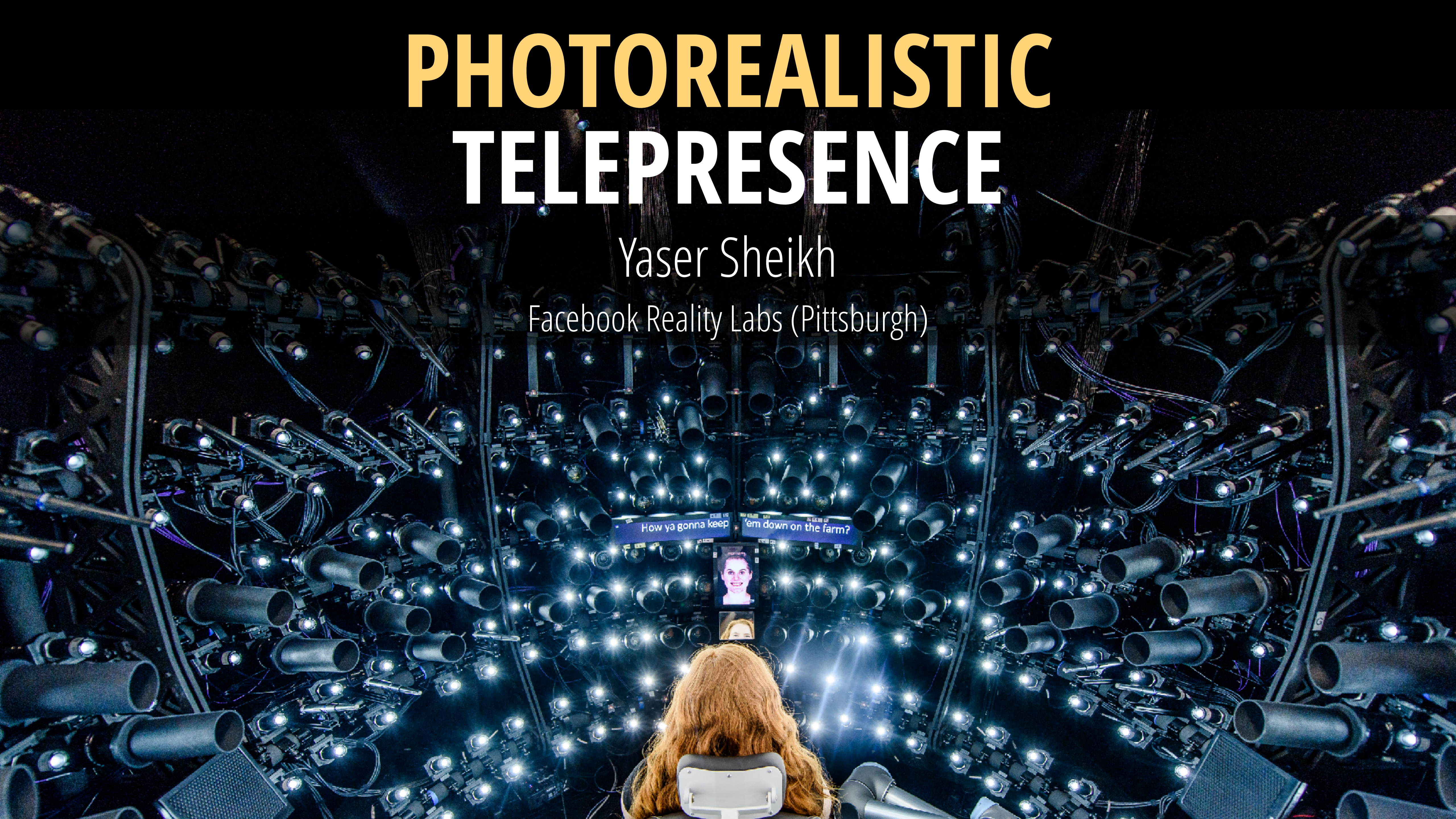


# PHOTOREALISTIC TELEPRESENCE

Yaser Sheikh

Facebook Reality Labs (Pittsburgh)





**HUMANS ARE SOCIAL ANIMALS**





# HUMANS ARE SOCIAL ANIMALS





"...an elaborate and secret code that is written nowhere,  
known by none,  
and understood by all"

-- Edward Sapir (1927)



**HUMANS ARE SOCIAL ANIMALS**





**PROXIMITY DETERMINES SOCIAL RELATIONSHIPS**





# PROXIMITY DETERMINES SOCIAL RELATIONSHIPS

Postal Service



550 BC

Telegraph



1840s

Telephone



1880s

Video Conferencing



1950s







Michael Findley



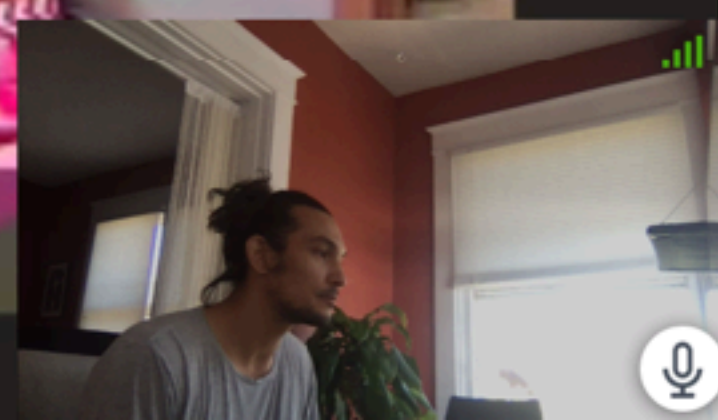
Mary Green



Saleem Muhammad



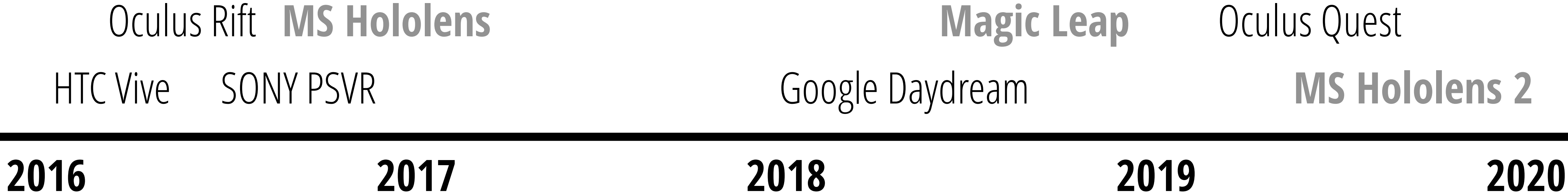
Erin Zimmerman





# 3D DISPLAYS FOR ARTIFICIAL REALITY

Virtual, Augmented, and Mixed Reality





Enable **Authentic** Communication in **Artificial** Reality







# THESIS

**METRIC TELEPRESENCE IS SUFFICIENT FOR AUTHENTIC COMMUNICATION**

Represent *true* (metric) presence,  
rather than “perceptually plausible”



# Metric Identity

How do we produce identity preserving avatars for billions of people?

# Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

# Metric Time

How do we do all this in realtime without access to artistic correction?



# Metric Identity

How do we produce identity preserving avatars for billions of people?

# Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

# Metric Time

How do we do all this in realtime without access to artistic correction?



What is the State-of-the-art in  
**DIGITAL HUMANS?**



**"Virtualized Reality," Kanade et al. 1997**



**"Tele-immersion," Bajscy et al. 2005**



**Orts-Escolano et al. "Holoportation: Virtual 3d teleportation in real-time," 2016**

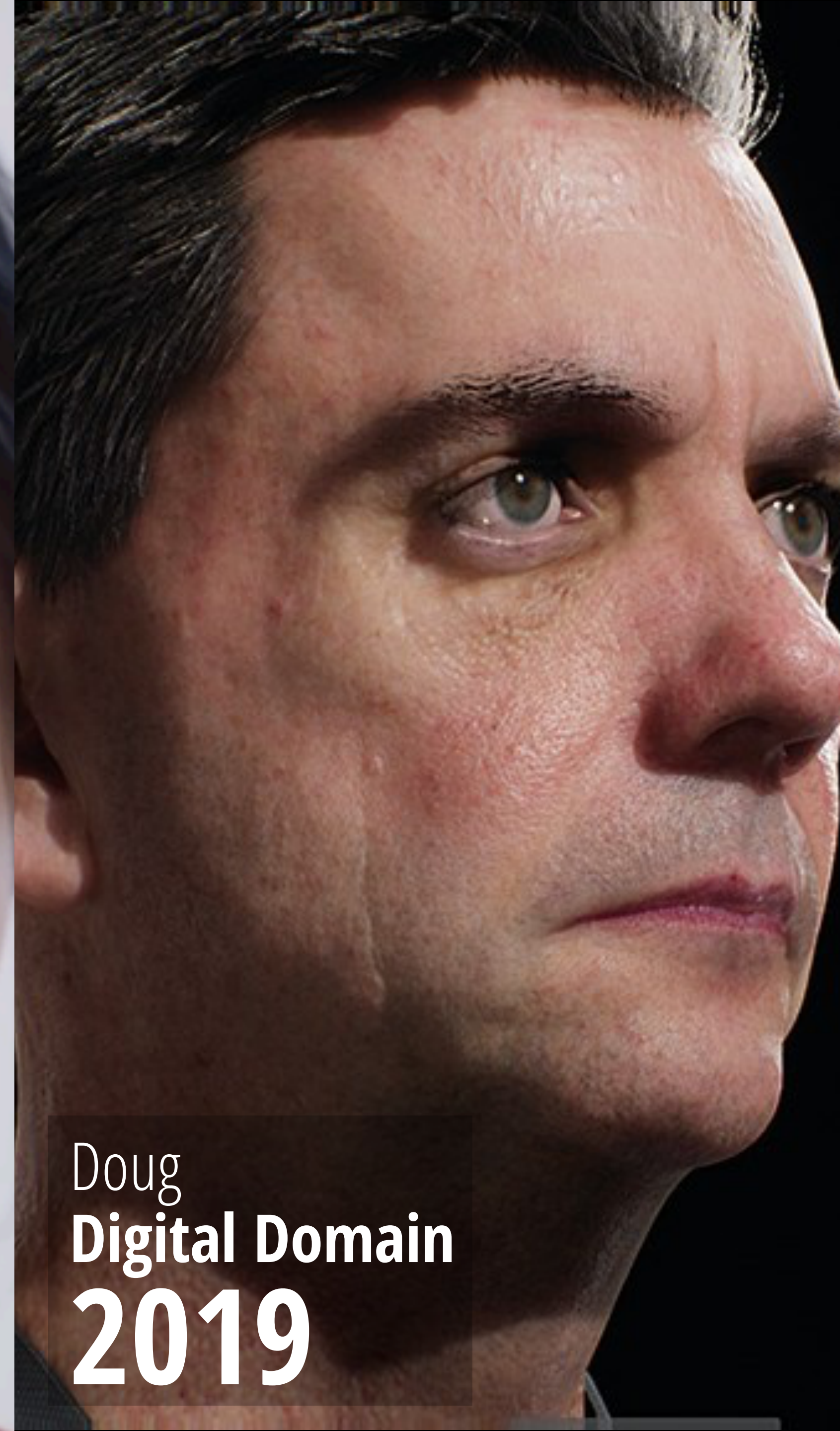




Mike  
**Wikihuman**  
**2017**



Siren  
**Epic Games**  
**2018**



Doug  
**Digital Domain**  
**2019**





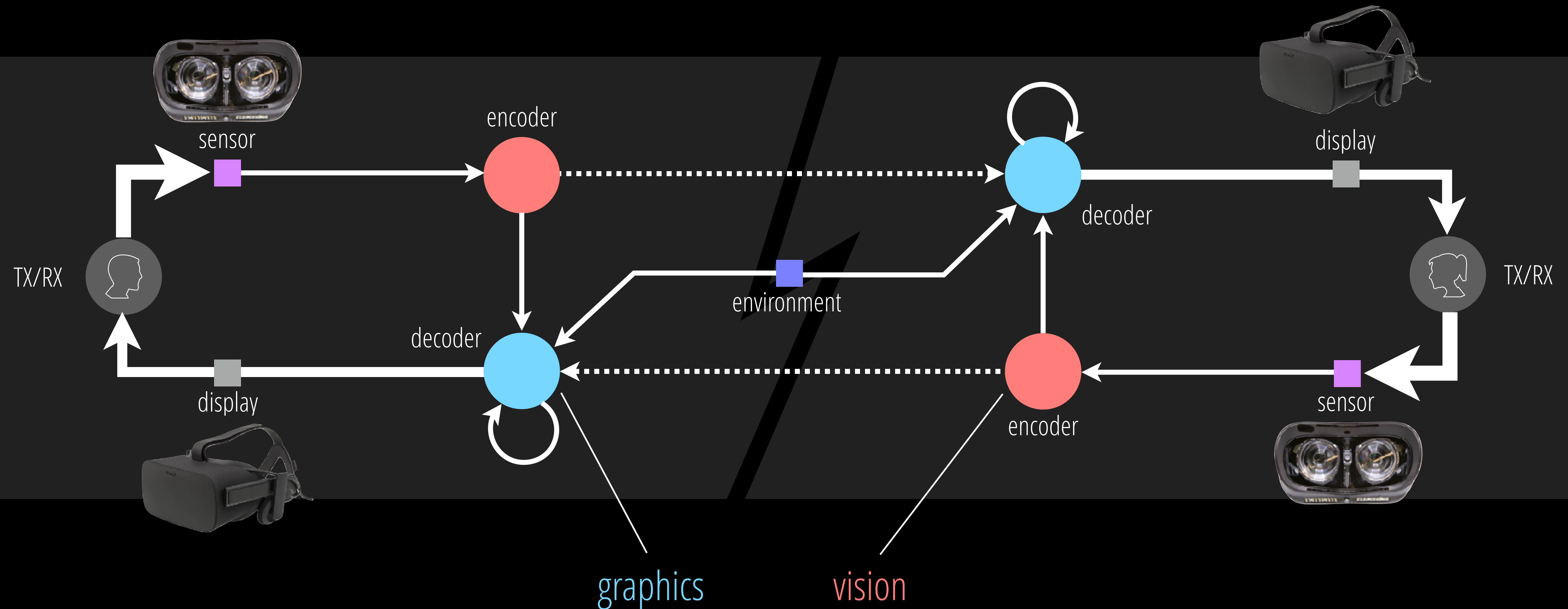






# WHAT IS A CODEC AVATAR?

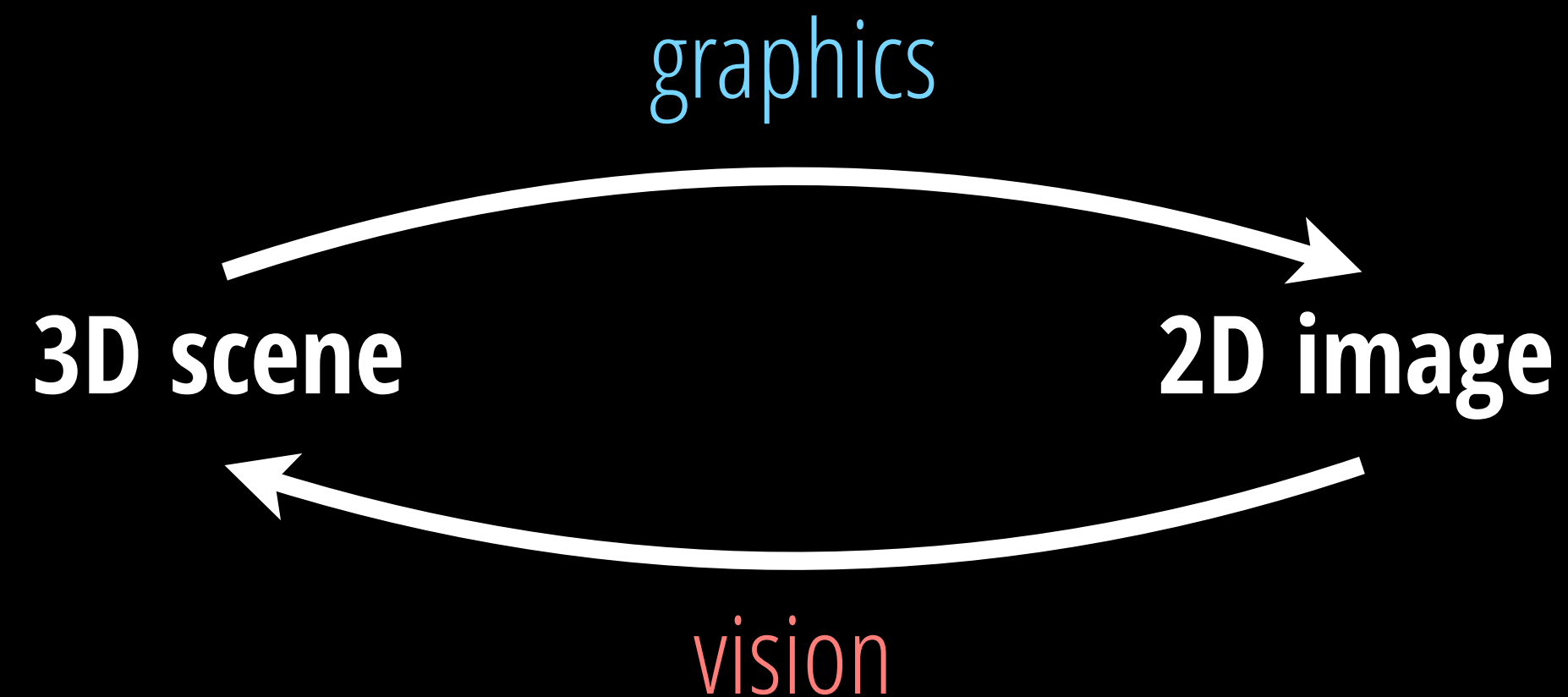
Social Interaction as a Communication Network





# VISUAL COMPUTING PIPELINE

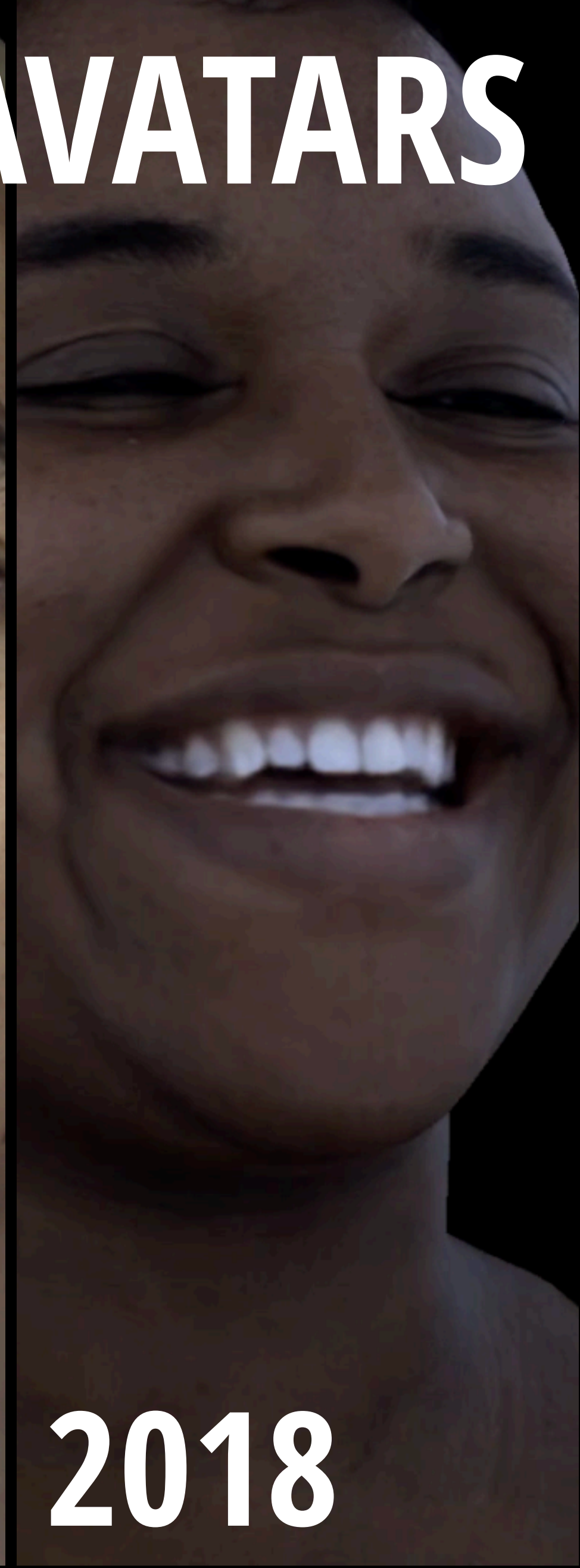
Geometry and Neural Networks Will Unify Vision and Graphics



# HPVC 2030?



# CODEC AVATARS



2017

2018

2018

2019

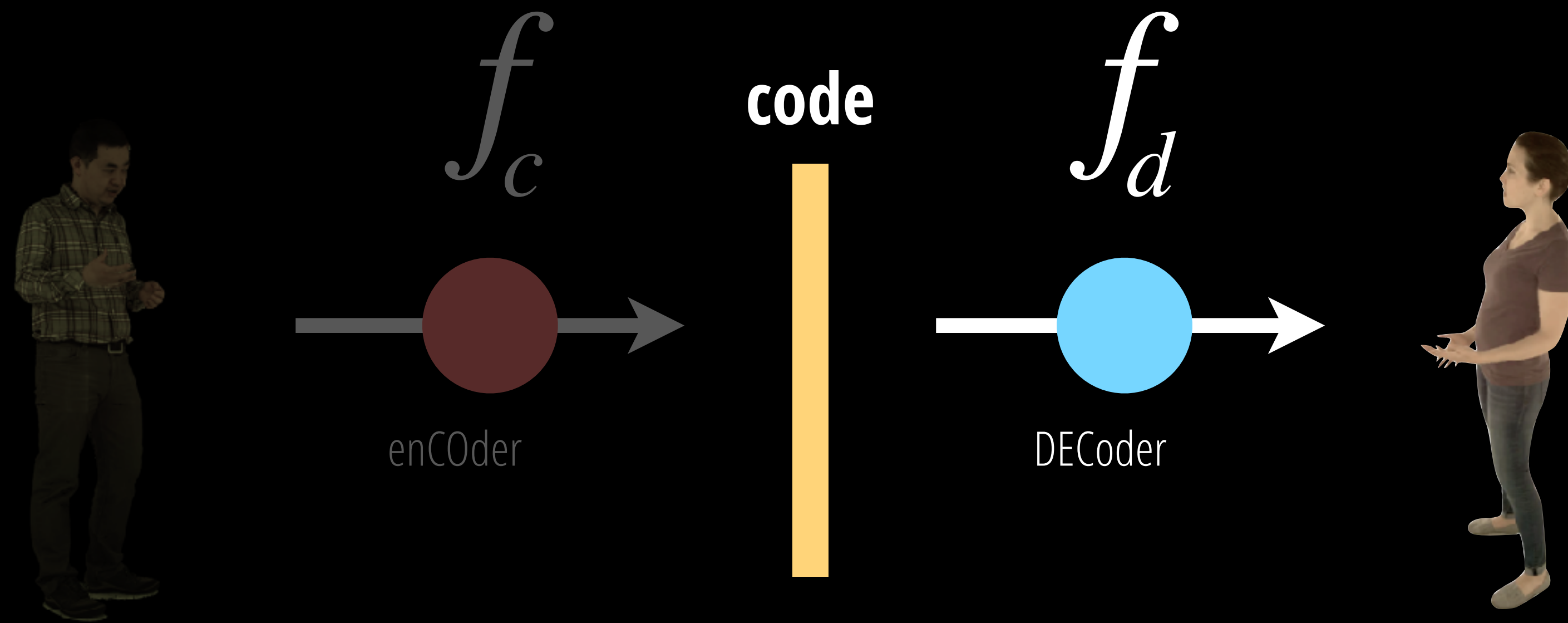
2019

2020



# WHAT IS A CODEC AVATAR?

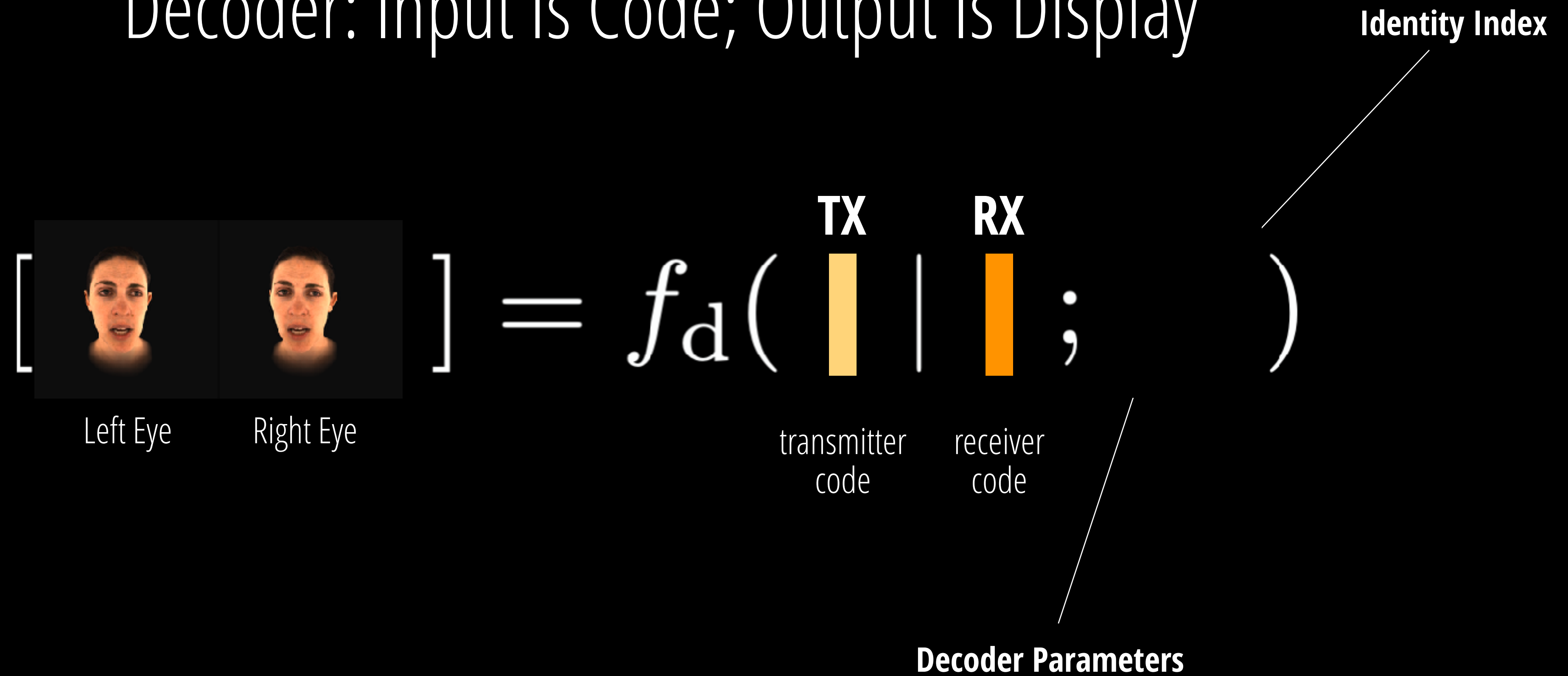
A Codec Avatar Is a Pair of Functions: an Encoder and a Decoder





