Habitus: Boosting Mobile Immersive Content Delivery through Full-body Pose Tracking and Multipath Networking

Anlan Zhang[1*] Chendong Wang[4] Yuming Hu[2] Ahmad Hassan[1*]
Zejun Zhang[1*] Bo Han[3] Feng Qian[1*] Shichang Xu[5]
[1]University of Southern California [2]University of Minnesota, Twin Cities
[3]George Mason University [4]University of Wisconsin-Madison [5]Google

April, 2024

Immersive Content is Everywhere

- 3-DoF (degree-of-freedom) to 6-DoF motion
  - x, y, z, yall, pitch, roll
- Bandwidth-intensive
  - Hard to deliver through common wireless links (e.g., 802.11ac)

360° Videos [1]

VR Games [2]

Volumetric Content [3]

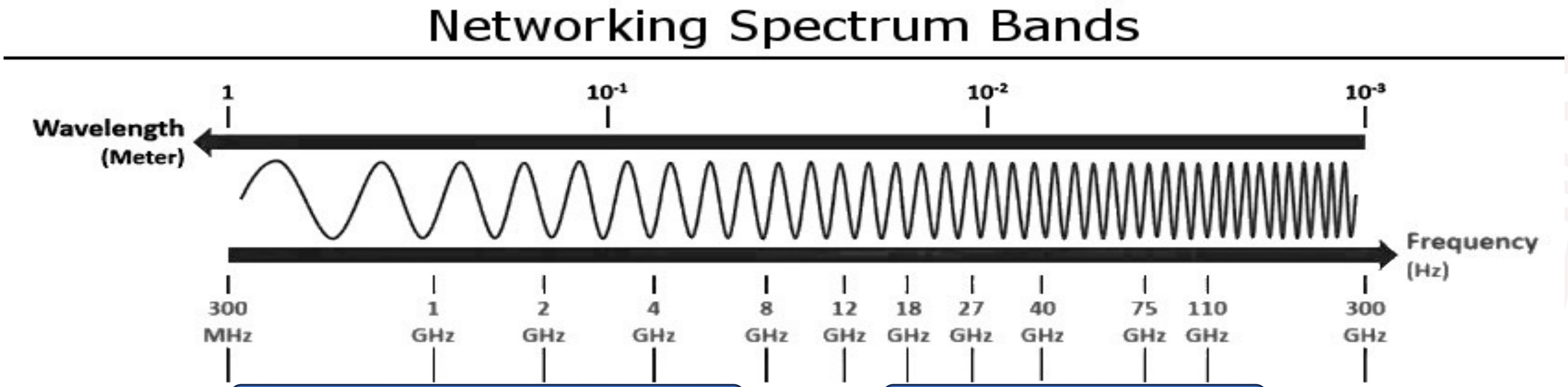
Media sources:

[1] https://giphy.com/gifs/virtual-tour-jkpg360-virtuell-rundtur-r2ddbd3VMZLpfrKkz7

[2] https://80.lv/articles/making-vfx-for-vr-first-person-shooter/

[3] https://www.youtube.com/watch?v=aO3TAke7_MI

Omnidirectional vs. mmWave Radio



Omnidirectional	mmWave
4G/LTE, low-band/midband 5G 802.11 (WiFi) a/b/g/n/ac/ax	mmWave 5G 802.11 (WiFi) ad/ay
Low Bandwidth Slow signal attenuation Less vulnerable to blockage	High Bandwidth Fast signal attenuation More vulnerable to blockage

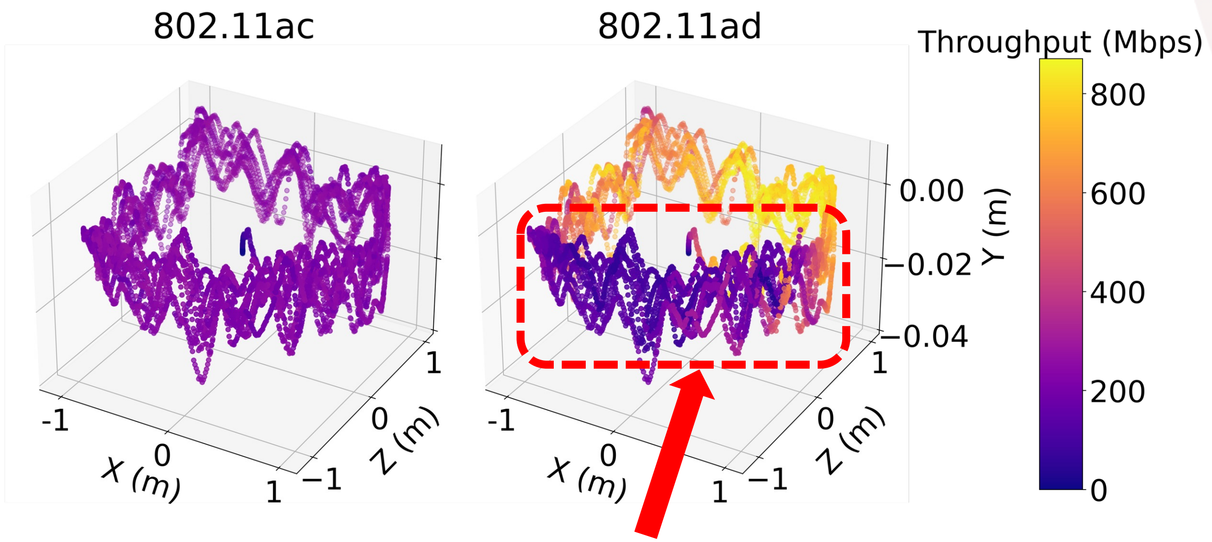
Promising for streaming bw-hungry immersive content!

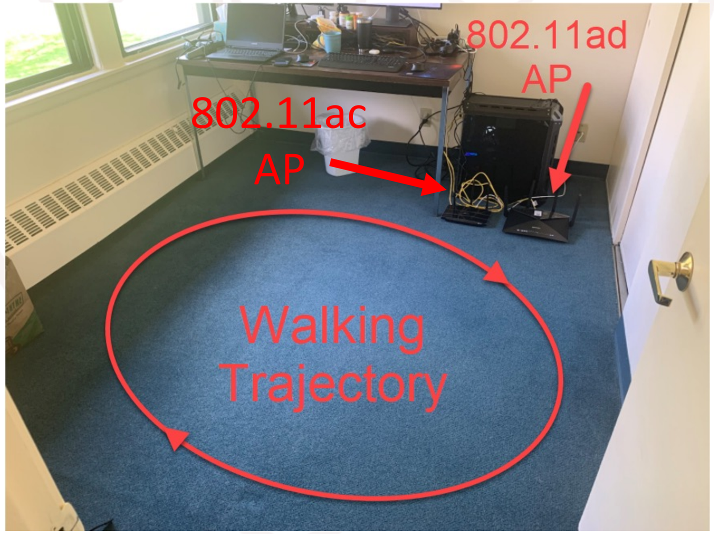
USC

Image sources:
[1] https://www.androidauthority.com/sub-6ghz-vs-mmwave-5g-3051246/

Case Study: Volumetric Video Streaming over mmWave

- 802.11ad (60 GHz) vs. 802.11ac (5 GHz)
- Test app: [ViVo, MobiCom'20]
- Impact on QoE (quality-of-experience)
    - Video quality +113%, stall +502%

