# Beyond image/video compression standards

Prof. Dr. Touradj Ebrahimi





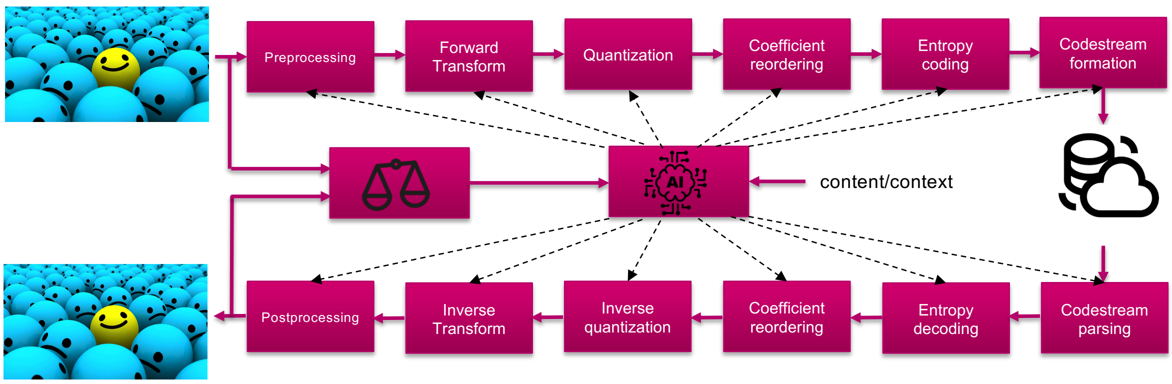
## How to take advantage of AI in image coding



- Three possible alternatives
  1. AI-assisted codec optimization
  2. Replacement of components of conventional model-based image coding by ML/AI tools (e.g. NN)
  3. End-to-end AI

