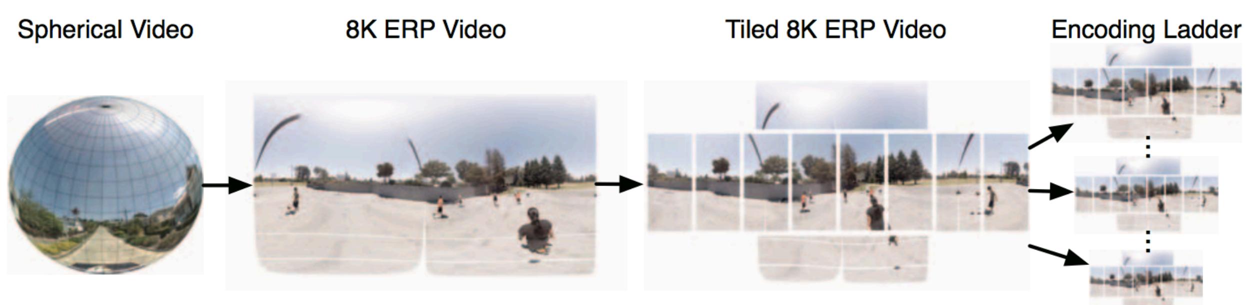
# Estimation of Optimal Encoding Ladder for Tiled 360° VR Video in Adaptive Streaming Systems

Cagri Ozcinar, Ana De Abreu, Sebastian Knorr, and Aljosa Smolic
Trinity College Dublin (TCD), Dublin 2, Ireland

IEEE International Symposium on Multimedia (ISM'17)

# Introduction

- 360◦ video streaming is significantly challenging owing to its resource-intensive encoding and storage requirements
- Adaptive streaming: each 360◦ video is divided into a set of tiles that includes different bitrate levels of the tiled video
- Representation sets for the video content forms the encoding ladder



# Cost-Optimal Encoding Ladders

- Tiles affects coding efficiency
- Tiles have a different level of contribution for the overall 360° video viewing quality

→ new encoding ladder configurations are required

⇒ focus on the configuration of cost-optimal encoding ladders in adaptive streaming systems

  - considering both the provider's and client's perspective
  - develop an encoding ladder estimation method for tiled 360◦ video streaming

# System Models: Distortion Modeling

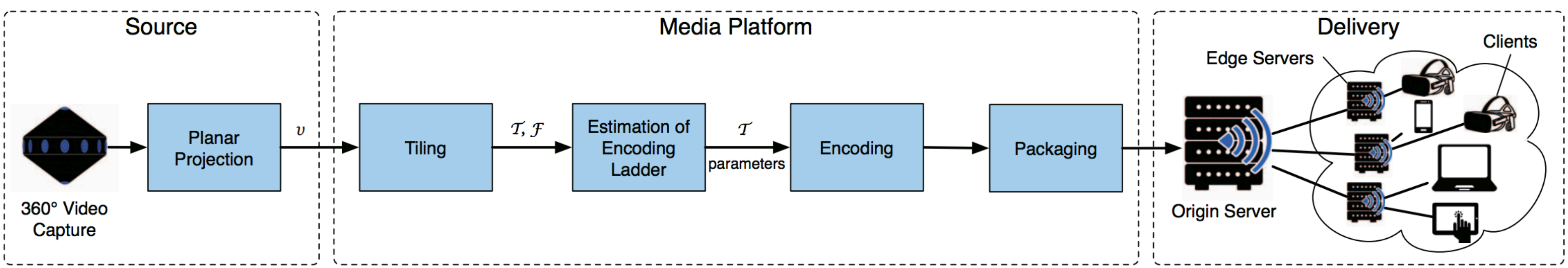
- Distortion function

$$FT_{ogB} = k_{og} Z_B^{\Omega_{og}} + \Phi_{og}$$

- Noise power for the i-th representation of the j-th tile

$$d_{ij} = \frac{\sum_{x \in W} \sum_{y \in H} \left( (t_j(x,y) - \tilde{t}_{ij}(x,y))^2 q_j(x,y) \right)}{\sum_{x \in W} \sum_{y \in H} q_j(x,y)}$$

$$q_j(x,y) = cos \frac{(y + 0.5 - H/2)\pi}{H}$$



# System Models: Cost Modeling

- Encoding cost

$$c_{ij}^{e} = \begin{cases} \mu_e, & r_{ij} \leq 720p \\ 2\mu_e, & 720p < r_{ij} \leq 1080p \\ 4\mu_e, & 1080p < r_{ij} \leq 4K \\ 8\mu_e, & 4K < r_{ij} \leq 8K \end{cases}$$