# WHAT IS A CODEC AVATAR?

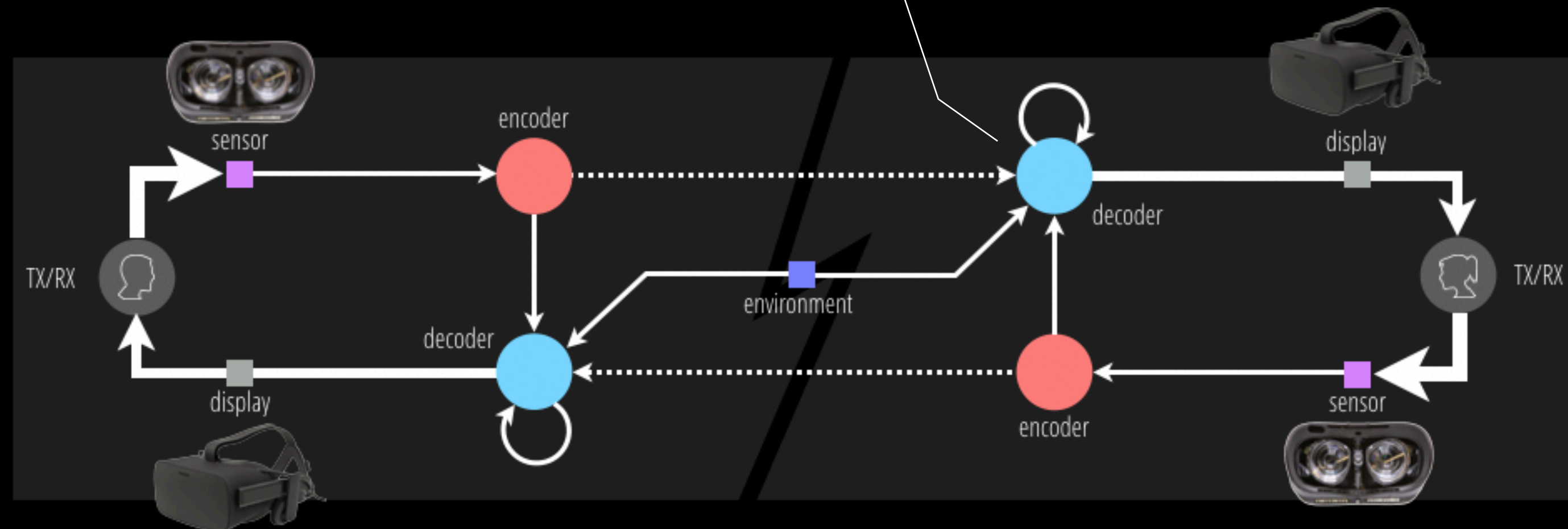
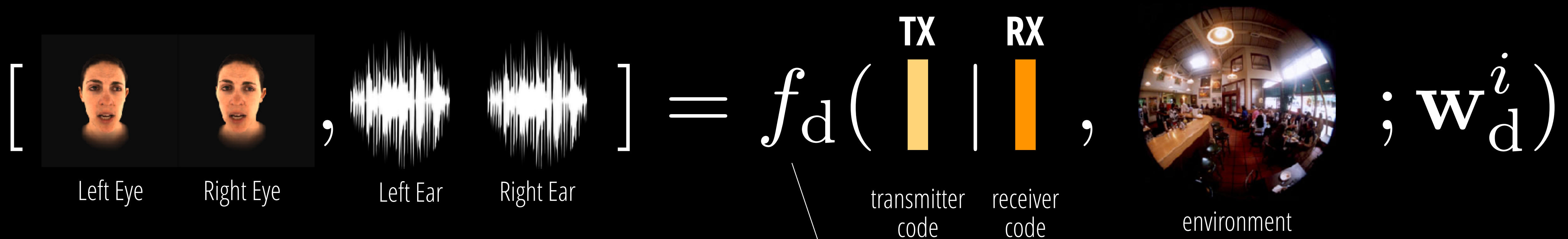
Decoder: Input Is Code; Output Is Display





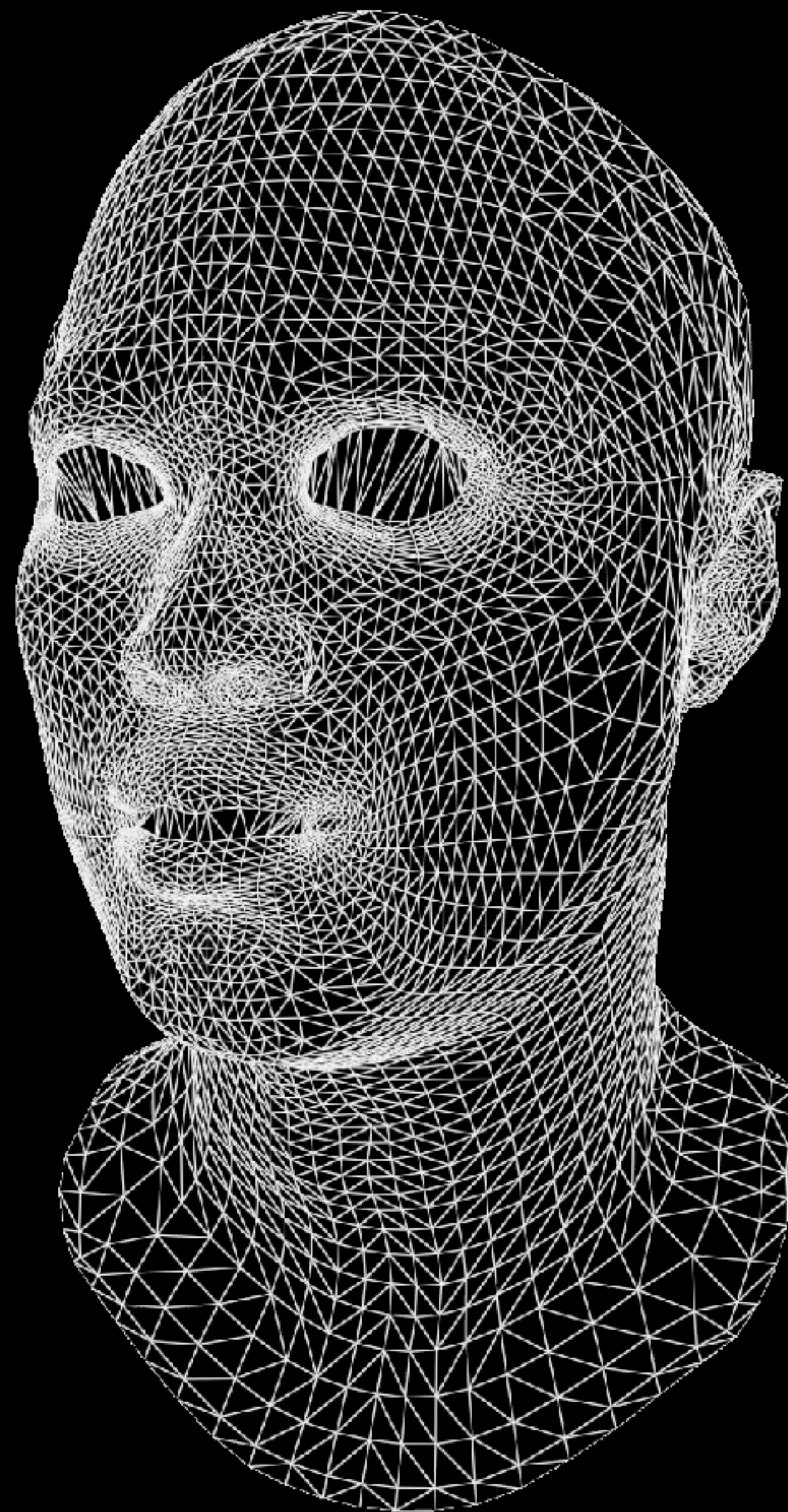
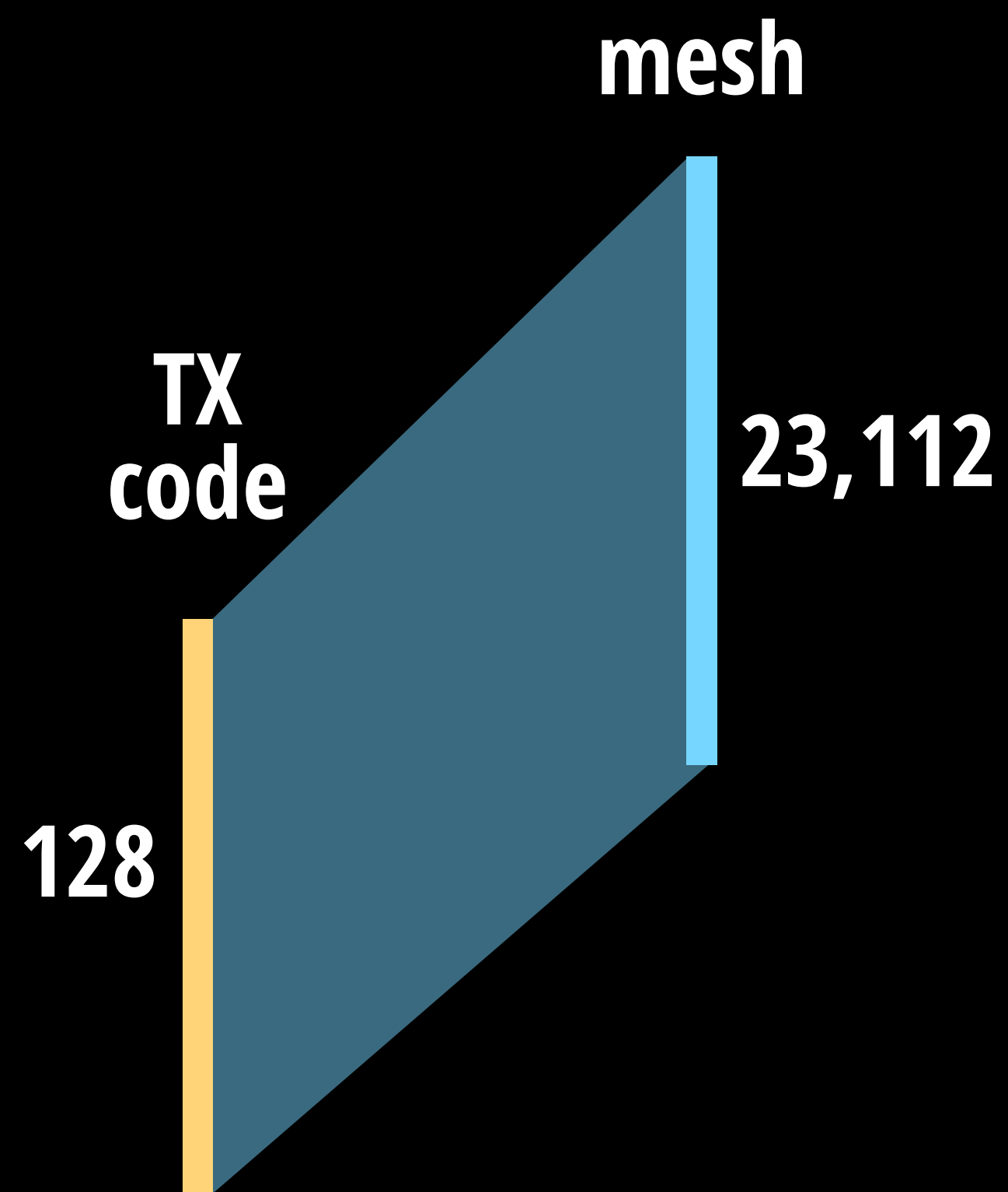
# WHAT IS A CODEC AVATAR?

Decoder: Input Is Code; Output Is Display





The mesh decoder  
is a nonlinear  
compositional (deep)  
function



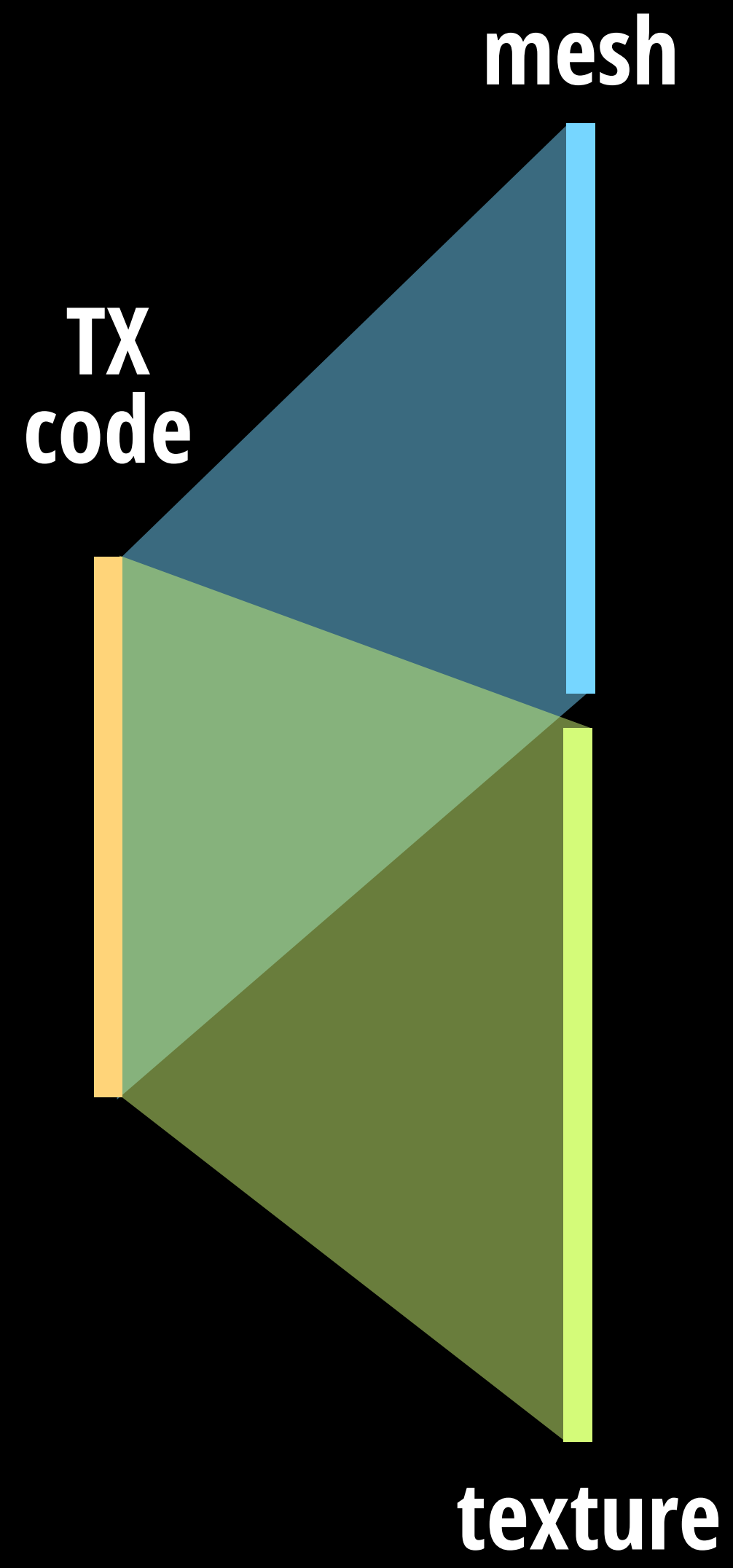


# CODEC AVATARS

Appearance Depends on Expression

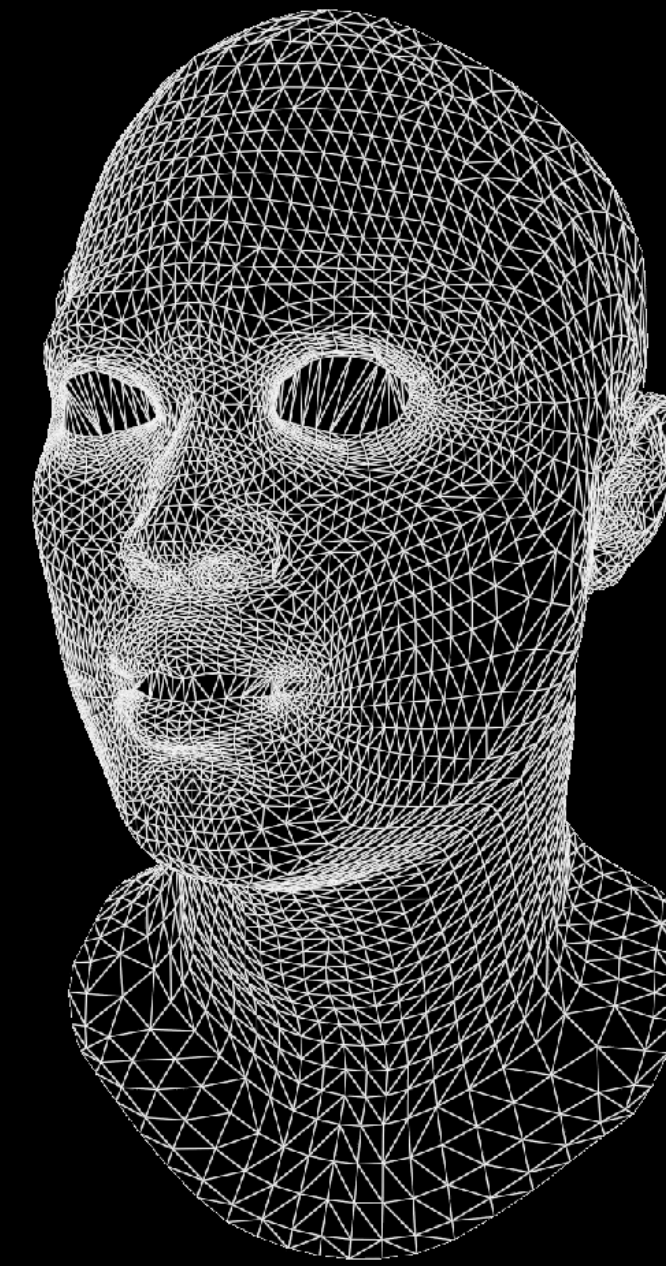
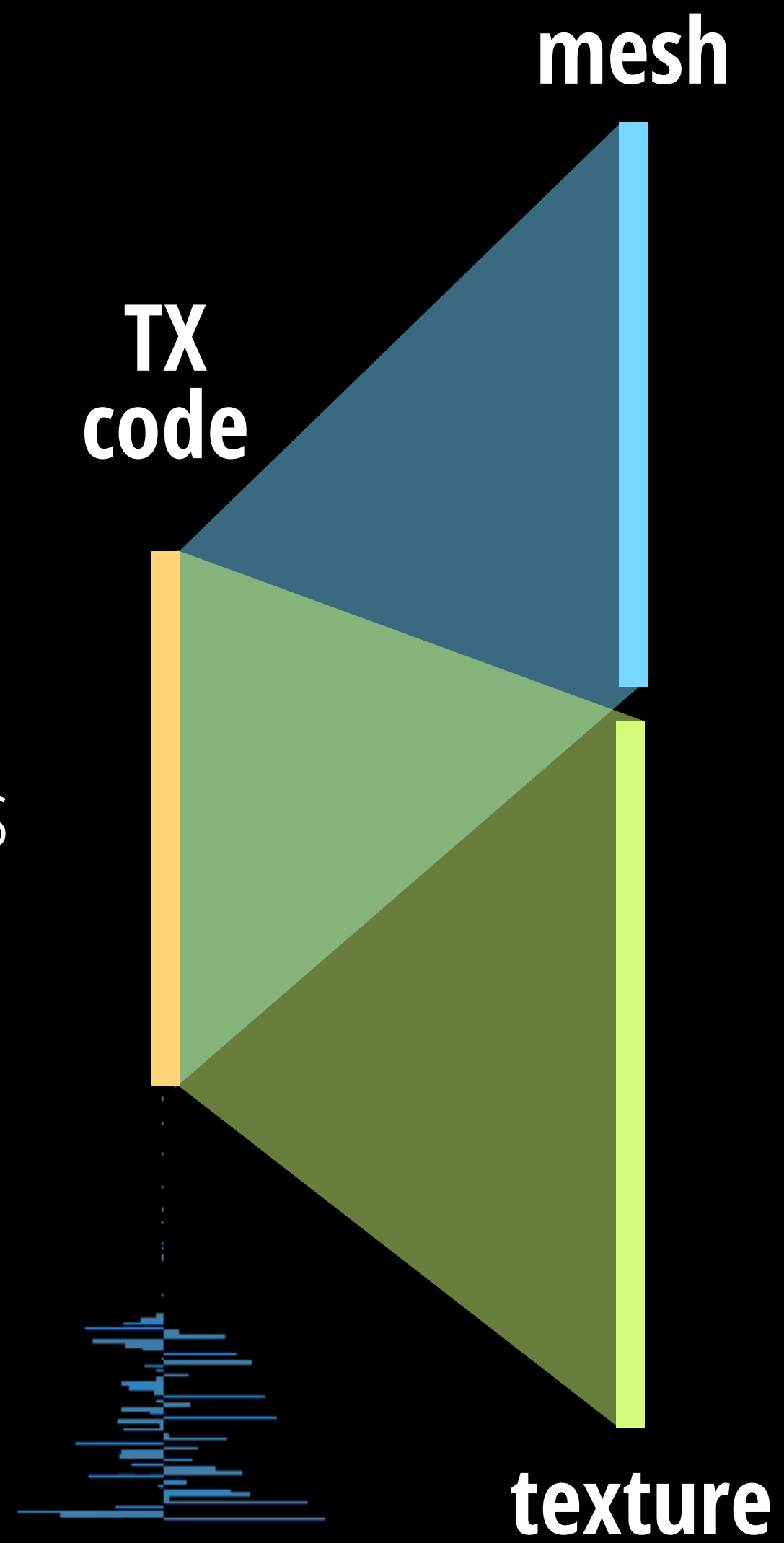






