

# JVET-AE0234 AHG9: Common text for proposed generative face video SEI message

B. Chen, J. Chen, Y. Ye (Alibaba)

S. Wang (CityU)

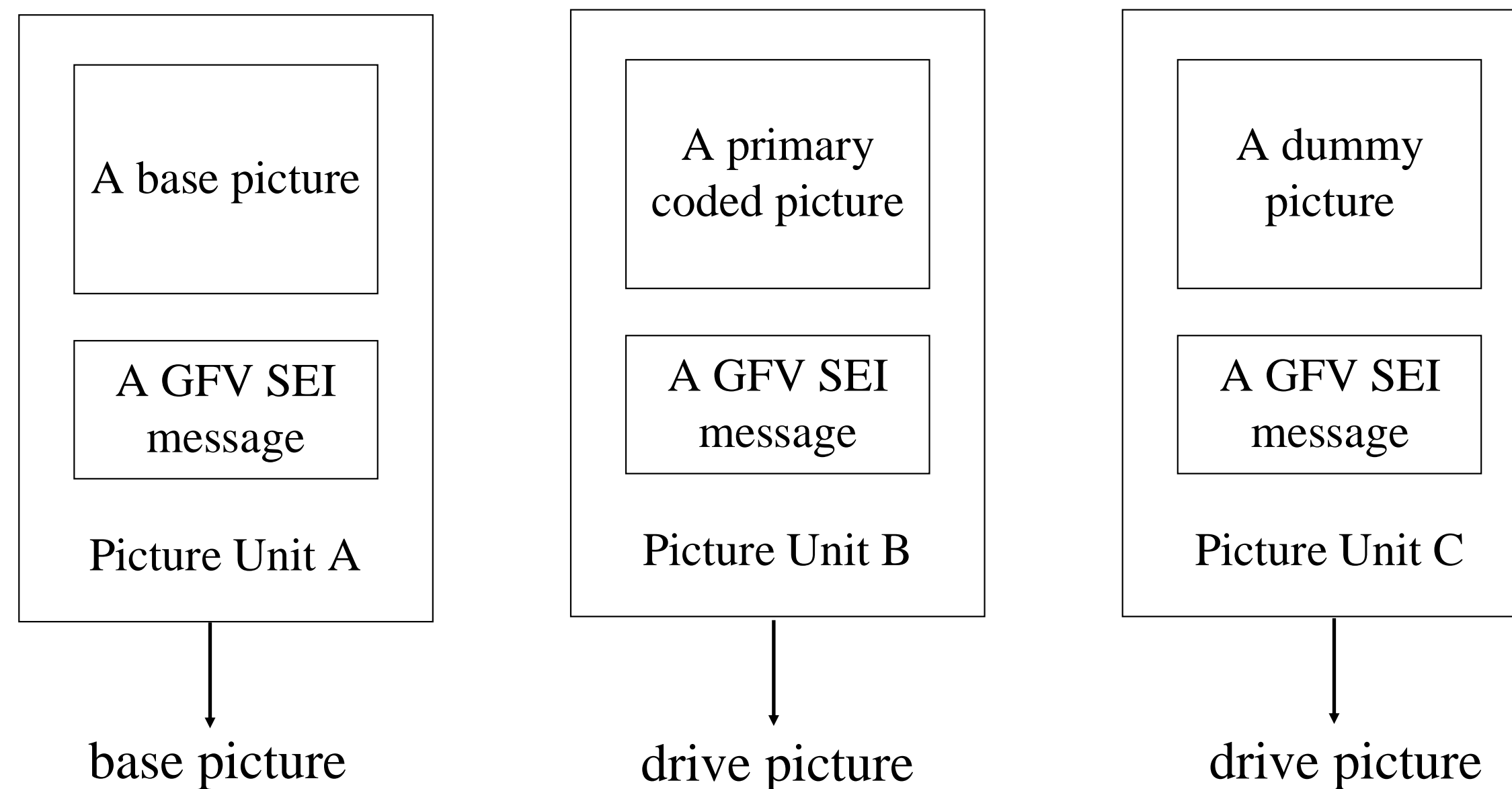
S. McCarthy, P. Yin, G.-M. Su, A. K. Choudhury, W. Husak, G. J. Sullivan(Dolby)