- Storage cost

$$c_{ij}^{s} = \mu_s b s_{ij}$$

| **Resolution** $\mathcal{G}$ | | $g_1$ | | | | | | $g_2$ | | | | | | $g_3$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | | Distortion | | | Data size | | | Distortion | | | Data size | | | Distortion | | | Data size | | |
| | | $k$ | $\Omega$ | $\Phi$ | $k$ | $\Omega$ | $\Phi$ | $k$ | $\Omega$ | $\Phi$ | $k$ | $\Omega$ | $\Phi$ | $k$ | $\Omega$ | $\Phi$ | $k$ | $\Omega$ | $\Phi$ |
| **Content type** $\mathcal{O}$ | $o_1$ | 1809 | -0.6959 | 5.649 | 0.7613 | 0.9901 | 52.54 | 4002 | -0.7558 | 2.723 | 0.8005 | 0.9859 | 52.25 | 1829 | -0.5587 | -3.266 | 0.8264 | 0.9846 | 214.9 |
| | $o_2$ | 220.1 | -0.3583 | 6.447 | 0.6467 | 1.003 | 29.36 | 191.9 | -0.2763 | -5.728 | 0.6078 | 1.009 | 71.15 | 480.6 | -0.3643 | -5.728 | 0.5654 | 1.015 | 269 |
| | $o_3$ | 820.4 | -0.4702 | 6.2 | 0.6631 | 1.001 | 10.69 | 643 | -0.3825 | -2.625 | 0.6691 | 1 | 17.46 | 616.9 | -0.2837 | -23.78 | 0.5943 | 1.012 | 203.8 |

# Problem Formulation

| Profiles: | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $B^{min}$ (*Mbps*) | 1 | 3 | 15 | 25 |
| $B^{max}$ (*Mbps*) | 4 | 20 | 30 | 40 |
| $\Lambda$ | 0.25 | 0.25 | 0.25 | 0.25 |

- Constraints
  - Bandwidth: $B_{min}$ and $B_{max}$
  - Computational and storage costs: $C_{max}$ and $S_{max}$
  - Encoding rate: minimum step size τ

$$\mathcal{L}^* : \underset{\mathcal{L}}{\operatorname{argmin}} \sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} \left(\gamma c_i + (1-\gamma) d_i\right) a_{ip}$$

$$c_i = \sum_{j \in \mathcal{T}} (c^e_{ij} + c^s_{ij}) \quad c_i \in \mathcal{P}$$

$$d_i = \sum_{j \in \mathcal{T}} d_{ij}$$

$$B^{min}_p \leq b_i a_{ip} \leq B^{max}_p \quad \forall i \in \mathcal{L} \text{ and } \forall p \in \mathcal{P},$$

$$\sum_{i \in \mathcal{L}} a_{ip} = \lfloor \frac{M \Lambda_p}{\sum_{p \in \mathcal{P}} \Lambda_p} \rfloor \qquad \forall p \in \mathcal{P},$$

$$\sum_{p \in \mathcal{P}} a_{ip} \leq 1 \qquad \forall i \in \mathcal{L},$$

$$\sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} s_i a_{ip} \leq S^{max},$$

$$\sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} c_i a_{ip} \leq C^{max},$$

$$\frac{b_i a_{ip}}{b^*_n} \geq \tau, \quad \forall i \in \mathcal{L}, \ \forall n \in \mathcal{L}^* \text{ and } \forall p \in P.$$

# Problem Formulation

| Profiles: | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $B^{min}$ (*Mbps*) | 1 | 3 | 15 | 25 |
| $B^{max}$ (*Mbps*) | 4 | 20 | 30 | 40 |
| $\Lambda$ | 0.25 | 0.25 | 0.25 | 0.25 |

- Constraints
  - Bandwidth: $B_{min}$ and $B_{max}$
  - Computational and storage costs: $C_{max}$ and $S_{max}$
  - Encoding rate: minimum step size τ

$$\mathcal{L}^* : \underset{\mathcal{L}}{\operatorname{argmin}} \sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} \left(\gamma c_i + (1-\gamma) d_i\right) a_{ip}$$

$$B_p^{min} \leq b_i a_{ip} \leq B_p^{max} \quad \forall i \in \mathcal{L} \text{ and } \forall p \in \mathcal{P},$$

$$\sum_{i \in \mathcal{L}} a_{ip} = \lfloor \frac{M \Lambda_p}{\sum_{p \in \mathcal{P}} \Lambda_p} \rfloor \quad \forall p \in \mathcal{P},$$

$$\sum_{p \in \mathcal{P}} a_{ip} \leq 1 \quad \forall i \in \mathcal{L},$$

$$\sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} s_i a_{ip} \leq S^{max},$$

$$\sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} c_i a_{ip} \leq C^{max},$$

$$\frac{b_i a_{ip}}{b_n^*} \geq \tau, \quad \forall i \in \mathcal{L}, \ \forall n \in \mathcal{L}^* \text{ and } \forall p \in P.$$

