Sub-sampled Cross-component Prediction for Chroma Component Coding

Junru Li[1], Meng Wang[2], Li Zhang[3], Kai Zhang[3],

Shiqi Wang[2], Shanshe Wang[1], Siwei Ma[1] and Wen Gao[1]

[1]Institute of Digital Media, Peking University, Beijing, China.

[2]Department of Computer Science, City University of Hong Kong, Hong Kong, China.

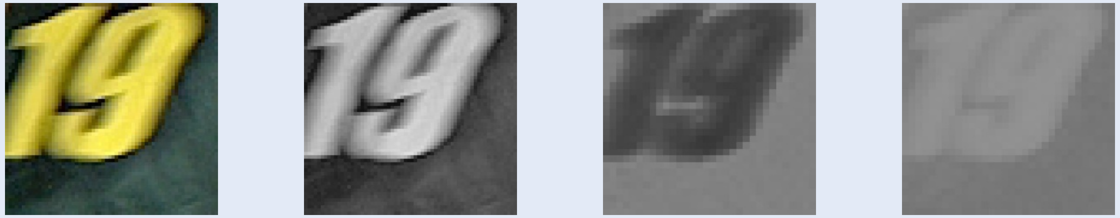
[3] Bytedance Inc., San Diego CA. 92122 USA.

Outline

- Background
- Sub-sampled Cross-component Prediction
- Experimental Results
- Conclusions

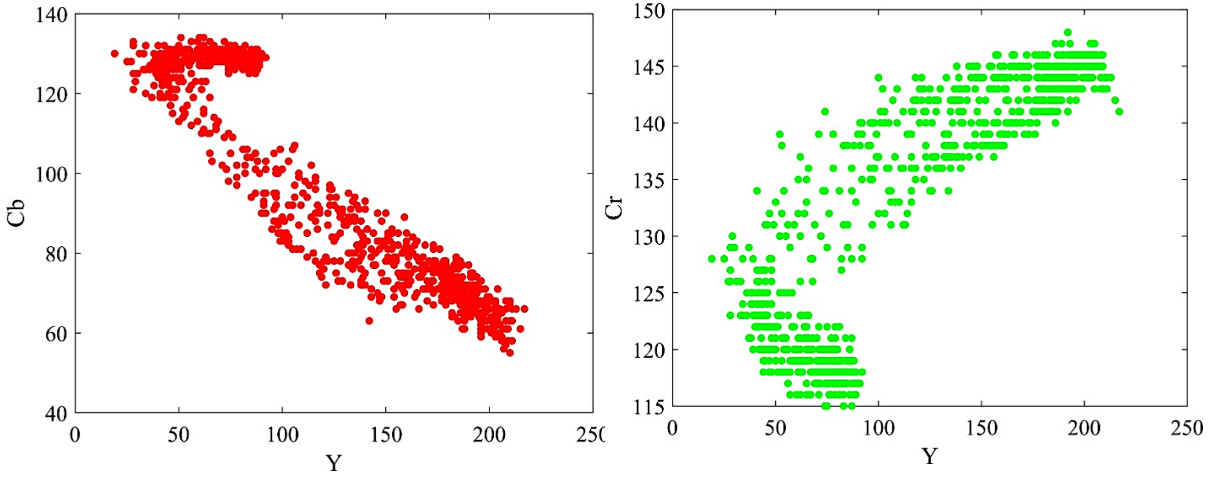
Cross-component Prediction[1][2]

- Inter-channel correlation
- Linear model (α, β)



Y-Cb-Cr Y Cb Cr

Illustration of the 3-channel block and single channel block



Relationships between Y-Cb and Y-Cr

[1] J. Kim, S.-W. Park, J.-Y. Park, and B.-M. Jeon, “Intra chroma prediction using inter channel correlation,” JCTVC-B021, Jul. 2010.
[2] J. Chen and V. Seregin, “Chroma intra prediction by scaled luma samples using integer operations,” JCTVC-C206, Oct. 2010.

Linear Model Derivation

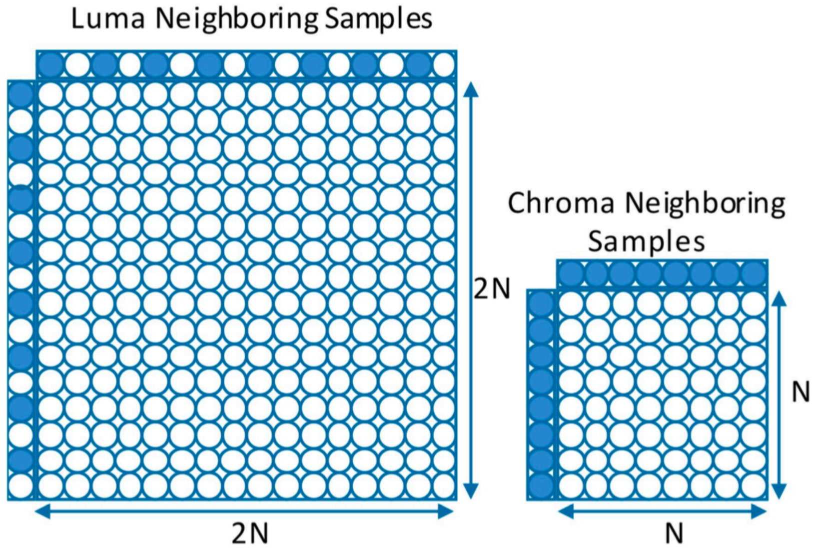
- Least Square Regression (LSR)[1][2].
- Luma down-sampling: $Rec_L^{(m)} \rightarrow \widehat{Rec}_L^{(m)}$
- Input: M neighboring sample pairs $\widehat{Rec}_L^{(m)}, Rec_C^{(m)}$
- Output: linear model parameters α, β

