Qualcomm

July 14, 2021

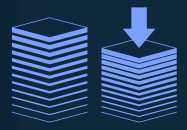
@QCOMResearch

How AI research is enabling next-gen codecs

Qualcomm Technologies, Inc.

Qualcomm

Agenda

1. 

   The demand for improved data compression is growing

2. 

   AI is a compelling tool for compression

3. 

   Our latest AI voice and video codec research

4. 

   Our on-device neural video decoder demo

5. 

   Future research work and challenges

The scale of video and voice being created and consumed is massive

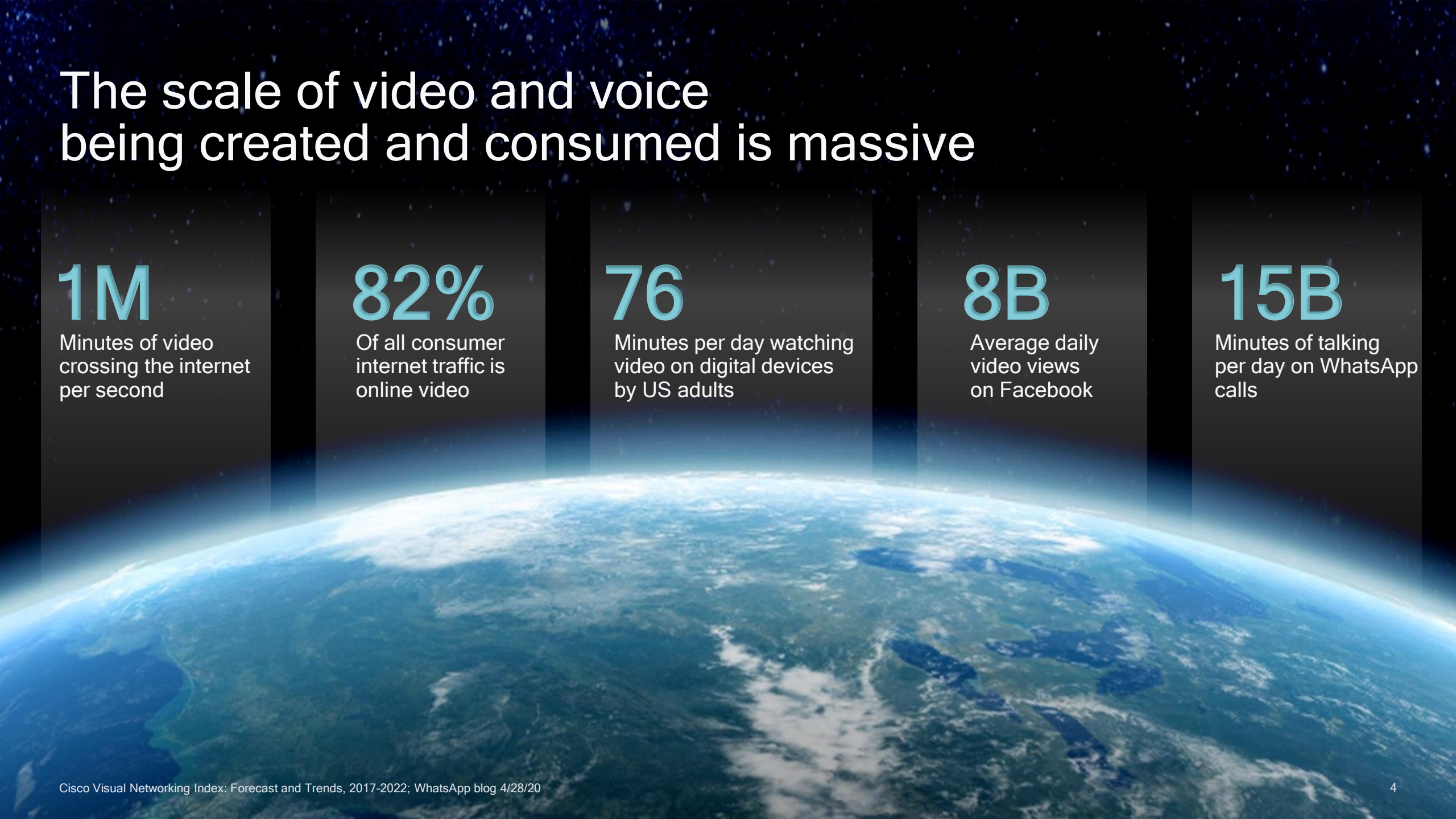
1M
Minutes of video crossing the internet per second

82%
Of all consumer internet traffic is online video

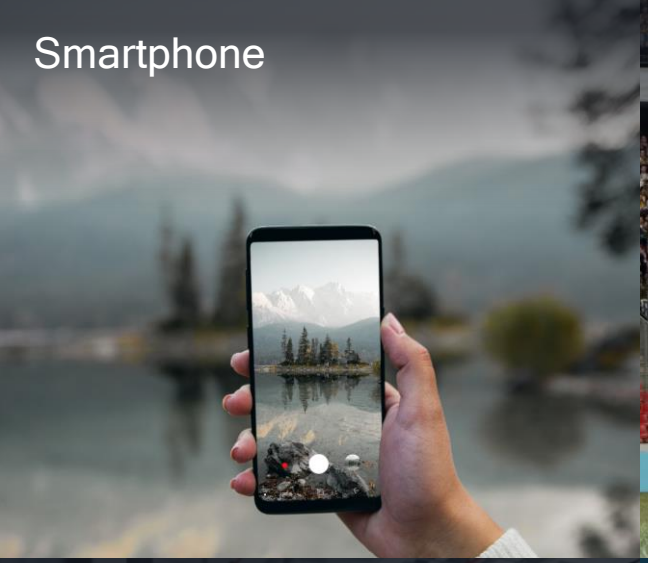
76
Minutes per day watching video on digital devices by US adults

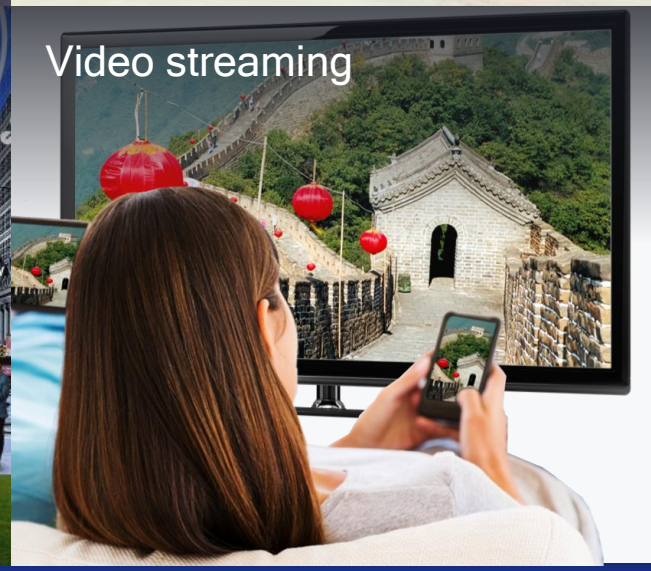
8B
Average daily video views on Facebook

15B
Minutes of talking per day on WhatsApp calls



Cisco Visual Networking Index: Forecast and Trends, 2017-2022; WhatsApp blog 4/28/20

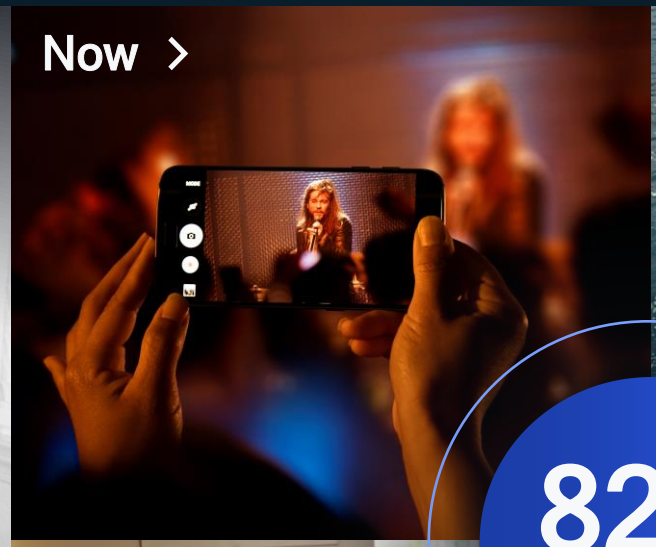


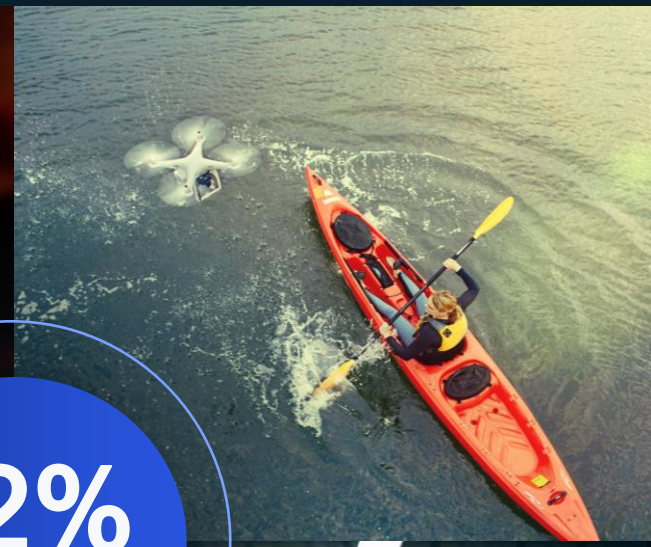


Increasingly, video is all around us – providing entertainment, enhancing collaboration, and transforming industries

Video technology revolutionized how we create and consume media

Enhanced video quality with less bits led to broad adoption across a wide range of devices and services