# Introduction

- Generative Face Video (GFV) SEI message were proposed in JVET-AC0088, JVET-AD0051 JVET-AE0080 and JVET-AE0083 in the previous meetings.
- In last meeting, JVET-AE0280 proposed a common syntax design of GFV SEI message and defined the interface between the decoder and the generative neural network.
- In this meeting, JVET-AF0048 further analyzed the decoder design of GFV coding system, and utilized a flow translator or a parameter translator to improve the decoder interoperability and parameter translatability.
- This contribution updates the common text of GFV SEI message proposed in JVET-AE0280 by introducing the parameter translator proposed in JVET-AF0048 which attempts to solve the interoperability issue between different facial representations and their associated generative networks.

# The design overview

- Each picture unit (PU) may contain a GFV SEI message.



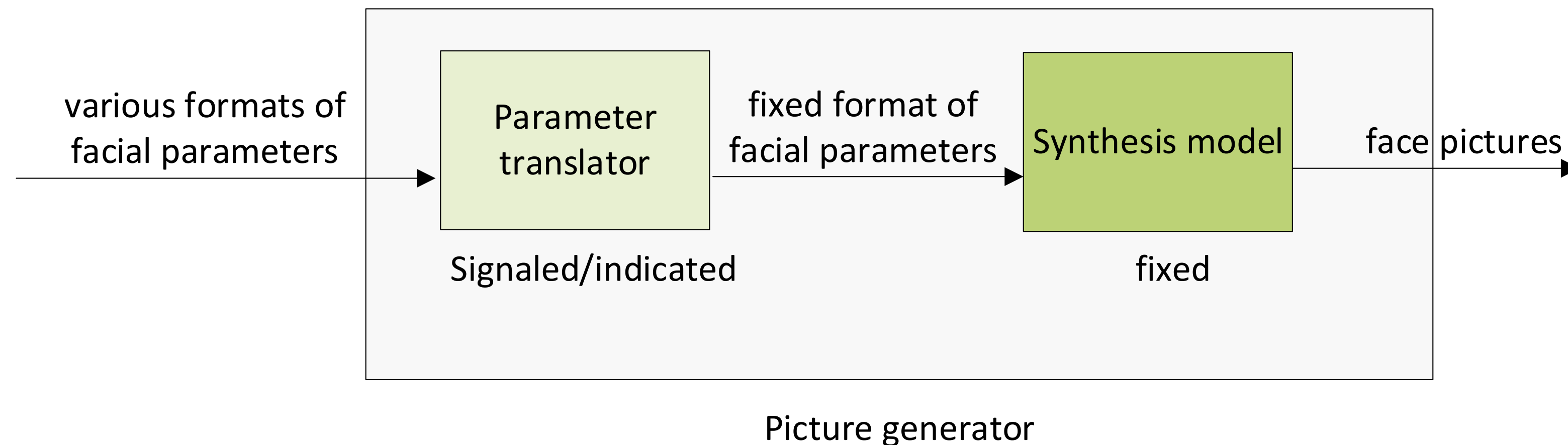
- ✓ a base picture: a picture that may be used by a network to generate a new face picture
- ✓ a primary coded picture: a picture can be fused by a network to improve background texture and facial details
- ✓ a dummy picture: a picture that is not used and the PU is only for transmitting the SEI message

\*Both the dummy picture and the primary coded picture should be marked as a non-output picture

- A GFV SEI message may contain:
  - The network (coded with ISO/IEC 15938-17 ) or the URI of the network
  - Various formats of facial parameters to support the current mainstream facial feature representation methods and the potential representation methods which might be used in the future