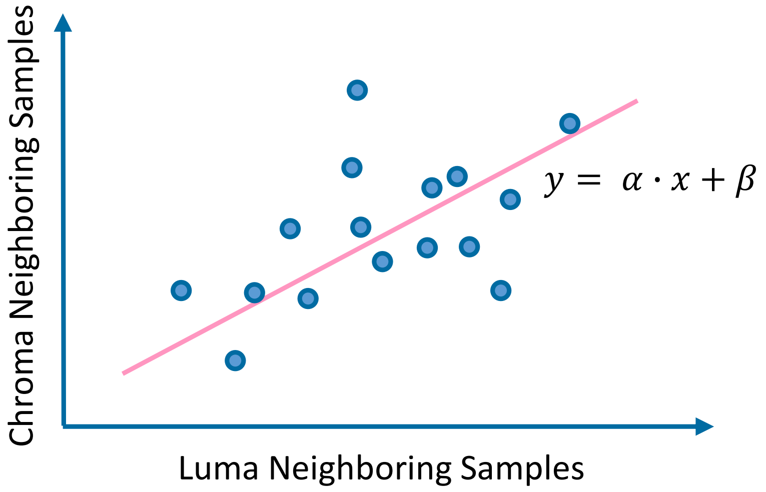
$$\alpha = \frac{M \cdot \sum_{m=1}^{M} Rec_C^{(m)} \widehat{Rec}_L^{(m)} - \sum_{m=1}^{M} Rec_C^{(m)} \sum_{m=1}^{M} \widehat{Rec}_L^{(m)}}{M \cdot \sum_{m=1}^{M} \left(\widehat{Rec}_L^{(m)}\right)^2 - \left(\sum_{m=1}^{M} \widehat{Rec}_L^{(m)}\right)^2}$$

$$\beta = \frac{\sum_{m=1}^{M} Rec_C^{(m)} - \alpha \sum_{m=1}^{M} \widehat{Rec}_L^{(m)}}{M}$$

$$Pred_C(i,j) = \alpha \cdot \widehat{Rec}_L(i,j) + \beta$$

High Complexity





Operations	Multiplication	Add	Shift	Comparison	Down Sampling
LSR	2M+4	7M+3	2	-	M

Linear Model Derivation

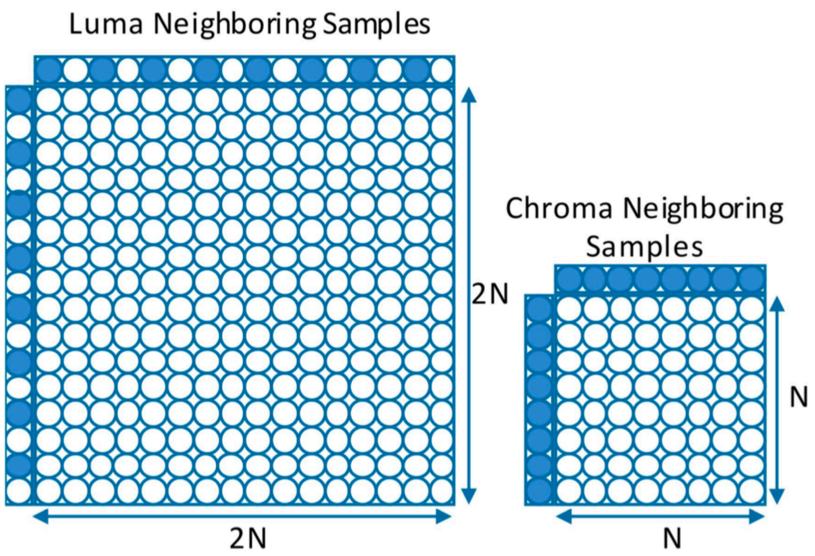
- Max-Min method[1] in VTM4.0.
- Luma down-sampling: $Rec_L^{(m)} \rightarrow \widehat{Rec}_L^{(m)}$
- Input: M neighboring sample pairs $\widehat{Rec}_L^{(m)}, Rec_C^{(m)}$
- Output: linear model parameters α, β

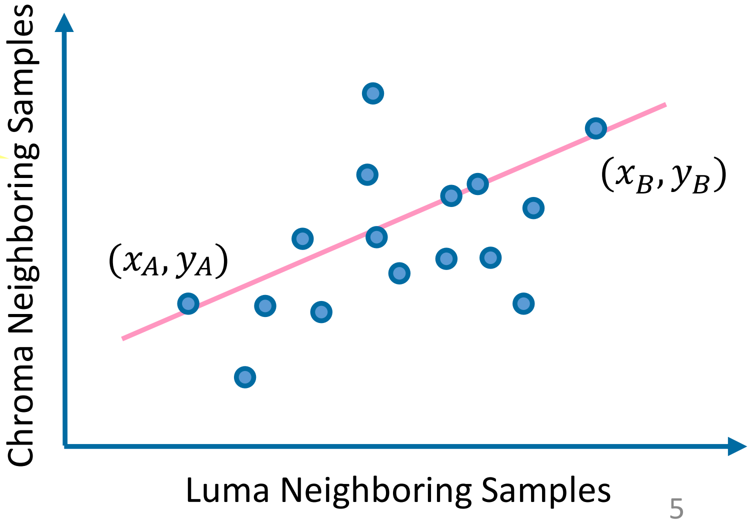
$$\alpha = \frac{y_A - y_B}{x_A - x_B}$$