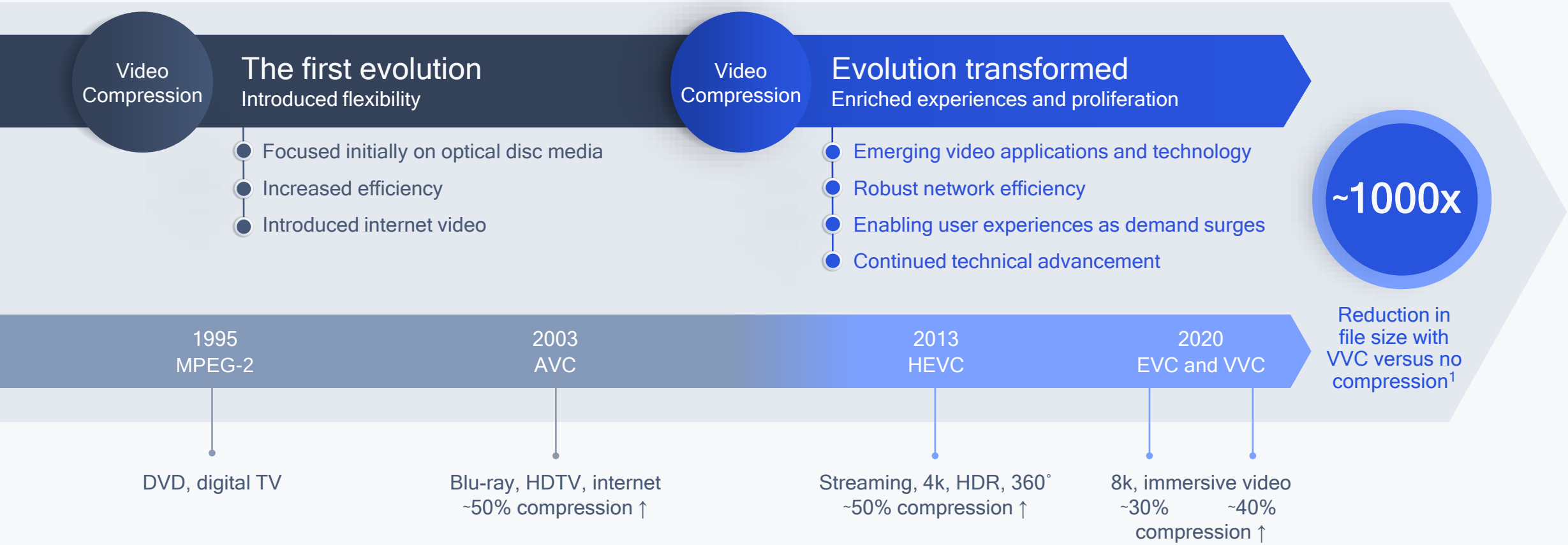


82%
of internet traffic will be video by 2022[1]

1. Cisco Annual Internet Report, 2018-2023

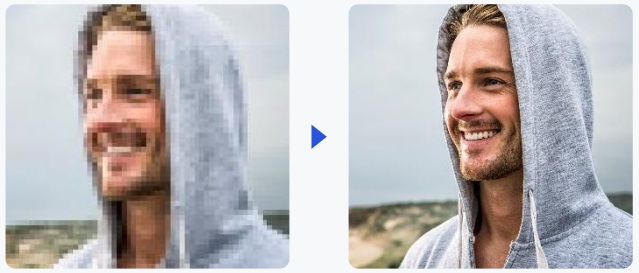
Technical innovation drives video evolution

A regular cadence of technical advancement in video codecs has led to massive reduction in file size



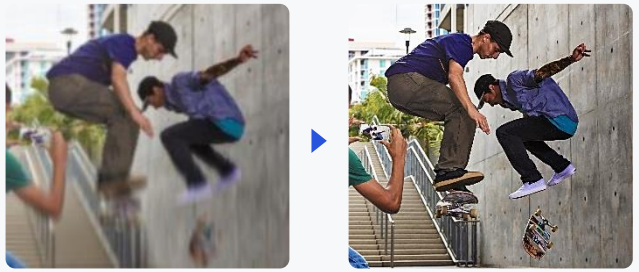
1. On a 1080p video at 30 frames per second; EVC = Essential Video Coding, VVC = Versatile Video Coding,

Pixel quantity



Resolution

Increased definition and sharpness

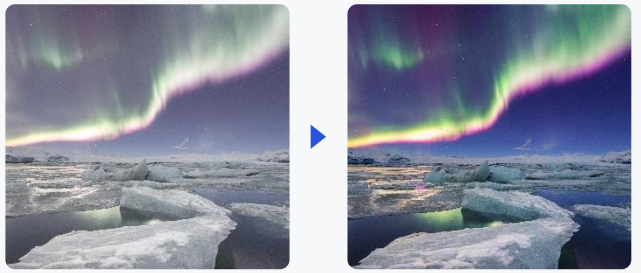


Frame rate

Reduced blurring and latency

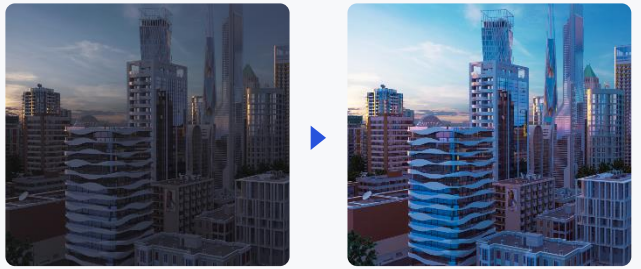
Broad interoperability
across all our screens

Pixel fidelity



Color accuracy

More realistic colors through an expanded color gamut, depth, and temperature



Contrast and brightness

Increased detail through a larger dynamic range and lighting enhancements

Visual quality is much more than resolution and frame rate

Preserving color accuracy, contrast, and brightness is also crucial

Deep generative model research for unsupervised learning

Given unlabeled training data, generate new samples from the same distribution

Generative models

Variational auto encoder (VAE)*

Generative adversarial network (GAN)

Auto-regressive

Invertible

Powerful capabilities

Extract features by learning a low-dimension feature representation

Sampling to generate, restore, predict, or compress data

Broad applications


Speech/video compression


Text to speech


Graphics rendering


Computational photography


Voice UI

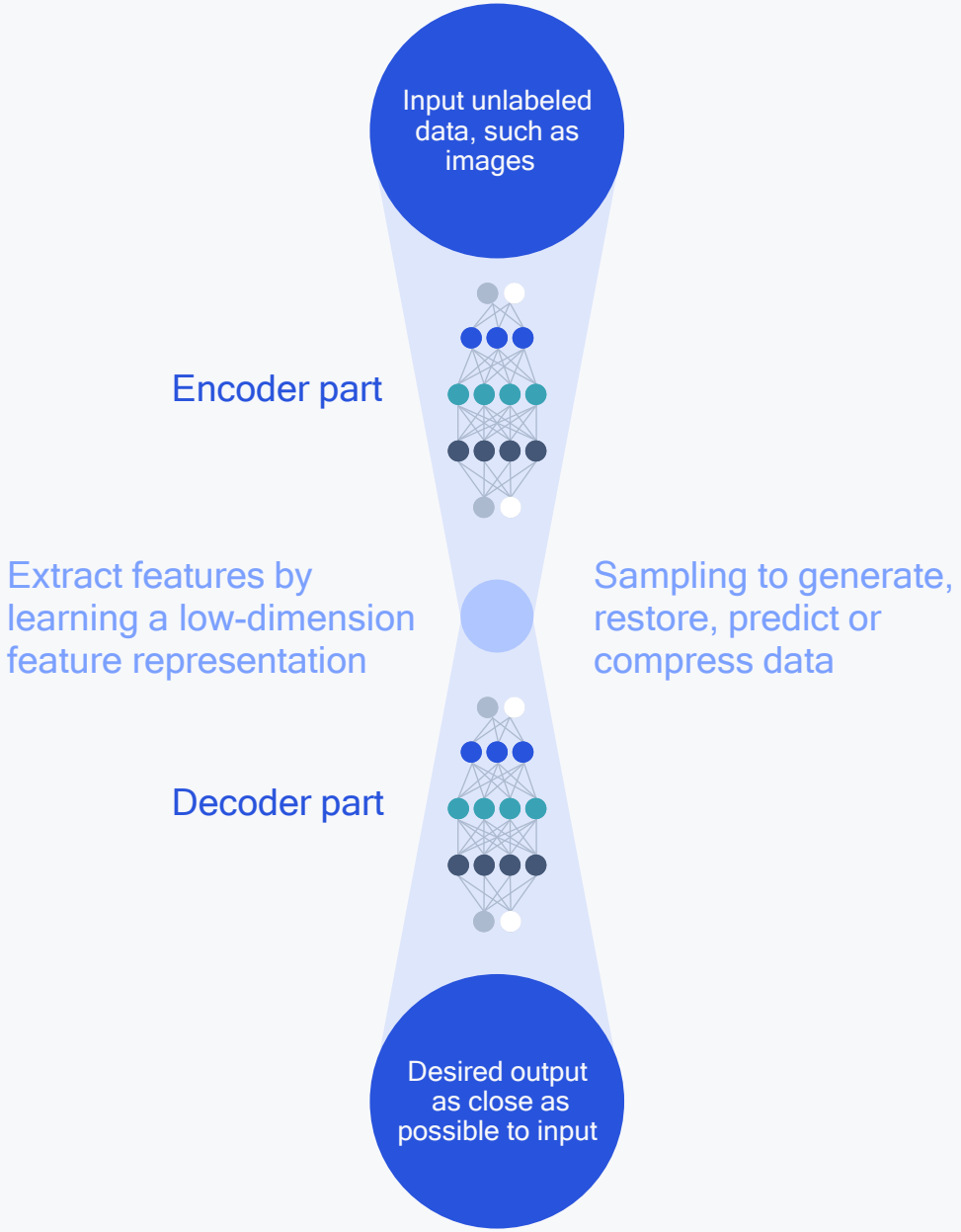
* VAE first introduced by D. Kingma and M. Welling in 2013