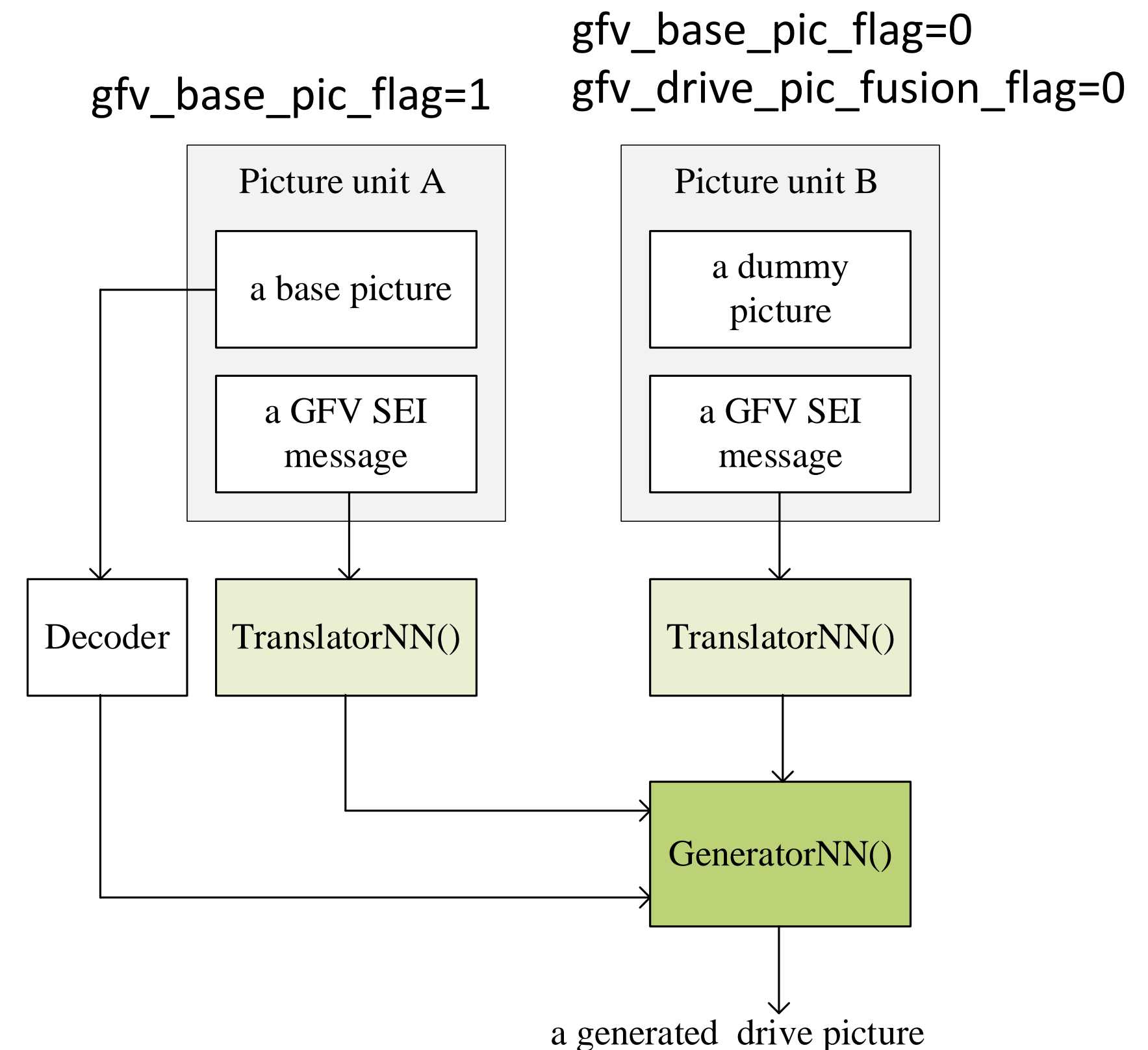
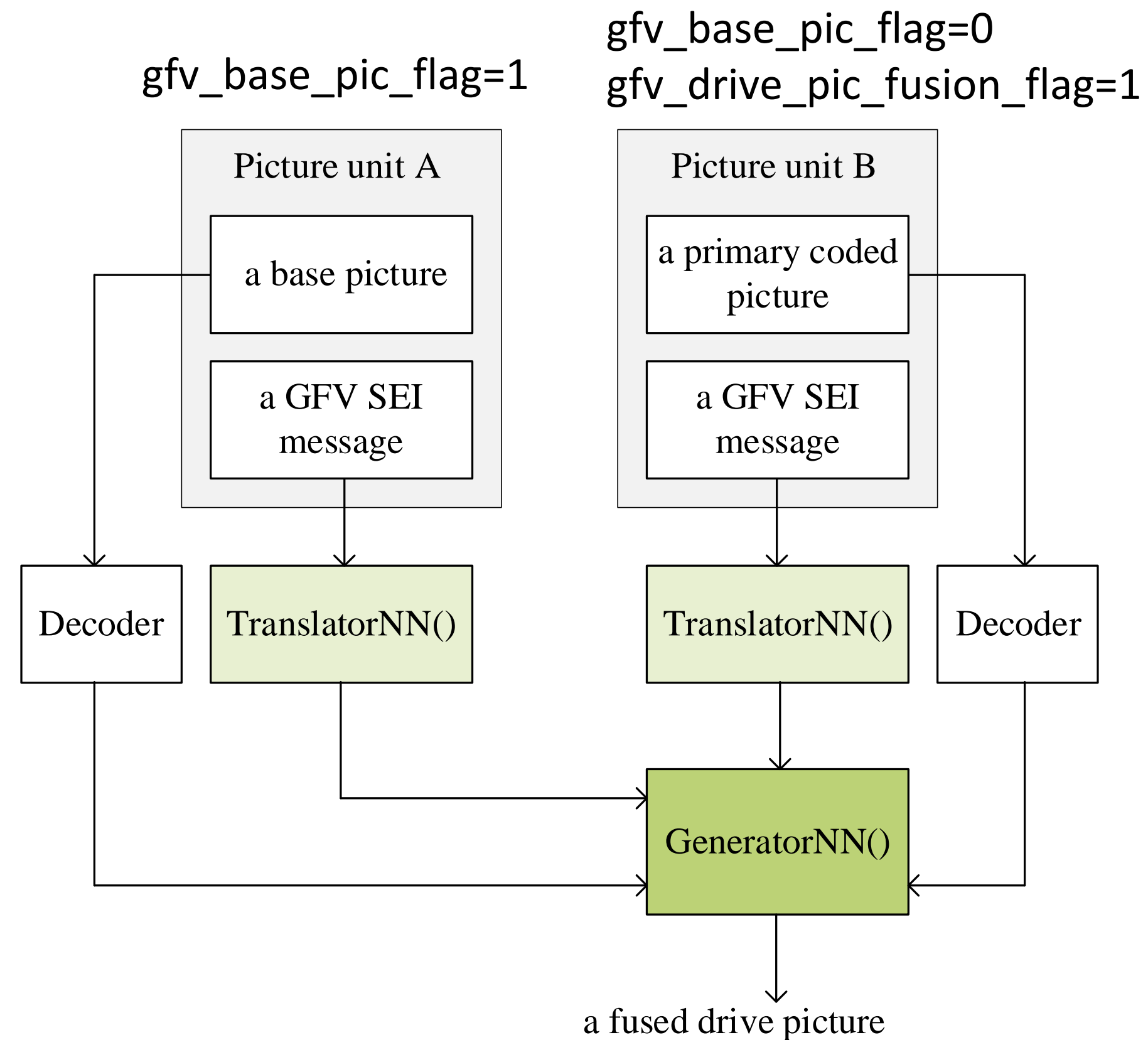
# Updates to JVET-AE0280

- Syntax element “gfv\_key” is removed to make the design cleaner.
- The generative network (picture generator) signaled/indicated by the SEI is replaced with a parameter translator.
  - ✓ The parameter translator is much lighter than picture generator (param: 40M+ vs. 1.2M, see JVET-AF0048 )
  - ✓ The various formats of the facial parameters signaled in SEI message are converted into a fixed format of parameters by the parameter translator and then input into the synthesis model to generate the face picture.
  - ✓ The synthesis model is fixed and needn't be signaled.



# The proposed text defines...

- Denote the parameter translator as **TranslatorNN()**
- Denote the synthesis model as **GeneratorNN()**
- Define **interfaces** of TranslatorNN() and GeneratorNN()
- Define the **processing order** as follows



# Summary

- The proposed generative face video SEI message
  - enables ultra-low bitrate compression at a fraction of VVC's bitrate.
- To address comments previously received, the common text provided in this contribution
  - further cleanups the syntax design
  - only signals or indicate a “translator” model which is much lighter than a picture generation model
  - fixes the synthesis model for picture generation to solve interoperability issues
- Software was already provided and crosschecked in the last meeting.
- Suggest to adopt the proposed SEI message to next draft of TuC for VSEI.

**Thanks**