$$\beta = y_A - \alpha \cdot x_A$$

$$Pred_C(i,j) = \alpha \cdot \widehat{Rec}_L(i,j) + \beta$$

Still with high complexity

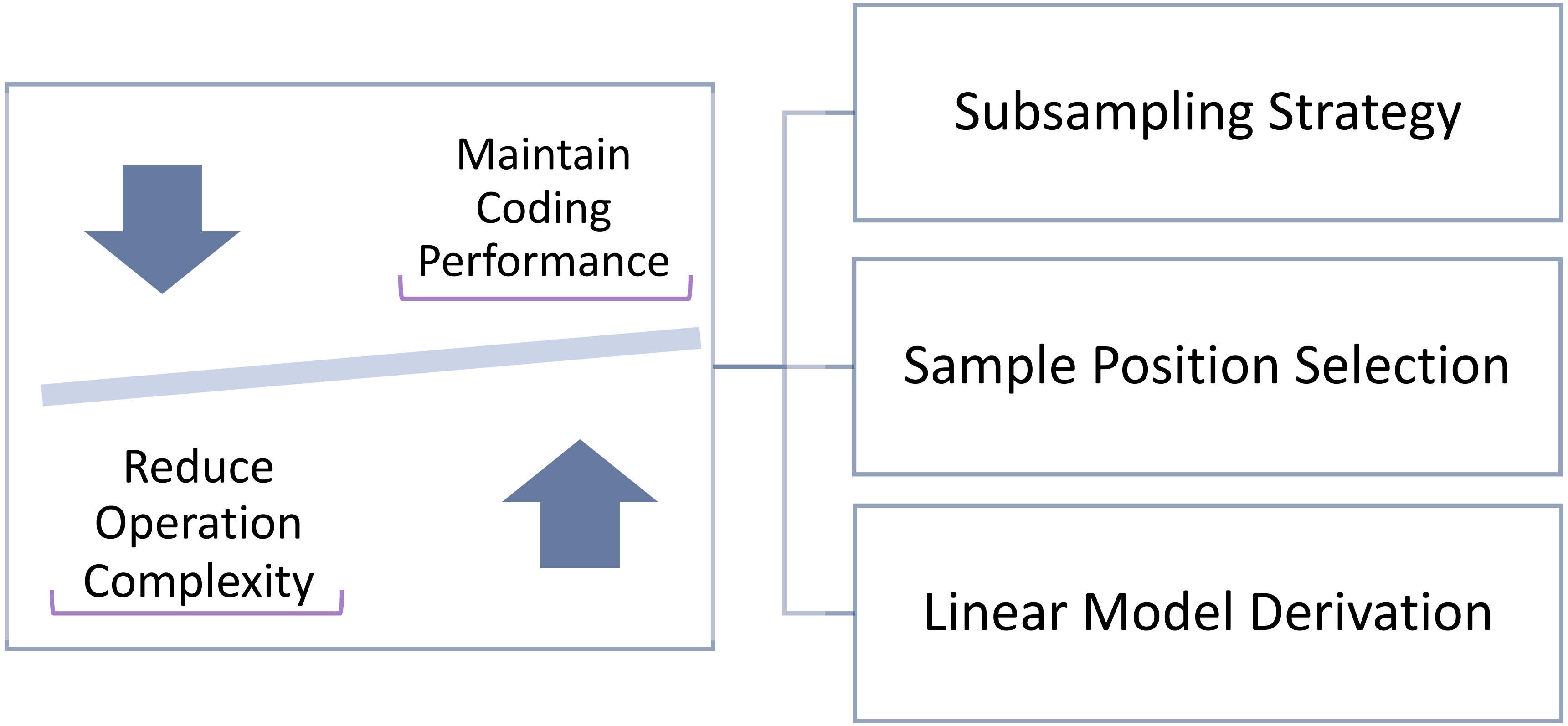




Operations	Multiplication	Add	Shift	Comparison	Down Sampling
LSR	2M+4	7M+3	2	-	M
Max-Min	1	3	1	2M	M

[1] G. Larche, J. Taquet, C. Gisquet, and P. Onno, “CE3-5.1: On cross-component linear model simplification,” JVET-L0191, Oct. 2018.

Sub-sampled Cross-component Prediction



Sub-sampled Cross-component Prediction

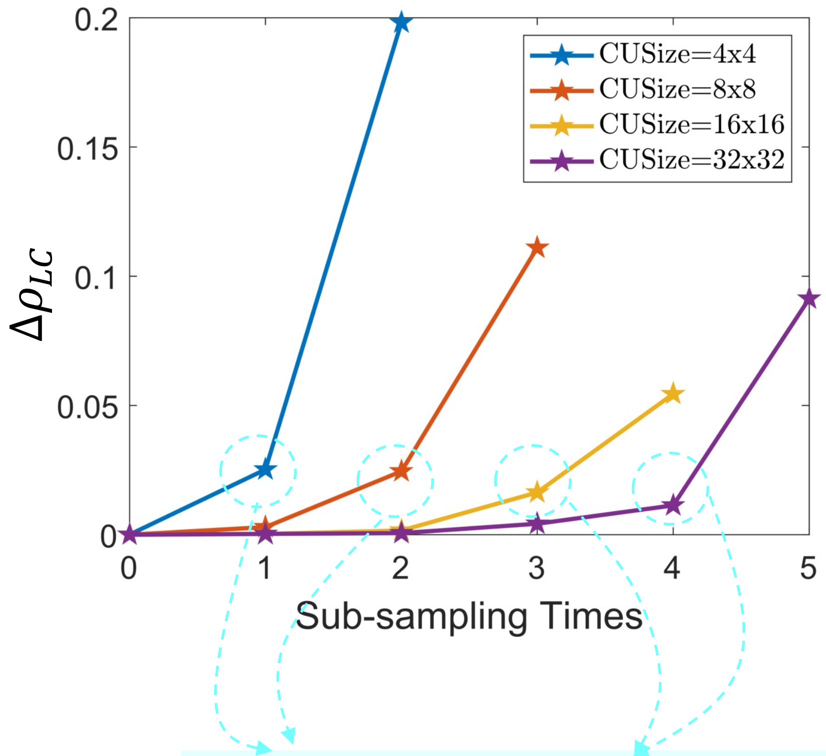
- Sub-sampling strategy: only 4 luma-chroma pairs

