Lecture 11:

The Present and Future of Video Conferencing Systems

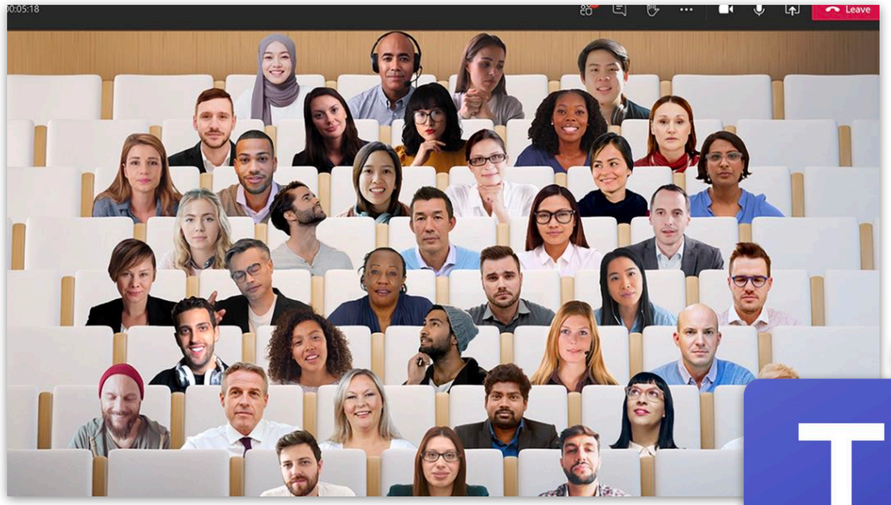
Visual Computing Systems
Stanford CS348K, Spring 2023

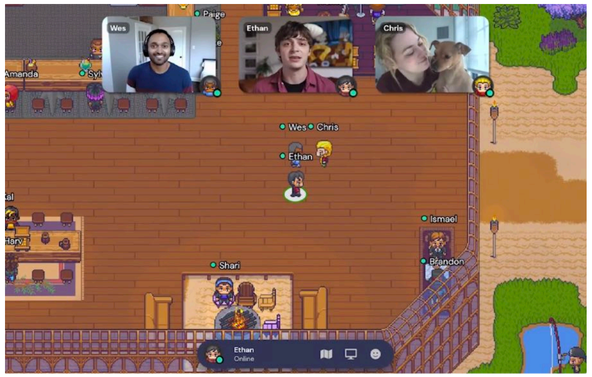
Today's agenda

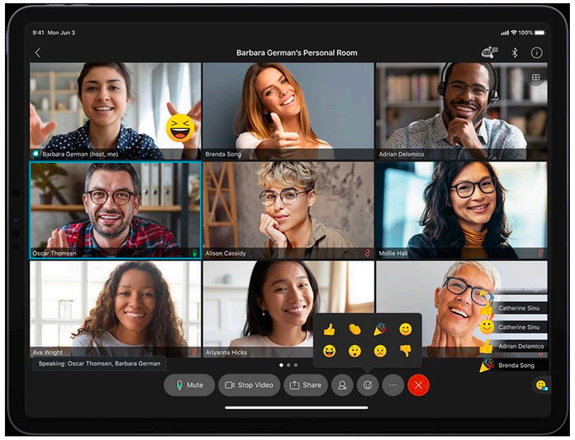
- Google VCU paper discussion (Ranganathan et al. 2021)
- Design of modern video conferencing systems
- Discussion of privacy in an always on video world

Videoconferencing systems

As you can imagine, a lot of modern interest in video conferencing (big and small!)







Webex Meetings

Gather

BlueJeans

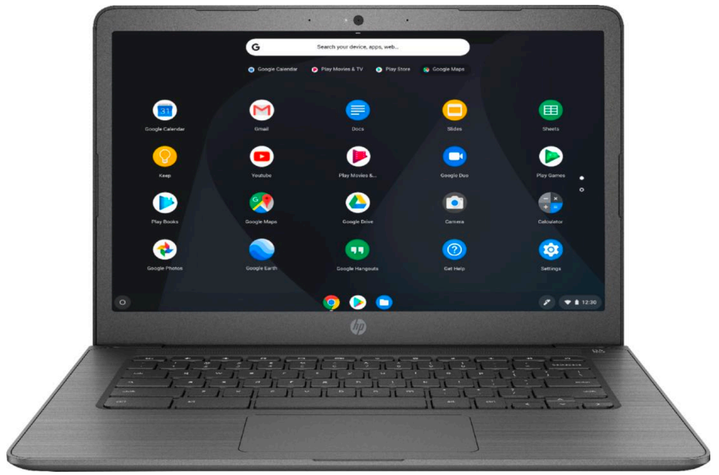
nooks

ohyay

Let’s design a video conferencing system

We want to deliver a visually rich experience similar to features of modern platforms