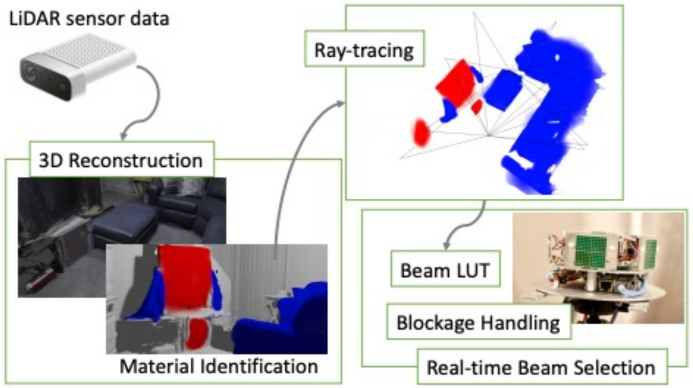




Human body blocks mmWave signal

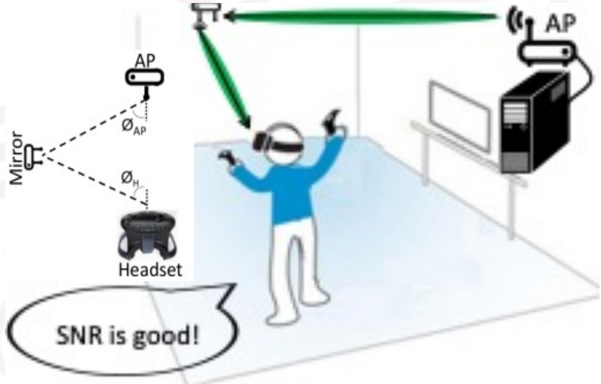
USC

Existing Systems Using mmWave



Improving the PHY layer,
e.g., SpaceBeam [MobiSys'21]

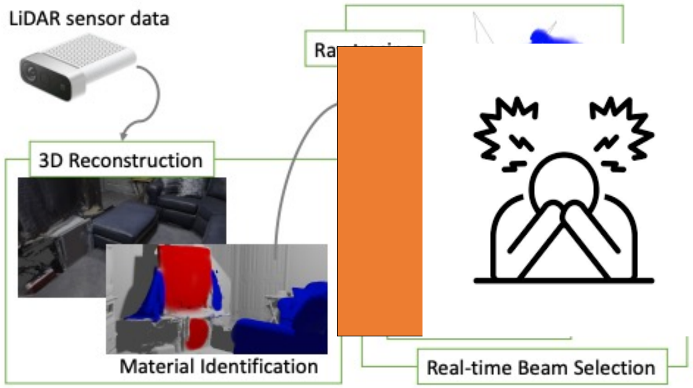
Enhancing line-of-sight (LoS),
e.g, VIVE Wireless Adapter [2]



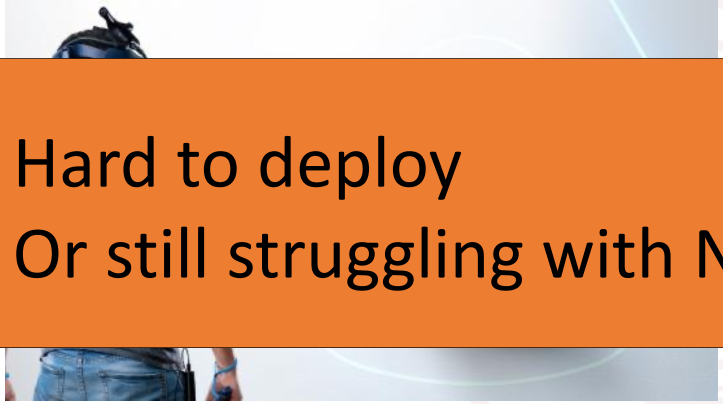
Using specialized device,
e.g., MoVR [NSDI'17]

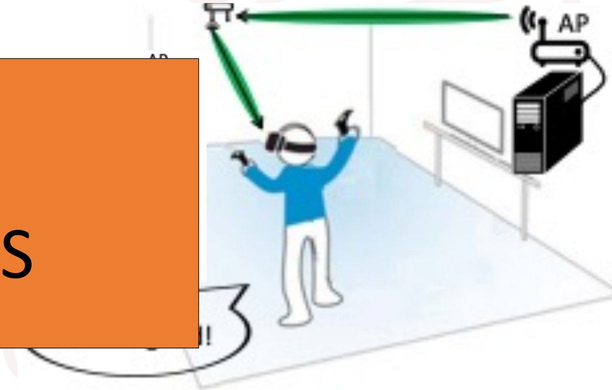
[1] Woodford, Timothy, et al. "Spacebeam: Lidar-driven one-shot mmwave beam management." *ACM MobiSys*. 2021.
[2] https://www.pcgamesn.com/vive-wireless-adapter-is-how-much
[3] Abari, Omid, et al. "Enabling {high-quality} untethered virtual reality." NSDI. 2017.

Existing Systems Using mmWave



- Hard to deploy
- Or still struggling with NLoS





Improving the PHY layer,
e.g., SpaceBeam [MobiSys'21]

Enhancing line-of-sight (LoS),
e.g, VIVE Wireless Adapter [2]

Using specialized device,
e.g., MoVR [NSDI'17]

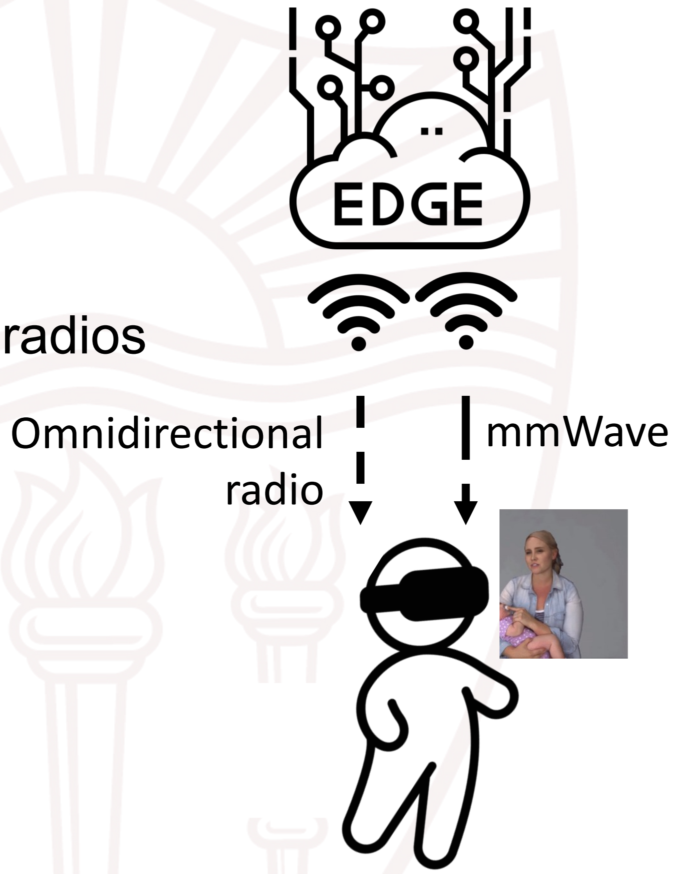
We need a robust and easy-to-deploy solution!

USC

[1] Woodford, Timothy, et al. "Spacebeam: Lidar-driven one-shot mmwave beam management." *ACM MobiSys*. 2021.
[2] https://www.pcgamesn.com/vive-wireless-adapter-is-how-much
[3] Abari, Omid, et al. "Enabling {high-quality} untethered virtual reality." NSDI. 2017.