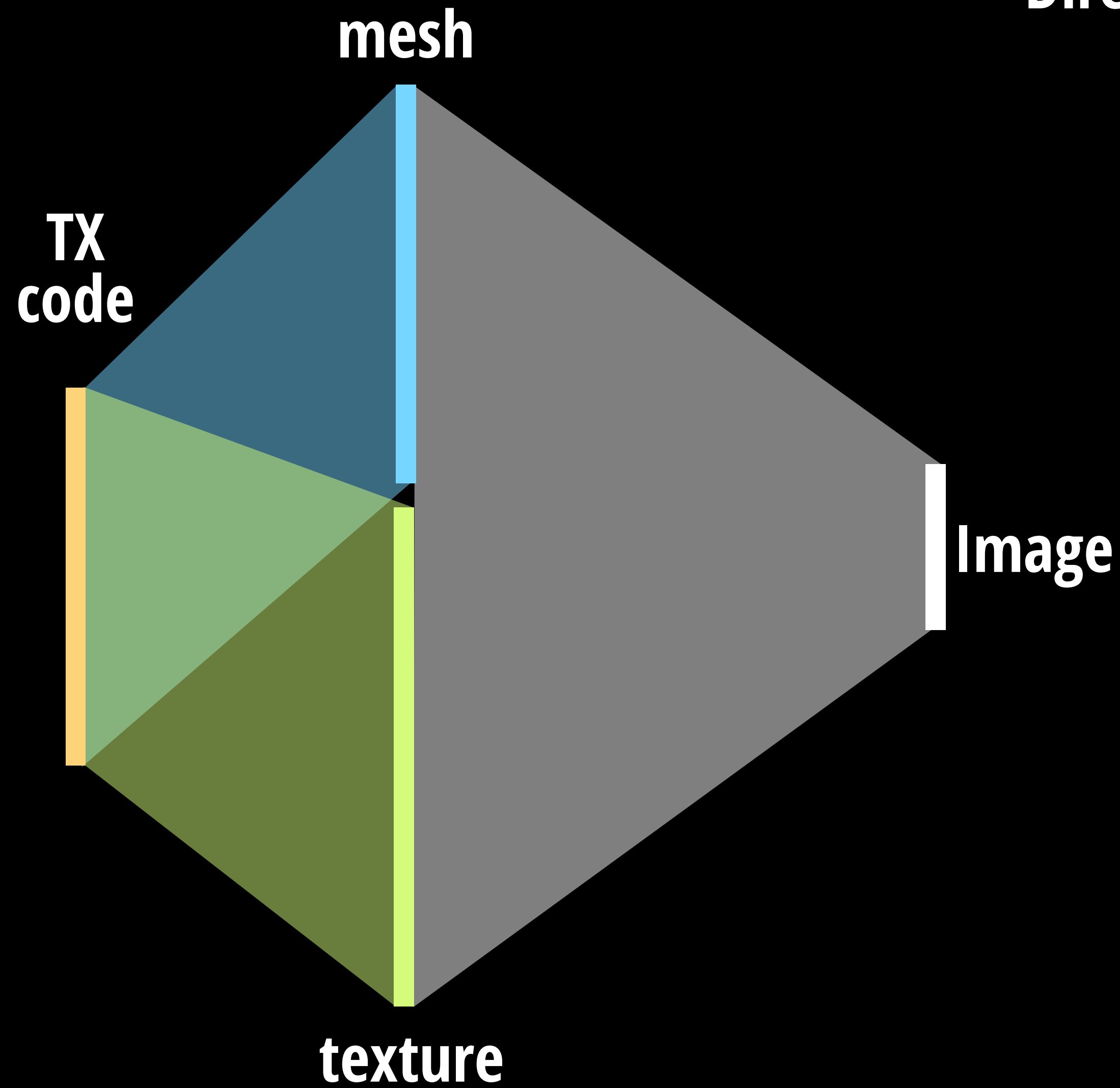


A shared code generates geometry and texture





**METRIC TELEPRESENCE**  
Directly minimize the difference  
from real image pixels





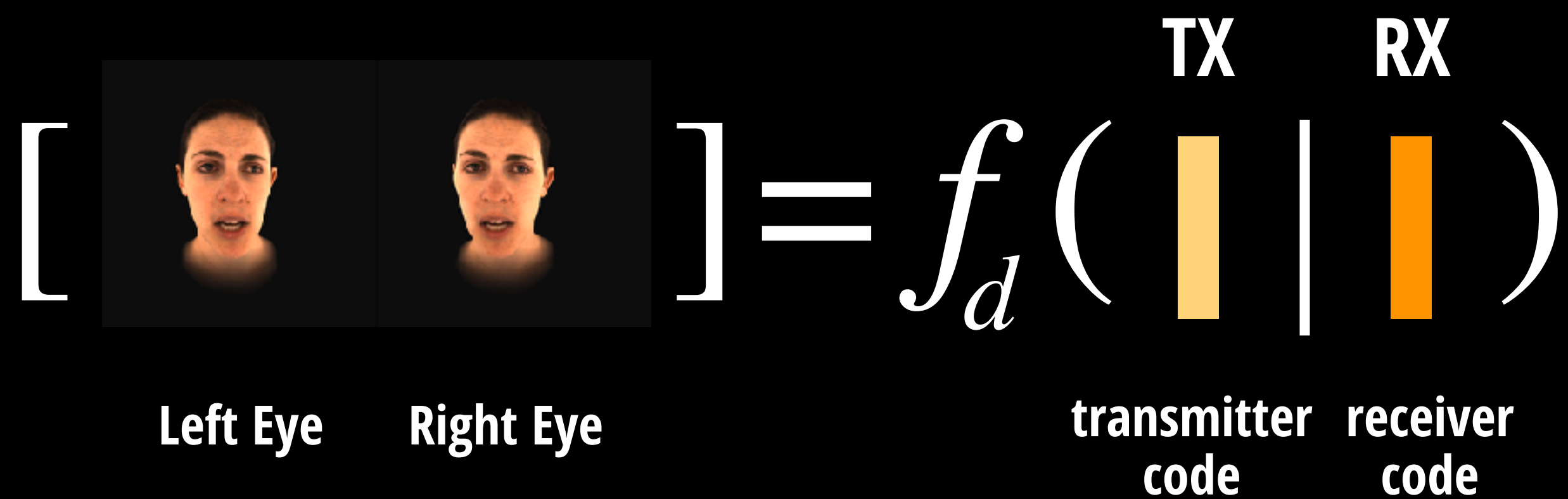
code



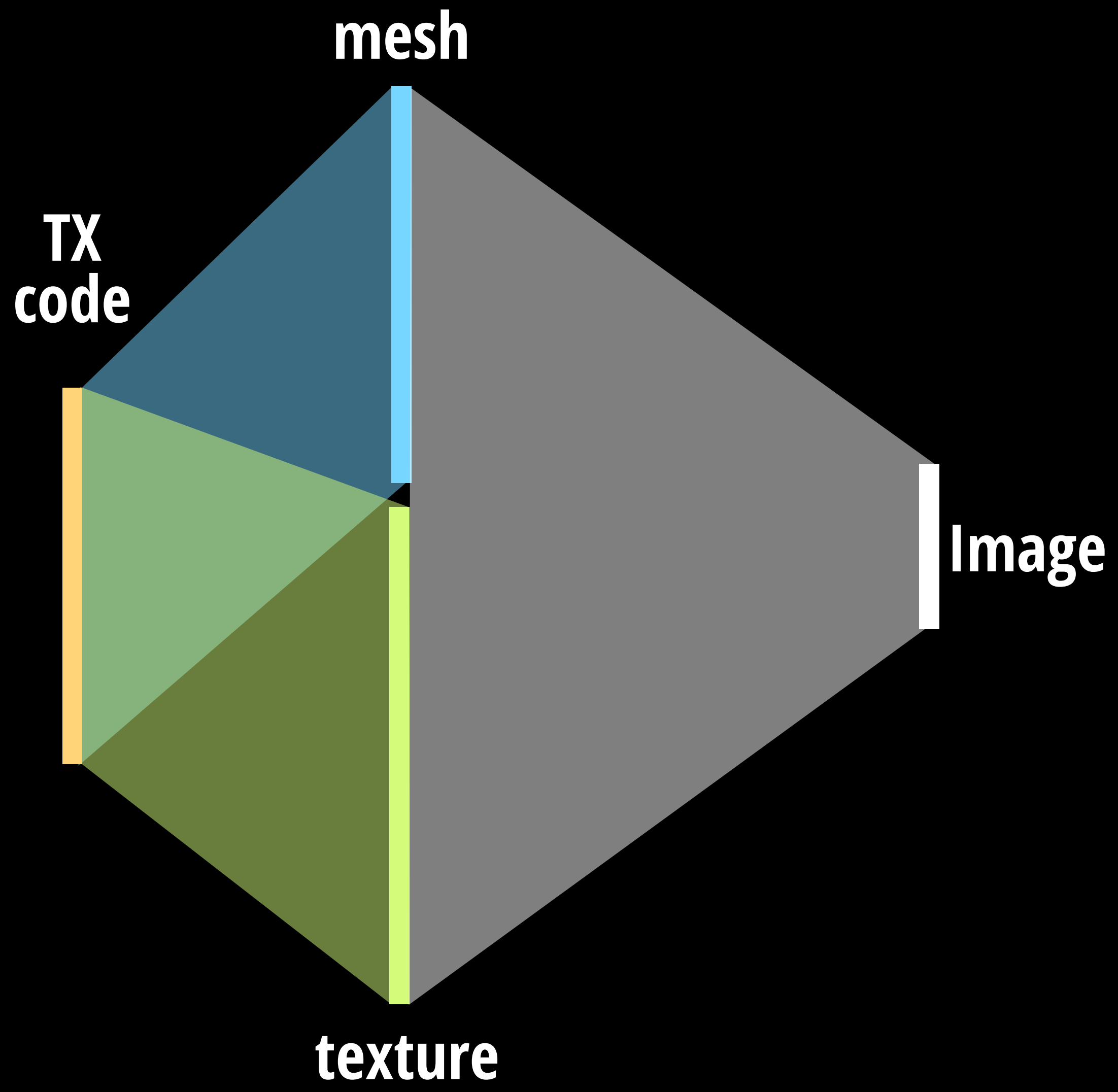


# WHAT ARE CODEC AVATARS?

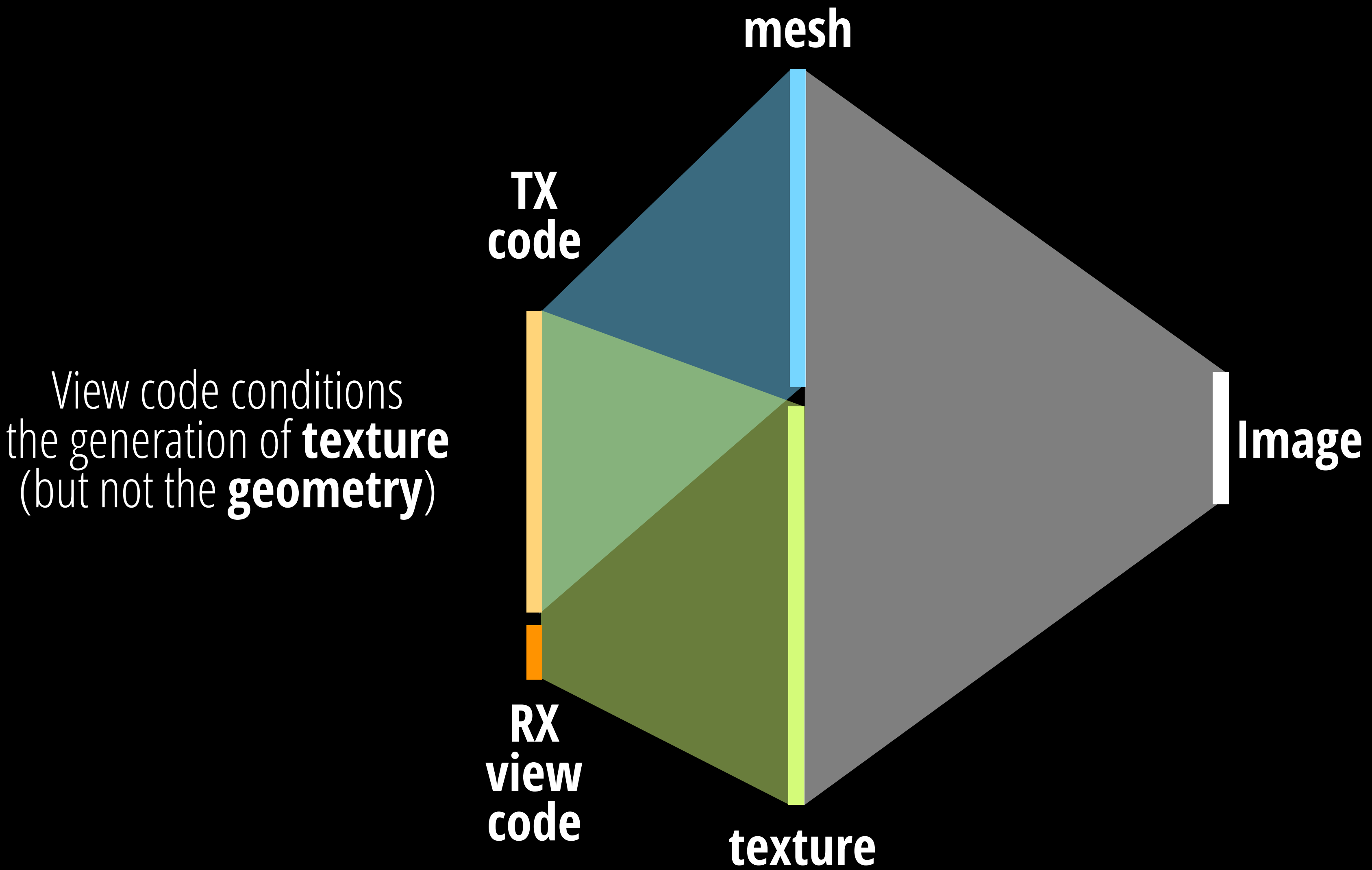
Appearance Depends on **Receiver's** Viewpoint













**view 1**



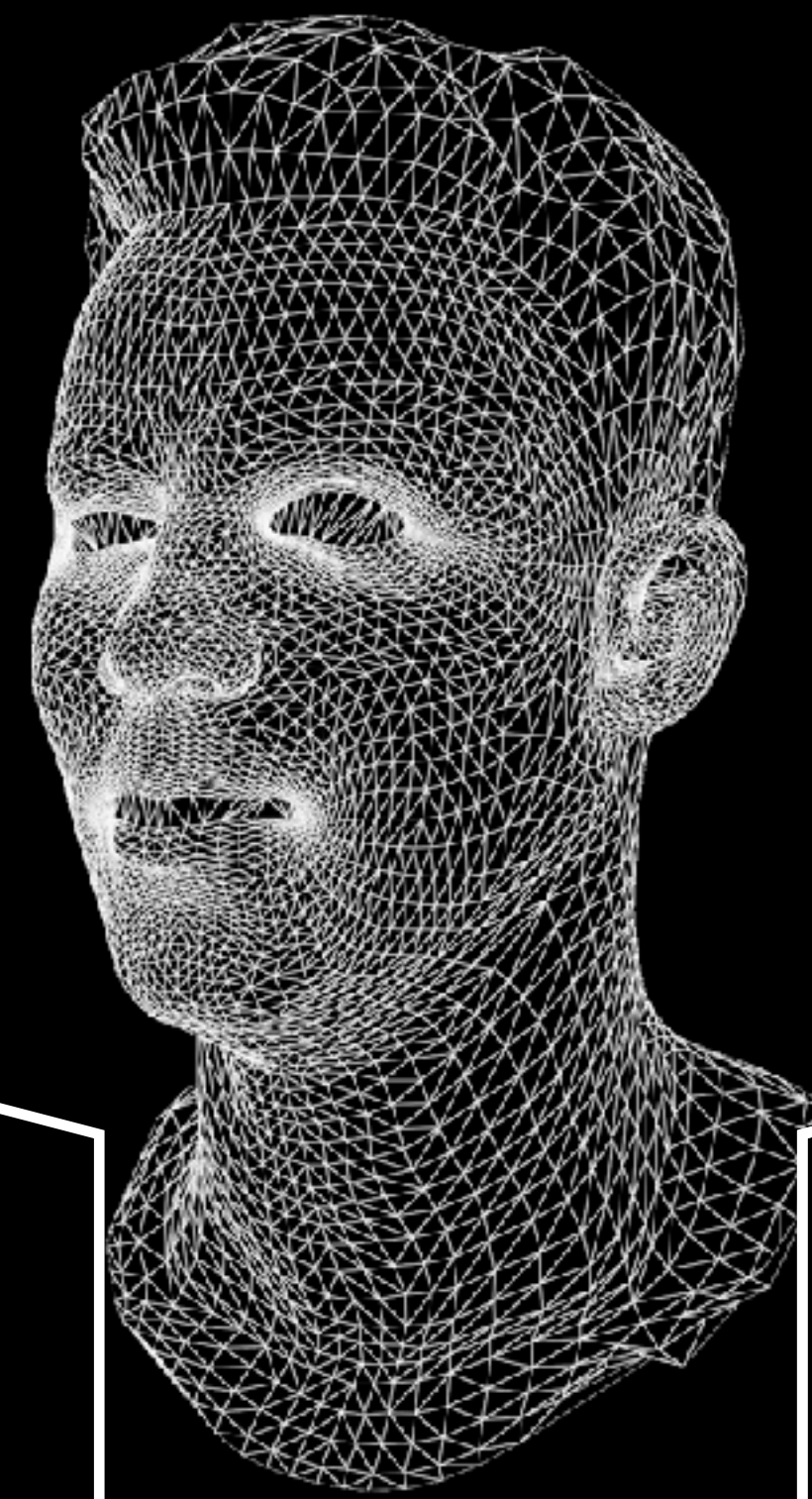
**view 2**



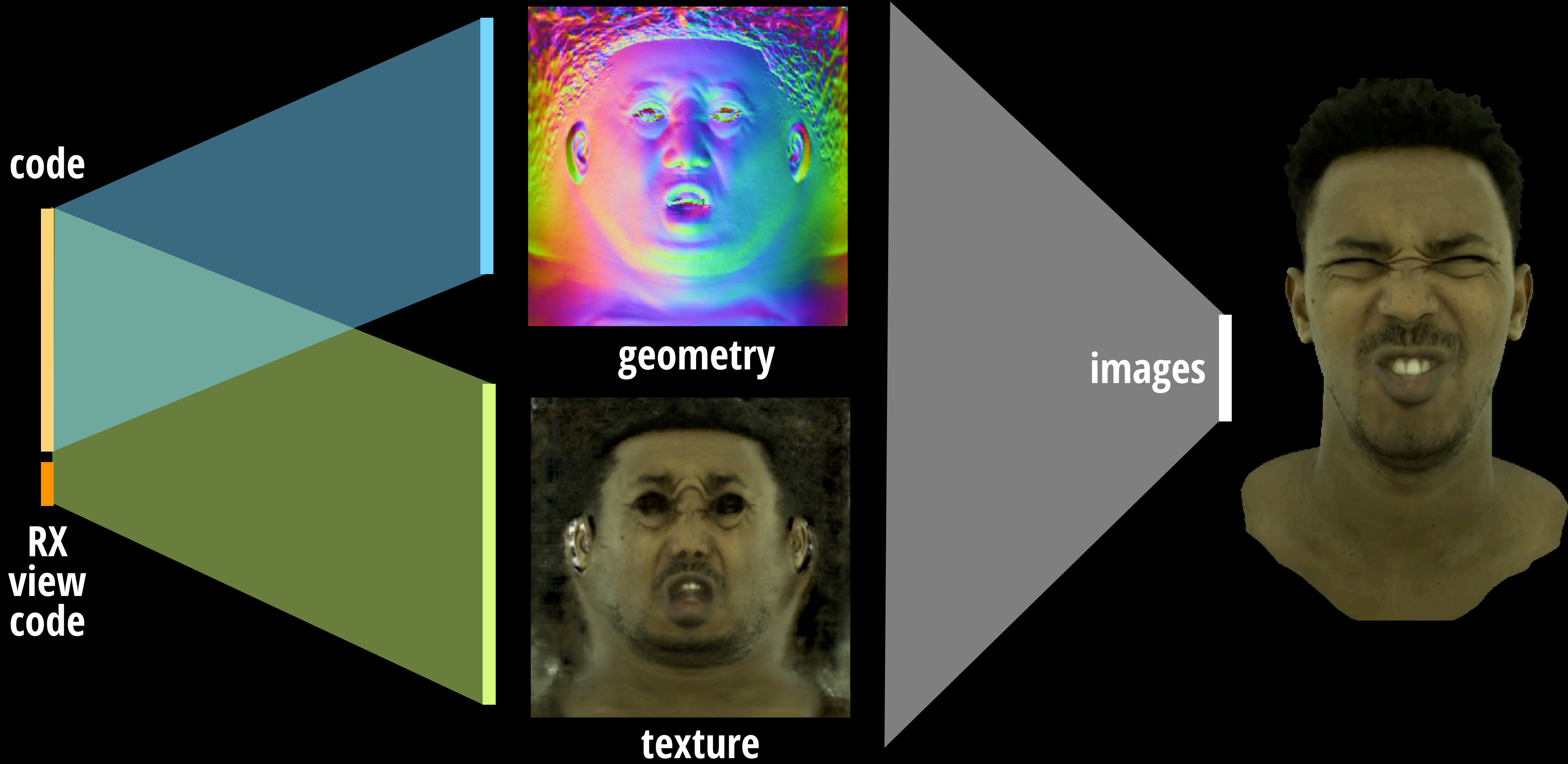
**view 3**



**view 4**









# WHY SO MANY CAMERAS?

"Measure what is measurable. Make measurable what is not." (Galileo)



100 Microphones

160 Cameras

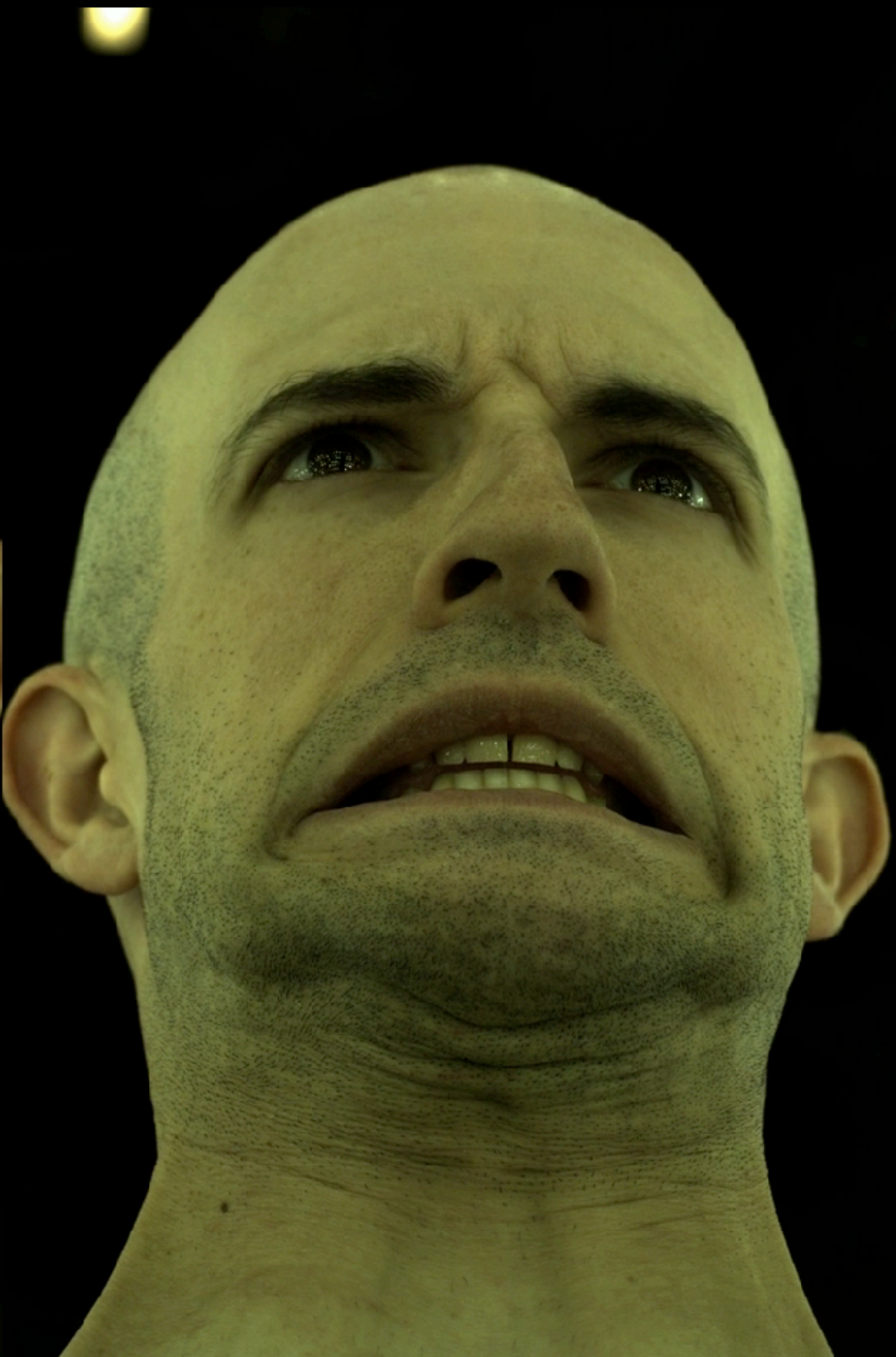
450 Lights

# Mugsy

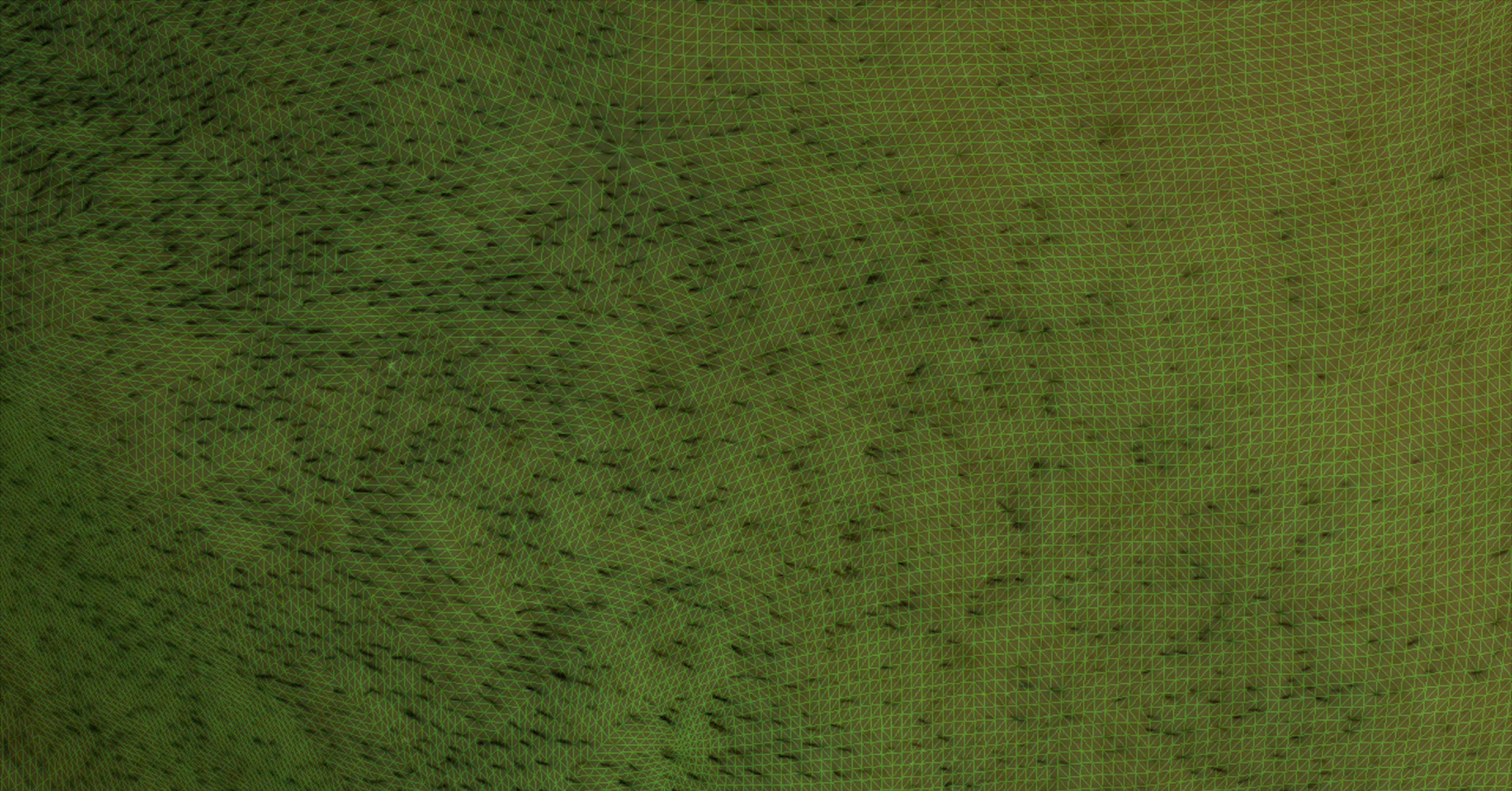






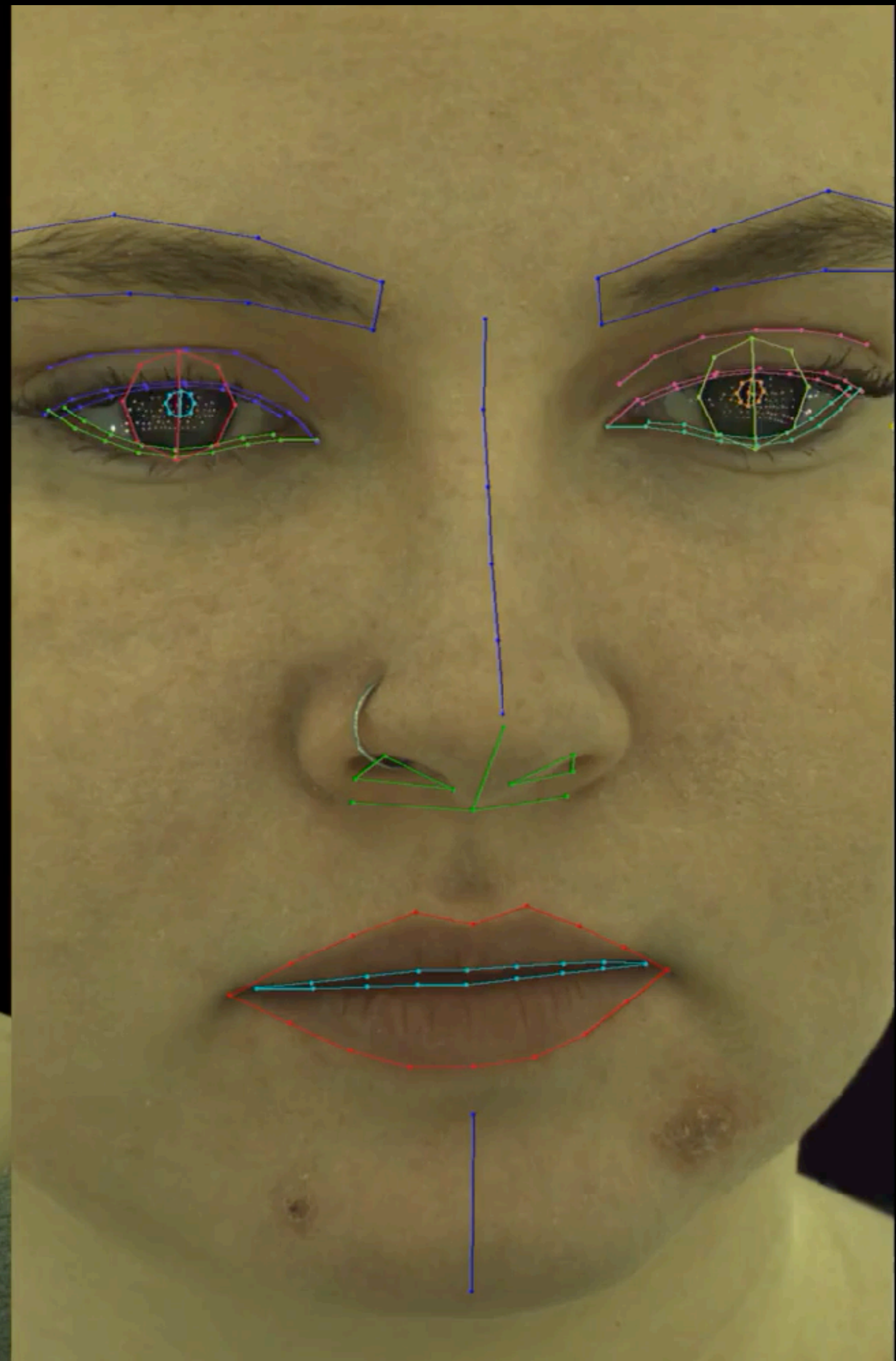






[Wu, Shiratori, Sheikh, "Deep incremental learning for efficient high-fidelity face tracking," SIGGRAPH Asia 2018]







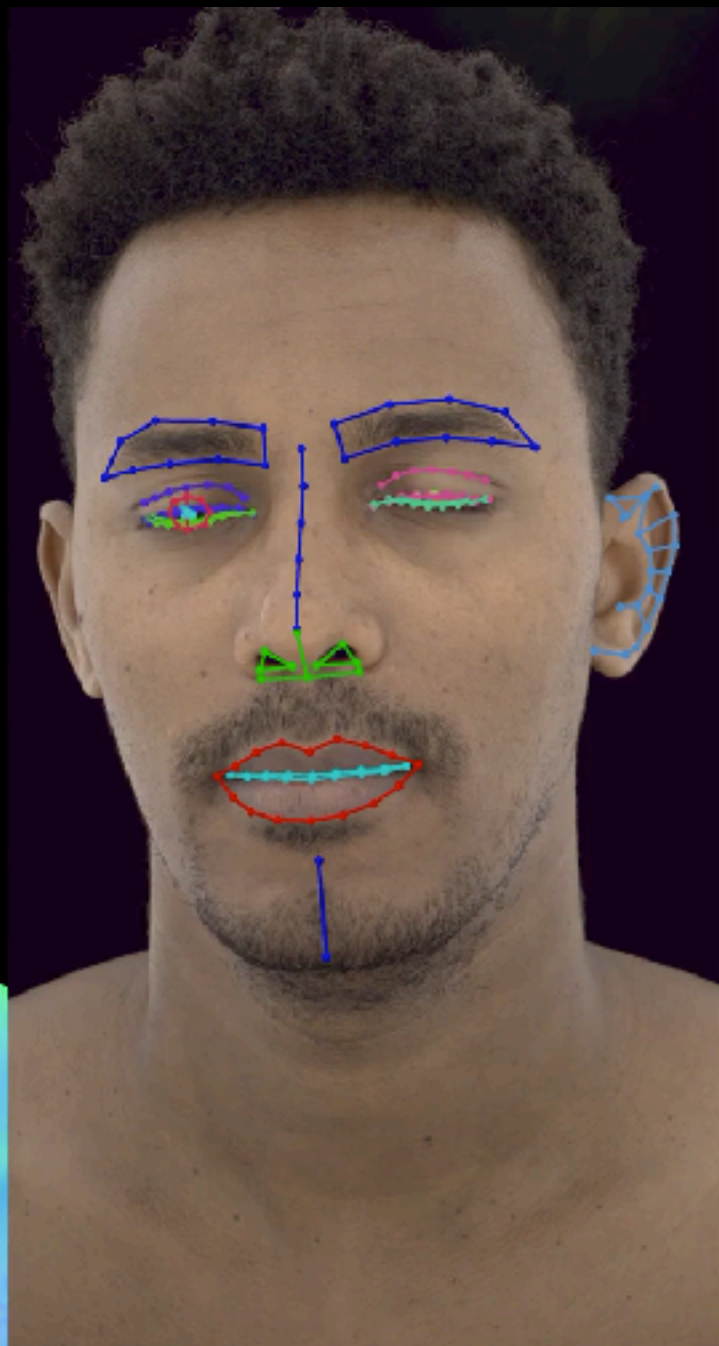
# LOCAL AVATAR PIPELINE PROCESSING STATUS