Deliver to wide range of clients and network settings



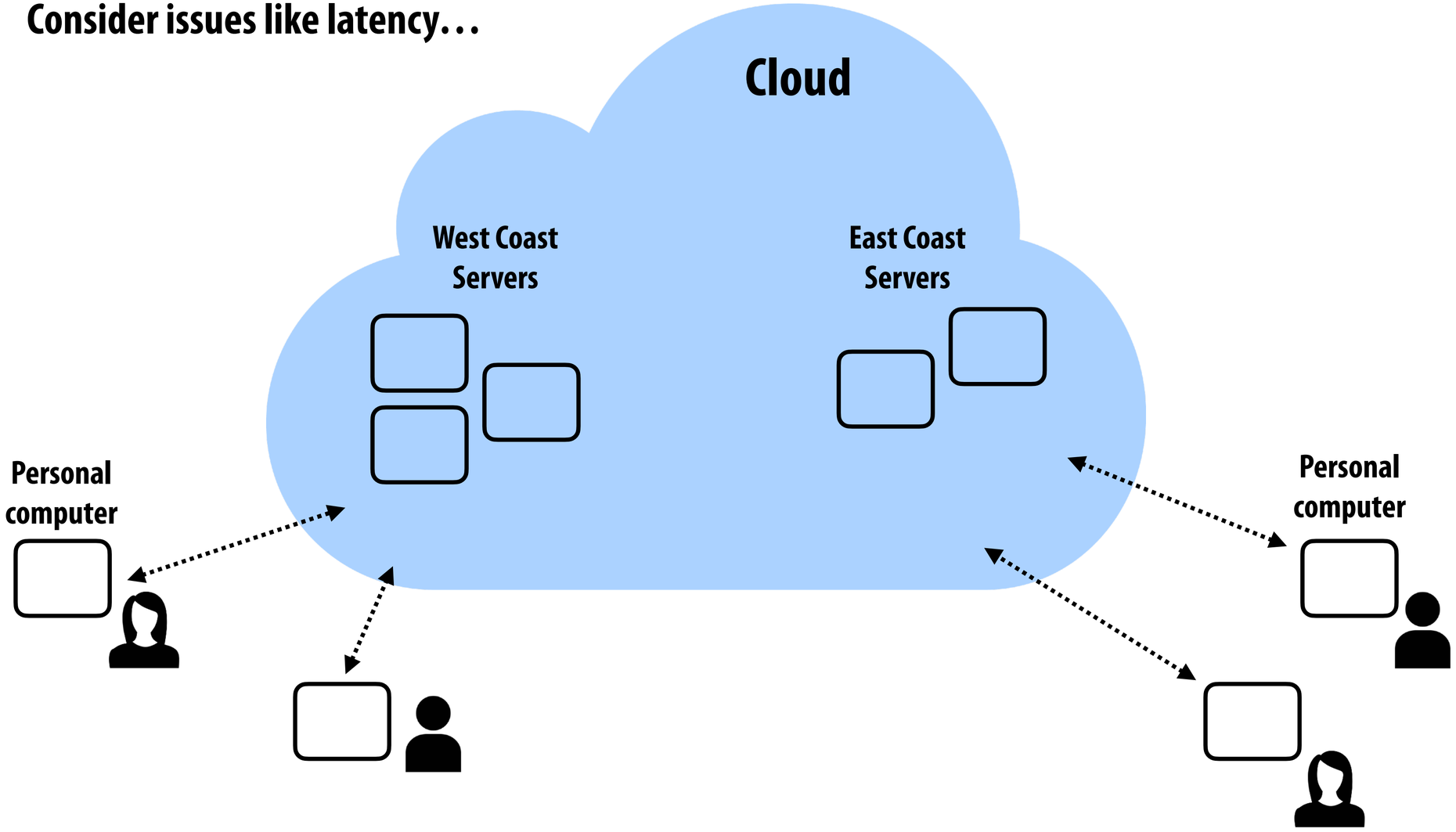
Let's design a video conferencing system

Large gallery views: companies raced to provide 7x7 gallery in 2020 *

*** we'll question whether this is a good idea later**

Setup…

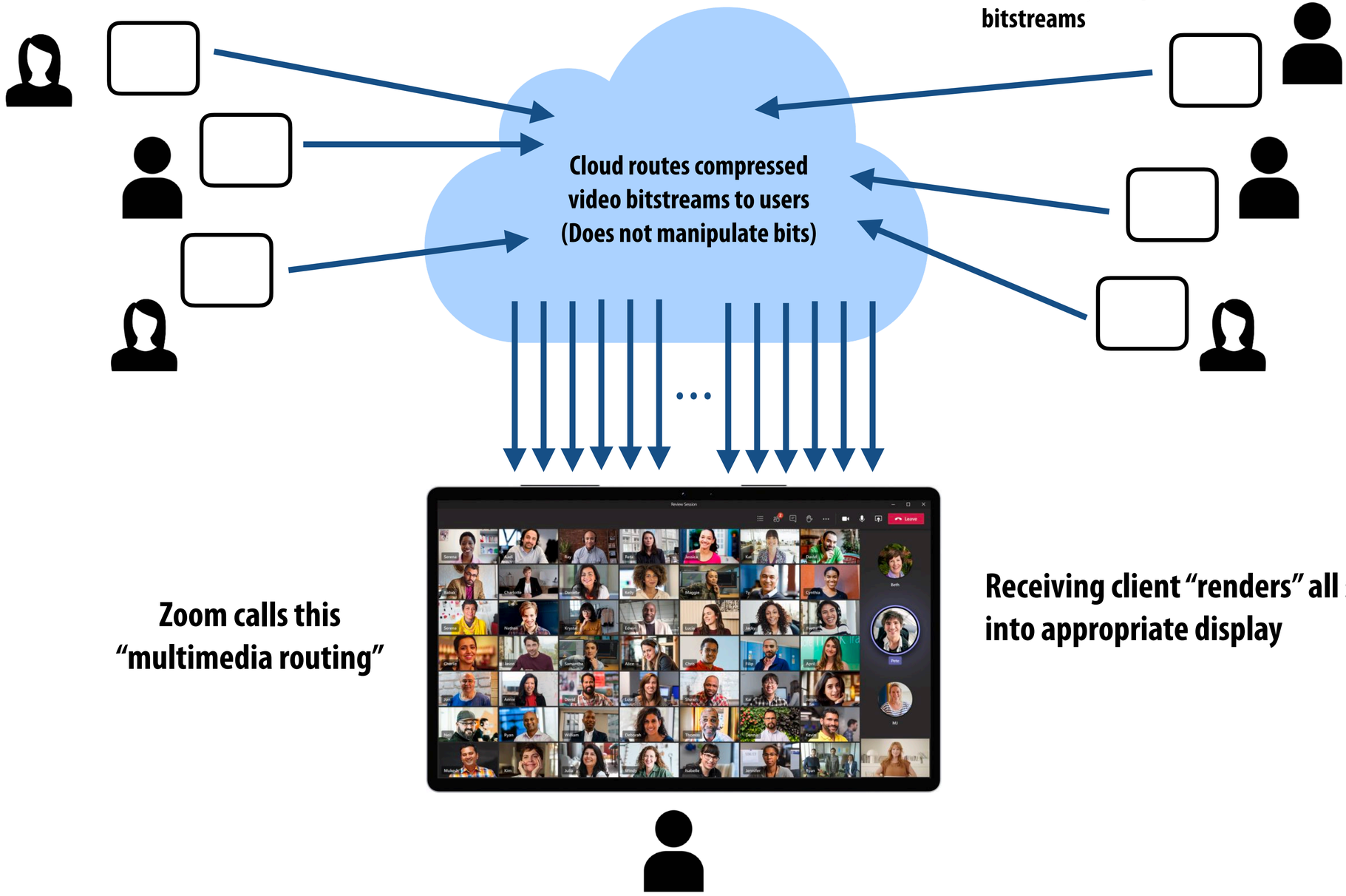
Consider issues like latency…



Q. Should we transcode/process video on our cloud servers?

- What are advantages (to users? To the service provider)?
- What are disadvantages?

