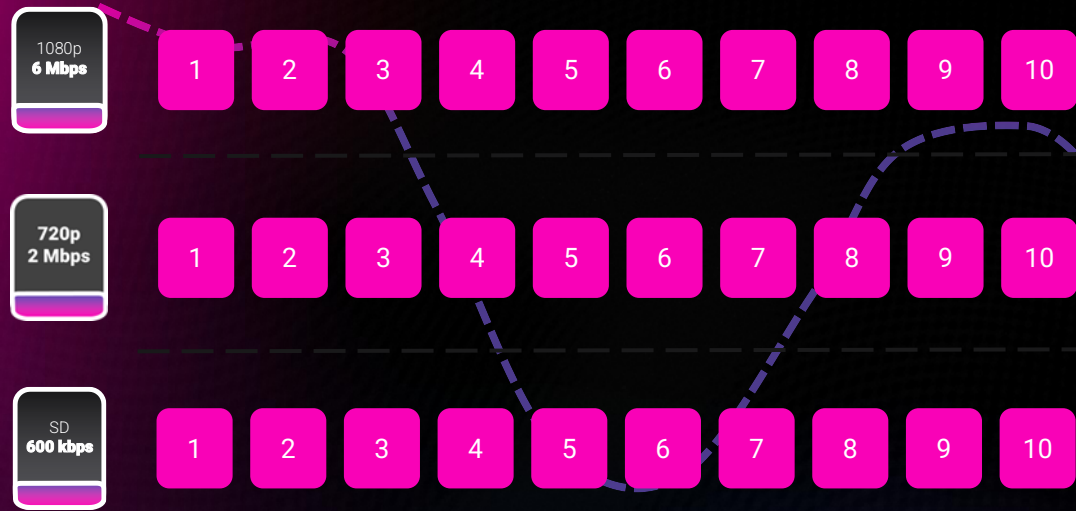
COGICOM

ENCODING OTT FRAGMENTS

MAKING THEM INDEPENDENT

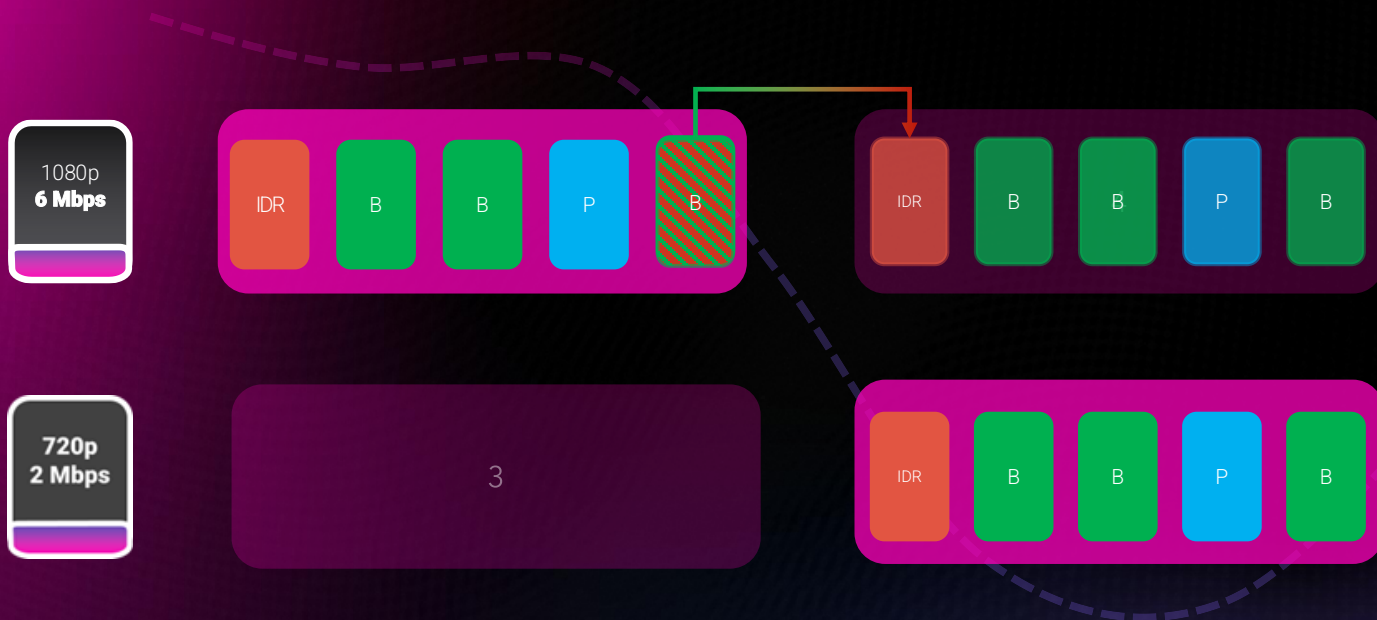
COGICOM

OTT SEGMENTS MUST BE DECODABLE INDEPENDENTLY



Each fragment/segment should be **decodable independently**

SEGMENTS SHOULDN'T END WITH AN OPEN GOP



The last GOP cannot be open; the last displayed frame cannot be a B-frame

SEGMENTS SHOULDN'T END WITH AN OPEN GOP

Display order



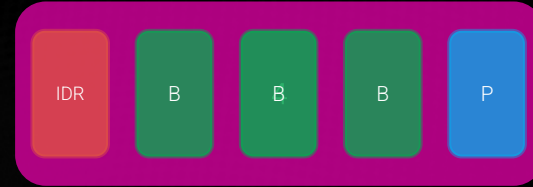
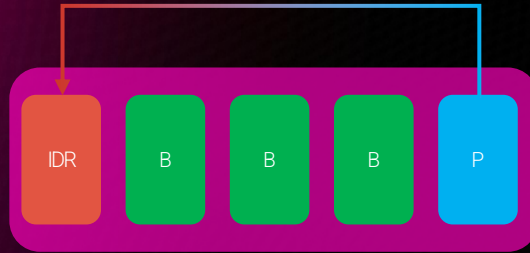
Transmission order



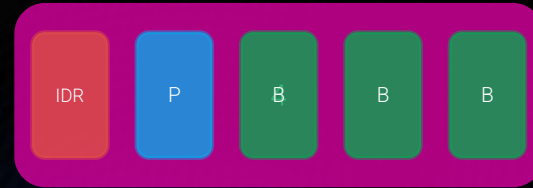