Original Image



3D Reconstruction



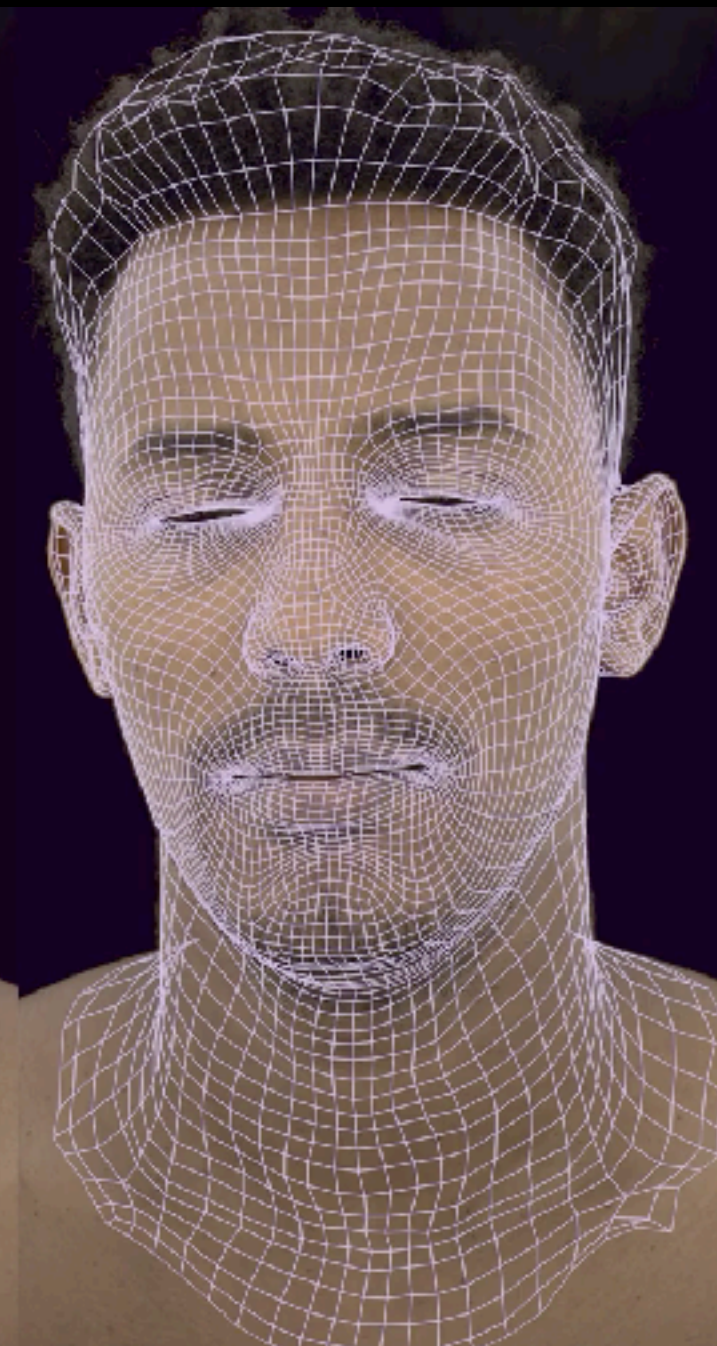
Keypoint Detection



Model-free  
Mesh Tracking



Personalized  
keypoint detection

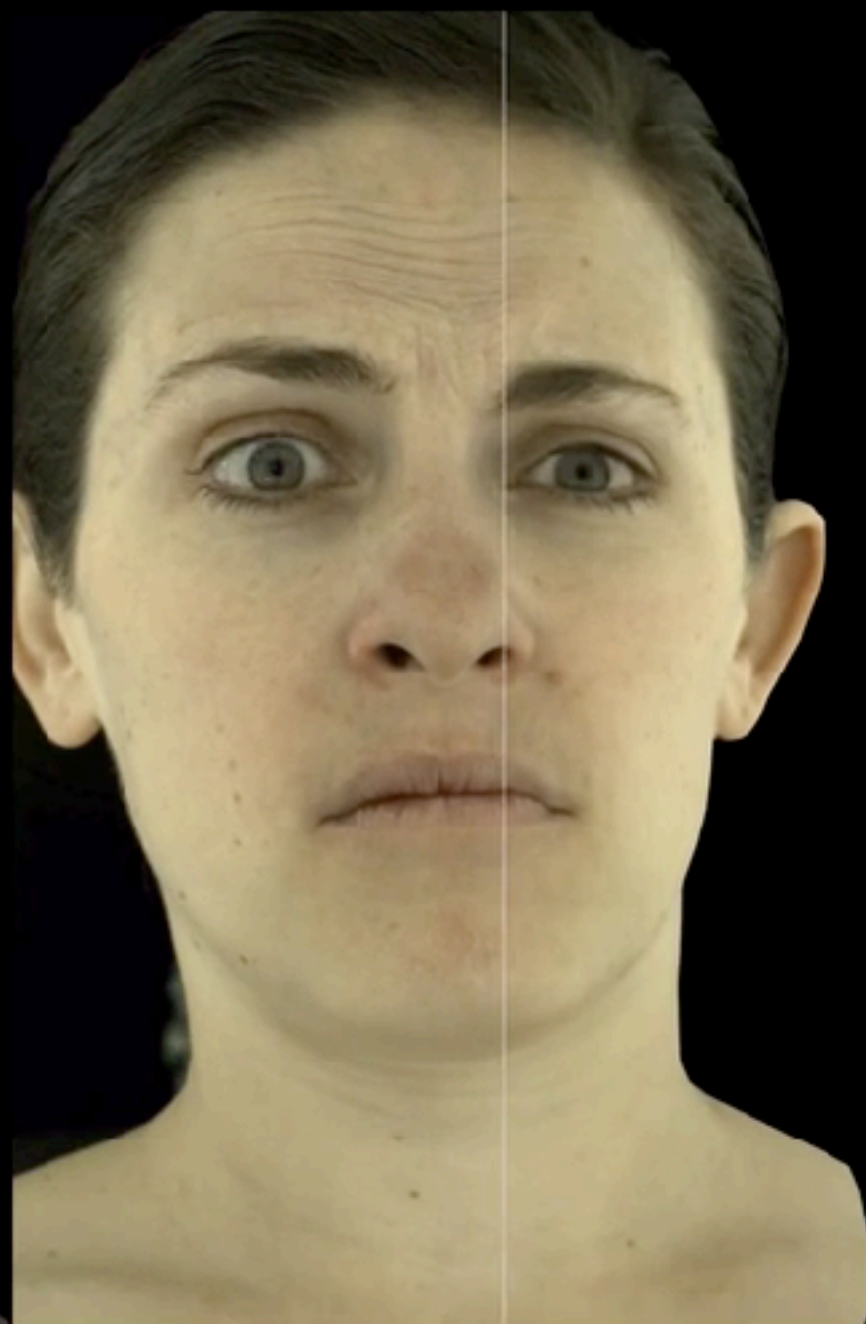
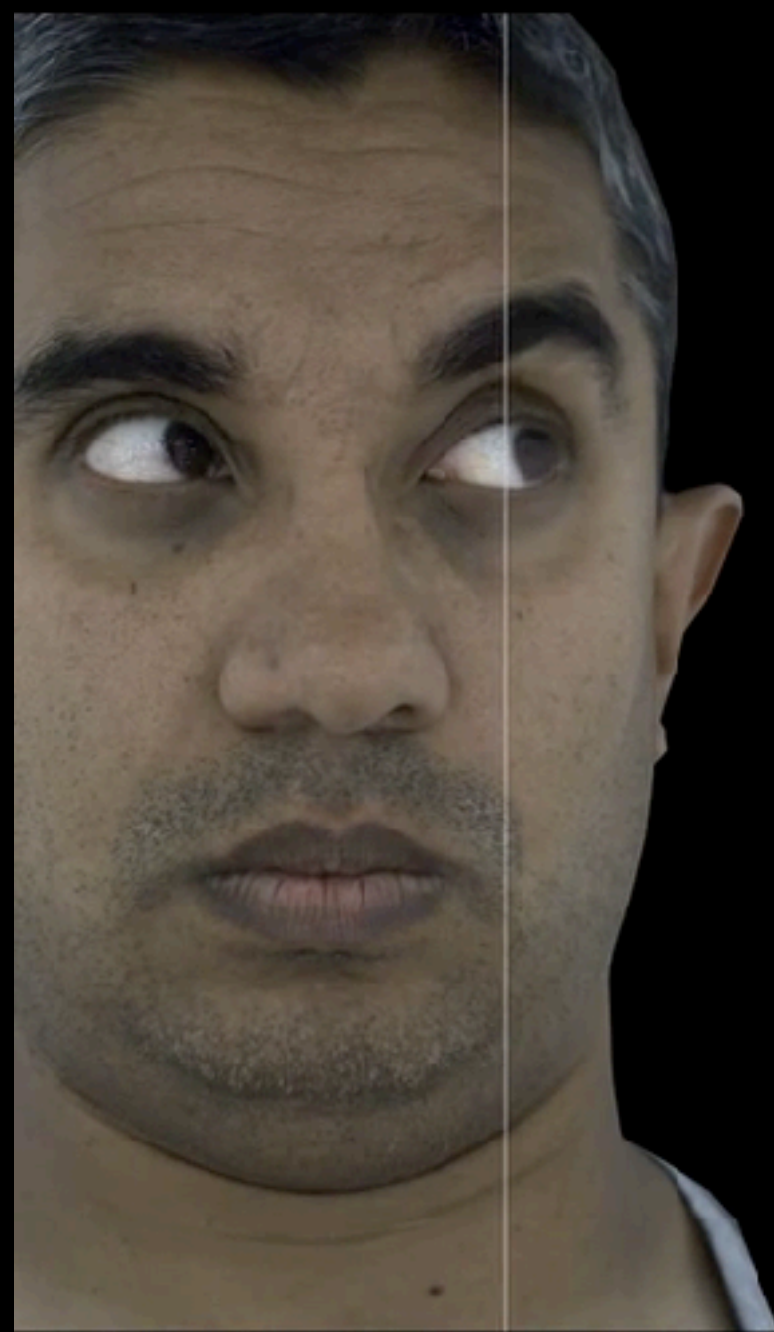


Model-based  
Mesh Tracking



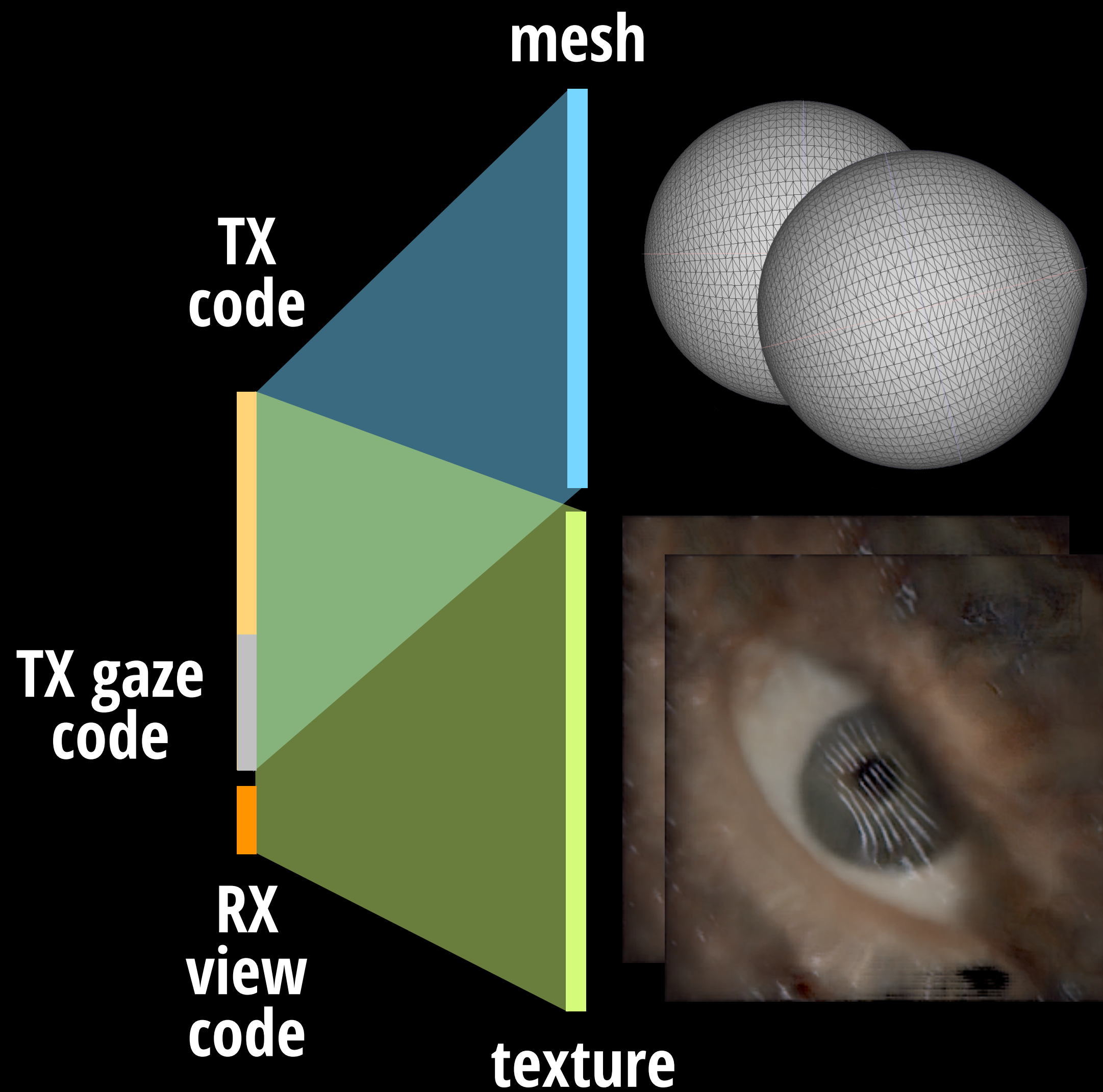
Avatar Decoder







Schwartz et al. "The Eyes Have It: An Integrated Eye and Face Model for Photorealistic Facial Animation," SIGGRAPH 2020

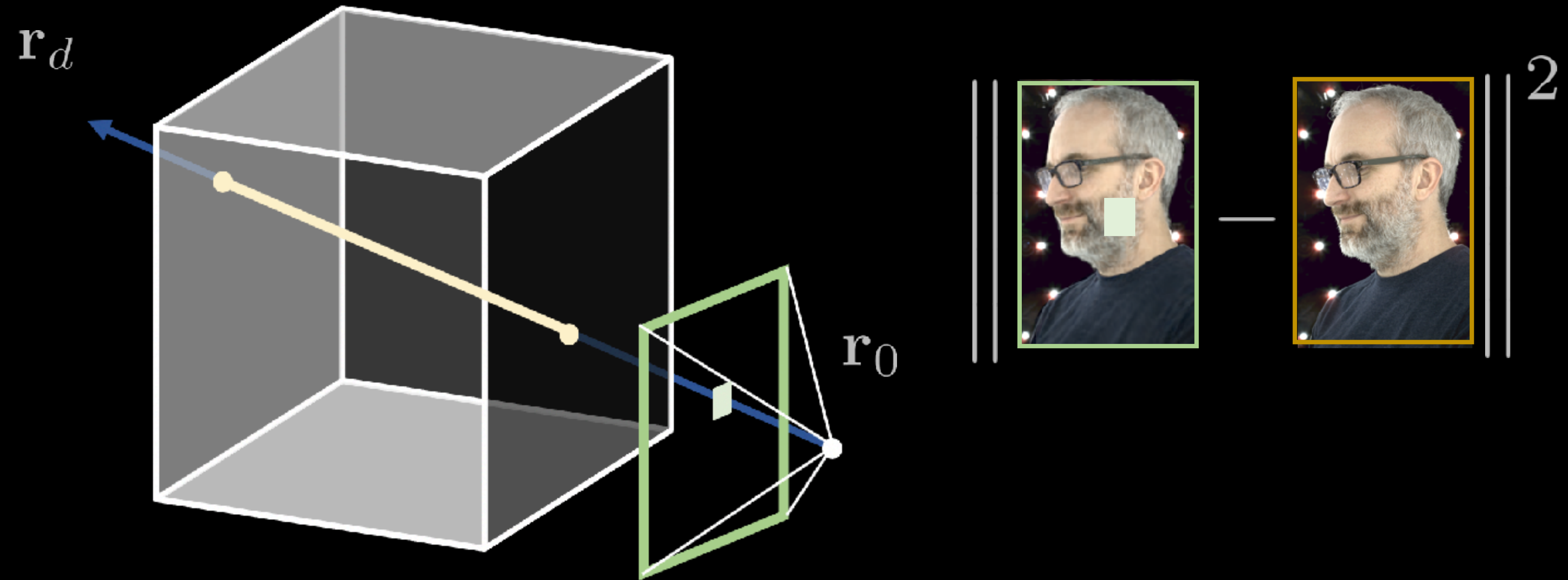




# VOLUMETRIC CODEC AVATARS

Neural Volumes

Volume Ray Marching





# VOLUMETRIC CODEC AVATARS

Neural Volumetric Rendering



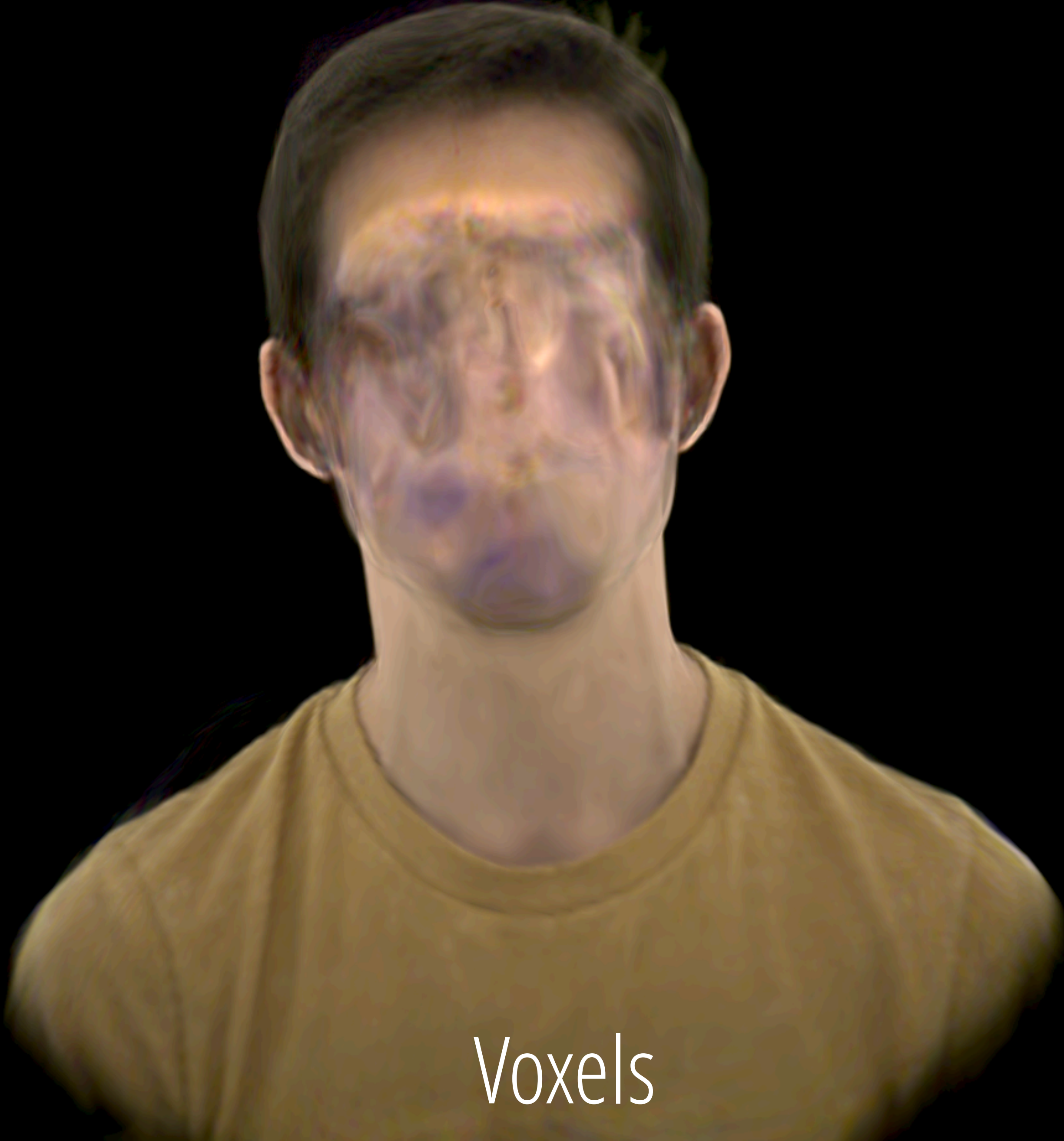
[Lombardi et al. "Neural Volumes: Learning Dynamic Renderable Volumes from Images," SIGGRAPH 2019]



# Hybrid Rendering with 3D Meshes



Mesh



Voxels



# Metric Identity

Identity preserving avatars for billions of people







# Metric Identity

How do we produce identity preserving avatars for billions of people?



# Metric Identity

How do we produce identity preserving avatars for billions of people?

## Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

## Metric Time

How do we do all this in realtime without access to artistic correction?



# Metric Identity

How do we produce identity preserving avatars for billions of people?

# Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

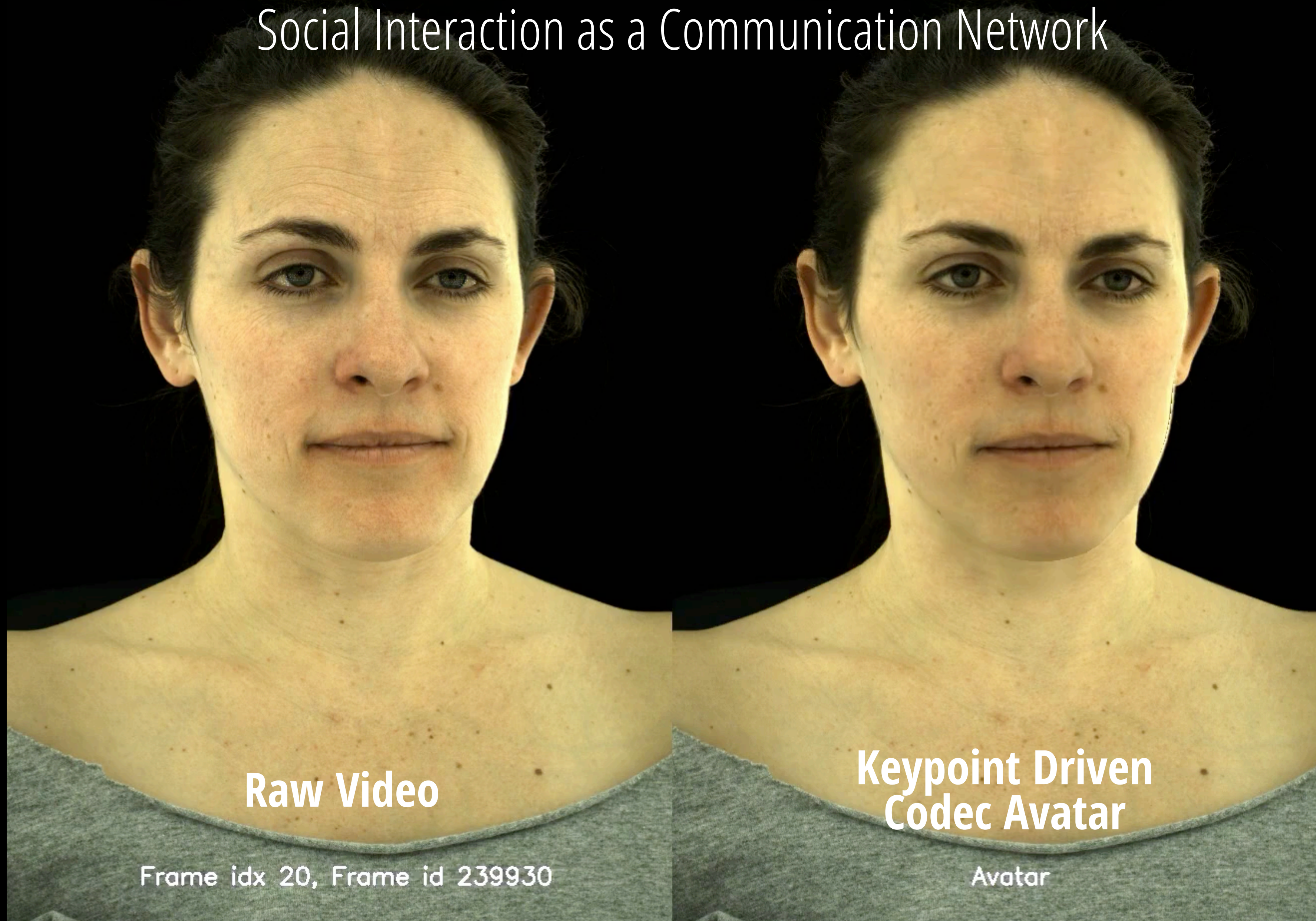
# Metric Time

How do we do all this in realtime without access to artistic correction?



# TRUTHFUL TELEPRESENCE

Social Interaction as a Communication Network



Raw Video

Frame idx 20, Frame id 239930

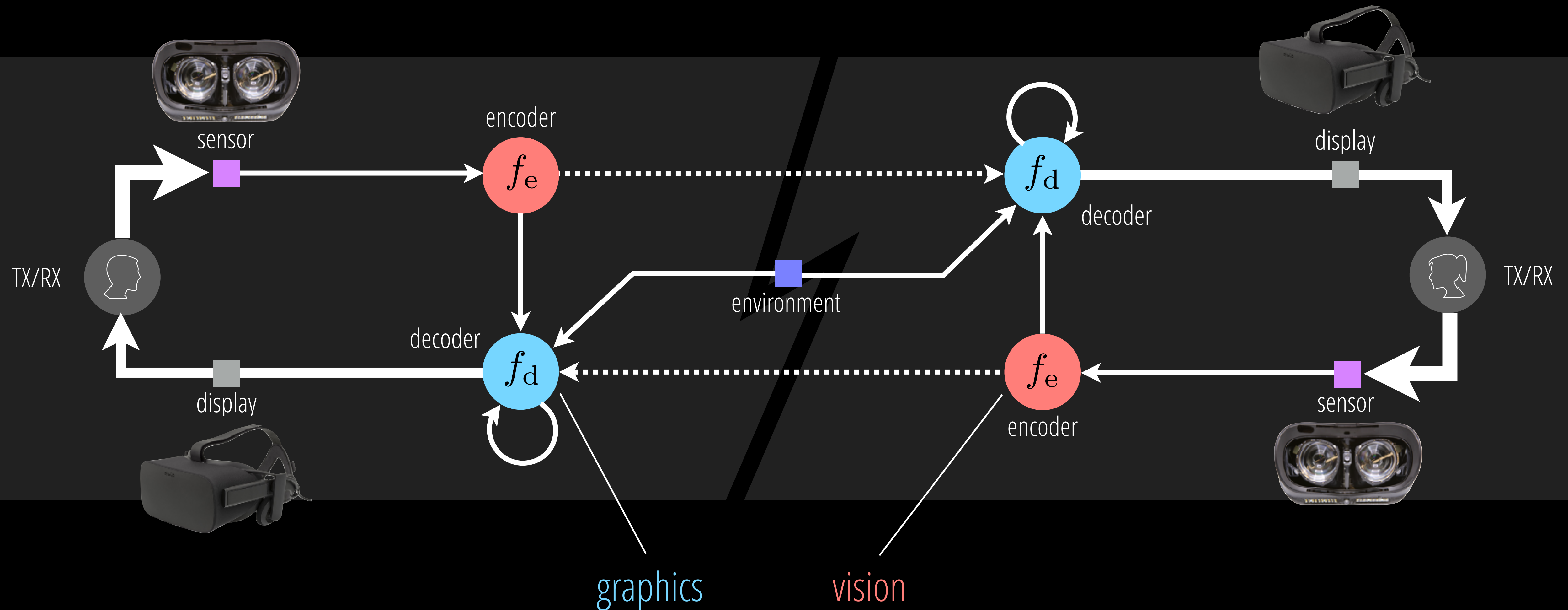
Keypoint Driven  
Codec Avatar

Avatar



# WHAT IS A CODEC AVATAR?

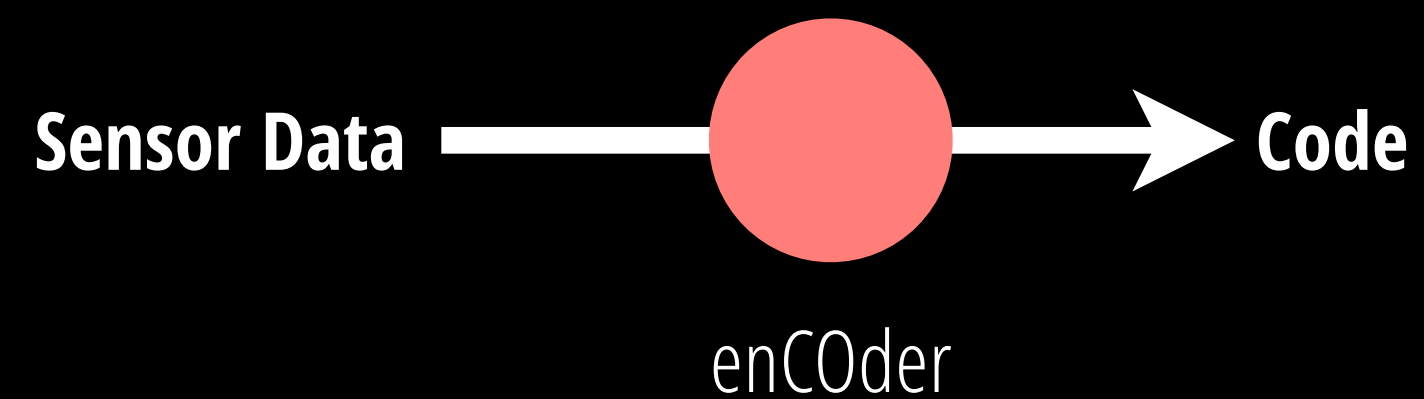
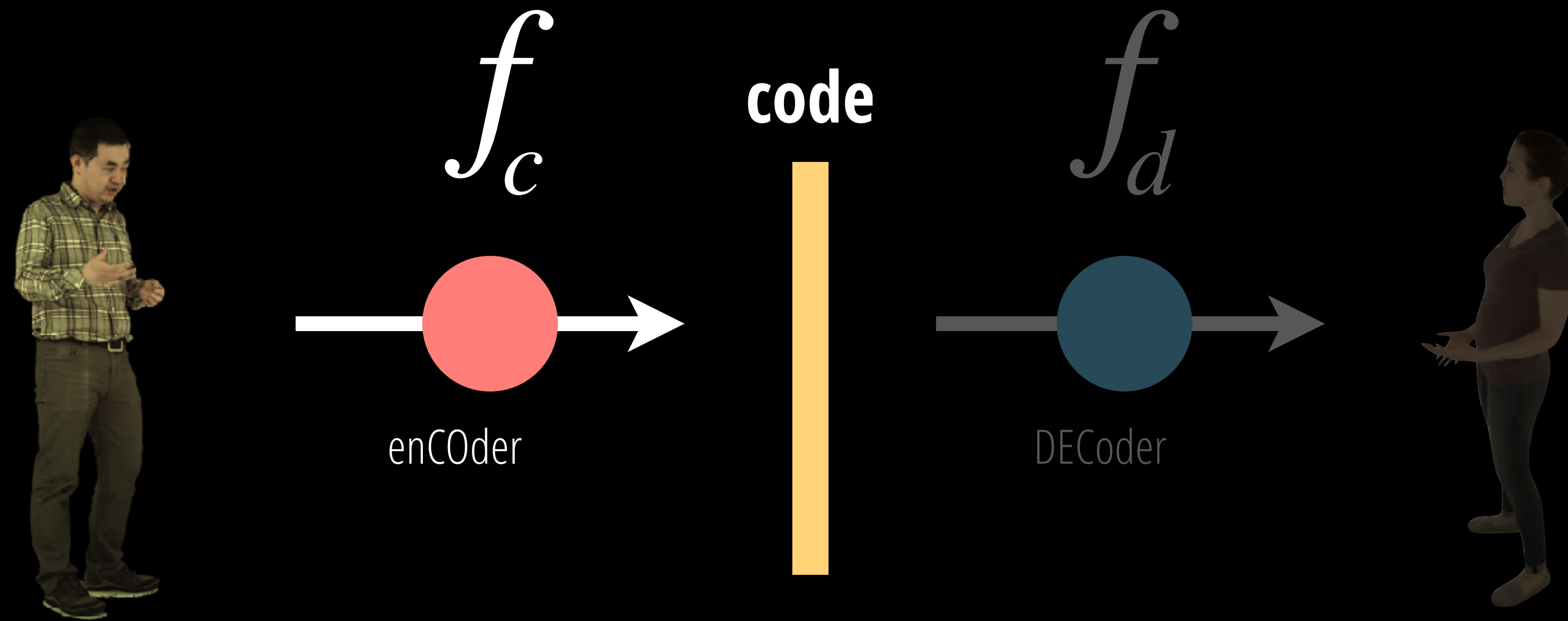
Social Interaction as a Communication Network