# AI assited codec optimization

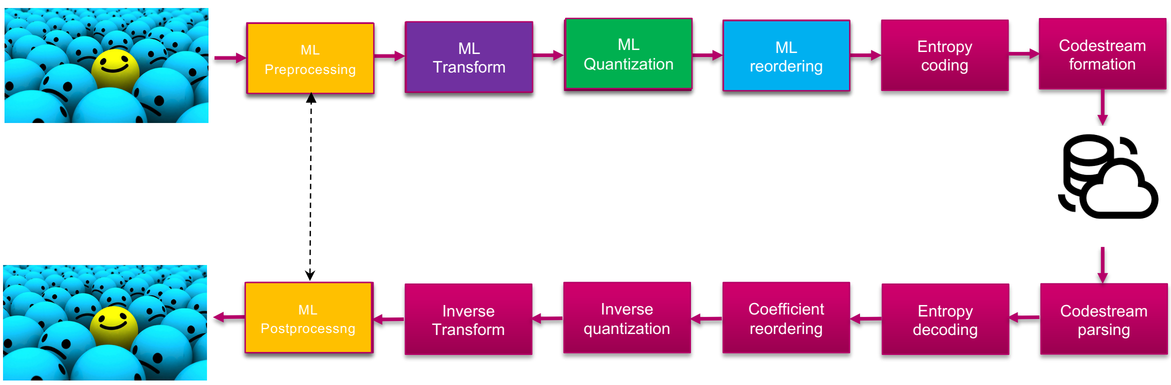
3






# Component replacement by ML/AI tools

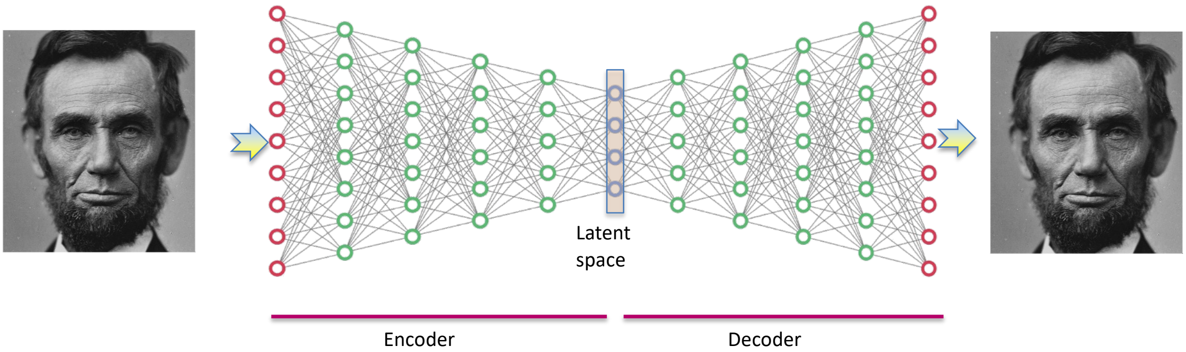
4



5

# End-to-end AI





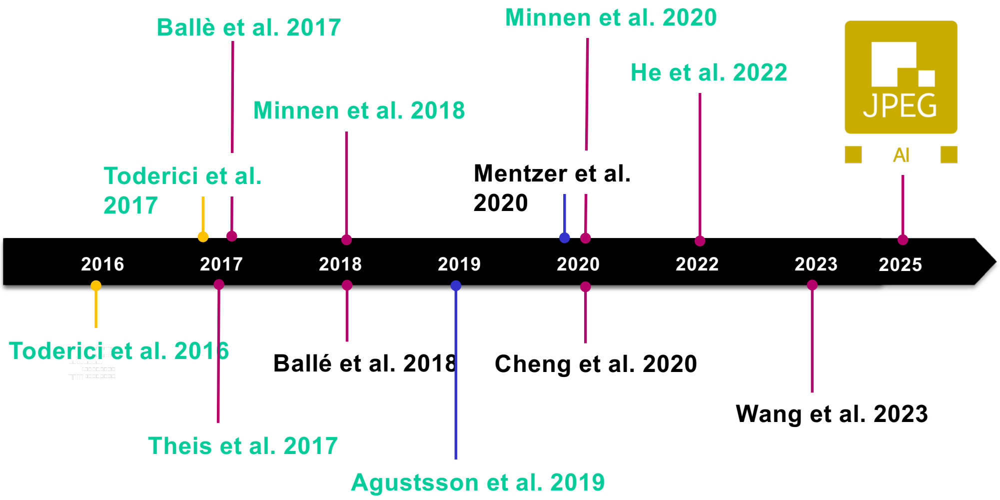

**Autoencoder**





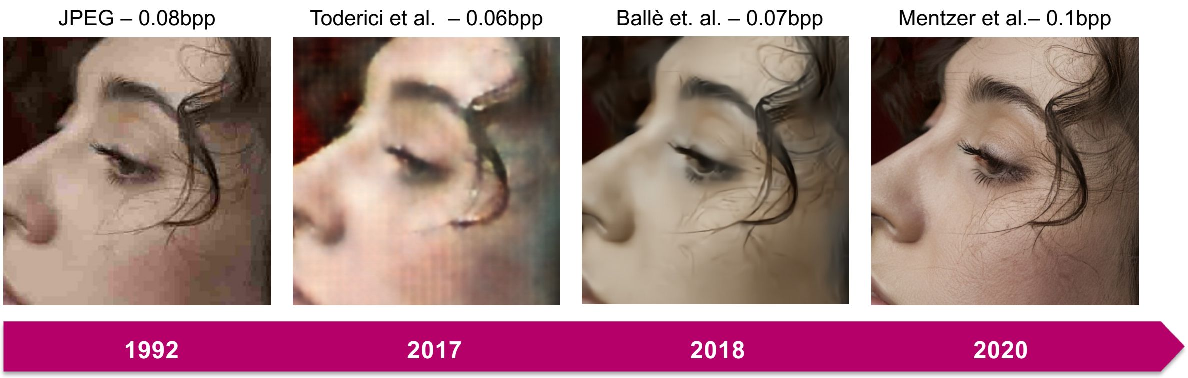
# End-to-end AI-powered image coding





## A radically new and powerful coding paradigm

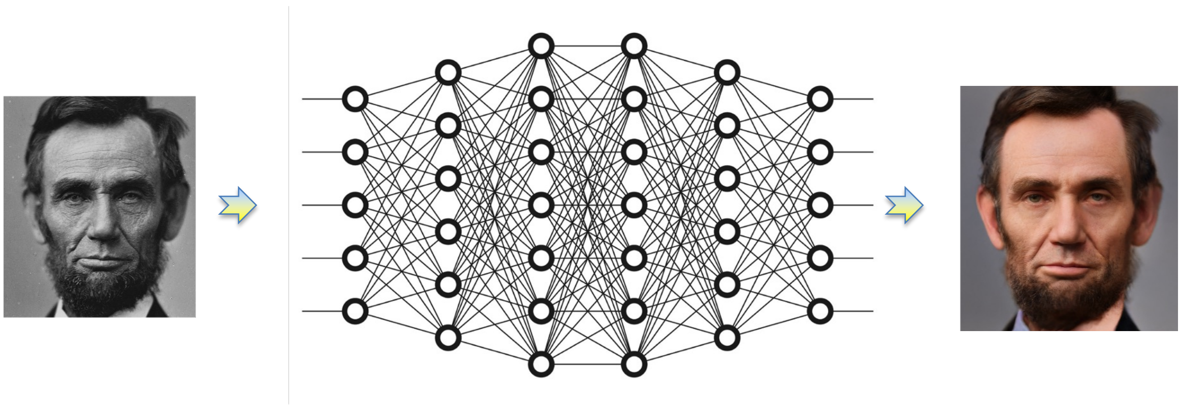






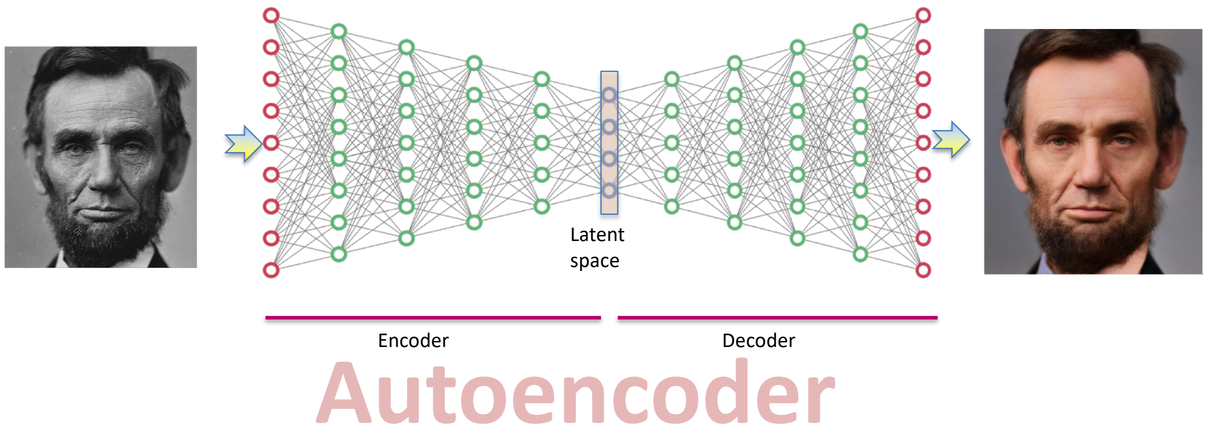

## AI-powered image enhancement





## AI-powered image enhancement

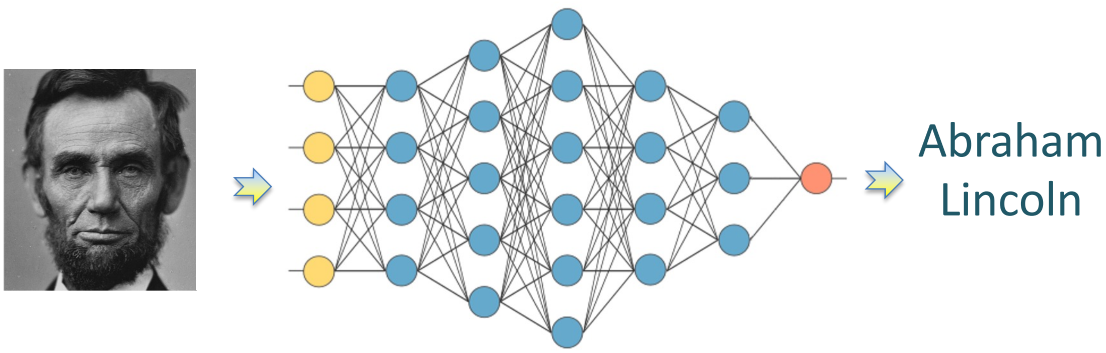
9





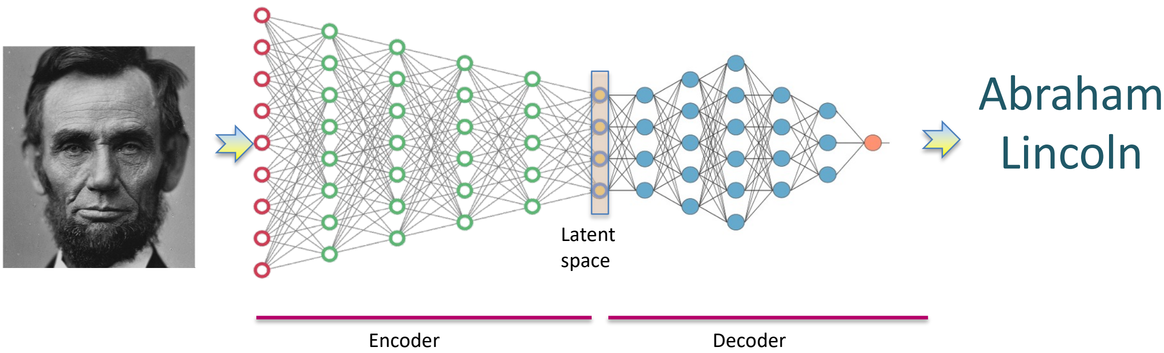

## AI-powered machine vision

10



## AI-powered machine vision





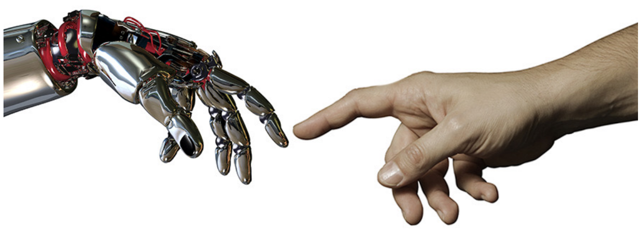

Autoencoder



## Image coding for humans and machines



- There is more in AI based image coding that meets the eyes
  - AI is extensively used in image enhancement and machine vision
  - An image coding for machines and another for humans divides the ecosystem
  - Can we put both under the same framework?



# JPEG AI scope



The JPEG AI scope is the creation of a learning-based image coding standard offering a **single-stream**, **compact**, compressed domain representation, targeting both **human visualization**, with significant compression efficiency improvement over image coding standards in common use at equivalent subjective quality, as well as effective performance for **image processing** and **computer vision** tasks, with the goal of supporting a **royalty-free baseline**

**The JPEG AI is a joint standardization effort between ISO/IEC JTC1/SC29/WG1 and ITU-T SG21**

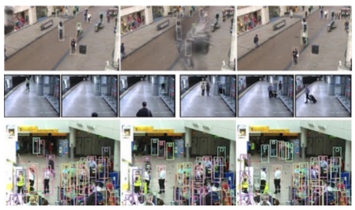
*JPEG AI is the first coding standard that leverages artificial intelligence for superior coding efficiency while taking into account the requirements of AI-powered content processing and analytics*

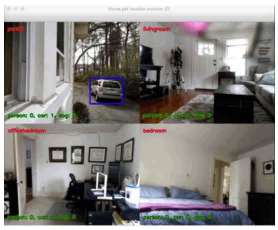


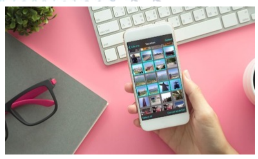
# Applications & Use Cases



- Mobile Imaging
- Cloud storage
- Visual surveillance
- Autonomous vehicles and devices
- Image collection storage and management
- Live monitoring of visual data
- Media distribution
- 360º photo sharing







## Requirements



- **High coding efficiency is important for many applications such as cloud storage or media distribution**
  - *AI-powered architectures have demonstrated to be particularly efficient in coding*
- **Content understanding is vital for many applications such as visual surveillance, autonomous vehicles, image collection management, etc.**
  - *Objects may need to be recognized*
  - *Images may need to be classified for organization purposes*
  - *Actions or events may need to be recognized*
- **Content enhancement is desirable in many applications such as in media distribution**
  - *Noise can be reduced*
  - *Resolution can be increased*
  - *Colors can be corrected*
- **Separate content coding for machines and content coding for humans hinders interoperability and increases complexity in applications that require both**
  - *Why two codecs when one can do both?*
  - *In many applications, the same content is consumed by both humans and machines*
  - *Compatible with user-centric AI (human in the loop) approach*