Deep generative model research for unsupervised learning

Given unlabeled training data, generate new samples from the same distribution

Generative models

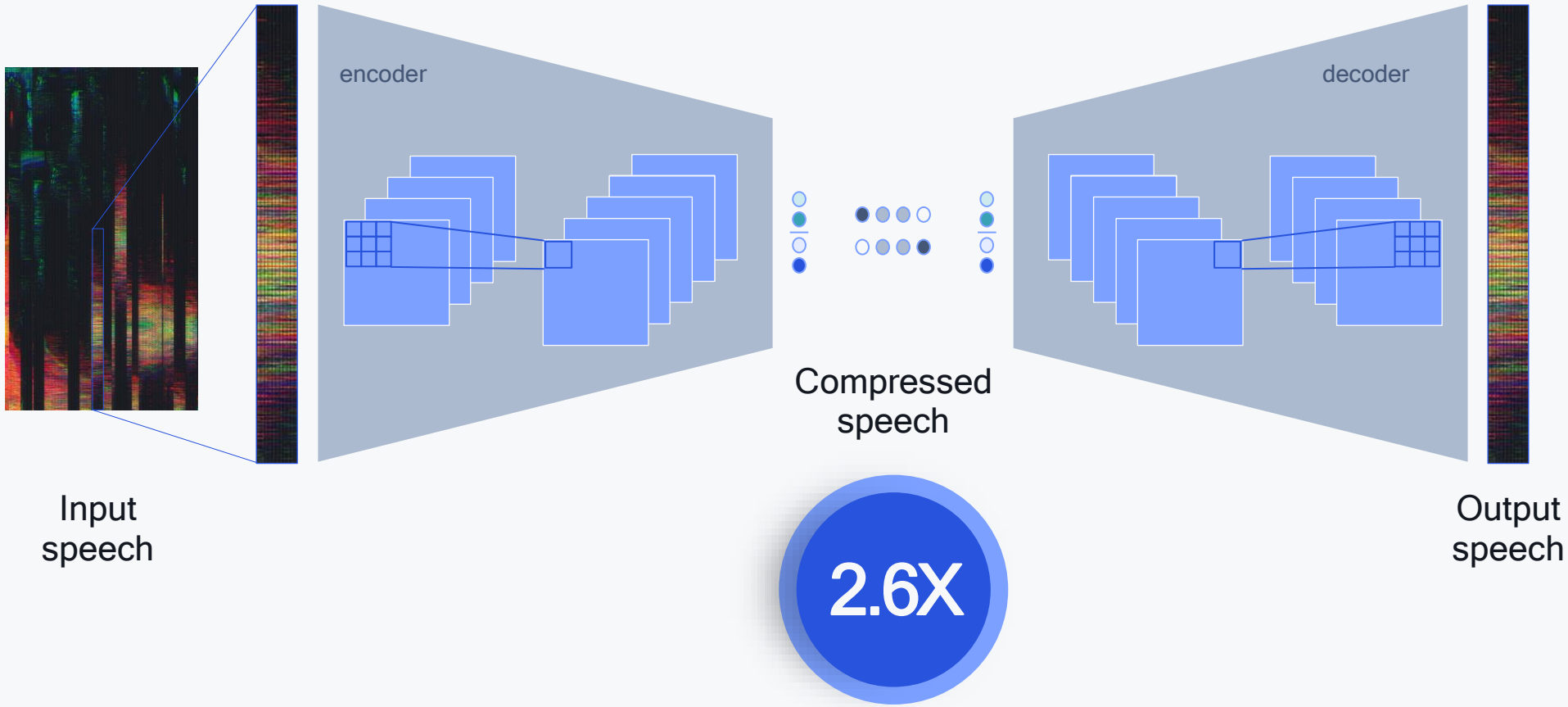
- Variational auto encoder (VAE)*
- Generative adversarial network (GAN)
- Auto-regressive
- Invertible



* VAE first introduced by D. Kingma and M. Welling in 2013

AI-based compression has compelling benefits

- Improved rate-distortion trade-off
- Semantics aware for human visual perception
- Specialized to a specific data distribution
- Optimized for advanced perceptual quality metrics
- No special-purpose hardware required, other than an AI acceleration
- Can generate visual details not in the bitstream
- Easy to upgrade, standardize, and deploy new codecs
- Easy to develop new codecs for new modalities

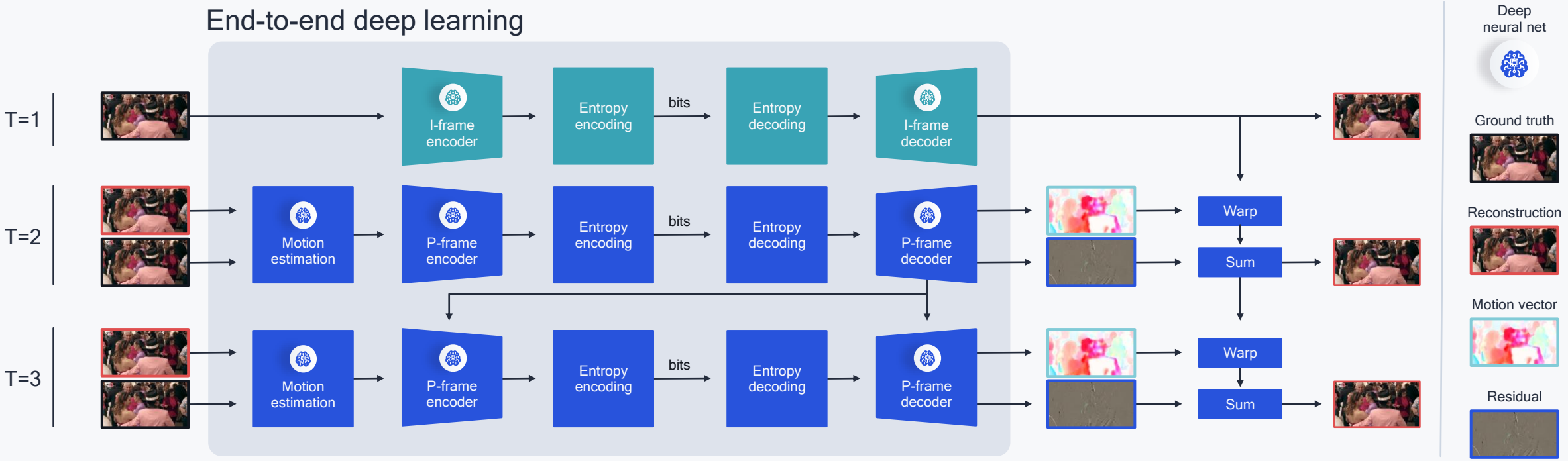


Achieving state-of-the-art data compression with AI

1 Comparison between state-of-the-art traditional speech coding algorithm and AI speech compression

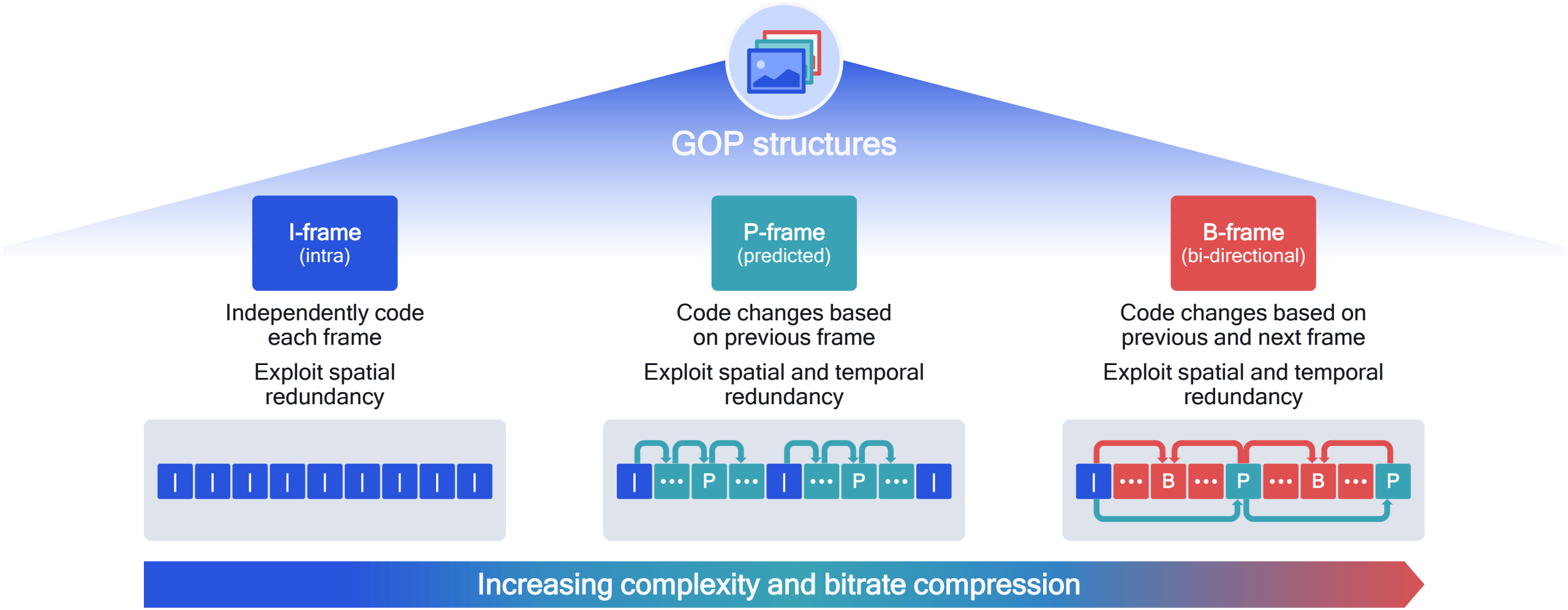
Novel machine learning-based video codec research

Neural network based I-frame and P-frame compression



Achieving state-of-the-art rate-distortion compared with other learned video compression solutions

Feedback Recurrent Autoencoder for Video Compression (Golinski, et al., arXiv 2020)



Video compression utilizes temporal and spatial redundancy

Group of pictures (GOP) is a sequence of video frame types used to efficiently encode data

Existing research for B-frame codecs has limitations

Our solution combines the best of existing P-frame and B-frame codecs while allowing them to share weights

