Estimation of cost-optimal encoding ladders for tiled 360-degree videos in adaptive streaming systems

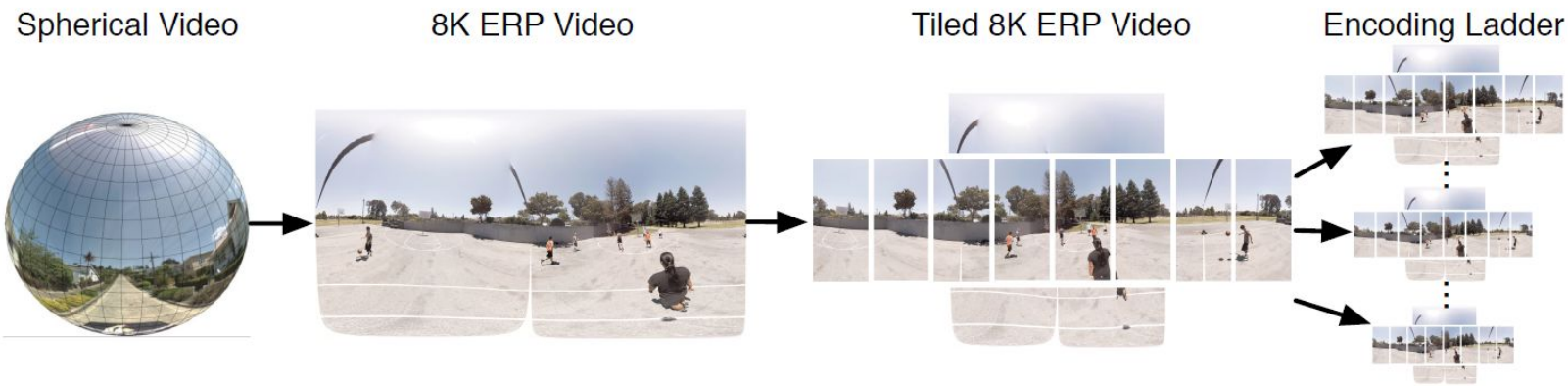
Cagri Ozcinar, Ana De Abreu, Sebastian Knorr, and Aljosa Smolic
Trinity College Dublin (TCD), Dublin 2, Ireland.

Outline

- Introduction
- Problem to be solved
- Proposed system model
  - Classification of the content type
  - Distortion modeling
  - Cost modeling
  - Problem formulation
- Evaluation
- Conclusion

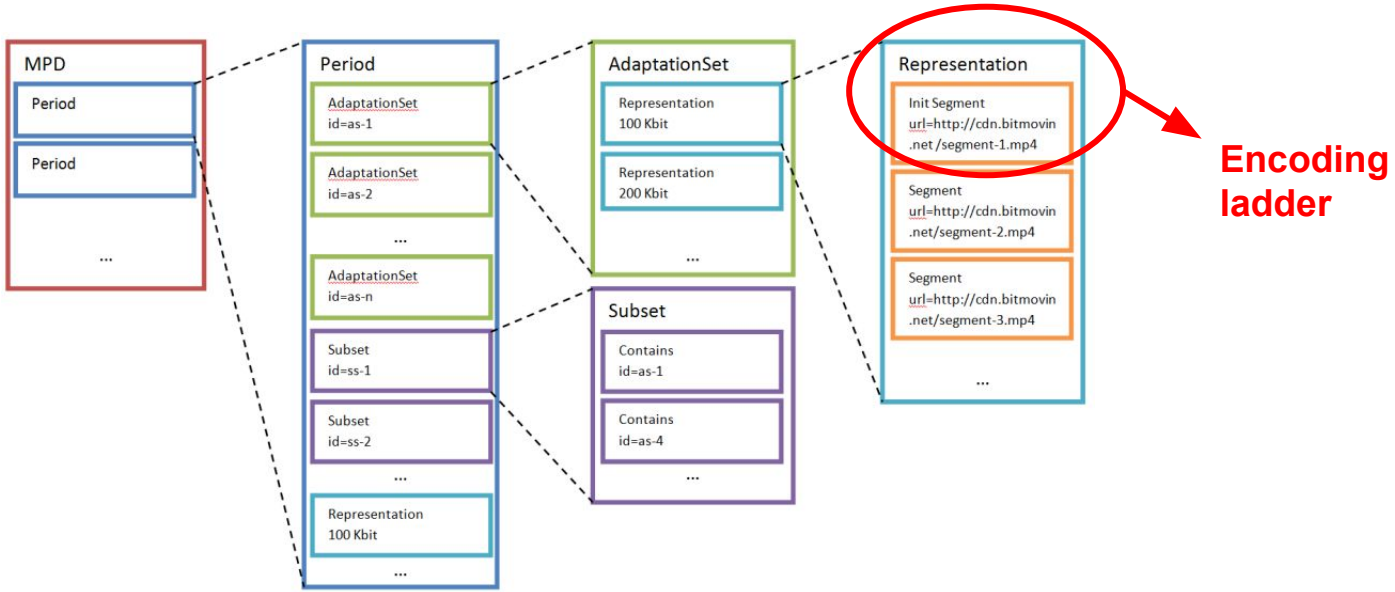
Introduction

- Streaming 360-degree video is challenging
- It is very high resolution, such as 4K, 8K equi-rectangular projection (ERP) or higher
- We only stream the user's field of view using tile-based encoding and adaptive streaming (DASH)



What do people usually do?

- ERP
- Tile-based encoding
- DASH with its extension, Spatial Relationship Description (SRD)



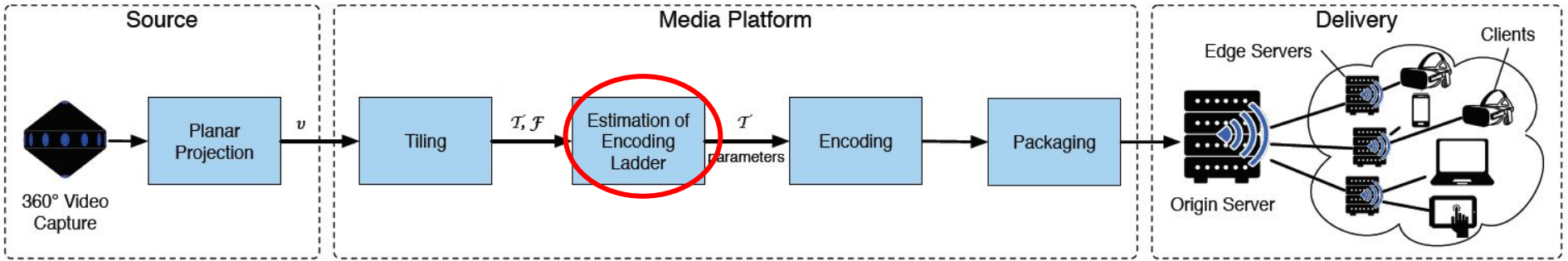
Problem to be solved

- Most recent work focused on the client's perspective without considering the service providers' perspective
- Client's perspective: end-users' latency, bandwidth, distortion, QoE
- Provider's perspective: computation cost and storage utilization

⇨ Estimation method of cost-optimal encoding ladders in adaptive streaming systems by considering both the provider's and client's perspective for tiled 360 video streaming

Proposed system

- Estimation of Encoding ladder contains 4 major components:
  - Classification of the content type
  - Distortion modeling
  - Cost modeling
  - Problem formulation
- Minimize the service provider's resource costs while providing high quality 360° video streaming experience (distortion)



Classification of the content type

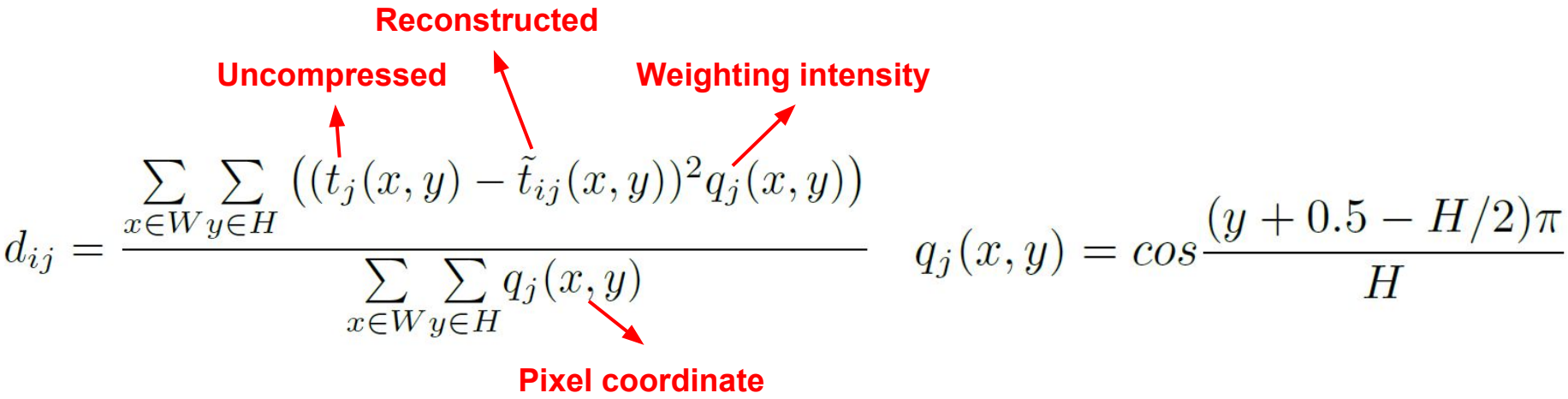
- Spatial complexity and temporal complexity

$$\mathcal{F} = \{f_{spa}, f_{tmp}\}$$

- 2-pass constant rate factor (CRF) encoding, which has the QPs slightly varied across the time based on the scene complexity, action, and motion
- The average size of I- and P- frames can be used to determine the complexity features
- Content type: o1, o2, and o3 (simple->complex)

Distortion modeling

- Spherical distortion (spherical content onto the planar surface)
- Weighted-to-spherically uniform mean square error (WS-MSE) [1]
- The noise power for the i-th representation of the j-th tile



[1] JVET, “AHG8: WS-PSNR for 360 video objective quality evaluation,” Tech. Rep. JVET-D0040, JTC1/SC29/WG11, ISO/IEC, Chengdu, CN, Oct. 2016.

Cost modeling - Encoding cost

- Encoding cost ce can be described for the j-th tile of the i-th representation
- The cost calculation model used by the Amazon cloud service [2]

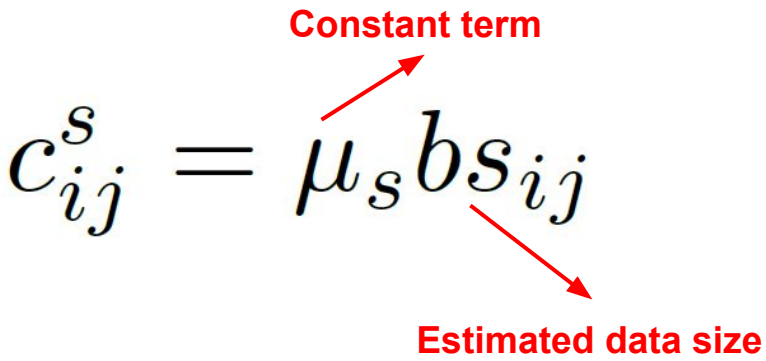
Constant term

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

[2] Amazon webservices, “Amazon elastic transcoder pricing, ”https://aws.amazon.com/elastictranscoder/pricing/, Jul 2017

Cost modeling - Storage cost

- Linear cost model
- The storage cost for the j-th tile of the i-th representation



Problem formulation

- To minimize both the total spherical distortion and the total resource cost
- Constraints
  - **Bandwidth**, a set of given network bandwidth profiles {P}
  - **Computational and storage costs**, limitations for the encoding and storage costs
  - **Encoding rate**, the bitrate levels of the representations

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

Pre-defined constant [0,1] **Decision variable {0,1}**

Evaluation settings

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

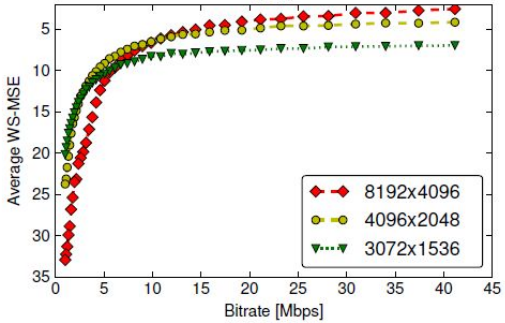
- Eight 8Kx4K resolution 360° ERP video test sequences
  - V = {Train, Stitched left Dancing360 8K, Basketball, KiteFlite, ChairLift, SkateboardInLot}
- Each video was split into N = 10 tiles
- Content type = {O} = {o1, o2, o3}
- Resolution = {G} = {g1, g2, g3} = {3072x1536, 4096x2048, 8192x4096}
- Bandwidth = {P} = {p1, p2, p3, p4}

Sequence	f_{spa}	f_{tmp}	$\mathcal{O}$
Train	234	0.501	o_1
Stitched_left_Dancing360_8K	313	0.501	
Basketball	1167	0.502	o_2
KiteFlite	1547	0.502	
ChairLift	2842	0.502	o_3
SkateboardInLot	3977	0.502	

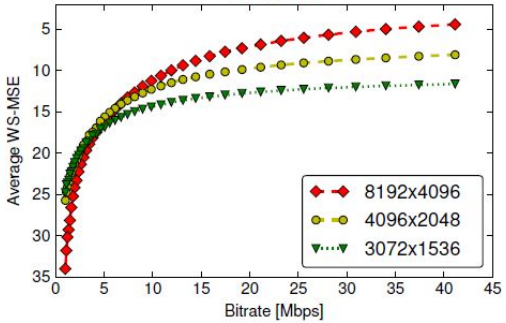
Evaluation results

- WS-MSE versus bitrate (in Mbps) performance
  - Each content type has various content dependencies for each encoding resolution and bitrate
  - The lowest complex encoding features, achieves a low distortion score
  - High-resolution version has a higher sensitivity for unpredictable motions, which requires further residuals to avoid visual distortions

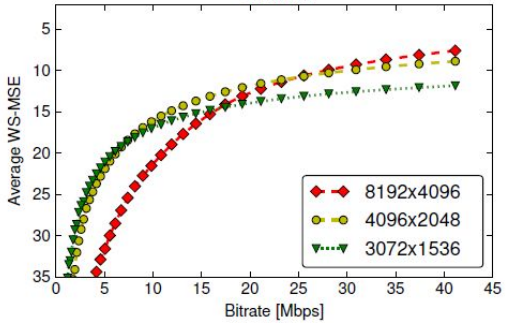
o1



o2



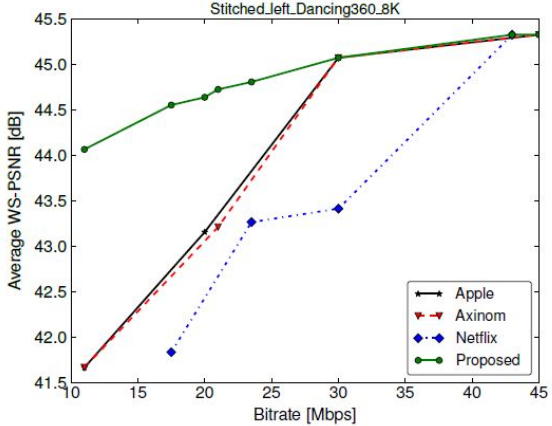
o3



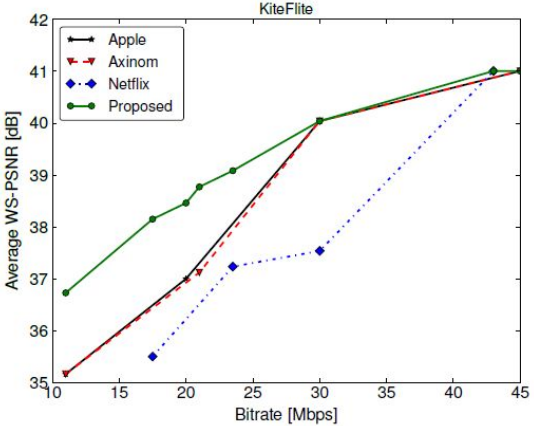
Evaluation results

- RD performance gain
  - Proposed method considerably increases the objective video quality (i.e., WS-PSNR) [1]
  - High bitrate savings between 10-30 Mbps bandwidth ranges for the content types o1 and o2

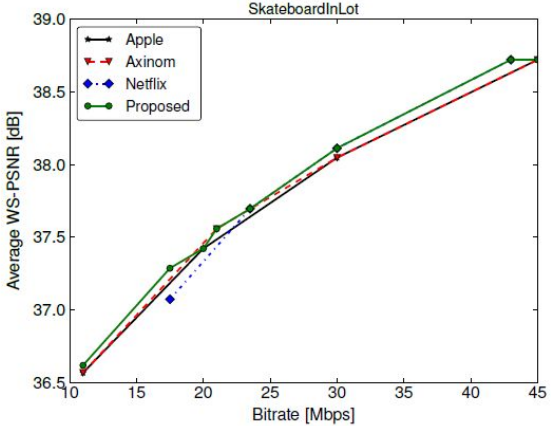
o1



o2



o3



[1] JVET, “AHG8: WS-PSNR for 360 video objective quality evaluation,” Tech. Rep. JVET-D0040, JTC1/SC29/WG11, ISO/IEC, Chengdu, CN, Oct. 2016.

Evaluation results (cont.)

- Bjøntegaard metric [1] (BD-rate)
- A negative BD-rate indicates a decrease of bitrate at the same quality
- Proposed method provides considerable bitrate savings compared to the recommended encoding ladders at the same bitrates.

Sequence v	**Streaming vendor**		
	Apple	Axinom	Netflix
Stitched_left_Dancing360_8K	-4.838	-7.070	-1.102
KiteFlite	-13.937	-20.395	-68.299
SkateboardInLot	-1.659	-1.094	-1.144

Table 5: BD-rate saving (%) of the proposed method.

[1] G. Bjøtegaard, “Calculation of average PSNR differences between RD-curves (vceg-m33),” Tech. Rep. M16090, VCEG Meeting (ITUT SG16 Q.6), Austin, Texas, USA,, Apr 2001

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

Evaluation results (cont.)

- Content type o1 increases its encoding resolution and decreases its target encoding rate
- Content type o3, decreases both its encoding resolution and target encoding rate
- Resolution = {G} = {g1,g2,g3} = {3072x1536, 4096x2048, 8192x4096}

Sequence v	γ	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)

Table 6: Results of the proposed encoding ladder estimation for $\gamma = 0$, $\gamma = 0.1$, and $\gamma = 0.5$.

Conclusion

- A novel encoding ladder estimation method for tiled 360 video streaming systems, considering both the provider's and client's perspectives
- Proposed method provides cost-optimal and enhanced video streaming experiences for VR end-users

Sequence v	Δcost (%)		Δdistortion (%)	
	$\gamma = 0.1$	$\gamma = 0.5$	$\gamma = 0.1$	$\gamma = 0.5$
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

Table 7: Total cost saving and distortion gain with respect to γ=0.0.