## JPEG AI strategy



**VERSION 1 (already published)**

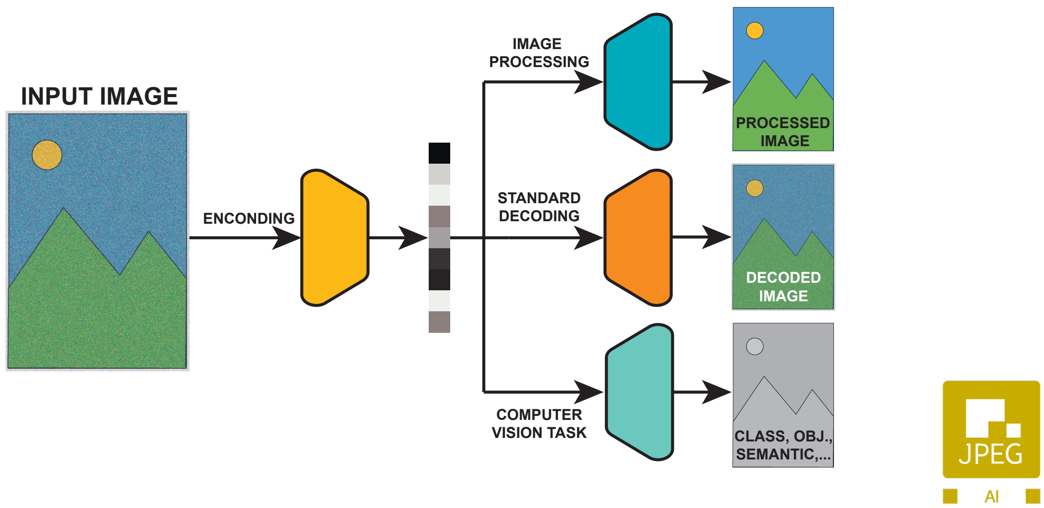
- Version 1 addresses several (but not all) JPEG AI ‘core’ and ‘desirable’ requirements with emphasis on compression efficiency.

**VERSION 2 (in progress)**

- Version 2 will address/include:
  1. JPEG AI requirements not yet addressed in Version 1, e.g. processing and computer vision tasks
  2. Significantly improved tools for JPEG AI requirements already addressed in Version 1, e.g. compression efficiency.

## JPEG AI triple-purpose framework



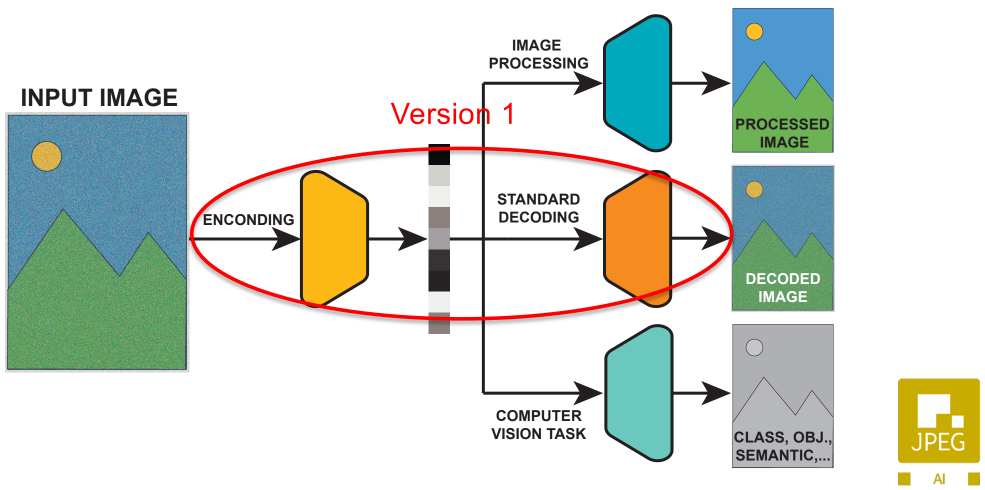







## JPEG AI (ISO/IEC 6048, ITU-T Rec.840)





# JPEG AI Version 1



| Part | Title | WD | CD | DIS | FDIS | IS |
|---|---|---|---|---|---|---|
| **1** | **JPEG AI: Core coding system** | - | 23/11 | 24/04 | - | 24/10 → 25/06 |
| 2 | JPEG AI: Profiling | - | 24/07 | 25/01 | - | 25/07 |
| 3 | JPEG AI: Reference software | - | 24/07 | 25/01 | - | 25/07 |
| 4 | JPEG AI: Conformance | - | 24/10 | 25/04 | - | 25/10 |
| 5 | JPEG AI: File format | - | 24/07 | 25/01 | - | 25/07 |