Implementing gallery view



One drawback of this design

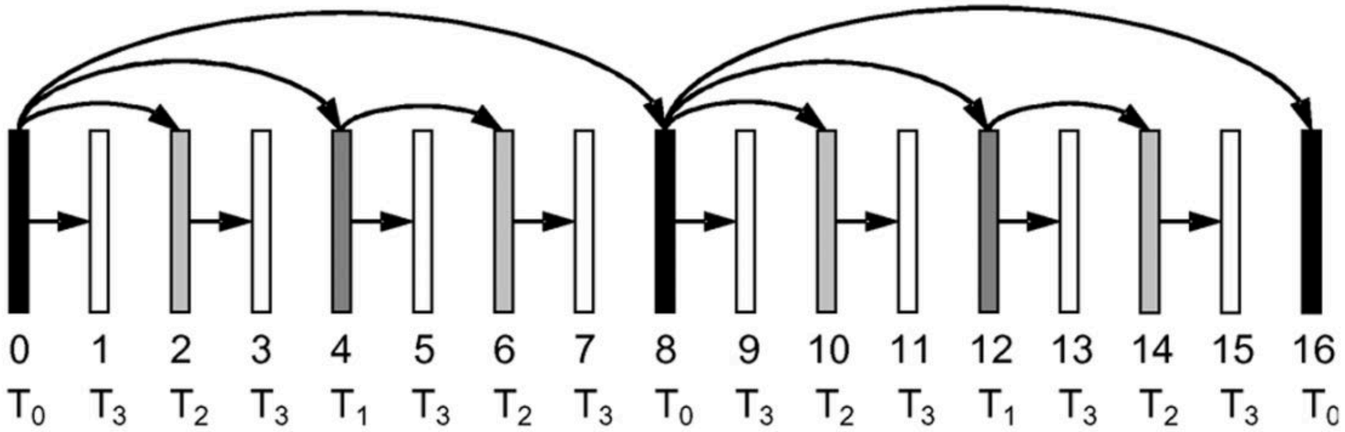
- If each client is providing a single compressed video stream, that means each person on the video call must receive the same bits right? (What if they are on different network connections?)

Scalable video codec (SVC)

SVC is an extension of H.264 standard

- "Scalable" compressed video bitstream: subsets of the bitstream encode valid video streams for a decoder
  - Implication: if packets get lost, the remaining packets form a valid H.264 bitstream, albeit at lower resolution or quality

Example: temporal scalability



Layer 0: (T_0) defines valid video at frame rate R
Layer 1 (T_1) defines bumps frame rate to 2R
…

Note how layer 0 information is used to predict higher layer information

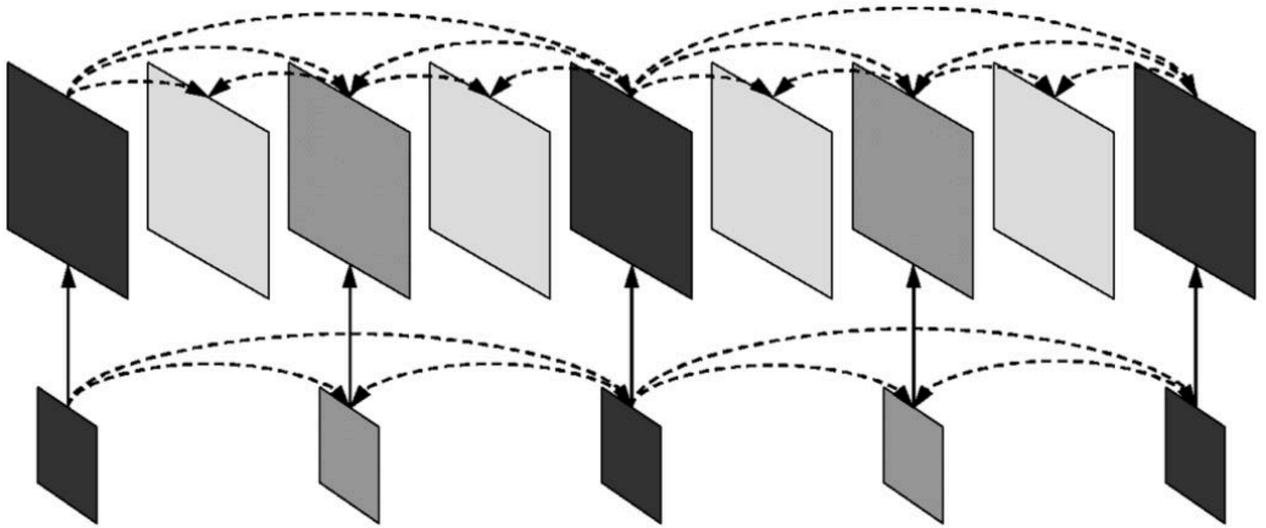
Scalable video codec (SVC)

SVC is an extension of H.264 standard

Example: spatial scalability

Layer 1:
(Higher res)