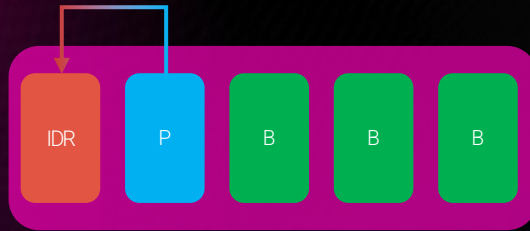
B-frame ending fragment 1 (*display order*),
is actually transmitted in fragment 2

SEGMENTS SHOULD END WITH A CLOSED GOP

Display order



Transmission order



The last GOP of a fragment needs to be closed, the last displayed frame should be a P-frame

THE IDR FRAME

INSTANTANEOUS DECODER REFRESH

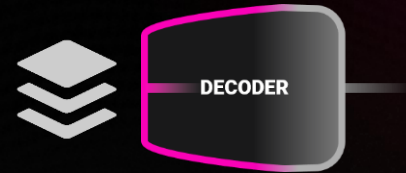
COGICOM

IDR FRAME IS A **TYPE OF INTRA FRAME**

- **DELIMITS OTT FRAGMENT/SEGMENTS**
- TELLS THE DECODER TO **FLUSH** ITS *DECODED PICTURE BUFFER*
- HELPS **INCREASE ROBUSTNESS ON OTT SEGMENTATION**