Our Solution: Habitus

- Multipath networking over heterogeneous links
  - Omnidirectional (e.g. 802.11ac) + mmWave (e.g., 802.11ad) radios
- Actively predicting the fluctuating mmWave throughput
  - Under constant motion of the viewer

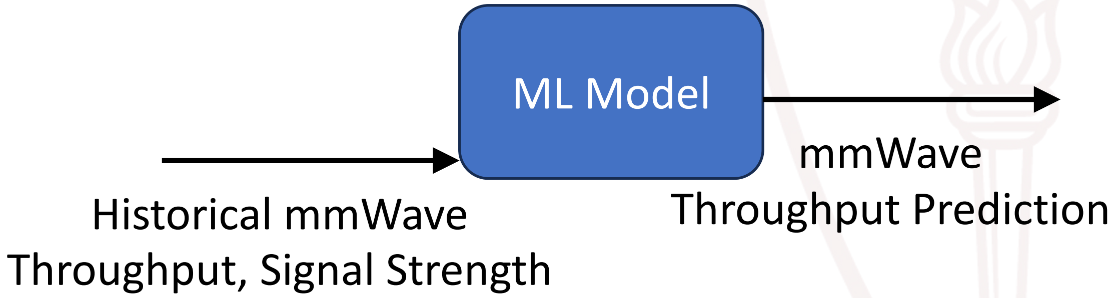


Challenges

- Predicting mmWave throughput w/ ML models
  - How to improve the accuracy under constant motion
- Applying offline pre-trained ML-based prediction models at runtime
  - How to react to unseen changes deviating from training data
- Heterogeneous (omnidirectional + mmWave) links
  - How to do multipath scheduling

Basic ML-based mmWave Throughput Prediction

- History network measurements → future network condition (e.g., throughput)



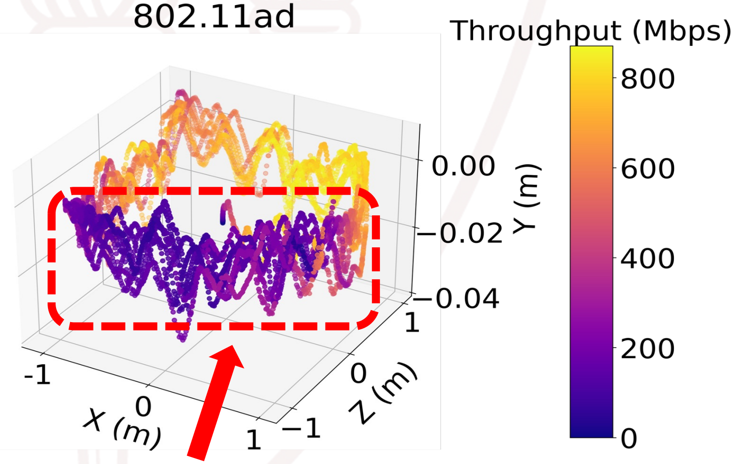
Motion-enhanced mmWave Throughput Prediction

- Insight 1: mmWave throughput is correlated w/ 6-DoF motions
  - Also validated by previous work: Lumos5G [IMC'20], Aggarwal *et al.* [PAM'21]

Historical 6-DoF head motion (yaw, pitch, roll, X, Y, Z)

Historical mmWave Throughput, Signal Strength

mmWave Throughput Prediction

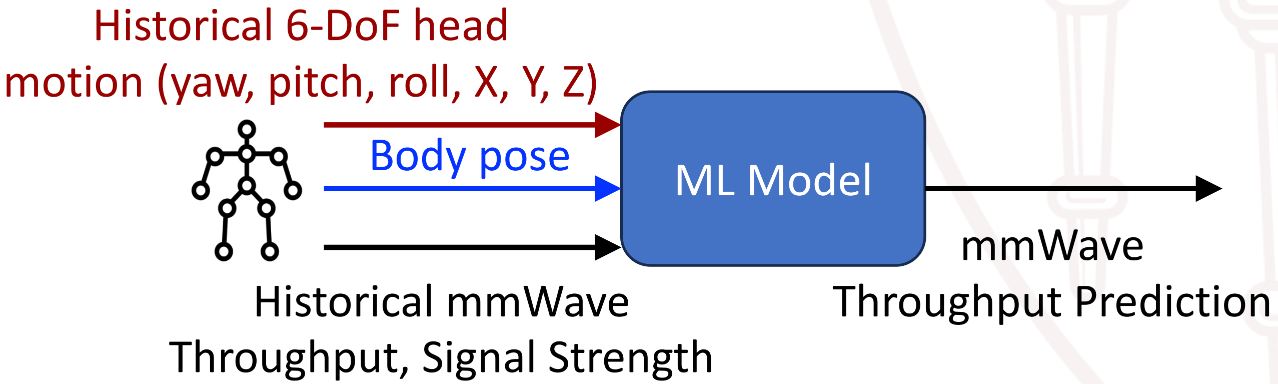


Human body blocks mmWave signal

[1] Narayanan, Arvind, et al. "Lumos5G: Mapping and predicting commercial mmWave 5G throughput." IMC. 2020.
[2] Aggarwal, Shivang, et al. "Throughput prediction on 60 GHz mobile devices for high-bandwidth, latency-sensitive applications." PAM, 2021.

Full-body Pose Guided mmWave Throughput Prediction

- Insight 2: Spatial correlation among body parts during human motion [1, 2]
  - Example: hand holding controller moves → head movement → throughput changes
  - Example: leg moves (e.g., viewer turns left) → head rotation → throughput changes
- Tracking full-body pose can improve mmWave throughput prediction
  - Body pose: a set of 3D key points



[1] Bak, Sławomir, et al. "Person re-identification using spatial covariance regions of human body parts." 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2010.
[2] Xu, Xinyu, and Baoxin Li. "Exploiting motion correlations in 3-D articulated human motion tracking." IEEE transactions on image processing 18.6 (2009): 1292-1303.

Full-body Pose Guided mmWave Throughput Prediction

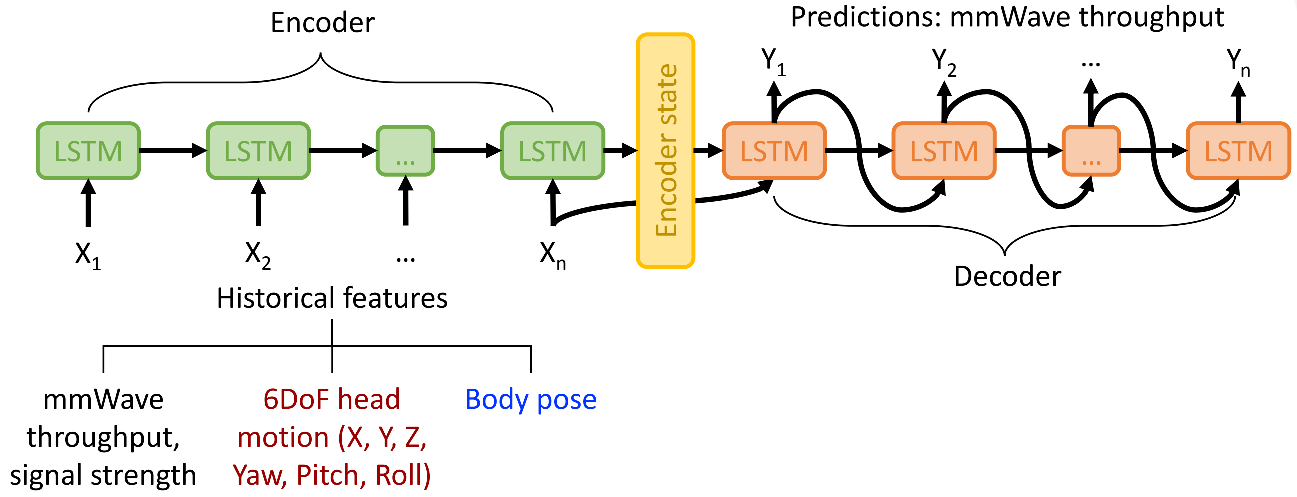
- Data collection at 4 locations w/ 802.11ac/ad APs + a stereo camera (for tracking pose)
  - 3 viewers (1.6m, 1.7m, 1.8m / 1 Female, 2 Males) exercise 10 motion patterns
    - Collect: ① 802.11ac/ad throughput & signal strength, ② 6DoF head motion, ③ full-body pose
  - More details are in the paper