The correlation between luma sample x and chroma sample y with sub-sample time of s,

$$\rho_{LC}^{s} = \frac{cov(L,C)}{\sigma_L \sigma_C}$$

$$\Delta\rho_{LC} = \rho_{LC}^{s} - \rho_{LC}^{0}$$

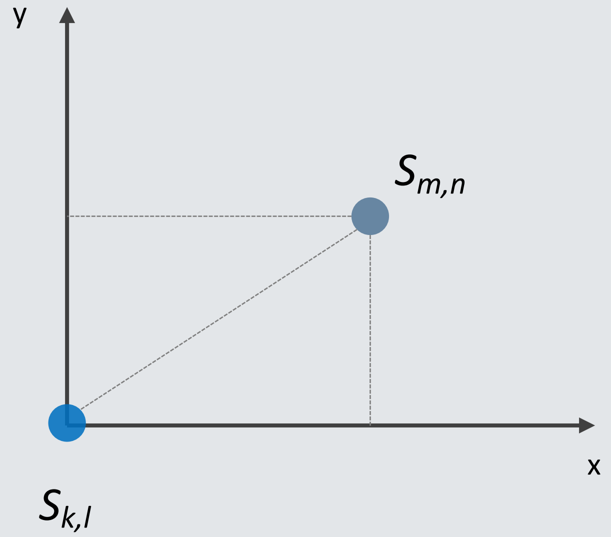
Sub-Sample Times	0	1	2	3	4	5
4x4	8	4	2	-	-	-
8x8	16	8	4	2	-	-
16X16	32	16	8	4	2	-
32X32	64	32	*16*	8	4	2



Four sample pairs are sufficient to maintain the existing correlation

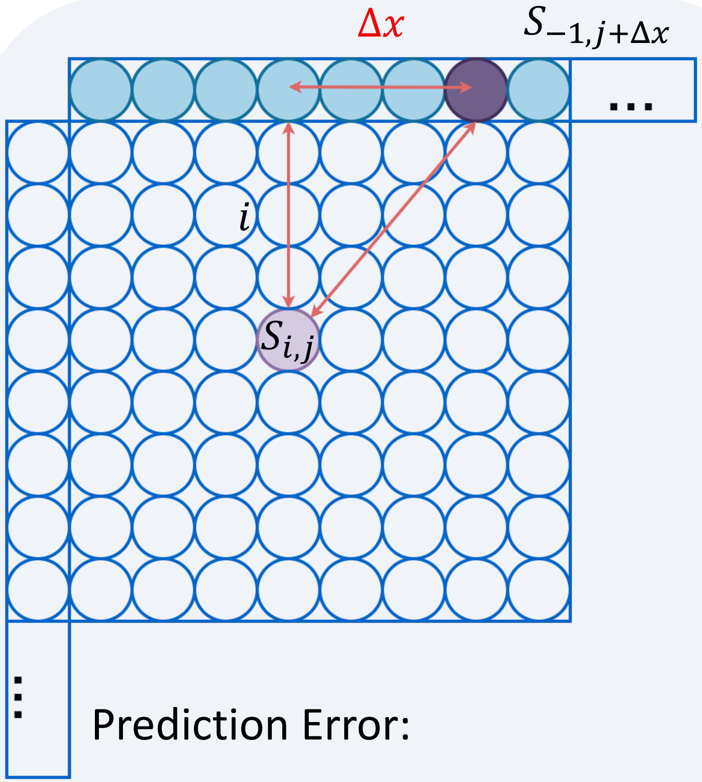
Sub-sampled Cross-component Prediction

- Sample position selection



Correlation of two pixels:

$$\rho\left(S_{k,l}, S_{m,n}\right) = \rho_y^{|k-m|} \cdot \rho_x^{|l-n|}$$



Prediction Error:

$$e_{i,j} = S_{i,j} - S_{-1,j+\Delta x}$$

The variance of the prediction error.

$$\rho\left(e_{i,j}, e_{p,q}\right) = \frac{Cov\left(e_{i,j}, e_{p,q}\right)}{\sigma_{e_{i,j}}\sigma_{e_{p,q}}}$$

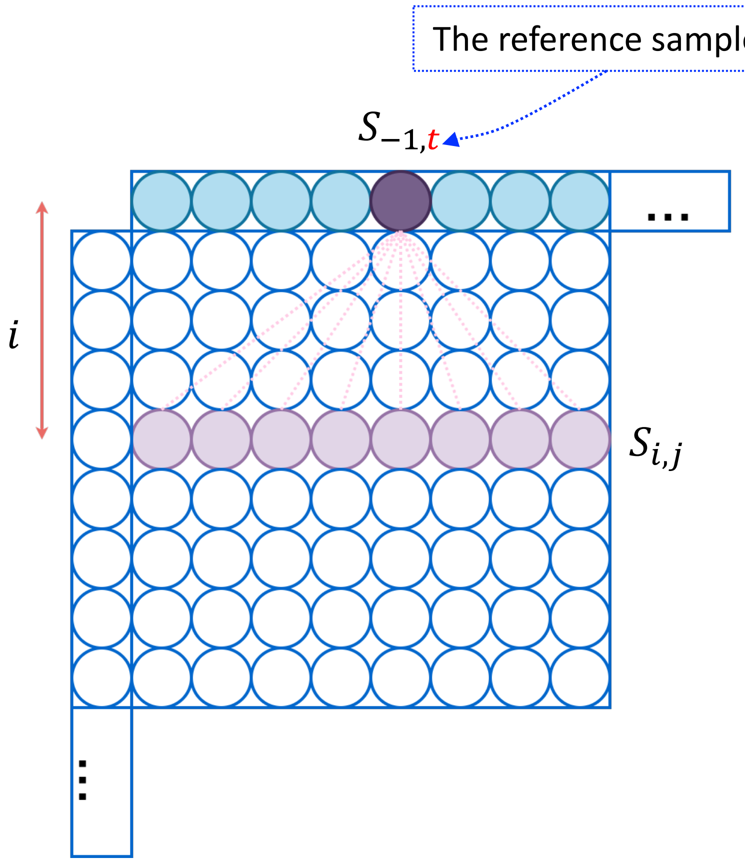
$$= \rho_x^{|j-q|}\rho_y^{|i-p|} - \rho_x^{|j+\Delta x-q|}\rho_y^{p} - \rho_x^{|j-q-\Delta x|}\rho_y^{i} + \rho_x^{|j-q|}$$