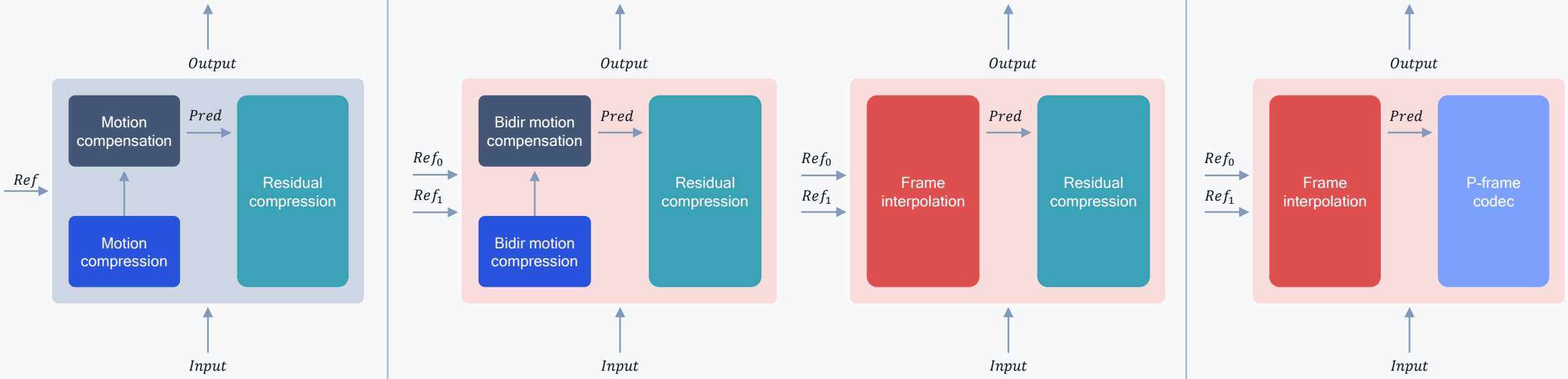


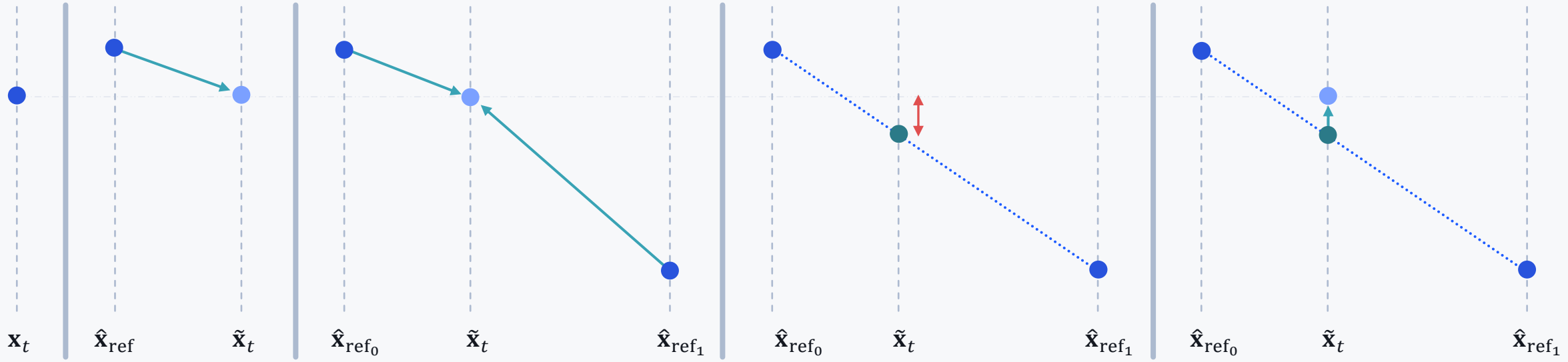
Illustration of different inter-frame coding methods

B-Frame compression through Extended P-frame & Interpolation Codec (B-EPIC)

$\mathbf{x}_t$ Input frame
$\tilde{\mathbf{x}}_t$ The corresponding prediction
$\hat{\mathbf{x}}_{\text{ref}}$ The reference

Object location
Prediction
Motion vector

Prediction error
Interpolation
Interpolation guide



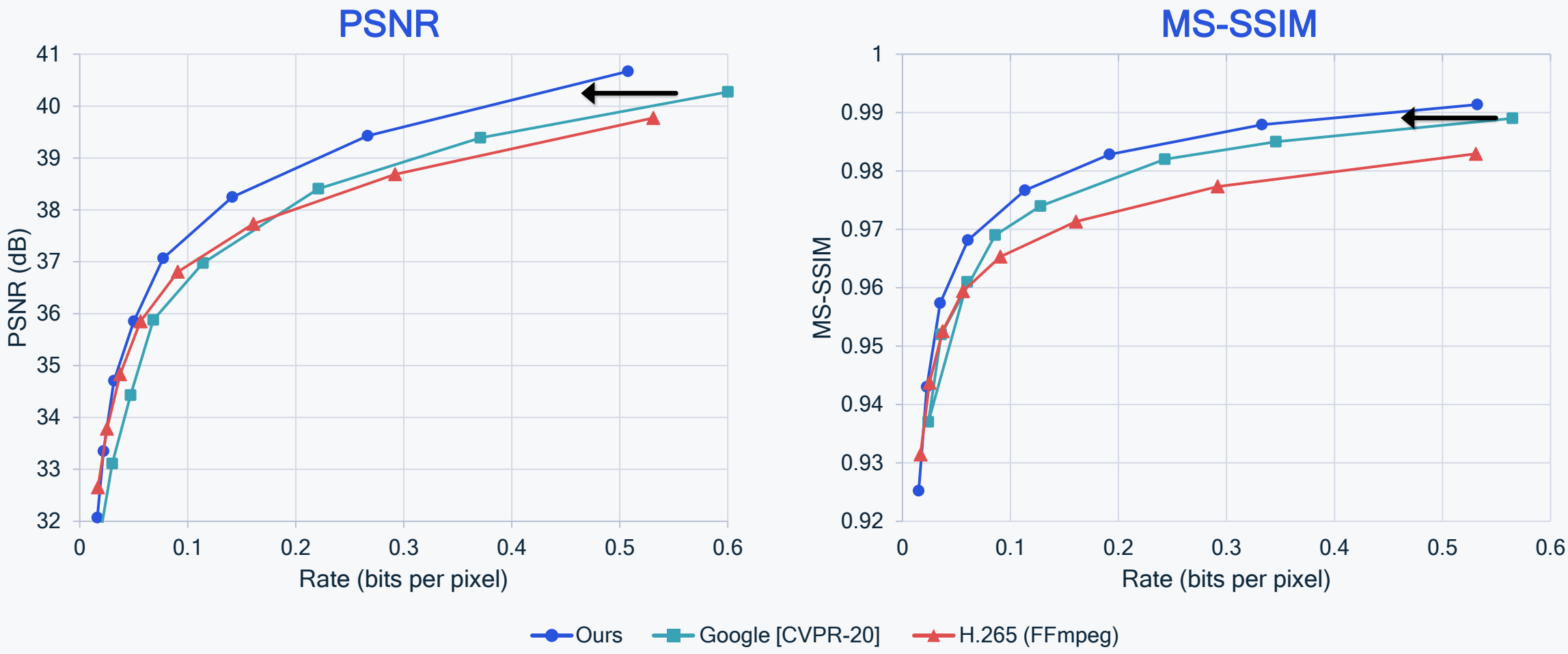
Actual object location

P-frame prediction: a motion vector with respect to a single reference is transmitted

B-frame prediction based on bidirectional flow/warp: two motion vectors with respect to two references are transmitted.

B-frame prediction based on frame interpolation: the interpolation result is treated as the prediction. No motion information is transmitted.

Our B-frame prediction approach: the interpolation result is corrected using a unidirectional motion vector similar to P-frame



Our B-frame coding provides state-of-the-art results

Improved rate-distortion tradeoff by extending neural p-frame codecs for B-frame coding

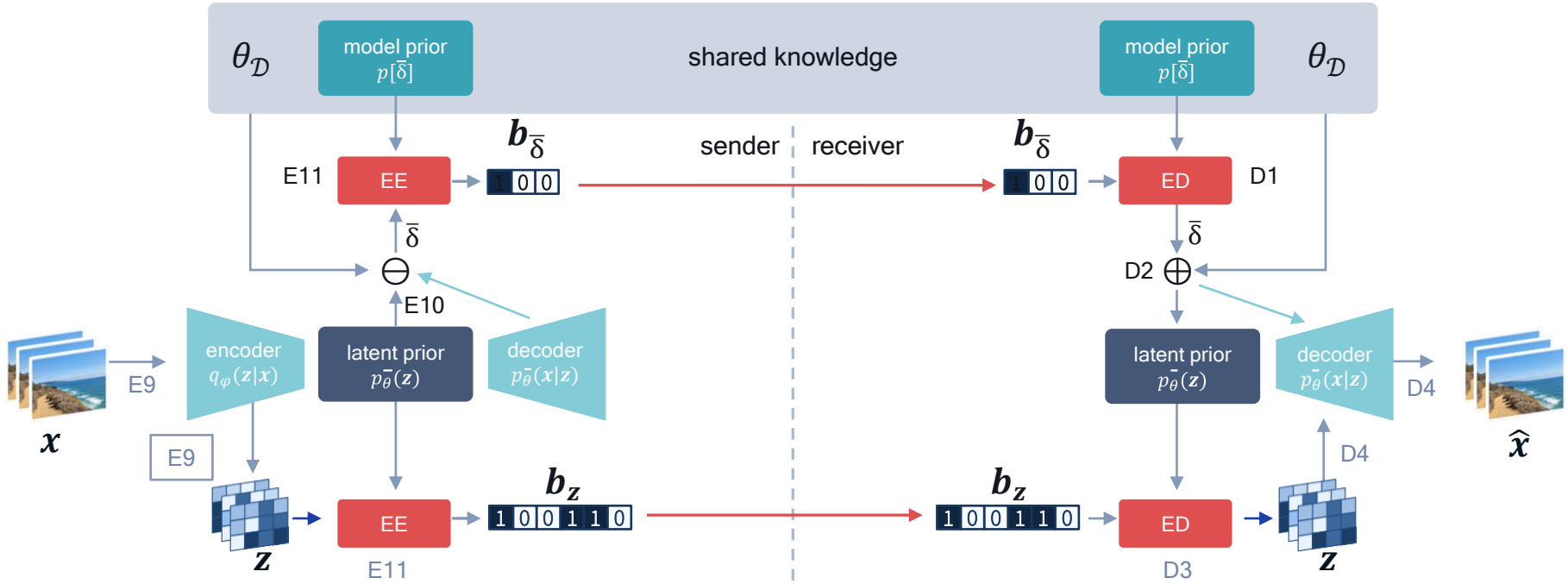
R. Pourezza, T. Cohen, Extending Neural P-frame Codecs for B-frame Coding, Under review at ICCV 2021

Overfitting through instance-adaptive video compression

Improved compression via continual learning

Send weight-deltas
based on overfitting

Send smaller encoded bitstream
based on overfitting



T. van Rozendaal*, I.A.M. Huijben*, T. Cohen, *Overfitting for Fun and Profit: Instance-Adaptive Data Compression*, ICLR 2021
T. Van Rozendaal, Y. Zhang, J. Brehmer, T. Cohen, *Instance-Adaptive Video Compression: Improving Neural Codecs by Training on the Test Set*, Under review at ICCV 2021