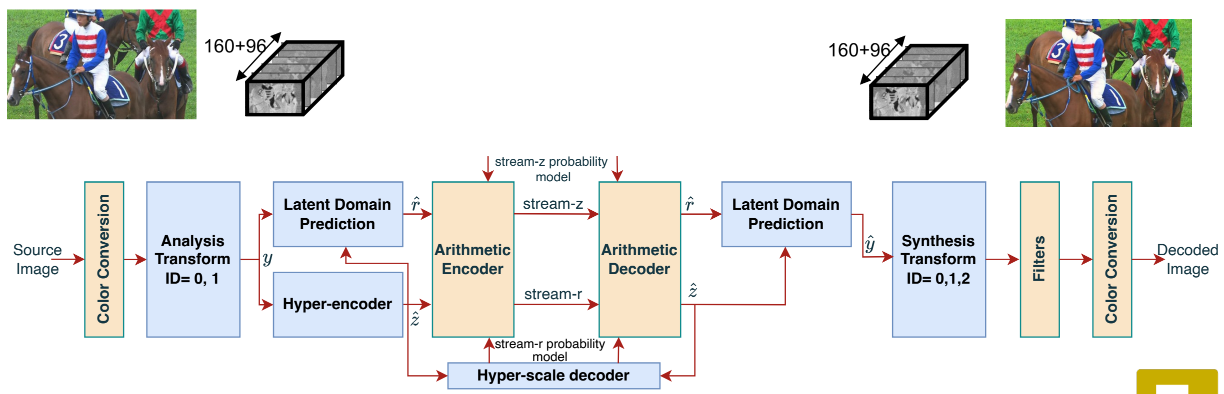
INFORMATION TECHNOLOGY STANDARDS

ITU-T Rec. T.840.1 ISO/IEC 6048-1



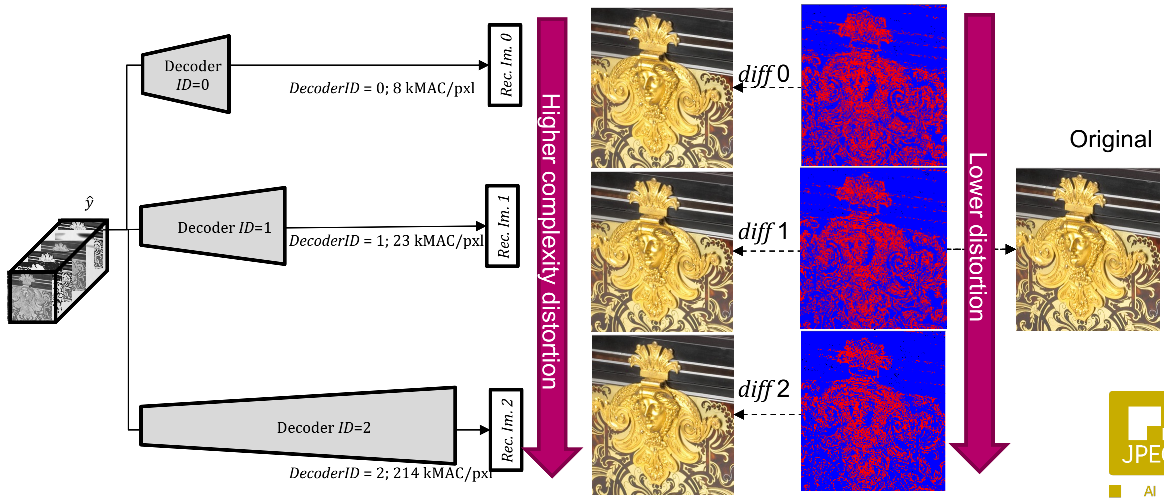
# JPEG AI Architecture





## JPEG AI: Single bitstream but multiple decoders








## Overall Performance

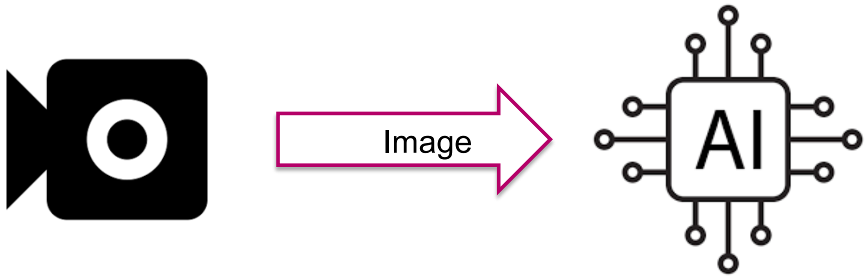


| Test | BD-Rate AVG | Dec. kMAC/pxl | Dec. Time GPU, × | Dec. Time CPU, × | Enc. Time GPU, × |
|---|---|---|---|---|---|
| **HEVC-SCC Intra** | 7.5% | - | 0.8 (CPU) | 0.8 (CPU) | 0.8 (CPU) |
| **VTM-11 Intra** | 0.0% | - | 1 (CPU) | 1 (CPU) | 1 (CPU) |
| **Enc0Dec0-tools-off** | **-12.0%** | **8** | **0.36** | **1.05** | **0.0005** |
| **Enc0Dec0-tools-on** | -16.2% | 14 | 0.41 | 2.4 | 0.0011 |
| **Enc0Dec1-tools-off** | **-16.7%** | **23** | **0.38** | **2.1** | **0.0005** |
| **Enc0Dec1-tools-on** | -20.2% | 28 | 0.41 | 3.3 | 0.0011 |
| **Enc1Dec2-tools-off** | **-24.0%** | **212** | **0.61** | **214** | **0.0012** |
| **Enc1Dec2-tools-on** | -27.0% | 215 | 0.64 | 215 | 0.0018 |

| Codec | 8K Encoding, s |
|---|---|
| JPEG | 5 |
| HEVC / H.265 | 2689 |
| VVC / H.266 | 18725 |
| JPEG AI | **5 (Enc0) or 20 (Enc1)** |

## An observation about today's AI-powered image coding paradigm





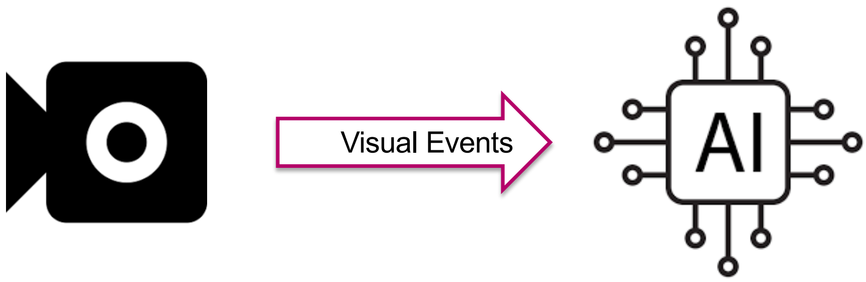

Not inspired by biological sensing

Inspired by biological computation (autoencoder / DNN)



## An observation about today's AI-powered image coding paradigm

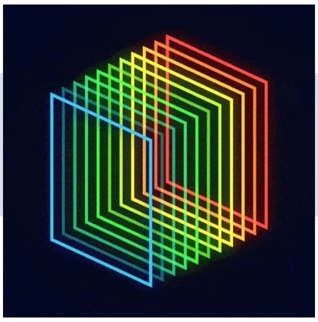




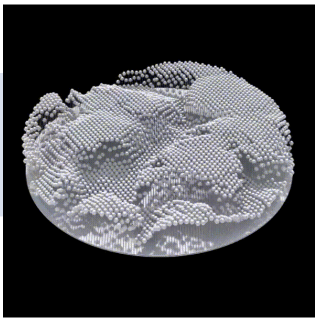

~~Not inspired~~ by biological sensing
Inspired

Inspired by biological computation (autoencoder / DNN)

## Event sensing



FRAME-BASED



EVENT-BASED

1. Sequential static pictures (frames)
2. Clock-driven (pre-defined frame rate)
3. Requires minimum exposure time
4. Fixed amounts of data in raw format
5. Appealing to human consumption

1. Continuous events (asynchronous intelligent pixels)
2. Scene-driven (1μs time resolution - 10’000 fps equivalent)
3. Little exposure time (120dB HDR / 40mlux low light sensitivity)
4. Amount of data varies with scene dynamics (10x to 1000x less)
5. Efficient data for machine vision





## Event-based imaging

- A lot of activities in machine vision and vision-based robotics
- Some activities in image processing
  - Event-based lossless and lossy coding
  - Event-based imaging performance assessment and metrics
  - Event-based image enhancement


Super-Resolving Blurry Images with Events, C Zhang, M Lin, X Zhang, C Jiang, L Yu
arXiv preprint arXiv:2405.06918, 2024•arxiv.org

## Scope



The scope of JPEG XE is the creation and development of a standard to represent Events in an efficient way allowing interoperability between sensing, storage, and processing, targeting machine vision applications.

**The JPEG XE is planned to become a joint standardization effort between ISO/IEC JTC1/SC29/WG1 and ITU-T SG21."**

*JPEG XE will be the first international standard for coding of events*



## JPEG XE use cases



- Embedded edge computing
  - Eye tracking, SLAM, AR/VR, obstacle avoidance
- Mobile imaging
  - Still image deblur, video deblur, super slow-motion, AR, gesture recognition
- Industrial machine vision
  - Vibration monitoring, object counting, object scanning, object tracking, quality control
- Long-term storage of event data
  - On-demand slow motion, AI training, computer vision benchmarking
- Movie production
  - Motion capture, virtual production, live capture
- Gaming
  - Visual input interfaces, motion capture, gesture control
- Scientific and engineering measurement
  - High speed observations, space launch monitoring, lunar eclipse, falling starts, Space Situational Awareness

## Requirements



- **Efficient coding, storage, compression and transmission of event data**
  - Better compression
  - Low memory requirements
  - Low power requirements
  - Low end-to-end latency
  - Low complexity
  - Recovery of absolute time-base
  - Interoperability with neural network processing architectures
  - Random access
- **Interoperability** with major **interface** specifications and protocols
  - Ease of **adoption** by **industry protocols** (DCMI, MIPI, …)



## JPEG XE strategy



- Lossless coding
  - Urgently needed in many of today's use cases
  - Two scenarios
    - *Constrained scenario: Extreme low computational complexity, willing to sacrifice compression efficiency (phase 1)*
    - *Unconstrained scenario: More computational resources, aim for the highest possible compression efficiency (phase 2)*
- Lossy coding (phase 3)
  - Desired by many use cases and tasks in the near future
    - *Tasks allow assessing the impact of lossy compression.*
    - *Difficult to assess the impact of loss. Requires dedicated new metrics*

## Structure of DNA



- Double-stranded helix with four types of nucleotides
  - A (adenine)
  - C (cytosine)
  - G (guanine)
  - T (thymine)



## Applications and use cases



- Long Term Media Archives and Cultural Heritage Preservation
- Social Networks Cold Media Storage
- Preservation of Medical Images
- Preservation of Large-scale Repositories of Biomedical Data: Beyond Local Data Storage
- DNA Coding for Traceability

## Archival on DNA support



- An artificial DNA based on digital information
  - Digital: binary 0 or 1
  - DNA: quaternary A, C, G, or T

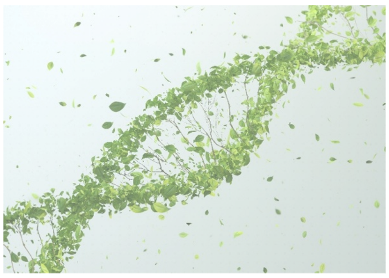
| | |
|---|---|
| 00 | A |
| 01 | C |
| 10 | G |
| 11 | T |



## Advantages of archival on DNA support



- Extremely **high density**
  - Roughly **hundreds of PB (=$10^{15}$ B)** can be stored **in 1 gram of DNA**
- Extremely **stable**
  - long-term preservation over a million years is possible
- **Eco-friendly**
  - No extra power sources are needed in an ideal condition
  - Very small carbon footprint