DECODED PICTURE BUFFER

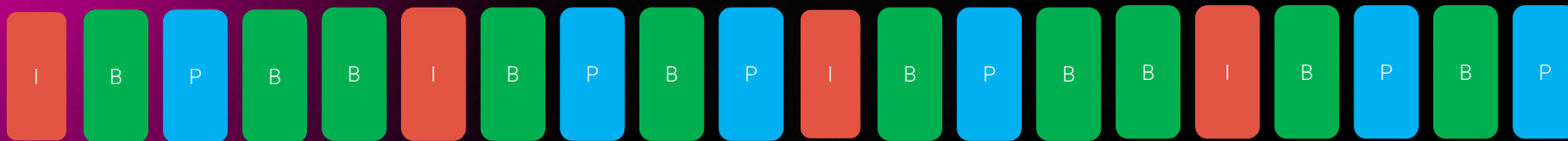
**BUFFER STORING REFERENCE PICTURES USED FOR TEMPORAL
PREDICTION**



COGICOM

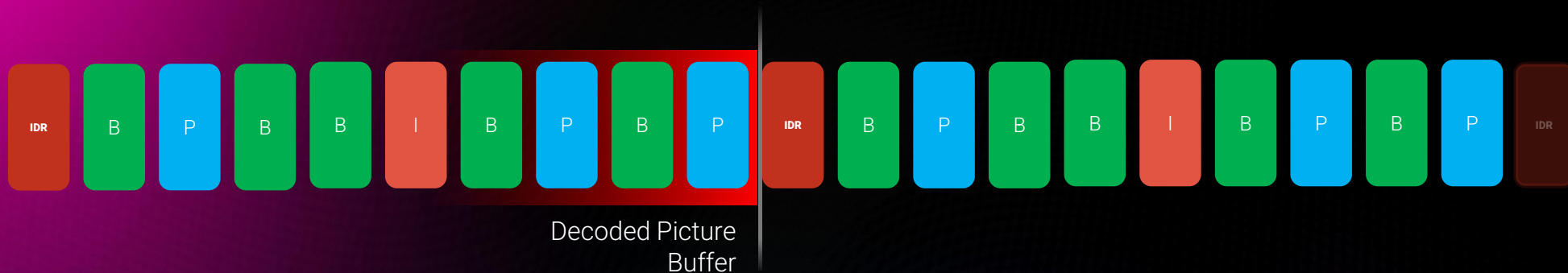
INSERTING AN IDR AT FRAGMENT START

5 fps



INSERTING AN IDR AT FRAGMENT START

5 fps



INSERTING AN IDR AT FRAGMENT START

5 fps



Any OTT fragment/segment should
start with an IDR frame

**AN OTT FRAGMENT IS
A TIME INTERVAL BETWEEN TWO CONSECUTIVE IDR FRAMES**

COGICOM

GOP vs OTT FRAGMENT

GOP CONFIGURATION ON THE OTT ENCODER

COGICOM