Overfit a model on one video instance and encode weight-deltas in the bitstream

Model delta + Encoded bitstream < baseline bitstream

SOTA results from instance-adaptive video compression

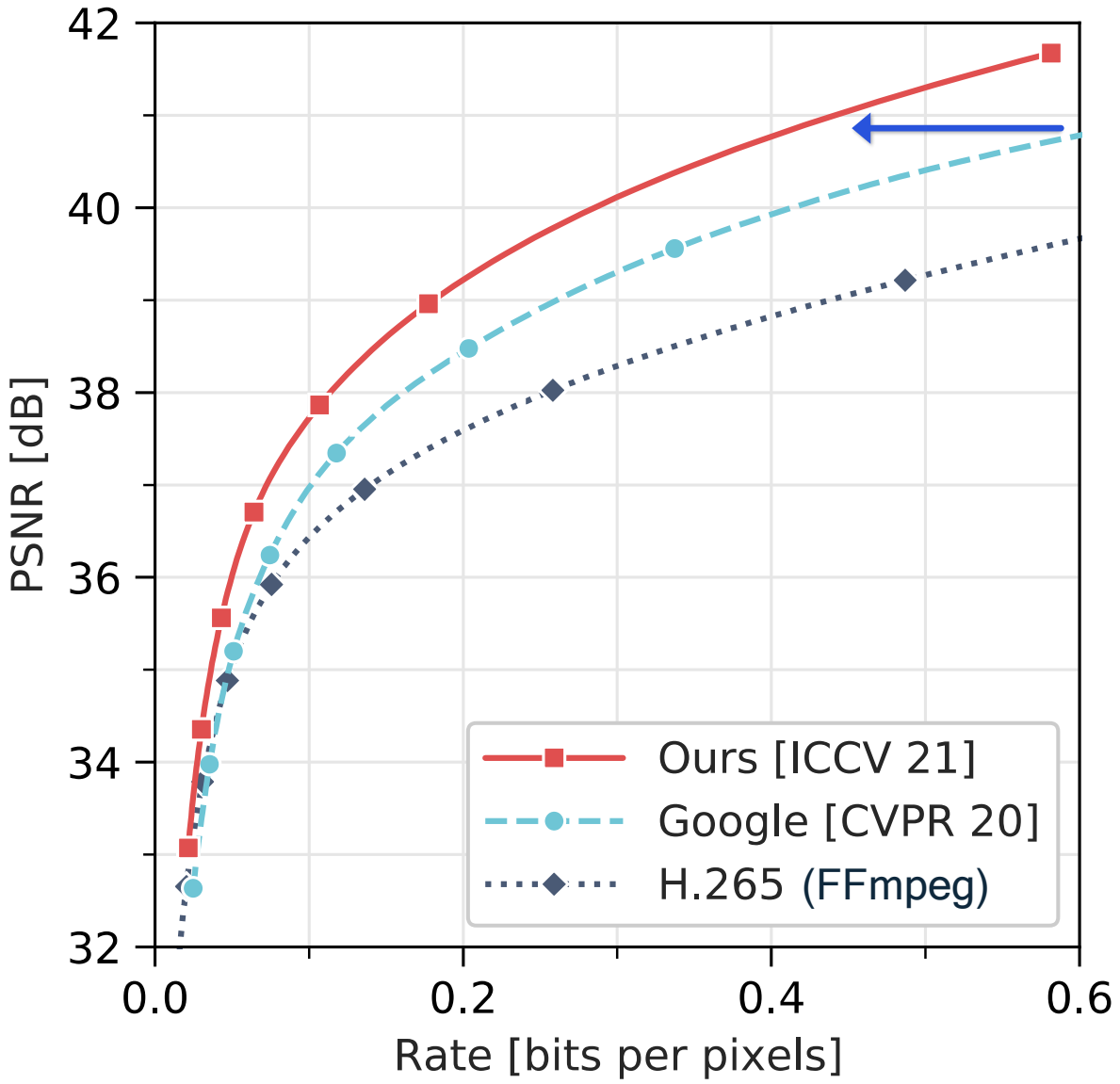
24%
BD-rate savings over leading neural codec by Google

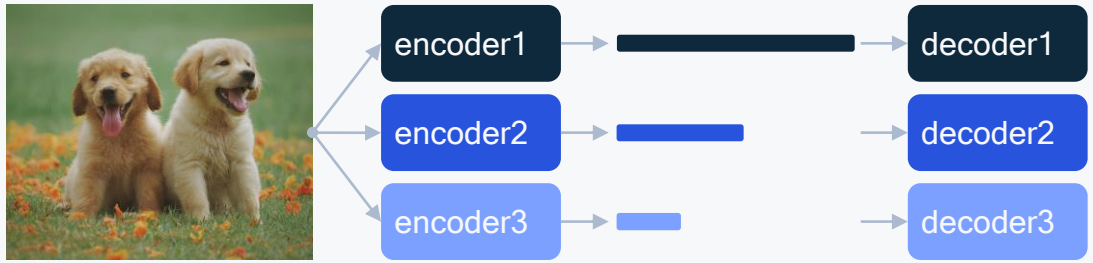
29%
BD-rate savings over FFmpeg H.265

Mobile friendly deployment

Decoding complexity can be reduced by

72% while still maintaining SOTA results

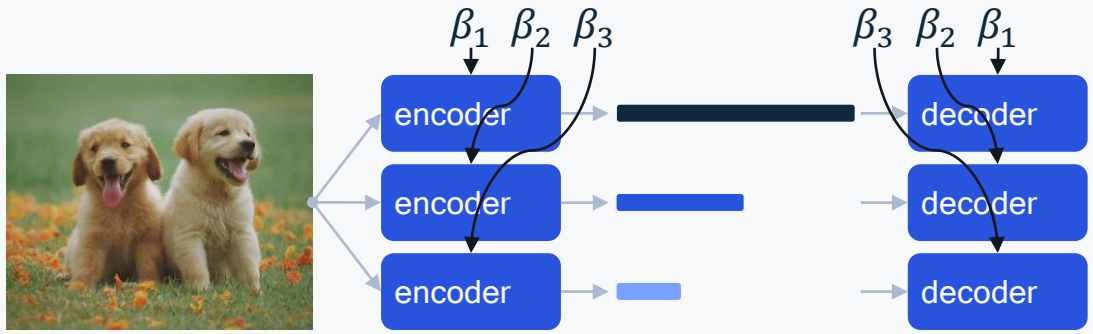




Level 0:
Multi-model multi-bitstream

Each model is optimized for a fixed β

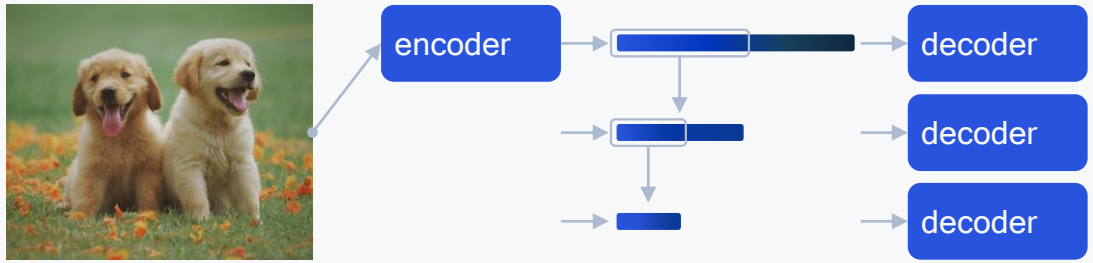
N models are needed for N rate-distortion tradeoffs



Level 1:
Single-model multi-bitstream

Single encoder-decoder model

Certain params /input are used to steer R-D tradeoff



Level 2:
Single-model single-bitstream

Single encoder-decoder model

Encode once, embed all bitrates

A.k.a. embedded/ progressive/ scalable coding

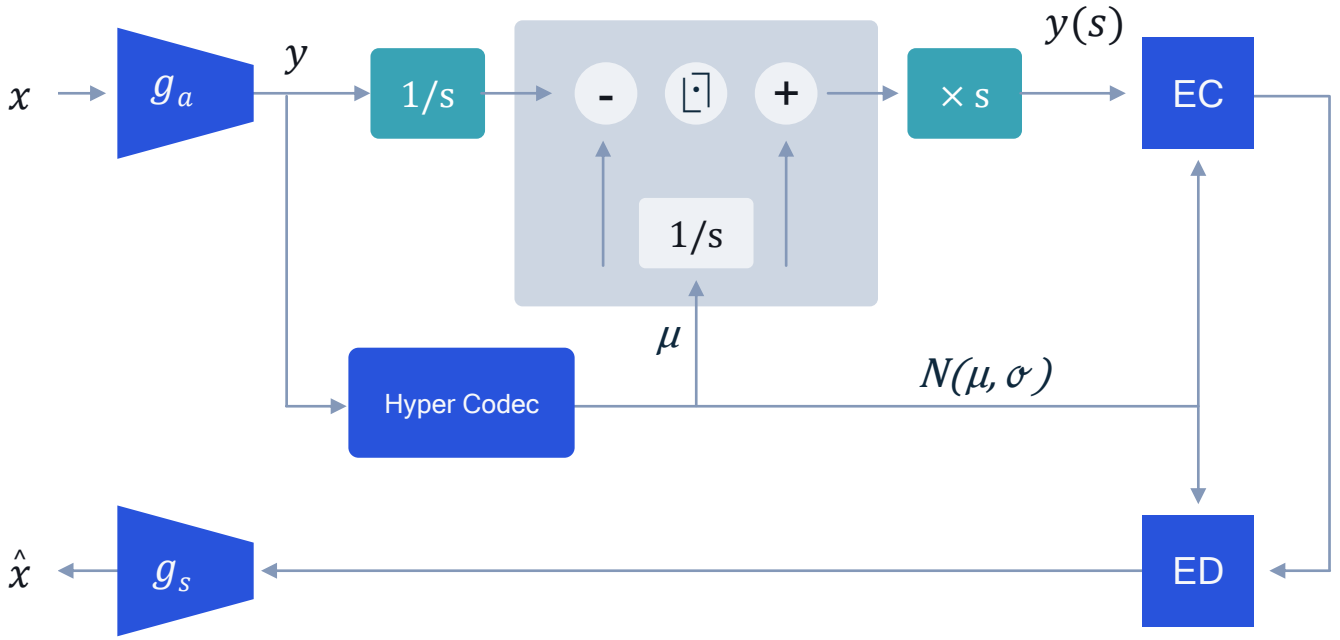
Increasing order of design difficulty and ease of deployment

Variable bitrate image compression for simpler deployment

Three levels of solutions with the goal of single model producing a single bitstream

