Energy-Efficient Video Processing for Virtual Reality

Yue Leng*, UIUC

Chi-chun Chen*, University of Rochester

Qiuyue Sun, University of Rochester

Jian Huang, UIUC

Yuhao Zhu, University of Rochester

* Co-primary authors

Virtual Reality

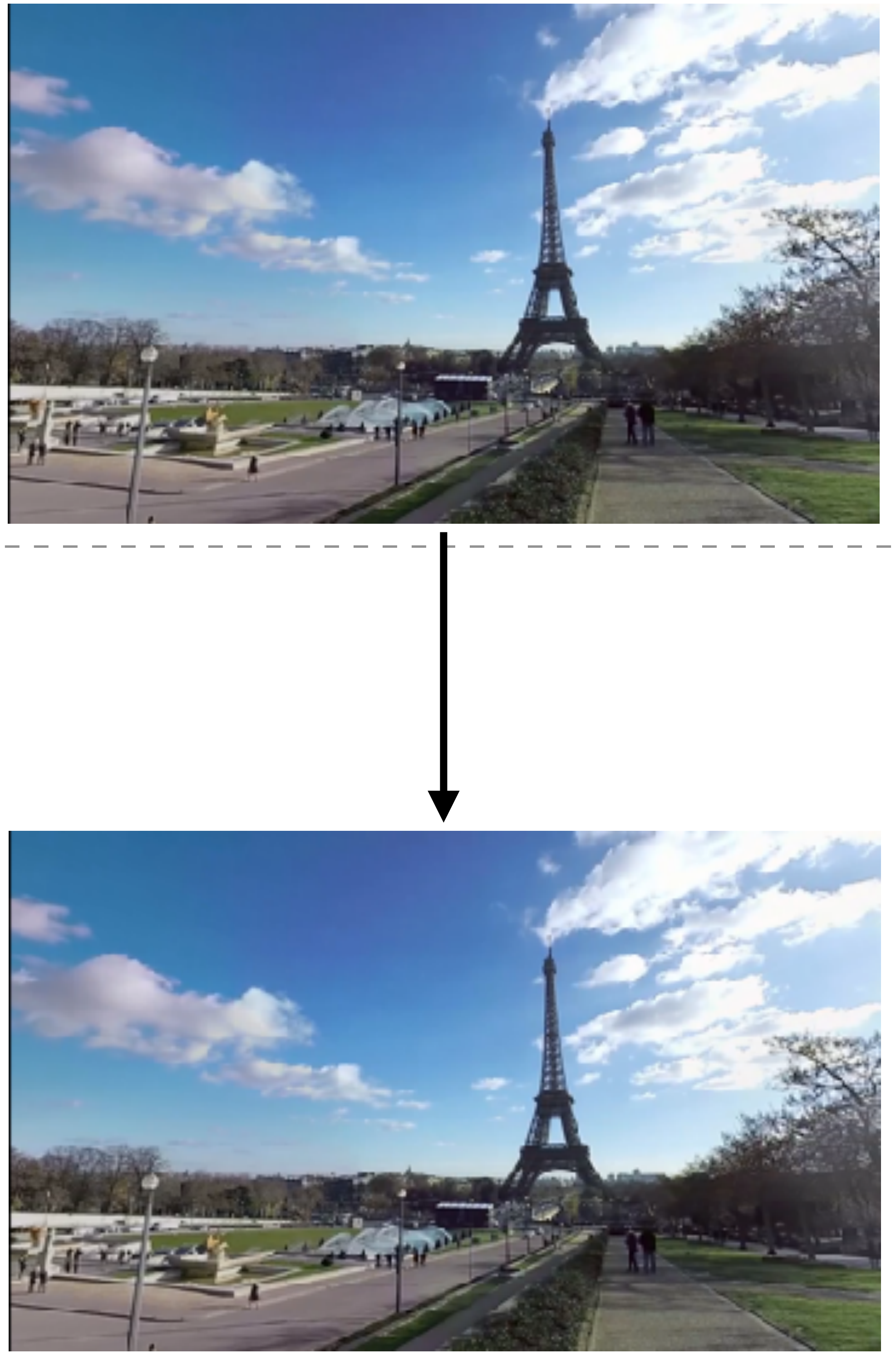
Conventional Video Processing

Server

Client

Conventional Video Processing

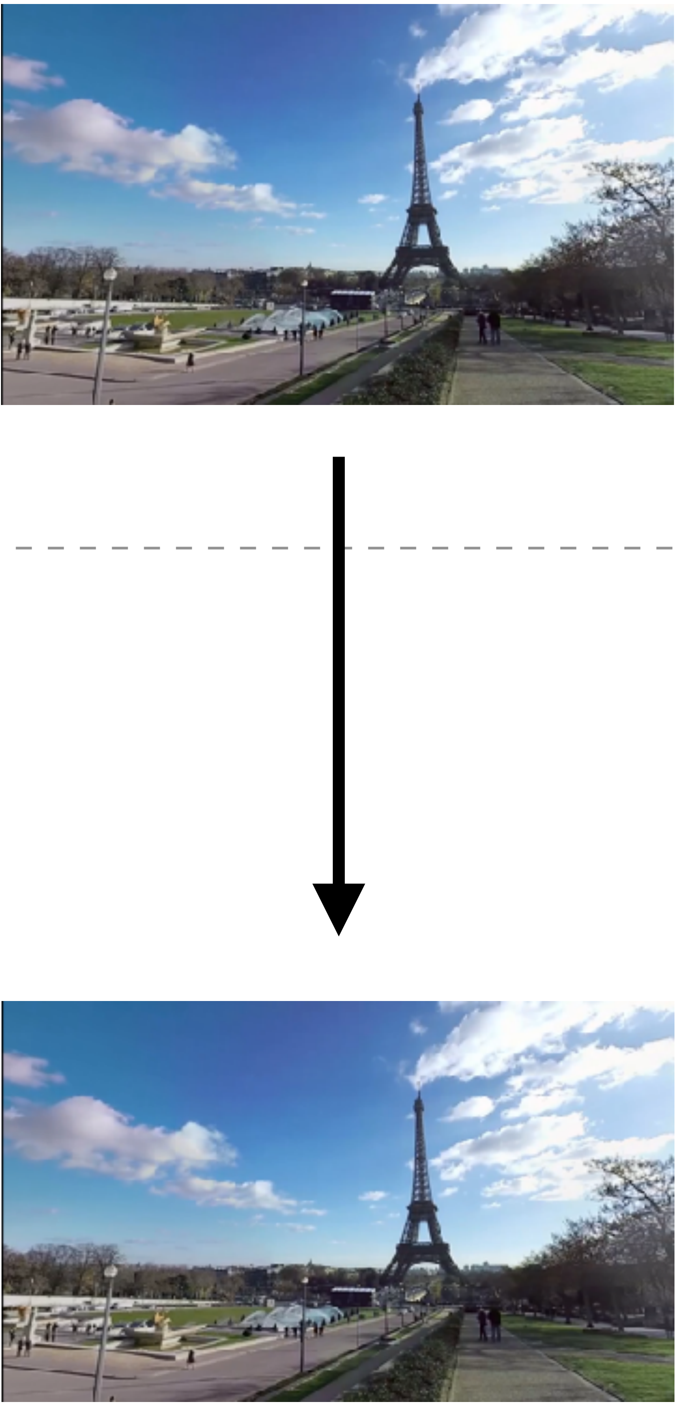
Server



Client

Conventional Video Processing

VR Video Processing



Server

Client

ourdet
Films
Frédéric
VR/360°

Conventional Video Processing

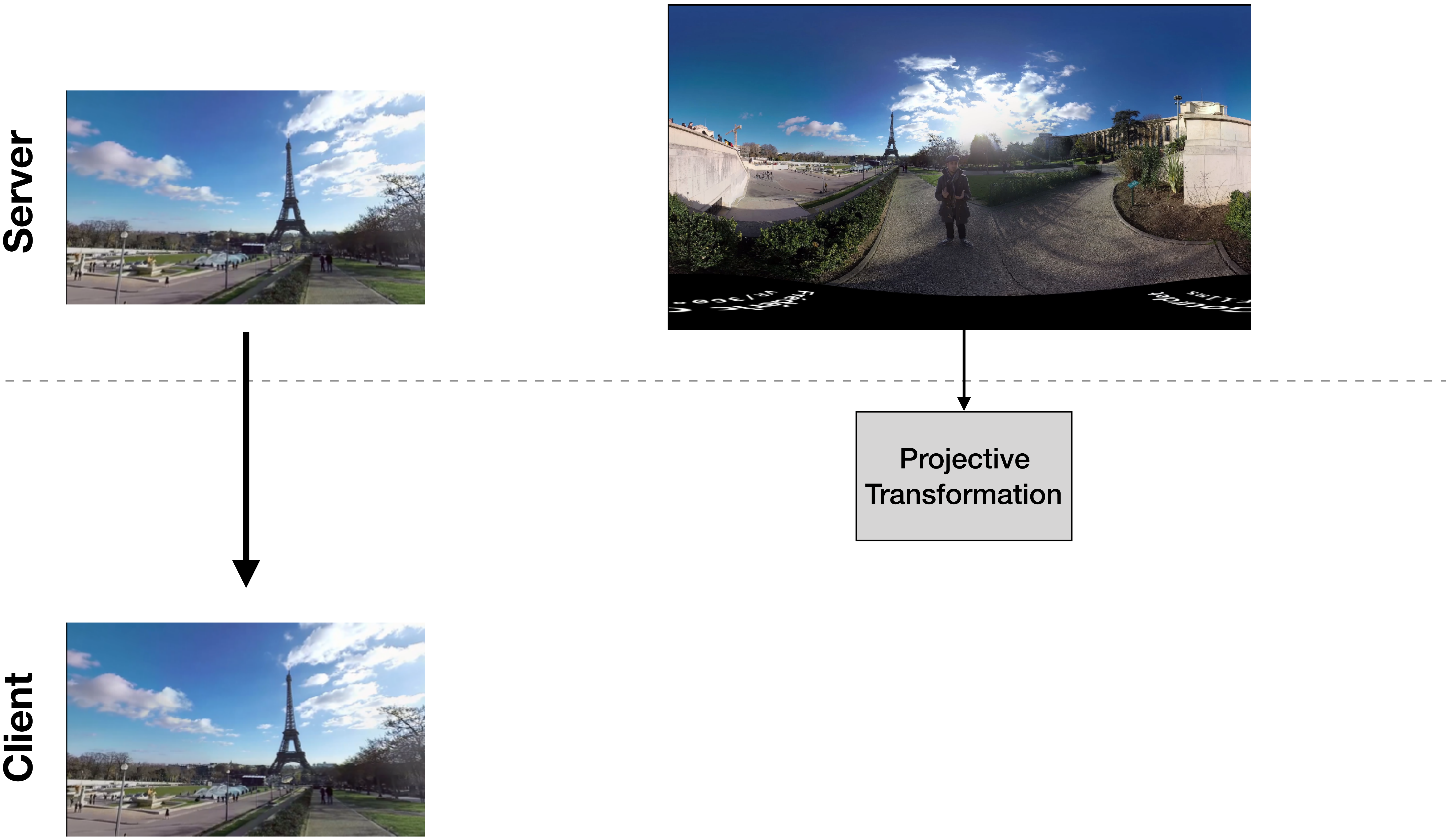
VR Video Processing

Server

Client

Conventional Video Processing

VR Video Processing



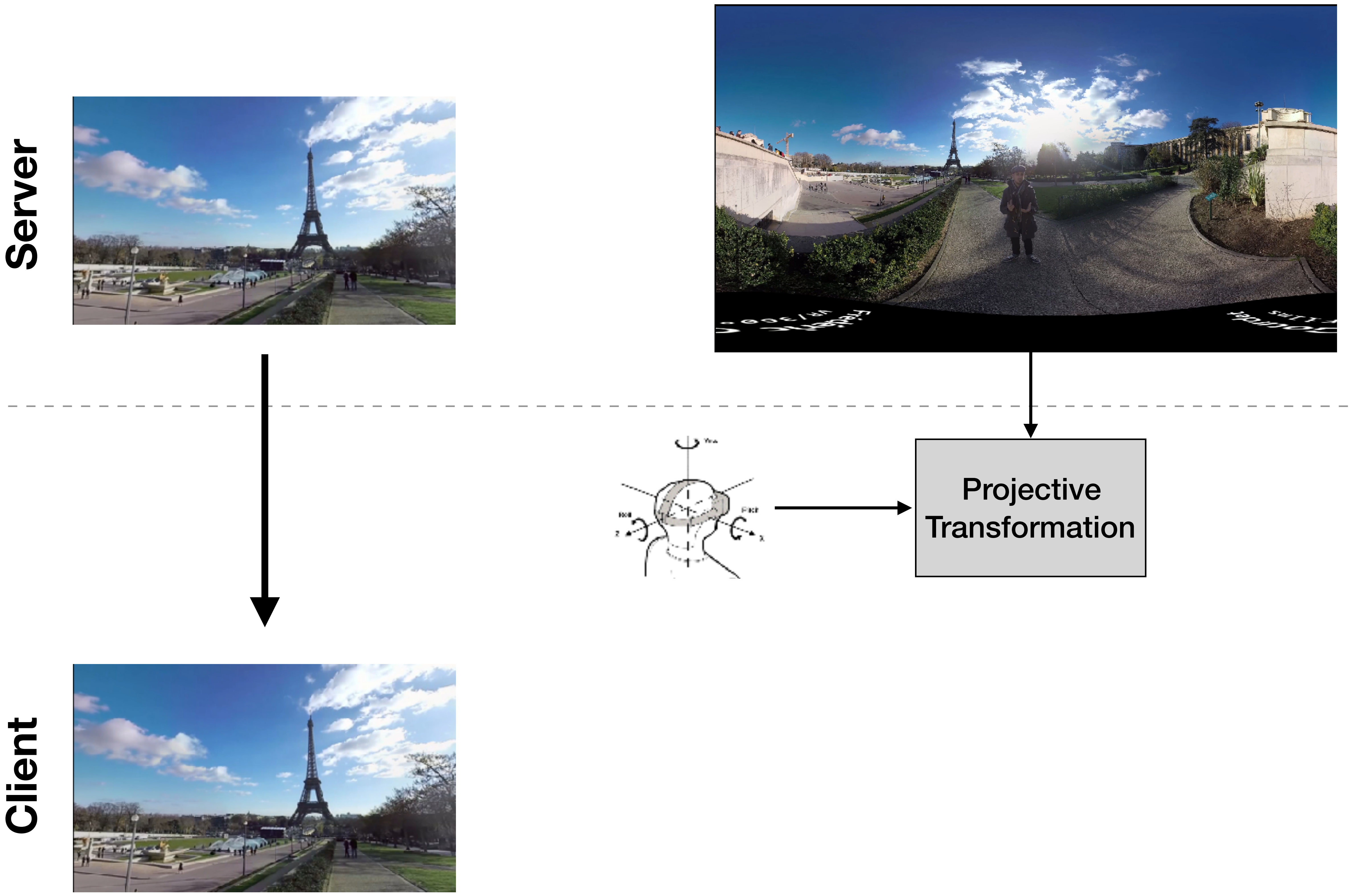
Server

Client

Projective Transformation

Conventional Video Processing

VR Video Processing



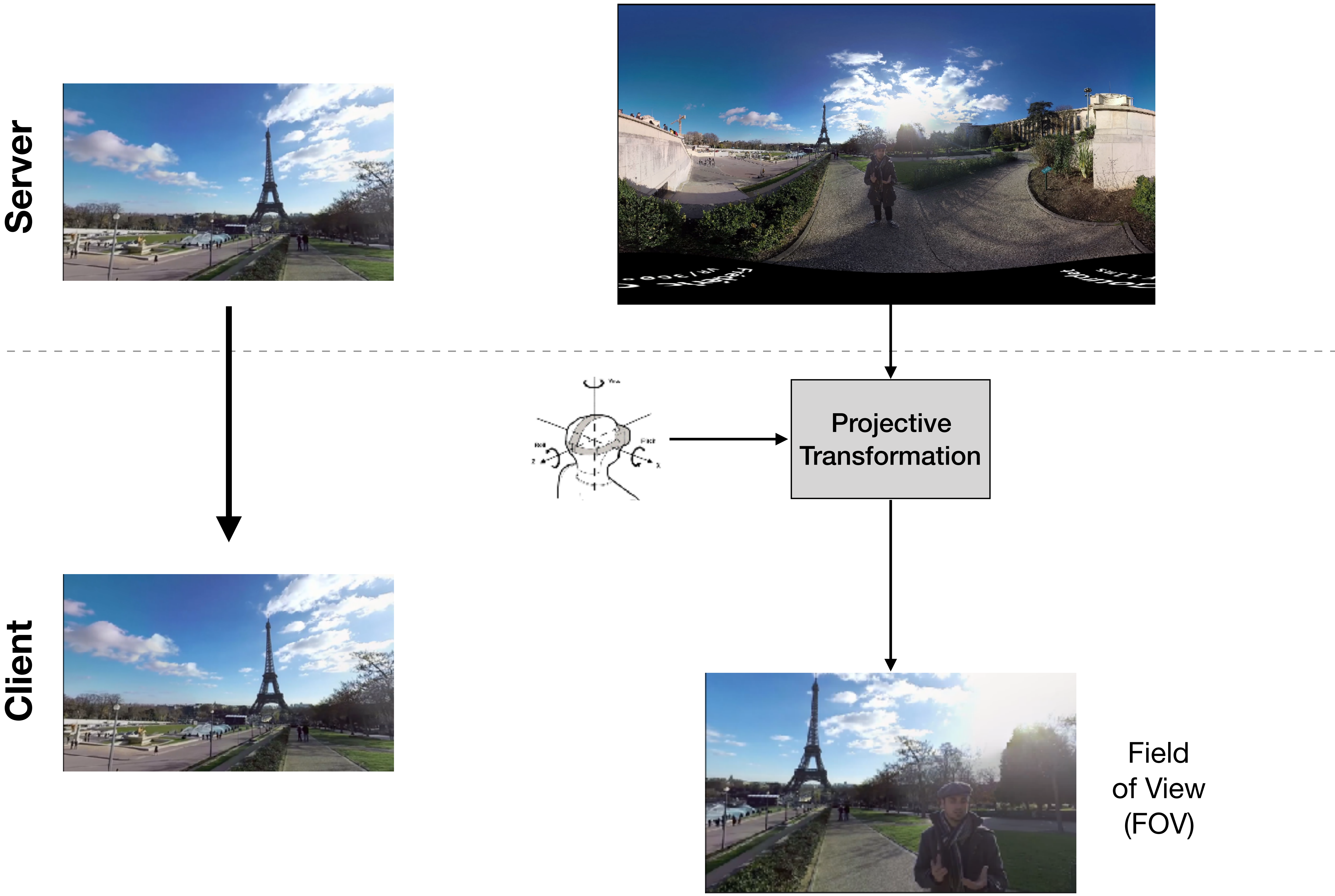
Server

Client

Projective Transformation

Conventional Video Processing

VR Video Processing



Server

Client

Projective Transformation

Field of View (FOV)

Excessive Power Consumption

Excessive Power Consumption

- Custom VR client device on Nvidia Tegra X2

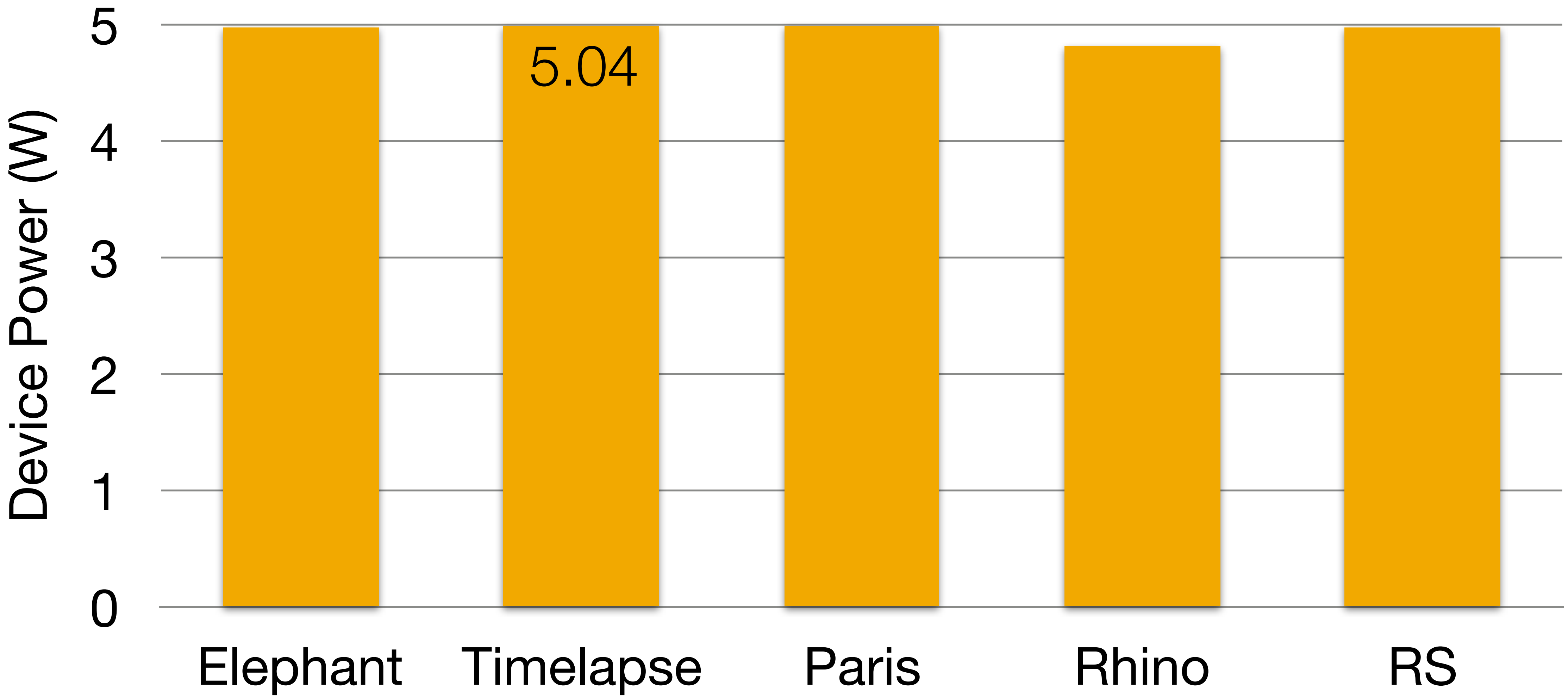
Excessive Power Consumption

- Custom VR client device on Nvidia Tegra X2
- Benchmark: 360-Degree Video Head Movement Dataset

[1] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204. DOI: https://doi.org/10.1145/3083187.3083215

Excessive Power Consumption

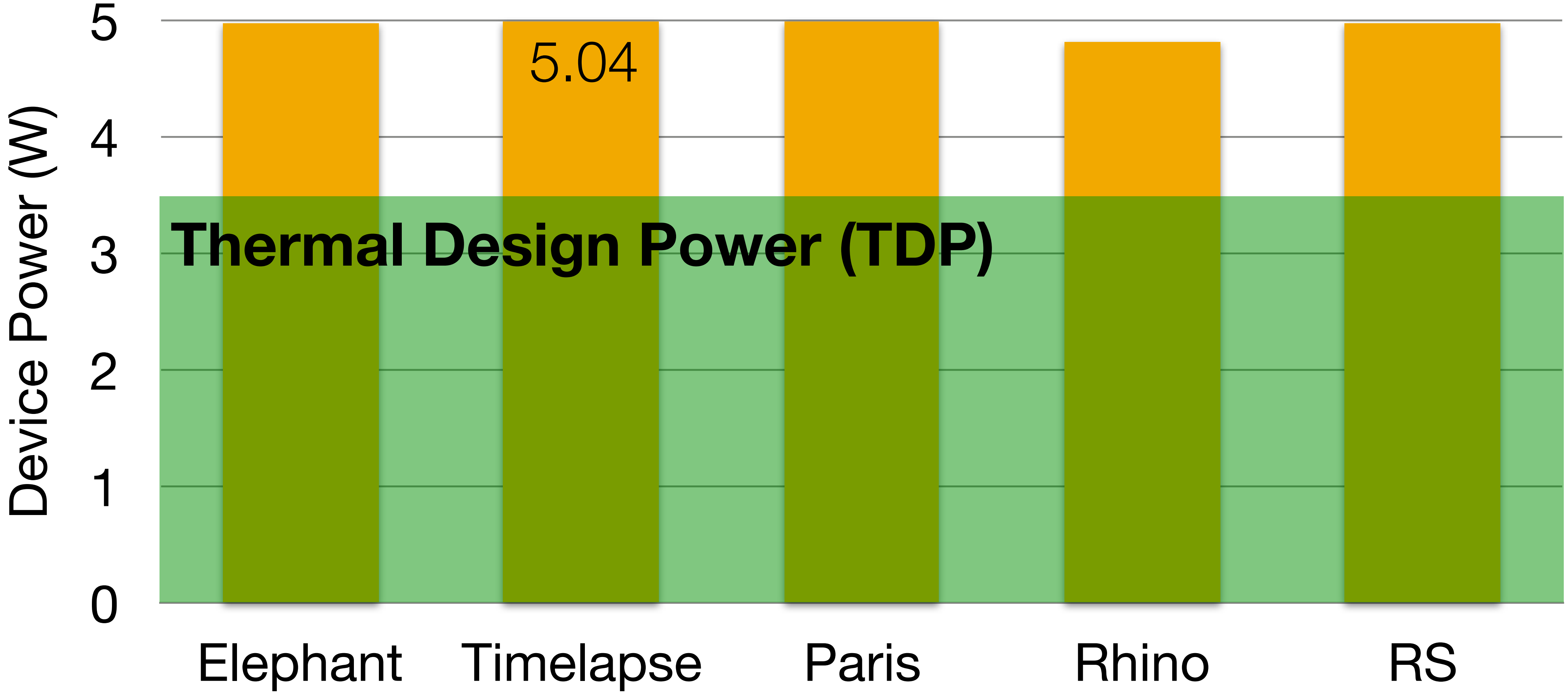
- Custom VR client device on Nvidia Tegra X2
- Benchmark: 360-Degree Video Head Movement Dataset



[1] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204. DOI: https://doi.org/10.1145/3083187.3083215

Excessive Power Consumption

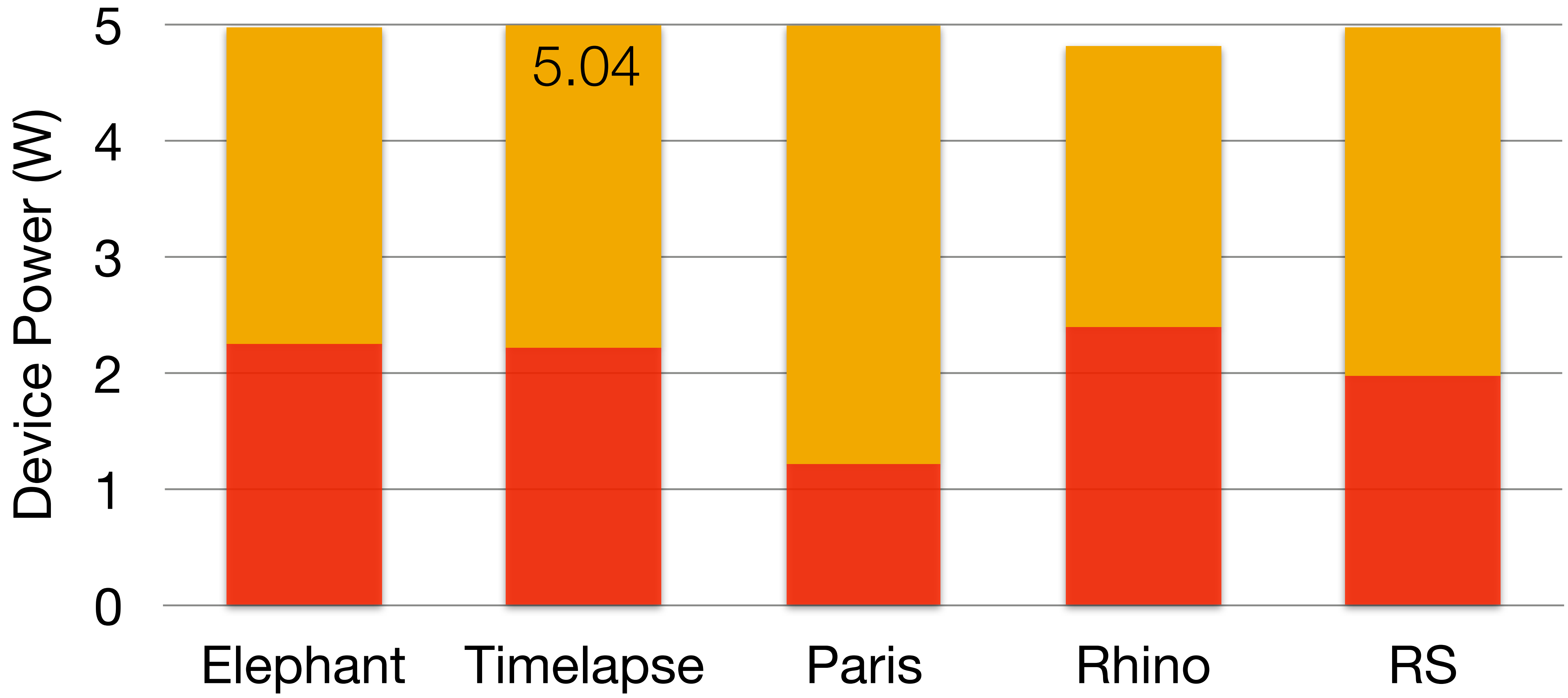
- Custom VR client device on Nvidia Tegra X2
- Benchmark: 360-Degree Video Head Movement Dataset



[1] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204. DOI: https://doi.org/10.1145/3083187.3083215

Excessive Power Consumption

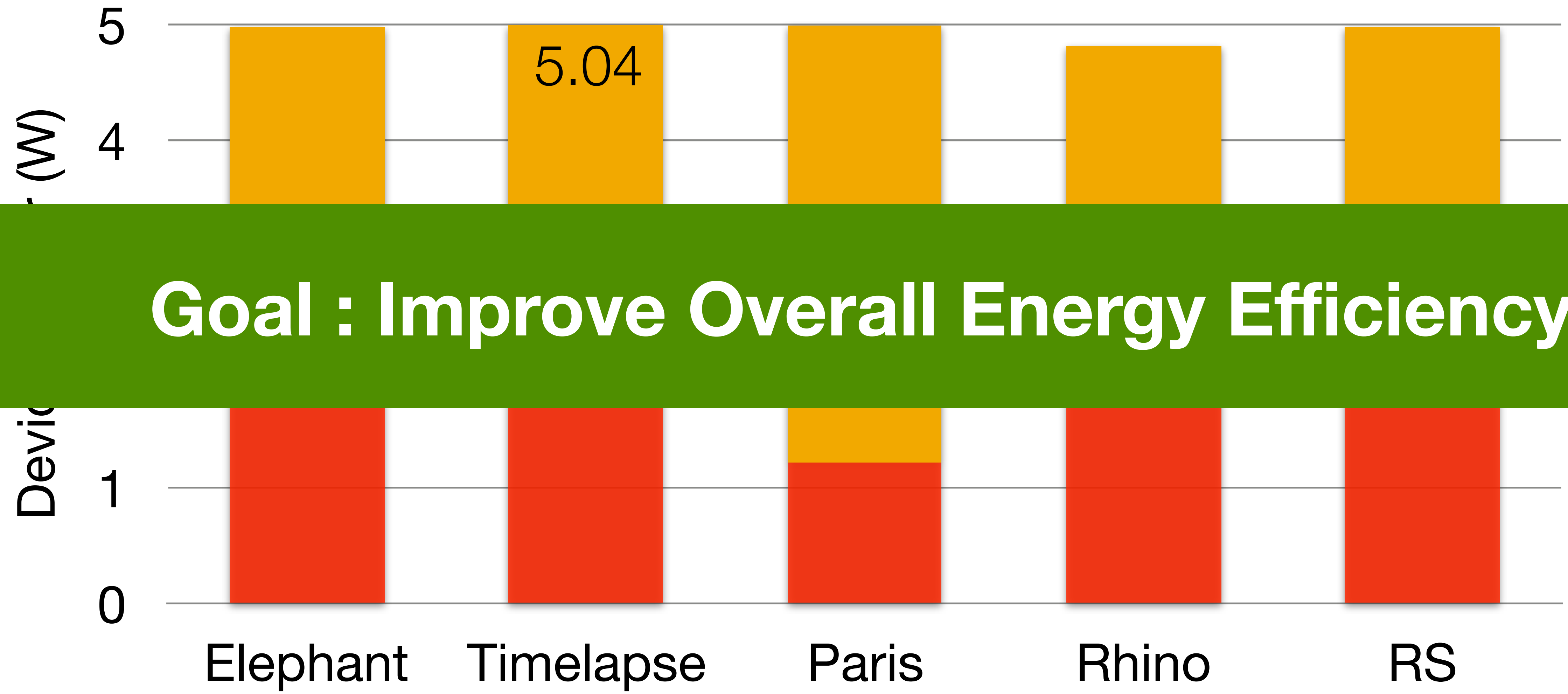
- Custom VR client device on Nvidia Tegra X2
- Benchmark: 360-Degree Video Head Movement Dataset
- Projective transformation (PT) is a major contributor (40%)



[1] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204. DOI: https://doi.org/10.1145/3083187.3083215

Excessive Power Consumption

- Custom VR client device on Nvidia Tegra X2
- Benchmark: 360-Degree Video Head Movement Dataset
- Projective transformation (PT) is a major contributor (40%)



Goal : Improve Overall Energy Efficiency

[1] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 2017. 360-Degree Video Head Movement Dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 199-204. DOI: https://doi.org/10.1145/3083187.3083215

EVR

A Cloud-Client Collaborative VR Video System

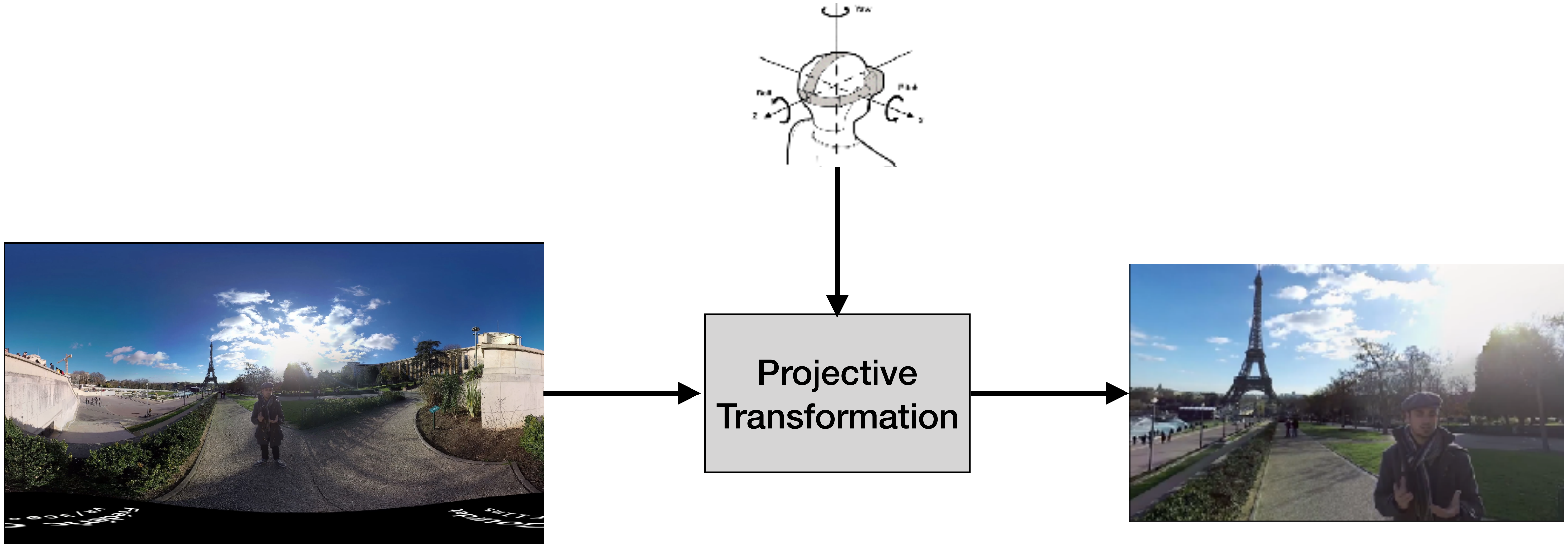
Cloud	Semantic Aware Streaming
Client	Hardware Accelerated Rendering
Result	Up to 58% Energy Reduction & Only 1% Frame Rate Drops

EVR

A Cloud-Client Collaborative VR Video System

Cloud	Semantic Aware Streaming
Client	Hardware Accelerated Rendering
Result	Up to 58% Energy Reduction & Only 1% Frame Rate Drops

Pre-render FOVs in the cloud

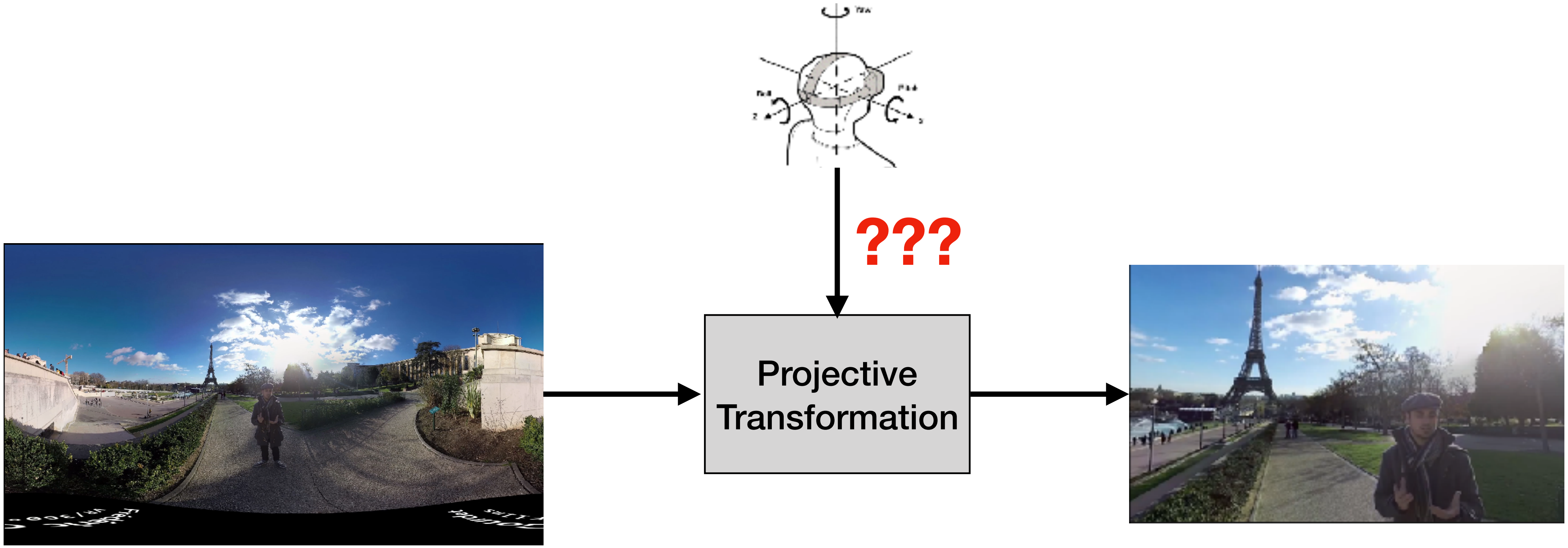


Projective Transformation

Field of View (FOV)

Pre-render FOVs in the cloud

- Challenge: How to obtain the viewing area without communicating with sensors on the client?



Field of View (FOV)

Leveraging Video Semantics Information

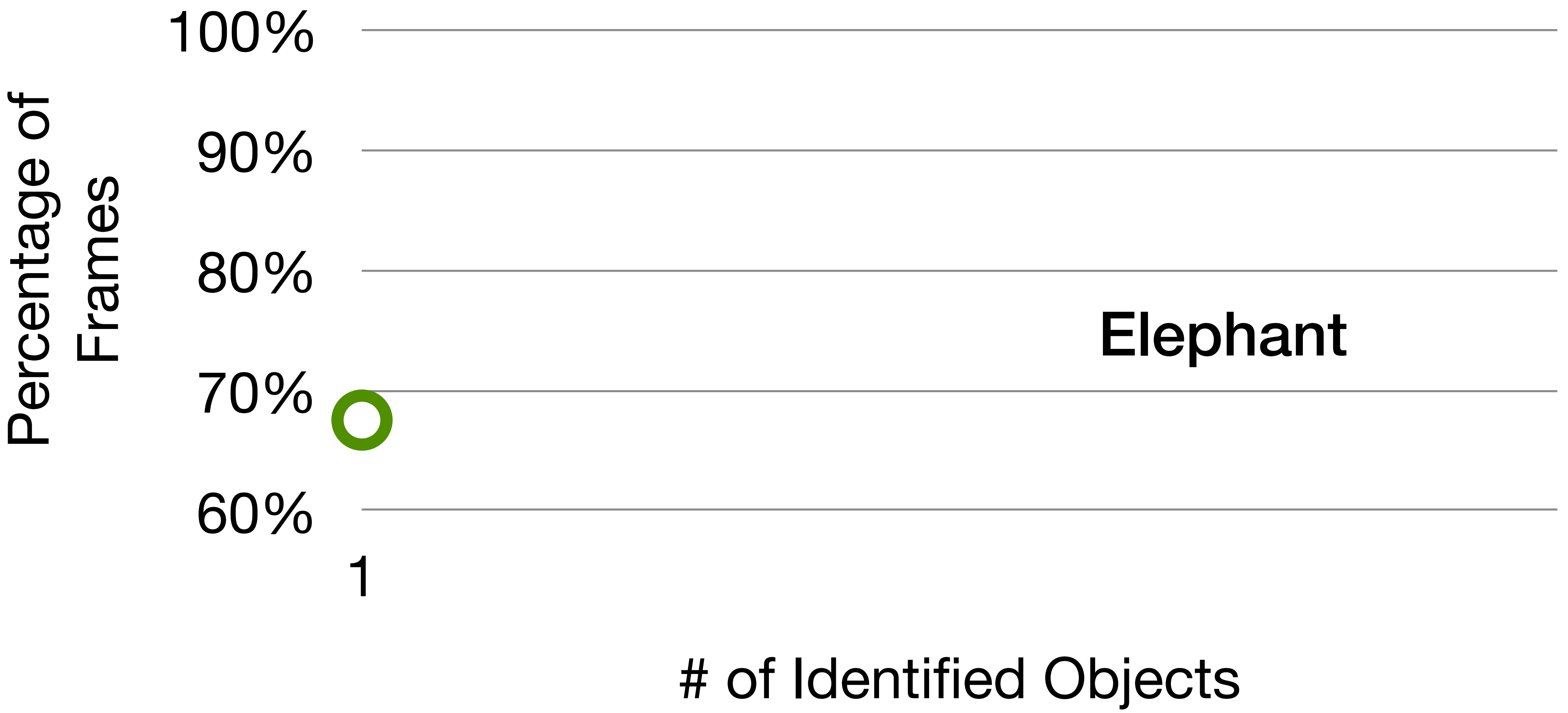
- Observations:

Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects

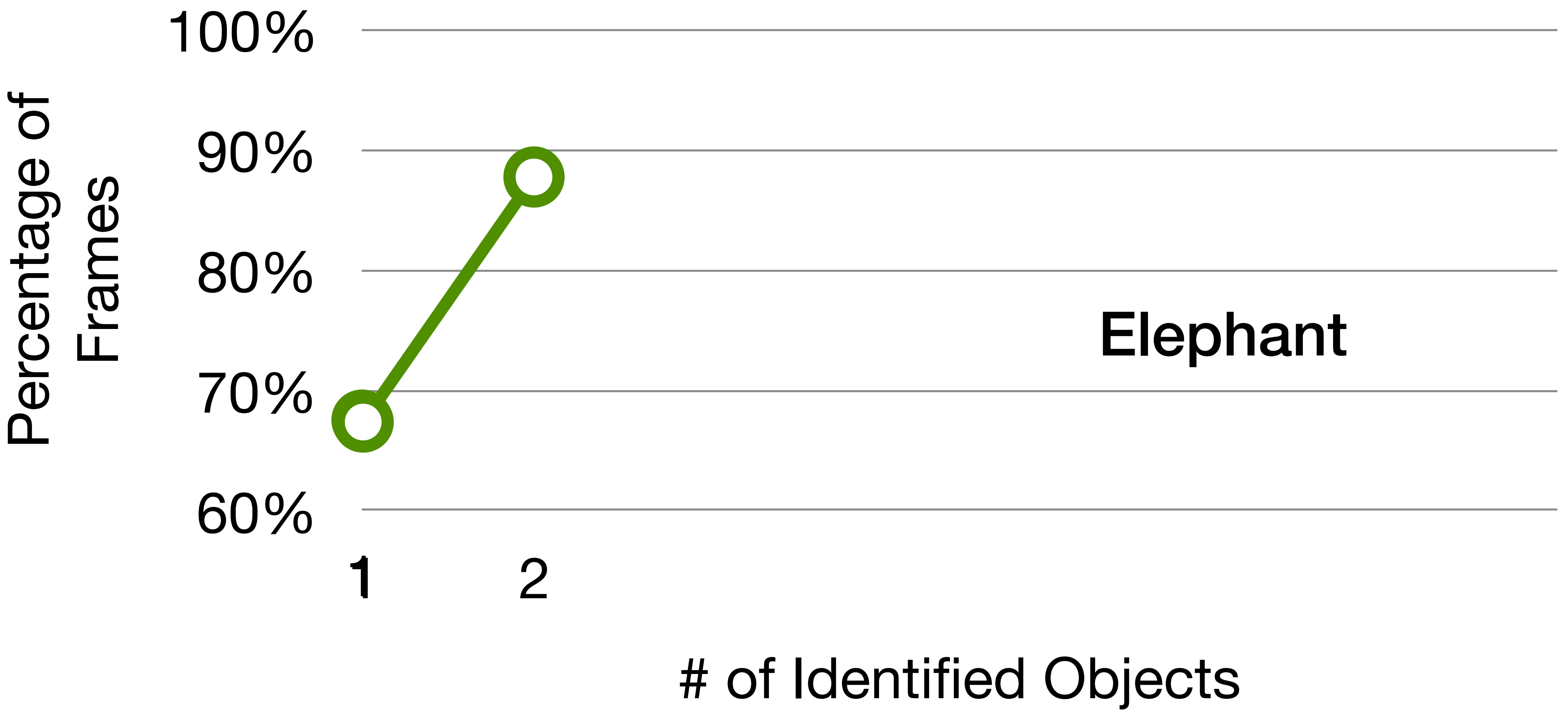
Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects



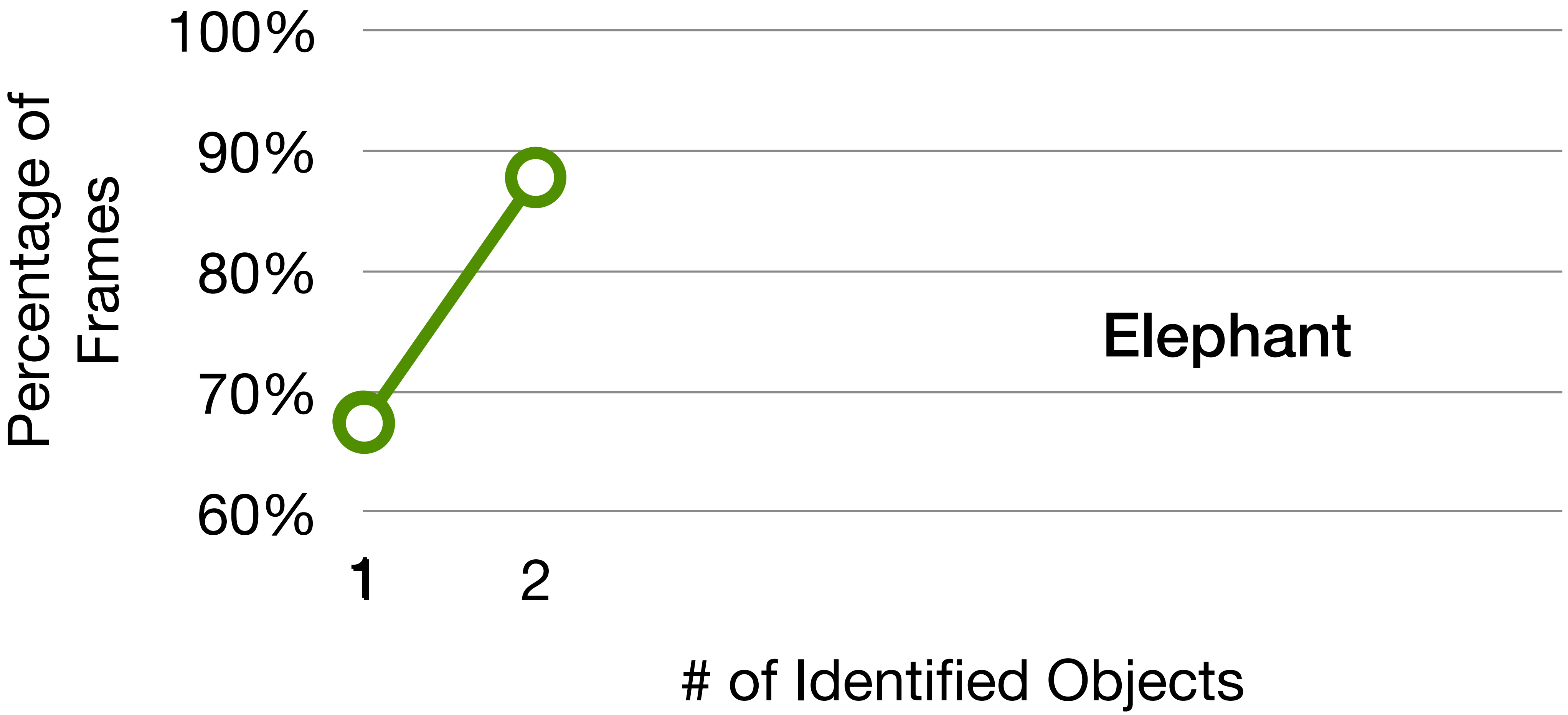
Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects



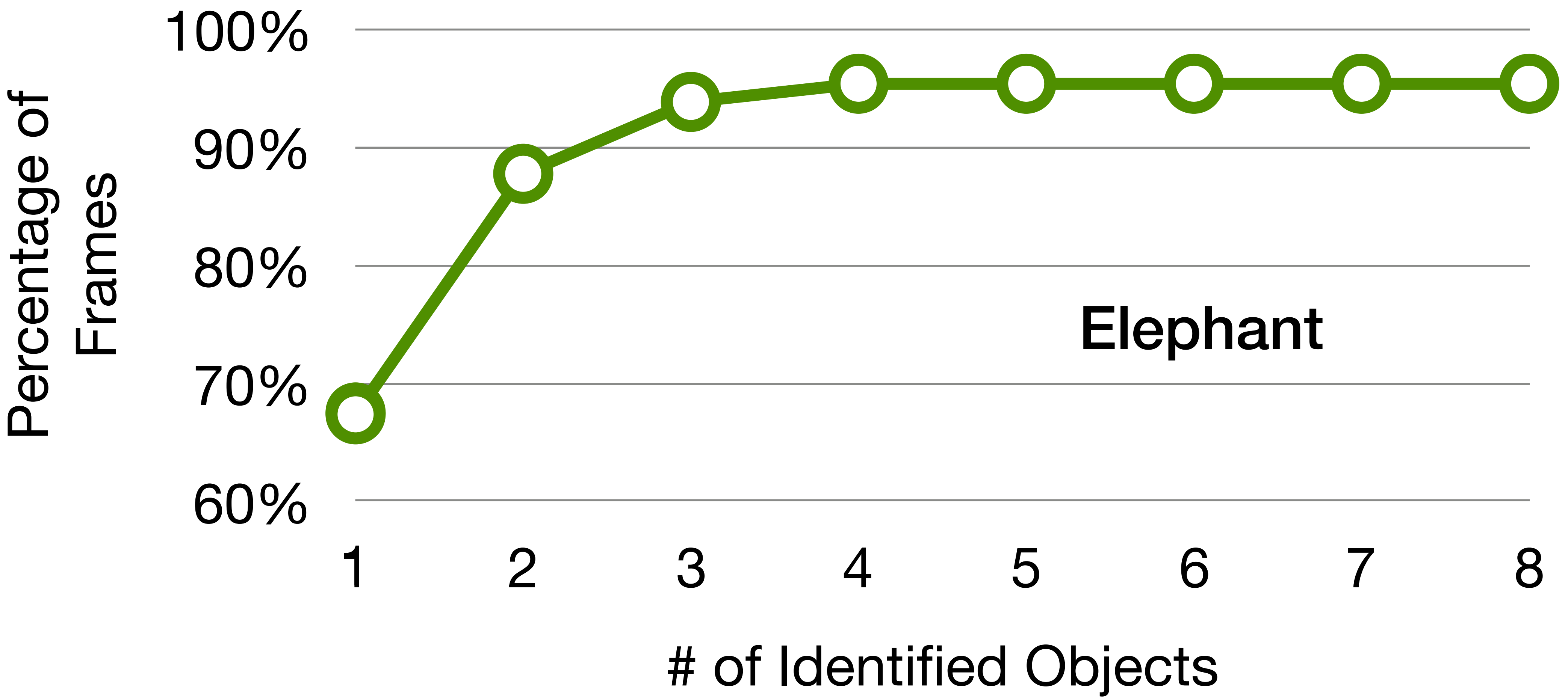
Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area



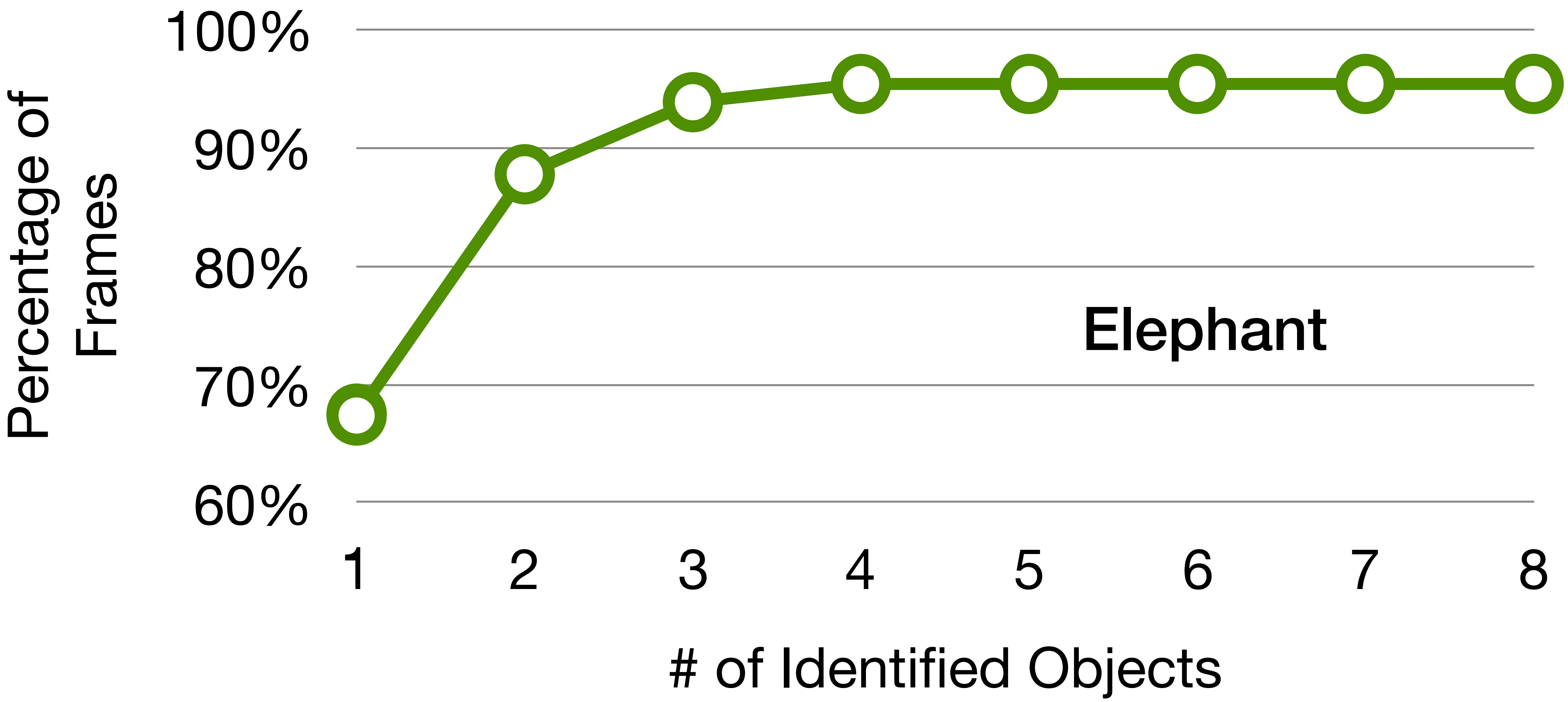
Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area



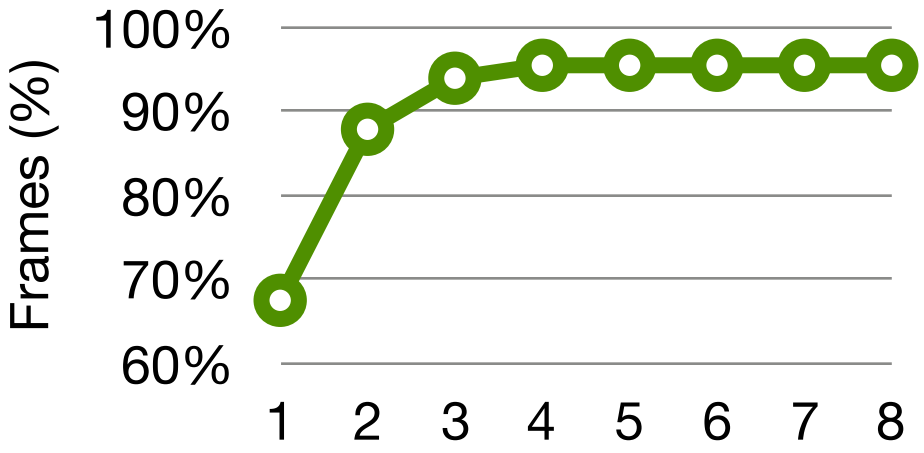
Leveraging Video Semantics Information

- Observations:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



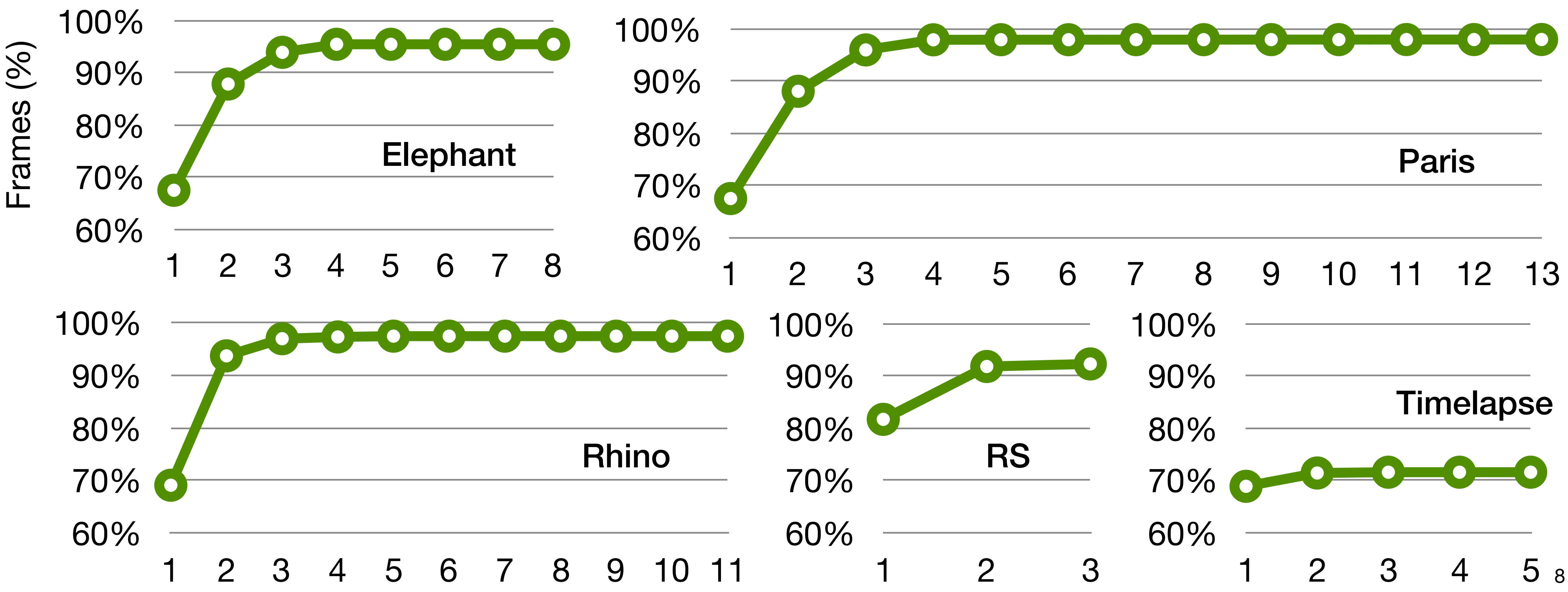
Leveraging Video Semantics Information

- Observation:
  - User tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



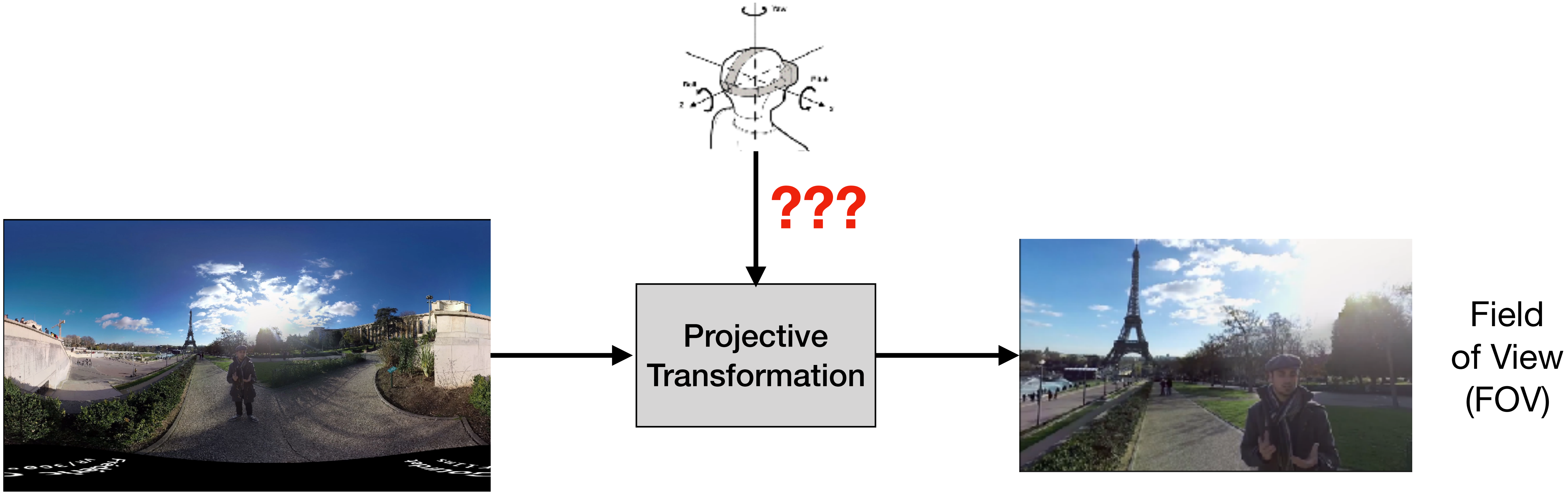
Leveraging Video Semantics Information

- Observation:
  - User tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



Leveraging Video Semantics Information

- **Observation:**
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



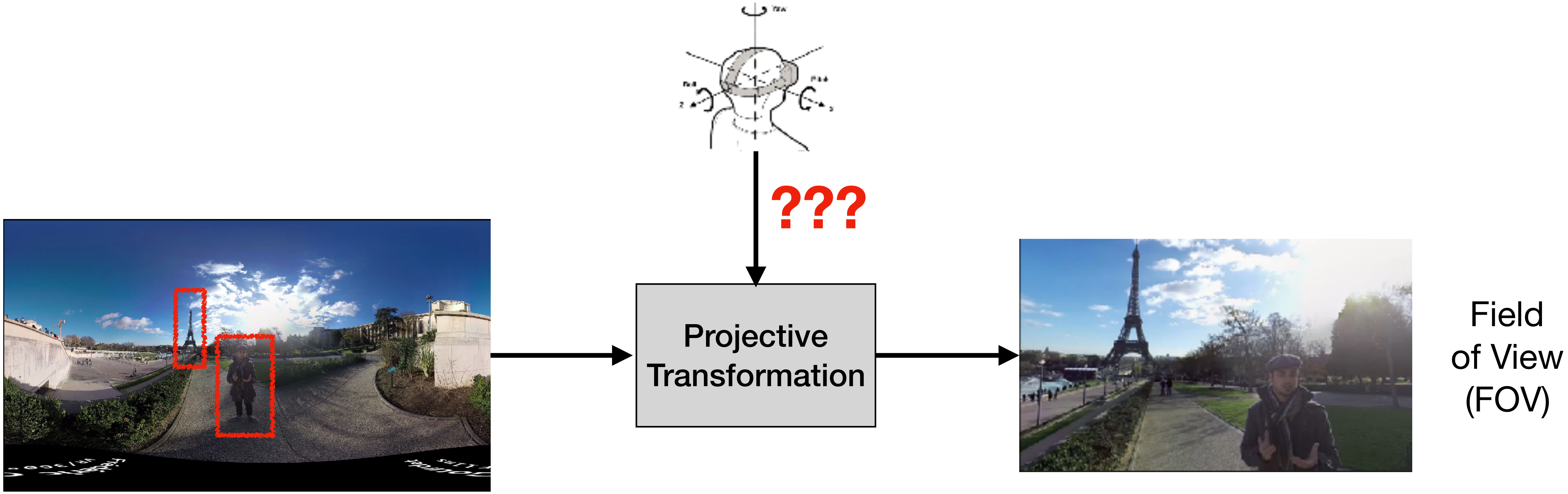
???

Projective Transformation

Field of View (FOV)

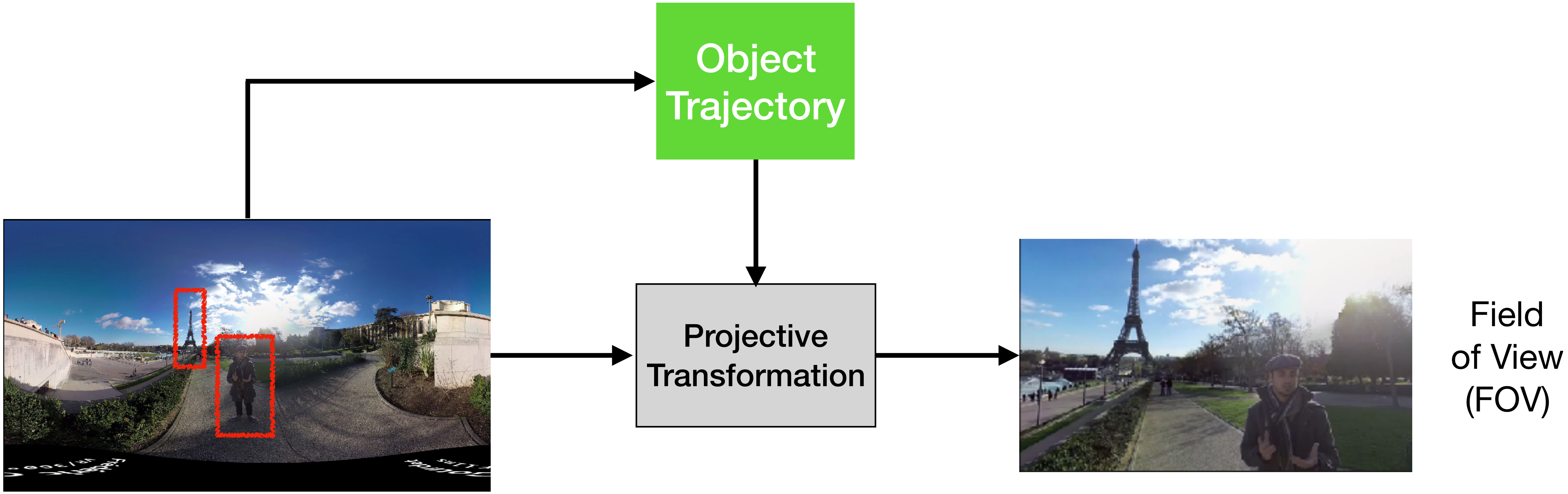
Leveraging Video Semantics Information

- Observation:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



Leveraging Video Semantics Information

- Observation:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction



Object Trajectory

Projective Transformation

Field of View (FOV)

Leveraging Video Semantics Information

- Observation:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction

- Advantage:
  - Client device avoids power-hungry PT operation

Leveraging Video Semantics Information

- Observation:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction

- Advantage:
  - Client device avoids power-hungry PT operation
  - Object trajectory can be generated offline once

Leveraging Video Semantics Information

- Observation:
  - Users tend to track objects
  - Object trajectory accurately predicts user viewing area
  - Tracking clustered objects result in more accurate prediction

- Advantage:
  - Client device avoids power-hungry PT operation
  - Object trajectory can be generated offline once
  - Object trajectory can be reused for multiple times

SAS Architecture Overview

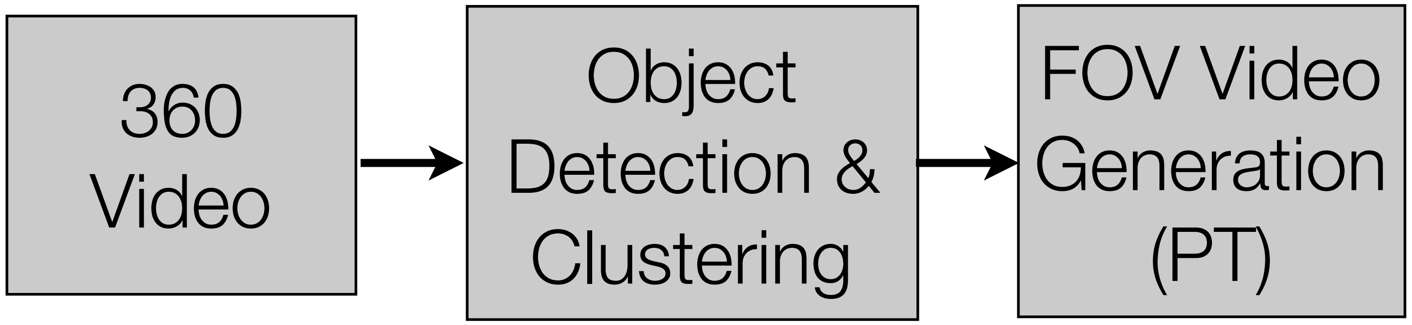
Server

360 Video

Client

SAS Architecture Overview

Server



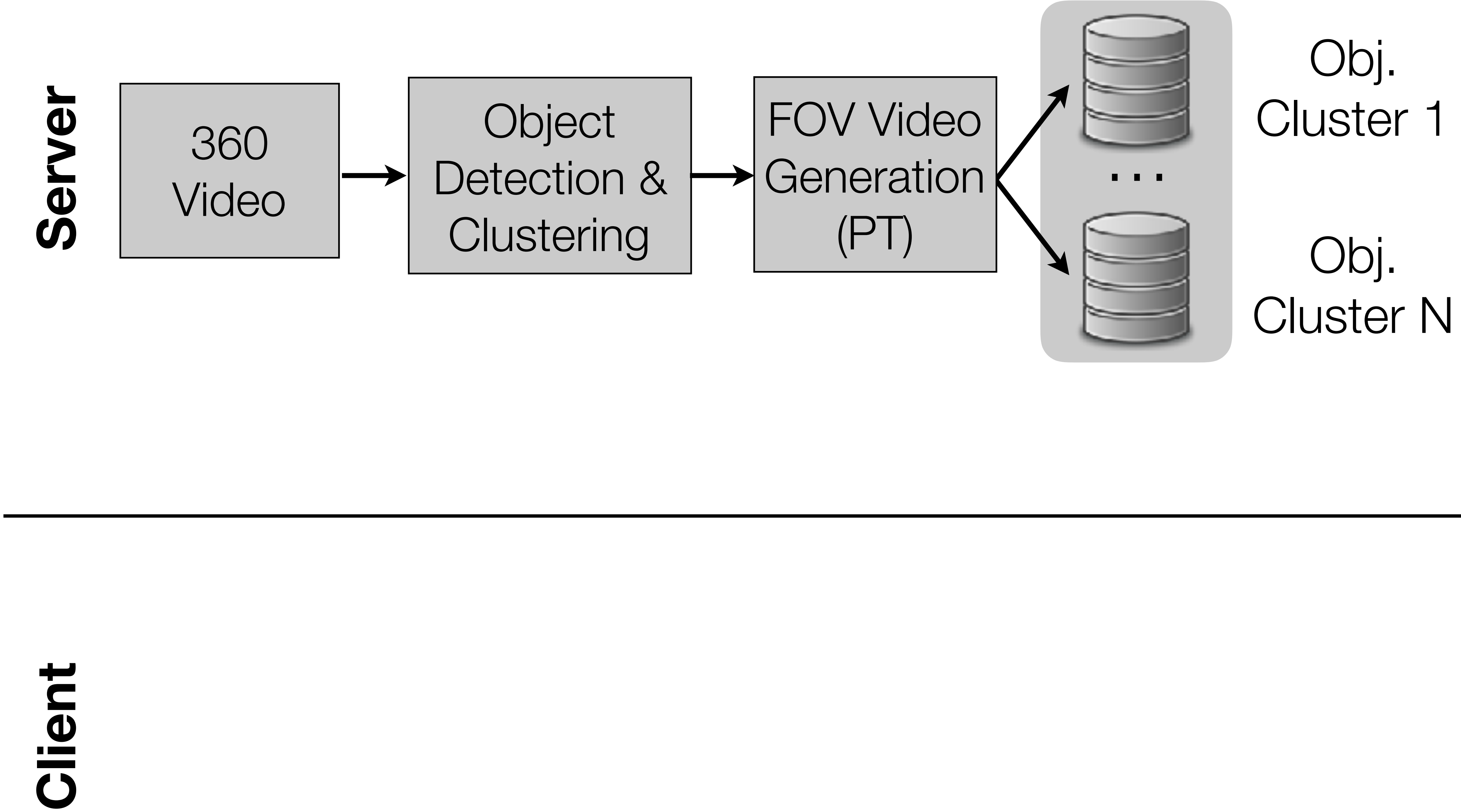
360 Video

Object Detection & Clustering

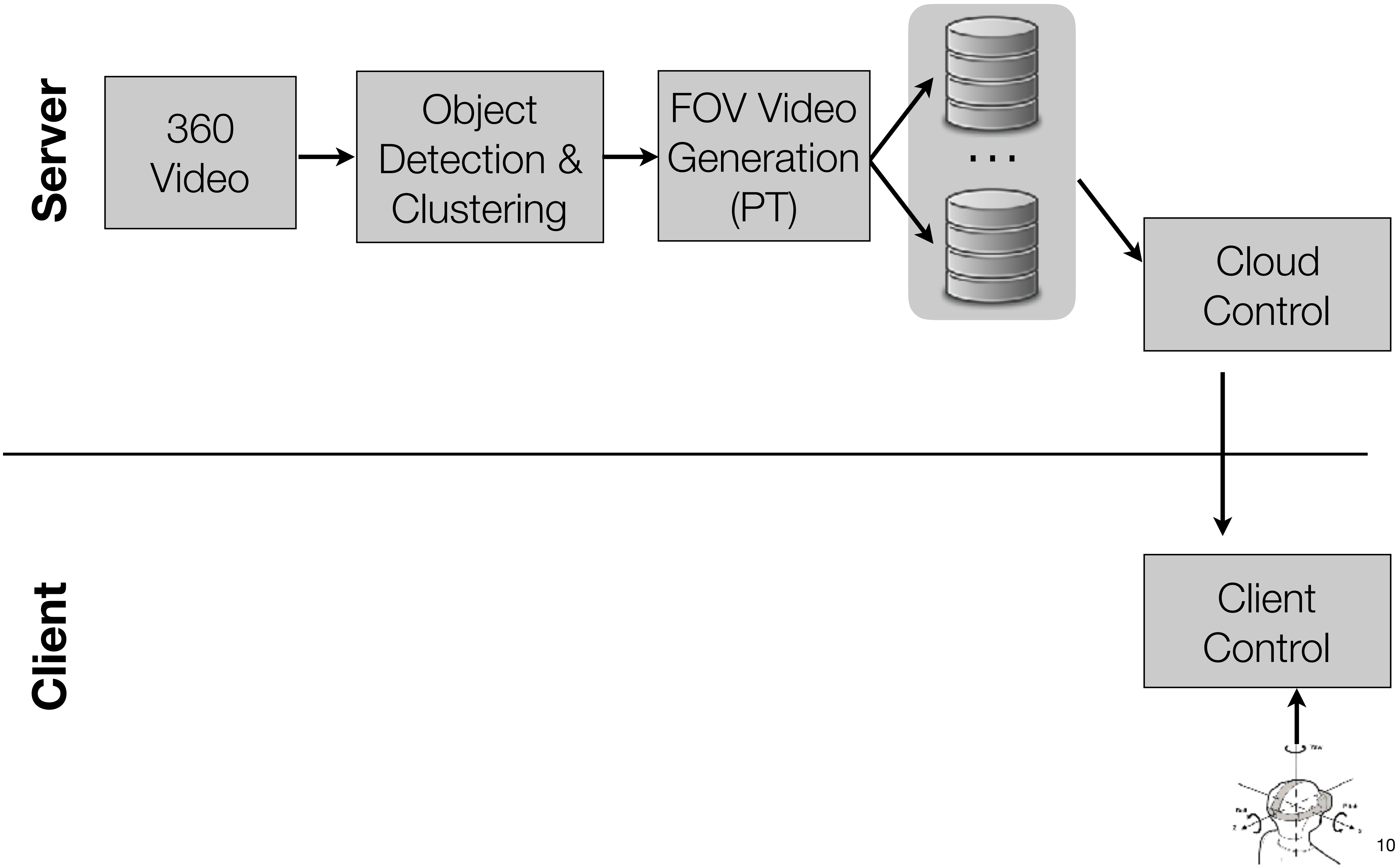
FOV Video Generation (PT)

Client

SAS Architecture Overview



SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

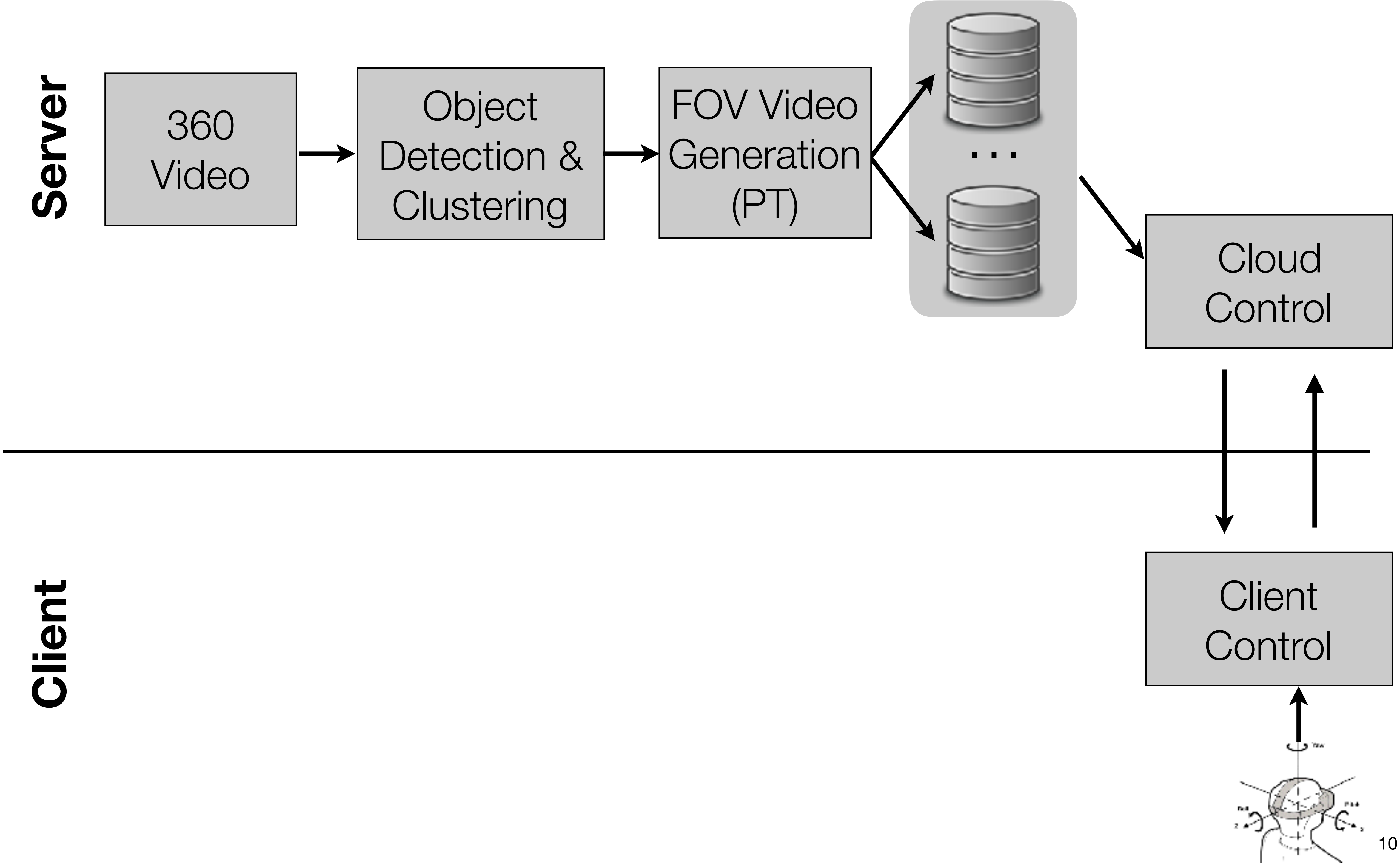
…

Cloud Control

Client

Client Control

SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

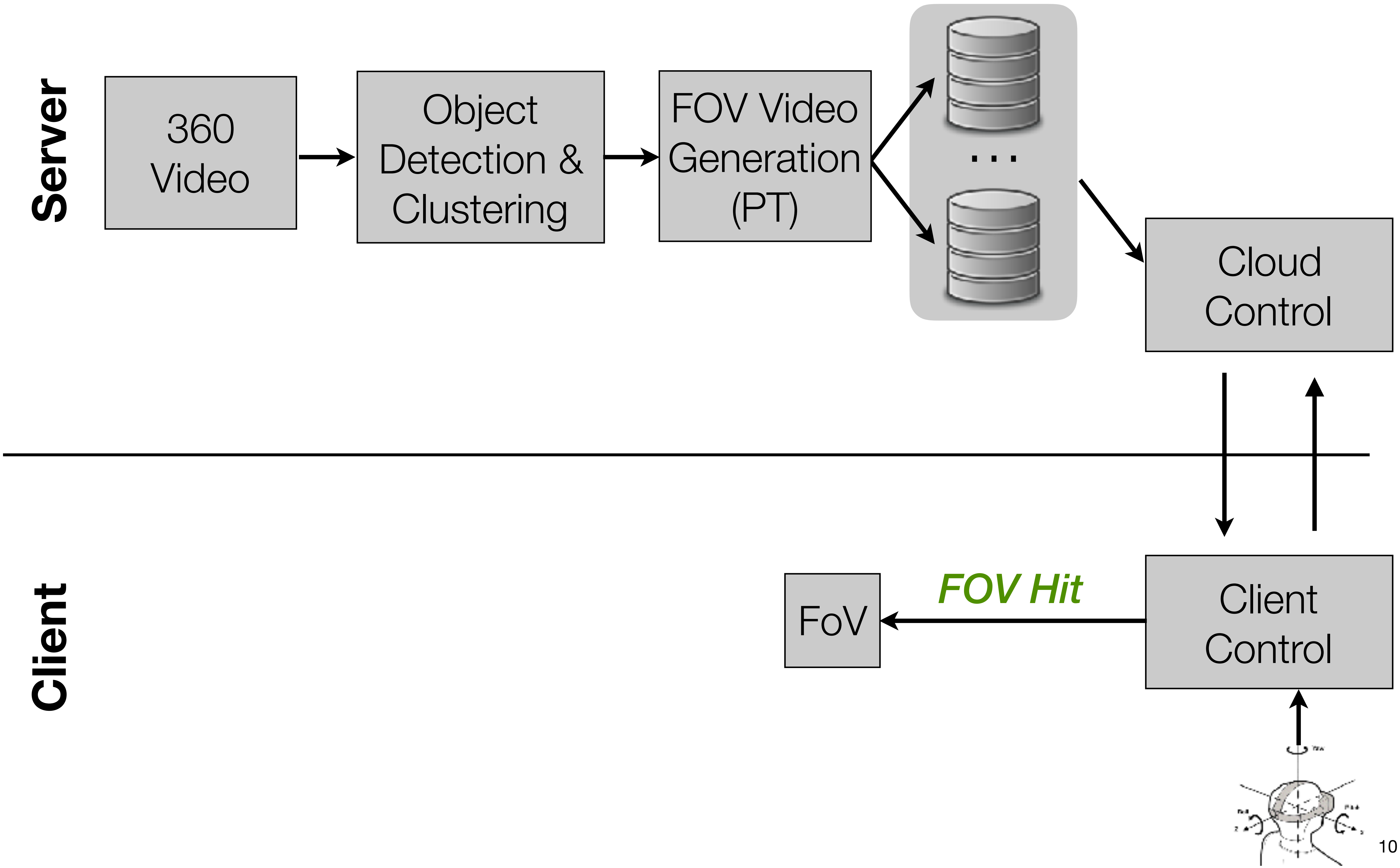
...

Cloud Control

Client

Client Control

SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

...

Cloud Control

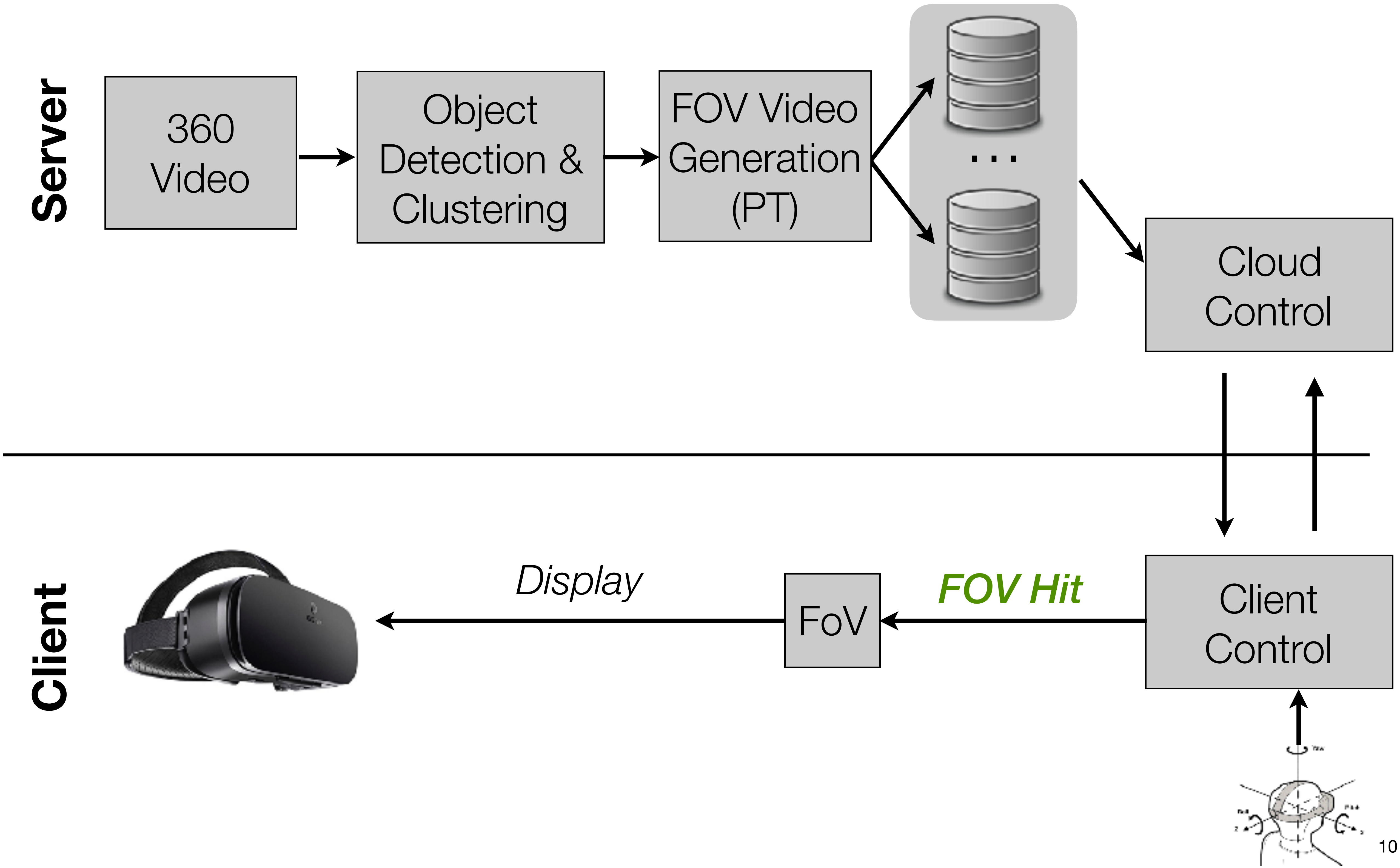
Client

FoV

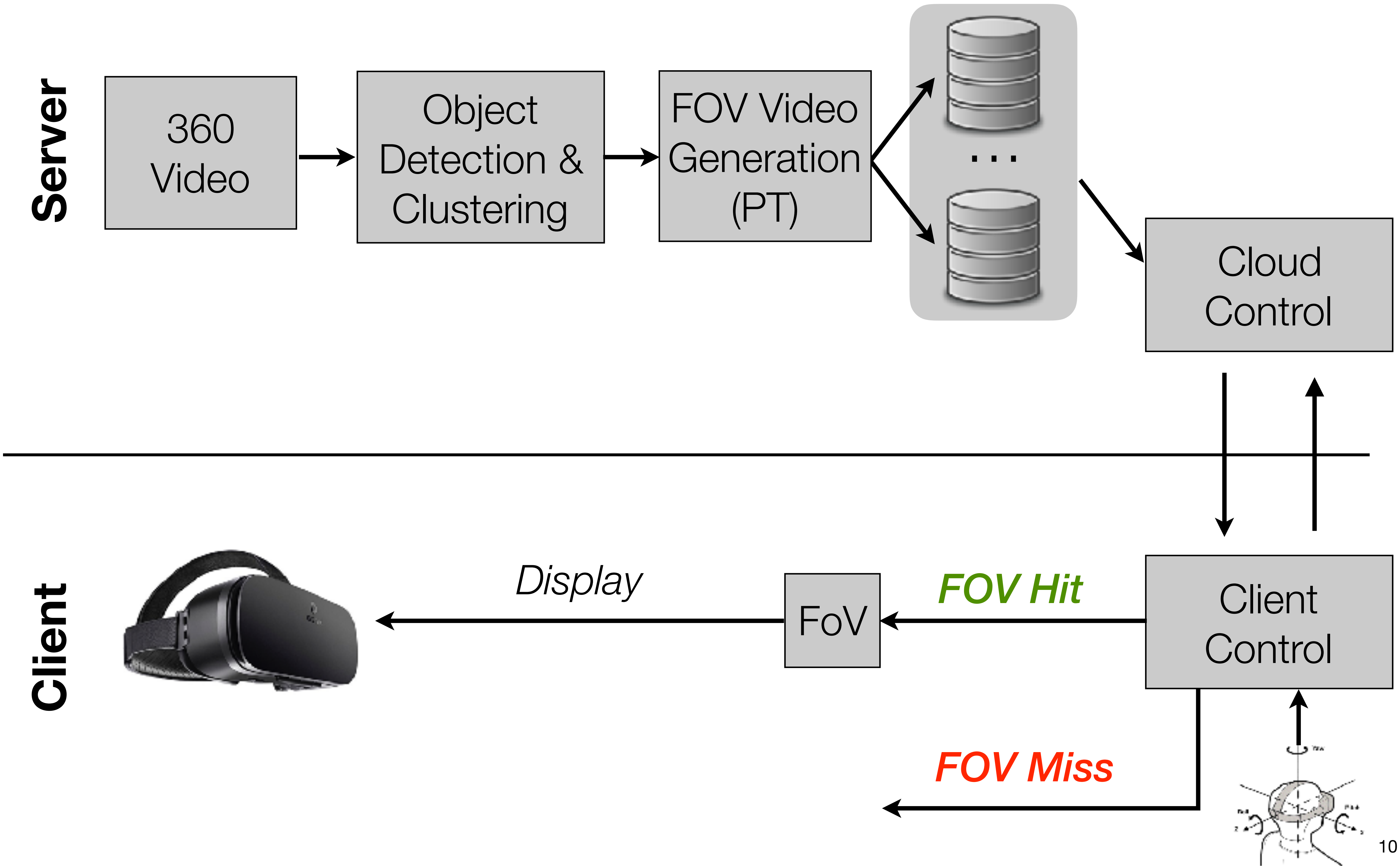
FOV Hit

Client Control

SAS Architecture Overview



SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

. . .

Cloud Control

Client

Display

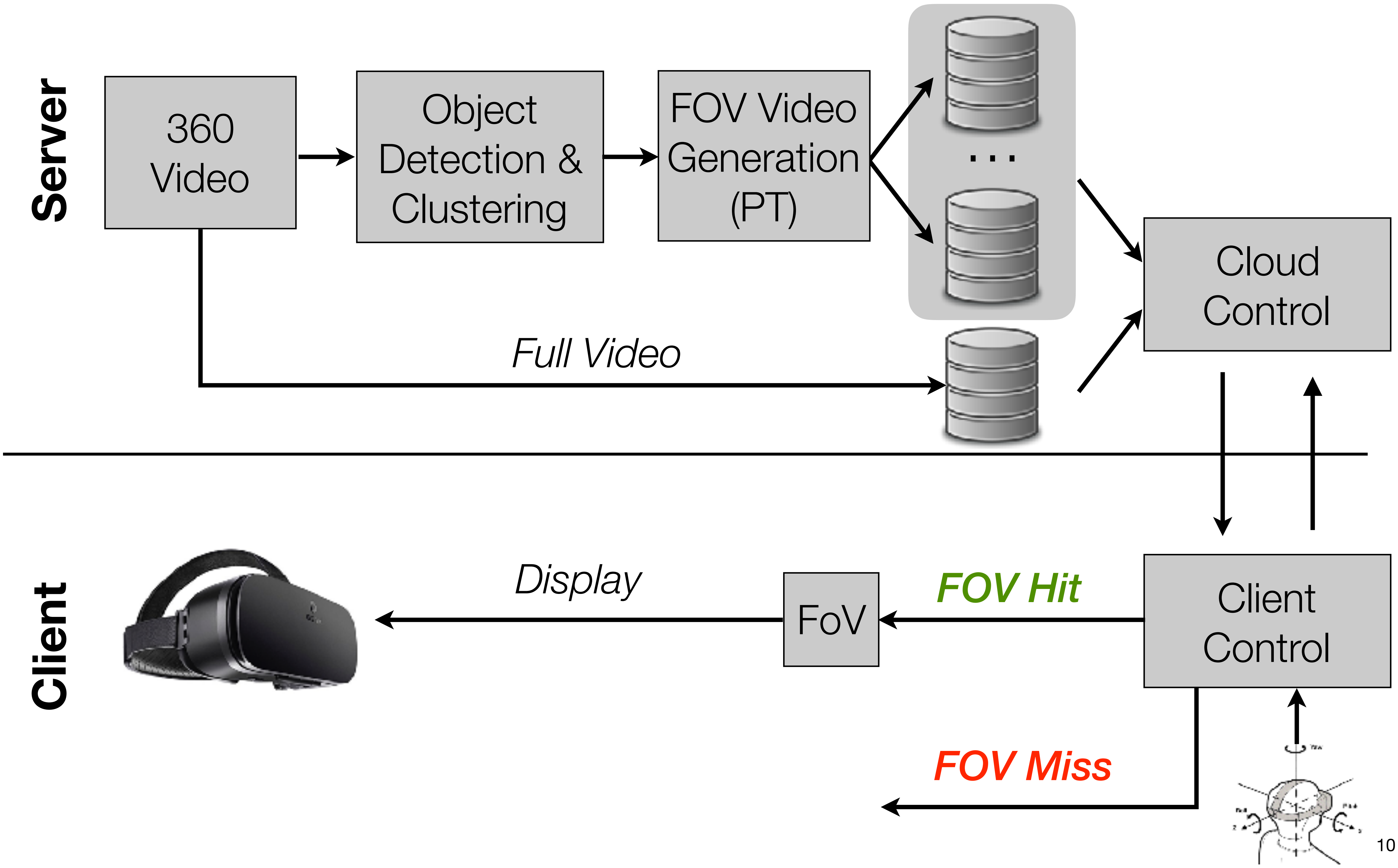
FoV

FOV Hit

Client Control

FOV Miss

SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

...

Full Video

Cloud Control

Client

Display

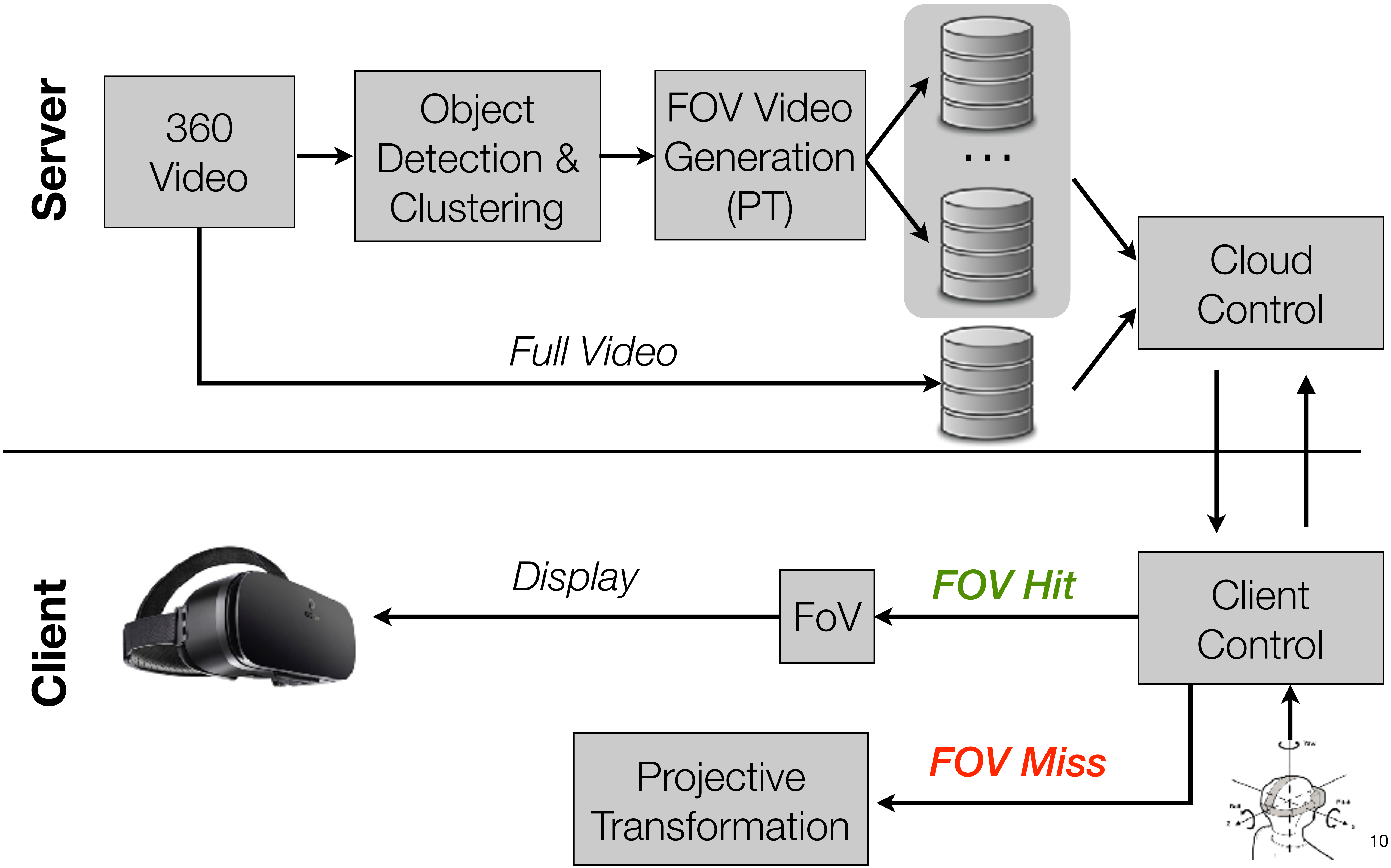
FoV

FOV Hit

Client Control

FOV Miss

SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

...

Full Video

Cloud Control

Client

Display

FoV

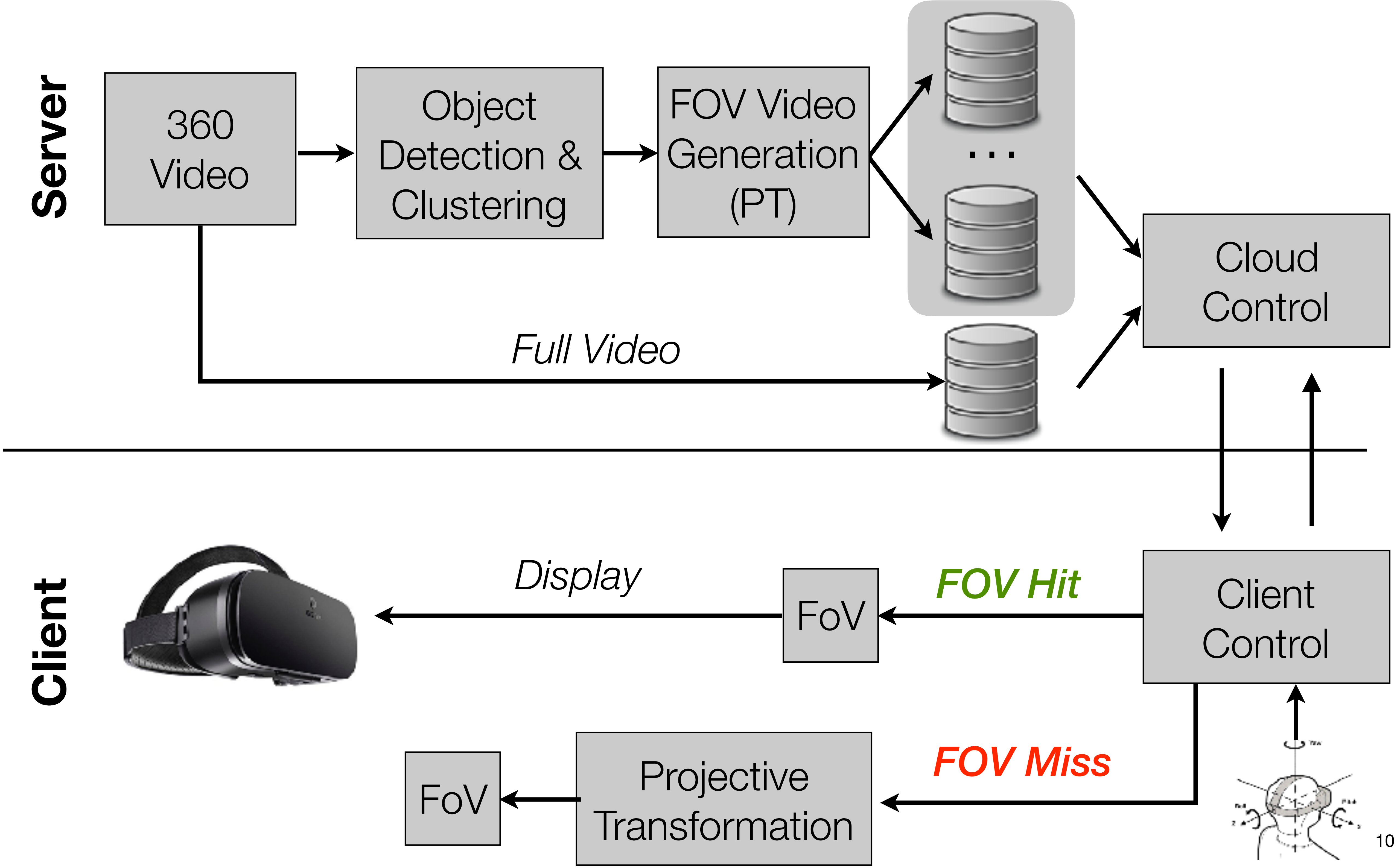
FOV Hit

Client Control

Projective Transformation

FOV Miss

SAS Architecture Overview



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

. . .

Full Video

Cloud Control

Client

Display

FoV

FOV Hit

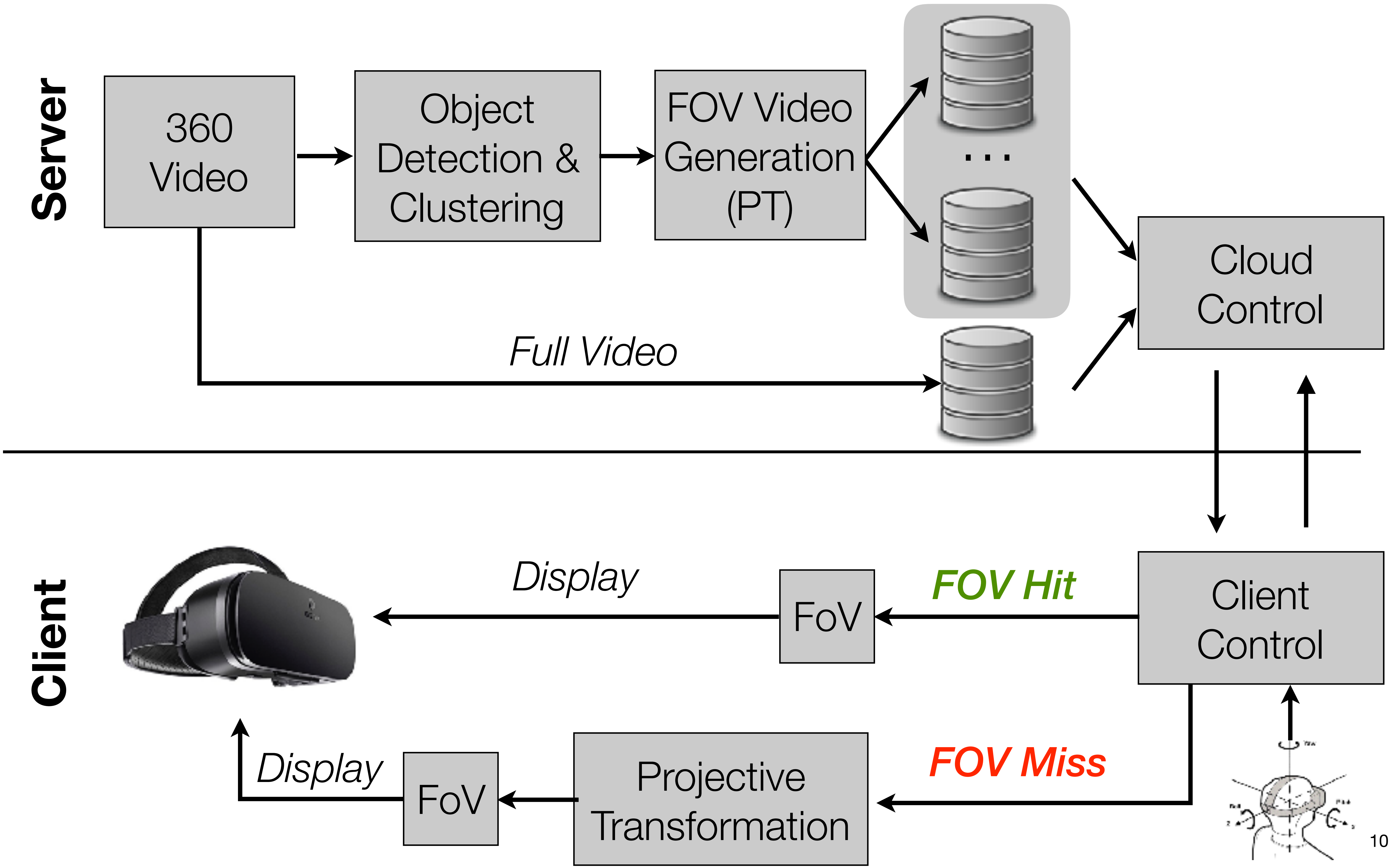
Client Control

FoV

Projective Transformation

FOV Miss

SAS Architecture Overview



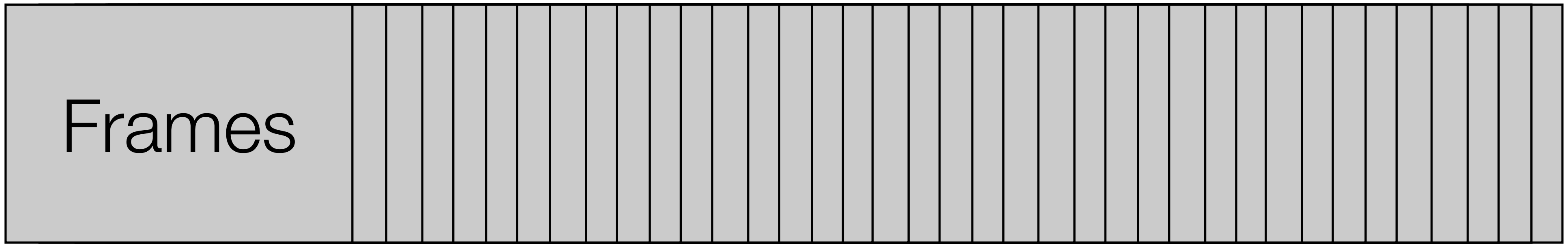
SAS: Design Decisions/Optimizations

SAS: Design Decisions/Optimizations

- Naive implementation

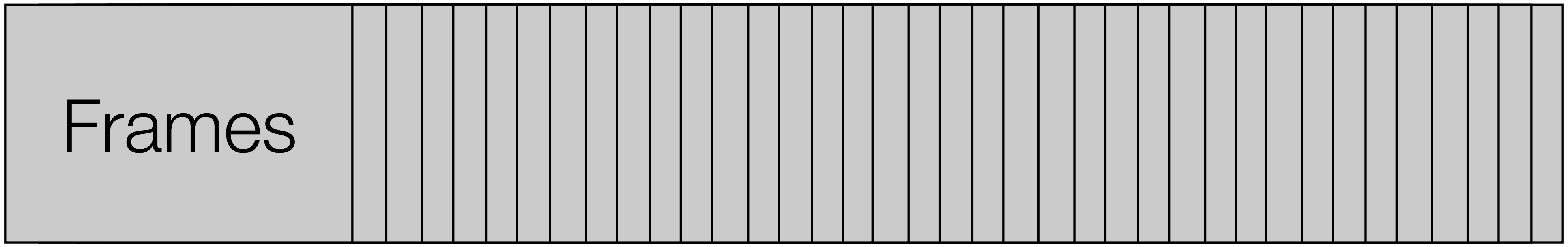
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video



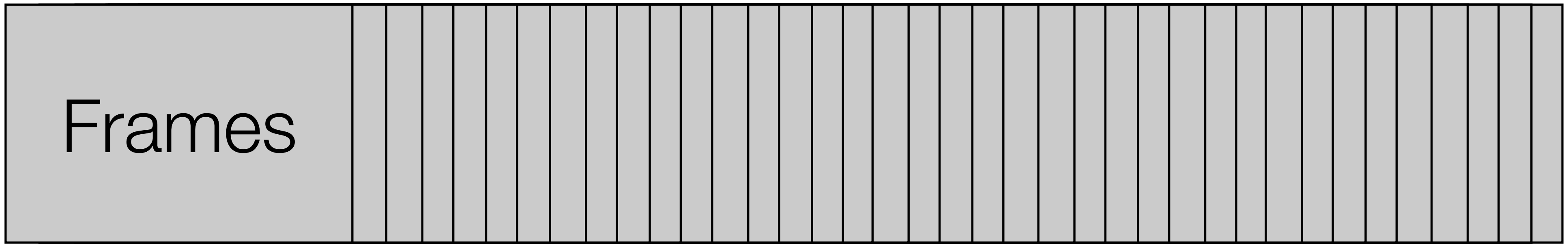
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate



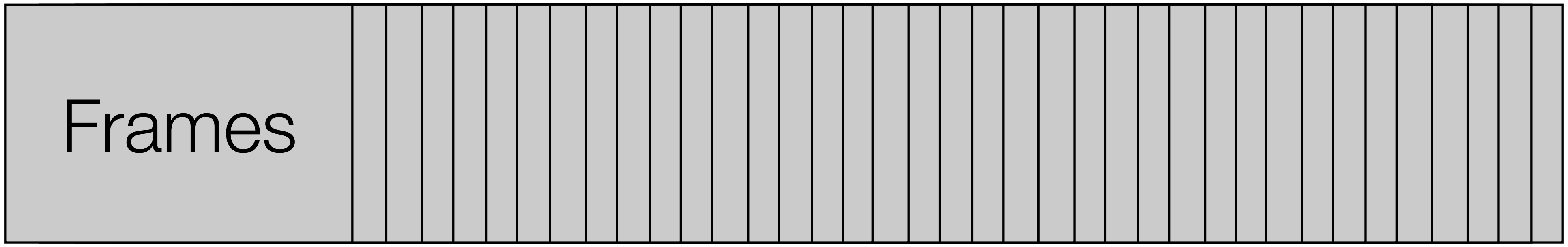
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate
  - FOV miss → Re-streaming entire original video



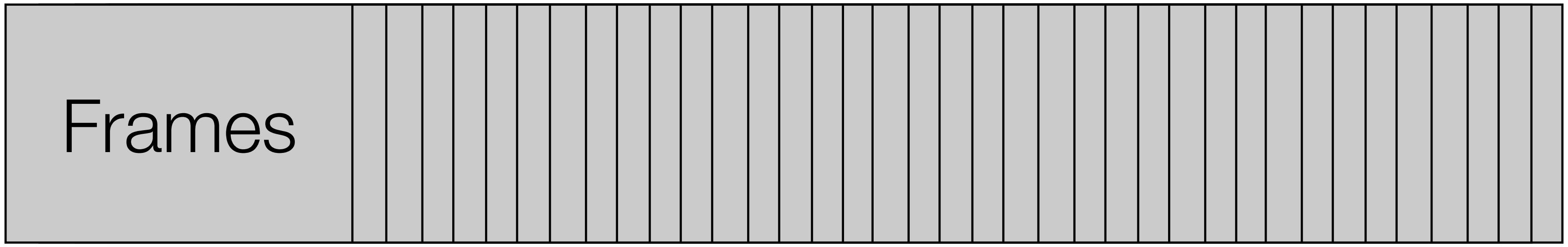
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate
  - FOV miss → Re-streaming entire original video
- Recovery Penalty



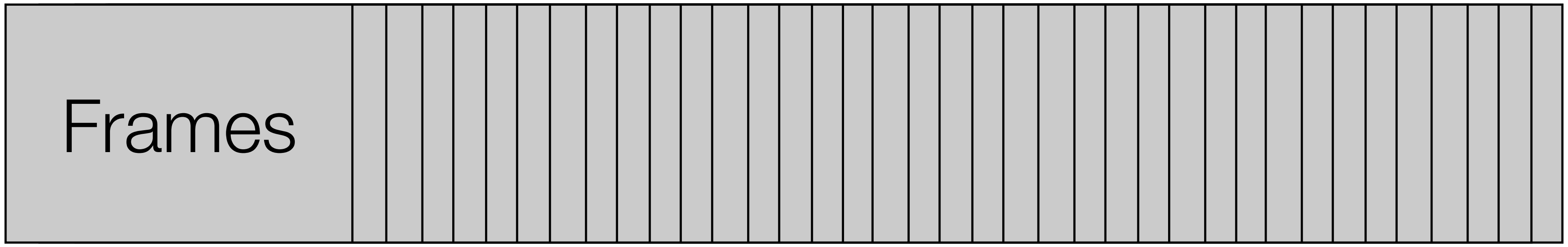
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate
  - FOV miss → Re-streaming entire original video
- Recovery Penalty
  - Streaming time



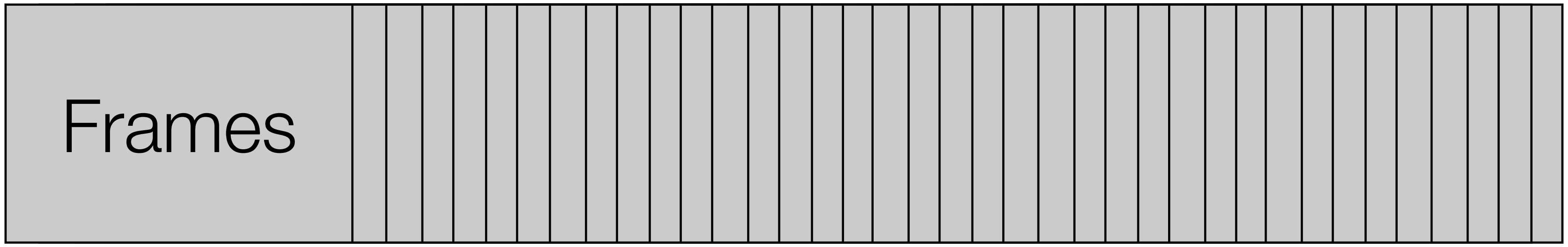
SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate
  - FOV miss → Re-streaming entire original video
- Recovery Penalty
  - Streaming time
  - Bandwidth usage

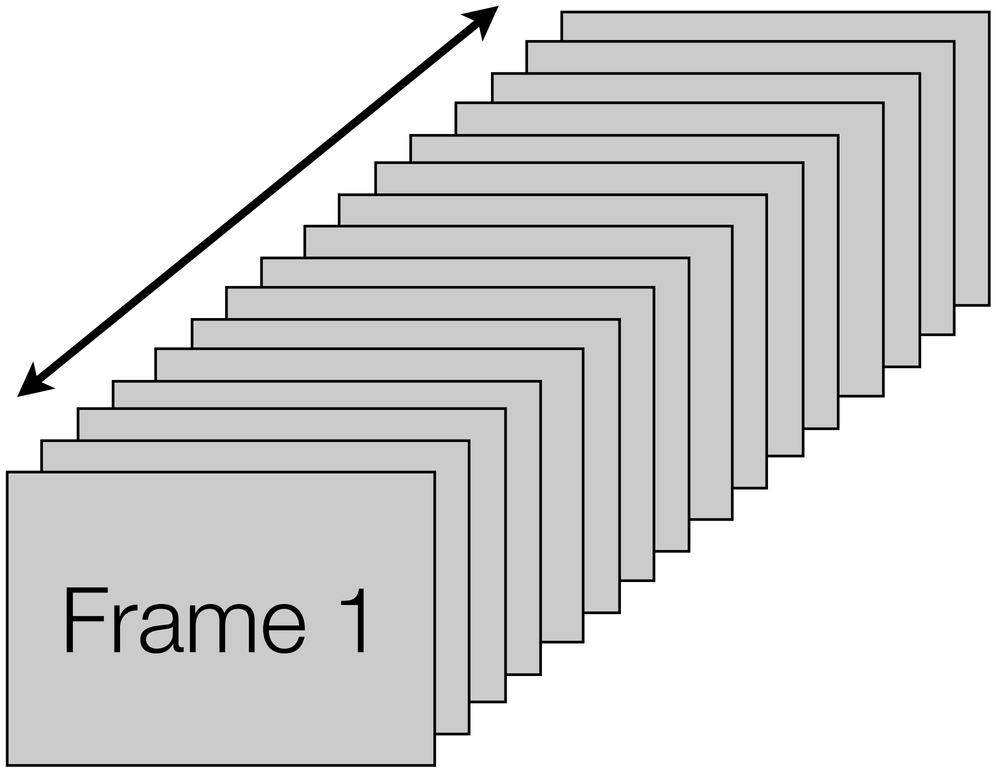


SAS: Design Decisions/Optimizations

- Naive implementation
  - Generate FOV videos for entire video
  - High video compression rate
  - FOV miss → Re-streaming entire original video
- Recovery Penalty
  - Streaming time
  - Bandwidth usage
- Trade-off: Recovery penalty vs. Video compression rate

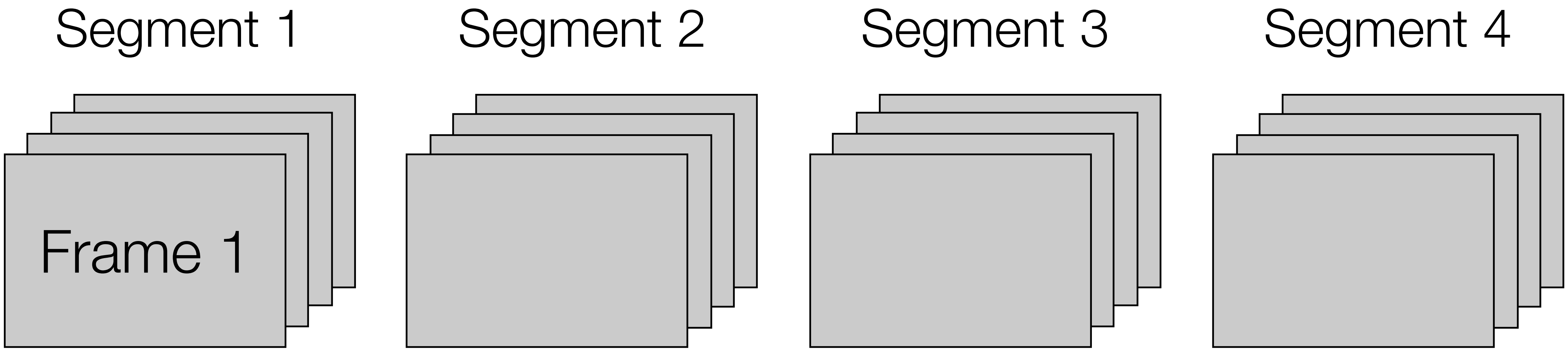


Temporal Segmentation



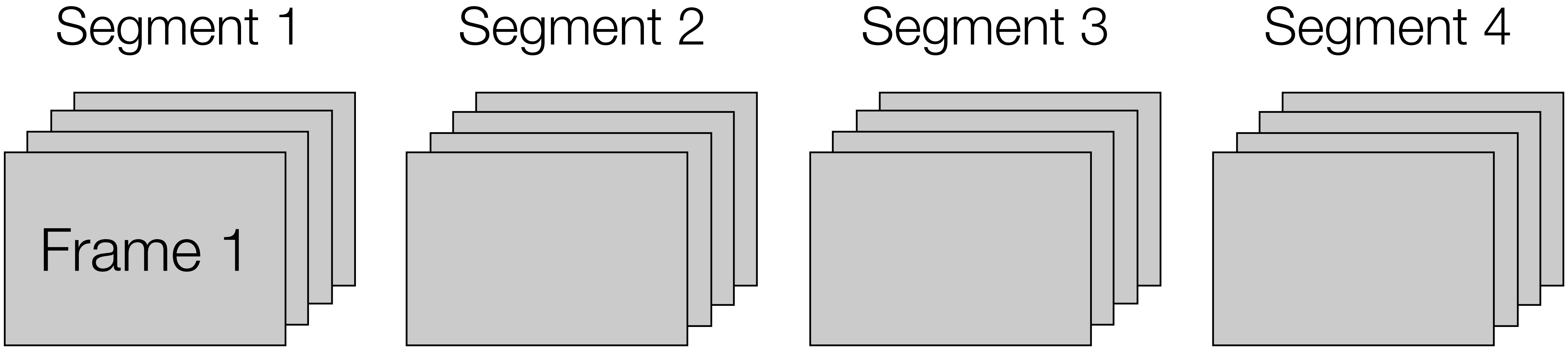
Temporal Segmentation

- Split video into several temporal segments



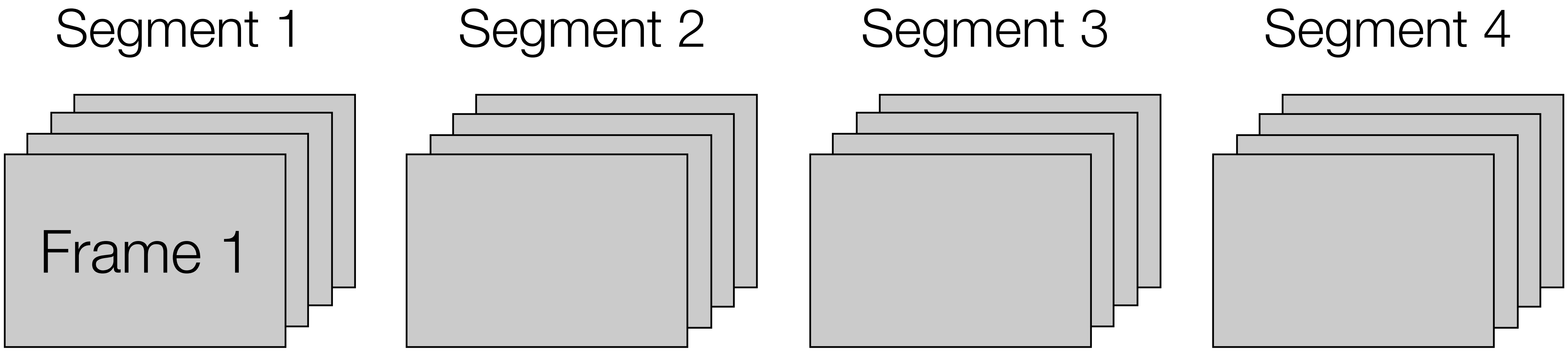
Temporal Segmentation

- Split video into several temporal segments
- Generate FOV videos for each segment



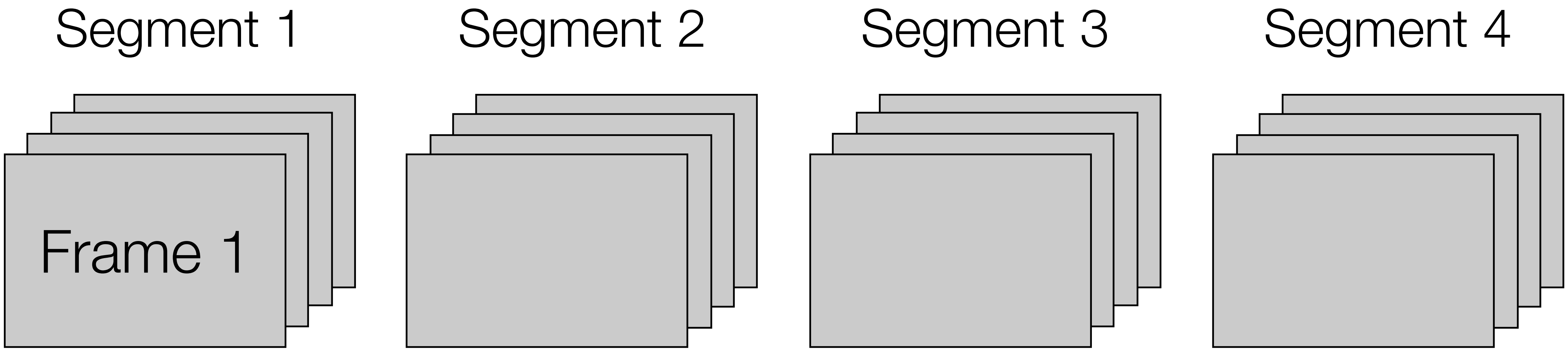
Temporal Segmentation

- Split video into several temporal segments
- Generate FOV videos for each segment
- Every FOV-miss only requests one segment



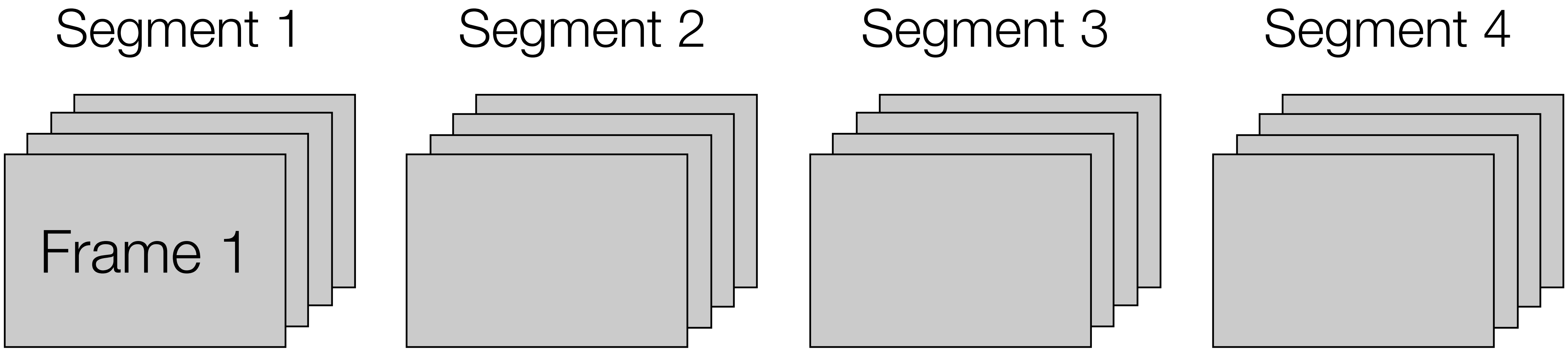
Temporal Segmentation

- Split video into several temporal segments
- Generate FOV videos for each segment
- Every FOV-miss only requests one segment
- To balance recovery penalty with video compression rate



Temporal Segmentation

- Split video into several temporal segments
- Generate FOV videos for each segment
- Every FOV-miss only requests one segment
- To balance recovery penalty with video compression rate
    - Design decision: 30 frames/segment



EVR

A Cloud-Client Collaborative VR Video System

Cloud

Semantic Aware Streaming

Client

Hardware Accelerated Rendering

Result

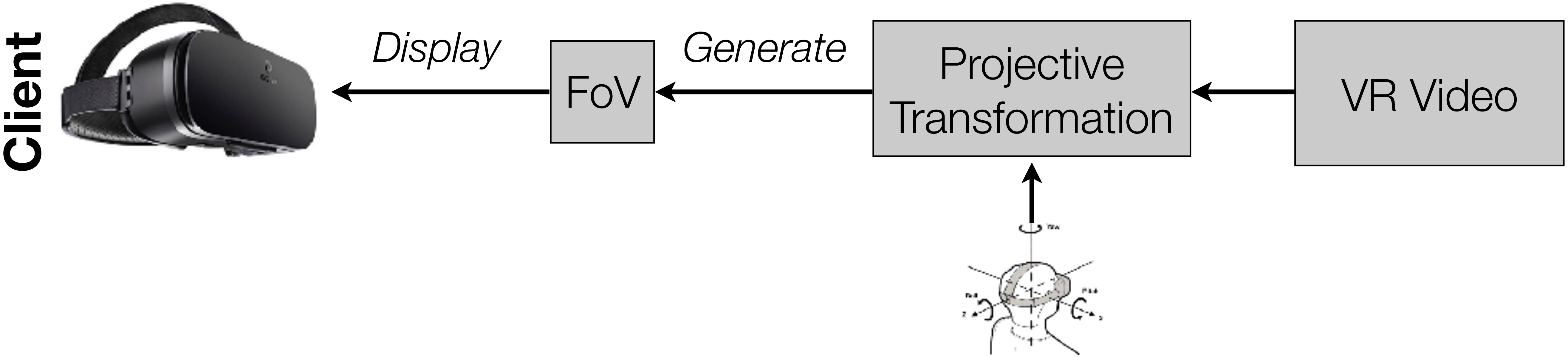
Up to 58% Energy Reduction & Only 1% Frame Rate Drops

EVR

A Cloud-Client Collaborative VR Video System

Cloud	Semantic Aware Streaming
Client	Hardware Accelerated Rendering
Result	Up to 58% Energy Reduction & Only 1% Frame Rate Drops

Current Hardware Usage



Client

Display

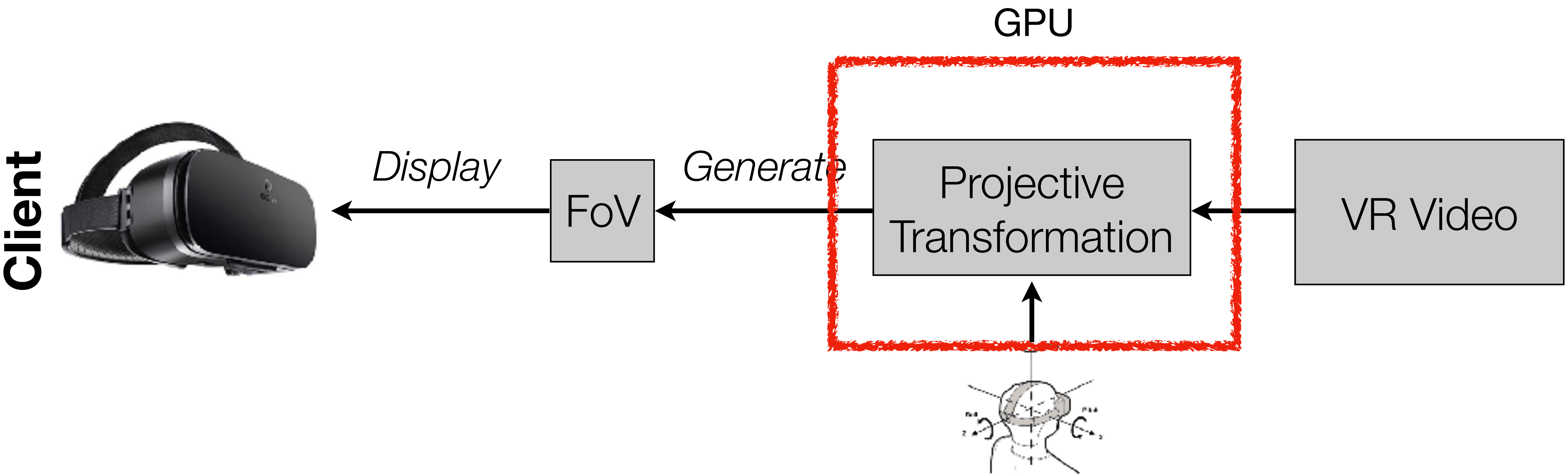
FoV

Generate

Projective Transformation

VR Video

Current Hardware Usage



GPU

Client

Display

FoV

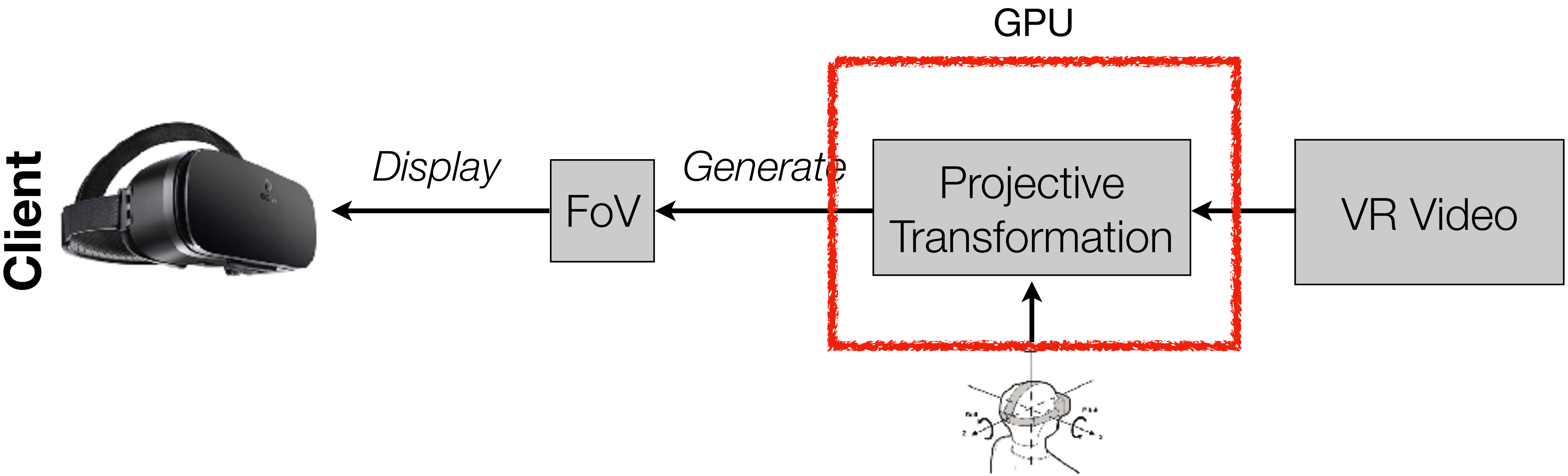
Generate

Projective Transformation

VR Video

Current Hardware Usage

- PT is treated as Texture Mapping



GPU

Client

Display

FoV

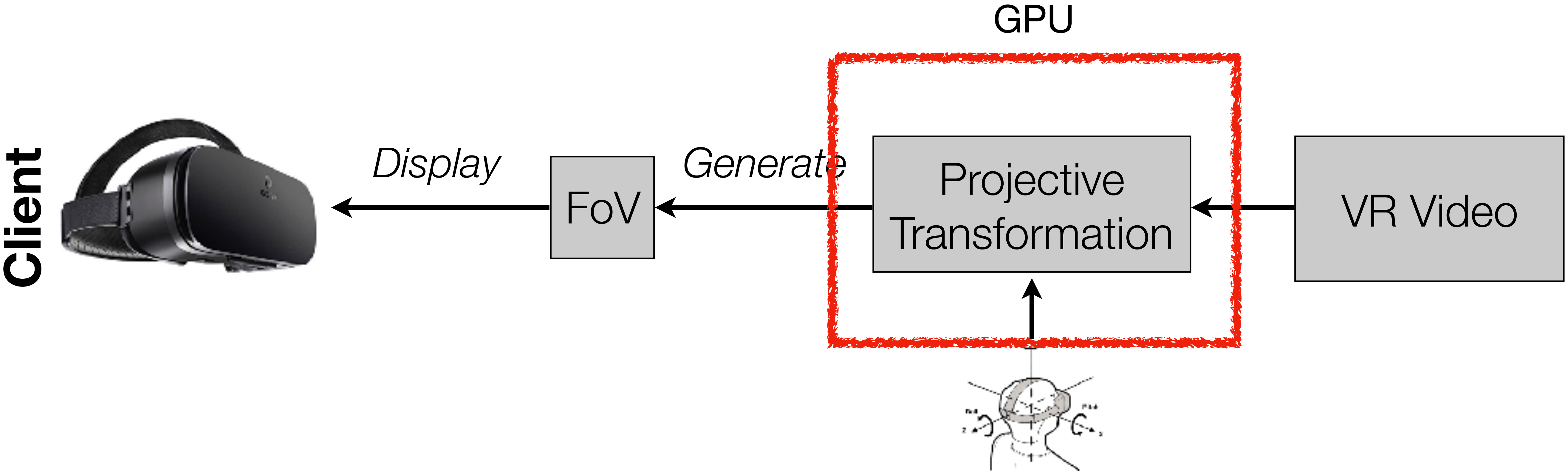
Generate

Projective Transformation

VR Video

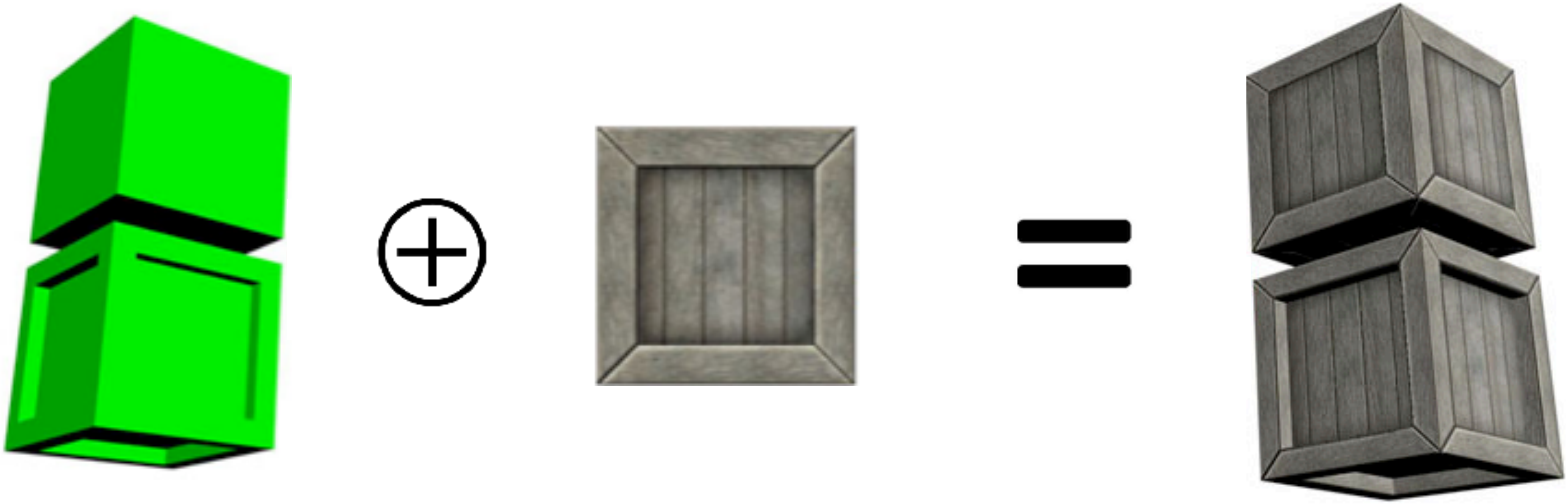
Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)



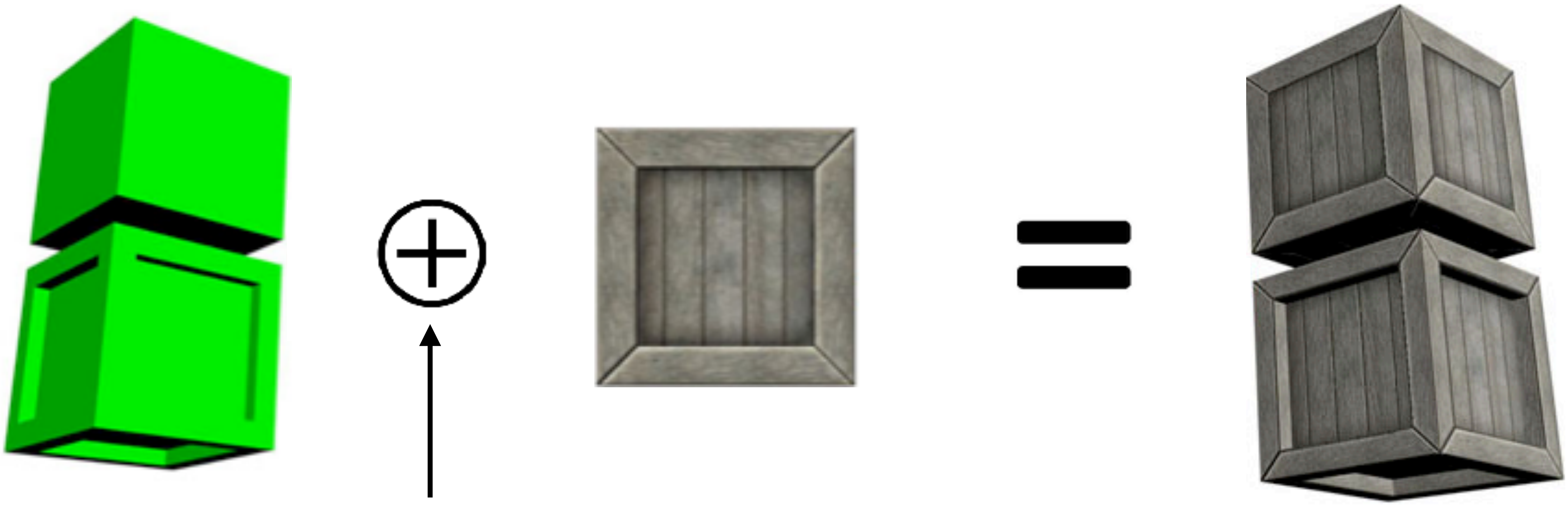
Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)



Current Hardware Usage

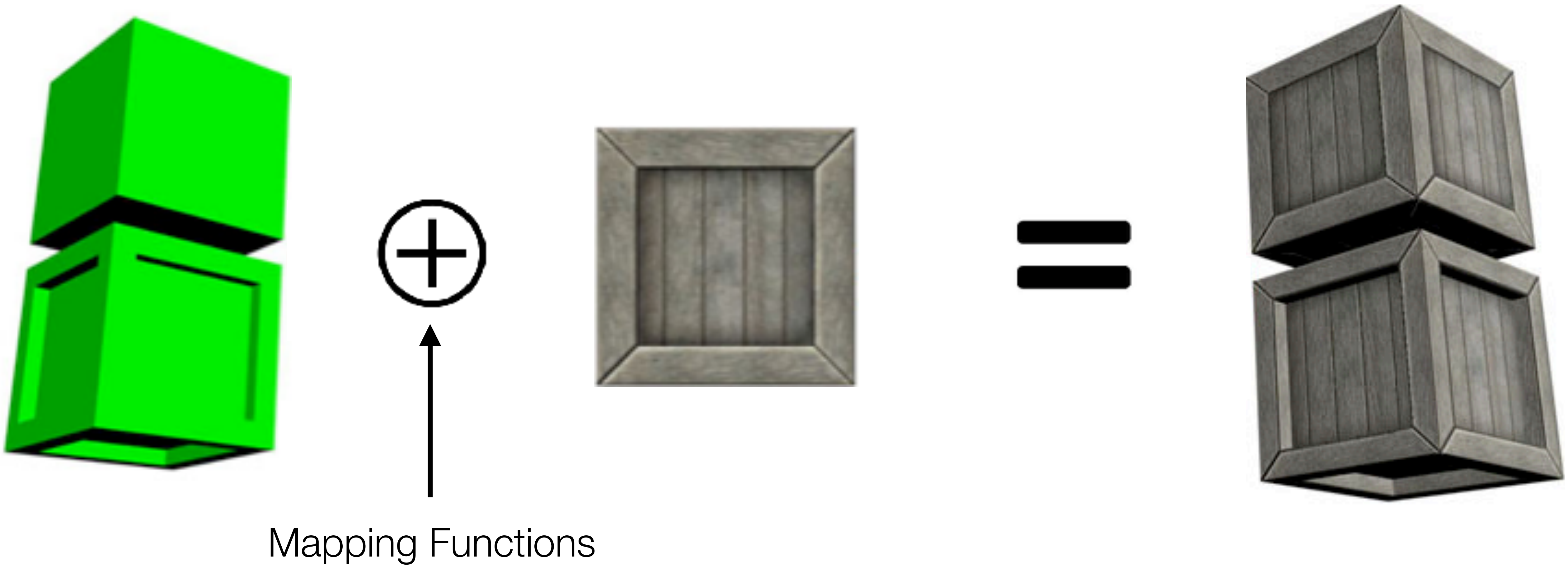
- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)



Mapping Functions

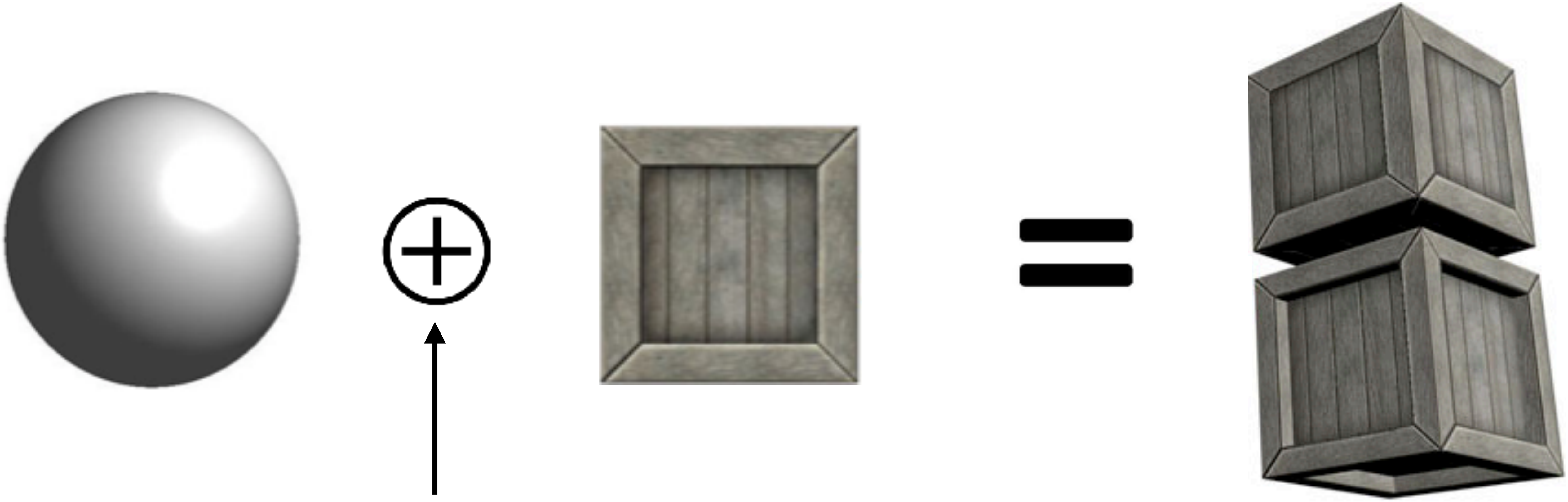
Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions



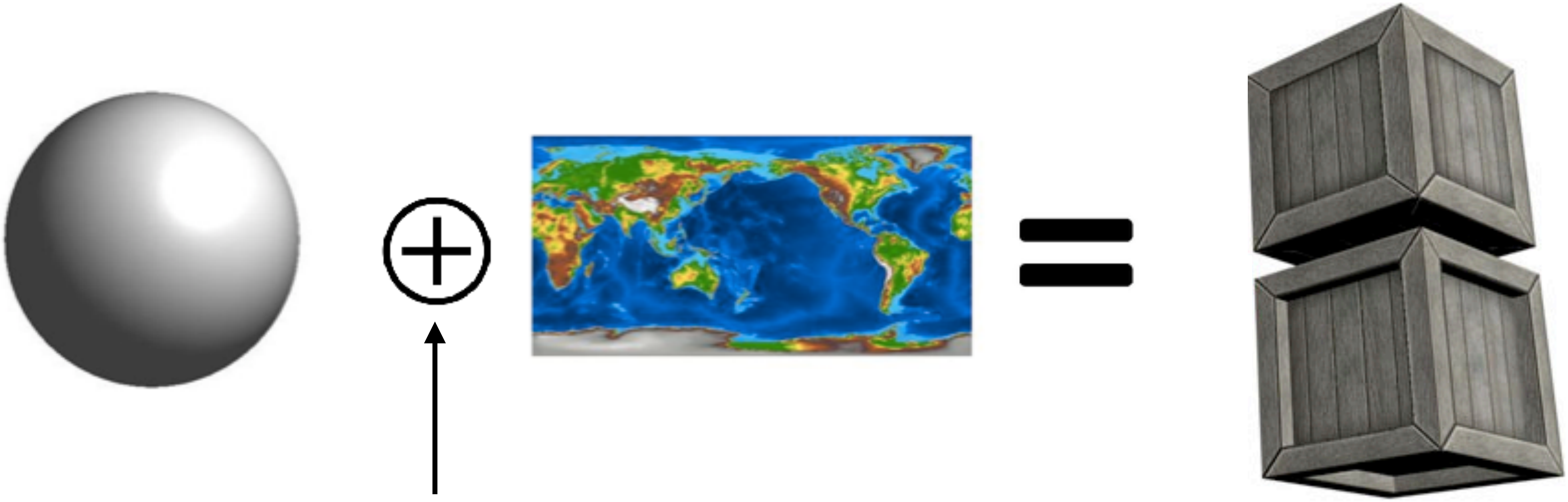
Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions



Current Hardware Usage

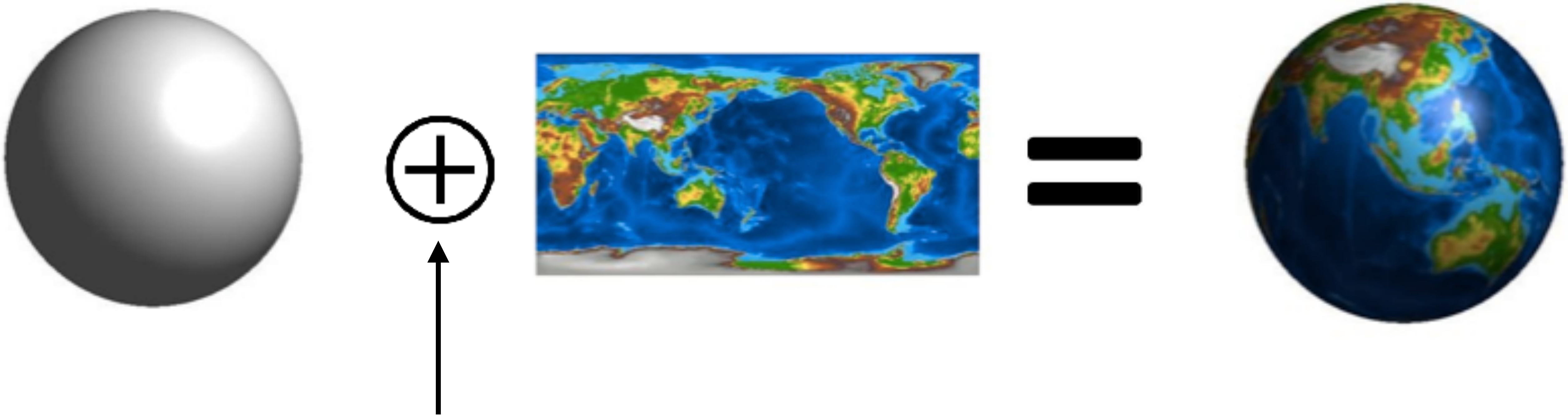
- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions



Mapping Functions

Current Hardware Usage

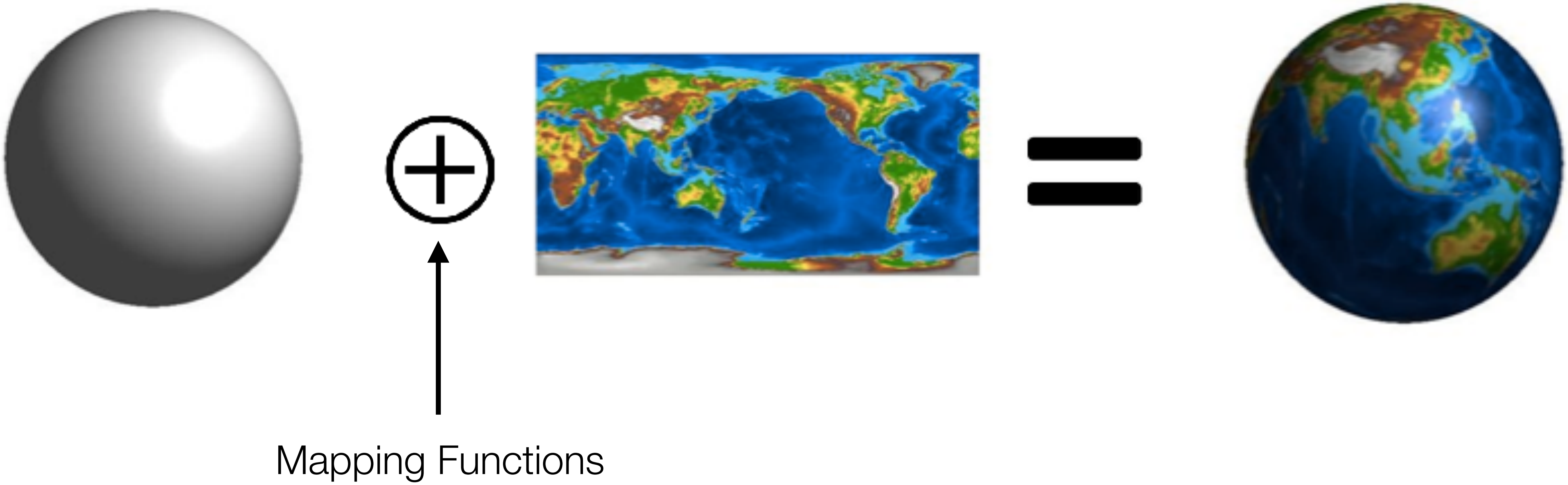
- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions



Mapping Functions

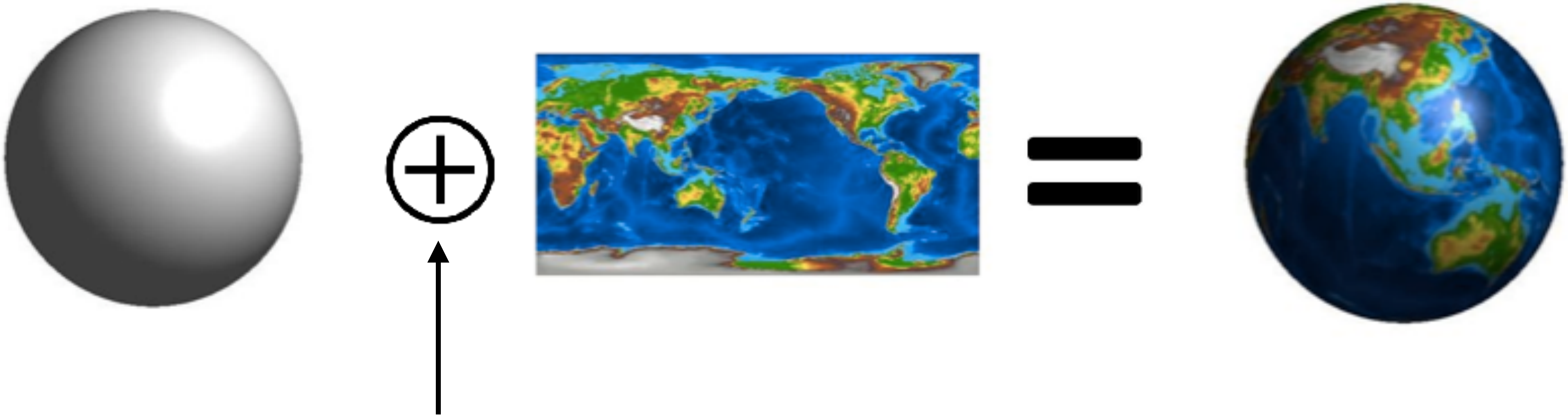
Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions
- Too “general” for VR applications



Current Hardware Usage

- PT is treated as Texture Mapping
- GPU has Texture Mapping Unit (TMU)
- OpenGL supports various object shapes and customized mapping functions
- Too “general” for VR applications



Equirectangular Projection, CubeMap Projection.
Equi-Angular Projection

HAR: Hardware Accelerated Rendering

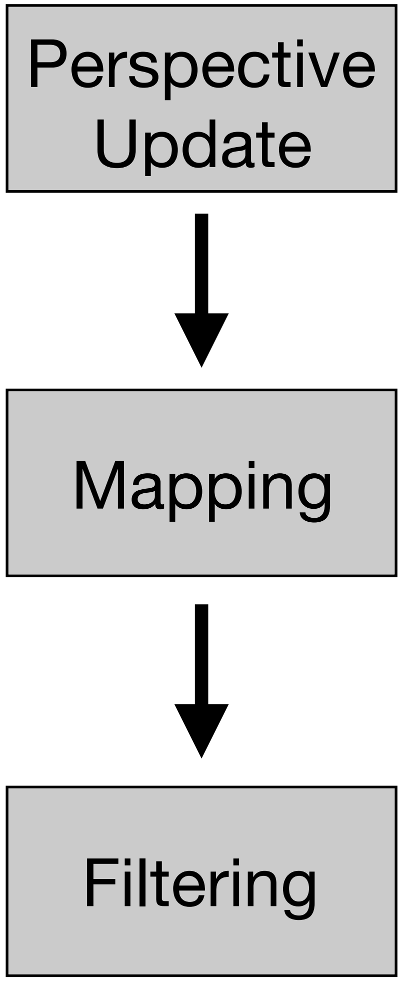
- Key Idea: VR-specific Texture Mapping Accelerator

HAR: Hardware Accelerated Rendering

- Key Idea: VR-specific Texture Mapping Accelerator
  - A specialized hardware for projective transformation