**Highly Sustainable**

## Challenges in archival on DNA support

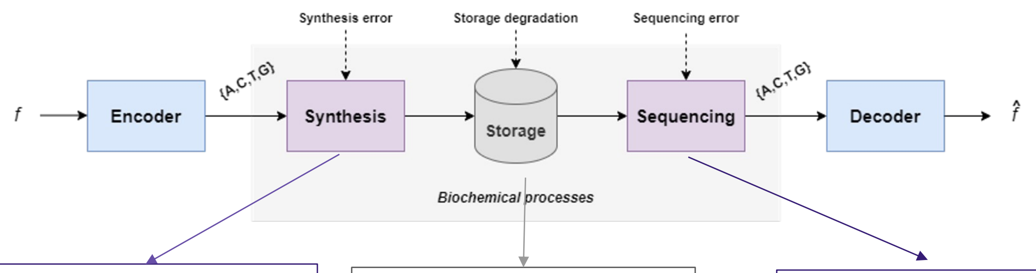


- Dealing with errors in synthesizing/sequencing DNA
  - Substitution: ATTGCAGC → ATTACAGC
  - Insertion: ATTGCAGC → ATTGACAGC
  - Deletion: ATTGCAGC → ATTCAGC
- Compliance with biochemical constraints
  - Simple conversion from binary to quaternary can't solve the problems!



## A biologically constrained problem






– Length limited to 300 nts
– Errors increase exponentially for longer oligos

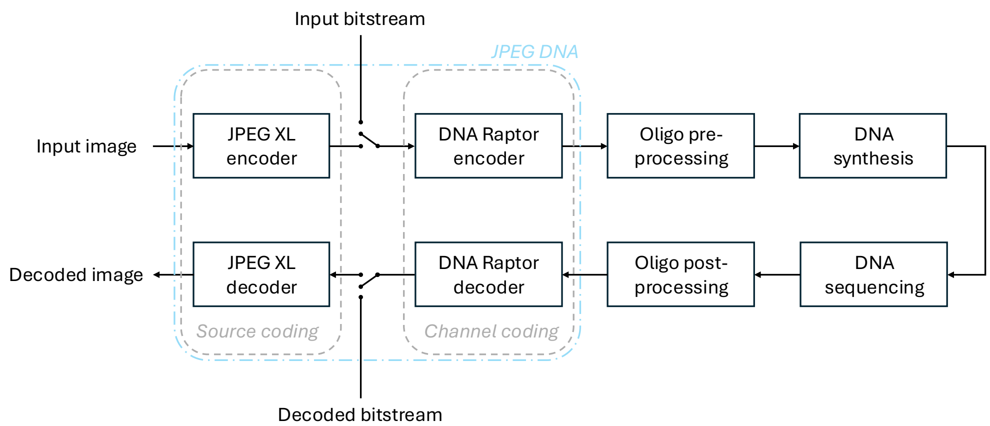
– DNA is prone to chemical decay if not stored under optimal conditions.

– Main source of errors
– Higher error if: homopolymers, patterns and unbalanced %GC

- “The scope of JPEG DNA is the creation of a standard for **efficient coding of images** that considers biochemical constraints and offers robustness to noise introduced by the different stages of the storage process that is based on DNA synthetic polymers”



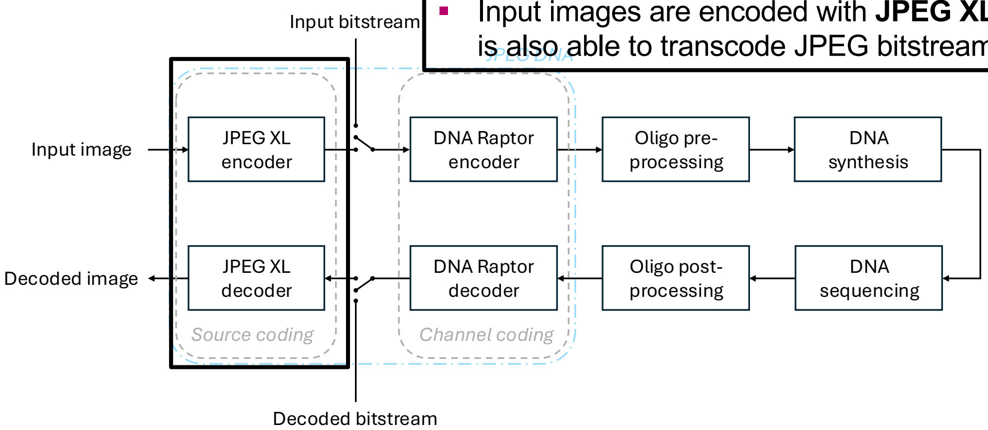




## JPEG DNA coding workflow



- Source and channel coding are **separated**
- Input images are encoded with **JPEG XL**, which is also able to transcode JPEG bitstreams




JPEG DNA

Multimedia Signal Processing Group
Ecole Polytechnique Fédérale de Lausanne

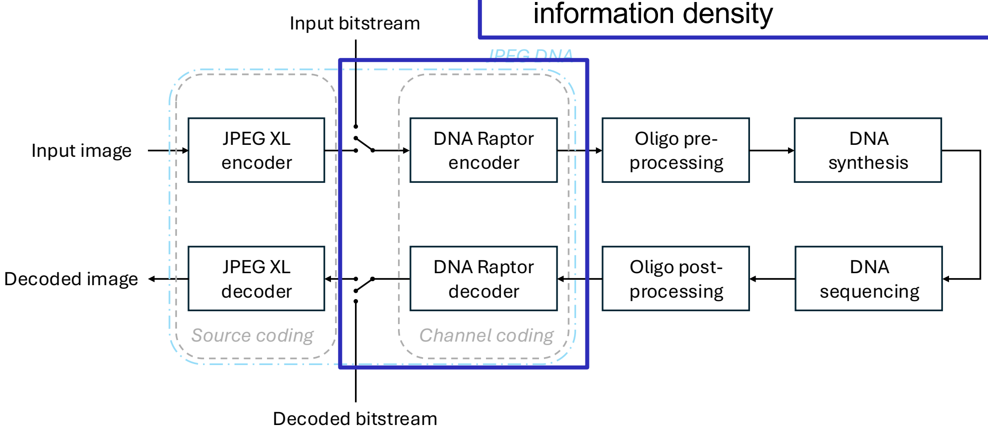
EPFL



## JPEG DNA coding workflow



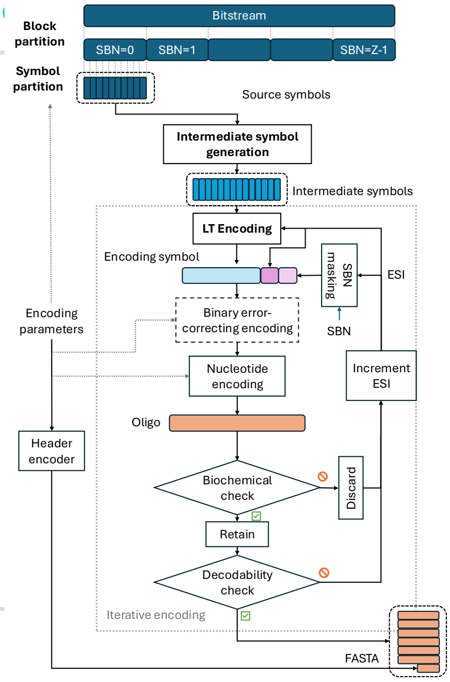
- Source encoding can be bypassed enabling encoding of **arbitrary binary inputs**
- **DNA Raptor codes** are used to encode the bitstream into nucleotides with high information density