OTT FRAGMENTS CAN CONTAIN ONE OR MORE GOP
DEPENDING ON THE DURATION SETTING
GOP CAN BE FIXED/VARIABLE AND CLOSED/OPEN

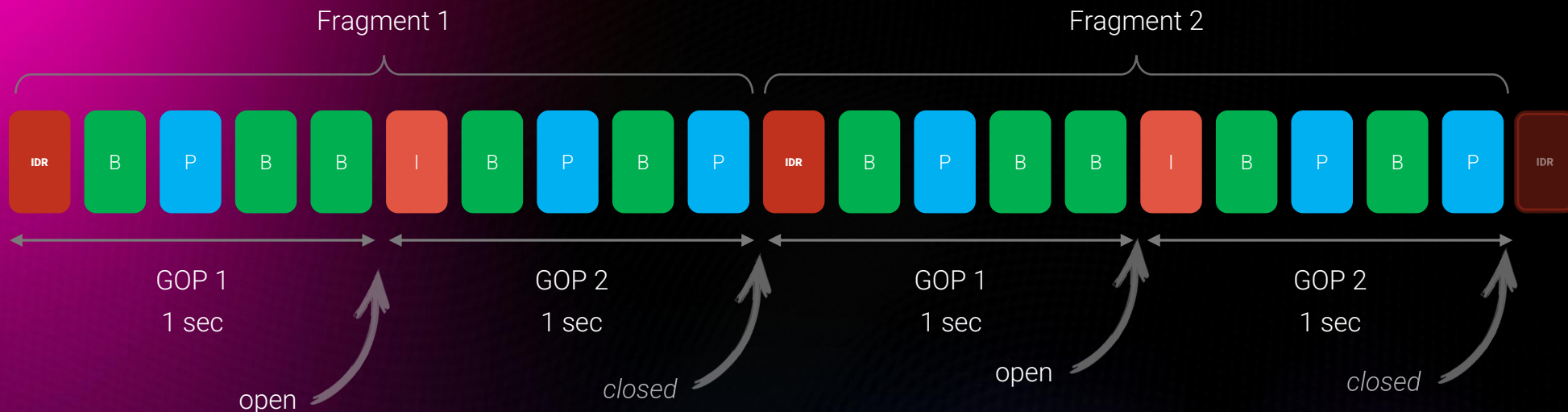
COGICOM

**ONLY THE LAST GOP OF EACH FRAGMENT MUST BE CLOSED
TO ENSURE SEAMLESS BITRATE SWITCHING**

COGICOM

GOP CONFIGURATION INSIDE OTT FRAGMENT

5 fps



OTT encoding
configuration

Fragment duration

2 sec

GoP

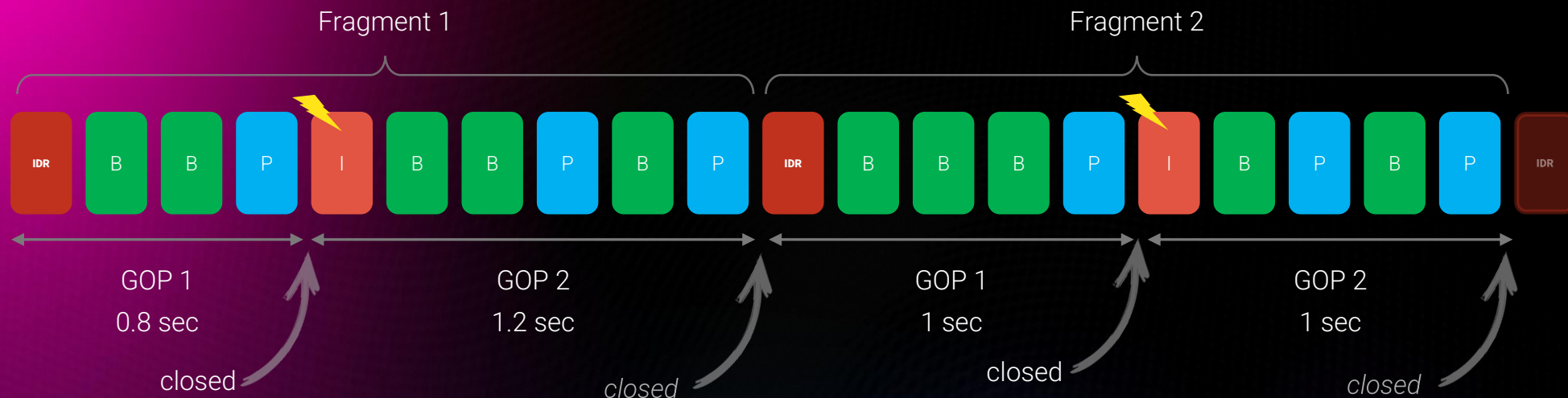
fixed 1 sec open IBBP

COGICOM

GOP vs OTT FRAGMENT

GOP CONFIGURATION INSIDE OTT FRAGMENT

5 fps



OTT encoding