# Experimental Setup

- N=10
- H.264/AVC
- Quality: WS-MSE and WS-PSNR
- Baselines: one-size-fits-all encoding ladders
- ILP using Pyomo

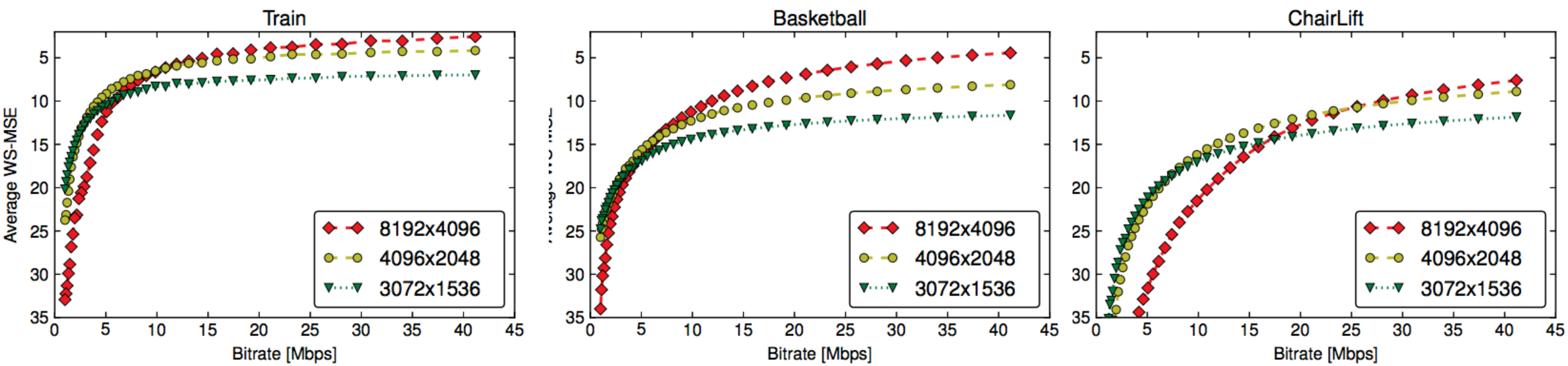
| Profiles: | $p_1$ | $p_2$ | $p_3$ | $p_4$ |
|---|---|---|---|---|
| $B^{min}$ (*Mbps*) | 1 | 3 | 15 | 25 |
| $B^{max}$ (*Mbps*) | 4 | 20 | 30 | 40 |
| $\Lambda$ | 0.25 | 0.25 | 0.25 | 0.25 |

| Sequence | $f_{spa}$ | $f_{tmp}$ | $\mathcal{O}$ |
|---|---|---|---|
| *Train* | 0.977 | 0.065 | $o_1$ |
| *Stitched_left_Dancing360_8K* | 0.884 | 0.110 | |
| *Basketball* | 0.843 | 0.090 | $o_2$ |
| *KiteFlite* | 0.861 | 0.090 | |
| *ChairLift* | 0.789 | 0.212 | $o_3$ |
| *SkateboardInLot* | 0.827 | 0.521 | |

| **Apple** [15] | | **Axinom** [21] | | **Netflix** [16] | |
|---|---|---|---|---|---|
| $Z$ (Mbps) | $W \times H$ | $Z$ (Mbps) | $W \times H$ | $Z$ (Mbps) | $W \times H$ |
| 45 | 8192 × 4096 | 45 | 8192 × 4096 | 43 | 8192 × 4096 |
| 30 | 8192 × 4096 | 30 | 8192 × 4096 | 30 | 4096 × 2048 |
| 20 | 4096 × 2048 | 21 | 4096 × 2048 | 23.5 | 4096 × 2048 |
| 11 | 3072 × 1536 | 12 | 3072 × 1536 | 17.5 | 3072 × 1536 |

# RD Performance Gain

- High diversity in video content characteristics
- One-size-fits-all schemes cannot provide cost-optimal and high-quality streaming performances