$i = p, j = q,$

$$\sigma_{i,j}^2 = 2 - 2\rho_y^{i}\rho_x^{|\Delta x|}.$$

Sub-sampled Cross-component Prediction

- Sample position selection



The variance of the prediction error within a row:

$$\sum_{j=0}^{N-1} \sigma_{i,j}^2 = 2N - 2\rho_y^i \sum_{j=0}^{N-1} \rho_x^{|j-t|} \Rightarrow f(t) = \sum_{j=0}^{N-1} \rho_x^{|j-t|} = \frac{2 - \rho_x^{t+1} - \rho_x^{N-t}}{1 - \rho_x}$$

$$\text{Minimize } \sum_{j=0}^{N-1} \sigma_{i,j}^2 \Rightarrow \text{Maximize } f(t)$$

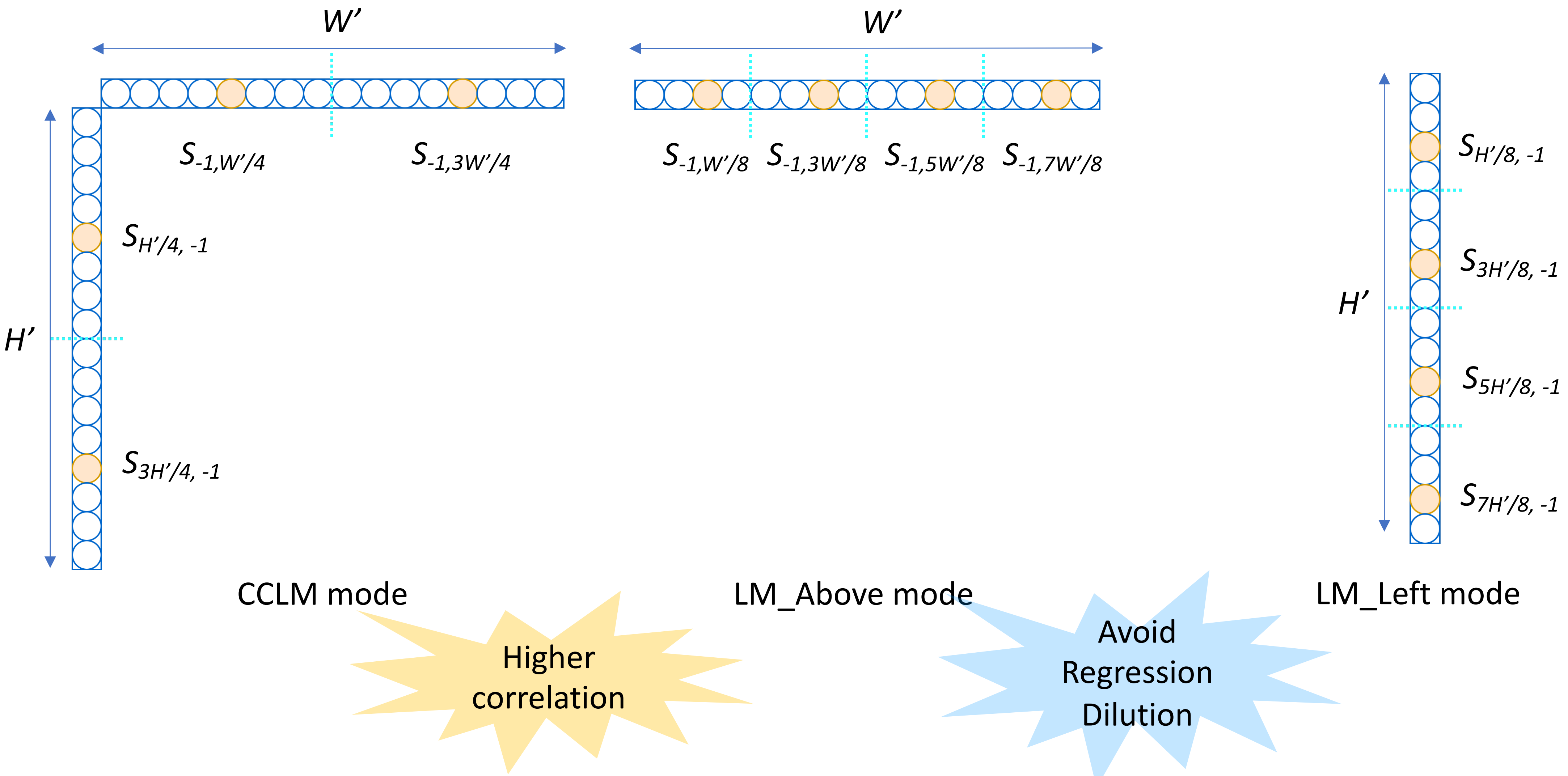
$$f'(t) = \frac{-(\rho_x^{t+1} - \rho_x^{N-t}) \cdot \ln \rho_x}{1 - \rho_x} = 0 \Rightarrow t_0 = \frac{N-1}{2}$$

$$t < t_0, f'(t) > 0$$
$$t > t_0, f'(t) < 0$$
$$t = t_0, f'(t) = 0$$

[Max]

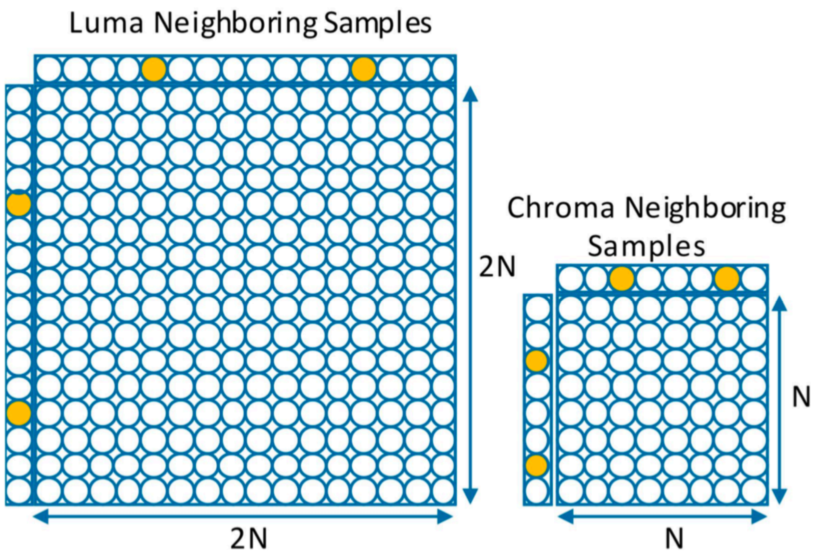
$$\text{Minimize } \sum_{j=0}^{N-1} \sigma_{i,j}^2 = 2N - 2\rho_y^i \sum_{j=0}^{N-1} \rho_x^{|j-t_0|}$$

Sub-sampled Cross-component Prediction

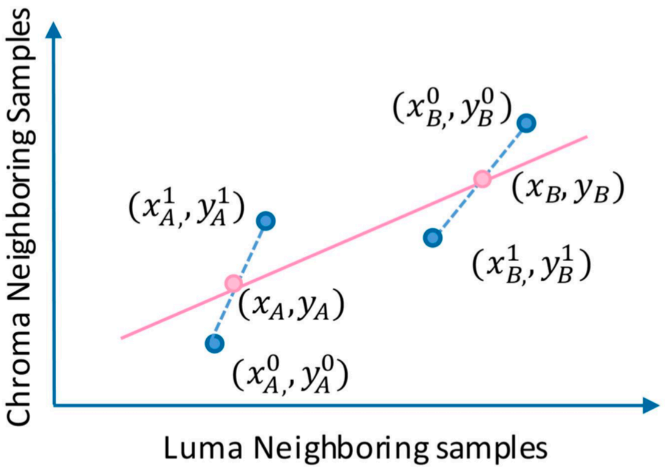


Sub-sampled Cross-component Prediction

- Input: at most 4 luma-chroma sample pairs
- Output: linear model parameters: α, β



Locations of the samples



Derivation of the linear model

$$x_A = (x_A^0 + x_A^1 + 1) \gg 1$$

$$x_B = (x_B^0 + x_B^1 + 1) \gg 1$$

$$y_A = (y_A^0 + y_A^1 + 1) \gg 1$$

$$y_B = (y_B^0 + y_B^1 + 1) \gg 1$$

$$\alpha = \frac{y_A - y_B}{x_A - x_B} \qquad \beta = y_A - \alpha \cdot x_A$$

$$Pred_C(i,j) = \alpha \cdot \widehat{Rec}_L(i,j) + \beta$$

Sub-sampled Cross-component Prediction

- Sensitivity of the model parameters.

Standard deviation of Luma

Standard deviation of chroma

Correlation of luma and chroma

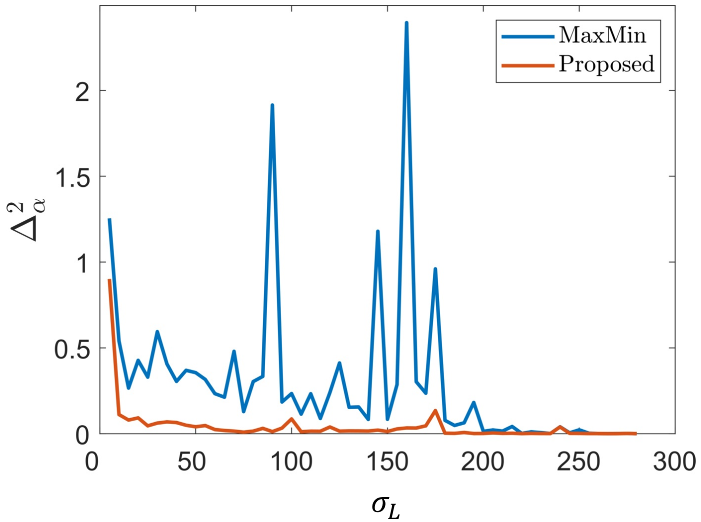
Mean square prediction error:

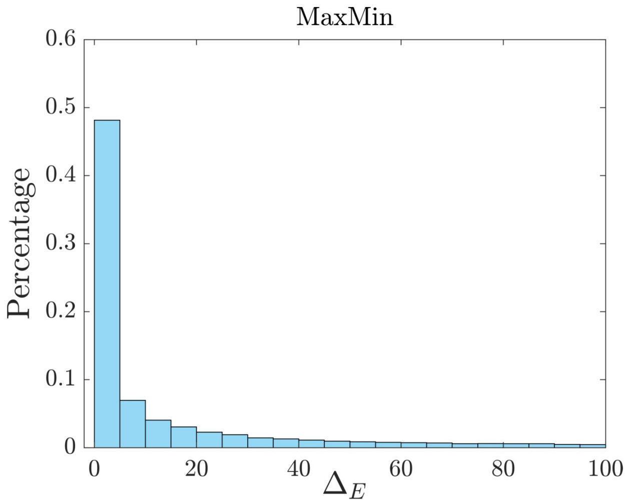
$$E_{\text{err}} = E\left[\left(C_{org} - \alpha_{LSR} \cdot L - \beta_{LSR}\right)^2\right] = E\left[\left(y_{org} - \frac{\sigma_C}{\sigma_L} \cdot \rho \cdot L - \left(\mu_C - \mu_L \cdot \frac{\sigma_C}{\sigma_L} \cdot \rho\right)\right)^2\right] = \sigma_C^2(1-\rho^2).$$

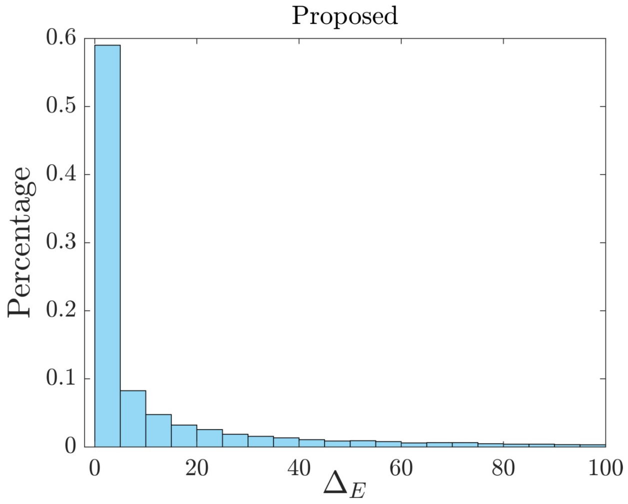
When the estimated α‘ and β’ differ from the α_{LSR} and β_{LSR}, $\alpha' = \alpha_{LSR} + \Delta_\alpha$ $\beta' = \beta_{LSR} + \Delta_\beta$

$$E_{\text{err}}' = E\left[\left(C_{org} - \alpha' \cdot L - \beta'\right)^2\right] = \sigma_C^2(1-\rho^2) + \left(\mu_L\Delta_\alpha + \Delta_\beta\right)^2 + \sigma_L^2\Delta_\alpha^2$$

By substituting Δ_β with $\mu_x\Delta_\alpha$, the prediction error will be increased by $\Delta E = E_{err}' - E_{err} = \sigma_L^2\Delta_\alpha^2$

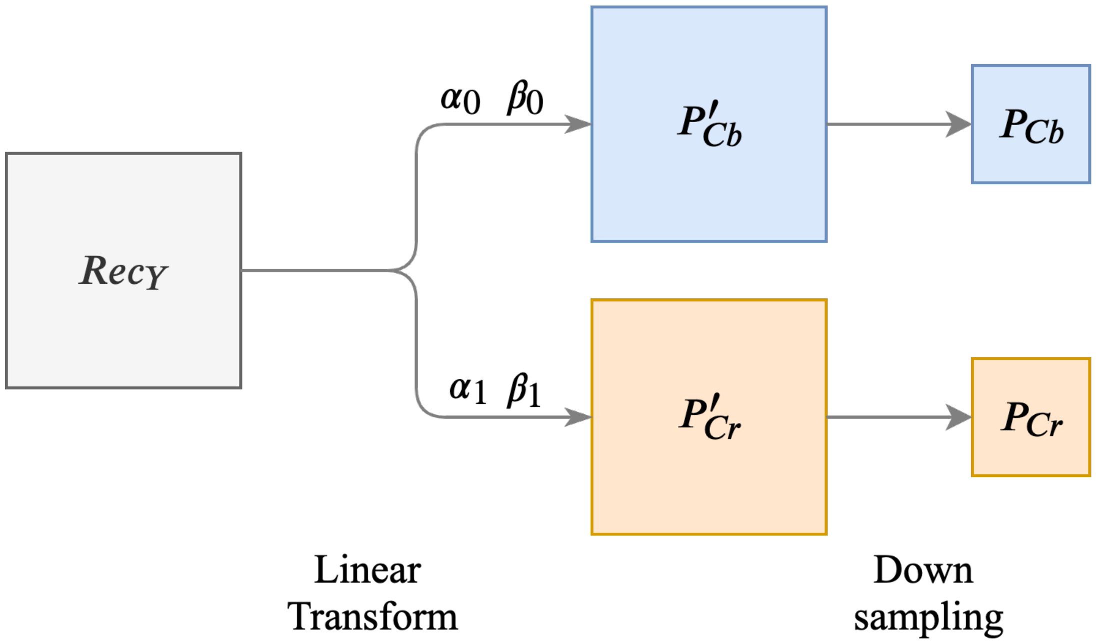






Implementations to AVS3 Standard

- Two-step cross-component prediction mode (TSCPM)
- Subsampled cross-component prediction



Mode Index	Chroma Intra Mode	Bin String
0	DM	1
5	TSCPM	01
1	DC	001
2	Horizontal	0001
3	Vertical	00001
4	Bi-linear	00000

Experimental Results

- Evaluations on VVC test model VTM-4.0[1]
    - Configurations: AI and RA
    - JVET Common Test Conditions[2]
    - QP = {22, 27, 32, 37}

- Evaluations on AVS3 test model HPM-5.0[3]
    - Configurations: AI and RA
    - AVS3 Common Test Conditions[4]
    - QP = {27, 32, 37, 42}

[1] “VVCsoftwareVTM-4.0,”https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_ VTM/tags/VTM-4.0.
[2] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Suhring, “JVET common test conditions and software reference configurations for SDR video,” Joint Video Exploration Team (JVET), doc. JVET-M1010, 2019.
[3] “AVS3 software repository,” /Public/codec/video_codec/HPM/HPM-5.0.
[4] J. Chen and K. Fan, “AVS3-P2 common test conditions v8.0,” AVS-N2727, Sep. 2019.