configuration

Fragment duration

2 sec

GoP

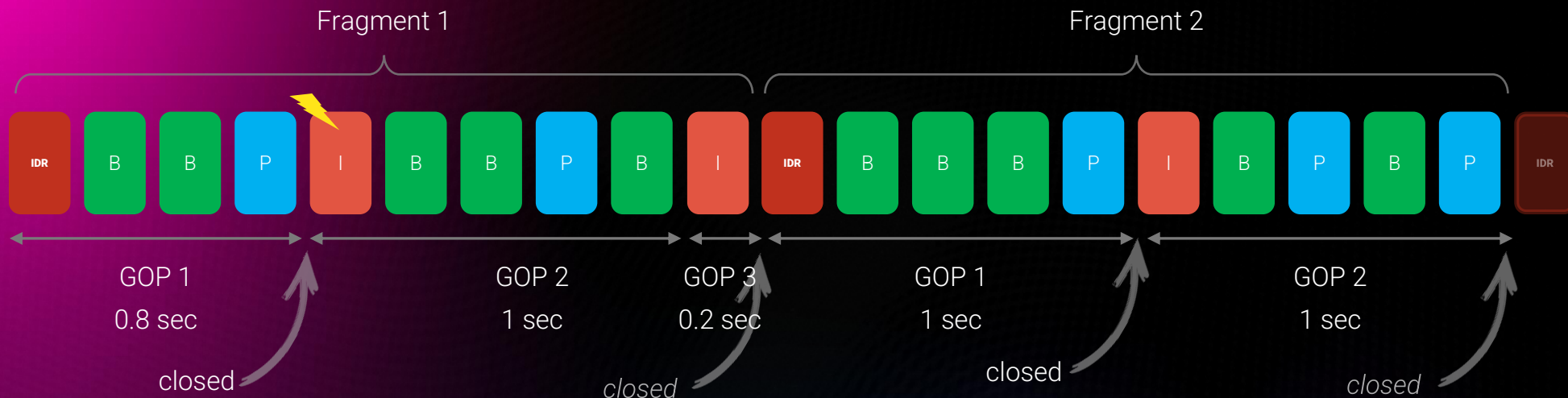
variable 1.2 sec closed IBBBP

COGICOM

GOP vs OTT FRAGMENT

GOP CONFIGURATION INSIDE OTT FRAGMENT

5 fps



OTT encoding
configuration

Fragment duration

2 sec

GoP

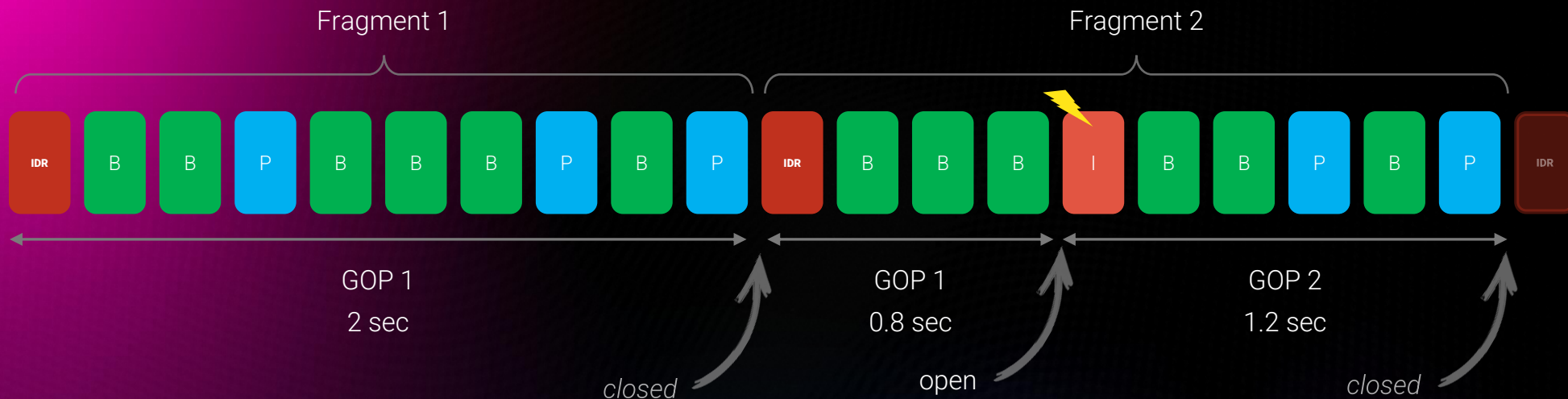
variable 1 sec closed IBBBP

COGICOM

GOP vs OTT FRAGMENT

GOP CONFIGURATION INSIDE OTT FRAGMENT

5 fps



OTT encoding
configuration

Fragment duration

2 sec

GoP

variable

2 sec

open

IBBBP

COGICOM

GOP vs OTT FRAGMENT

Quiz 3

COGICOM