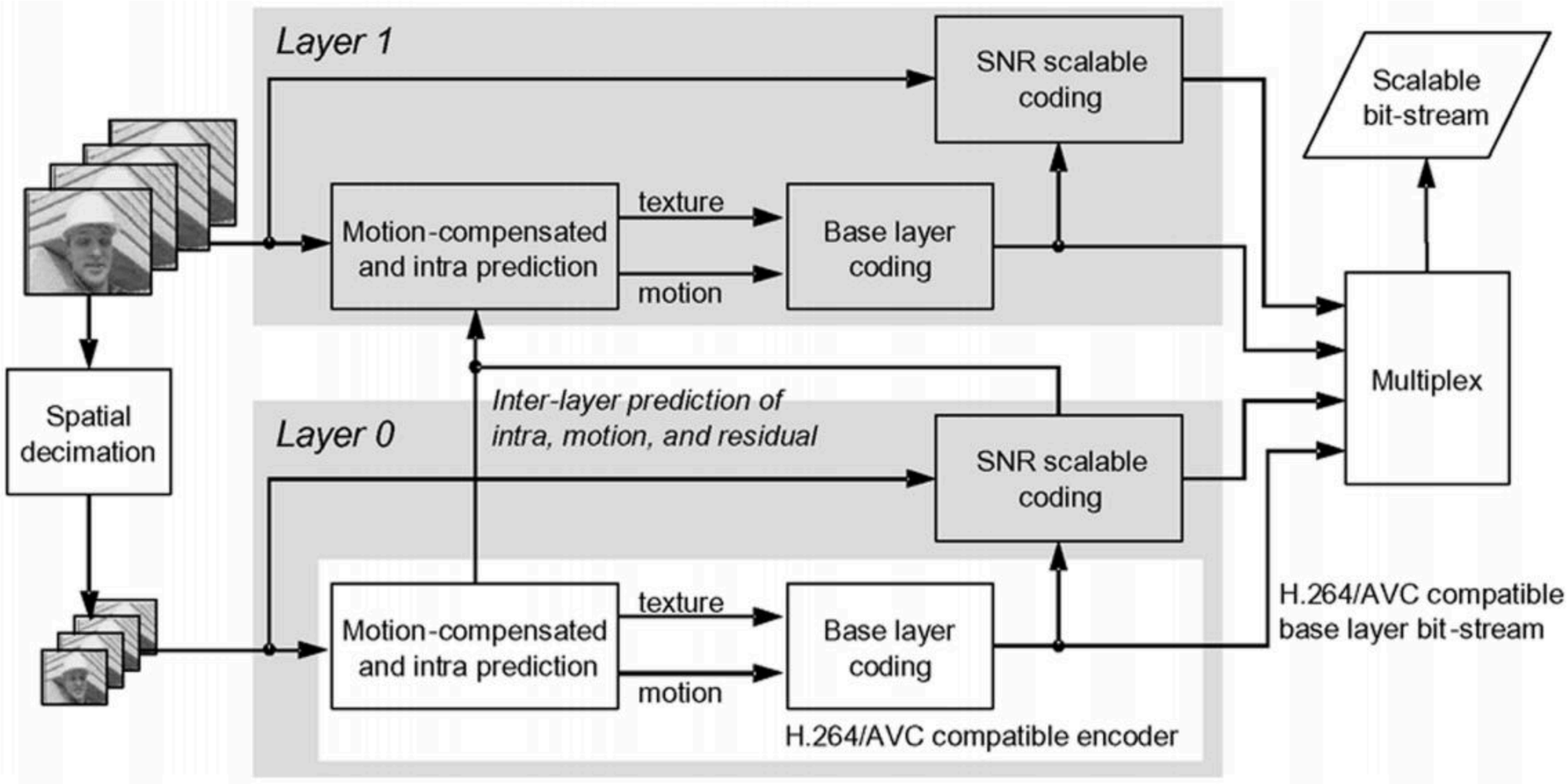
Layer 0:
(Low res)



Again, note how layer 0 information is used to predict higher layer information
(Higher efficiency than independently encoding two video streams)

Layer 0: defines valid video at low resolution (and low frame rate)
Layer 1: provides additional information for higher resolution (and higher frame rate) video

Scalable video codec (SVC) encoder



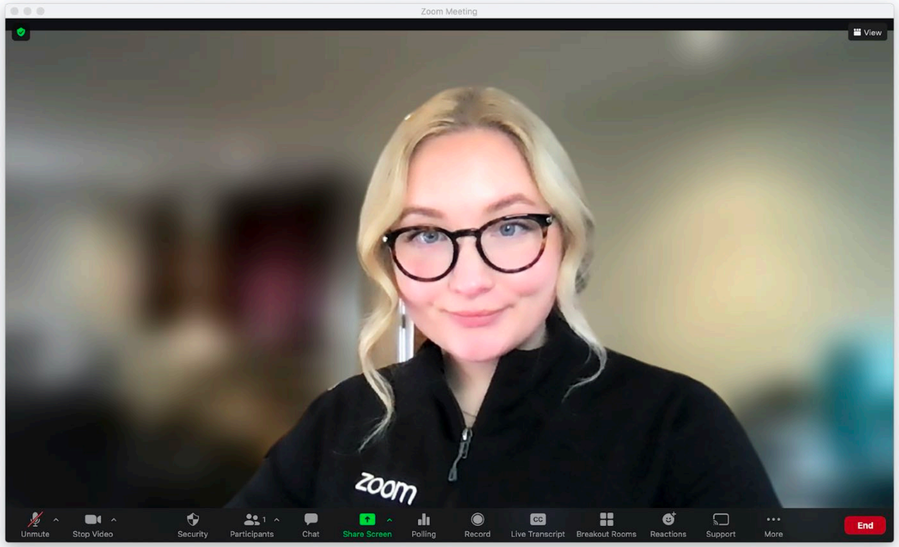
Costs: higher encoding/decoding costs

(But possible on modern clients as SVC is supported in hardware)

Need for audio/video analysis (e.g., effects)

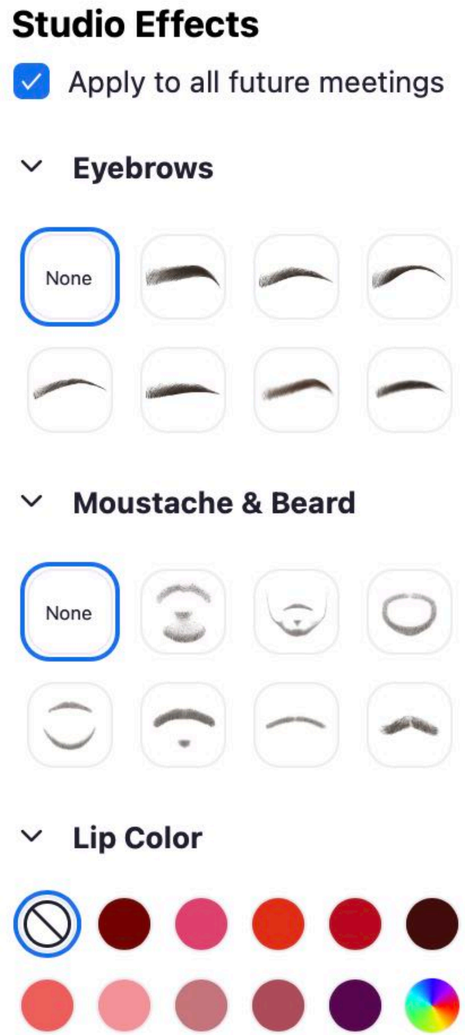
Effect: segment participant from background

