Empowered by Innovation NEC

JCTVC-D280/m19045

Performance report on DPCM-based memory compression on TMuC 0.9

Hirofumi Aoki, Keiichi Chono, Kenta Senzaki, Junji Tajime and Yuzo Senda

NEC Corporation

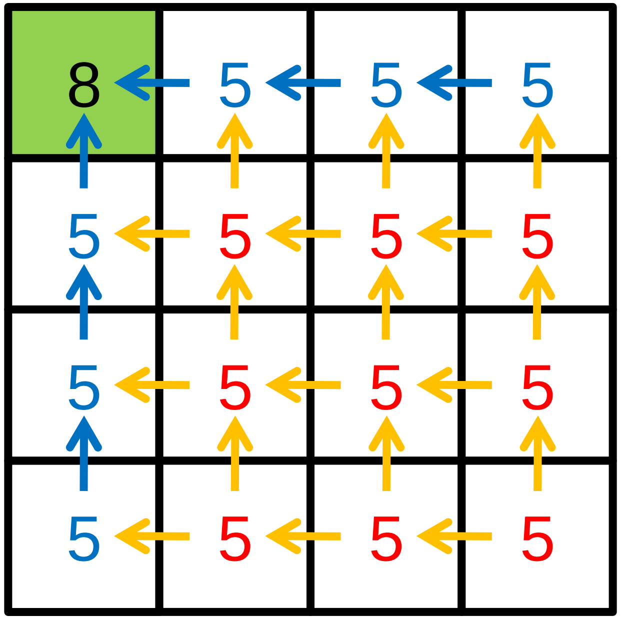
Reference frame memory compression

- A promising approach to reduce required memory both on size and access bandwidth
  - Decoded frames are compressed before stored in DPB, and decompressed after read from DPB
  - Memory required for MC is thus reduced
- NEC's approach:
  - *DPCM-based memory compression scheme*
    - JCTVC-B057: 1-D DPCM-based memory compression
    - JCTVC-C094: 2-D extension
    - JCTVC-C095: Adaptive quantization
  - Simple algorithm, decent compression efficiency, and fixed compression ratio guaranteed w/o rate control
- Experimental results on TMuC 0.9 are shown in this contribution

DPCM-based compression (JCTVC-C094)

Ex.) Compression of 8-bit 4x4 unit with 5-bit DPCM

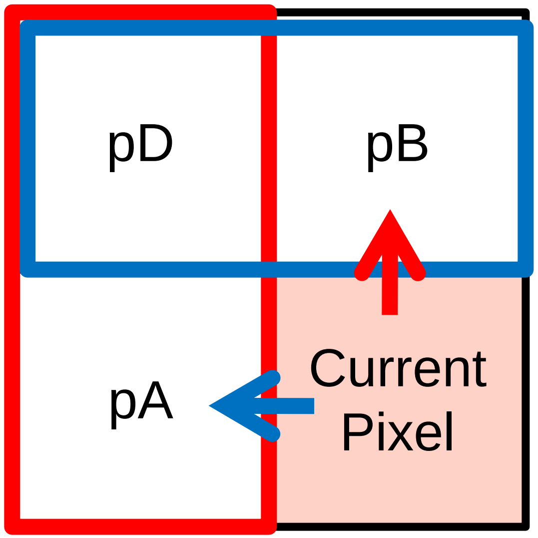
- The top-left-most pixel is coded with PCM
- Top-most pixels are coded with horizontal DPCM
- Left-most pixels are coded with vertical DPCM
- Other pixels are coded with Adaptive DPCM (ADPCM)
- The quantizer(s) for (A)DPCM-coded pixels: defined in SPS or PPS