Patterns	Description
S1	The user stands in the center of the room, turning around in a clockwise direction.
S2	The user stands in the center of the room, turning around in a counterclockwise direction.
S3	The user walks around in a clockwise direction.
S4	The user walks around in a counterclockwise direction in a normal speed.
S5	The same as S4, but in a slow speed.
S6	The same as S4, but in a fast speed.
S7	A chair occupies the front place of the access point. The user walks around in a counterclockwise direction.
S8	The same as S3, but the user does not change the orientation of his/her head.
S9	The same as S4, but the user does not change the orientation of his/her head.
S10	The user walks around following the walking trace in S7, but there is no chair.

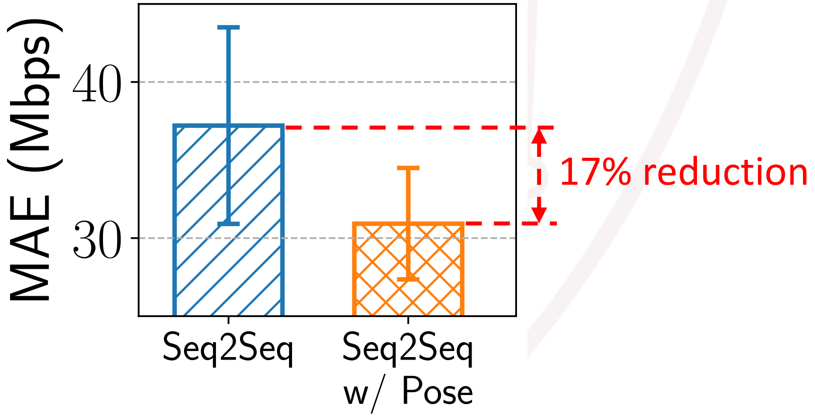
Table 1: User motion patterns.

Full-body Pose Guided mmWave Throughput Prediction

- Prediction target: mmWave throughput in the next 1 second
  - MAE (mean absolute error) of model w/o and w/ pose
    - Seq2Seq model
    - Other models: GBDT (gradient boosting decision tree), MLP, RNN in our paper
  - w/ Pose: 5% - 29% MAE reduction

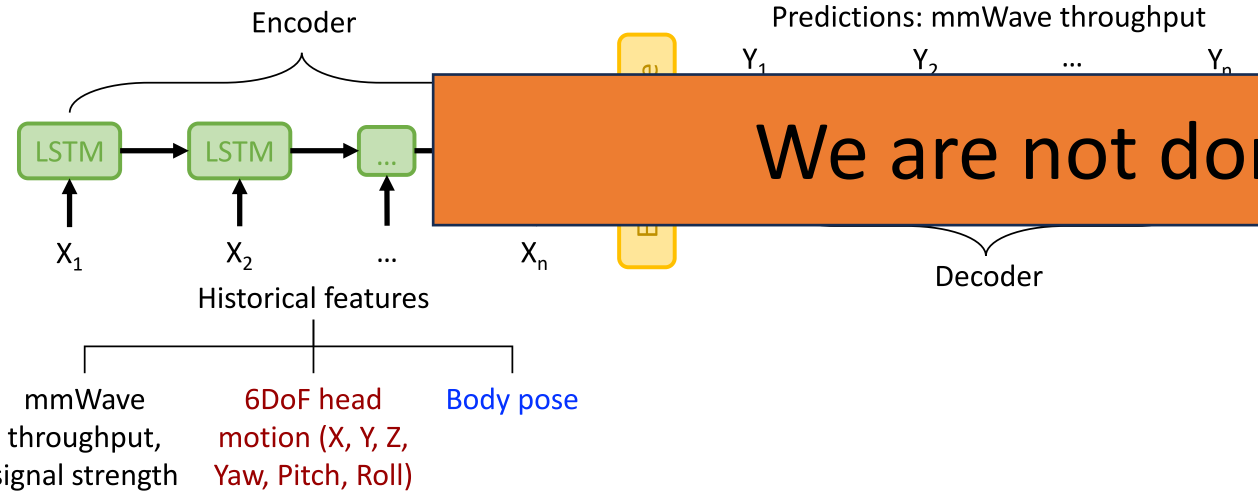


Seq2Seq model w/ Encoder-Decoder Architecture

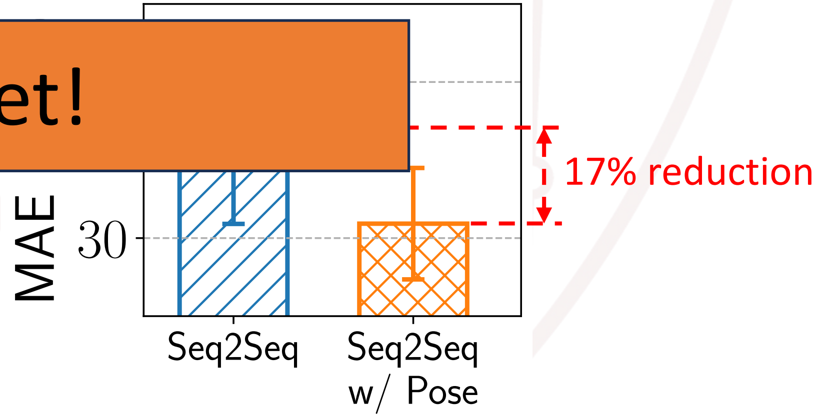


Full-body Pose Guided mmWave Throughput Prediction

- Prediction target: mmWave throughput in the next 1 second
  - MAE (mean absolute error) of model w/o and w/ pose
    - Seq2Seq model
    - Other models: GBDT (gradient boosting decision tree), MLP, RNN in our paper
  - w/ Pose: 5% - 29% MAE reduction



Seq2Seq model w/ Encoder-Decoder Architecture



We are not done yet!