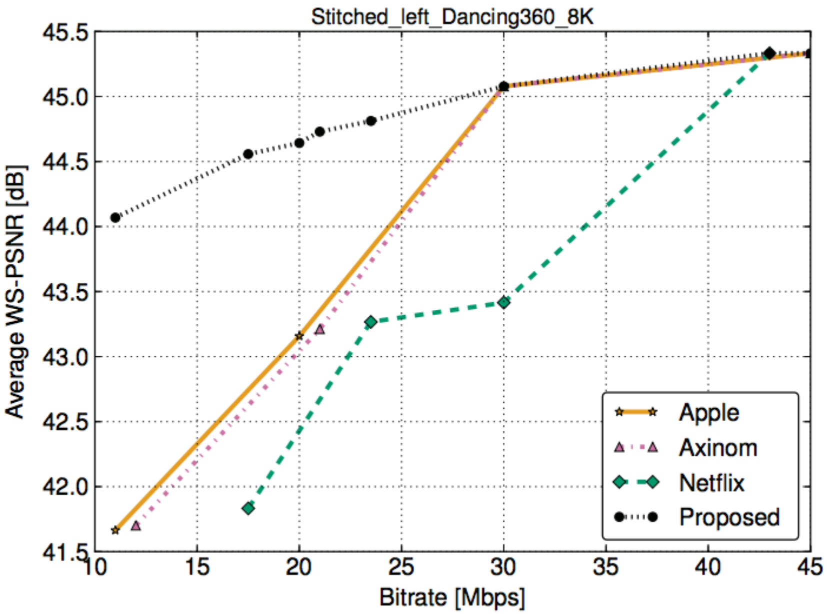


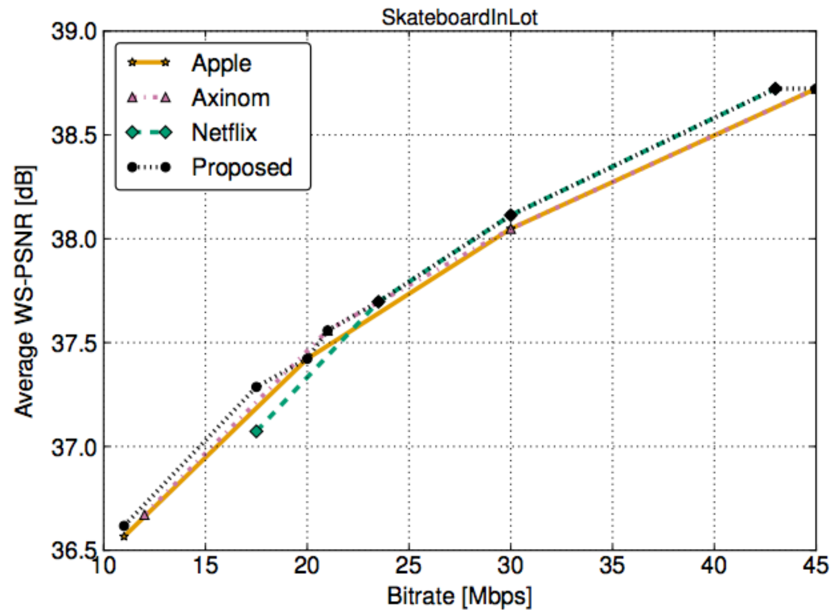
# RD Performance Gain

| Sequence | $f_{spa}$ | $f_{tmp}$ | $\mathcal{O}$ |
|---|---|---|---|
| *Train* | 0.977 | 0.065 | $o_1$ |
| *Stitched_left_Dancing360_8K* | 0.884 | 0.110 | |
| *Basketball* | 0.843 | 0.090 | $o_2$ |
| *KiteFlite* | 0.861 | 0.090 | |
| *ChairLift* | 0.789 | 0.212 | $o_3$ |
| *SkateboardInLot* | 0.827 | 0.521 | |

- γ = 0

$$\mathcal{L}^* : \operatorname*{argmin}_{\mathcal{L}} \sum_{i \in \mathcal{L}} \sum_{p \in \mathcal{P}} (\gamma c_i + (1-\gamma) d_i)\, a_{ip}$$



# BD Rate (%)

| **Sequence** $v$ | **Streaming vendor** | | |
|---|---|---|---|
| | Apple | Axinom | Netflix |
| *Stitched_left_Dancing360_8K* | -5.557 | -5.885 | -69.253 |
| *KiteFlite* | -13.876 | -14.436 | -69.178 |
| *SkateboardInLot* | -1.673 | -1.701 | -1.155 |

# Experiment Scenario

- Constraints of $S_{max}$ and $C_{max}$ are 8000, $\tau = 1.2$, and M = 12

| Sequence $v$ | $\gamma$ | Representation $i$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| *Stitched_left_Dancing360_8K* | 0.0 | ($g_1$,1.47) | ($g_1$,1.78) | ($g_1$,2.15) | ($g_1$,3.8) | ($g_1$,4.6) | ($g_1$,5.6) | ($g_2$,10.84) | ($g_2$,13.11) | ($g_2$,15.87) | ($g_2$,28.11) | ($g_3$,34.01) | ($g_3$,41.15) |
| | 0.1 | ($g_2$,1.34) | ($g_2$,1.61) | ($g_2$,1.95) | ($g_2$,2.60) | ($g_3$,3.14) | ($g_3$,3.80) | ($g_3$,6.12) | ($g_3$,7.40) | ($g_3$,8.96) | ($g_3$,17.45) | ($g_3$,21.12) | ($g_3$,25.55) |
| | 0.5 | ($g_2$,1.00) | ($g_2$,1.21) | ($g_2$,1.47) | ($g_2$,2.36) | ($g_3$,2.86) | ($g_3$,3.46) | ($g_3$,6.12) | ($g_3$,7.40) | ($g_3$,8.96) | ($g_3$,17.45) | ($g_3$,21.12) | ($g_3$,25.55) |
| *KiteFlite* | 0.0 | ($g_1$,1.47) | ($g_1$,1.78) | ($g_2$,2.15) | ($g_2$,3.80) | ($g_2$,4.60) | ($g_3$,5.56) | ($g_3$,10.84) | ($g_3$,13.11) | ($g_3$,15.87) | ($g_3$,28.11) | ($g_3$,34.01) | ($g_3$,41.15) |
| | 0.1 | ($g_1$,1.47) | ($g_1$,1.78) | ($g_2$,2.15) | ($g_2$,3.80) | ($g_2$,4.60) | ($g_3$,5.56) | ($g_3$,6.73) | ($g_3$,8.14) | ($g_3$,9.85) | ($g_3$,17.45) | ($g_3$,21.12) | ($g_3$,25.55) |
| | 0.5 | ($g_1$,1.00) | ($g_1$,1.21) | ($g_1$,1.47) | ($g_2$,2.36) | ($g_2$,2.86) | ($g_2$,3.46) | ($g_3$,6.12) | ($g_3$,7.40) | ($g_3$,8.96) | ($g_3$,17.45) | ($g_3$,21.12) | ($g_3$,25.55) |
| *SkateboardInLot* | 0.0 | ($g_1$,1.47) | ($g_1$,1.78) | ($g_1$,2.15) | ($g_1$,3.80) | ($g_1$,4.60) | ($g_1$,5.56) | ($g_2$,10.84) | ($g_2$,13.11) | ($g_2$,15.87) | ($g_2$,28.11) | ($g_3$,34.01) | ($g_3$,41.15) |
| | 0.1 | ($g_1$,1.47) | ($g_1$,1.78) | ($g_1$,2.15) | ($g_1$,2.86) | ($g_1$,3.46) | ($g_1$,4.18) | ($g_1$,6.12) | ($g_1$,7.40) | ($g_1$,8.96) | ($g_1$,17.45) | ($g_2$,21.12) | ($g_2$,25.55) |
| | 0.5 | ($g_1$,1.21) | ($g_1$,1.47) | ($g_1$,1.78) | ($g_1$,2.36) | ($g_1$,2.86) | ($g_1$,3.46) | ($g_1$,6.12) | ($g_1$,7.40) | ($g_1$,8.96) | ($g_2$,17.45) | ($g_2$,21.12) | ($g_2$,25.55) |

| **Sequence $v$** | $\Delta$cost (%) | | $\Delta$distortion (%) | |
|---|---|---|---|---|
| | $\gamma = 0.1$ | $\gamma = 0.5$ | $\gamma = 0.1$ | $\gamma = 0.5$ |
| *Stitched_left_Dancing360_8K* | 37.463 | 39.683 | -13.628 | -42.914 |
| *KiteFlite* | 33.165 | 39.206 | -9.564 | -25.326 |
| *SkateboardInLot* | 37.214 | 38.884 | -8.977 | -15.26 |

# Conclusion

- A novel encoding ladder estimation method for tiled 360◦ video streaming systems, considering both the provider’s and client’s perspectives
- The developed system included classification of the content type, distortion modeling, cost modeling, and problem formulation
- Achieved significant bitrate savings compared to the one-size-fits-all encoding ladders
- Automatically find cost-optimal encoding ladders using several practical constraint