JPEG DNA

Multimedia Signal Processing Group
Ecole Polytechnique Fédérale de Lausanne

EPFL



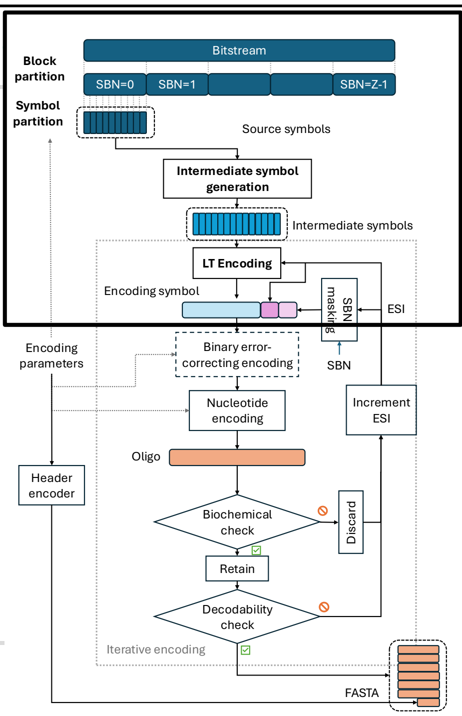


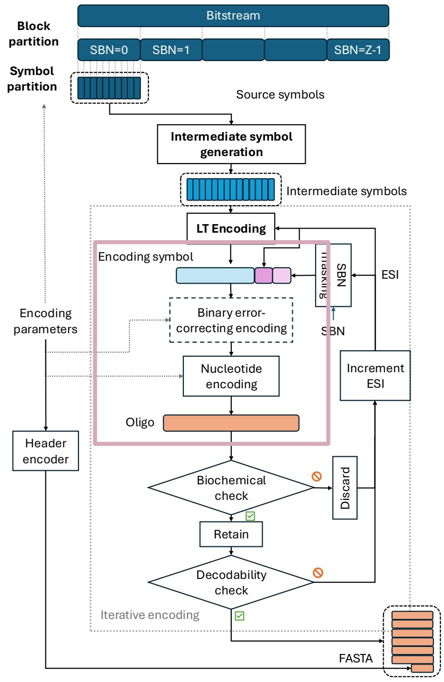







- The channel encoder partitions the bitstream into source symbols
- Encoding symbols are produced using Raptor codes as defined in RFC5053





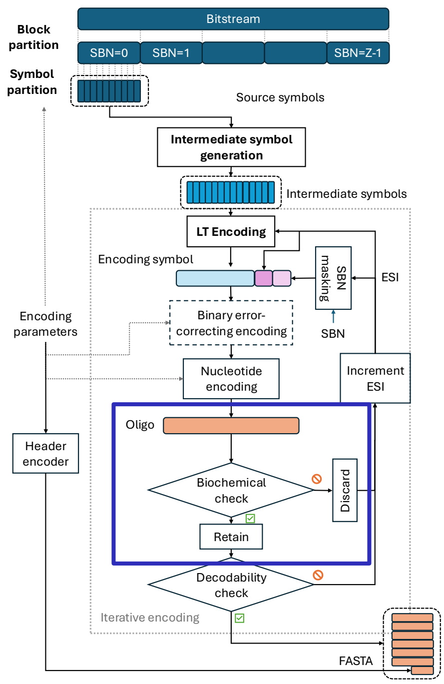

- Symbols with their identifiers are encoded into oligos







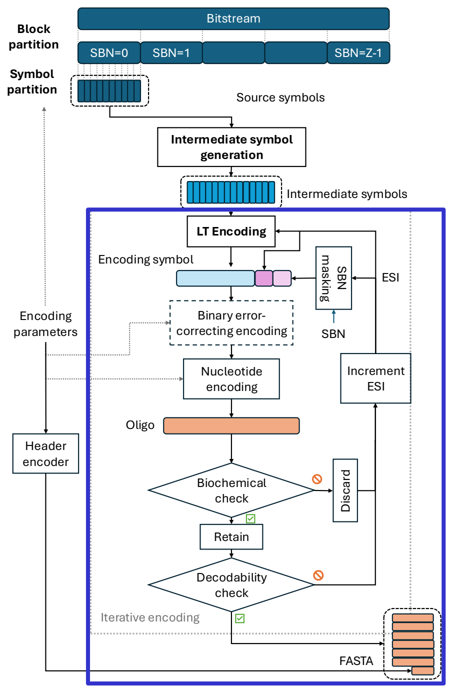





- Oligos not respecting imposed biochemical constraints are discarded

## JPEG DNA



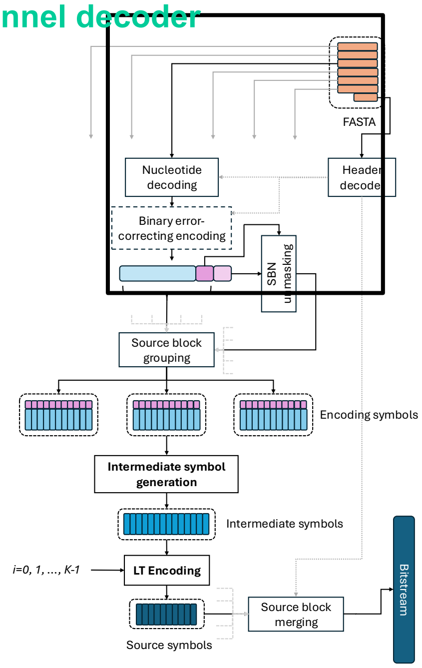



- The process is iterated until enough valid oligos are produced
- Encoding can be stopped when decodability is achieved to **minimize coding** rate or overhead can be produced for **error correction**



## JPEG DNA channel decoder



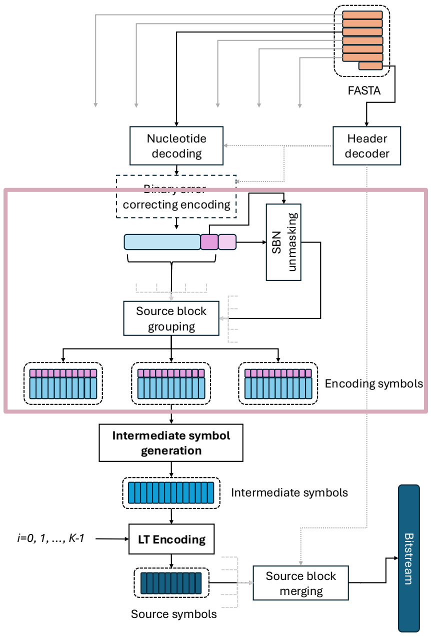



- Oligos are independently decoded back to binary representation

## JPEG DNA channel decoder



- Resulting packets are grouped according to their source block



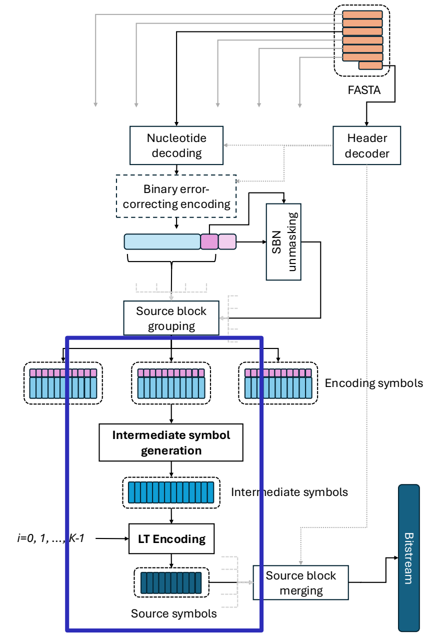





## JPEG DNA channel decoder



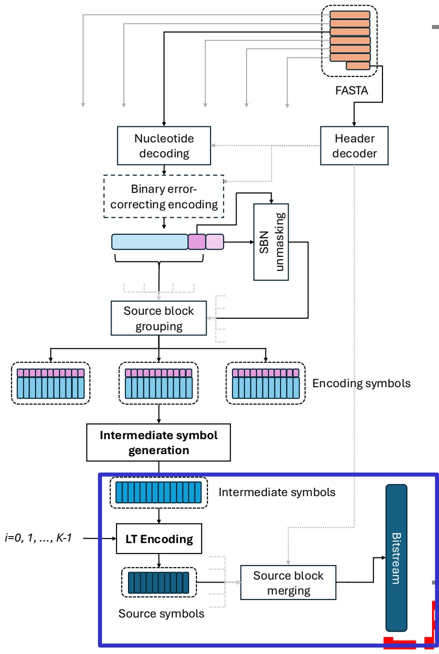
- Encoding symbols are decoded back to source symbols with RFC5053 Raptor codes