Adaptive prediction for ADPCM

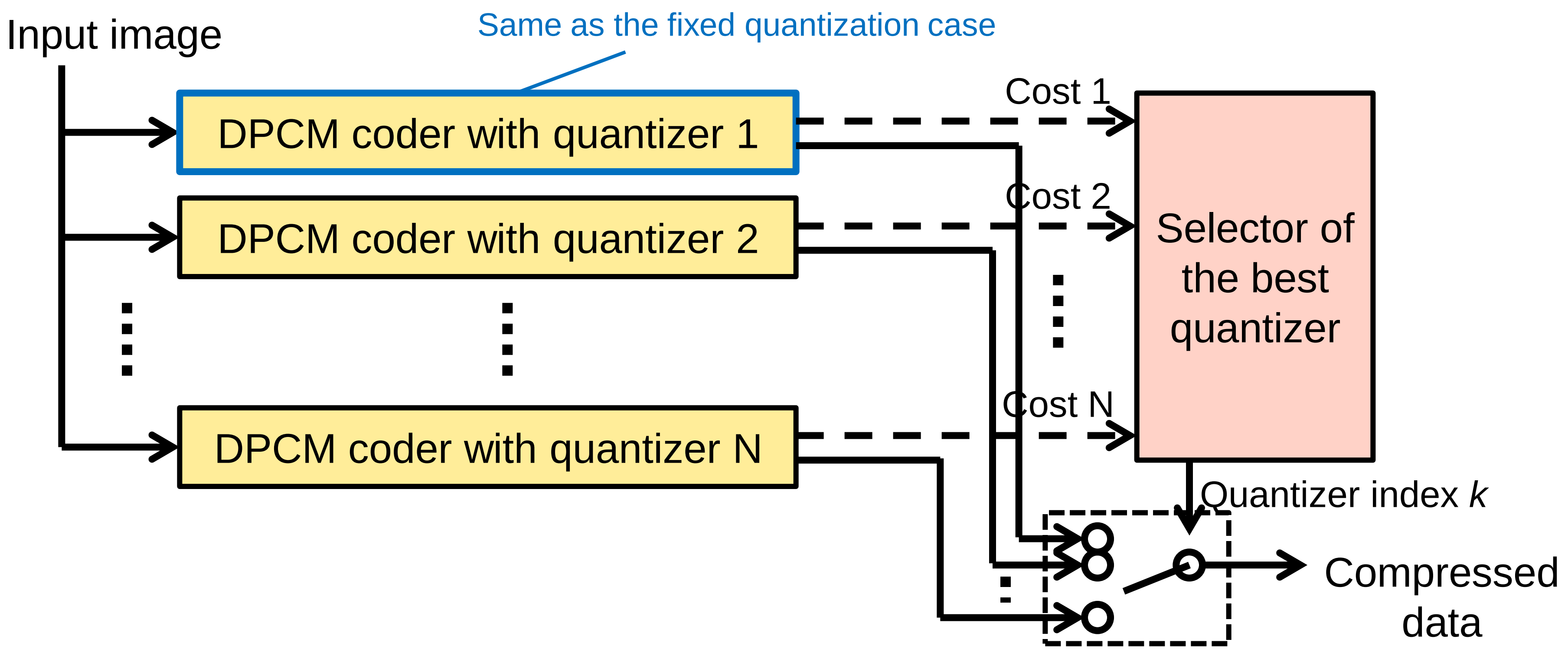
Based on comparison of pixel differences

- Calculate the vertical difference between pD and pA
- Calculate the horizontal difference between pD and pB
- If |pD-pA| < |pB-pC|, select pB as the prediction value
- Otherwise, select pA as the prediction value



Adaptive quantization (JCTVC-C095)