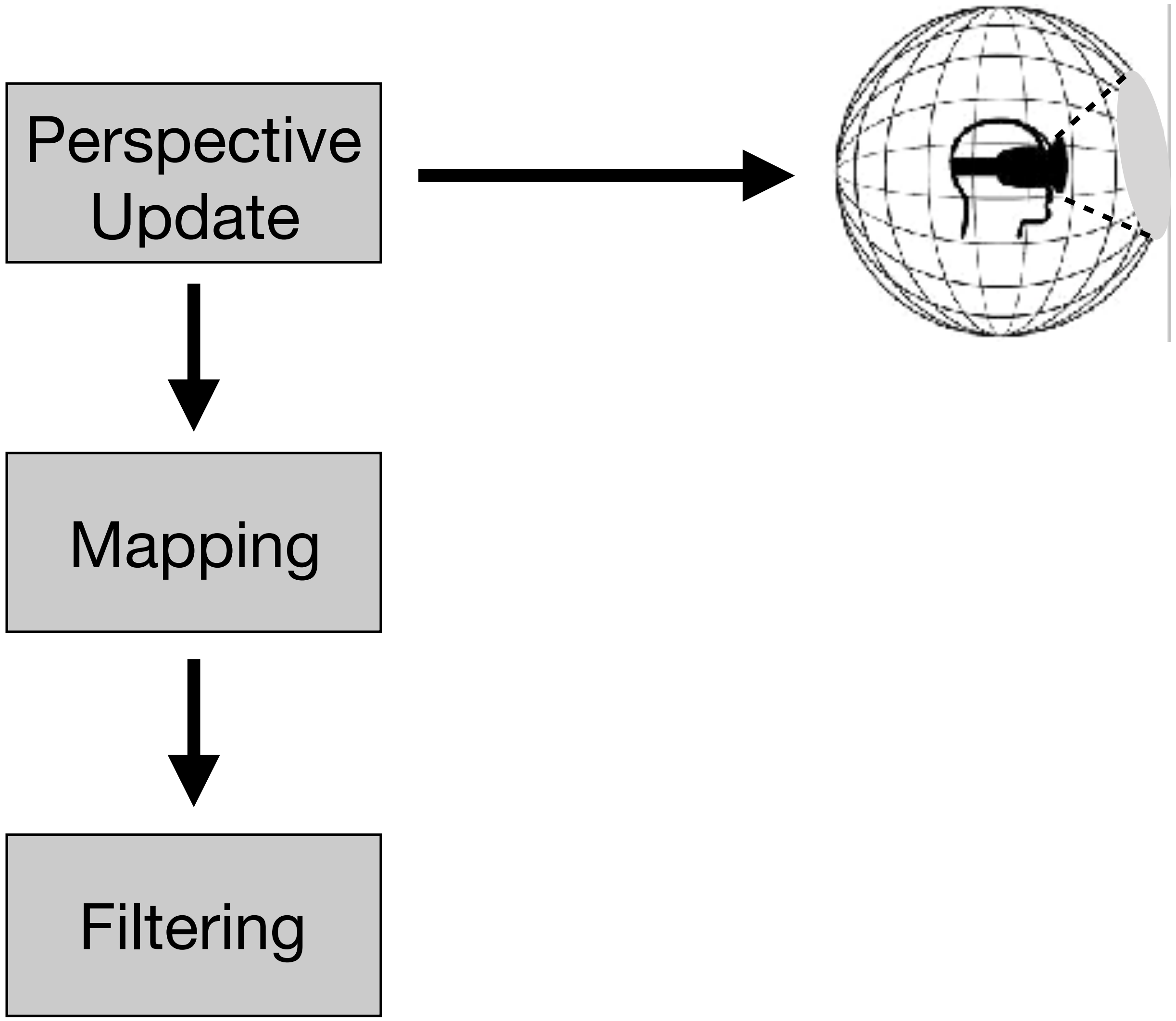
HAR: Hardware Accelerated Rendering

- Key Idea: VR-specific Texture Mapping Accelerator
  - A specialized hardware for projective transformation



HAR: Hardware Accelerated Rendering

- Key Idea: VR-specific Texture Mapping Accelerator
  - A specialized hardware for projective transformation



Perspective Update

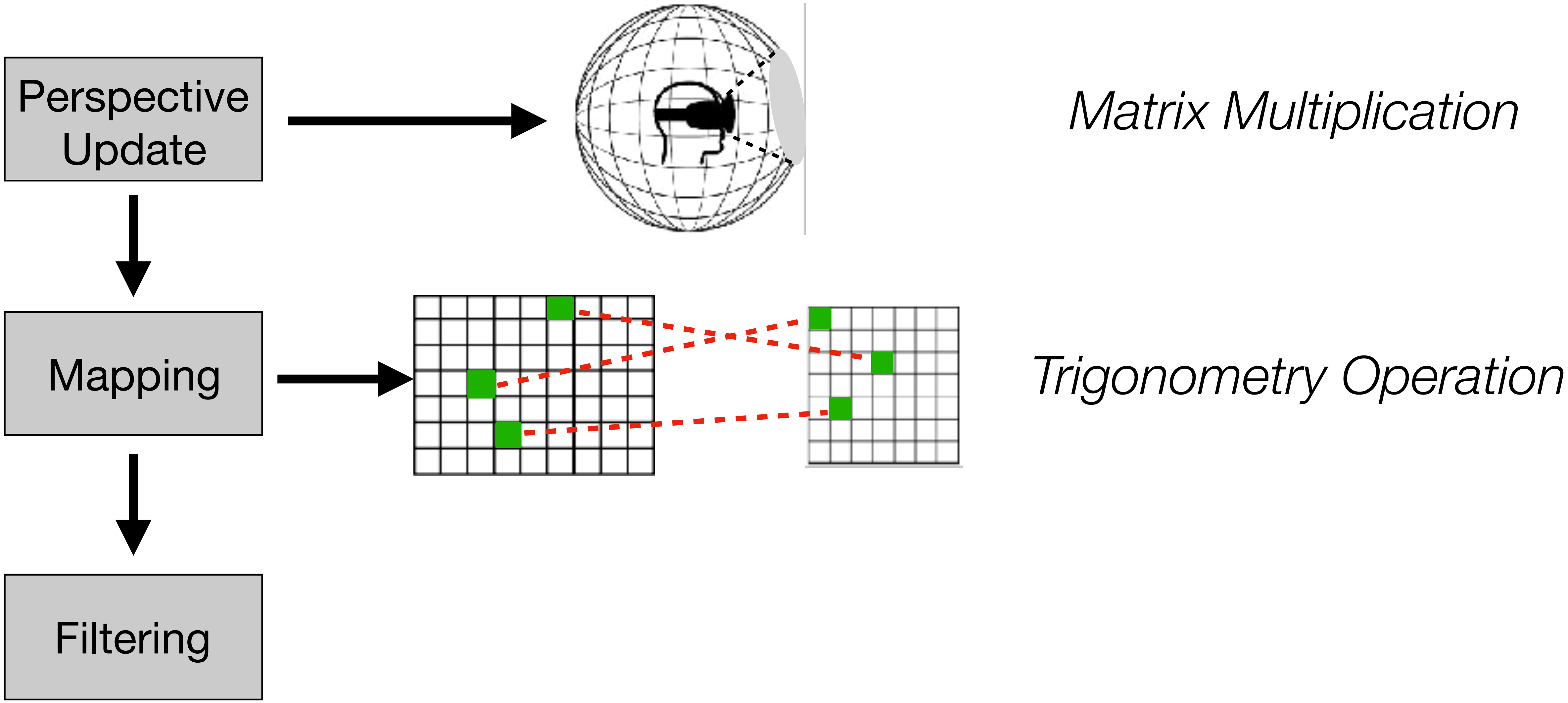
Mapping

Filtering

Matrix Multiplication

HAR: Hardware Accelerated Rendering

- **Key Idea: VR-specific Texture Mapping Accelerator**
  - A specialized hardware for projective transformation



Perspective Update

Mapping

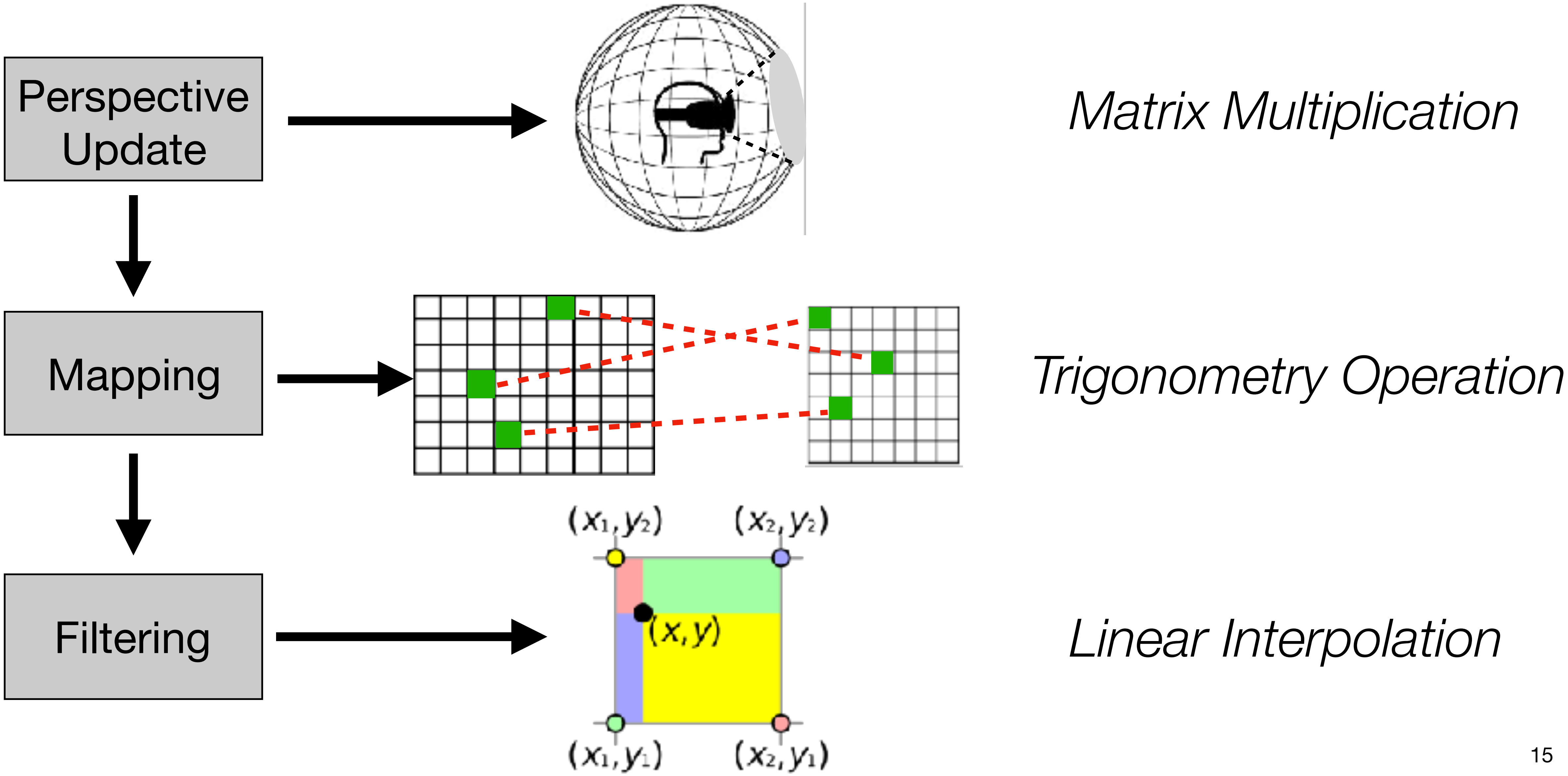
Filtering

Matrix Multiplication

Trigonometry Operation

HAR: Hardware Accelerated Rendering

- **Key Idea: VR-specific Texture Mapping Accelerator**
  - A specialized hardware for projective transformation



Perspective Update

Mapping

Filtering

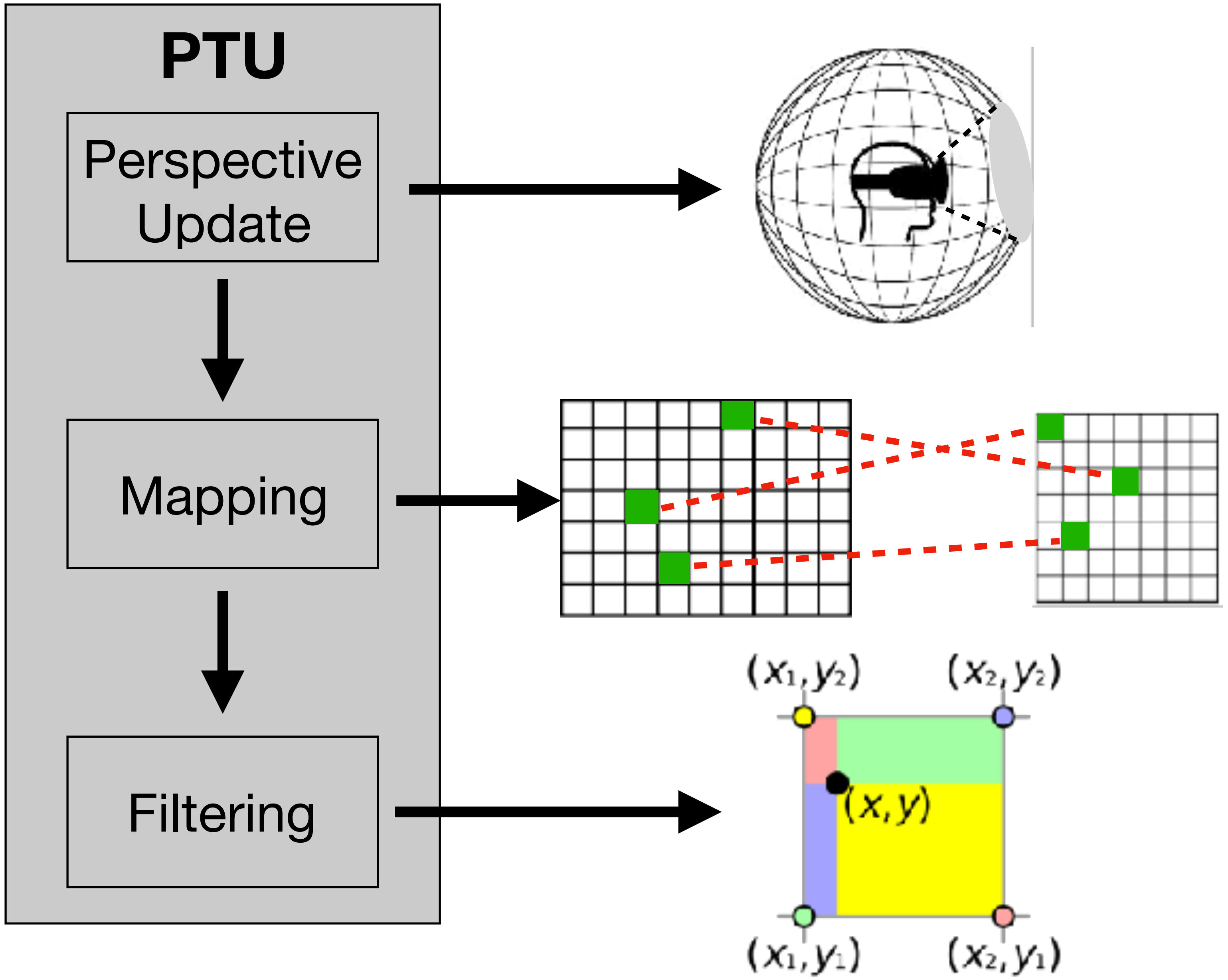
Matrix Multiplication

Trigonometry Operation

Linear Interpolation

HAR: Hardware Accelerated Rendering

- Key Idea: VR-specific Texture Mapping Accelerator
  - A specialized hardware for projective transformation
  - Projective Transformation Unit (PTU)



PTU

Perspective Update

Mapping

Filtering

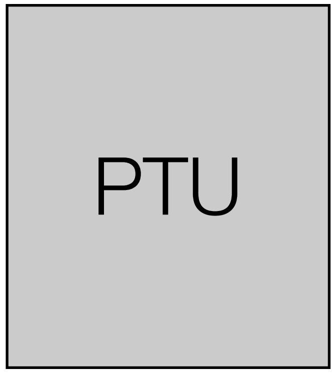
Matrix Multiplication

Trigonometry Operation

Linear Interpolation

Projective Transformation Engine (PTE)

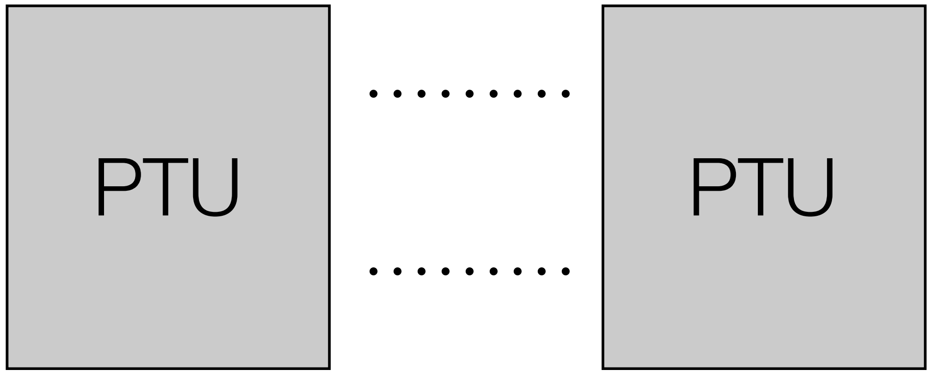
- Exploits pixel-level parallelism of projective transformation



PTU

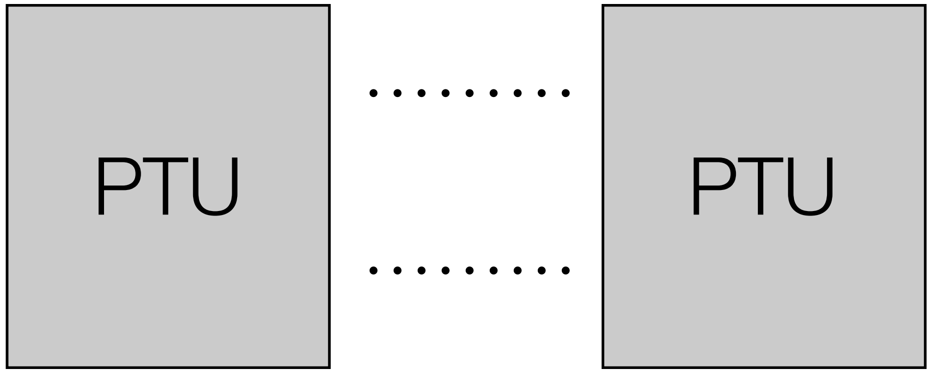
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation



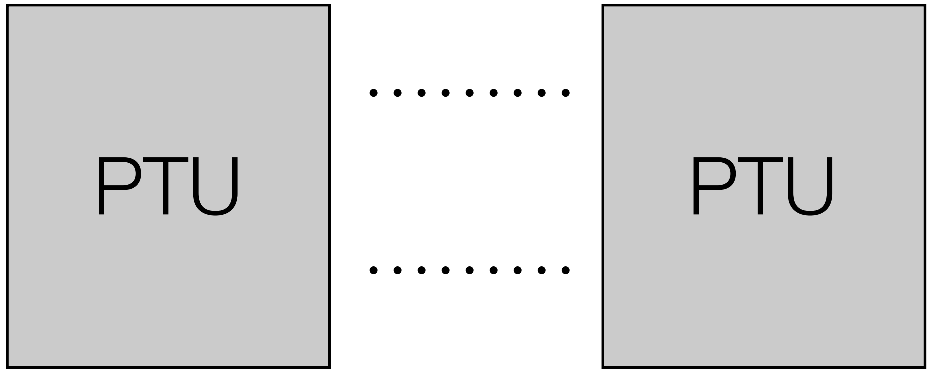
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs



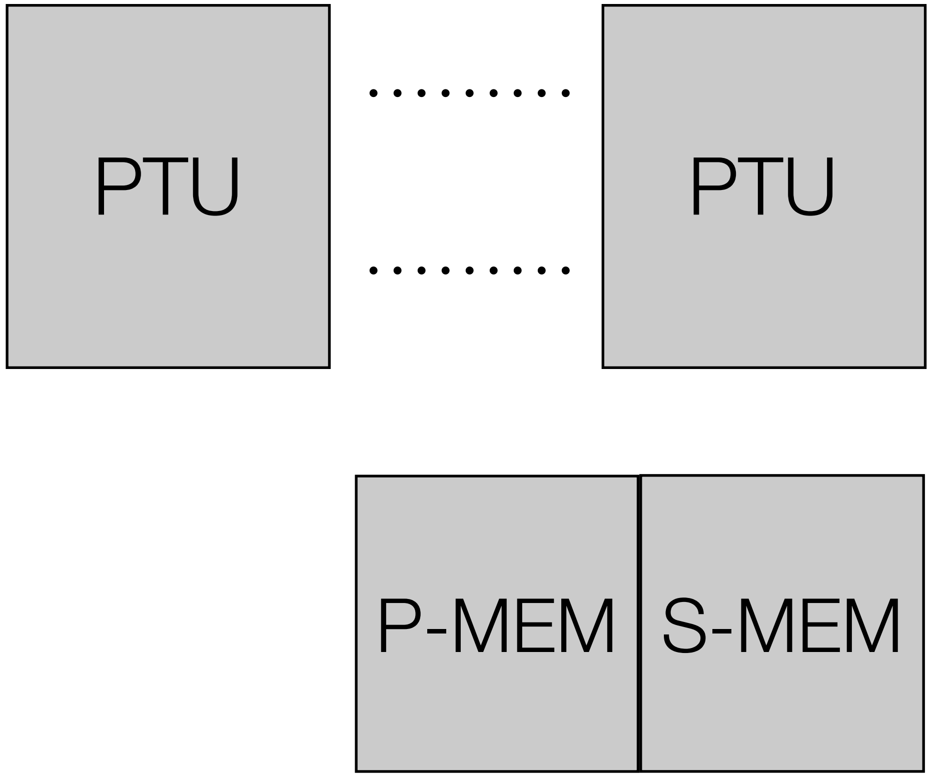
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs
  - Each PTU is responsible for one pixel



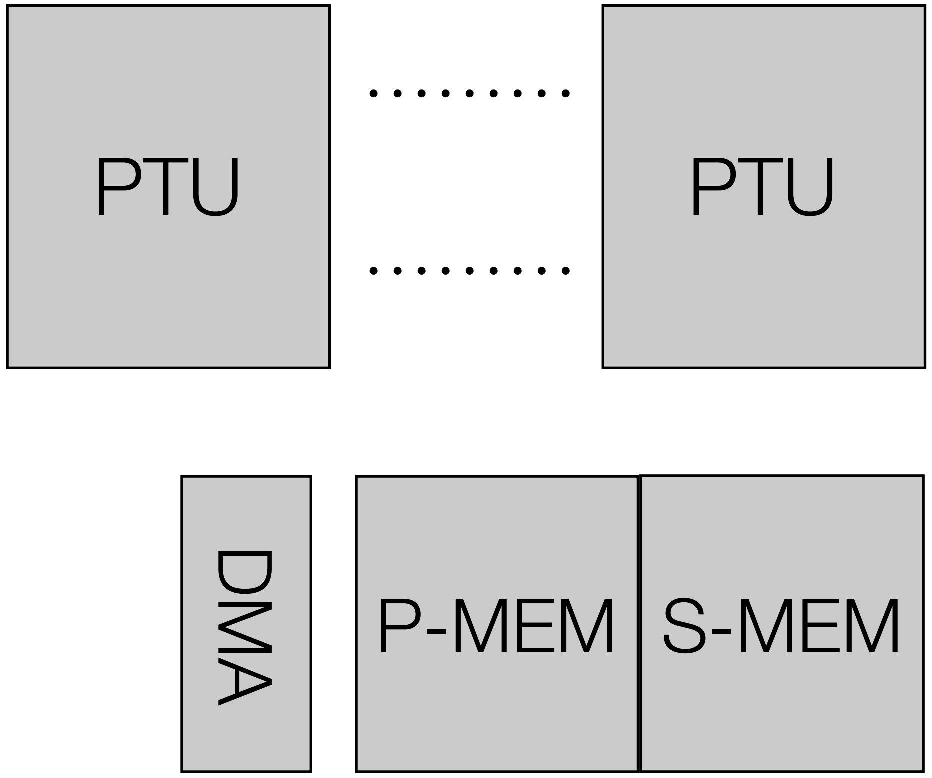
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs
  - Each PTU is responsible for one pixel



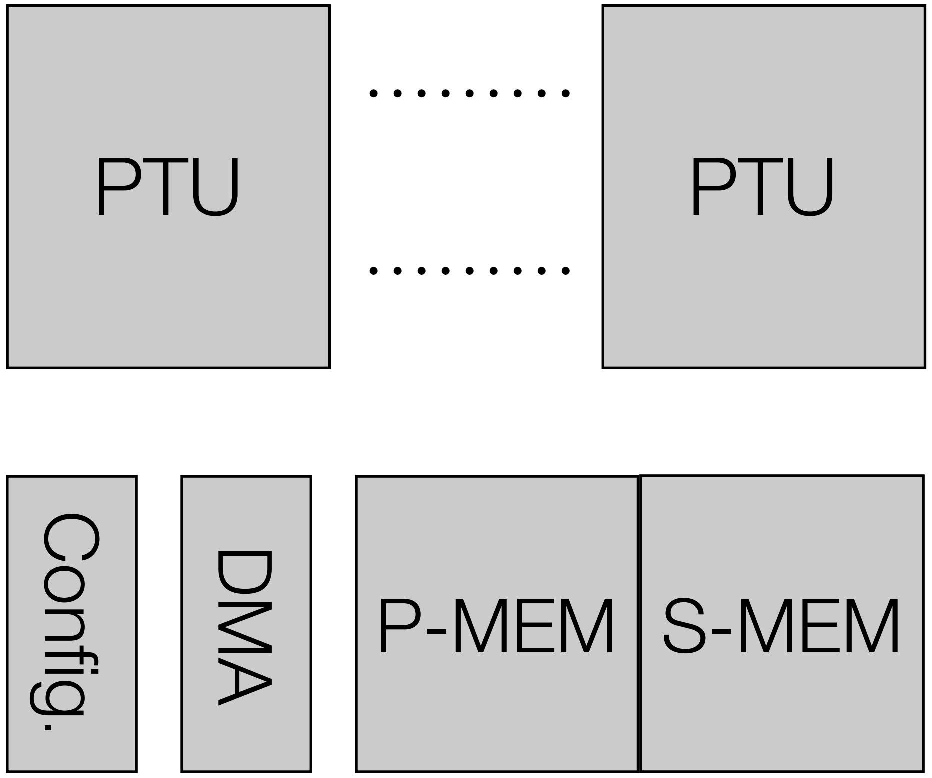
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs
  - Each PTU is responsible for one pixel



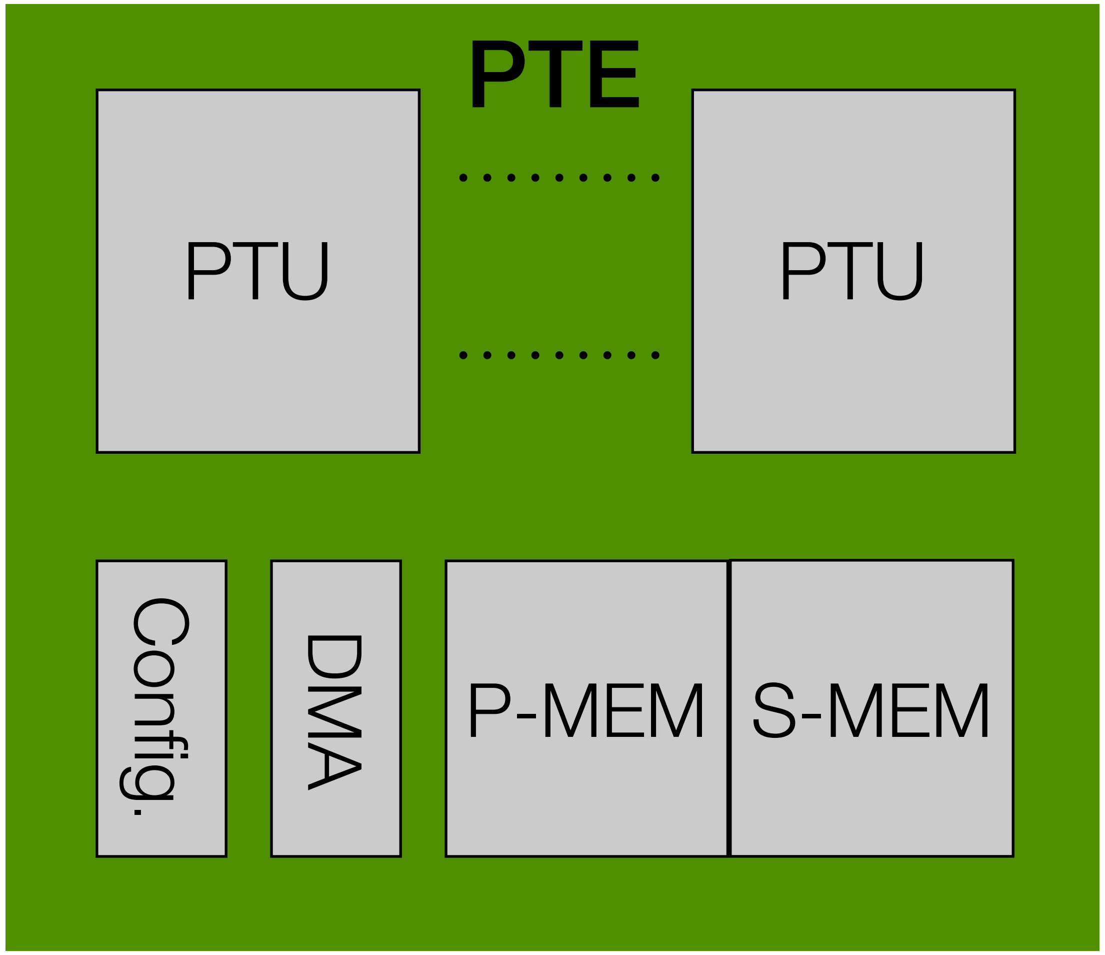
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs
  - Each PTU is responsible for one pixel



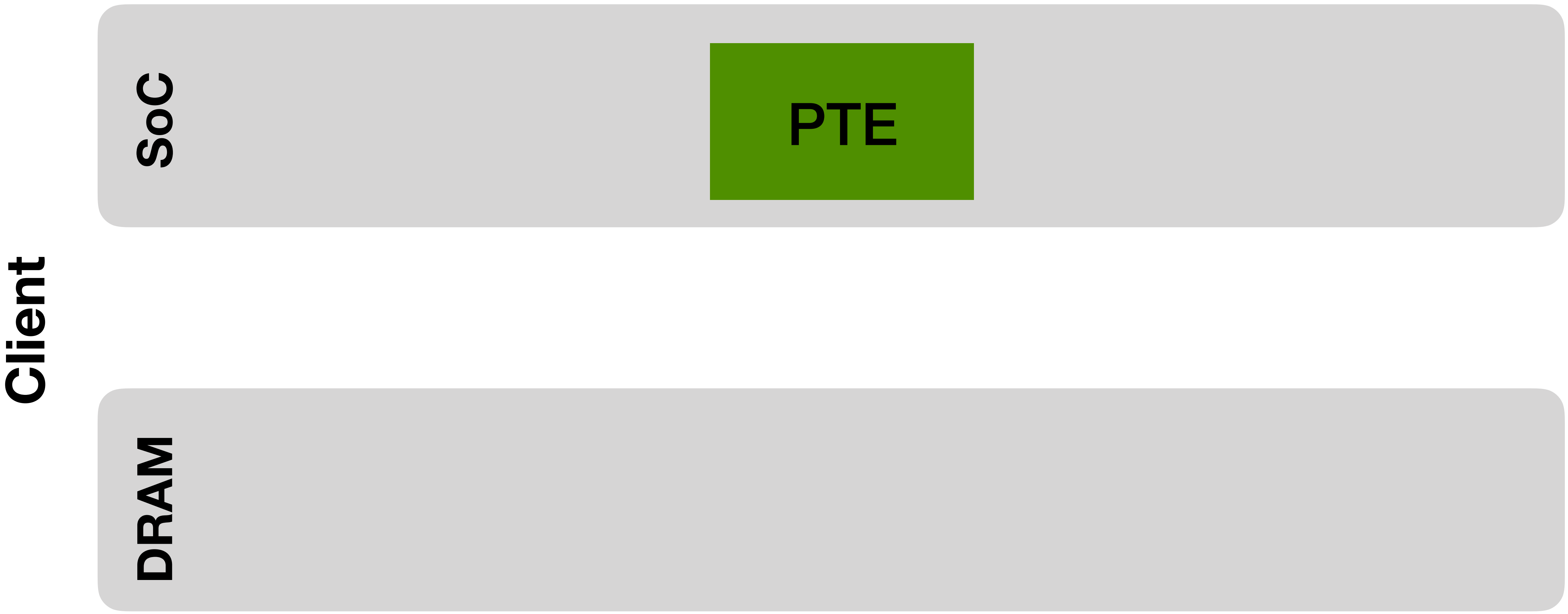
Projective Transformation Engine (PTE)

- Exploits pixel-level parallelism of projective transformation
  - Instantiate multiple PTUs
  - Each PTU is responsible for one pixel



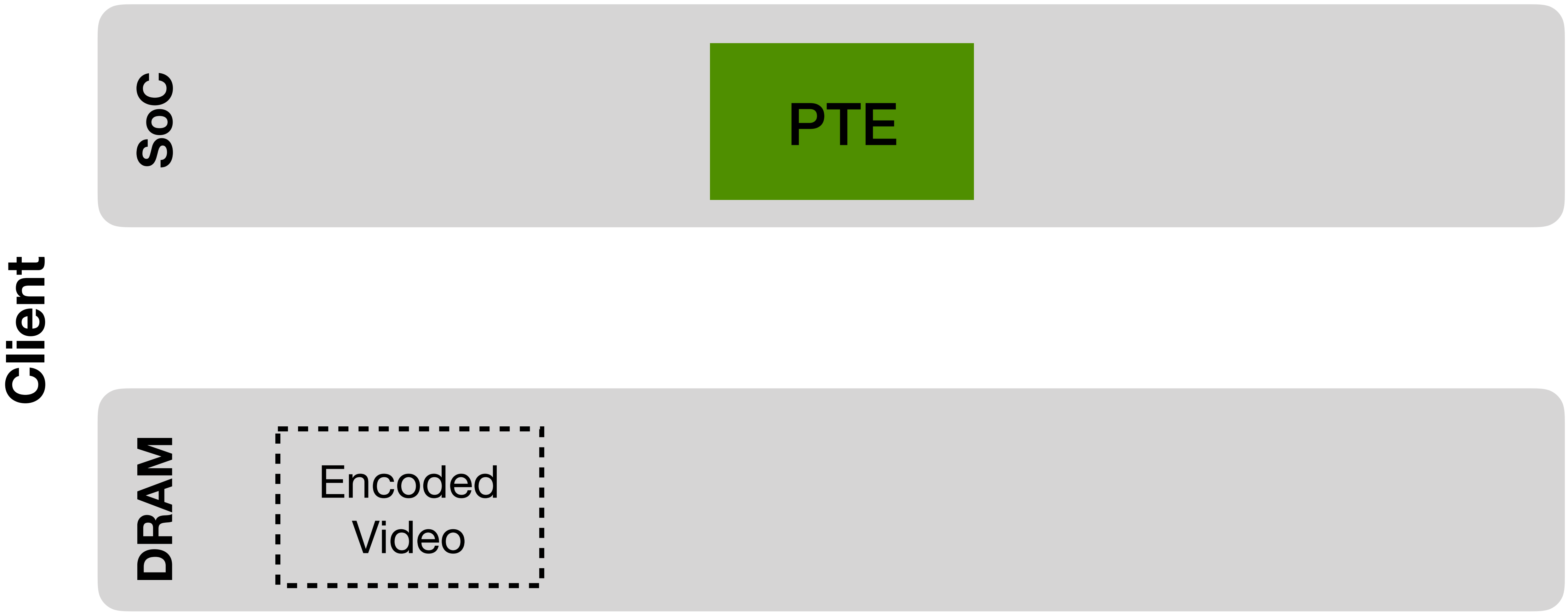
SoC Integration

- Collaborates with IPs in existing hardware architecture



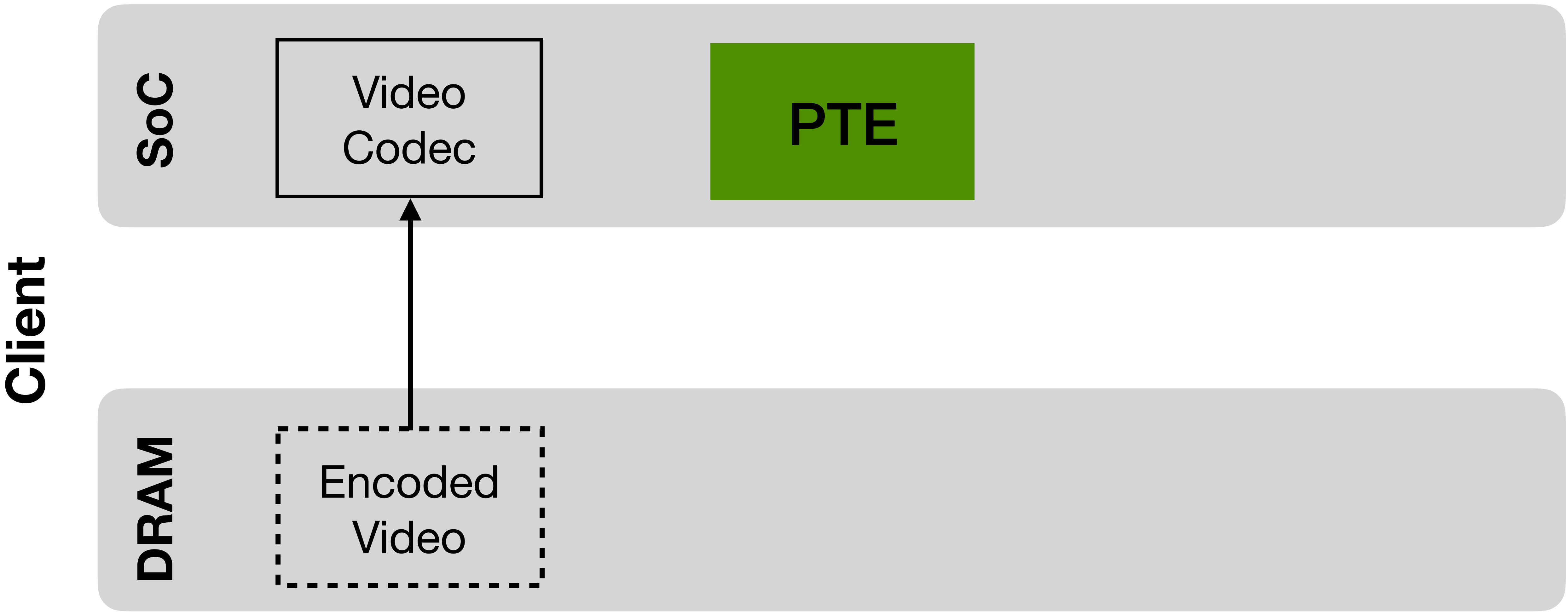
SoC Integration

- Collaborates with IPs in existing hardware architecture



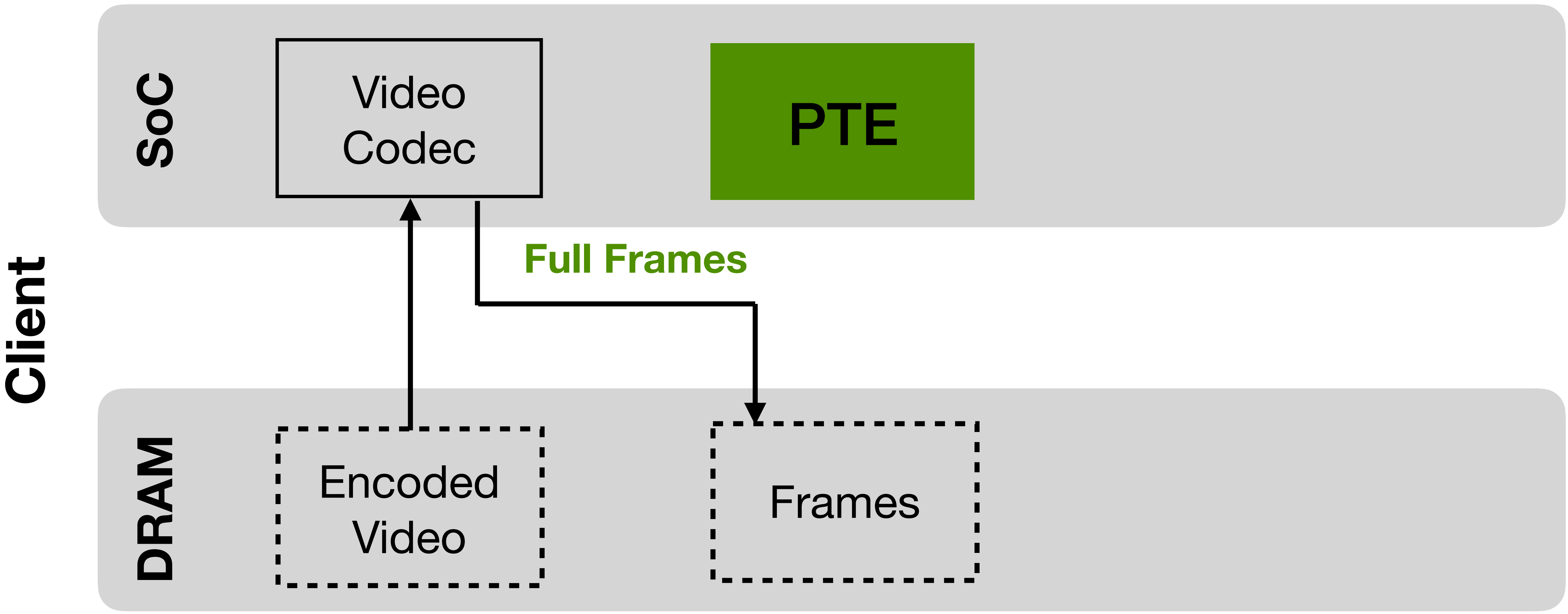
SoC Integration

- Collaborates with IPs in existing hardware architecture



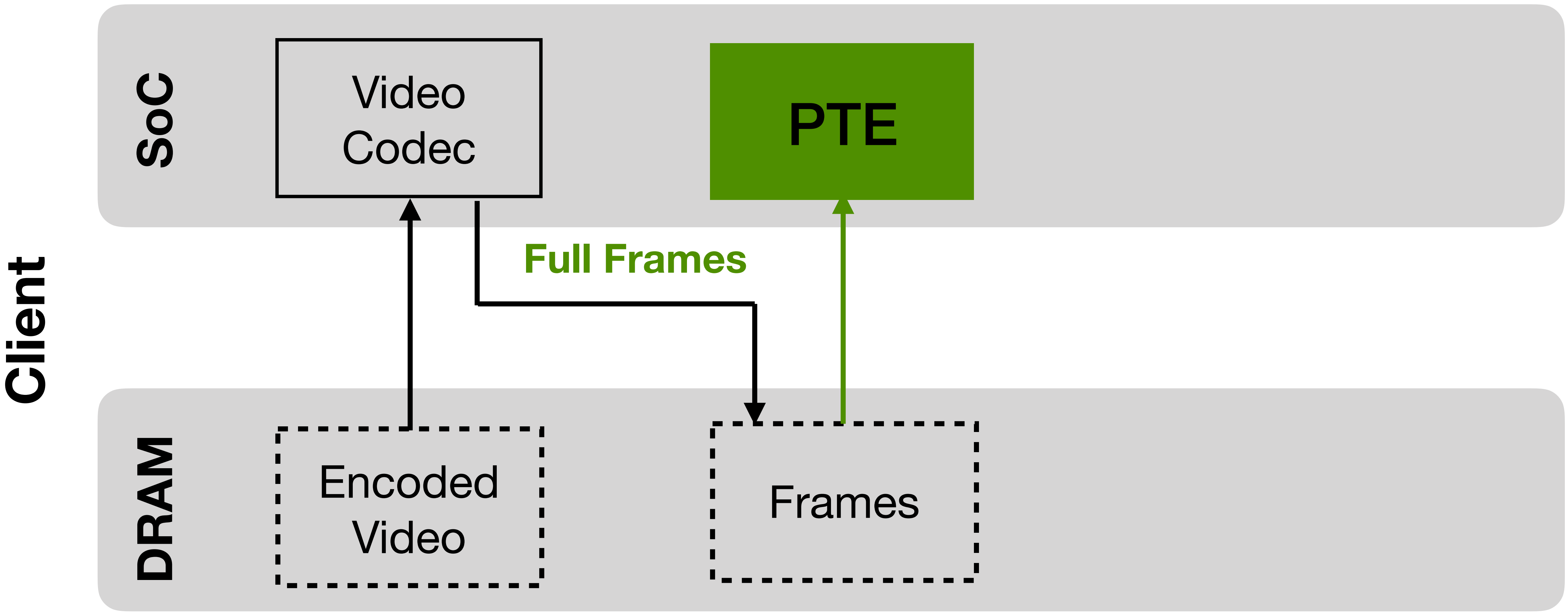
SoC Integration

- Collaborates with IPs in existing hardware architecture



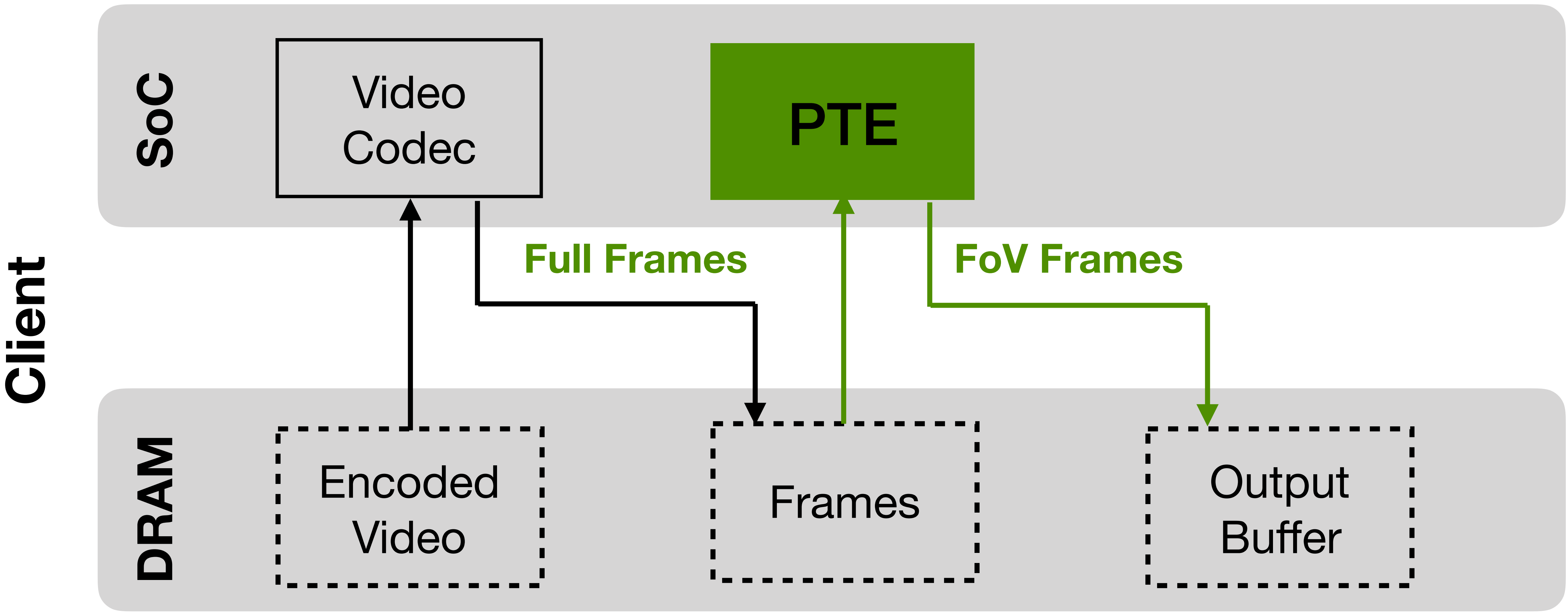
SoC Integration

- Collaborates with IPs in existing hardware architecture



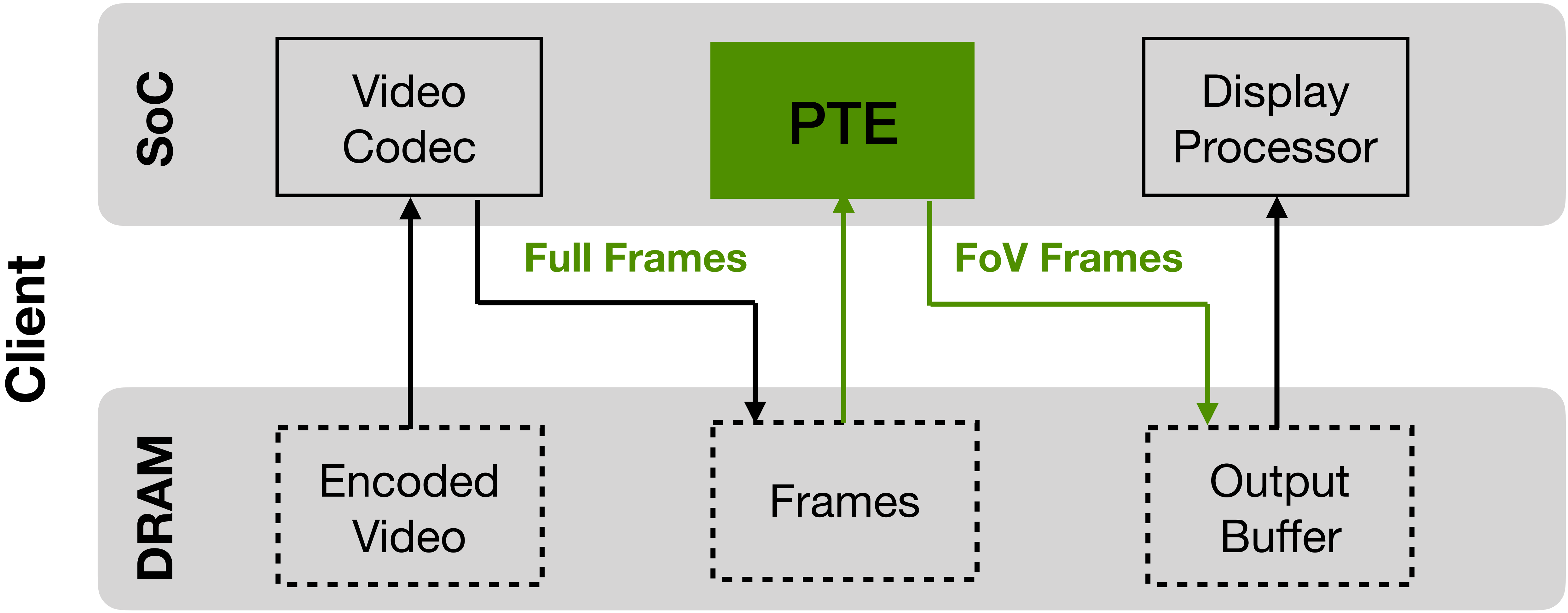
SoC Integration

- Collaborates with IPs in existing hardware architecture

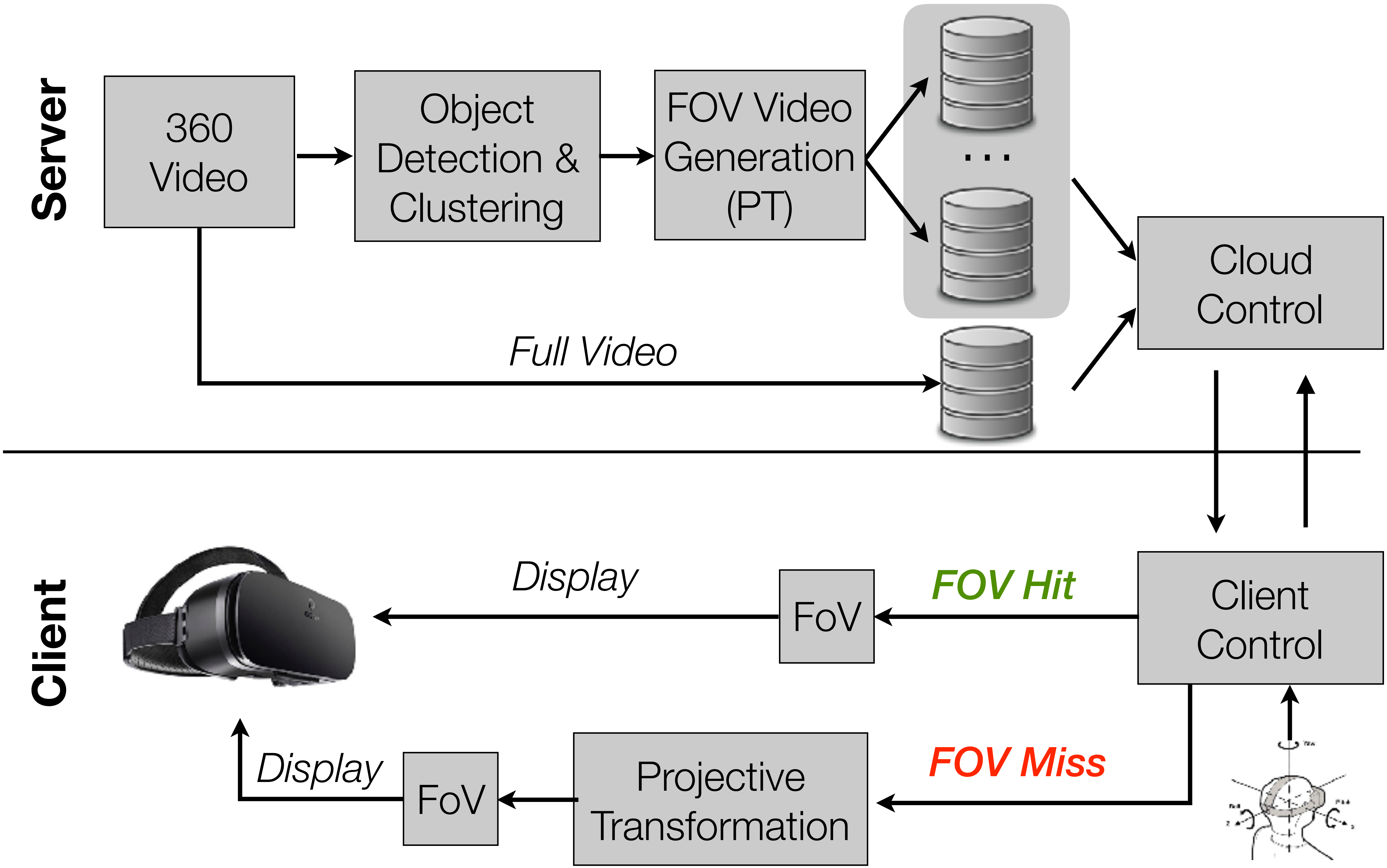


SoC Integration

- Collaborates with IPs in existing hardware architecture



Putting It Together



Server

360 Video

Object Detection & Clustering

FOV Video Generation (PT)

. . .

Full Video

Cloud Control

Client

Client Control

FOV Hit

FoV

Display

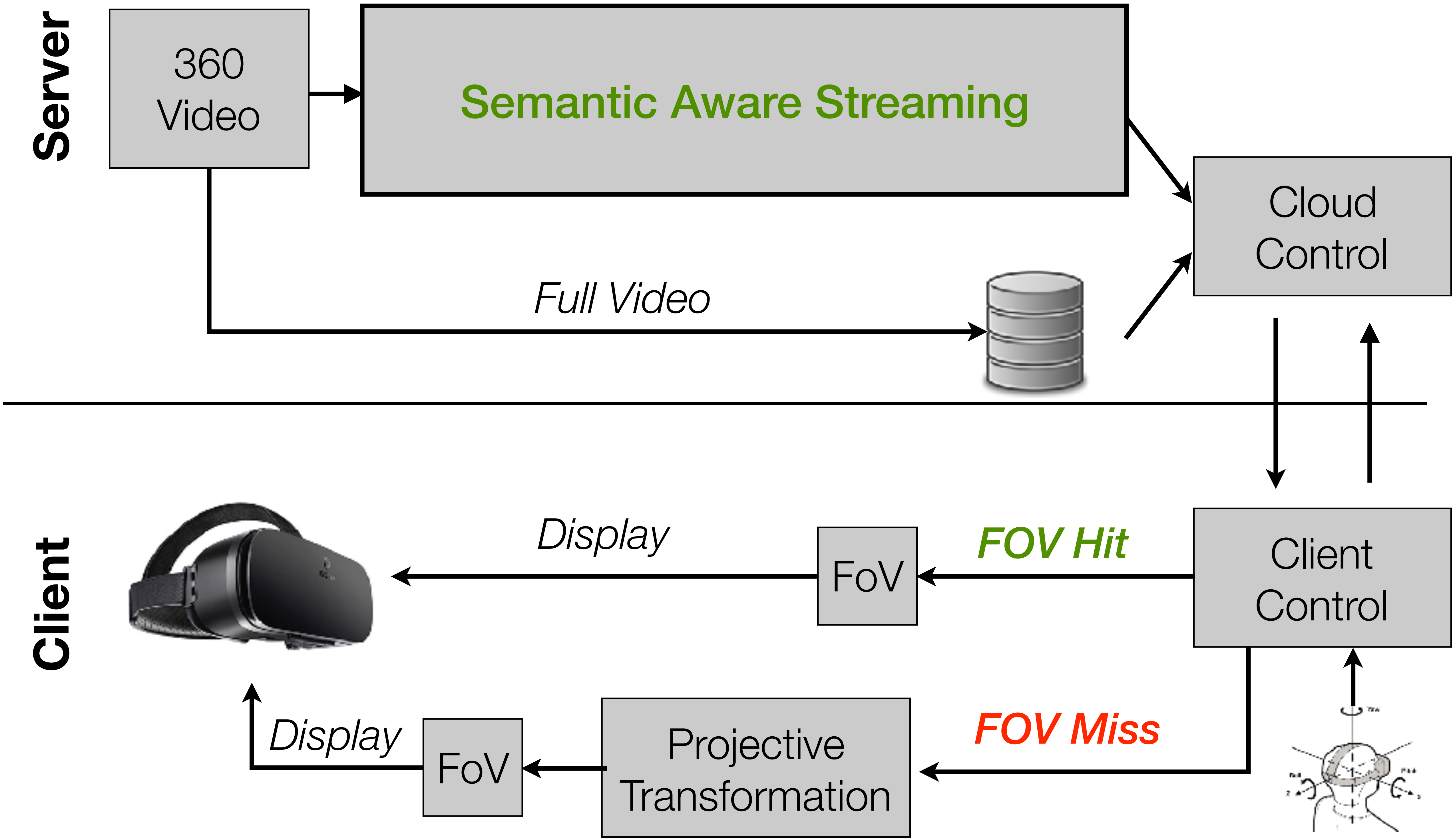
FOV Miss

Projective Transformation

FoV

Display

Putting It Together



Server

360 Video

Semantic Aware Streaming

Cloud Control

Full Video

Client

Display

FoV

FOV Hit

Client Control

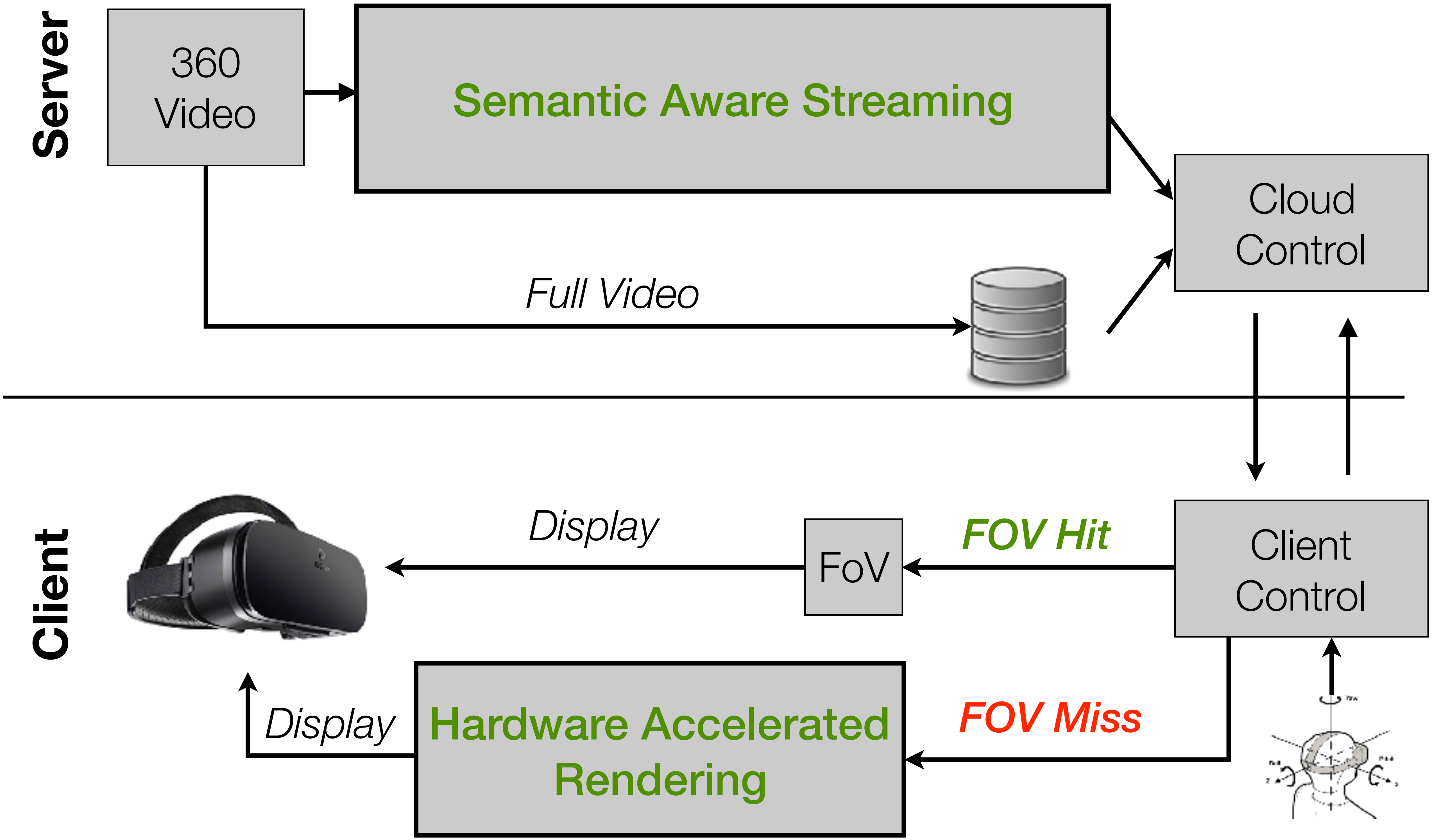
Display

FoV

Projective Transformation

FOV Miss

Putting It Together



Server

360 Video

Semantic Aware Streaming

Full Video

Cloud Control

Client

Display

FoV

FOV Hit

Client Control

Display

Hardware Accelerated Rendering

FOV Miss

EVR

A Cloud-Client Collaborative VR Video System

Cloud	Semantic Aware Streaming
Client	Hardware Accelerated Rendering
Result	Up to 58% Energy Reduction & Only 1% Frame Rate Drops

EVR

A Cloud-Client Collaborative VR Video System

Cloud

Semantic Aware Streaming

Client

Hardware Accelerated Rendering

Result

Up to 58% Energy Reduction & Only 1% Frame Rate Drops

Experimental Setup

- Dataset: 59 real VR user head movement traces from watching 5 YouTube VR videos

Experimental Setup

- **Dataset:** 59 real VR user head movement traces from watching 5 YouTube VR videos
- **SAS server:** Deployed on an AWS instance (EC2 + S3)

Amazon EC2

Amazon S3

Experimental Setup

- **Dataset:** 59 real VR user head movement traces from watching 5 YouTube VR videos
- **SAS server:** Deployed on an AWS instance (EC2 + S3)
- **SAS client:** Java-based in-house player. FFmpeg+NVDEC for video decoding.