# WHAT IS A CODEC AVATAR?

A Codec Avatar Is a Pair of Functions: an Encoder and a Decoder







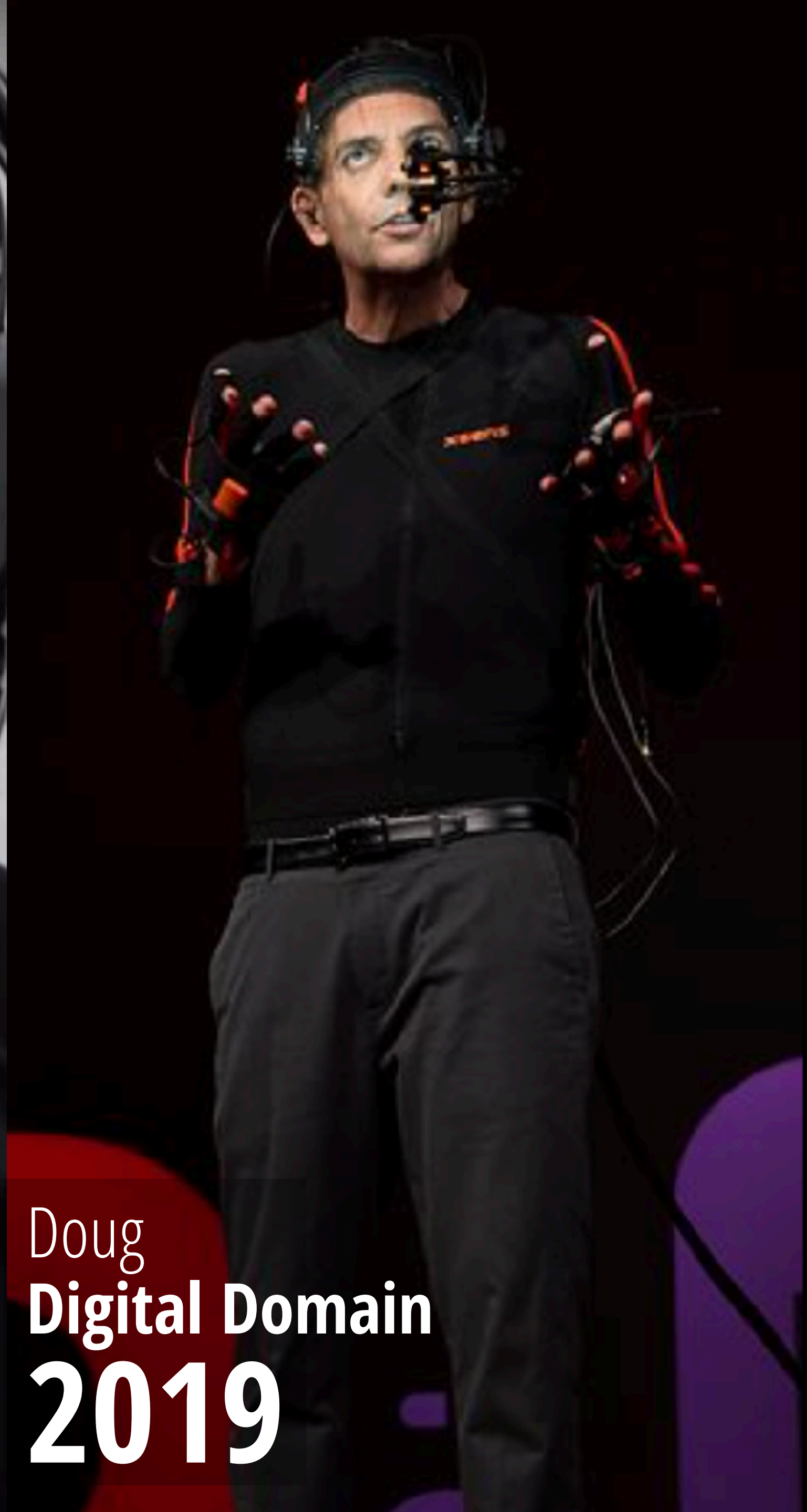




Mike  
**Wikihuman**  
**2017**



Siren  
**Epic Games**  
**2018**

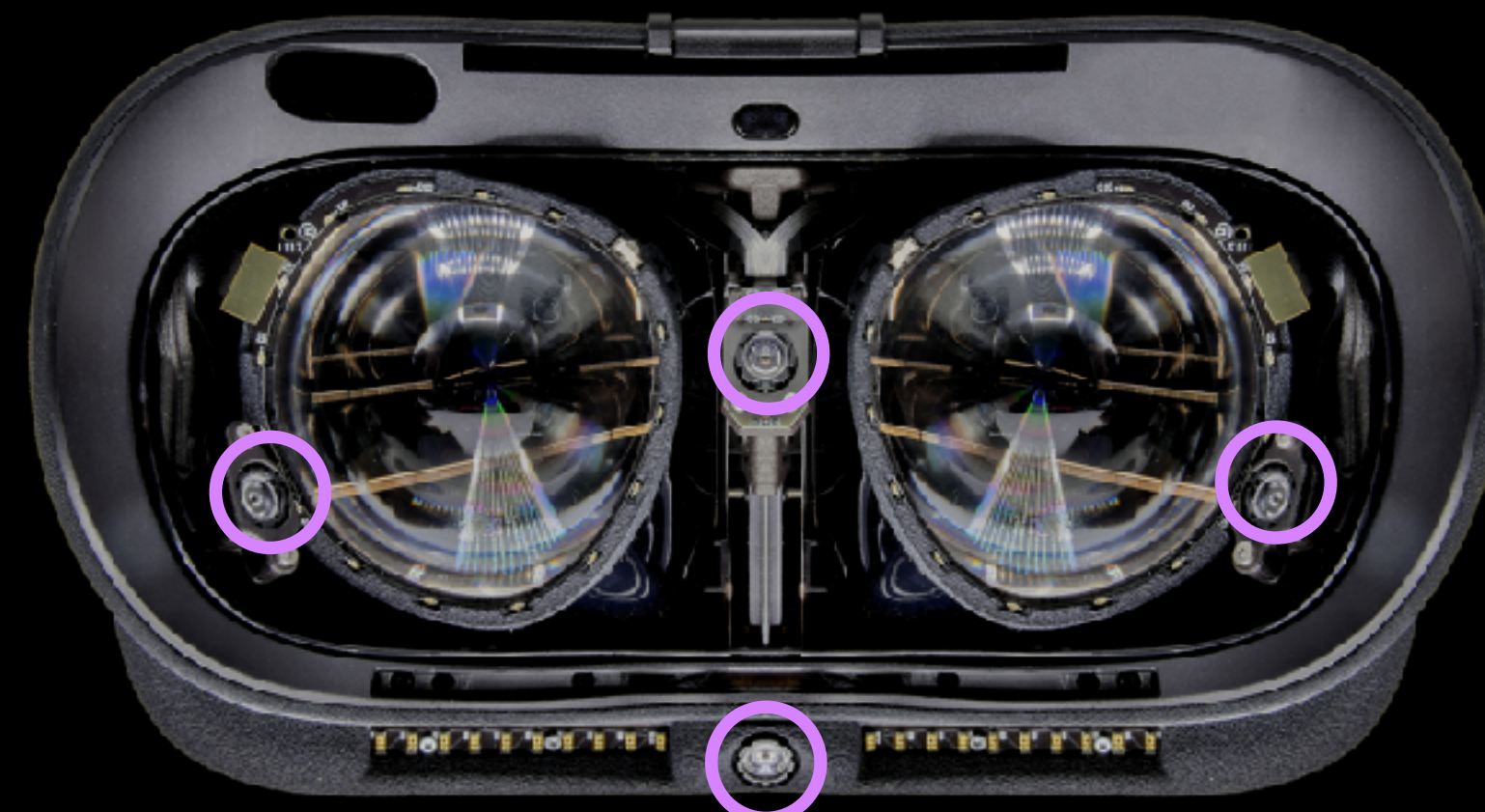


Doug  
**Digital Domain**  
**2019**



# HMC System

(Head-Mounted Capture)



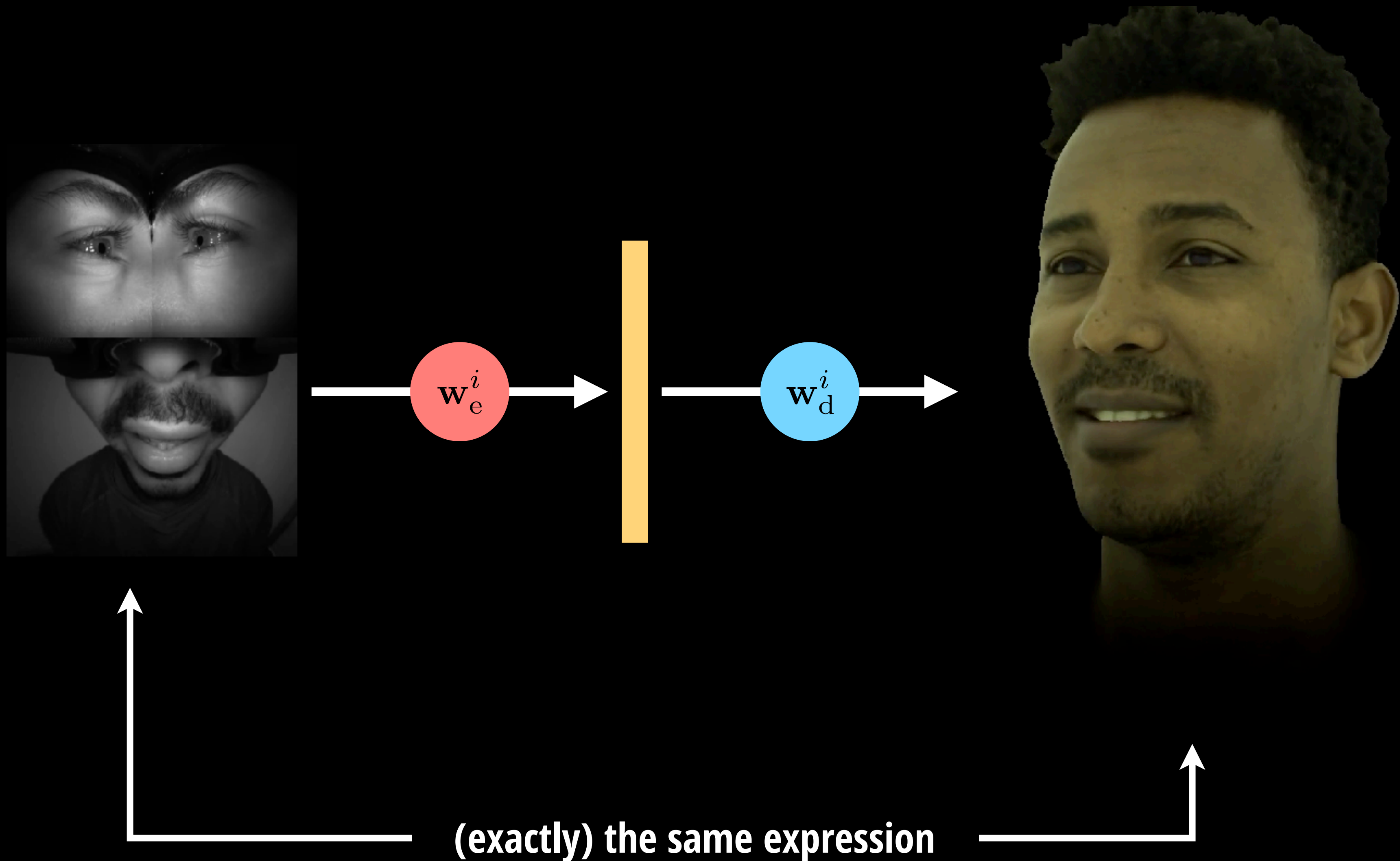
Cameras



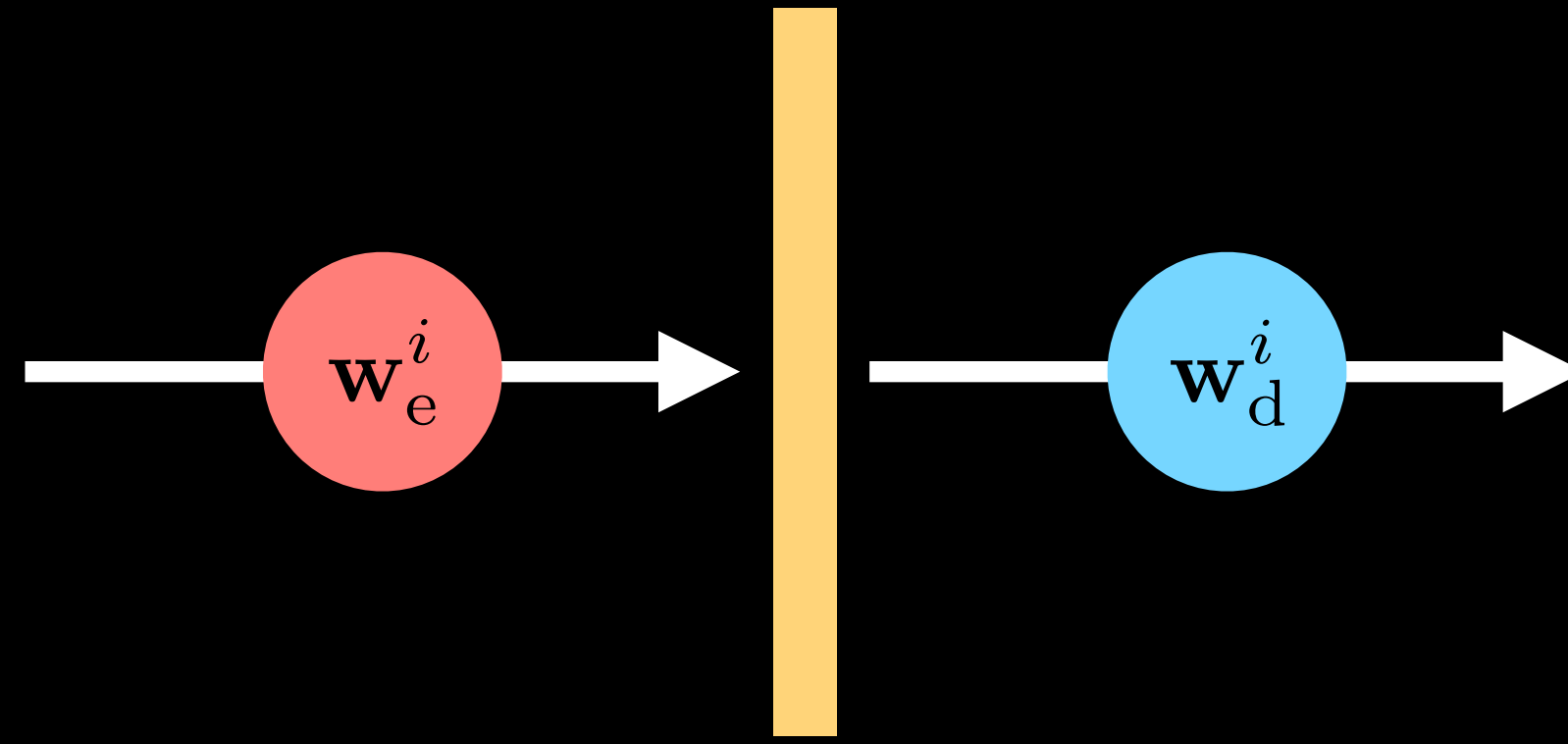


# THE CORRESPONDENCE PROBLEM

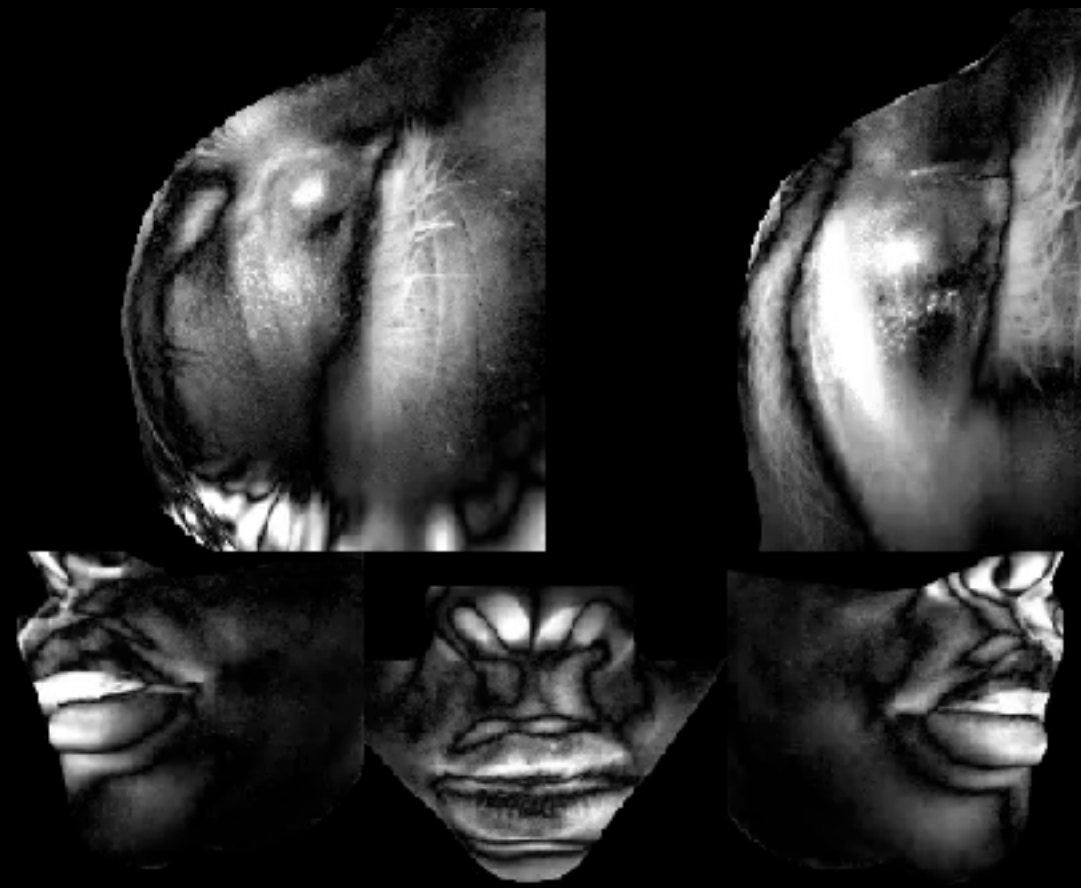
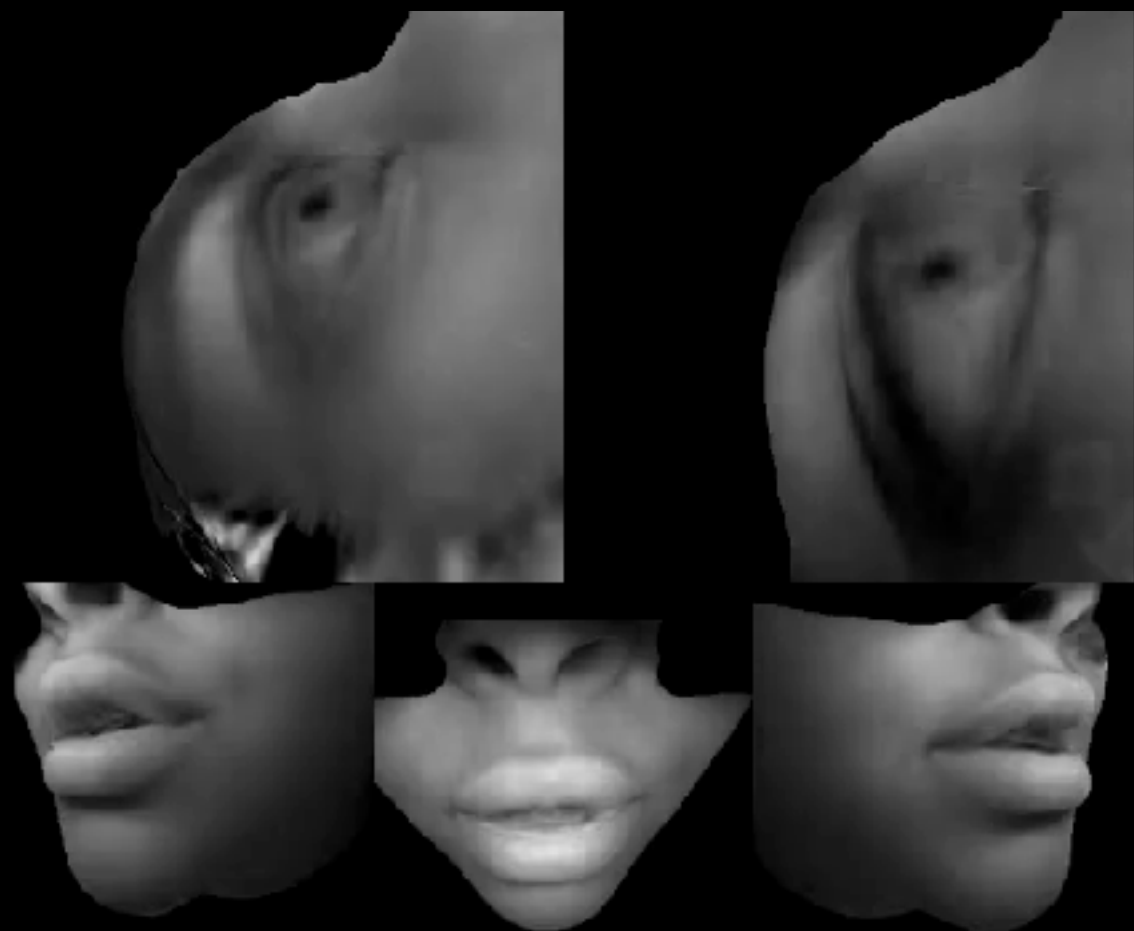
How do we get correspondences between sensor data and target display?







LOSS

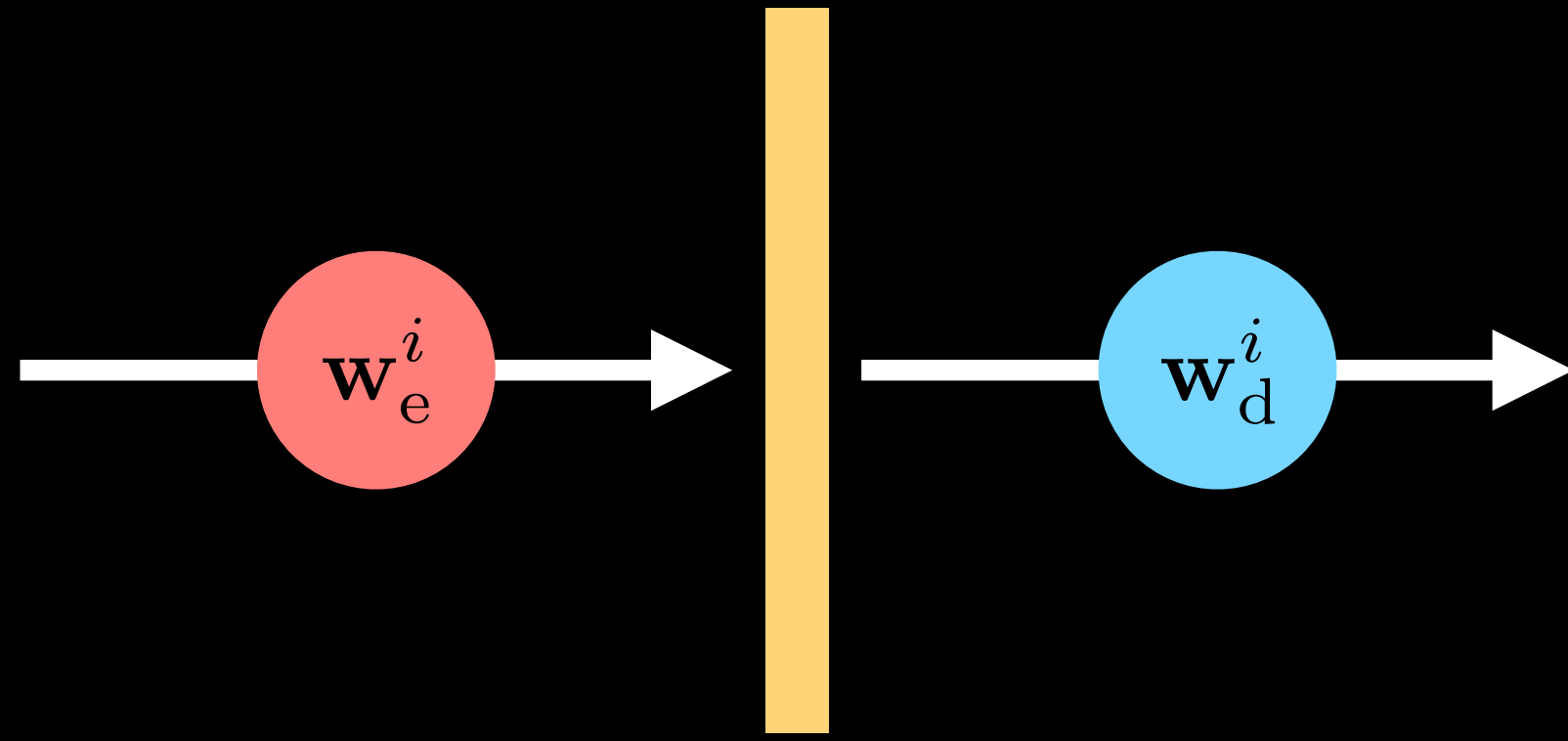


Domain Transfer

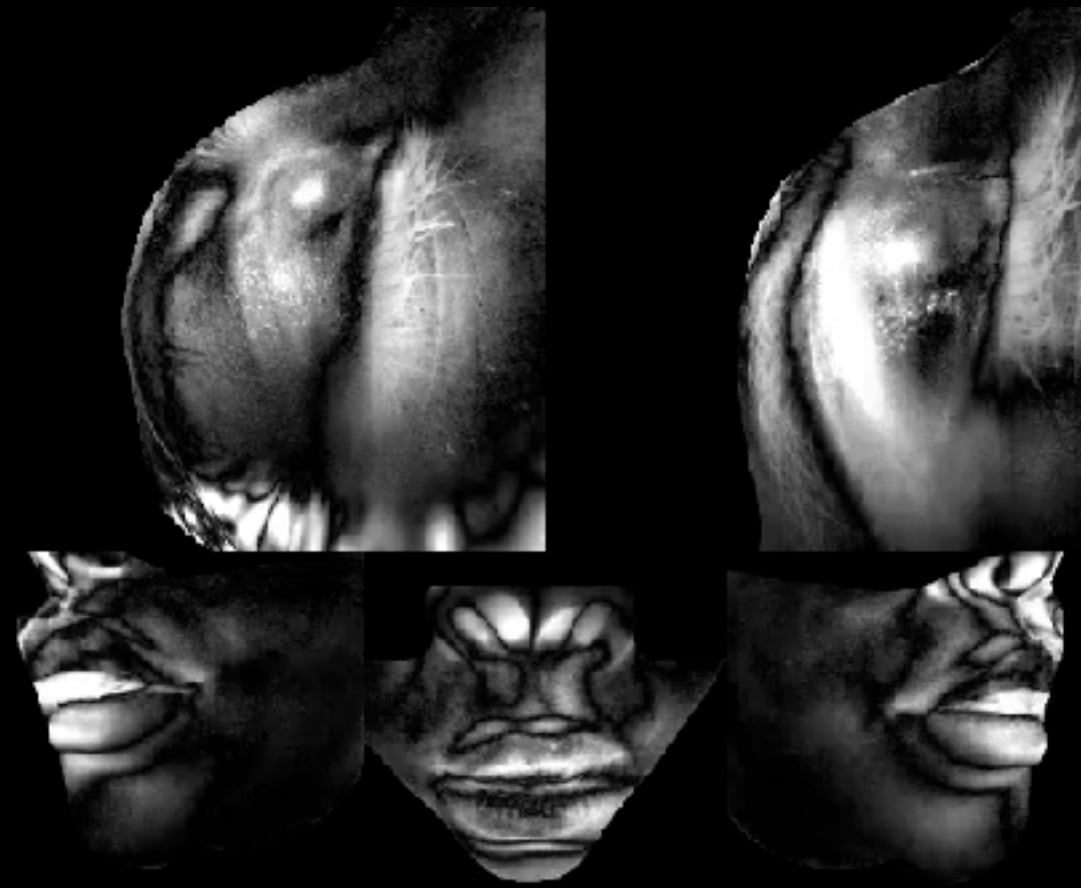
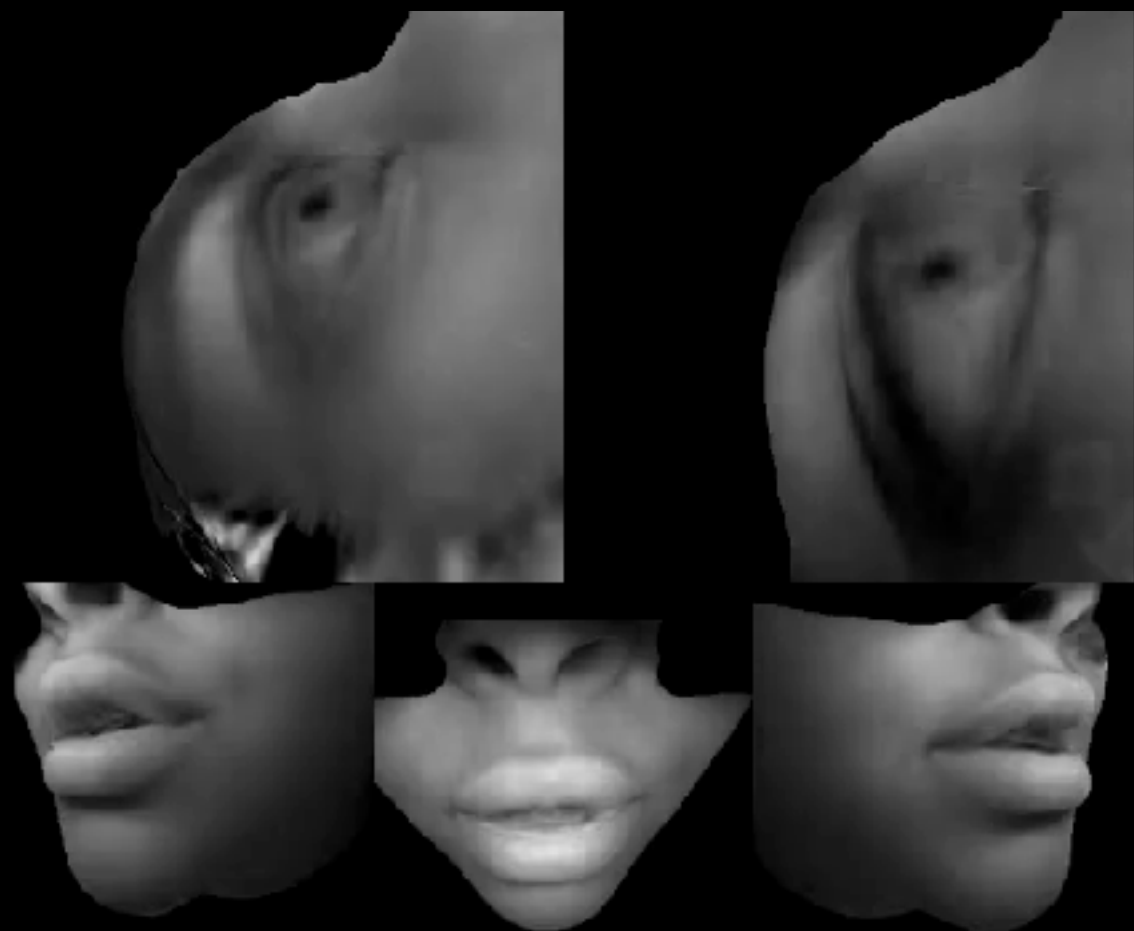


Differentiable Rendering





LOSS



Domain Transfer



Differentiable Rendering



# Metric Behavior

Measuring the Subtleties of True Multimodal Behavior From Minimal Sensing

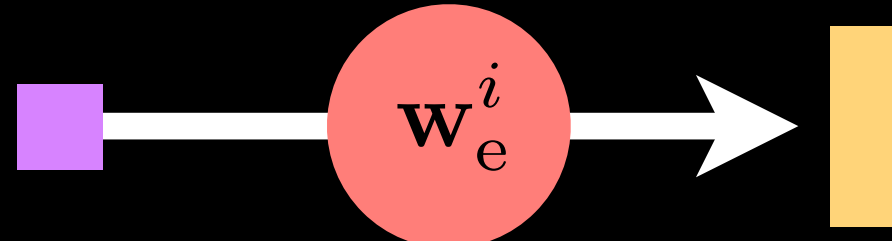






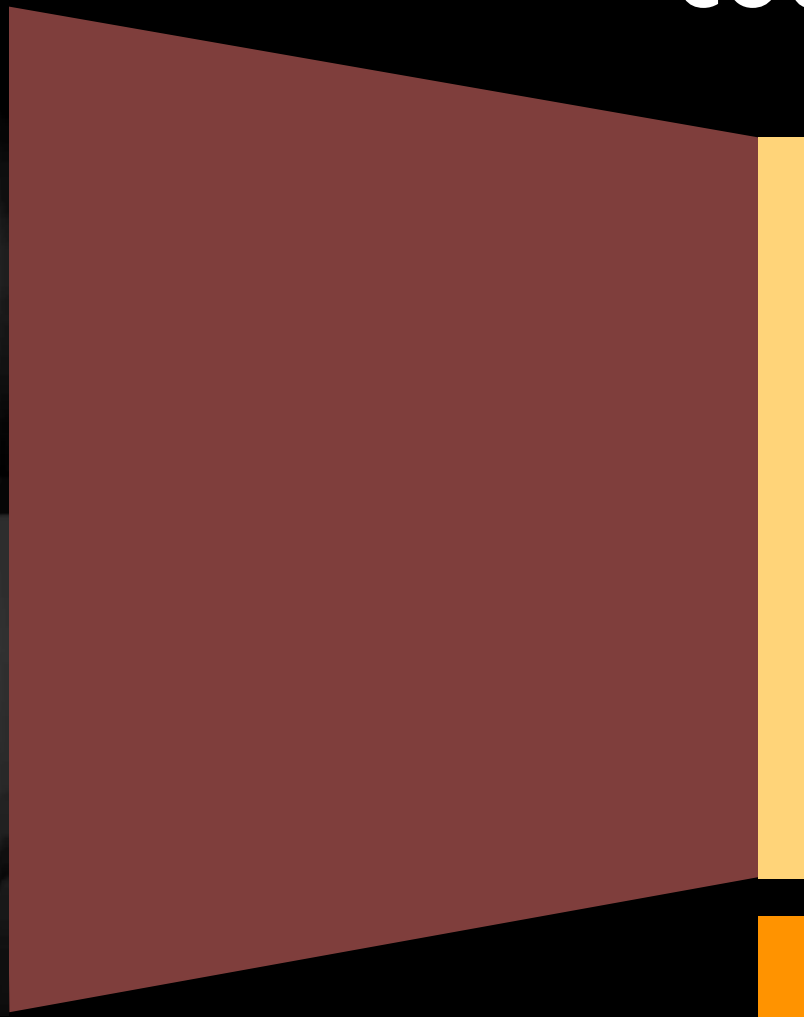


TX



Encoder

code



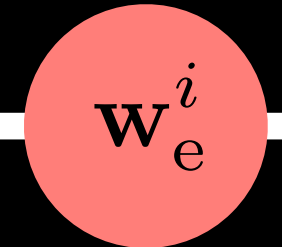
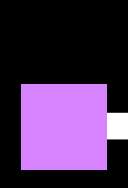
HMC Cameras

RX  
view  
code

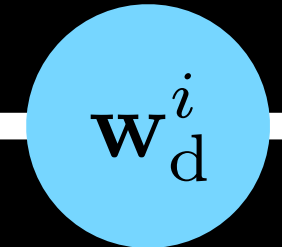
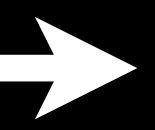




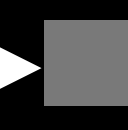
TX



Encoder



Decoder



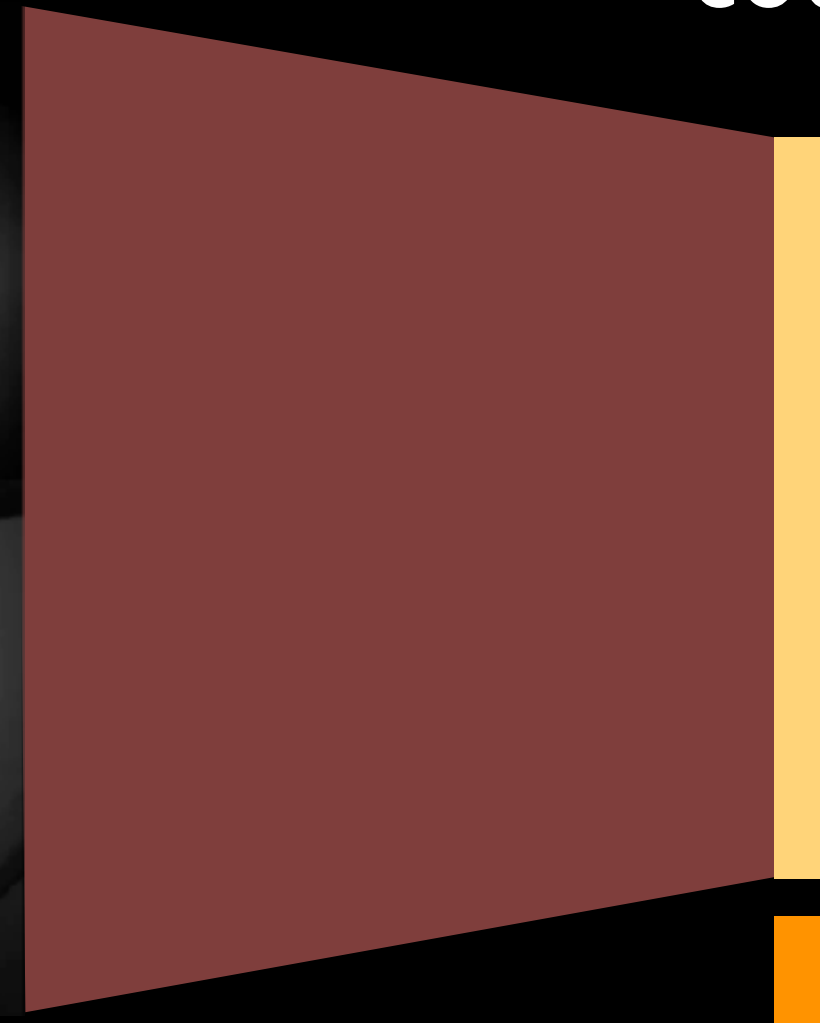
RX



HMC Cameras



code



RX view code



mesh



texture

RENDER



images



# Metric Behavior

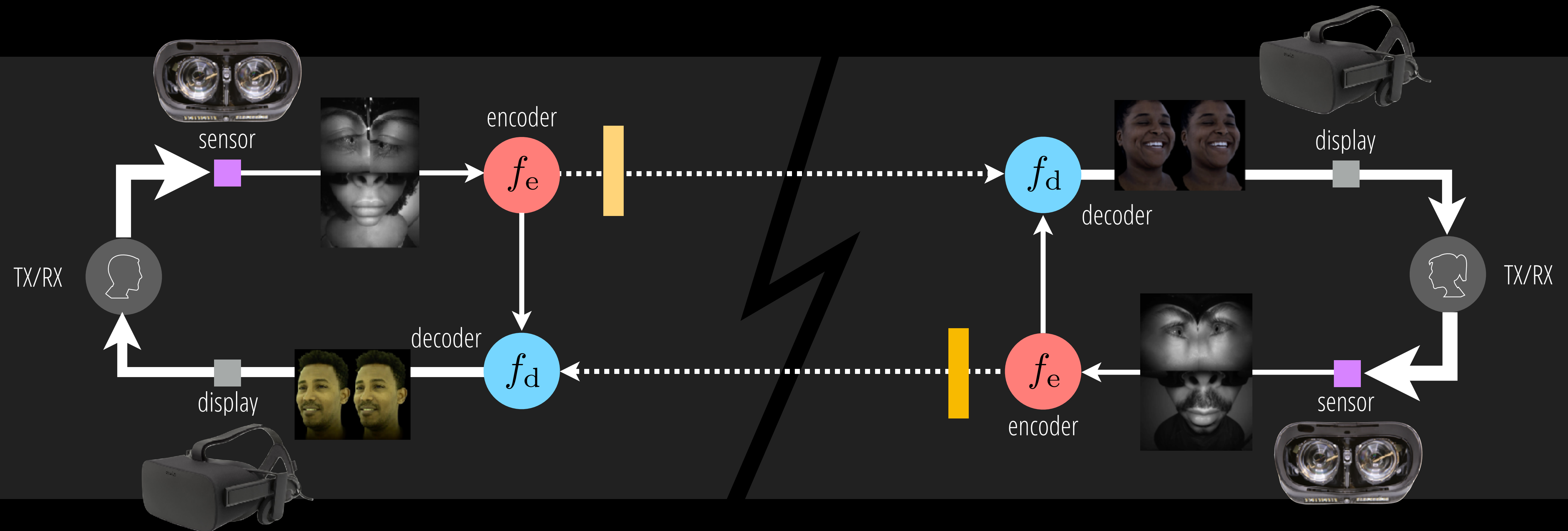
Measuring the Subtleties of True Multimodal Behavior From Minimal Sensing





# WHAT IS A CODEC AVATAR?

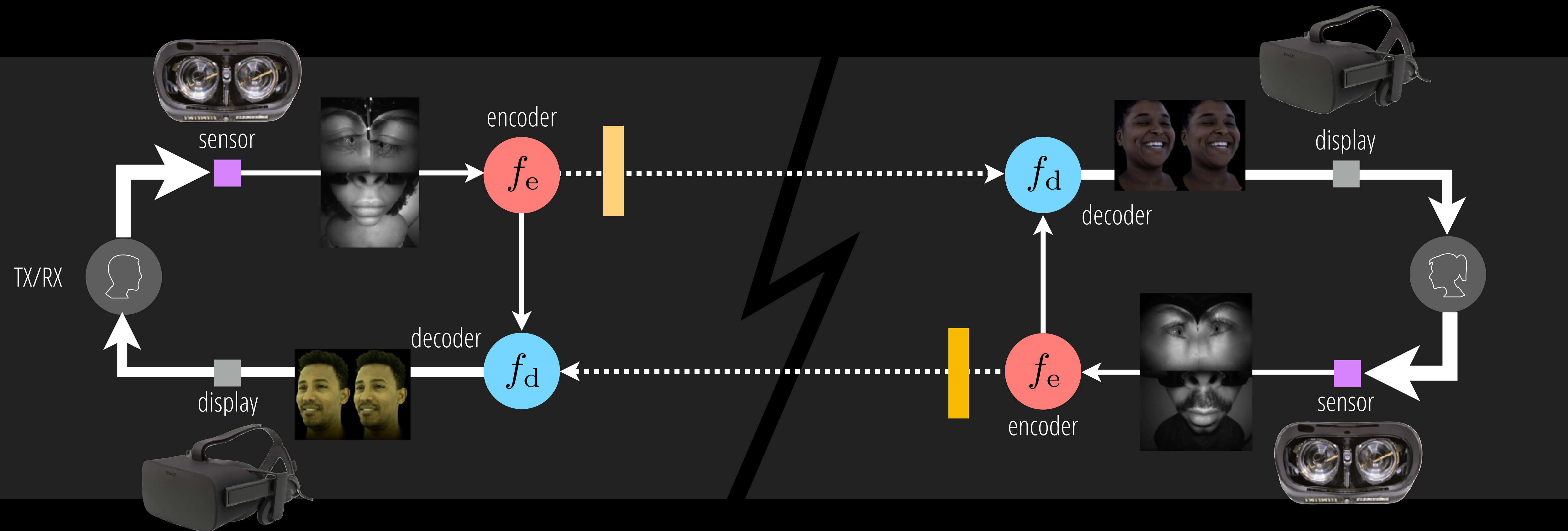
Social Interaction as a Communication Network





# WHAT IS A CODEC AVATAR?

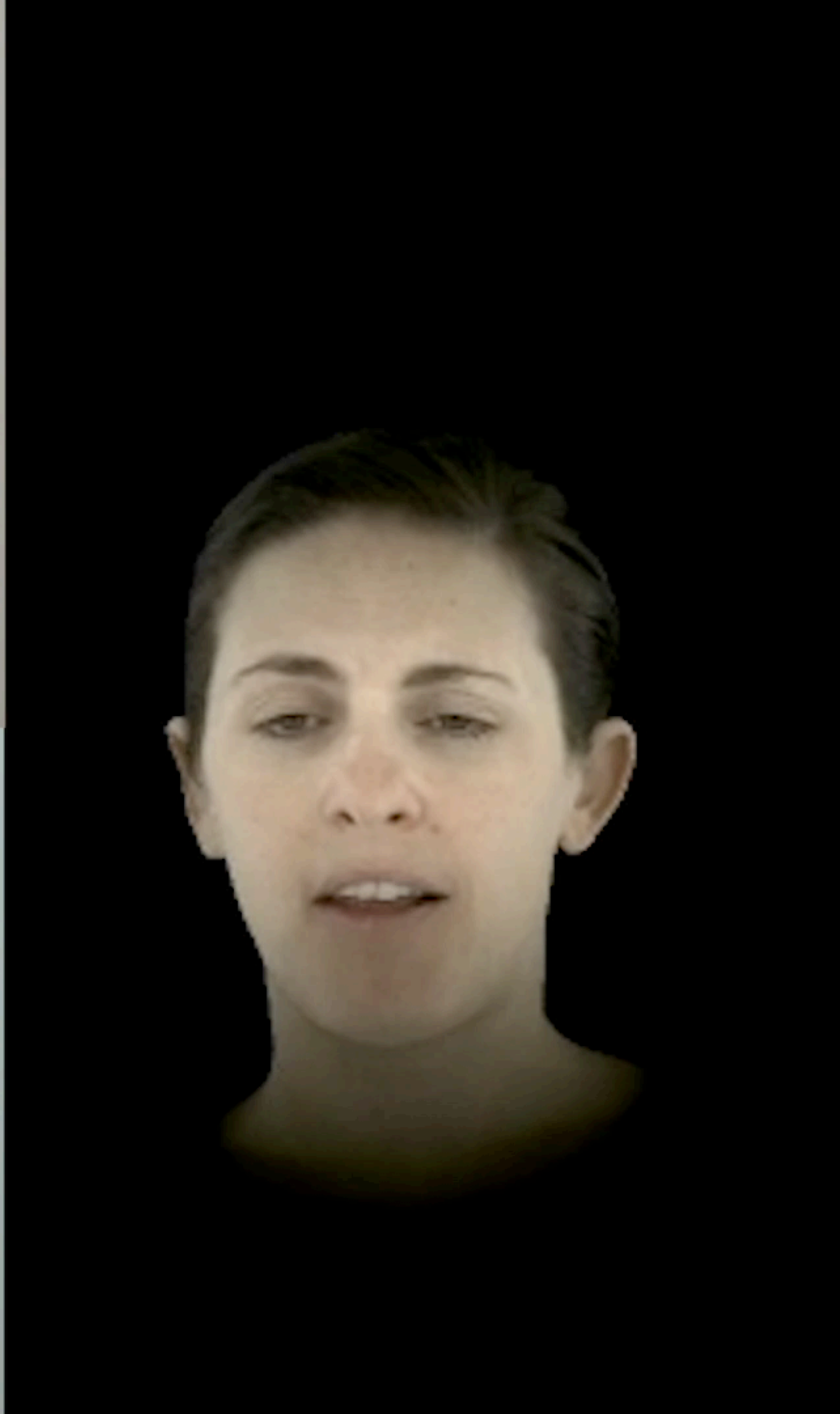
Social Interaction as a Communication Network













# Metric Identity

How do we produce identity preserving avatars for billions of people?

# Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

# Metric Time

How do we do all this in realtime without access to artistic correction?



# Metric Identity

How do we produce identity preserving avatars for billions of people?

# Metric Behavior

How do we measure the subtleties of true multimodal behavior from minimal sensing?

# Metric Time

How do we do all this in realtime without access to artistic correction?

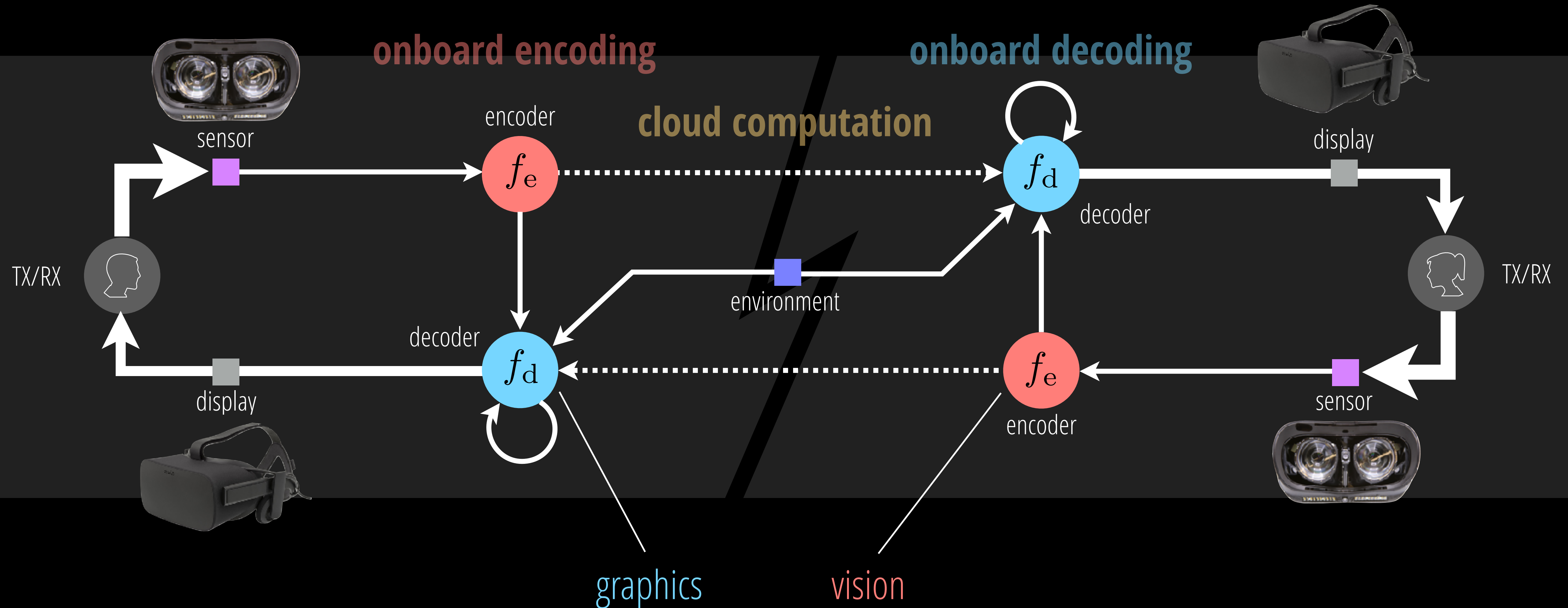


**NO SIGNAL  
IS A SIGNAL**



# WHAT IS A CODEC AVATAR?

## The Visual Computing Pipeline



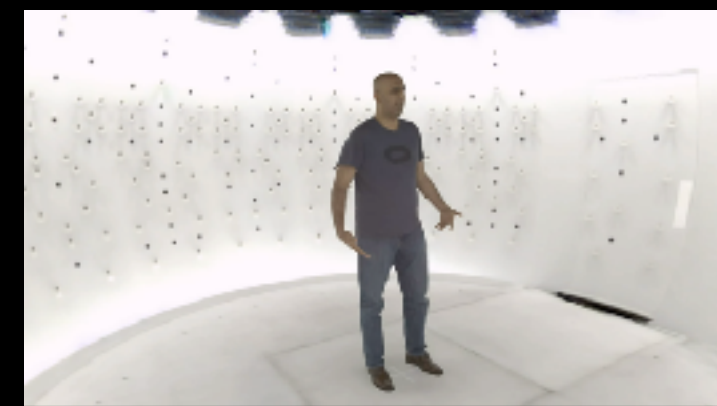
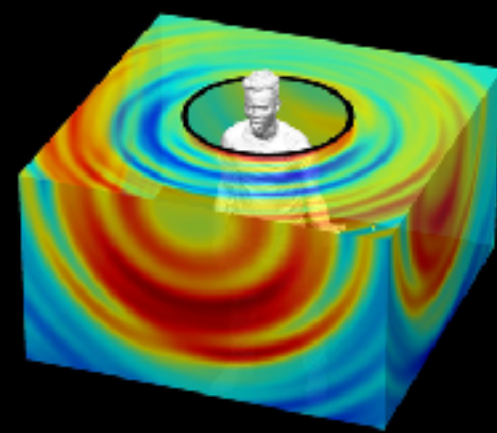
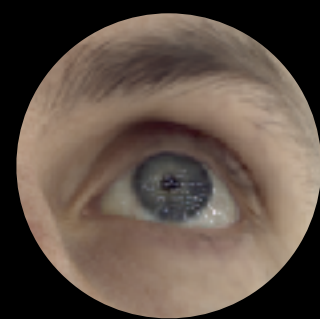


onboard decoding





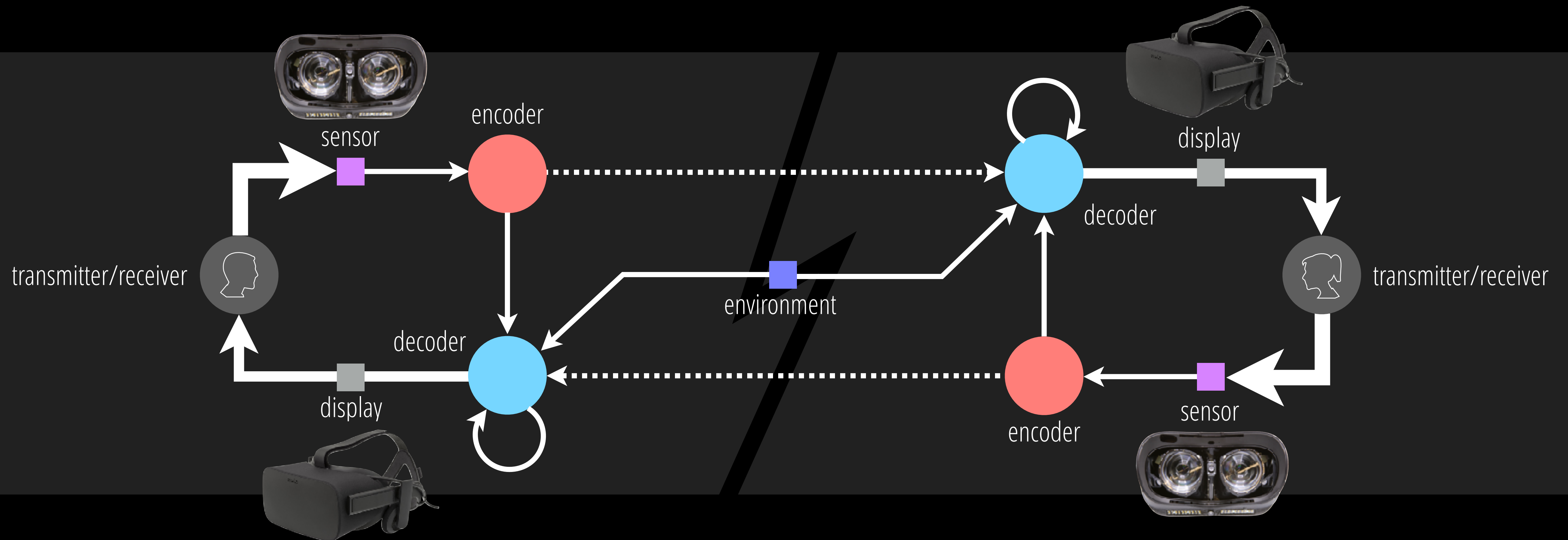
# WHAT'S NEXT?