Impact of Unseen Changes

- Use a pre-trained model at runtime: performance drop
  - Reason: cannot adapt to changes deviating from training data, e.g.,
    - C1: location change
    - C2: user change
    - C3: motion pattern change
    - C4: static environmental change (e.g., a chair being moved)
    - C5: dynamic environmental change (e.g., a walk-by spectator)
  - Intuition: There is fundamental knowledge learned by models
    - Physical property of mmWave
    - Throughput distribution in certain positions
  - Intuition: The pre-trained model has different sensitivities to these changes
    - The changes that reshape the fundamental knowledge have larger impact

USC

Impact of Unseen Changes

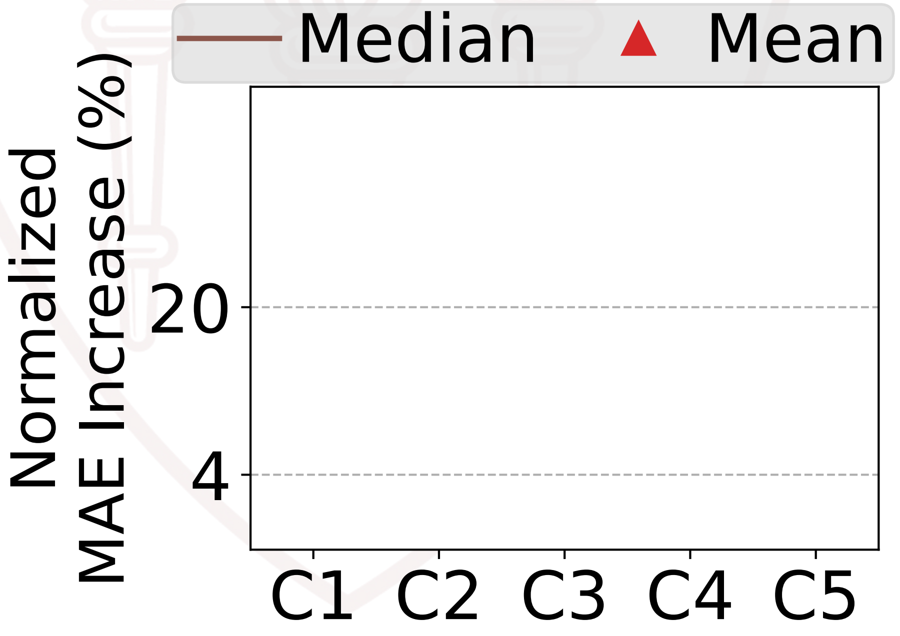
- Systematically quantify model sensitivity to the changes
  - Apply pre-trained models to
    - New location/user/motion pattern
    - Manually created static/dynamic environmental changes

Static environmental change

Dynamic environmental change

Impact of Unseen Changes

- Systematically quantify model sensitivity to the changes
  - C1: location change
  - C2: user change
  - C3: motion pattern change
  - C4: static environmental change
  - C5: dynamic environmental change
- Impact of the changes on MAE

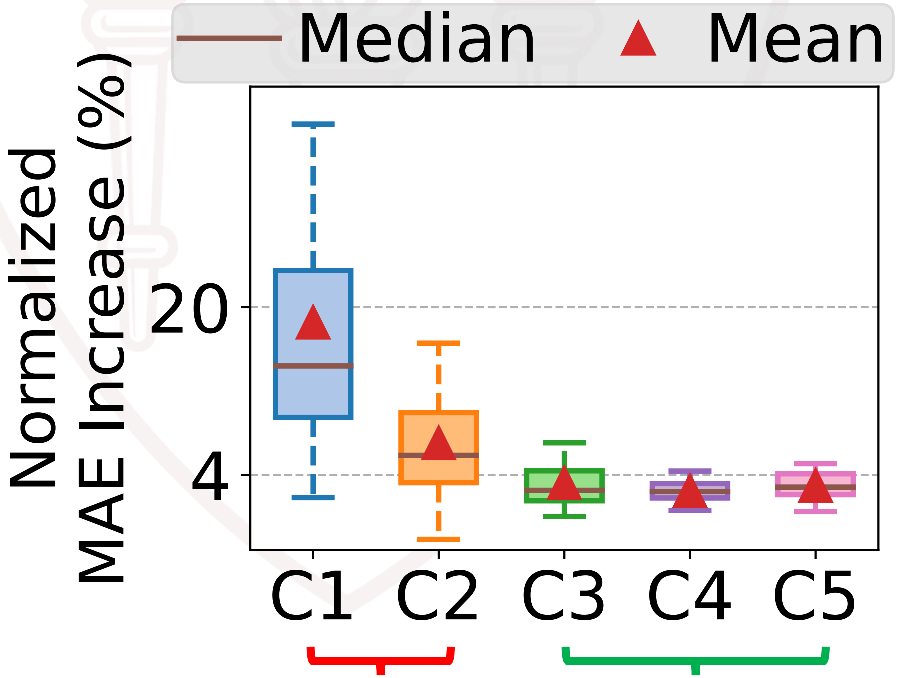


Impact of Unseen Changes

- Systematically quantify model sensitivity to the changes
  - C1: location change
  - C2: user change
  - C3: motion pattern change
  - C4: static environmental change
  - C5: dynamic environmental change
- Impact of the changes on MAE

Large impact

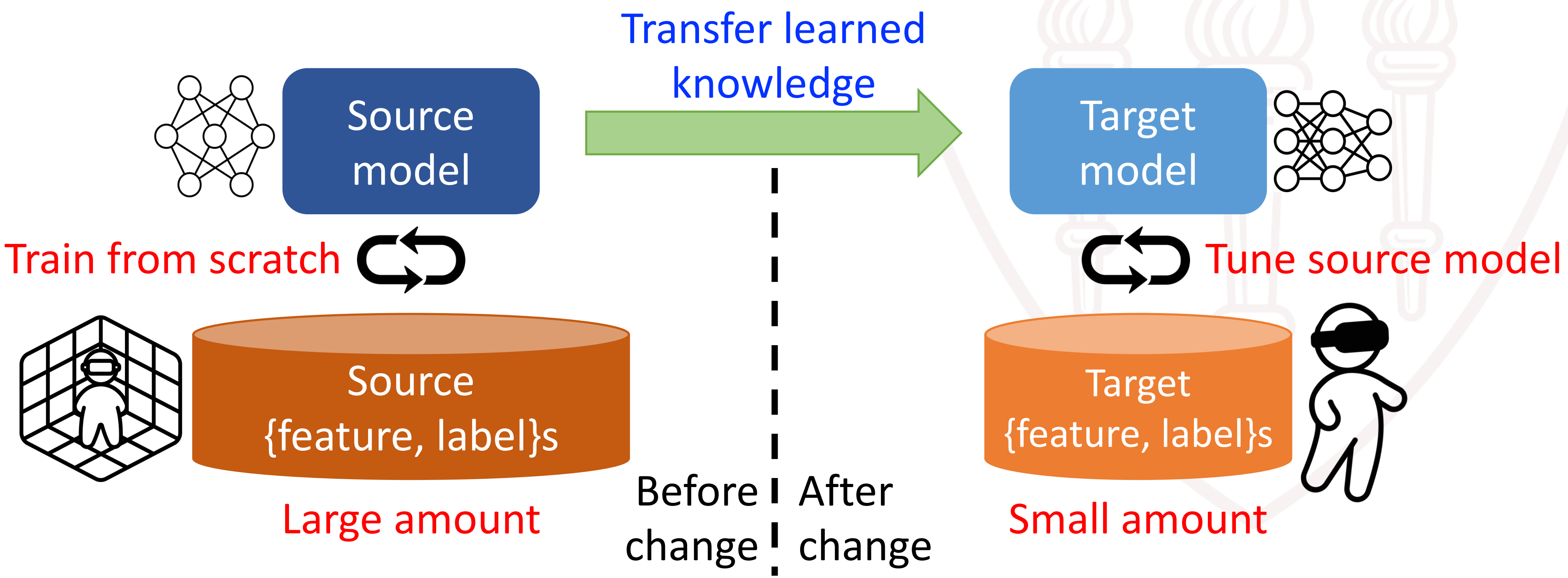
Small impact



USC

Handle Unseen Changes

- Our solution: Transfer Learning (TL)
  - Key assumption: there is invariant learned knowledge before & after a change
  - Benefit: adapt model to the change much faster than training a new model from scratch