Enhancing the appearance of video feeds

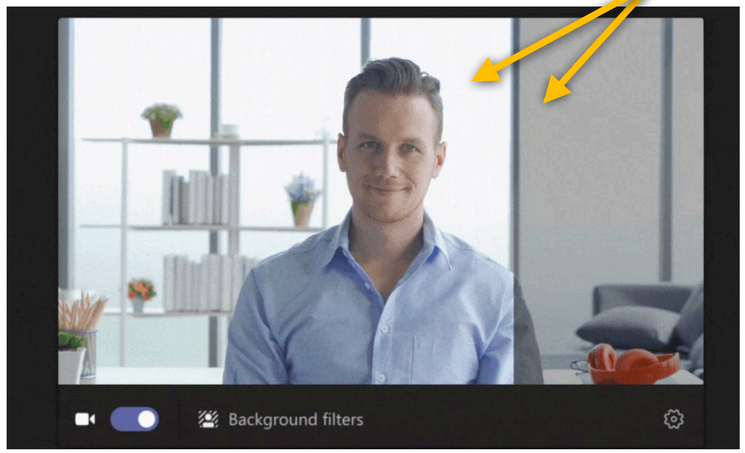


Blur background

Render additional content

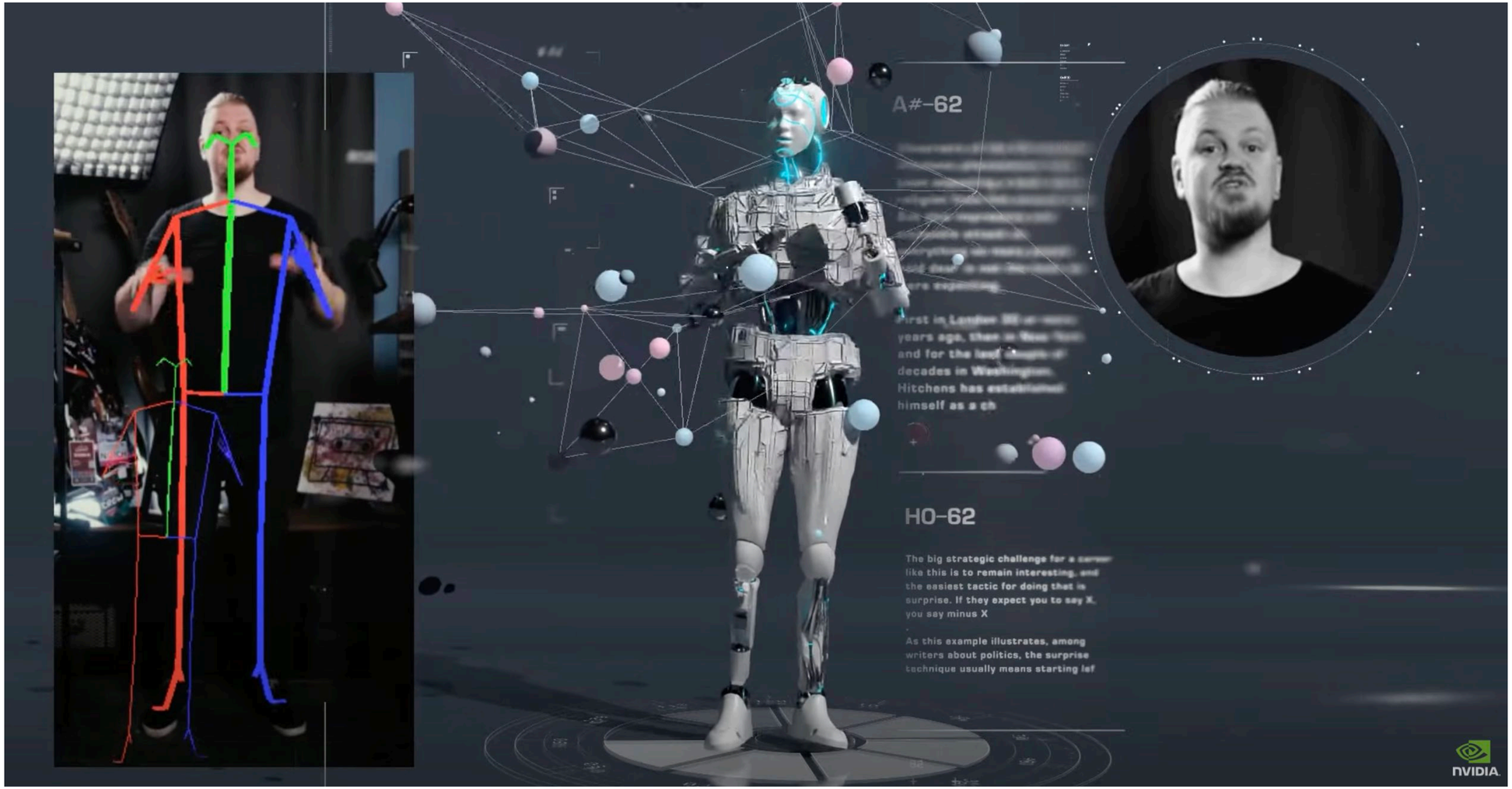


Adjust lighting



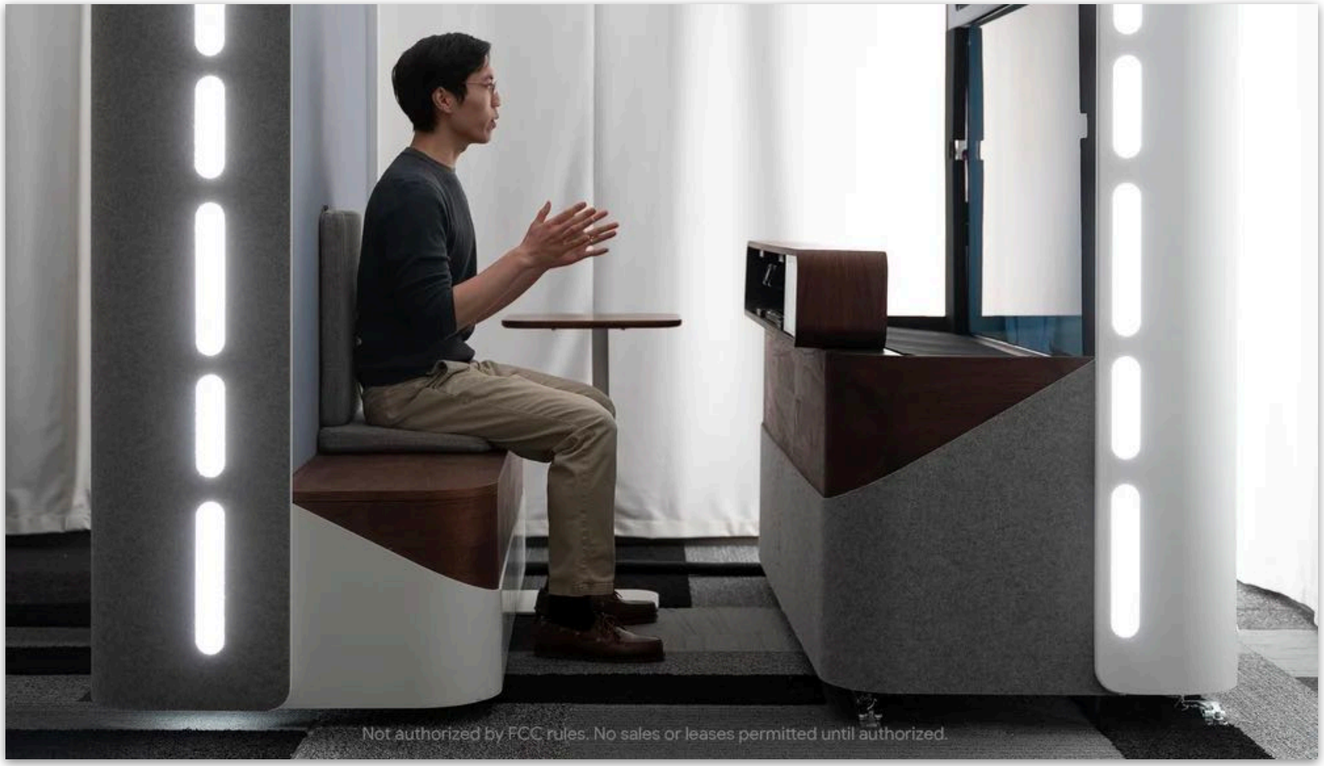
NVIDIA Maxine

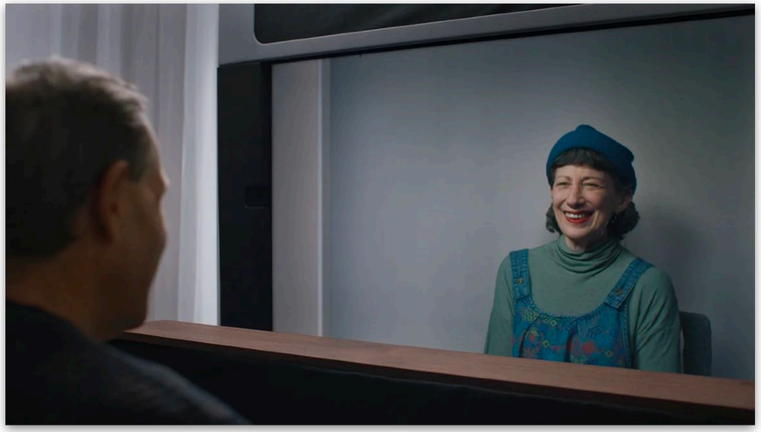
GPU-accelerated video processing for video conferencing applications



Examples: avatar control, video superresolution, advanced background segmentation

Project Starline





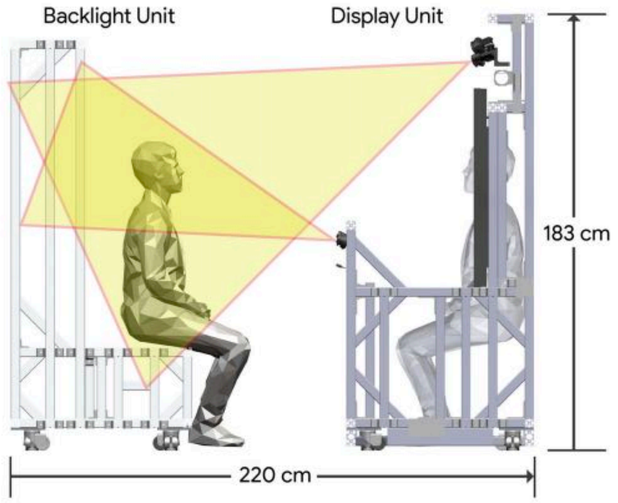
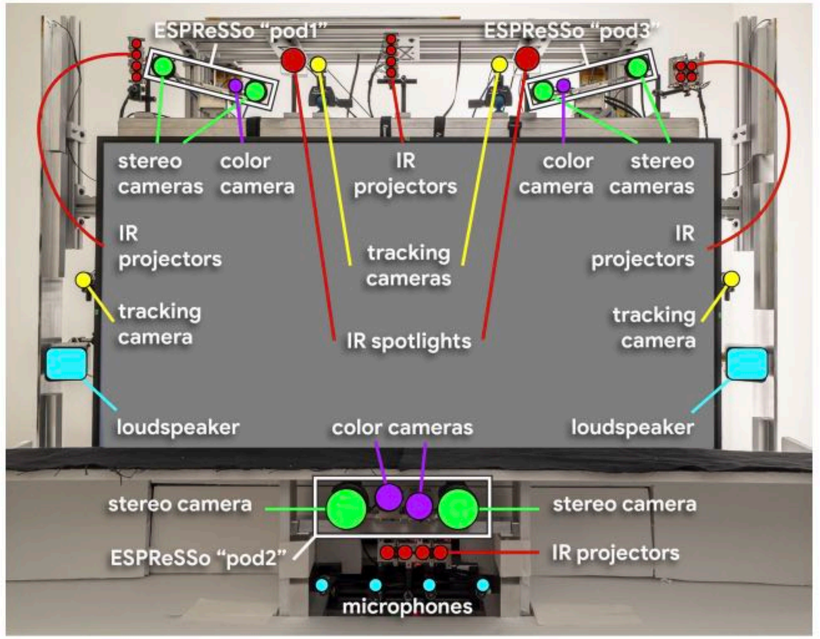


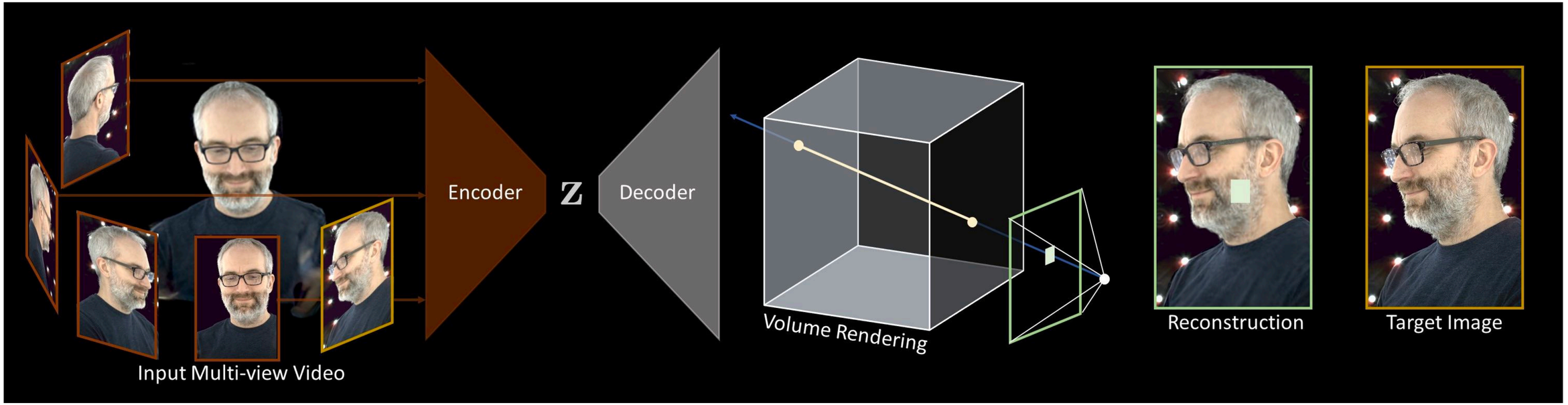
Fig. 4. Side-elevation view of our prototype system, illustrating the relative placement of the user, cameras, display, and virtual remote participant.