Latent scaling for a single-model multi-bitstream solution

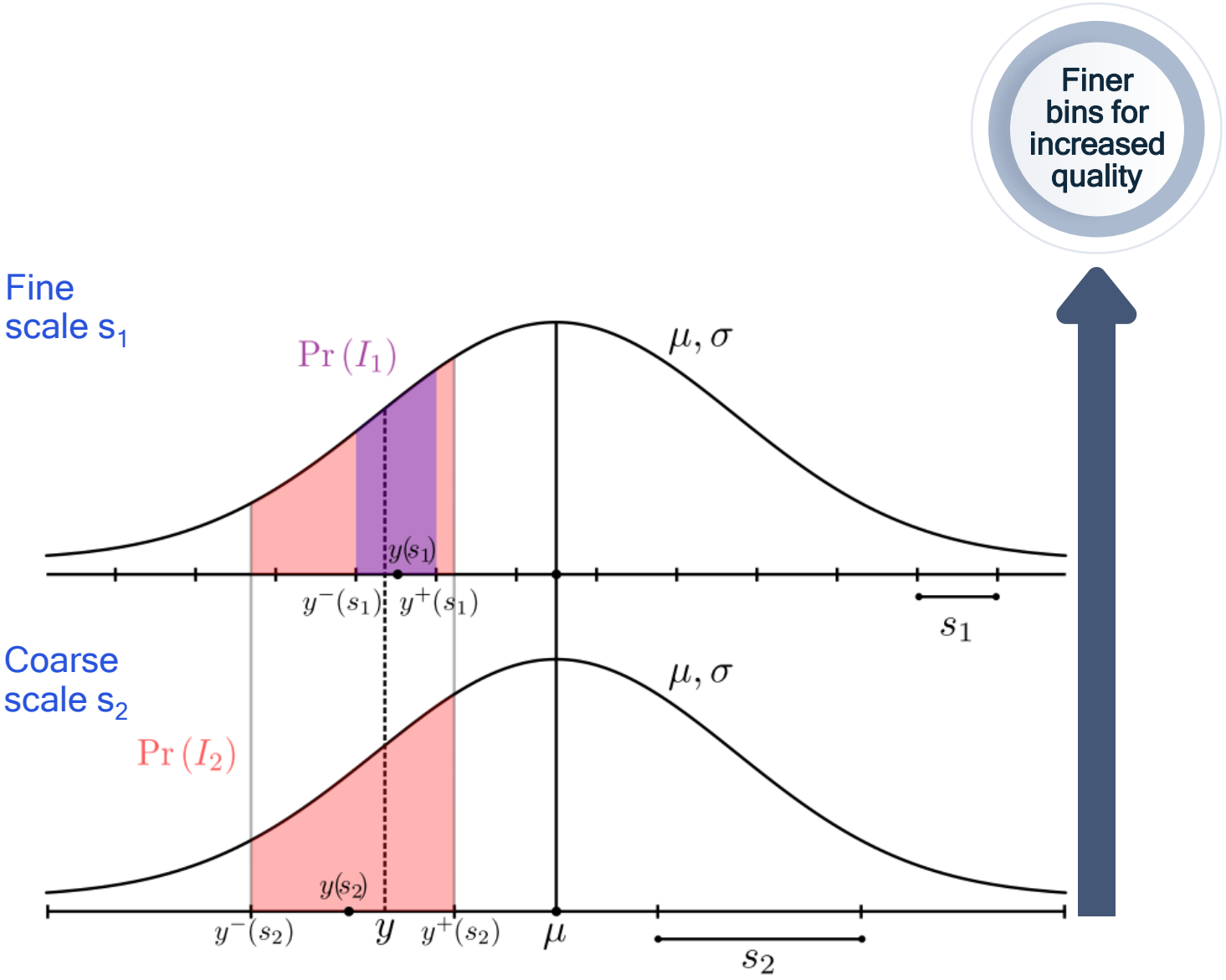
Level 1 approach that applies a scaling factor to the latent for a different trade-off between rate and distortion



g_a	Analysis transform / encoder
g_s	Synthesis transform / decoder
x	Input image
s	Scaling factor that determines the bitrate
μ, σ	Prior parameters
EC	Entropy coding
ED	Input image

Nested quantization for a single-model single-bitstream solution

Level 2 approach that uses multiple quantization levels and conditional probability to create a single bitstream with multiple quality levels



Variable-bitrate progressive neural image compression

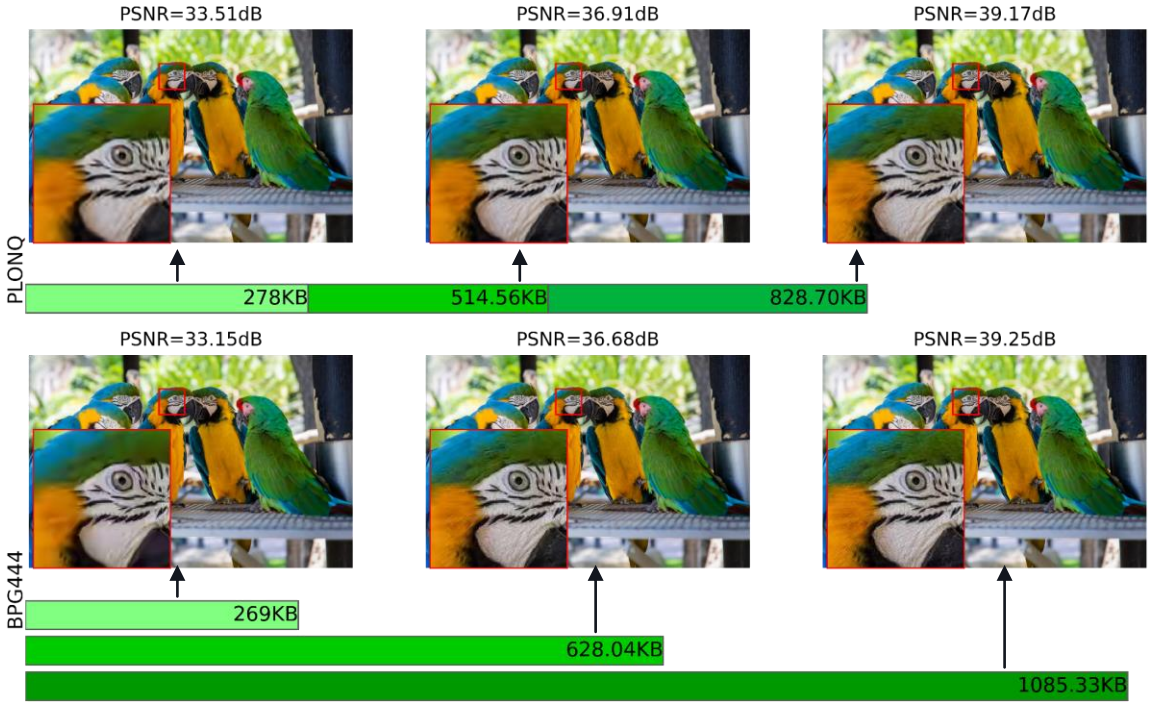
Achieves comparable performance to HEVC Intra but uses only a single model and a single bitstream