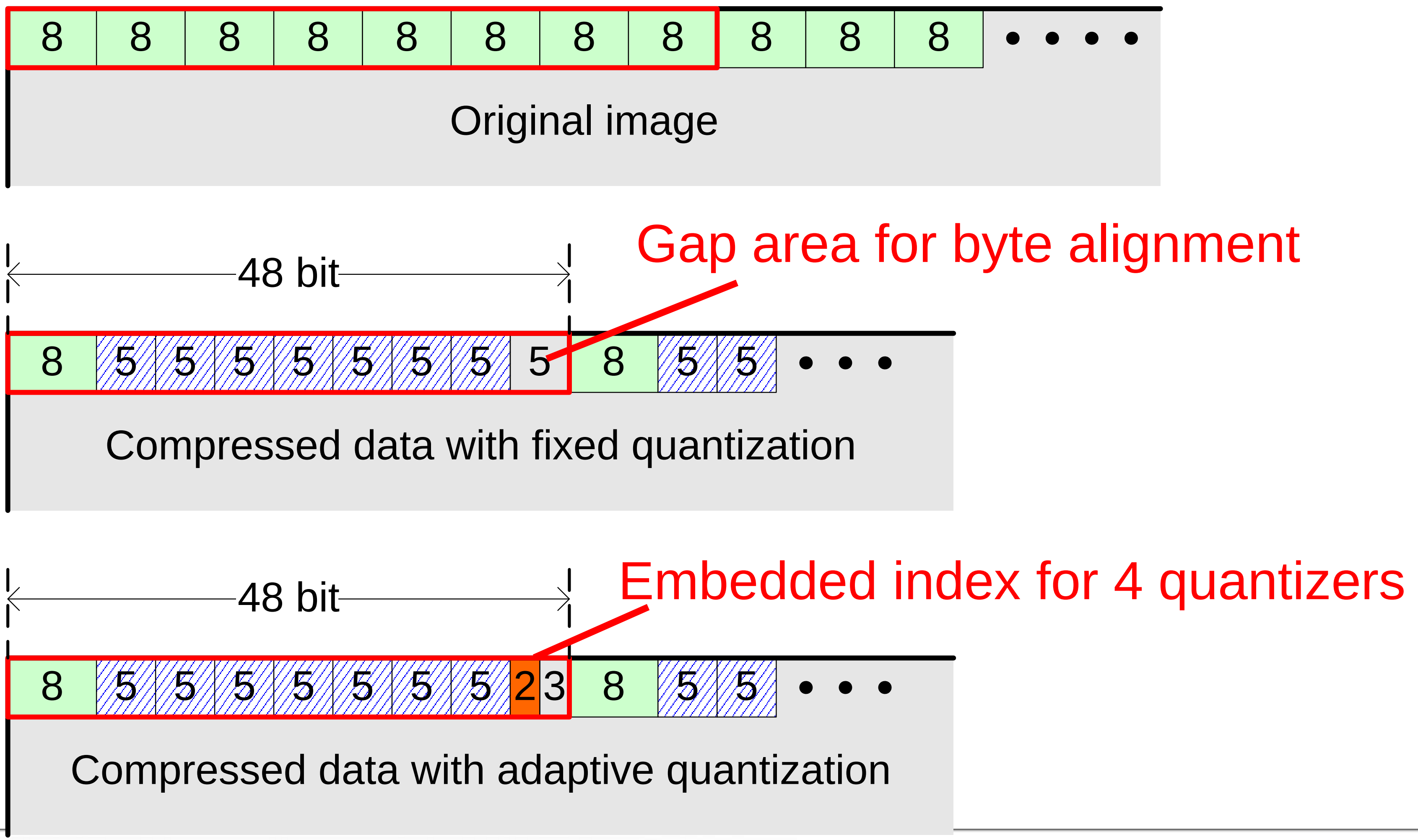
- Multiple quantizers are employed
  - Defined in PPS or SPS
- For each compression unit, one of quantizers that shows the best performance is chosen



- Quantizer index k is embedded into compressed data

Quantizer index embedding

- Generally memory size and memory access bandwidth does not increase



Decompression process

- Simply done by:
  - Getting a quantization representative value from a look-up table
  - Adding the value to a predicted value

$$r(x,y) = \begin{cases} c(x,y); & (x,y) = (0,0) \\ (\underbrace{p(x,y)}_{\text{Predicted value}} + \underbrace{Q_k(c(x,y))}_{\text{Quantization representative value}})\%2^B; & (x,y) \neq (0,0) \end{cases}$$

Predicted value   Quantization representative value

Experiments on TMuC 0.9

Test conditions

- Random Access, High Efficiency (with 12-bit IBDI)
- Low Delay, High Efficiency (with 12-bit IBDI)
- Random Access, Low Complexity (without IBDI)
- Low Delay, Low Complexity (without IBDI)

Evaluation measure

- Coding efficiency by BD metrics
- Complexity by decoding and encoding runtime
- Memory access bandwidth

Parameters of the proposed method:

- Memory Compression Unit size: 4x4
- # of quantizers for adaptive quantization: 1, 4
- IBDI on (12-bit): 7-bit DPCM -> (12+7*15)/8 ~= 7.5 bpp
- IBDI off (8-bit): 5-bit DPCM -> (8+5*15)/8 ~= 5.5 bpp

Coding efficiency

- BD-Rates (%) for HE configurations (12bpp -> 7.5bpp)
  - Better than those of fixed rounding

	Proposed method						Fixed Rounding		
	1 quantizers			*4 quantizers*					
	Y	U	V	Y	U	V	Y	U	V
RA	**0.6**	**0.5**	**0.6**	**0.3**	**0.1**	**0.2**	**0.8**	**1.7**	**1.5**
LD	**1.8**	**2.6**	**3.0**	**1.1**	**0.5**	**0.8**	**2.9**	**4.6**	**5.1**

- BD-Rates (%) for LD configurations (8bpp -> 5.5bpp)

	1 quantizers			*4 quantizers*		
	Y	U	V	Y	U	V
RA	**6.7**	**3.2**	**3.3**	**3.7**	**1.0**	**1.2**
LD	**9.7**	**4.0**	**4.6**	**5.6**	**1.6**	**1.7**

Complexity

- Encoding and decoding runtime (%)

Conf.		Enc. runtime (%)		Dec. runtime (%)	
		1 quantizers	*4 quantizers*	*1 quantizers*	*4 quantizers*
HE	RA	**100**	**100**	**130**	**202**
	LD	**100**	**100**	**128**	**193**
LC	RA	**101**	**101**	**161**	**321**
	LD	**102**	**101**	**158**	**308**

- Encoder complexity is negligible
- Decoder complexity is not low in the current implementation
  - Note: Cost calculations for multiple quantizers can be parallelized easily

Memory access bandwidth

Memory access bandwidth increase (%)

- "x/y" represents memory architecture setting where alignment size is x bits and burst-read size is y bits.

Conf.		w/o cache				w/ FIFO cache	
		32/64	32/128	64/256	64/512	64/256	64/512
HE	RA	**-29.3**	**-32.3**	**-38.9**	**-47.3**	**-40.1**	**-40.7**
	LD	**-25.7**	**-28.6**	**-38.2**	**-48.2**	**-40.9**	**-42.1**
LC	RA	**-22.4**	**-28.0**	**-38.1**	**-55.4**	**-35.1**	**-35.9**
	LD	**-18.8**	**-25.8**	**-36.9**	**-55.0**	**-39.2**	**-40.6**

Successfully reduced by up to approximately 50%

Conclusion

- DPCM-based memory compression scheme is proposed
- The effectiveness of the scheme on reduction of memory requirement is shown
  - Memory access bandwidth reduction: up to approximately 50%
- Coding loss is less than that of fixed rounding
  - Gain by IBDI is therefore not vanished

- Memory requirement reduction is critical in H/W implementation
  - Other related proposals: JCTVC-D023, D035 and D152

Proposal:

- Establish a CE on reference frame memory compression schemes, and further investigate this proposal in the CE

Empowered by Innovation

NEC