[Lawrence et al. 2021]

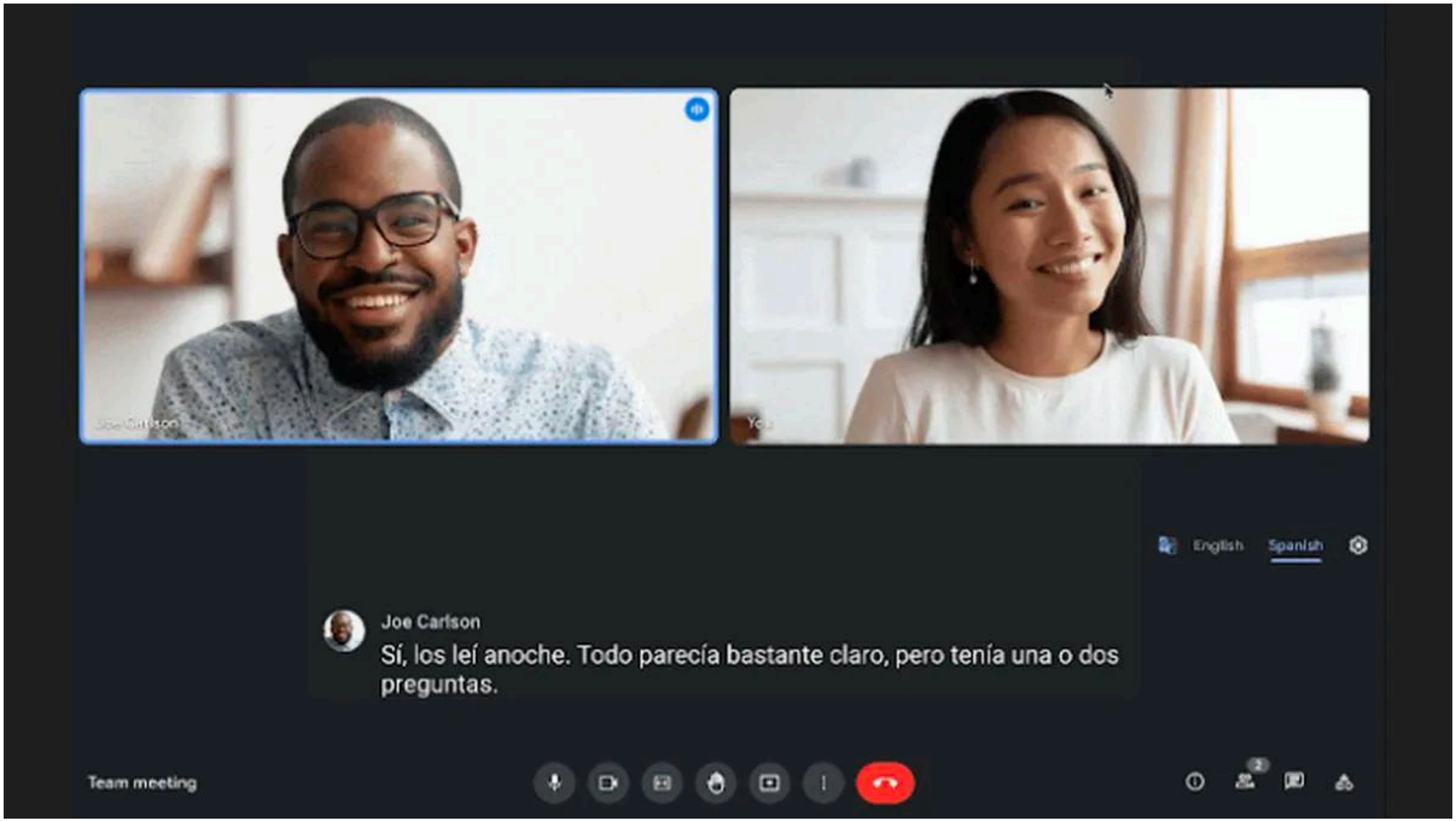
Neural volumes

- Learn to encode multiple views of a person into a latency code (z) that is decoded into a volume than can be rendered with conventional graphics techniques *from any viewpoint*



- Motivated by VR applications

Other forms of augmentation



Real-time translation and captioning

What do people *really need*?

“Zoom fatigue” is very real

[Bailenson 2021]

FEBRUARY 23, 2021

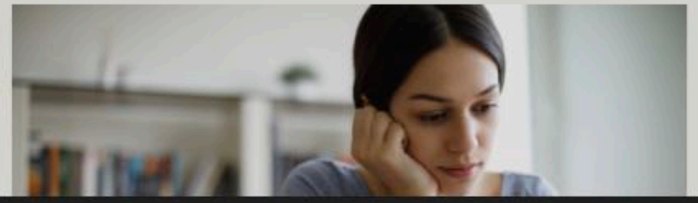
Stanford researchers identify four causes for ‘Zoom fatigue’ and their simple fixes

It’s not just Zoom. Popular video chat platforms have design flaws that exhaust the human mind and body. But there are easy ways to mitigate their effects.

BY VIGNESH RAMACHANDRAN

Even as more people are logging onto popular video chat platforms to connect with colleagues, family and friends during the COVID-19 pandemic, Stanford researchers have a warning for you: Those video calls are likely tiring you out.

Prompted by the recent boom in videoconferencing, communication Professor Jeremy Bailenson, founding director of the Stanford Virtual Human Interaction Lab (VHIL), examined the psychological



1) Excessive amounts of close-up eye contact is highly intense.

2) Seeing yourself during video chats constantly in real-time is fatiguing.

3) Video chats dramatically reduce our usual mobility.

4) The cognitive load is much higher in video chats.

The best camera is the one that's off?

Yes, you can make a Zoom background of yourself pretending to pay attention

And it's surprisingly easy to do, too.

Brian Lloyd
2 years ago

Share

HOW-TO

Best funny Zoom background trick: Put yourself in a looping video so you can skip the meeting

Now you can duck out on those hourlong conference calls.

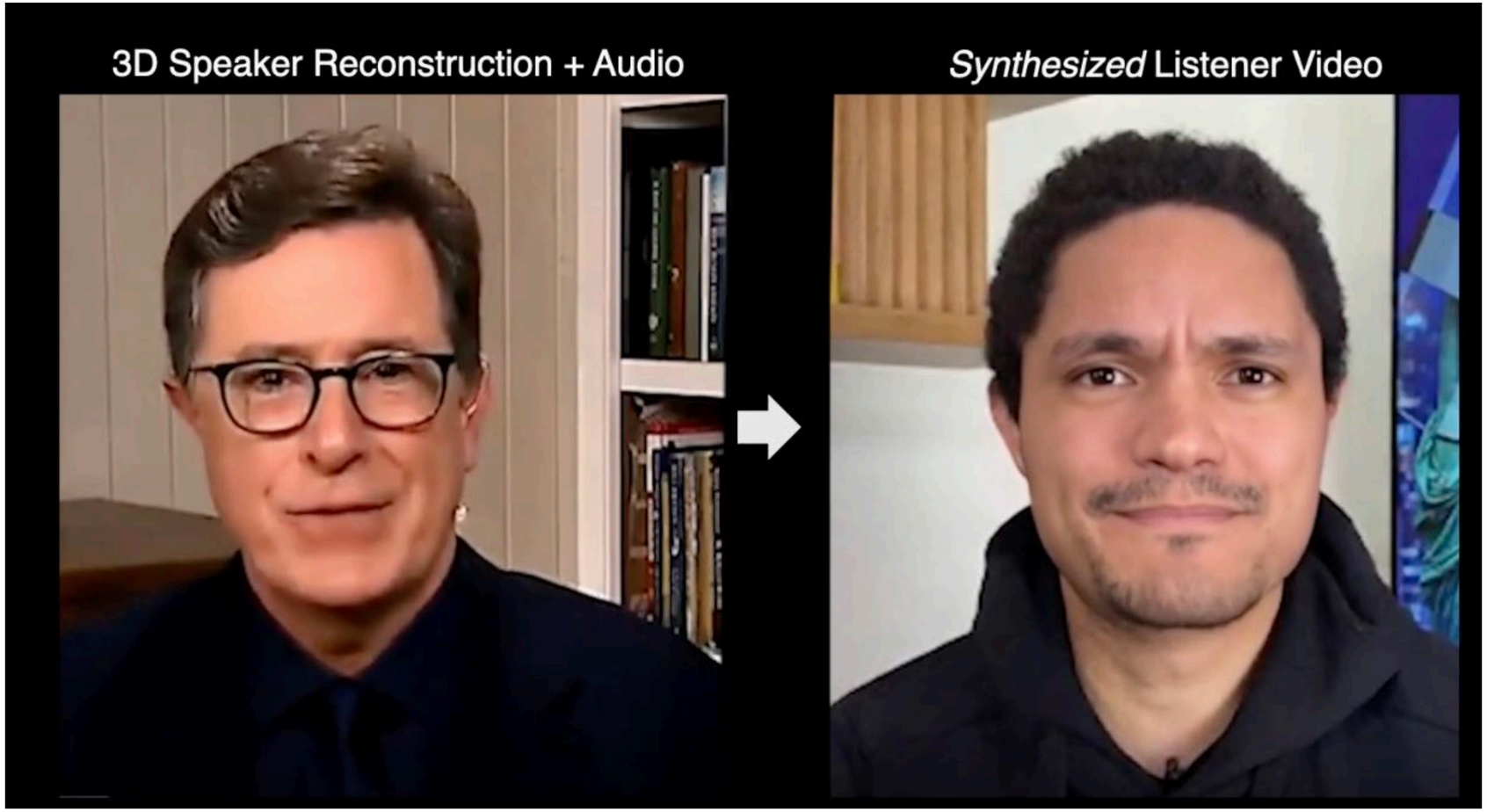
By Gordon Ung
Executive Editor, PCWorld | APR 13, 2020 3:30 AM PDT

We've all been in Zoom video conference meetings that drag on longer than a bad

Synthesizing reactions?

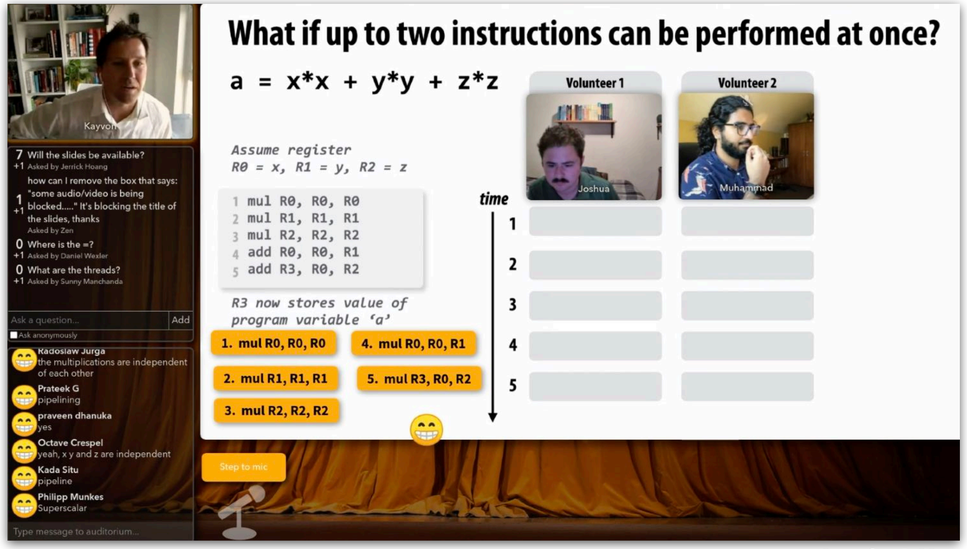
Input: audio of speaker

Output: video of listener's reaction



User-triggered effects (examples: audio clips, “reactions”)

Ways to fidget, visual distractions can be helpful…



Discussion:

Where is the ethical line between "augmenting or abstracting what's real" and "fake"?

How can more advanced technology help strike a better balance between facilitating better forms of communication + a sense of presence (e.g., working from home) vs. ensuring privacy and personal space?

What harms could widespread use of near-photorealistic digital personal avatars (e.g., for work calls) cause? What are possible benefits?