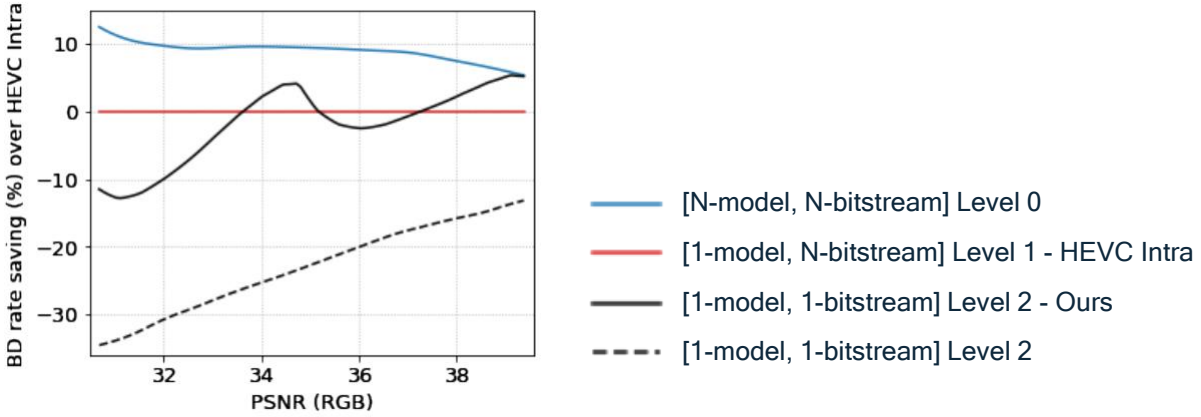
Our scheme

A single bitstream where the prefixes decode to different qualities

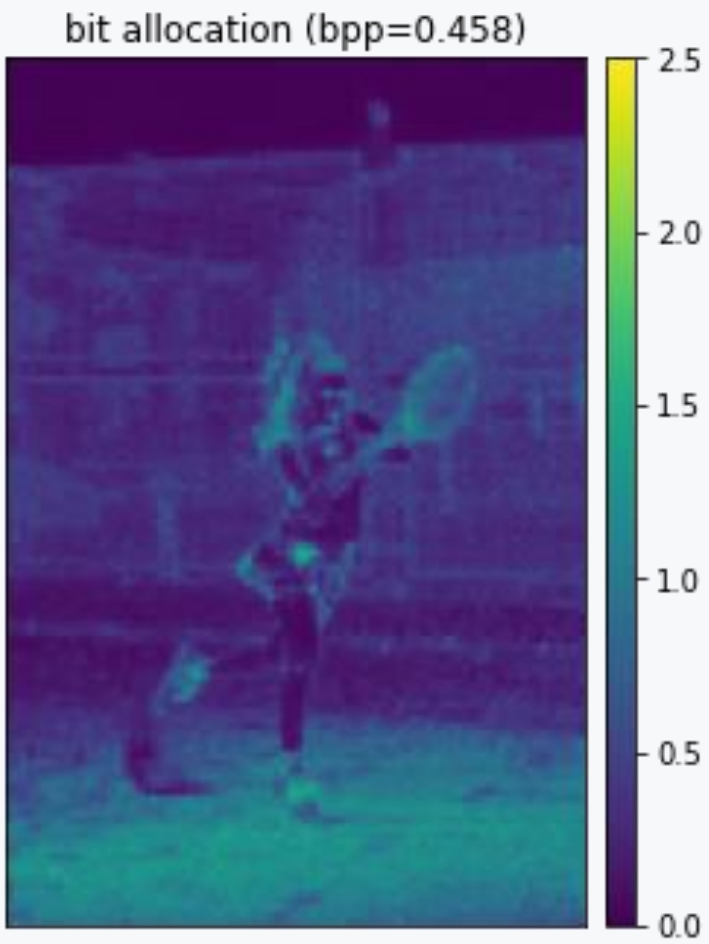
HEVC Intra

Separate bitstreams, one for each quality

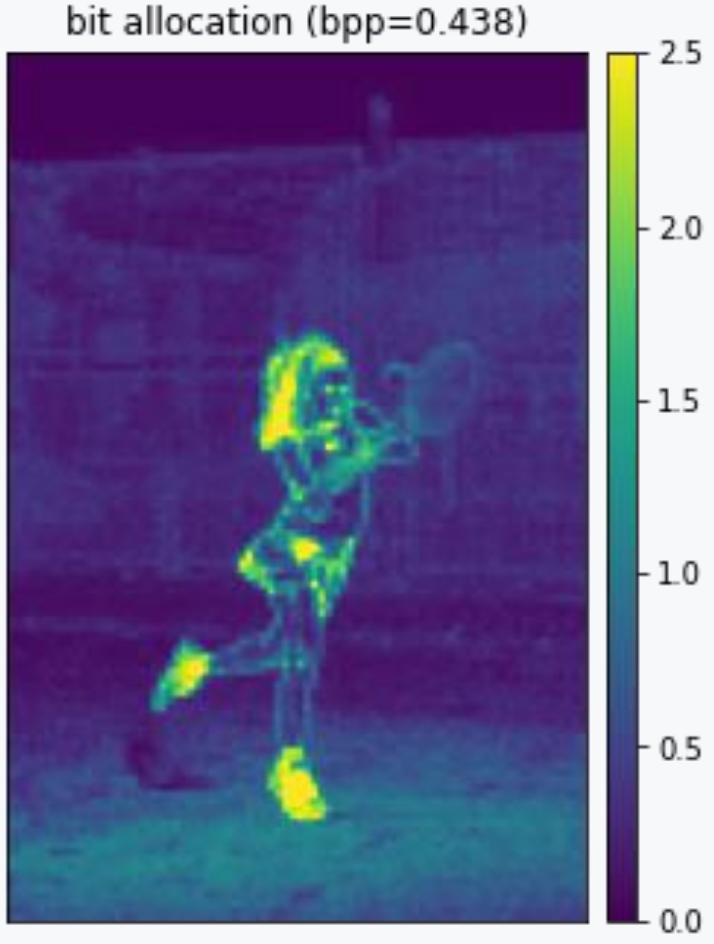




Source



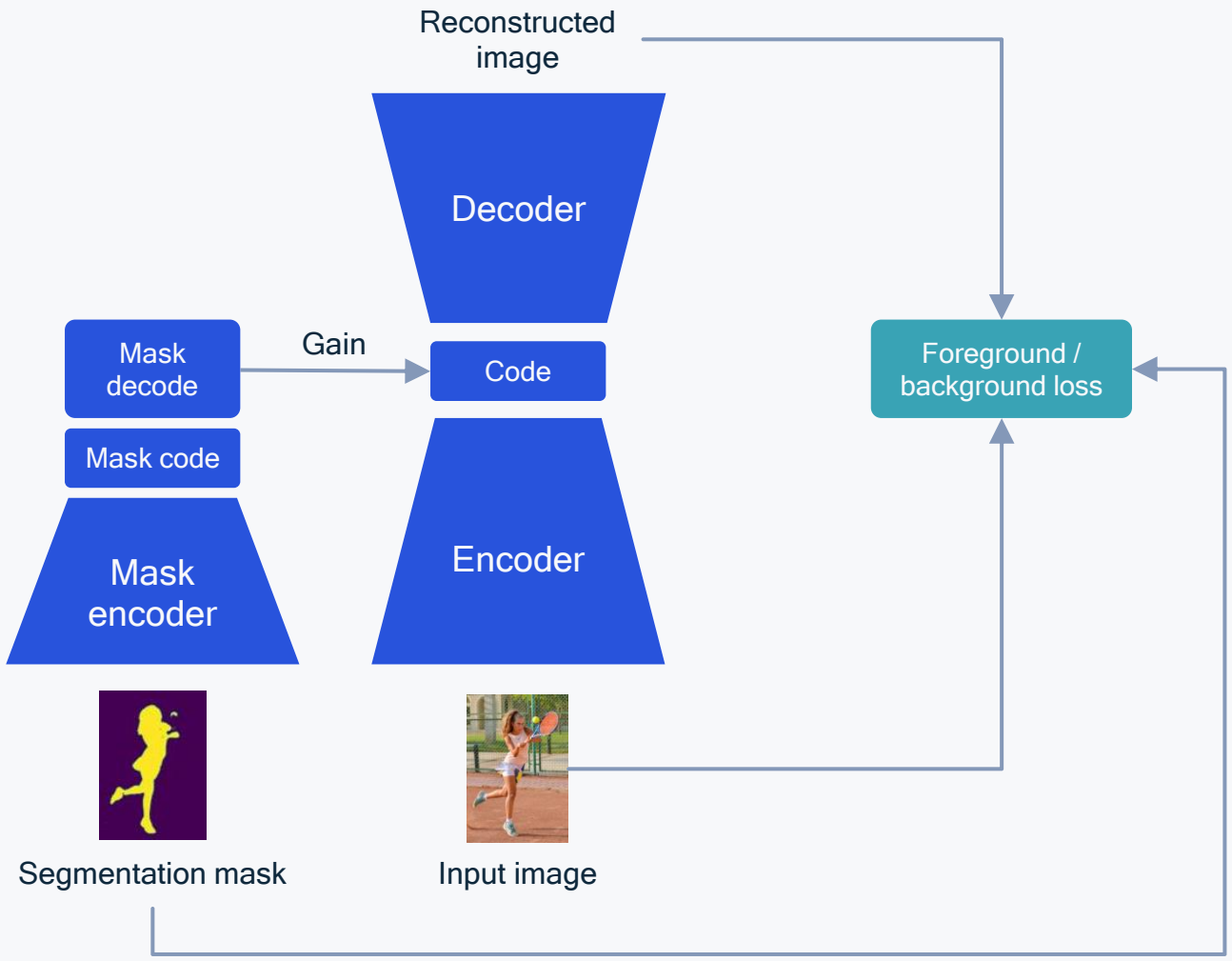
Baseline bit allocation



Semantic-aware bit allocation

Semantic-aware image compression for improved quality

Allocate more bits to regions of interest



Semantic compression focuses on regions of interest

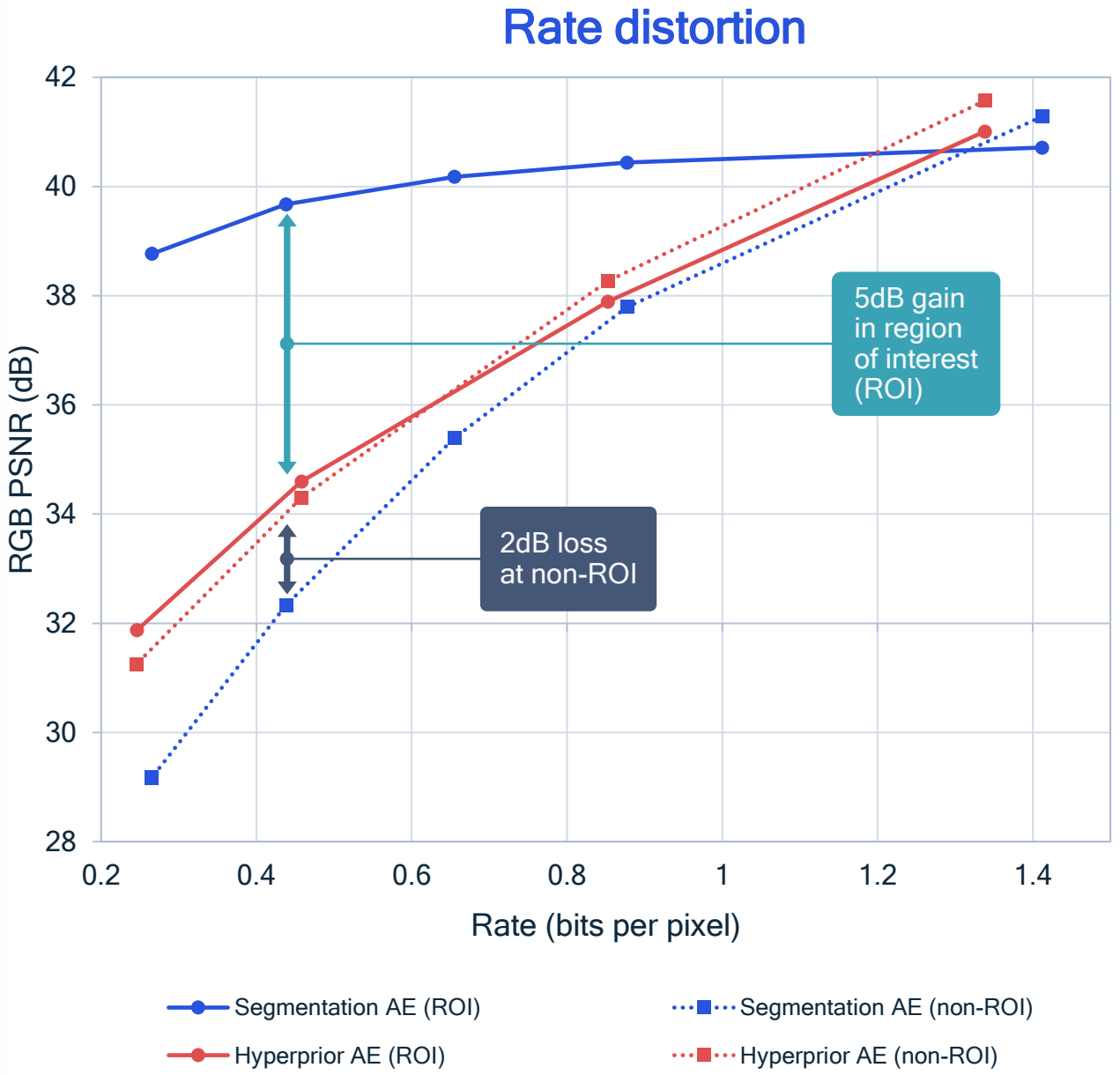
End-to-end deep learning on foreground/background weighted distortion loss

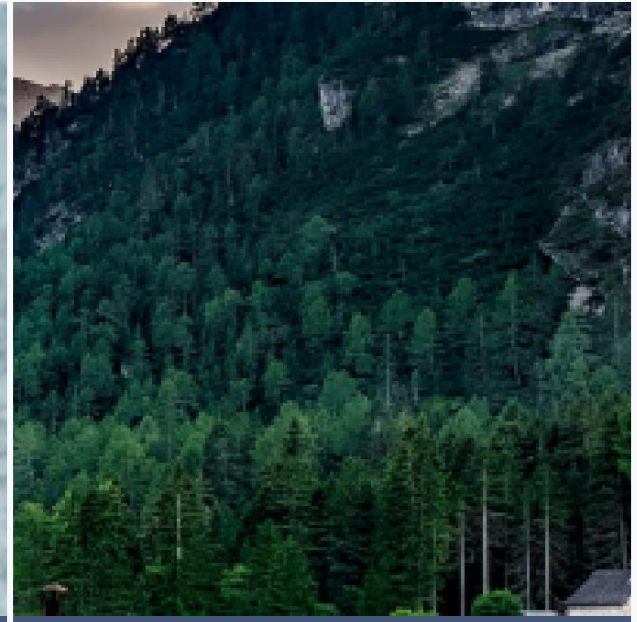
Semantic-aware image compression improves quality where it matters