Amazon EC2

Amazon S3

Experimental Setup

- **Dataset**: 59 real VR user head movement traces from watching 5 YouTube VR videos
- **SAS server**: Deployed on an AWS instance (EC2 + S3)
- **SAS client**: Java-based in-house player. FFmpeg+NVDEC for video decoding.
- **Network:** 300 Mbps on campus WiFi environment

Amazon EC2

Amazon S3

Experimental Setup

- **Dataset:** 59 real VR user head movement traces from watching 5 YouTube VR videos
- **SAS server:** Deployed on an AWS instance (EC2 + S3)
- **SAS client:** Java-based in-house player. FFmpeg+NVDEC for video decoding.
- **Network:** 300 Mbps on campus WiFi environment
- **Hardware:**
  - Nvidia TX2 and Xilinx ZC702 development board

Amazon EC2

Amazon S3

Nvidia TX2

Xilinx ZC702

Evaluation

Evaluation

- **Baseline:** VR client device built on Nvidia TX2

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
    - Running at 30 FPS

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- **Metrics:**

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- **Metrics:**
  - Energy Savings

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- **Metrics:**
  - Energy Savings
  - User Experience Loss

Evaluation

- Baseline: VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- Metrics:
  - Energy Savings
  - User Experience Loss
- Use Cases

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- **Metrics:**
  - Energy Savings
  - User Experience Loss
- **Use Cases**
  - Only Semantic Aware Streaming (SAS)

Evaluation

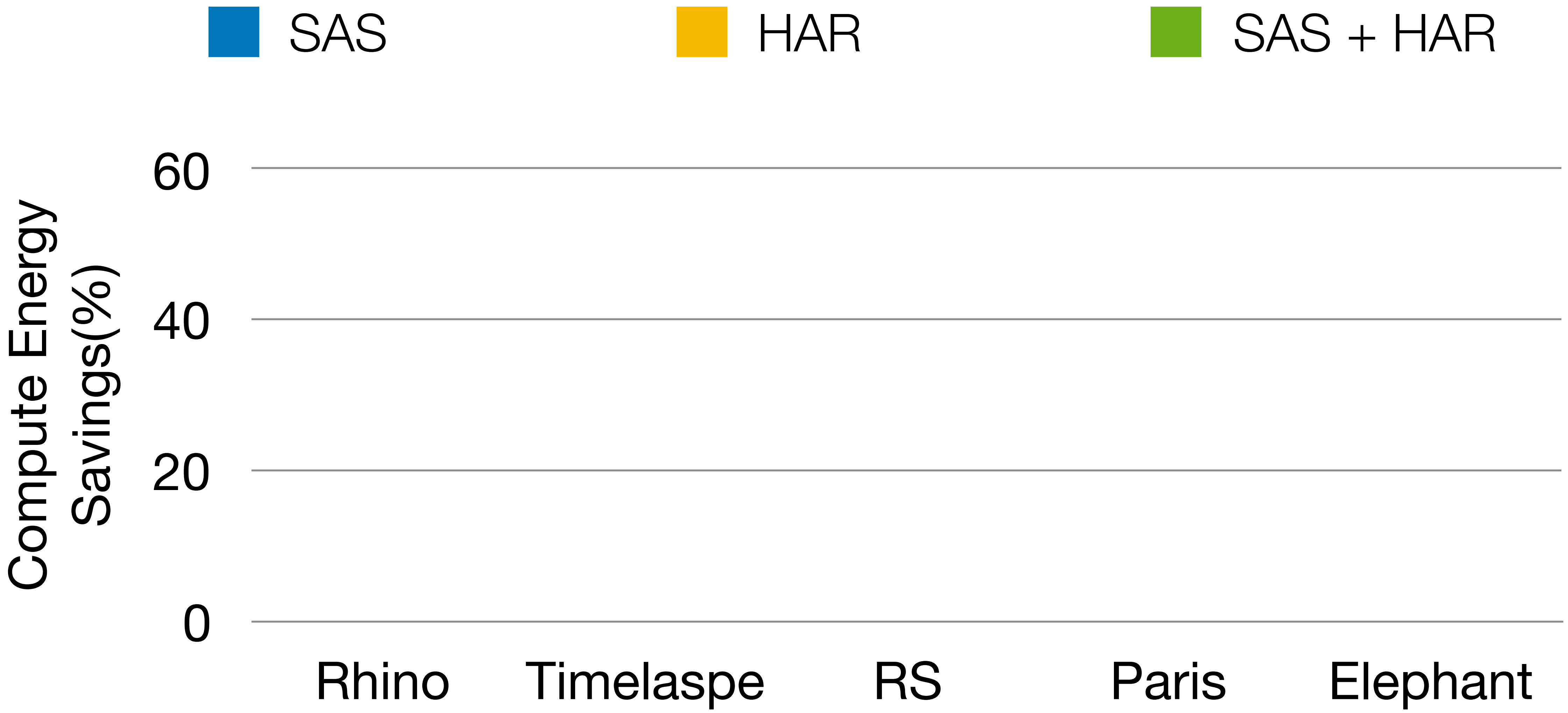
- **Baseline:** VR client device built on Nvidia TX2
    - Running at 30 FPS
    - No SAS and HAR optimizations
- **Metrics:**
    - Energy Savings
    - User Experience Loss
- **Use Cases**
    - Only Semantic Aware Streaming (SAS)
    - Only Hardware Accelerated Rendering (HAR)

Evaluation

- **Baseline:** VR client device built on Nvidia TX2
  - Running at 30 FPS
  - No SAS and HAR optimizations
- **Metrics:**
  - Energy Savings
  - User Experience Loss
- **Use Cases**
  - Only Semantic Aware Streaming (SAS)
  - Only Hardware Accelerated Rendering (HAR)
  - SAS + HAR

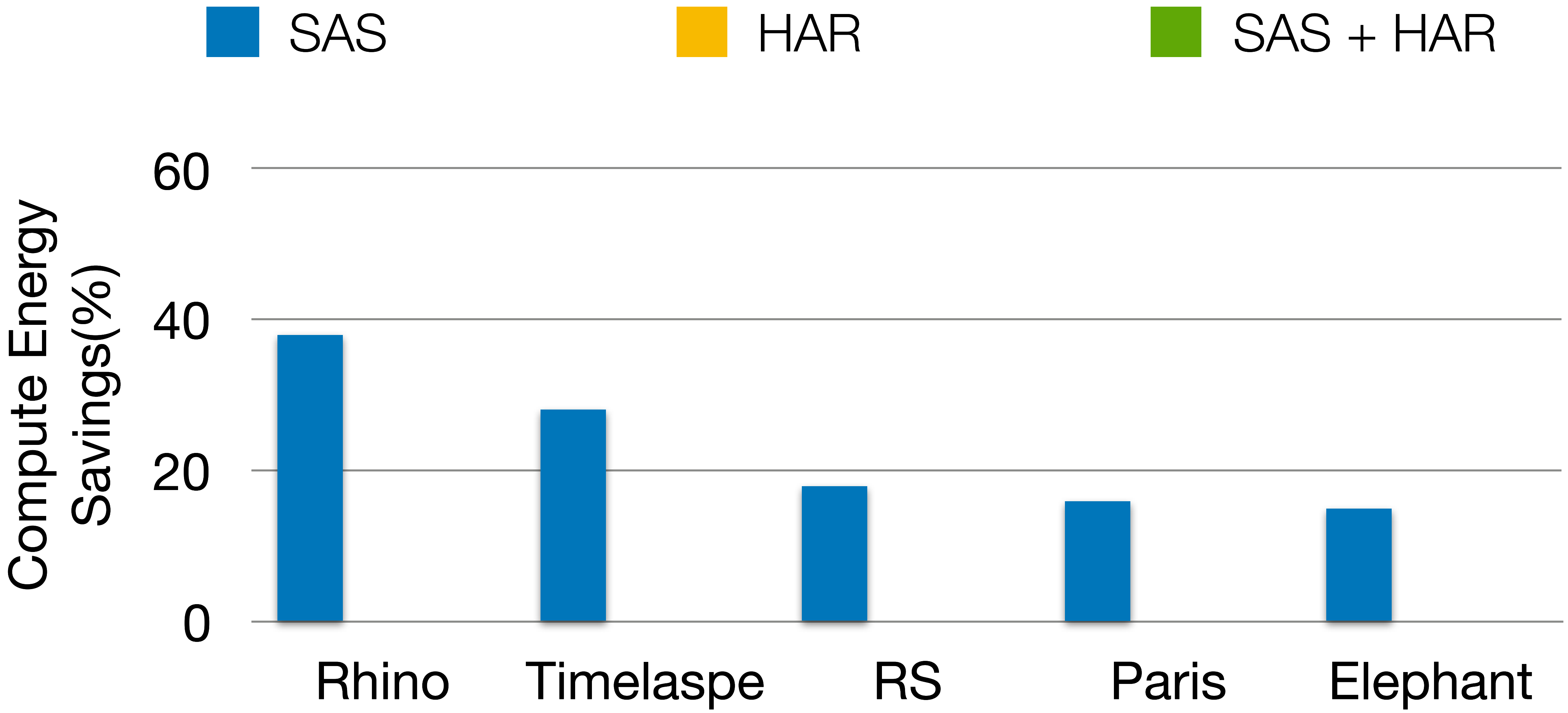
Energy Savings

Energy Savings



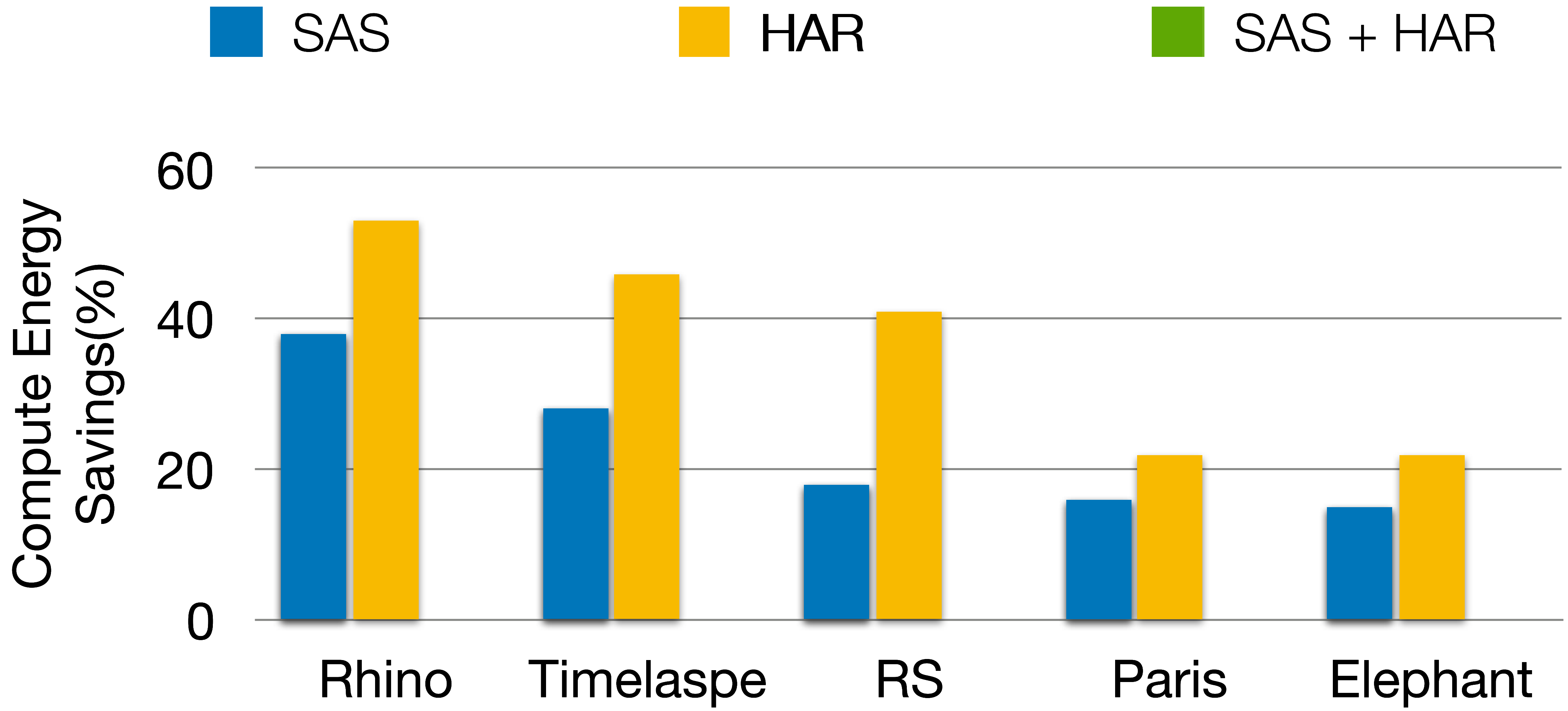
Energy Savings

- Applying only SAS achieves ~22% energy reduction



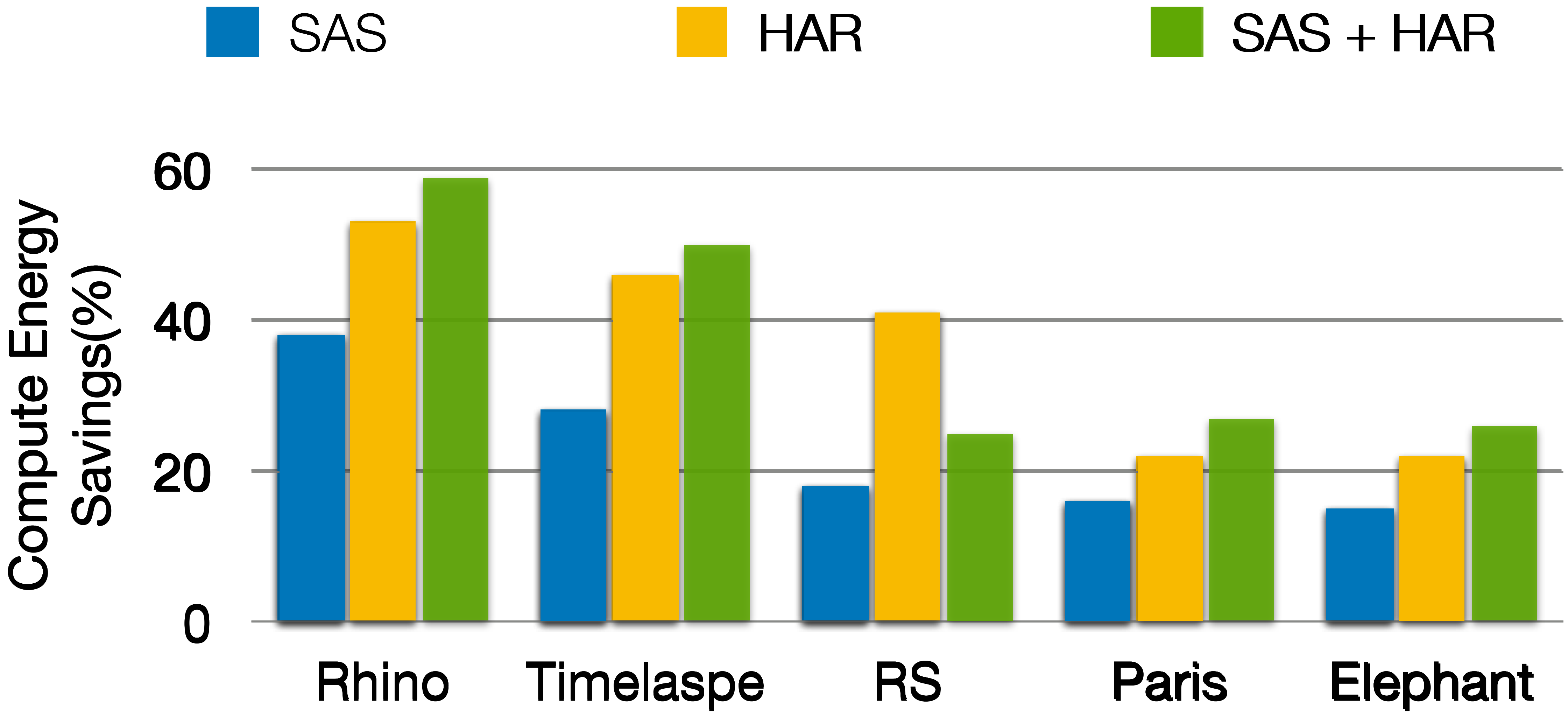
Energy Savings

- Applying only SAS achieves ~22% energy reduction
- Applying only HAR achieves ~38% energy reduction

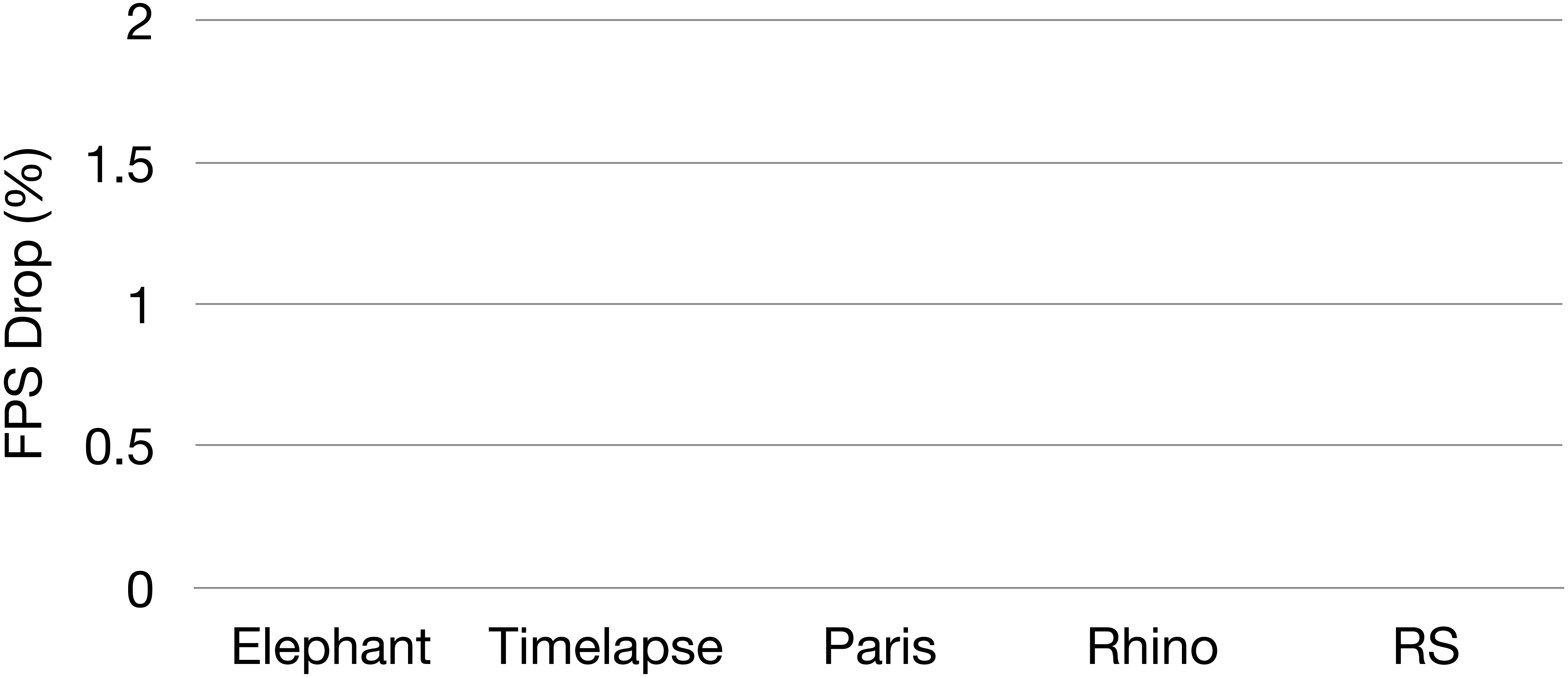


Energy Savings

- Applying only SAS achieves ~22% energy reduction
- Applying only HAR achieves ~38% energy reduction
- SAS + HAR save up to 58% and ~41% energy

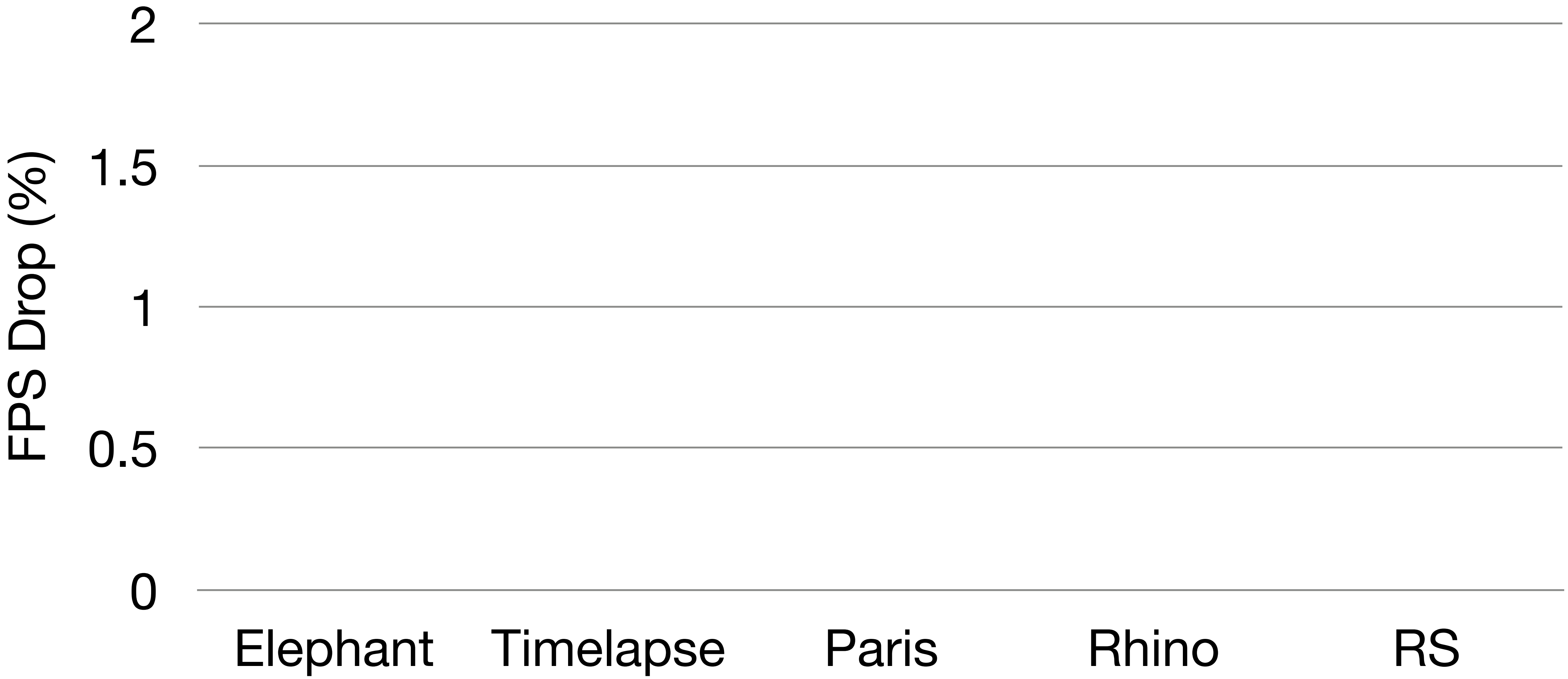


User Experience Loss



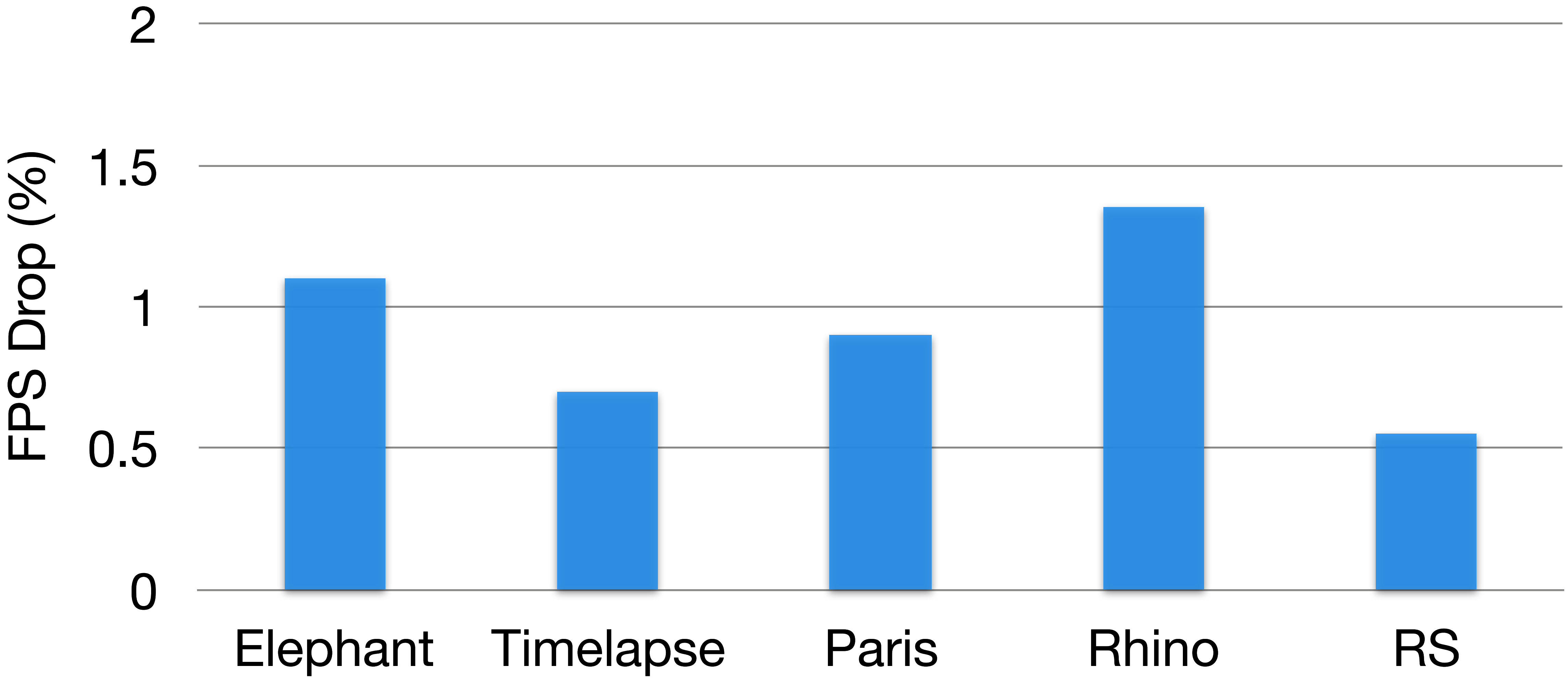
User Experience Loss

- SAS introduce a FOV-miss rate of ~7.7%



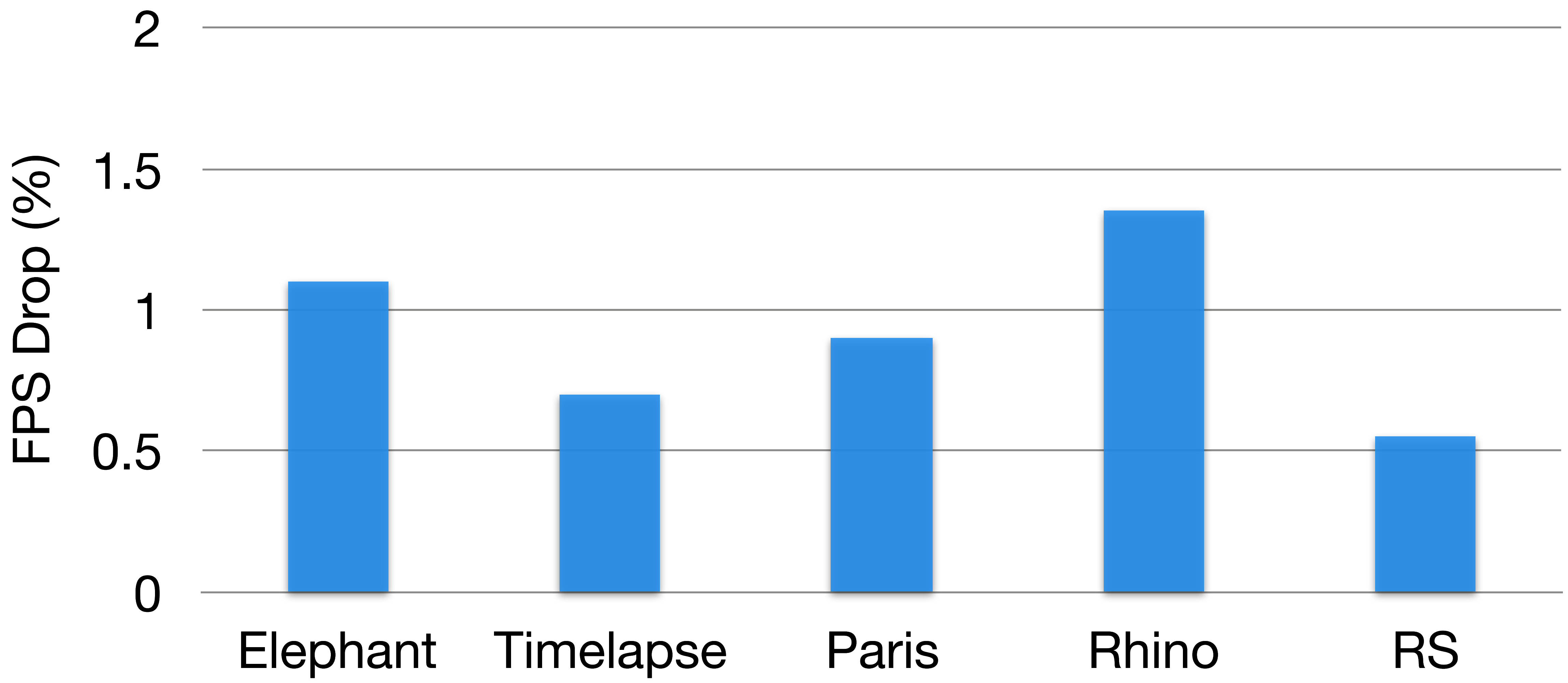
User Experience Loss

- SAS introduce a FOV-miss rate of ~7.7%
- FPS drops only ~1%



User Experience Loss

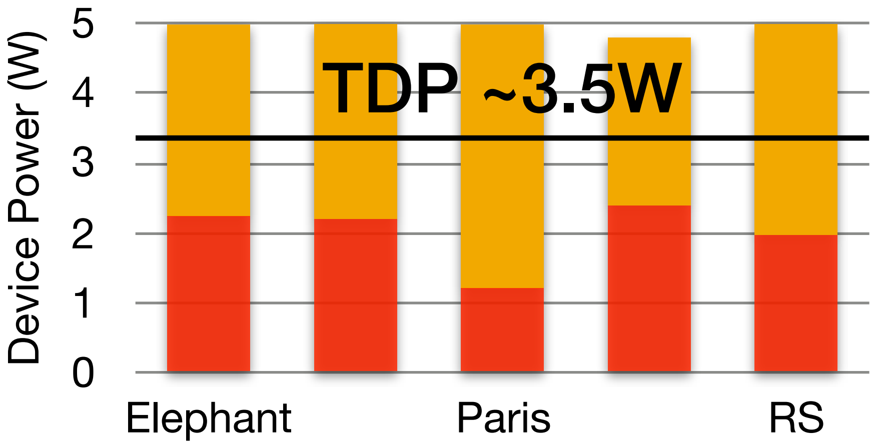
- SAS introduce a FOV-miss rate of ~7.7%
- FPS drops only ~1%
- Loss is negligible by users



Conclusion

- Virtual reality popularity is growing rapidly

Conclusion

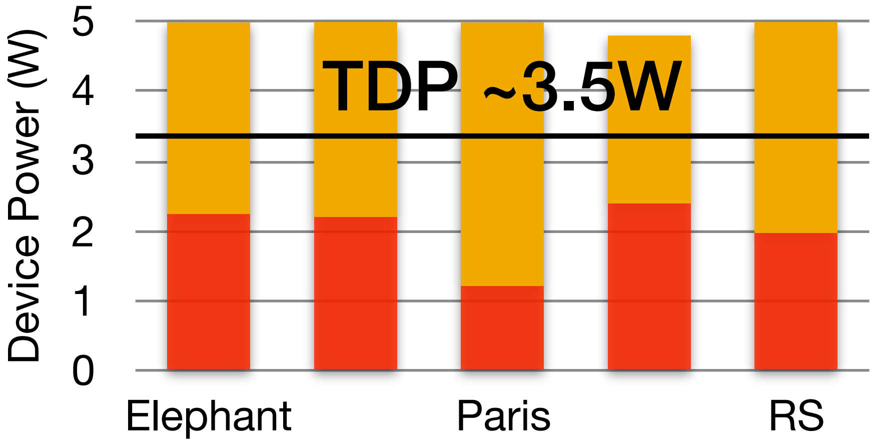


- Virtual reality popularity is growing rapidly

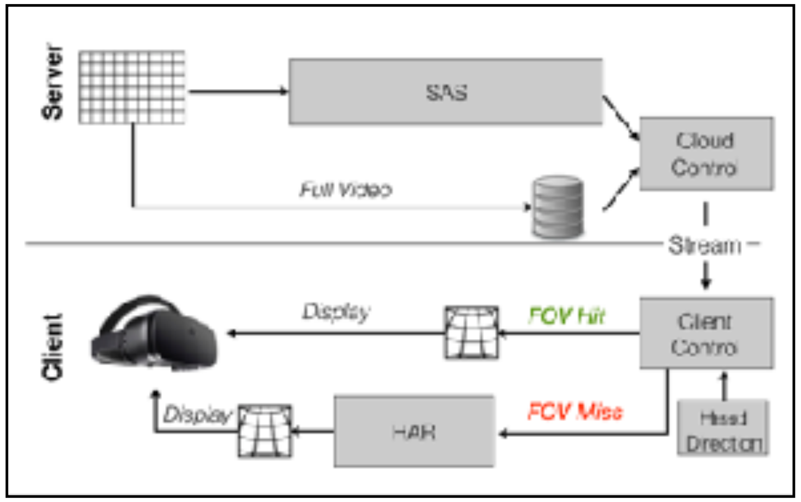
- Virtual reality devices consume excessive power

Conclusion

- Virtual reality popularity is growing rapidly



- Virtual reality devices consume excessive power



- EVR saves up 58% of energy
  - ▷ Semantic Aware Streaming
  - ▷ Hardware Accelerated Rendering