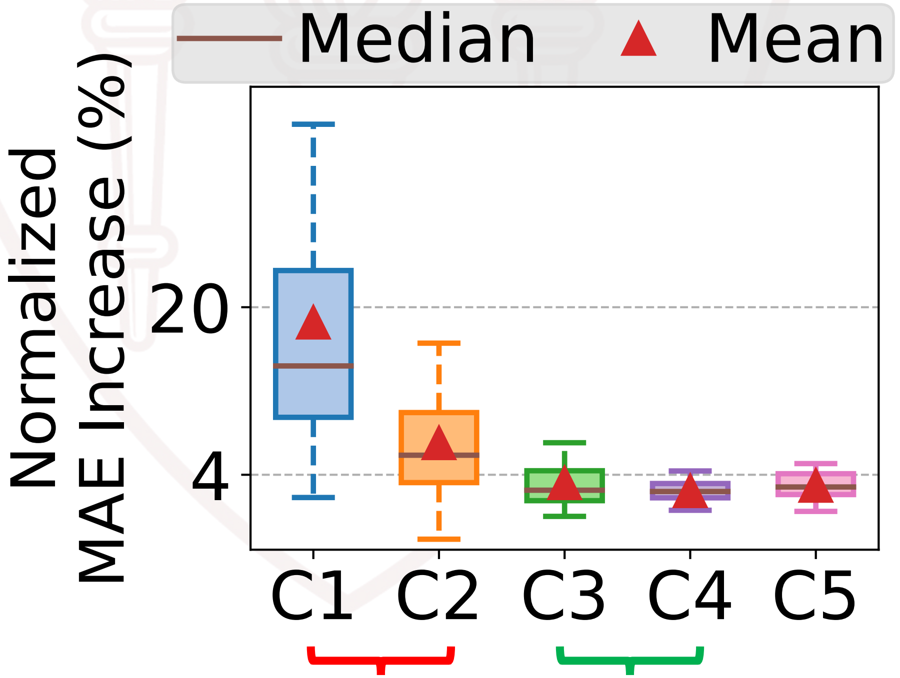
Handle Unseen Changes

- Our solution: Transfer Learning (TL)
  - C1: location change
  - C2: user change
  - C3: motion pattern change
  - C4: static environmental change
  - C5: dynamic environmental change
- More details are in the paper

Offline TL for bootstrapping

Online TL

Vision-based NLoS detection

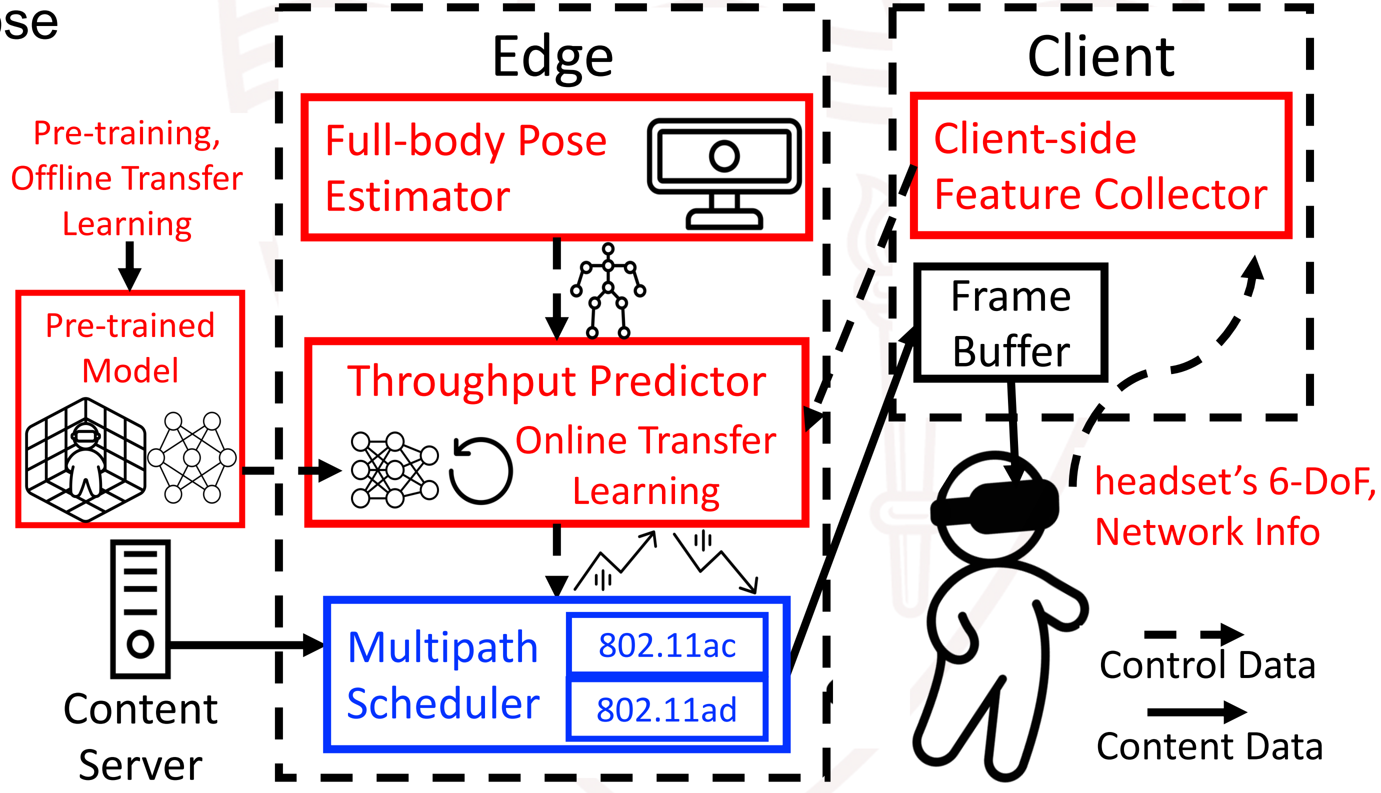


Multipath Scheduling

- Multipath: omnidirectional radio + mmWave
  - Prioritize omnidirectional radio
  - Opportunistically use mmWave
- Trend-aware scheduling
  - Conservatively or aggressively using mmWave
- See paper for details

Holistic View of Habitus

- mmWave throughput prediction
  - Enhanced by tracking full-body pose
  - React to unseen changes
    - Online/Offline transfer learning
    - NLoS detection
- Multipath networking
  - Omnidirectional radio + mmWave
  - Trend-aware scheduling