# WHAT IS A CODEC AVATAR?

## Environments for Codec Avatars





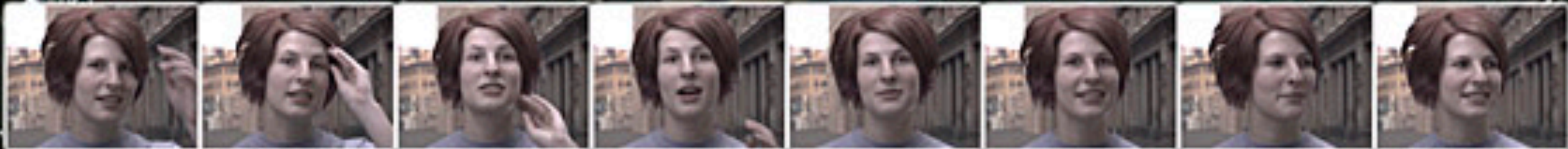
# 3D Reconstruction







[Guo et al. "The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting, TOG 2019]





# ENVIRONMENTS

Relighting Codec Avatars

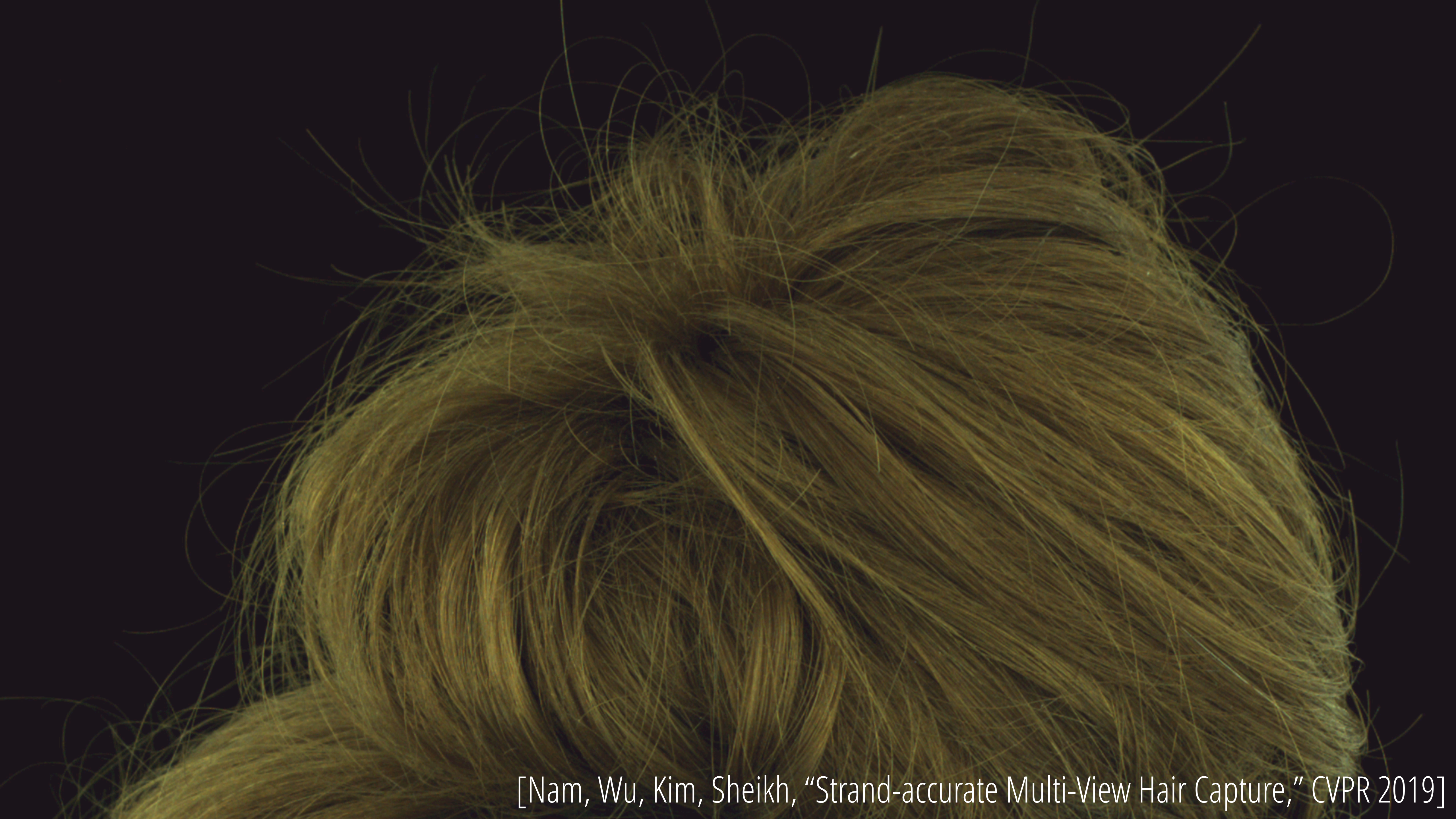






[Nam, Wu, Kim, Sheikh, "Strand-accurate Multi-View Hair Capture," CVPR 2019]



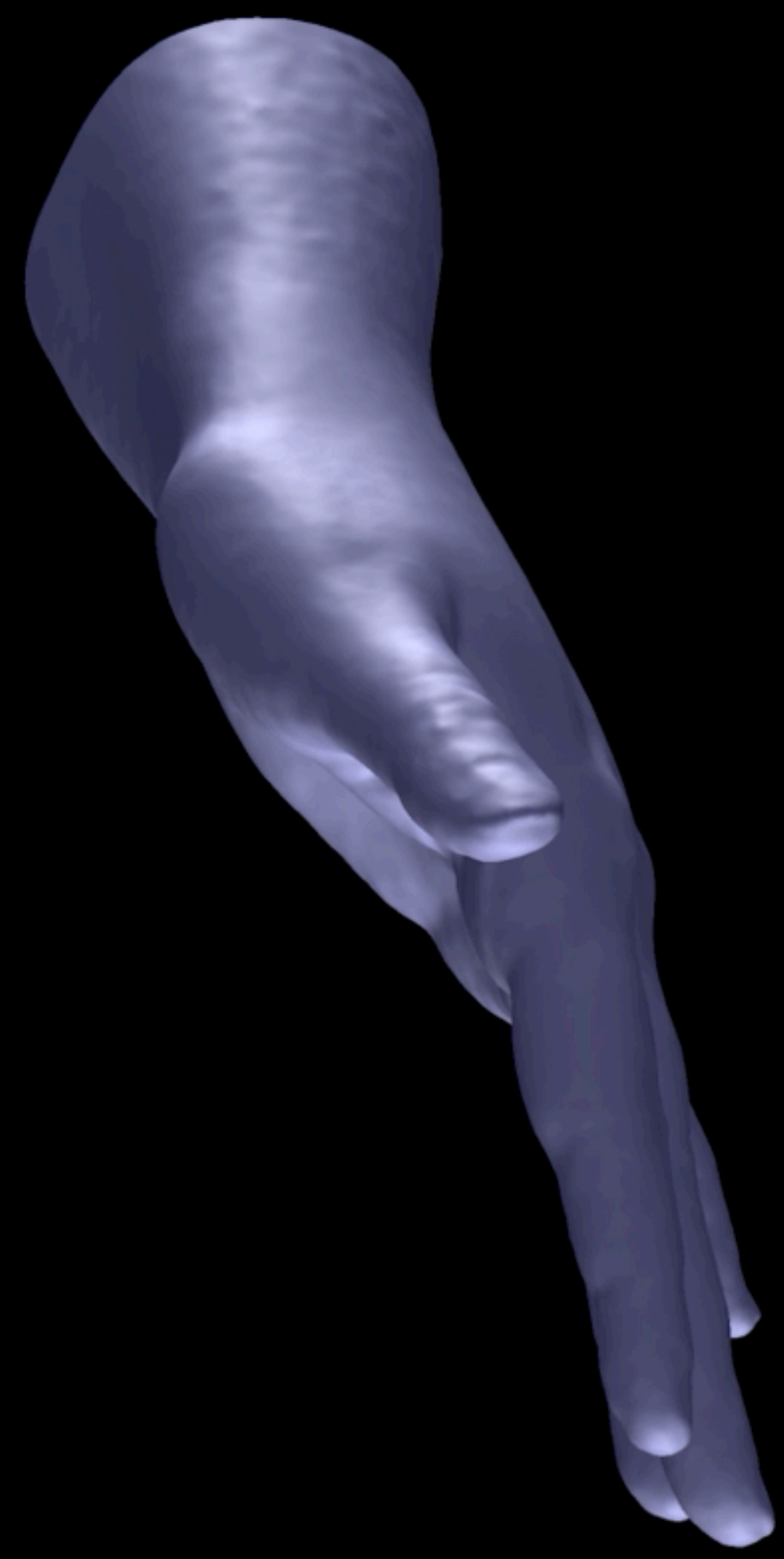


[Nam, Wu, Kim, Sheikh, "Strand-accurate Multi-View Hair Capture," CVPR 2019]

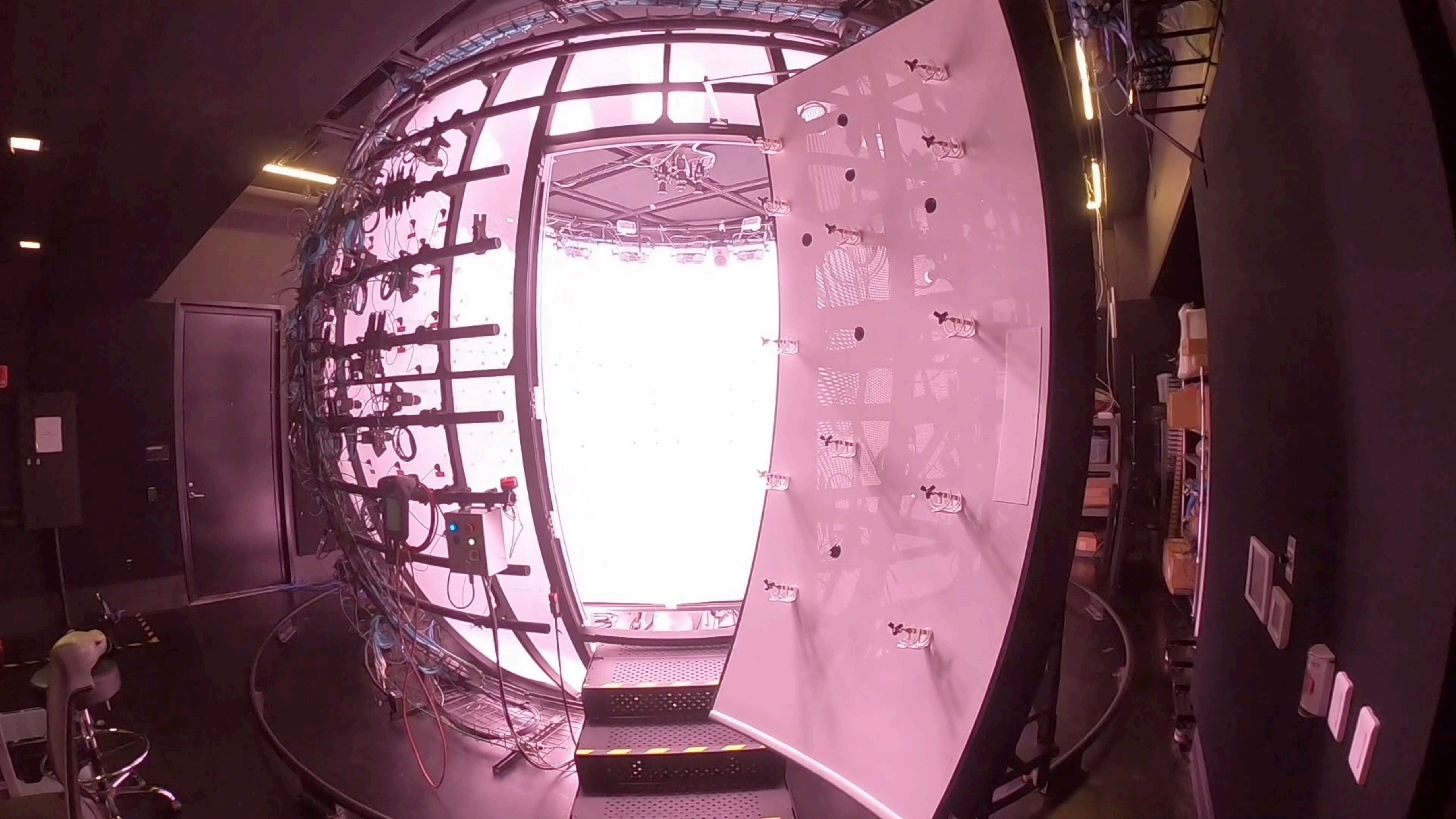




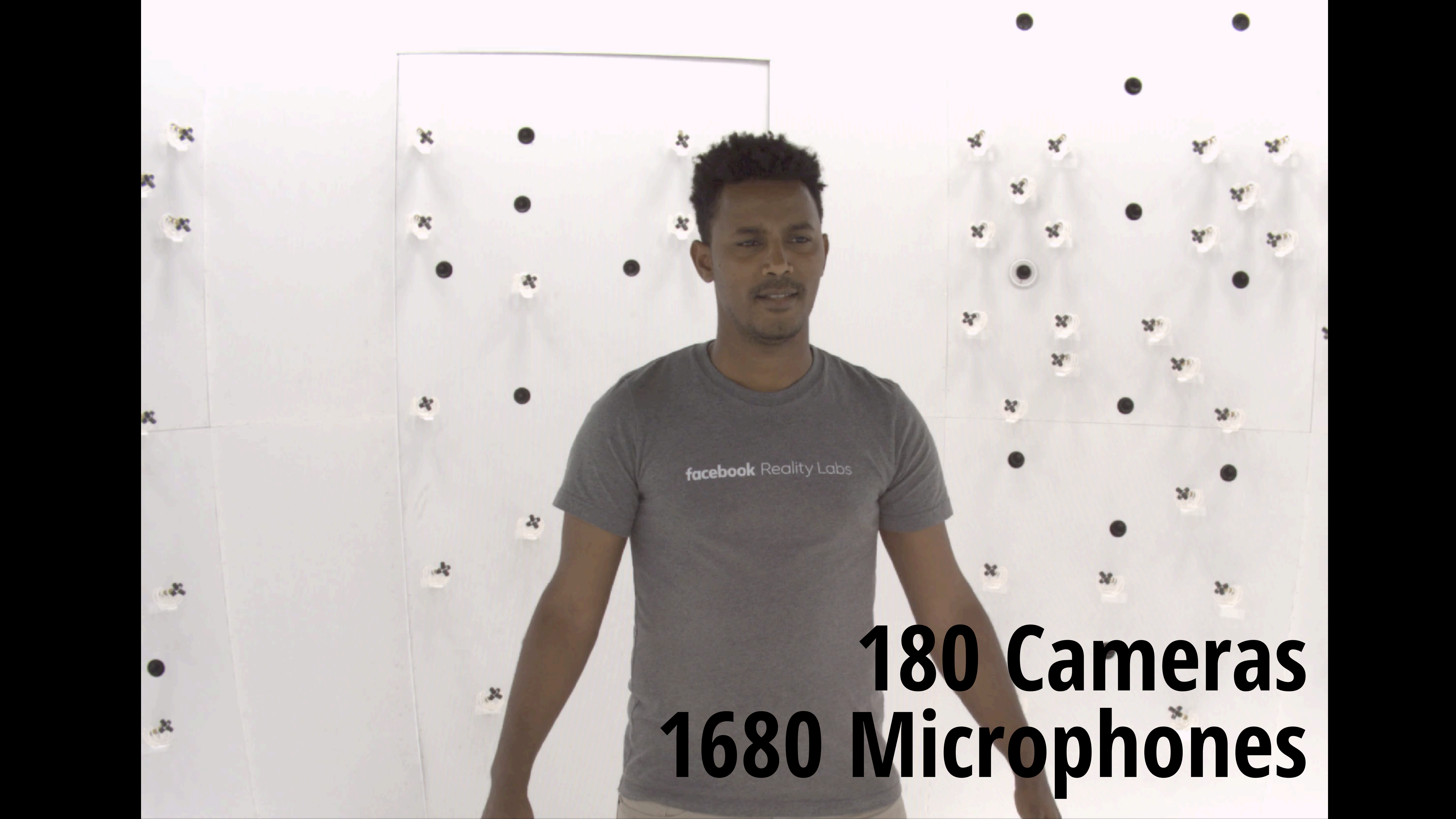












facebook Reality Labs

**180 Cameras**  
**1680 Microphones**









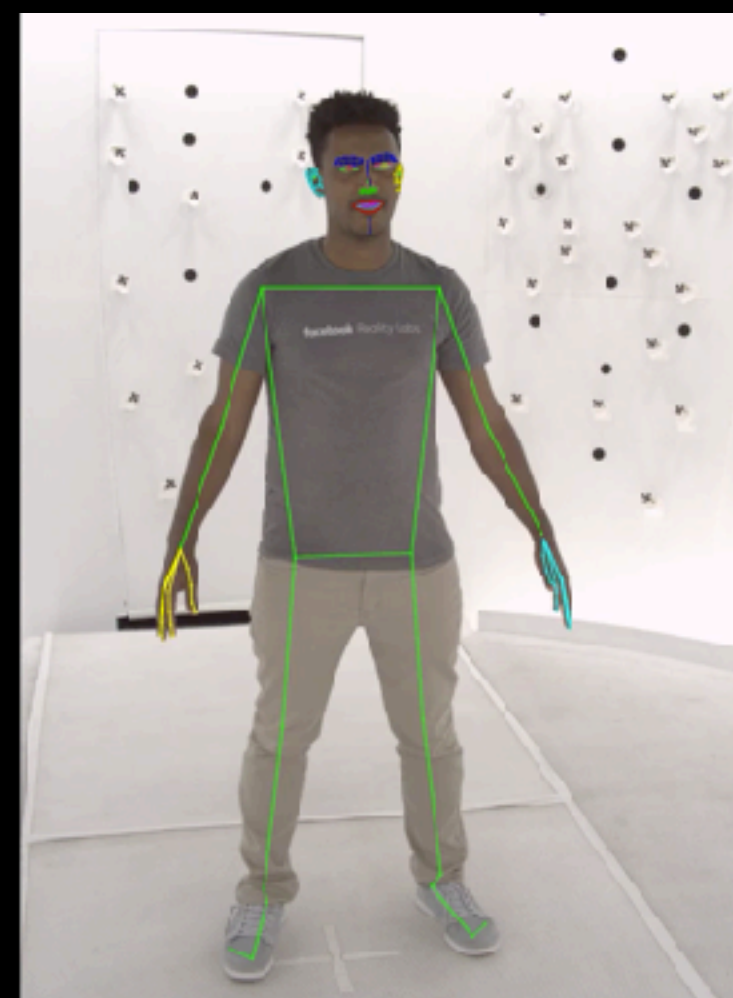
Original Image



Segmentation



3D Reconstruction



Keypoint  
Detection



Mesh Tracking\*



Model-free  
Mesh Tracking



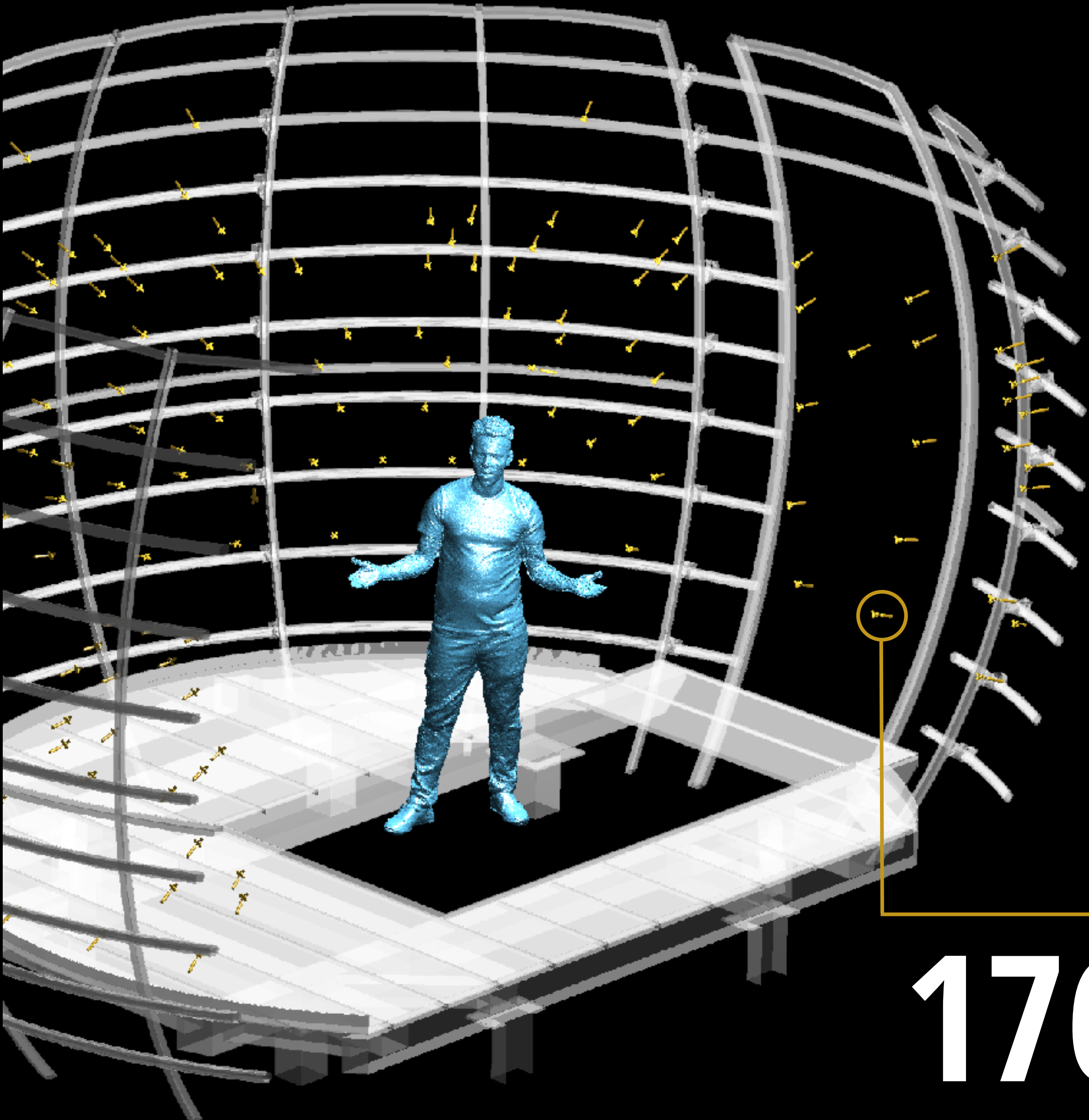
Avatar Decoder





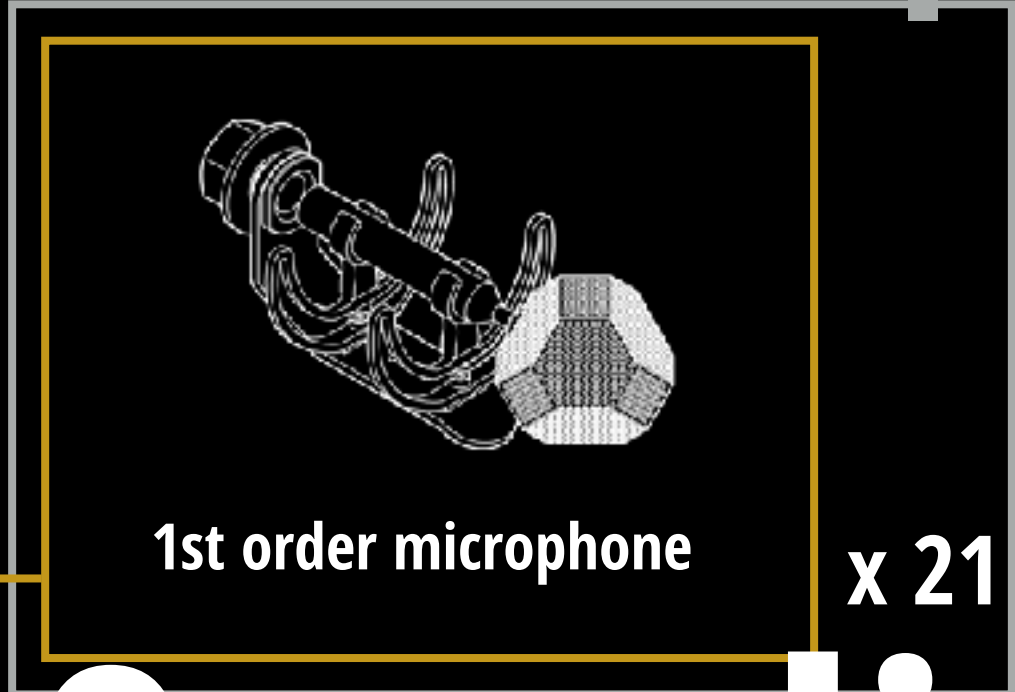
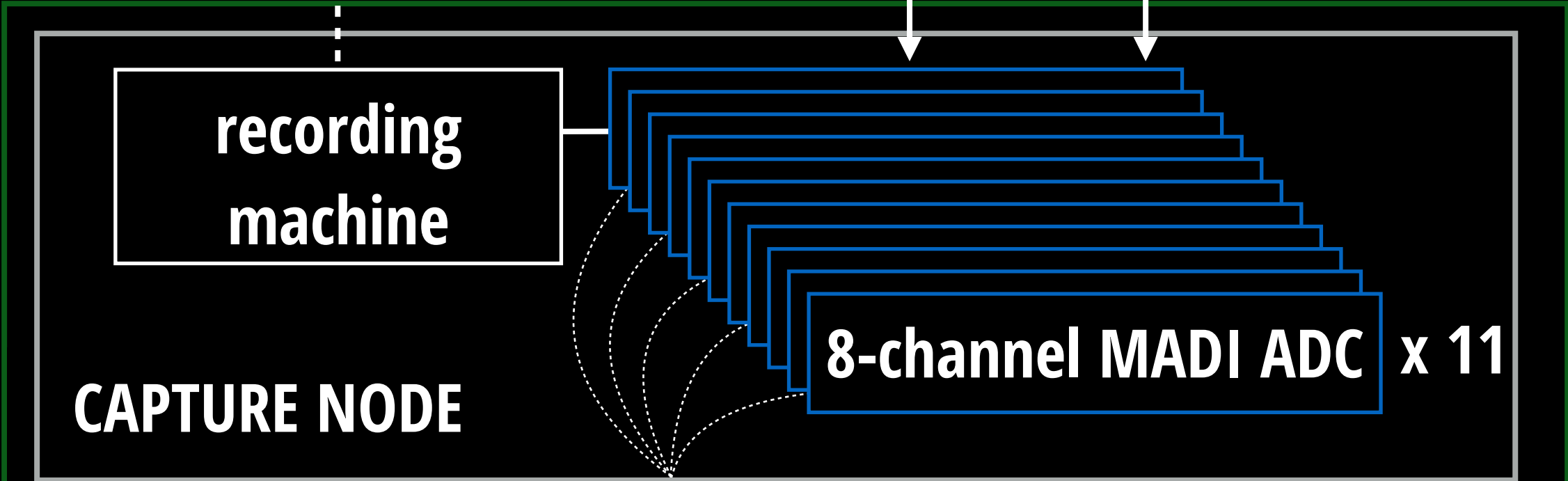


# SOCIOPTICON AUDIO SYSTEM



RECORDING SOFTWARE

IRIG world-clock



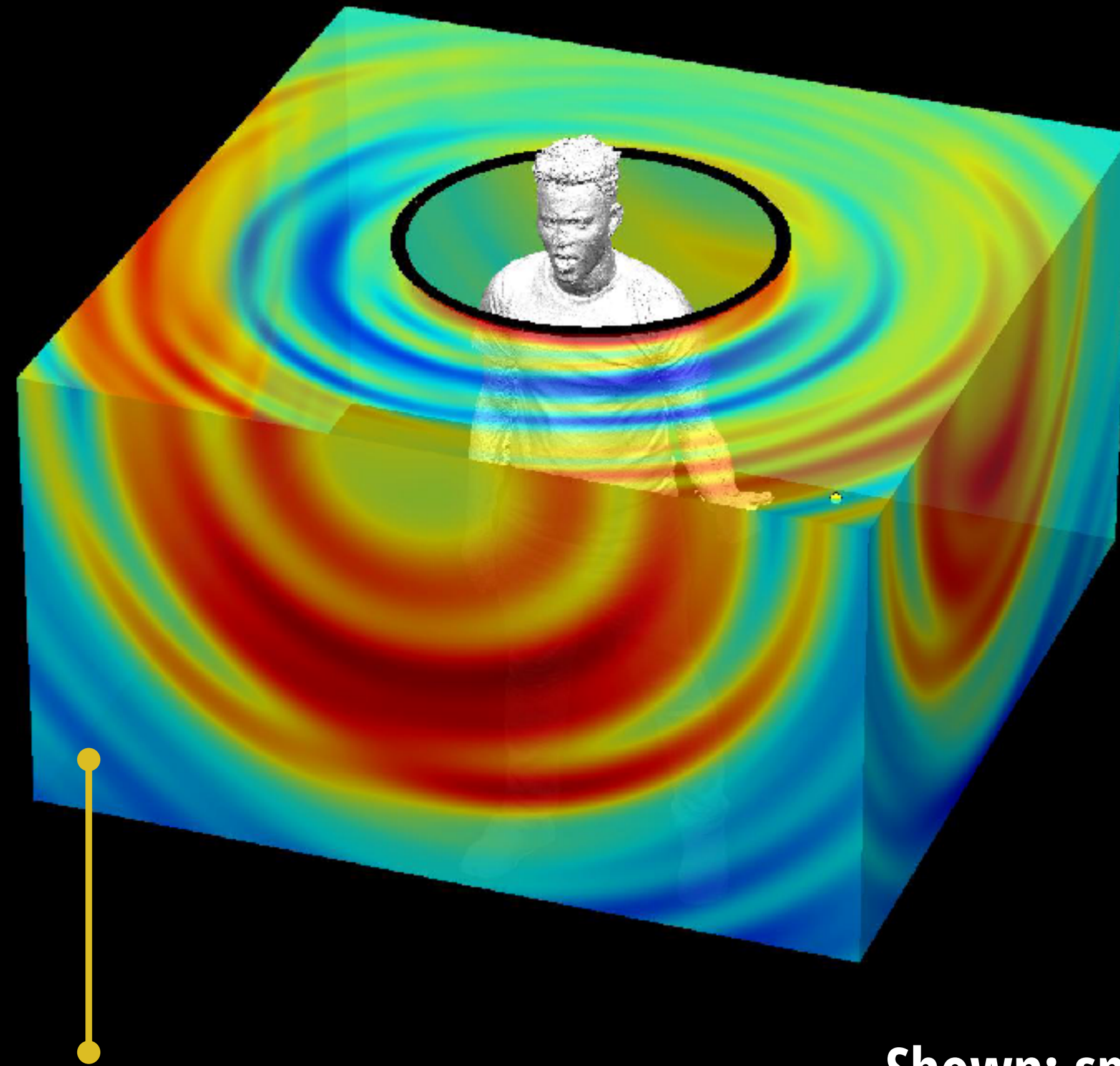
x 20

1760 audio signals

# 1760 audio channels

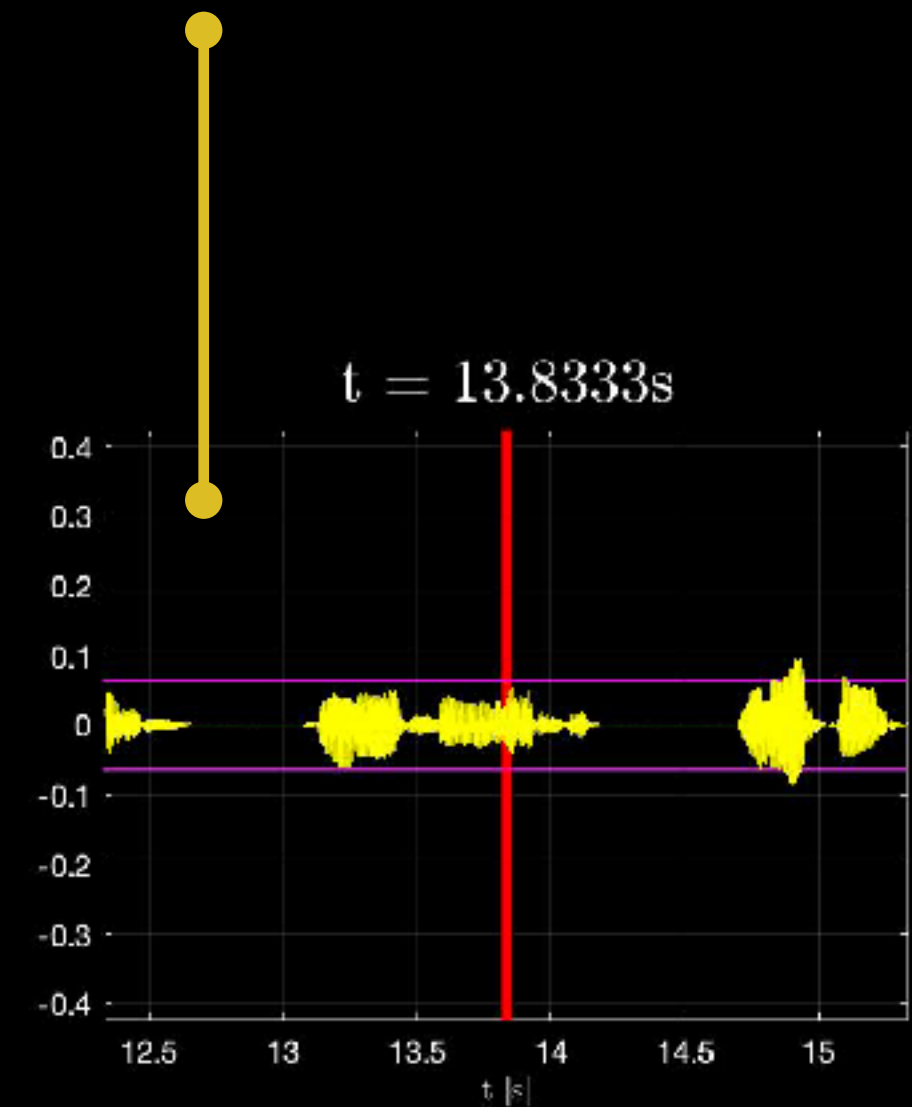


SoundField(x, y, z, t)

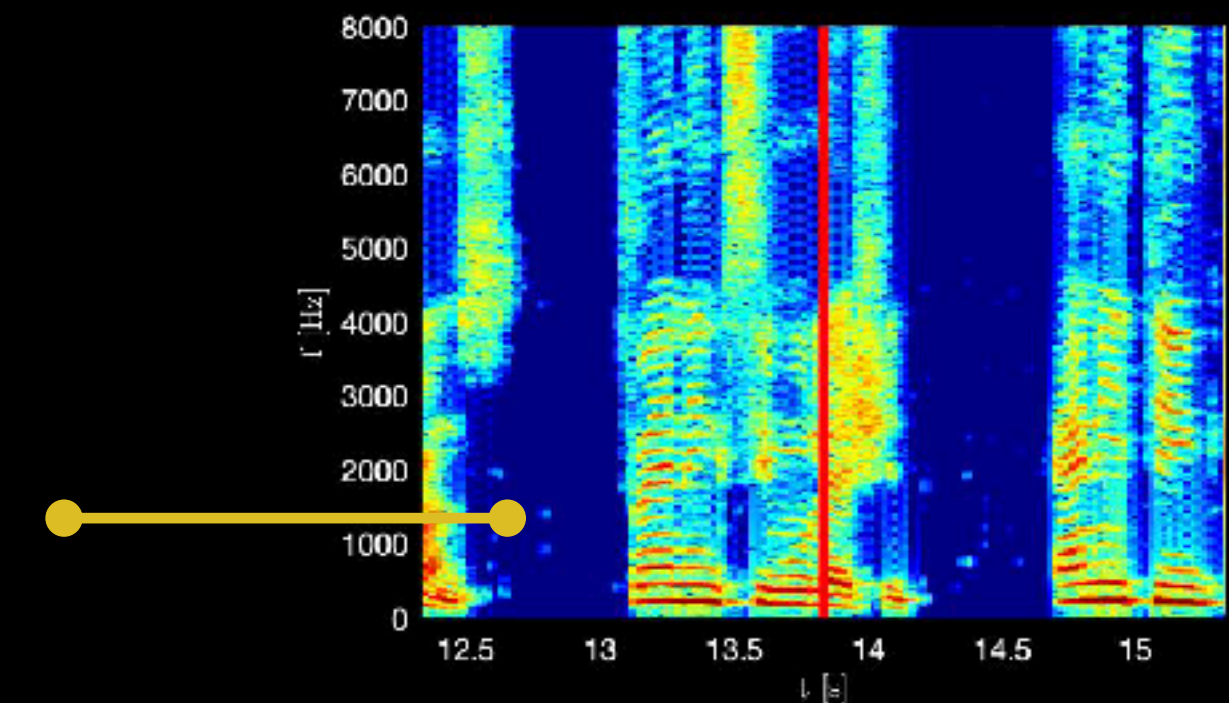


Shown: values of the sound pressure at the given positions in space; shown for a given time instant; computed on the surface of a volume box

Shown: **waveform** for a given virtual microphone position; **time instant** & **max/min colormap values** for the visualized sound pressure samples

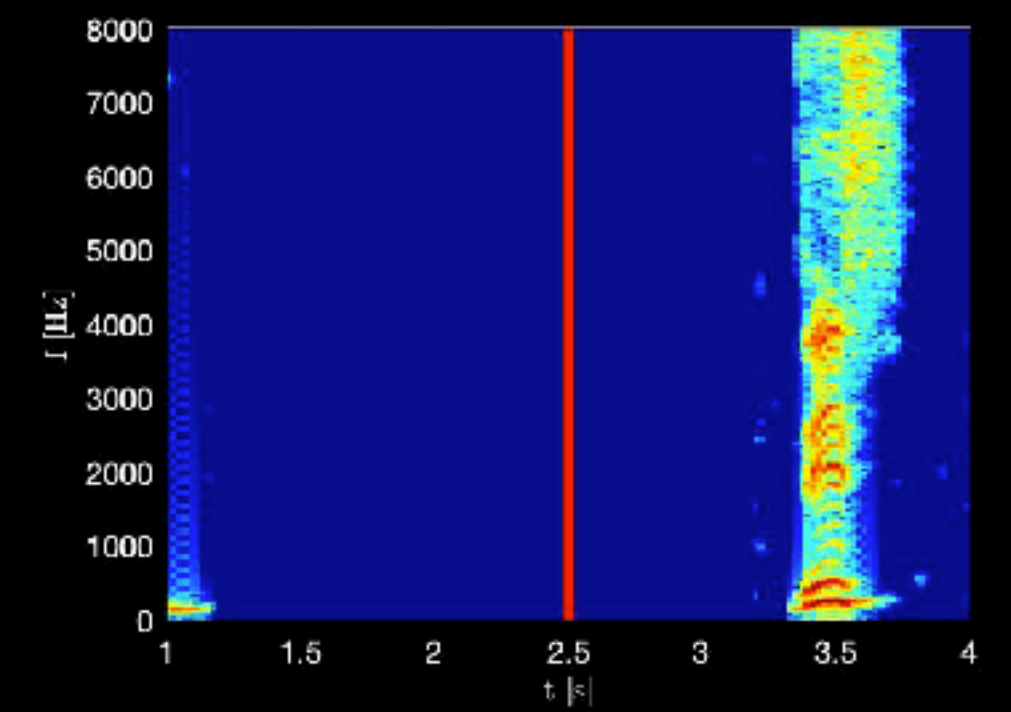
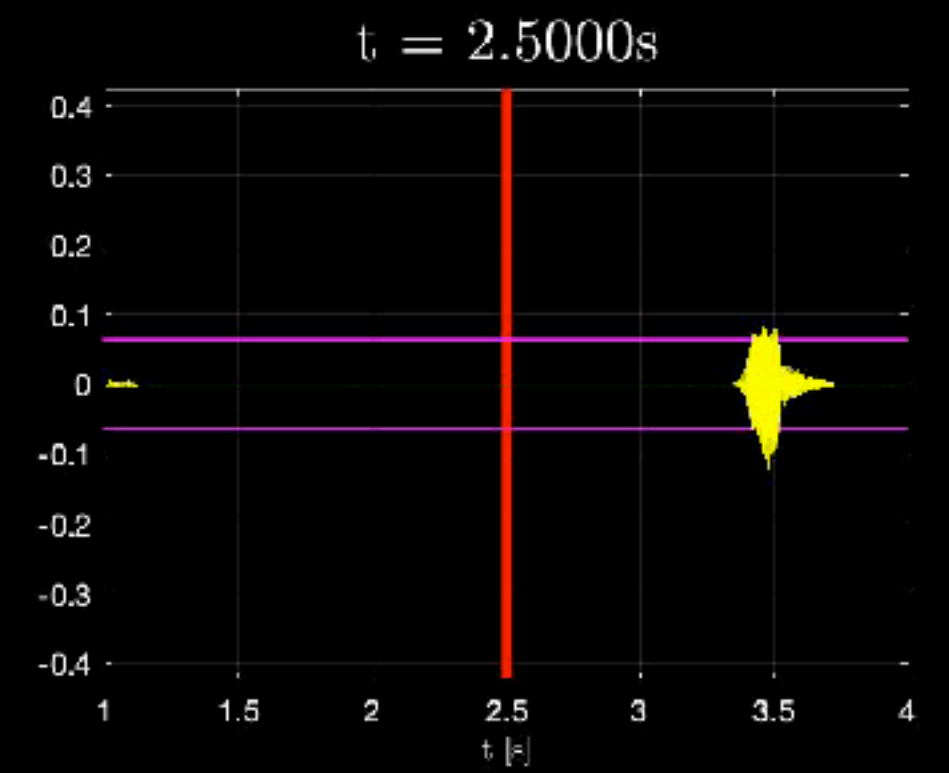
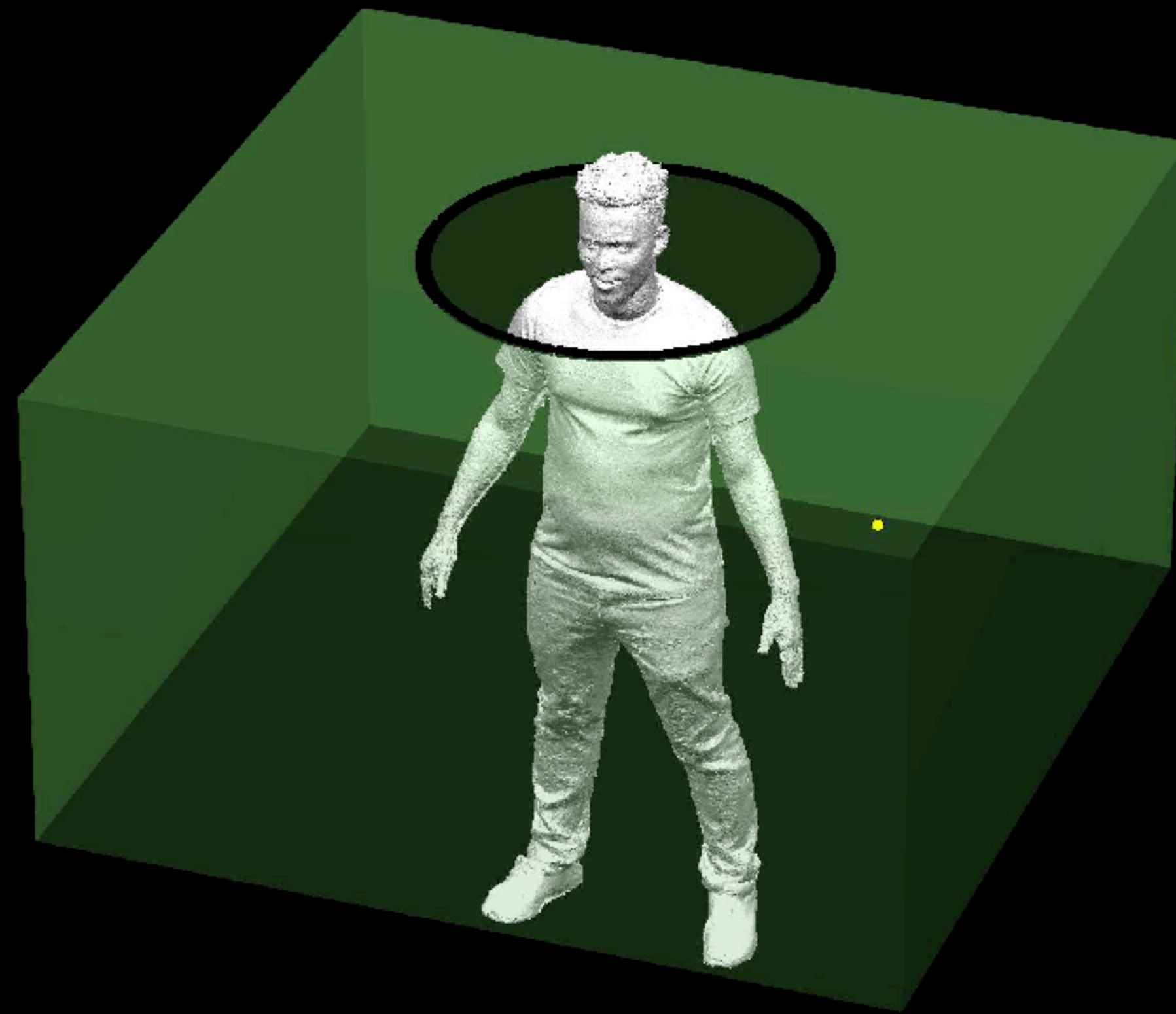


Shown: spectrogram of the virtual microphone signal





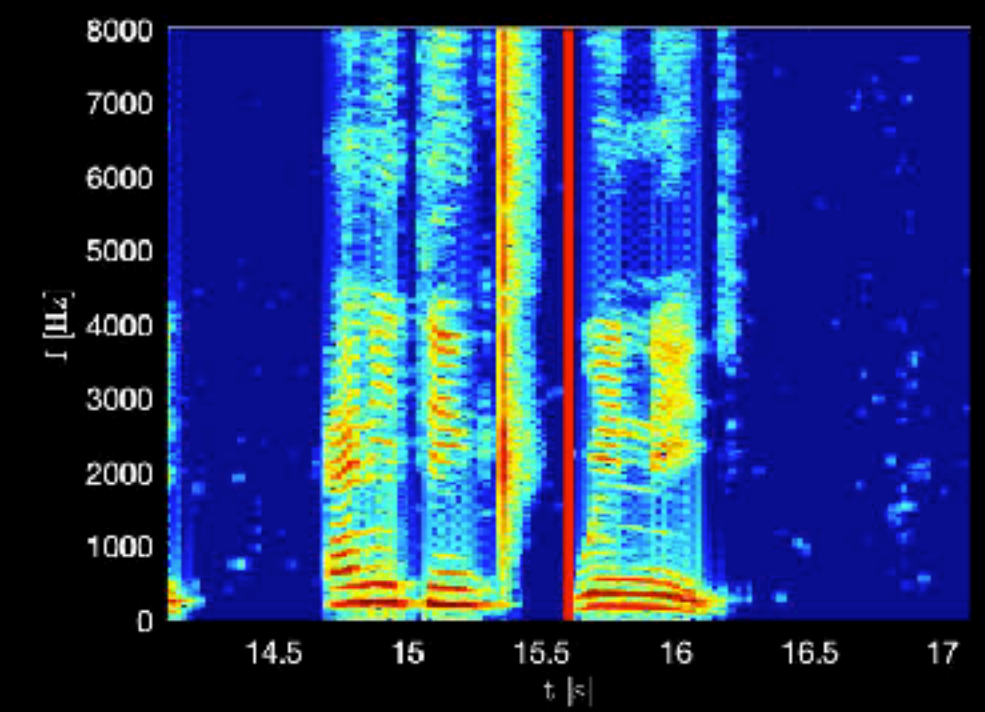
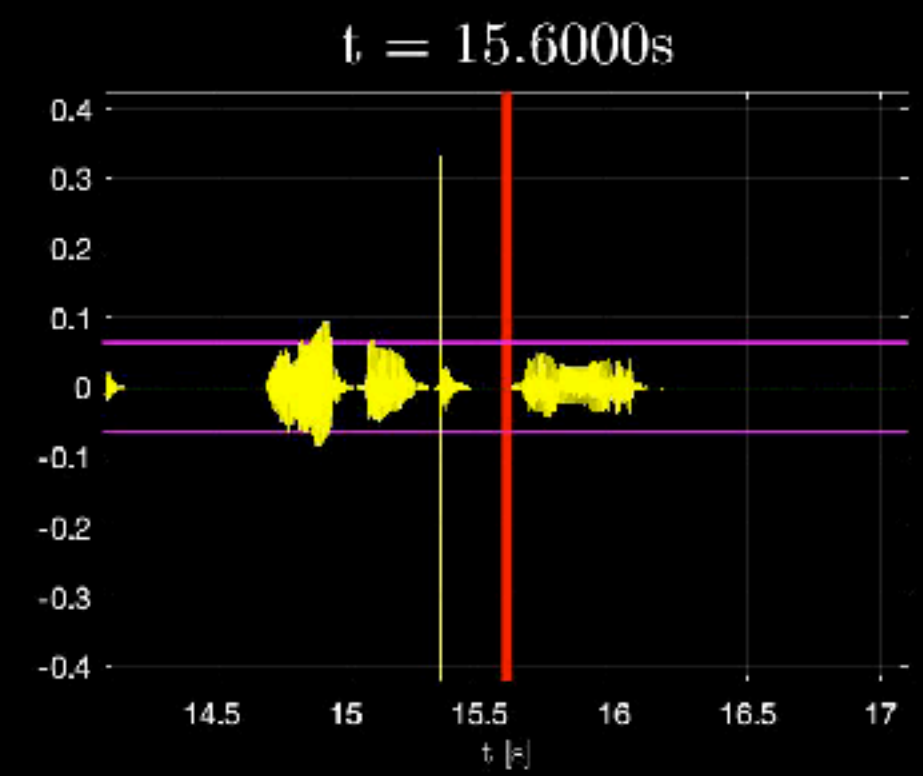
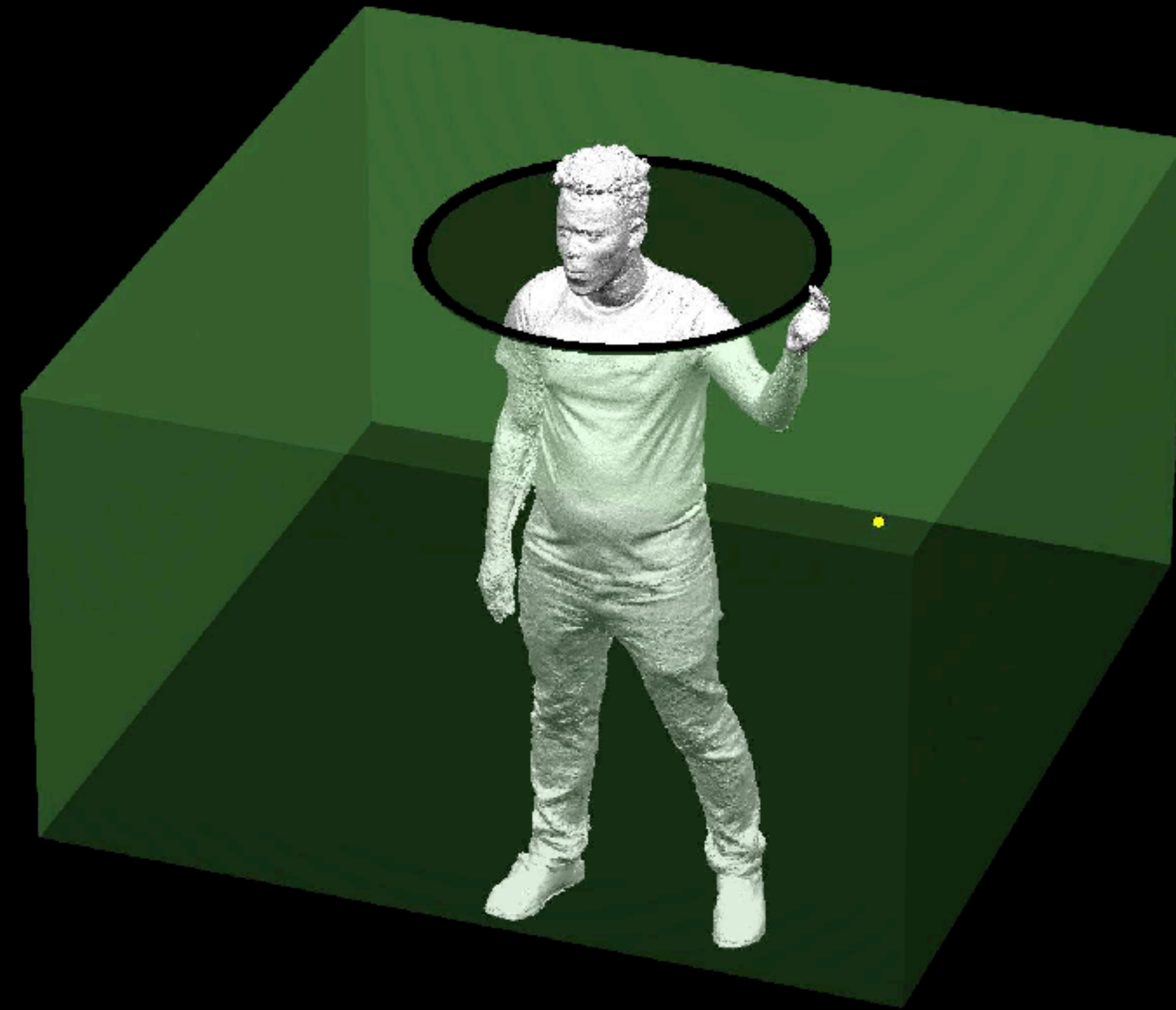
SoundField(x, y, z, t)



**SPEECH**



SoundField(x, y, z, t)

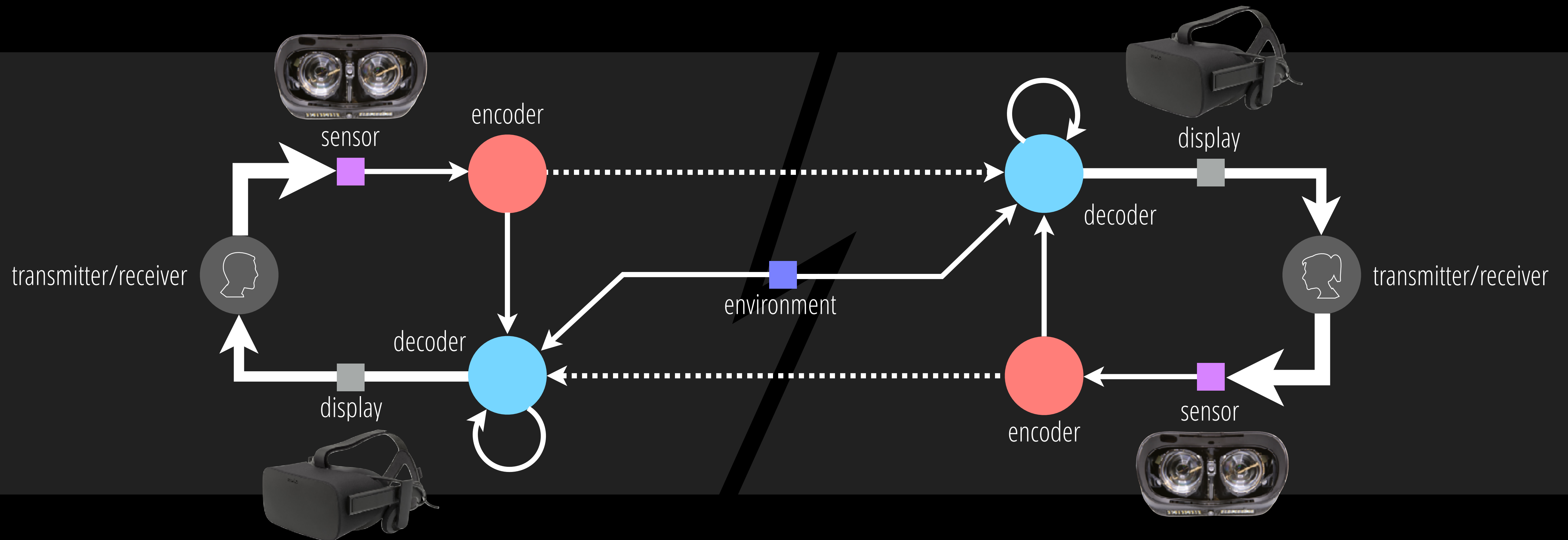


**FINGER SNAP**



# WHAT IS A CODEC AVATAR?

Social Interaction as an Information Network





“In a few years, men will be able to communicate more effectively through a machine than face to face...”

J. C. R. Licklider and R. W. Taylor

**The Computer As a Communication Device**

Science and Technology (1968)

Postal Service



550 BC

Telegraph



1840s

Telephone



1880s

Video Conferencing



1950s

Metric  
Telepresence









Schwartz et al. "The Eyes Have It: An Integrated Eye and Face Model for Photorealistic Facial Animation," SIGGRAPH 2020

Wei et al. "VR Facial Animation via Multiview Image Translation," SIGGRAPH 2019

Lombardi et al. "Neural Volumes: Learning Dynamic Renderable Volumes from Images," SIGGRAPH 2019

Nam, Wu, Kim, Sheikh, "Strand-accurate Multi-View Hair Capture," CVPR 2019

Lombardi, Simon, Saragih, and Sheikh, "Deep Appearance Models for Facial Rendering," SIGGRAPH 2018

Wu, Shiratori, Sheikh, "Deep incremental learning for efficient high-fidelity face tracking," SIGGRAPH Asia 2018

Bagautdinov, Wu, Saragih, Sheikh, "Modeling Facial Geometry using Compositional VAEs," CVPR 2018

Joo, Simon, Sheikh, "Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies," CVPR 2018

Poms, Wu, Yu, Sheikh, "Learning Patch Reconstructability for Accelerating Multi-View Stereo," CVPR 2018

Dong et al., "Supervision-by-Registration: An unsupervised approach to improve the precision ...," CVPR 2018

Bansal, Ma, Ramanan, Sheikh "Recycle GANs," ECCV 2018

[yasers@fb.com](mailto:yasers@fb.com)

Facebook Reality Labs (Pittsburgh)





P I T T S B U R G H