## JPEG DNA channel decoder



- Symbols are merged to reconstruct the bitstream

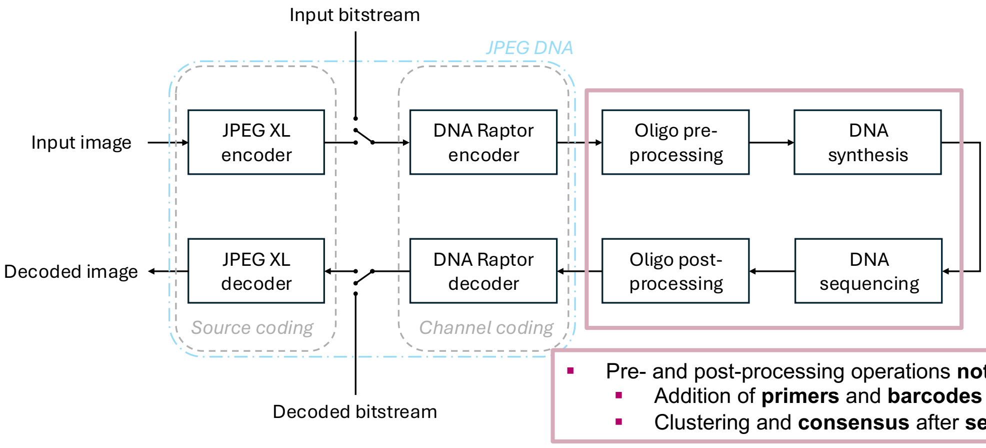







## JPEG DNA coding workflow



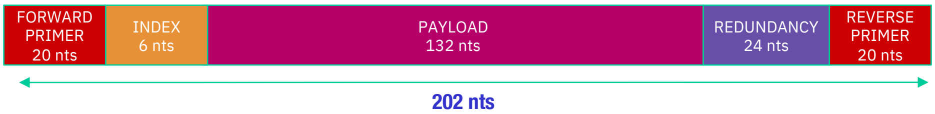



- Pre- and post-processing operations **not included** in the standard:
  - Addition of **primers** and **barcodes**
  - Clustering and **consensus** after **sequencing**

# Oligo (pre-/post-)processing



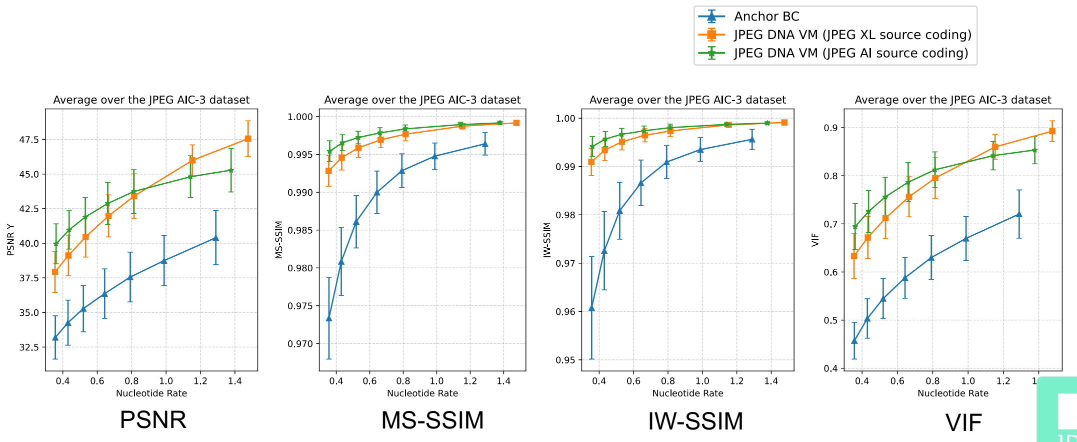
– Appropriate **oligo formatting is necessary** for synthesis and sequencing:
- ***Two primers*** *(forward and reverse): for PCR amplification*
  - **2x 20 nt**
- ***Index****: enable the addressing of different images inside the same dataset*
  - **6 nt**
- ***Redundancy****: parity nucleotides for error correction or error detection*
  - **24 Reed-Solomon (RS) nt**
- ***Payload****:*
  - **132 nt**





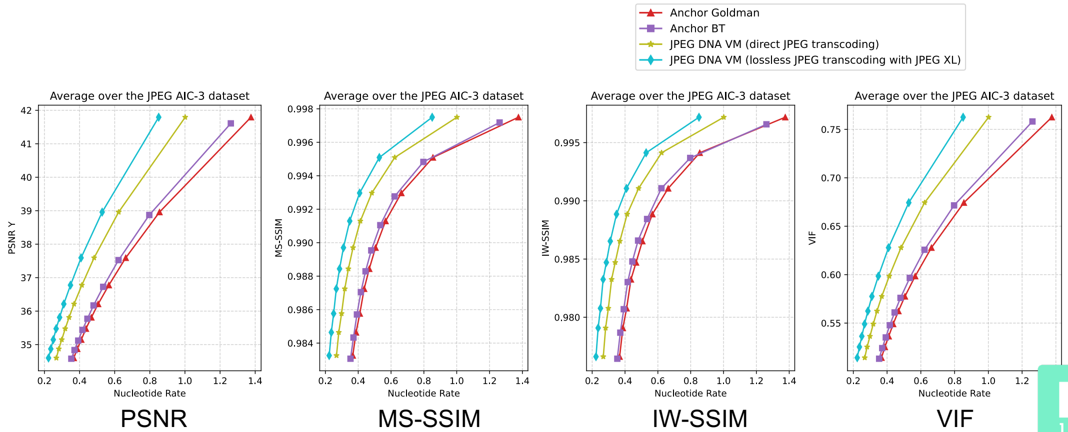

# Performance evaluation in direct encoding





## Performance evaluation in transcoding

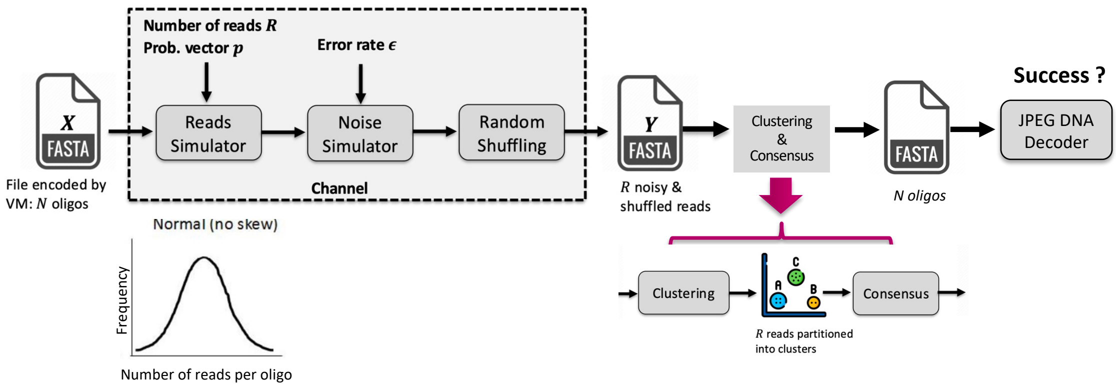







## Noisy channel model

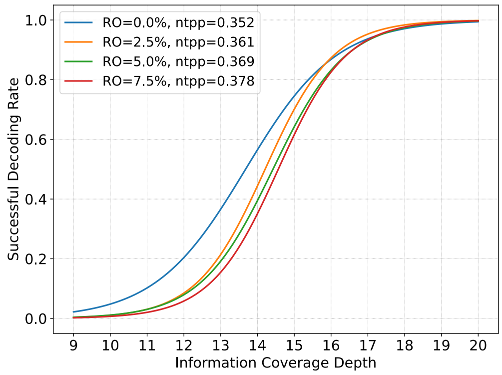




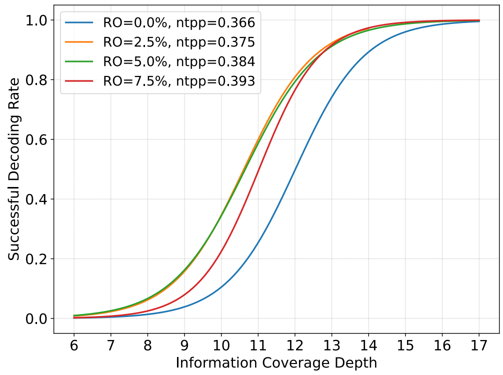

**Channel setup:** reads per oligo $(r_1, \ldots, r_N) \sim \mathrm{Multinomial}(R; \boldsymbol{p} = (\frac{1}{N}, \ldots, \frac{1}{N}))$; substitutions

## Performance evaluation in noisy channel






RS bytes = 0



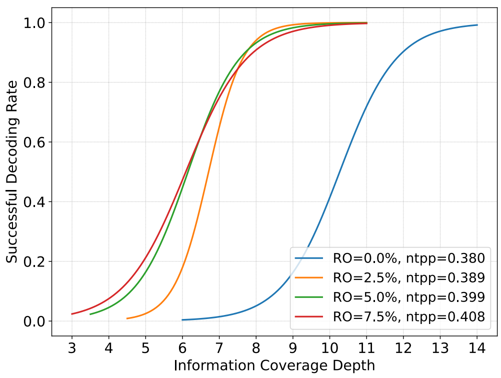

RS bytes = 2

- The JPEG DNA is able to achieve robust decoding at coverage depth of approximately 20 **without RS codes**
- The required **coverage** is **reduced** to approximately 17 with 2 bytes added per packet

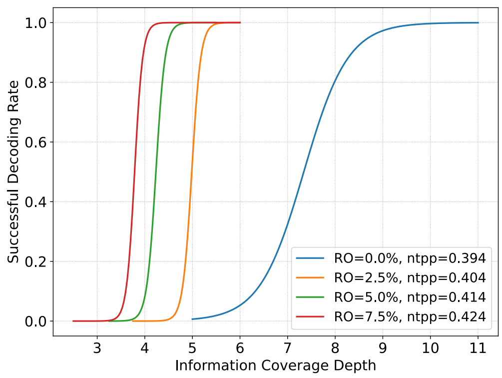


## Performance evaluation in noisy channel






RS bytes = 4

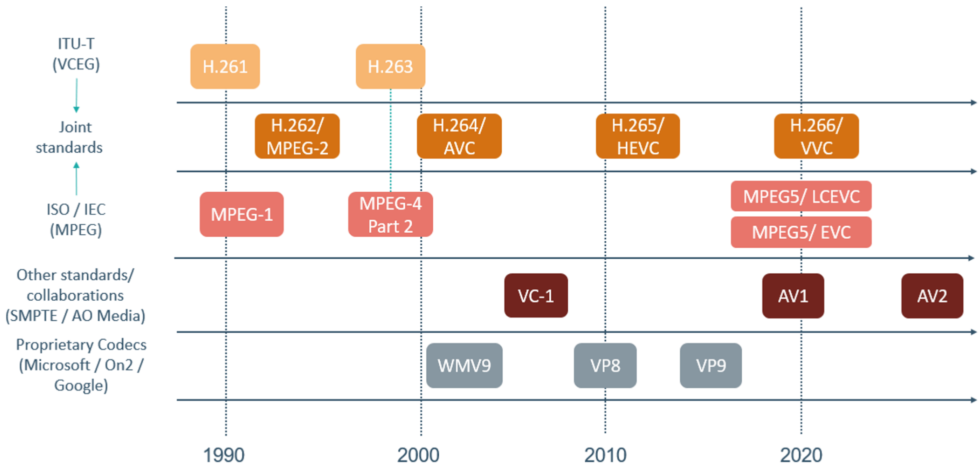



RS bytes = 6

- Increasing the number of redundant RS bytes **further reduces** the required coverage
- The use of **overhead** in the Raptor encoder is observed to be **beneficial** in combination with **intra-oligo error correction**

# Video coding trends

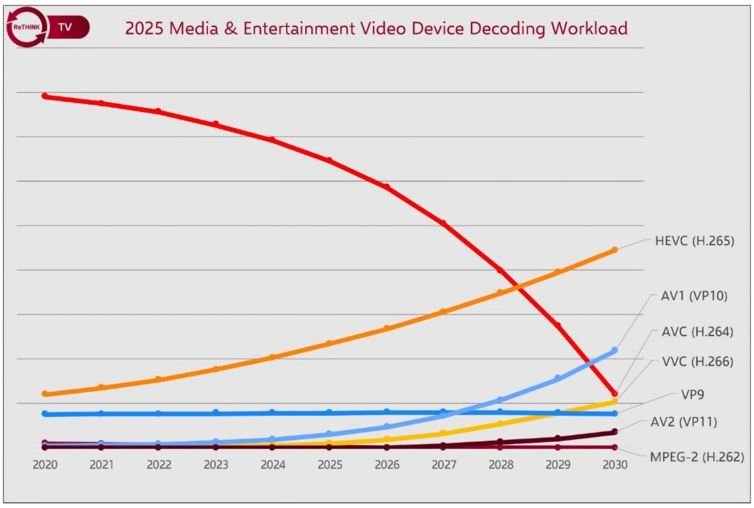






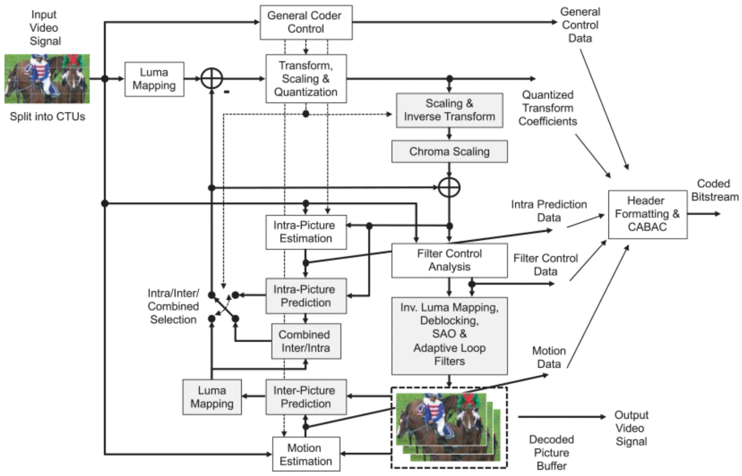

# Video compression trends





# H.266/VVC block diagram

59

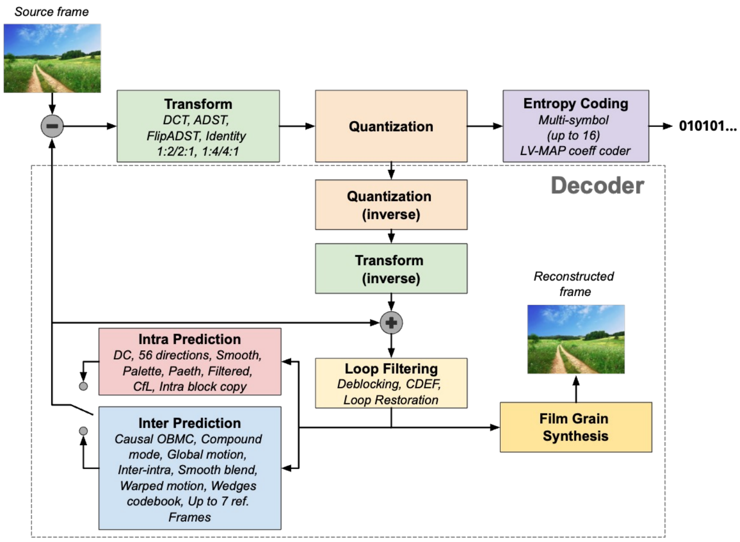






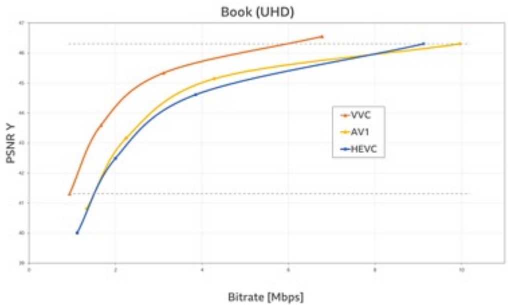

# AOM/AV1 block diagram

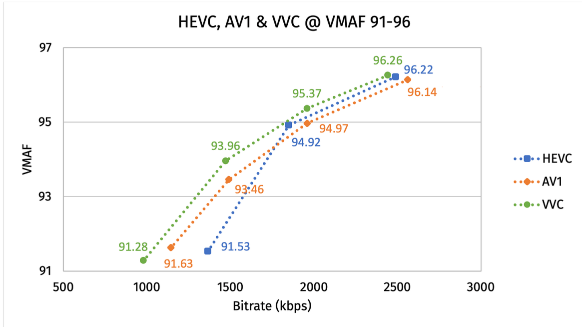
60



## HEVC vs AV1 vs VVC








## AI-powered video compression roadmap



- JVET – Neural Video Coding
- ITU-T SG16 – Video Coding + AI Media Systems
- AOMedia – AV2 Neural Tools
- MPAI – MPAI EVC + MPAI EEV

Multimedia Signal Processing Group
Ecole Polytechnique Fédérale de Lausanne
EPFL