Experimental Results

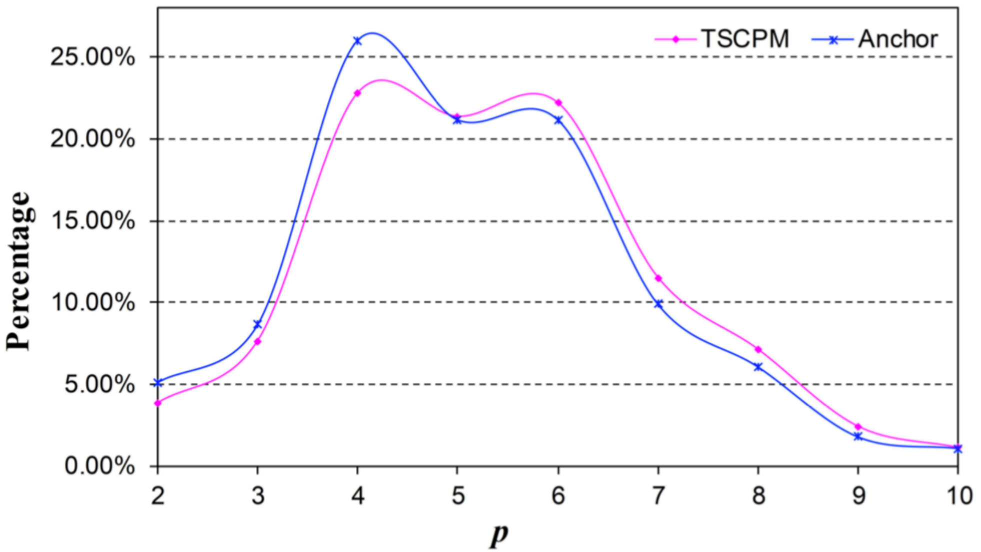
- Evaluations of the sub-sampled cross-component prediction on VTM-4.0
  - BD-Rate (AI)
    - Y/U/V: -0.05%/-0.30%/-0.38%
  - BD-Rate (RA)
    - Y/U/V: -0.05%/-0.36%/-0.63%
  - The largest gain comes from the 4K sequences in class A1.
    - *Tango2*: Y/U/V -0.16%/-0.12%/-1.05%
    - *Campfire*: Y/U/V -0.32%/-0.27%/-0.46%

Table 3: Experimental Results on VTM-4.0 under AI and RA configurations

Class	AI			RA		
	Y	U	V	Y	U	V
A1	-0.17%	-0.16%	-0.46%	-0.16%	-0.06%	-0.40%
A2	-0.03%	-0.04%	-0.08%	-0.04%	0.08%	-0.06%
B	-0.02%	-0.59%	-0.70%	0.00%	-0.82%	-1.13%
C	-0.02%	-0.47%	-0.33%	-0.03%	-0.32%	-0.59%
E	-0.02%	-0.01%	-0.15%			
Overall	**-0.05%**	**-0.30%**	**-0.38%**	**-0.05%**	**-0.36%**	**-0.63%**
D	-0.01%	-0.21%	-0.34%	-0.02%	-0.18%	-0.32%
F	0.01%	-0.04%	-0.21%	0.00%	-0.09%	-0.21%
Enc Time	99%			100%		
Dec Time	100%			100%		

Experimental Results

- Evaluations of TSCPM on HPM-5.0
  - BD-Rate (AI)
    - Y/U/V: -0.99%/-9.18%/-9.41%
  - BD-Rate (RA)
    - Y/U/V: -0.50%/-7.63%/-7.80%



Statistical analysis of chroma CB size with and without TSCPM, $p = \log_2(W \times H)$

Table 4: Performance of proposed TSCPM on AVS3 test model HPM-5.0

	Seq	AI			RA		
		Y	U	V	Y	U	V
720p	City	0.01%	-0.43%	-0.28%	0.09%	-1.20%	-0.13%
	Crew	-0.15%	-5.26%	-4.87%	0.11%	-2.15%	-4.80%
	Vidyo1	-0.05%	-4.08%	-2.41%	0.11%	-3.78%	-2.63%
	Vidyo3	0.03%	-2.66%	-7.72%	-0.24%	-6.27%	-10.26%
1080p	BasketballDrive	-0.34%	-6.00%	-6.29%	-0.27%	-5.68%	-5.16%
	Cactus	-0.40%	-7.08%	-6.47%	-0.27%	-9.23%	-6.26%
	MarketPlace	-0.54%	-11.03%	-11.11%	-0.03%	-12.55%	-9.59%
	RitualDance	-0.30%	-7.90%	-13.55%	-0.18%	-4.41%	-10.17%
4k	Tango2	-1.56%	-22.48%	-24.06%	-0.67%	-20.75%	-20.65%
	Campfire	-7.74%	-32.29%	-32.34%	-4.43%	-16.80%	-21.59%
	ParkRunning3	-0.61%	-1.69%	-1.06%	-0.22%	-0.87%	-0.38%
	DaylightRoad2	-0.18%	-9.21%	-2.74%	-0.04%	-7.85%	-1.95%
	720p	-0.04%	-3.11%	-3.82%	0.02%	-3.35%	-4.46%
	1080p	-0.40%	-8.00%	-9.36%	-0.19%	-7.97%	-7.80%
	4k	-2.52%	-16.42%	-15.05%	-1.34%	-11.57%	-11.14%
	Average	**-0.99%**	**-9.18%**	**-9.41%**	**-0.50%**	**-7.63%**	**-7.80%**
	Enc Time	100%			99%		
	Dec Time	100%			99%		

Complexity Analyses

Operations	Multiplication	Add	Shift	Comparison	Down Sampling
LSR	2M+4	7M+3	2	-	M
Max-Min	1	3	1	2M	M
The proposed method	1	7	5	4	4

Operations on a 32x32 chroma CB	Comparison	Down Sampling
VTM-4.0	128	64
The proposed method	4	4
1-(the proposed method/VTM-4.0)	94%	97%

Conclusion

- A novel sub-sampled cross-component prediction
  - Adopted to the VVC and the AVS3 standard

- Design with theoretical analyses
  - Pixel correlation
  - Inter-channel correlation

- Dramatically decrease the operation complexity
  - Save 94% comparisons, 97% down-sampling operations

- BD-Rate gains for the VVC
  - Y/U/V: -0.05%/-0.30%/-0.38% on average

- TSCPM achieves significant BD-Rate gains for the AVS3
  - Y/U/V: -0.99%/-9.18%/-9.41% on average
  - Y/U/V: -2.52%/-16.42%/-15.05% on 4K sequences

Thank You ！