The system architecture of Habitus

USC

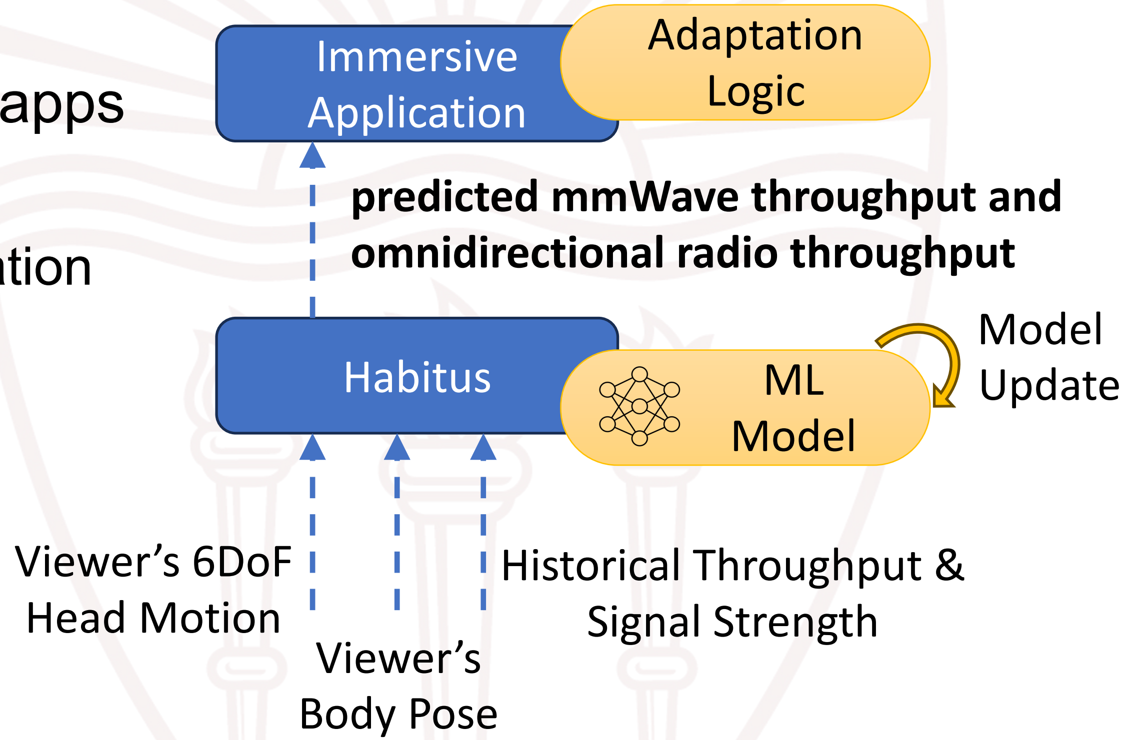
Habitus Prototype

360º Videos

VR Games

Volumetric Content

- Habitus is a general framework for immersive apps
- Implementation w/ commodity HW/SW
  - Challenges, e.g., accurate throughput estimation



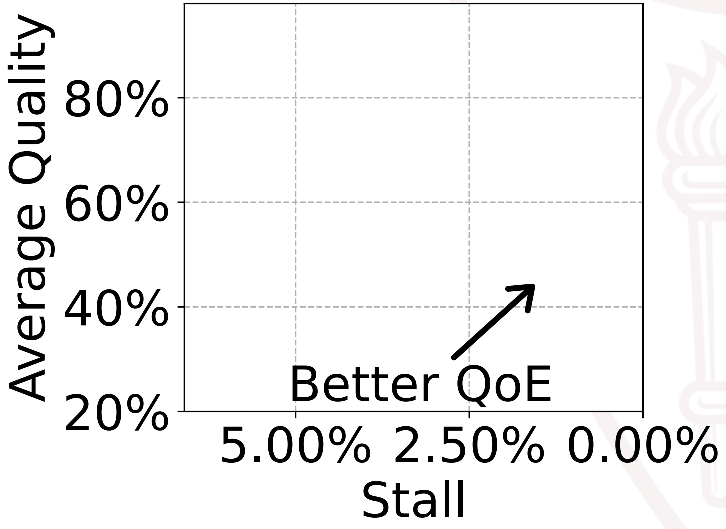
Application Interface

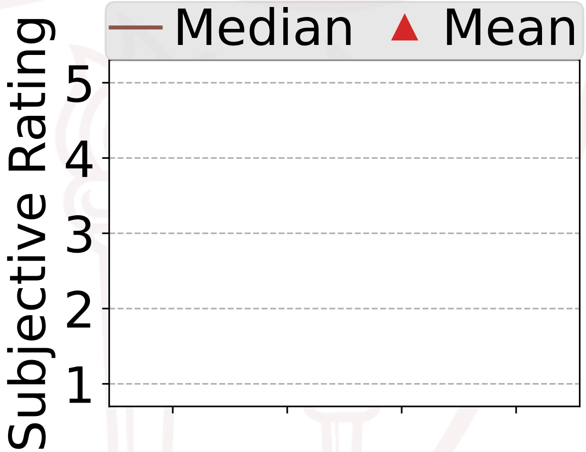
Integrate Habitus to ViVo [MobiCom'20]: only changing 47 LoC

Hardware: ROG Phone II, KCXGHYI VR Headset, Netgear Nighthawk X10 AP, ZED 2i Camera
Software: Linux iw, libpcap, Google ARCore, PyTorch & Torchscript, Zed OpenPose

Case Study: Volumetric Video Streaming

- Left: trace-driven emulation
- Right: user trial (N=12)

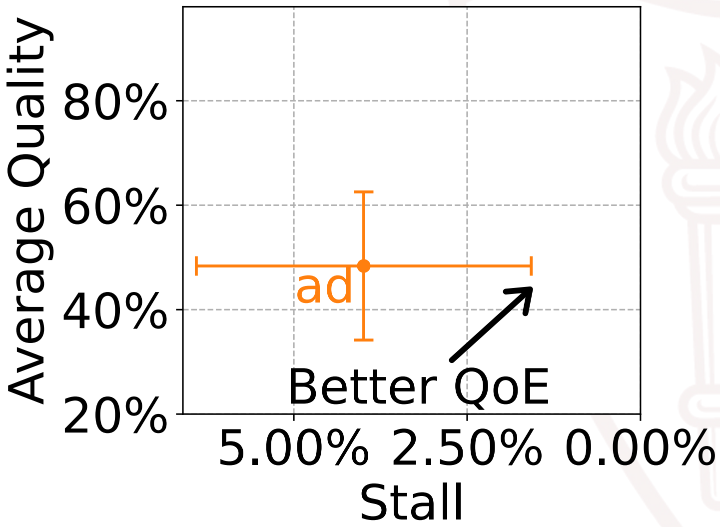


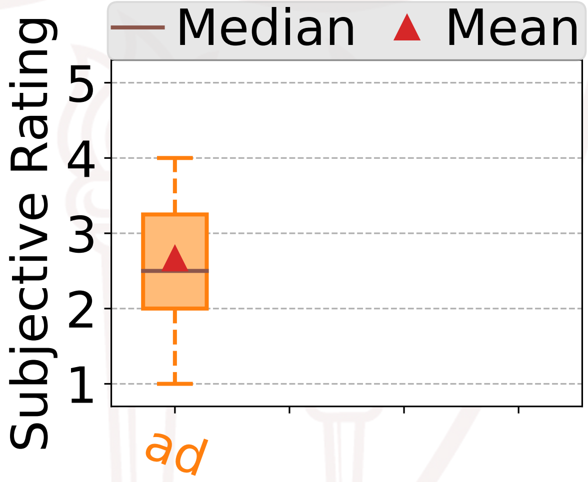


Case Study: Volumetric Video Streaming

- Left: trace-driven emulation
- Right: user trial (N=12)

- **ad**: only use 802.11ad (mmWave) w/o prediction

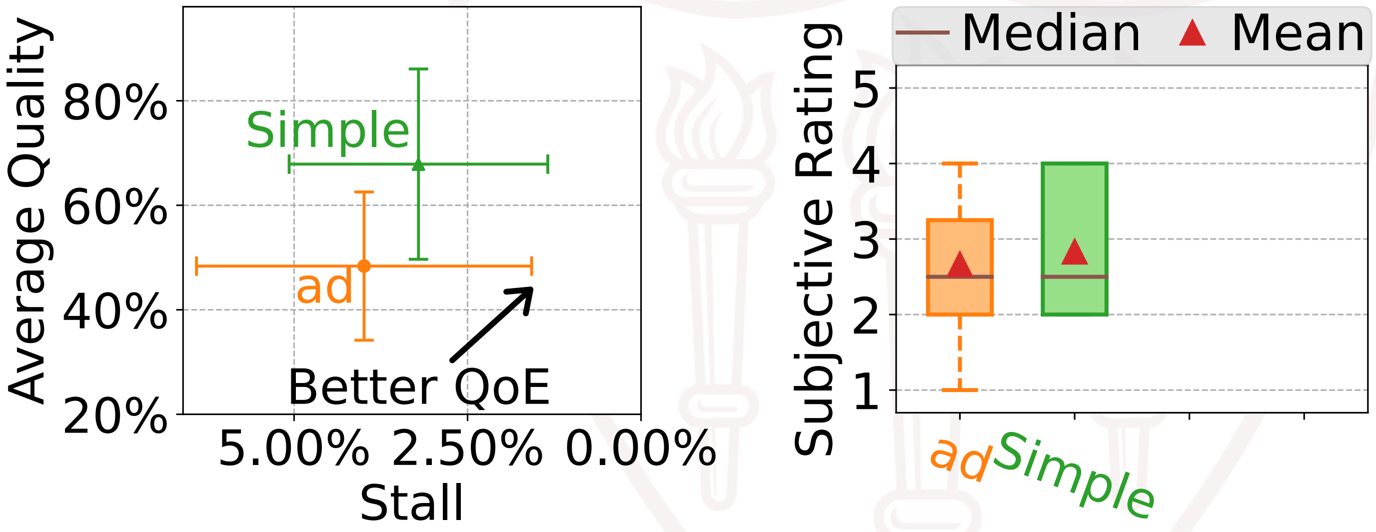




Case Study: Volumetric Video Streaming

- Left: trace-driven emulation
- Right: user trial (N=12)

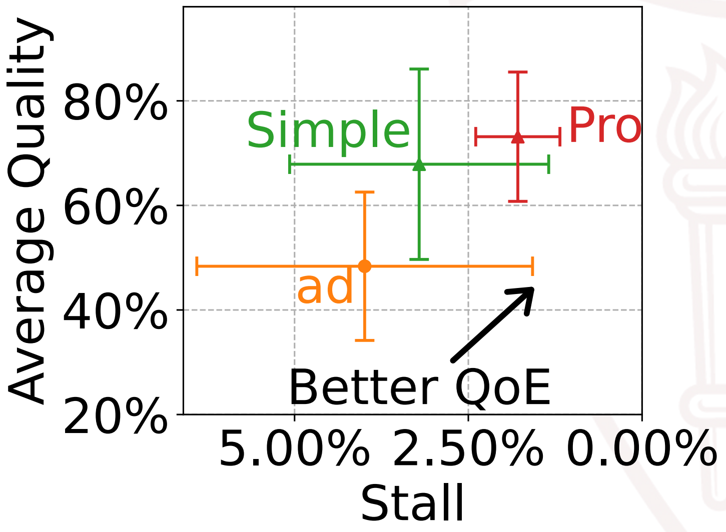
- **ad**: only use 802.11ad (mmWave) w/o prediction
- **Simple**: ac + ad w/o prediction

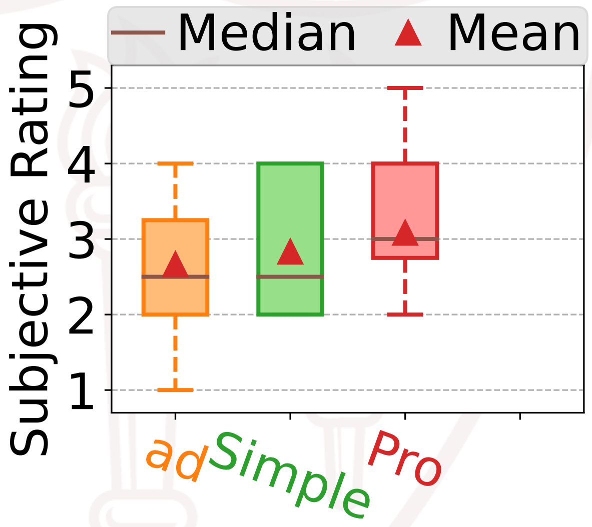


Case Study: Volumetric Video Streaming

- Left: trace-driven emulation
- Right: user trial (N=12)

- **ad**: only use 802.11ad (mmWave) w/o prediction
- **Simple**: ac + ad w/o prediction
- **Pro**: Habitus ac + ad w/ prediction (6DoF features only)

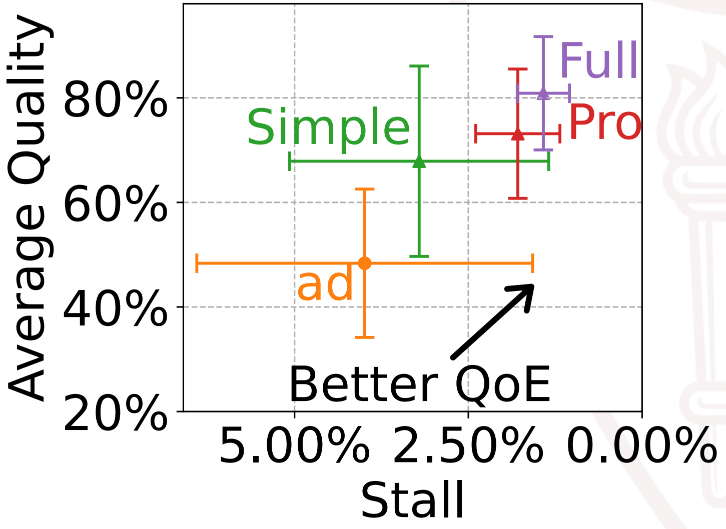


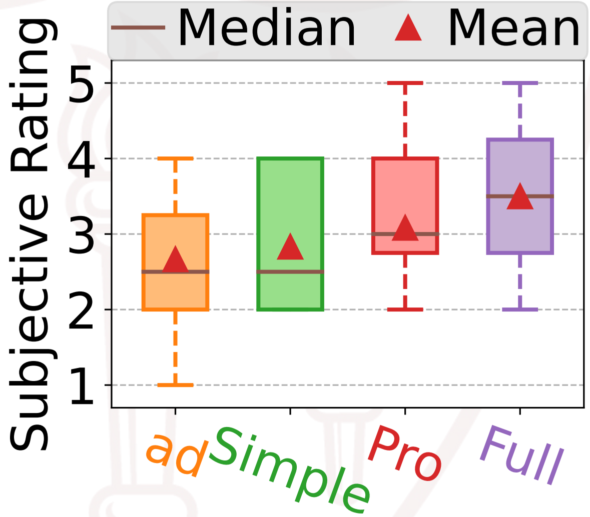


Case Study: Volumetric Video Streaming

- Left: trace-driven emulation
- Right: user trial (N=12)

- **ad**: only use 802.11ad (mmWave) w/o prediction
- **Simple**: ac + ad w/o prediction
- **Pro**: Habitus ac + ad w/ prediction (6DoF features only)
- **Full**: Habitus ac + ad w/ prediction (6DoF + full-body Pose features)





- Habitus (Pro, Full) considerably outperforms baseline approaches
- Using full-body pose (Full) further boosts the QoE
- Find more evaluation in our paper

USC

Summary

- Challenge of high-quality immersive content delivery over mmWave

- The design of Habitus
  - Multipath scheduling over omnidirectional radio and mmWave
  - Full-body pose guided mmWave throughput prediction
    - Handle unseen changes

- QoE improvement of Habitus demonstrated by trace-driven emulation & user trial
  - We release our dataset and the source code for data collection
    - ① 802.11ac/ad throughput & signal strength, ② 6DoF head motion, ③ full-body pose
  - See our paper for the links