State-of-the-art results for rate-distortion tradeoff

Next step

Extend to video compression

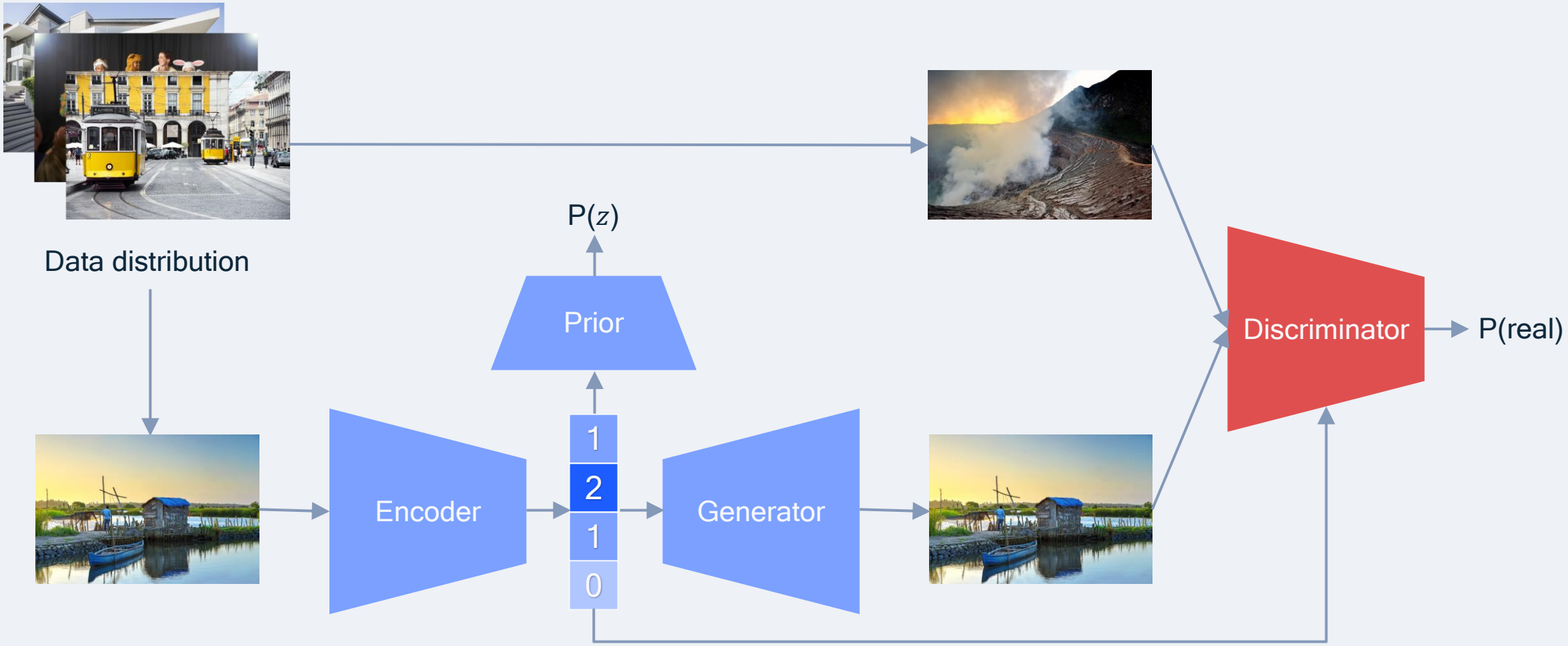




Original	Rate + Distortion	Rate + Perception	Rate + Distortion + Perception
Distortion	Very good	Very bad	Good
Perceptual quality	Very bad	Very good	Good

The tradeoff between perception and distortion metrics

Optimizing distortion and/or perceptual quality leads to different reconstructions



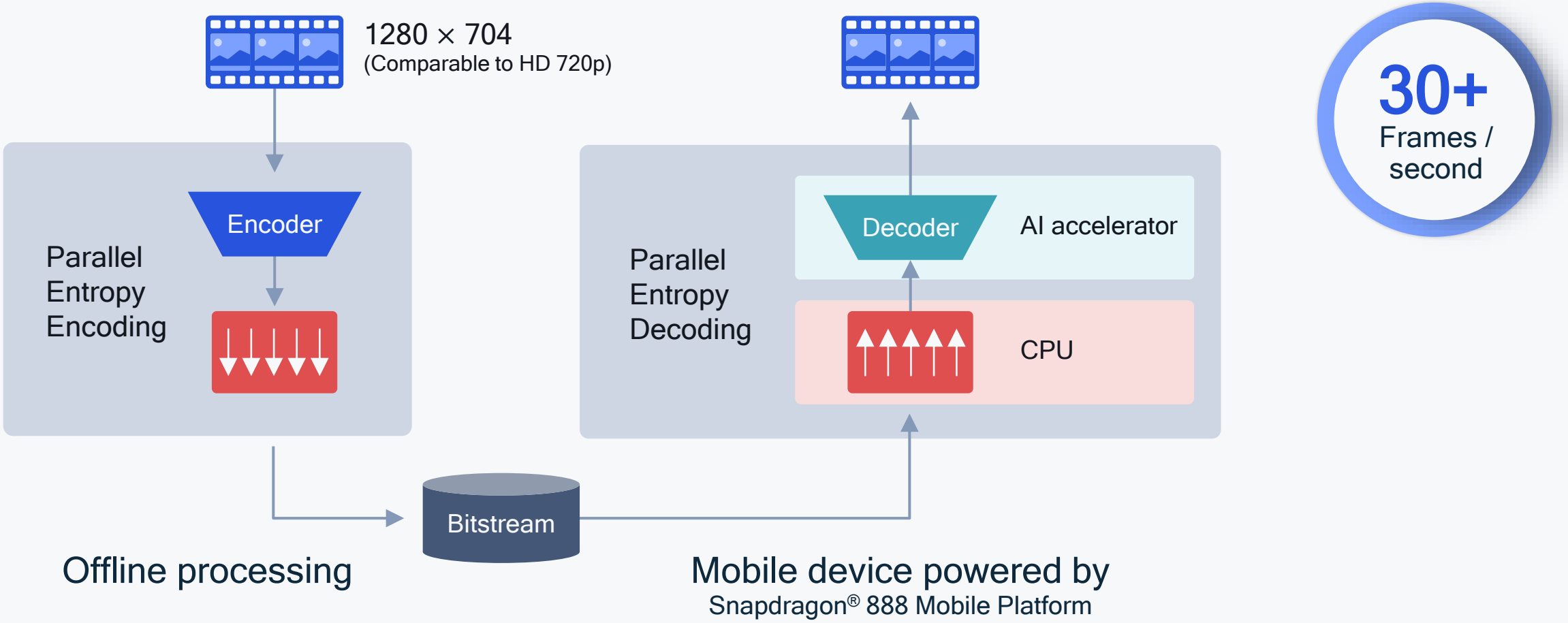
How GAN-based codecs work

GAN-based codecs can produce much more visually appealing content at very low bits per pixel


GAN
14.5 kB
0.221 bpp
145x reduction
Raw
2097.15 kB
8.0 bpp
512x1024
BPG QP=40
16.4 kB
0.250 bpp
128x reduction

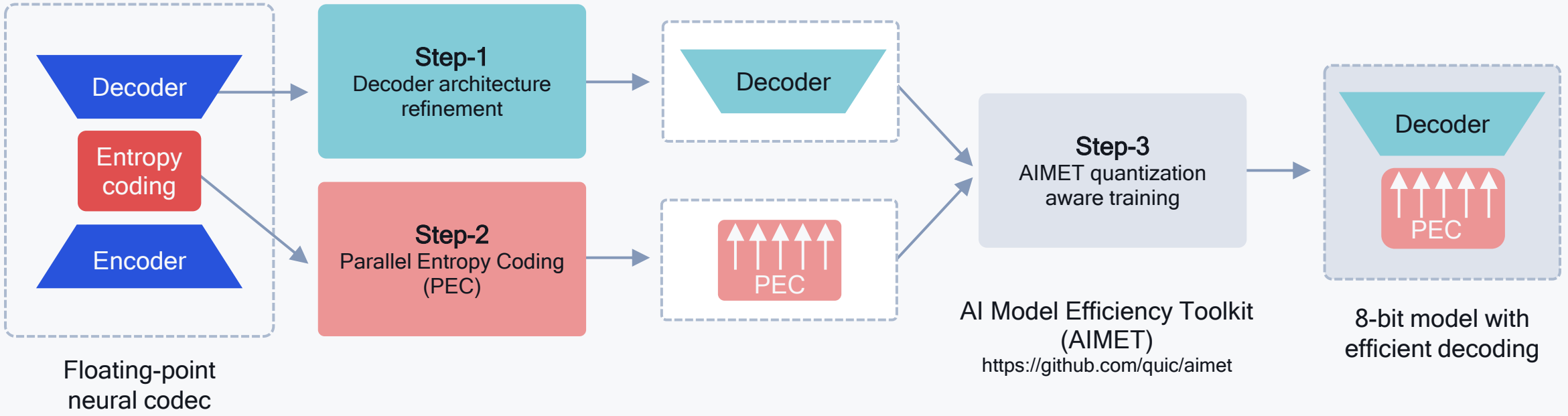
Real-time on-device neural video decode is now possible

Research demo showcases all-intra neural video decode on a smartphone. Inter-frame decoding is next



Efficient on-device neural decoding

Limit to I-frame (intra-frame) component of neural video decoding; Inter-frame decoding is coming soon



Parallel entropy coding + architecture refinement + 8-bit → Fast inference

AIMET quantization aware training → Improve quantized model accuracy

Qualcomm
AI research

Real-time neural video decoding on a mobile device

demo video

Challenges for mass deployment of neural video compression

Computational efficiency

Real-time on-device video encoding and decoding is still a challenge

Perceptual quality

Need to develop differentiable perceptual quality metrics

GANs are promising, but not perfect

Video GANs are challenging to train due to computational requirements

Rate distortion improvement

Bitrate needs to continue to get smaller for the foreseeable future as video demand increases

New modalities

Lots of work to do on new modalities like point clouds, omnidirectional video, multi-camera setups (e.g. in AV), etc.

Qualcomm

AI-based compression has the potential to more efficiently share video and voice across devices

We are conducting leading research and development in end-to-end deep learning for video and speech coding

We are solving implementation challenges and demonstrating a real-time neural video decoder running on a smartphone

Questions?

Connect with Us

www.qualcomm.com/ai

www.qualcomm.com/news/onq

@QCOMResearch

https://www.youtube.com/qualcomm?

http://www.slideshare.net/qualcommwirelessevolution

Qualcomm

Thank you

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog