EtherPose: Continuous Hand Pose Tracking with Wrist-Worn Antenna Impedance Characteristic Sensing

Daehwa Kim
Carnegie Mellon University
daehwak@andrew.cmu.edu

Chris Harrison
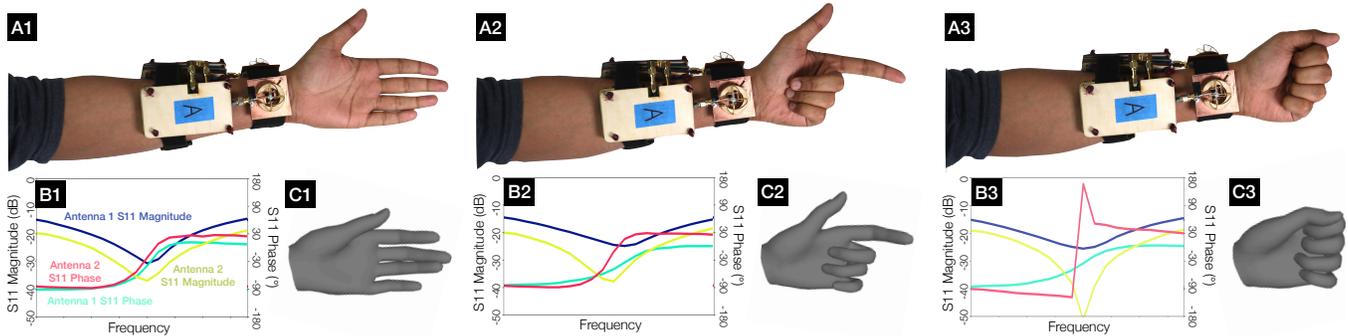Carnegie Mellon University
chris.harrison@cs.cmu.edu

Figure 1: EtherPose is a self-contained wrist band (A) measuring the swept frequency RF impedance of two antennas (B), which we use to pose a real-time 3D hand model (C).

ABSTRACT

EtherPose is a continuous hand pose tracking system employing two wrist-worn antennas, from which we measure the real-time dielectric loading resulting from different hand geometries (i.e., poses). Unlike worn camera-based methods, our RF approach is more robust to occlusion from clothing and avoids capturing potentially sensitive imagery. Through a series of simulations and empirical studies, we designed a proof-of-concept, worn implementation built around compact vector network analyzers. Sensor data is then interpreted by a machine learning backend, which outputs a fully-posed 3D hand. In a user study, we show how our system can track hand pose with a mean Euclidean joint error of 11.6 mm, even when covered in fabric. We also studied 2DOF wrist angle and micro-gesture tracking. In the future, our approach could be miniaturized and extended to include more and different types of antennas, operating at different self resonances.

CCS CONCEPTS

• **Human-centered computing → Interaction devices**.

KEYWORDS

Smartwatch, Input, Band, Wrist, Arm, Virtual and Augmented Reality, Natural User Interfaces

1 INTRODUCTION

Digitizing a user's hands for use in interactive systems has been a long standing research area, with seminal systems such as the Sayre Glove [12] and DataGlove [54] presented more than half a century ago. Since then, advanced in sensors and machine learning have allowed systems to be miniaturized and become less invasive to the user, with the ultimate aim to not encumber the hands at all. Uses of hand pose tracking are numerous, including domains as diverse as virtual and augmented reality [21], spatial user interfaces [33], sign language recognition [11, 42], and context awareness [31].

While hand pose sensing via external cameras and other remote sensors is possible, in this work we focus on worn systems that provide pervasive input capabilities. Today, the most capable worn hand pose systems in the literature use optical methods (e.g., RGB cameras, thermal cameras, range finders). While successful, they are sensitive to occlusion from clothing and the user's hand itself in certain poses. Secondarily, wrist-worn camera-based methods innately have privacy implications that can deter consumers. For this reason, researchers continue to explore new methods that can either stand alone or, in the future, contribute to multimodal sensing approaches.

To this literature, we contribute a new system called EtherPose (a homage to Etherphone, the original name of Leon Theremin's hand sensing musical instrument that broadly utilizes the same phenomena [50]). Instead of measuring proximity via capacitive coupling with an external antenna (as Theremin did), we use a small

worn antenna emitting a swept-frequency RF signal and measure the reflected signal's magnitude and phase shift (i.e., S11 parameter) with a compact, battery-powered vector network analyzer (VNA). As the user's hand changes geometry (i.e., to form different poses), the expanded antenna ground plane formed by the user tissue changes, therefore changing the antenna self-resonance and thus the impedance characteristic of the antenna observed at a predetermined frequency.

To inform the design of our final prototype, we conducted a series of simulation and empirical studies, which we detail in subsequent sections. The diameter of the wrist allowed us to include an additional, second antenna, which helps to capture other hand geometry changes. This process culminated in a proof-of-concept device, coupled with a machine learning backend, on which we ran user studies. Briefly here, for continuous hand pose tracking, we found a mean euclidean joint error of 11.6 mm across our nine participants. Inspired by recent work [17, 46], we also investigated wrist rotation estimation, finding a mean angular error of 5.87°.

The contributions of this paper are multifold. Foremost, EtherPose is the first demonstration of continuous hand pose tracking using antenna impedance characteristic. This signal is robust to varying clothing and lighting conditions, and is more privacy-preserving than comparable camera-based methods. Our iterative development approach is also uncommon, relying on tandem real-world experiments and computer simulations. These results informed the design of our untethered, battery-powered, and real-time EtherPose prototype. We then use this setup to evaluate three input modalities, whereas most prior systems explore a single modality.

2 RELATED WORK

Researchers have explored an array of methods to digitize users' hands, from multi-camera room installations [59] to worn gloves [12, 54]. More pertinent to this work are methods that are worn and mobile, and importantly, do not instrument the user's hands. This attention means that signals must be sensed from another body location, and in this work we focus on the wrist (and forearm), as this is a practical location where many users already wear a watch or jewelry. We note there are hand tracking systems that operate on the upper arm, shoulder, chest, or head (e.g., [22, 34]), but these are fundamentally different device form factors and utilize signals generally not present at the wrist, and thus are not immediately comparable. After reviewing hand-sensing wristbands, we more specifically discuss electrical and RF approaches applied to the problem of interactive hand input.

2.1 Hand Sensing Wristbands

The most popular approach for hand sensing near or on the hand are optical methods, as it generality offers high resolution data. While worn systems have employed depth (e.g., WatchSense [52]) and thermal cameras (e.g., Fingertrak [23], Pyro [18], Yamato et al. [67]), by far the most common camera variety used are those operating in the visible or near infrared light range. The latter systems include Digits [29], CyclopsRing [7], Hand with Sensing Sphere [2], Back-Hand-Pose [63], and Opisthenar [68]. Range-finding sensors (optical or acoustic) are also fairly common, and utilized in systems such as ThumbTrak [55], RotoWrist [46] and WristWhirl [17]. A

commonality in the above systems is the need for a sensor line of sight, which EtherPose does not require.

Non-optical approaches have also been explored, but generally offer more limited hand tracking functionality. Chief among these methods are electromyography (EMG) and electrical impedance (or capacitance) tomography methods, which we discuss briefly in the next section. Acoustic approaches include in-air ultrasound beamforming [24] and in-body interferometry [25]. Researchers have also looked at passive acoustic approaches, such as listening for touch events on the hand and vibrations induced from object use [32]. Finally, pressure or contour sensors (optical, capacitive or mechanical) have been used to sense deformations in wrist geometry (see e.g., WristFlex [13], Jung et al. [27], Fukui et al. [15], GestureWrist [43], Rudolph et al. [45]), from which certain hand poses can be recognized.

We note that among these systems, most demonstrate static hand "gesture" classification, with sets of around ten poses. More rare are systems demonstrating continuous hand pose tracking, which is more challenging. Systems that perform continuous hand pose tracking include Digits [29], Fingertrak [23], Back-Hand-Pose [63], WR-Hand [37], and NeuroPose [38]. Beyond hand pose, there are two other input modalities worth noting (which EtherPose implements and evaluates). First is wrist angle input, explored in systems such as WristWhirl [13] and RotoWrist [46]. Second is "micro-gesture" input, utilizing small and subtle movements of the fingers for interactive control, previously demonstrated in systems such as Soli [36], Pyro [18], AtaTouch [30], ElectroRing [28], and Serendipity [60].

2.2 Electrical & RF Hand Sensing Systems

There are several categories of electrical and radio frequency (RF) based hand sensing systems, which we review in order of increasing similarity to EtherPose.

Farthest from our work are electrical bio-sensing systems, such as electromyography (EMG) [37, 38], electrical impedance tomography [71, 73], and bio-capacitive sensing [57]. One step closer to EtherPose are radar-based hand tracking systems, such as Yu et al. [69], Paradiso et al. [40, 41], and Sluÿters et al. [48] (see [1] for an excellent survey of radar-based hand gesture methods). Of these radar systems, only Soli [36] has achieved sufficient miniaturization to be worn (the technology has shipped inside smartphones) and demonstrated fine-grained, micro-gestural input. Highly related to radar approaches are systems utilizing RF reflections from a user's body. This technique has been used with backscattered "wifi" signals for full body pose estimation (e.g., RF-Pose [75]), as well as with RFID tags worn on the body (e.g., RF-Wear [26]).

Next most related are worn hand input systems that inject RF signals inside the body. For instance, Touché [47] measured swept frequency impedance between two wrist bands for coarse two-hand gestures. SkinTrack [74] used an powered ring that injected 80MHz RF into the body, the phase of which could be captured using four electrodes operating on the skin-side of a smartwatch, enabling on-skin 2D touch tracking. ActiTouch [72] used a 10.5MHz RF signal injected into the wearers arm, and measures the inter-body and in-air radiated energy for on-skin touch detection (with finger position tracking done with computer vision). Essentially the

same approach is employed in ElectroRing [28], but with a finger-worn apparatus. Finally, there are systems that may be broadly categorized as in-air electric field sensing approaches, first explored for interactive use in seminal work by Smith et al. [49, 50]. More recently, eRing [62], and PeriSense [61] are ring-like devices that use several electrodes to capacitively sense around the finger for discrete hand pose classification. Cohn et al. [10] demonstrated a wrist worn electric field sensing implementation able to detect five user locomotion modes. AuraSense [76] used four receiver electrodes on the sides of a smartwatch to detect changes in the electric field caused by input from the user's other hand.

Finally, and most related to our work, are other electric field sensing systems, but which use a vector network analyser (VNA) to measure S-parameters of the holistic system (device, user body and environment). We provide a brief primer on the sensing principle in the next section. VNAs are generally large and expensive pieces of bench equipment. Nonetheless, they have been used in a tethered fashion (i.e., antennas worn on body, connected to a benchtop VNA) for sensing coarse human motion, such as arms swinging, rowing, sitting and hopping motions [35]. Xu et al. [64] also used a benchtop VNA, and demonstrated classification of four finger motions using single-frequency impedance over time. We note that we were inspired by this work's use of EM simulations, and included this method as part of our iterative design process. The latter work also proposed a compact folded cylindrical helix antenna design, which we included in our experiments. Finally, in the HCI literature, AtaTouch [30] used a worn VNA (but required wall power) and one dipole antenna to detect discrete pinching gestures. In contrast to the latter systems, EtherPose is comparatively small, self-contained, battery-powered, and demonstrates continuous, real-time 3D hand pose estimation.

2.3 Sensing Principle

EtherPose leverages "loading mode" electric field sensing, in which a radiating element is sufficiently proximate to a human-body that they capacitively couple (see Smith for an excellent primer on this subject [50]). In our system, the proximity of the radiating element to human tissue (e.g., wrist and hand) means the wearer becomes part of the radiated element ground plane. A ground plane with dimensions around 2 - 3 wavelengths can be approximated to a virtual electric infinite ground plane. In other words, adopting a frequency of operation which has a single wavelength approximately the same dimension as an average human hand, creates a condition where the radiated structure is coupled to a finite ground plane [39, 56]. Therefore, any change in hand pose (e.g., fingers moving closer or father from the antenna) manifests as a change in the antenna structure that, in turn, changes the self-resonance frequency and performance. Depending on the antenna topology, this coupling effect can be varied and enhanced. As illustrated in Figures 2 and 4, the hand and forearm are capacitively loaded to the antenna. As the hand pose changes, the distance between antenna and hand affects the capacitance between them (C_2). Additionally, the relative distances between the fingers varies the capacitance between each digit (C_1). This ground plane effect also modifies the mutual inductance (L_1) between the antenna and hand as the amount of carrying current in the opposite direction is varying according to
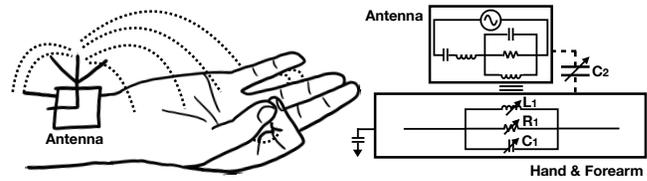

Figure 2: Simplified equivalent circuit model of a dipole in a lossy medium and a user's hand and forearm.

the ground plane's size, distance, and shape (e.g., loops when fingers pinch). Also, different gestures limit the extent of the hand covered by the electric field, changing the resistance (R_1). The S11 parameter (i.e., scattering parameter, RF transmitted from port 1, reflected RF measured at port 1) describes the ratio between returned signal and incident signal reflected by an impedance discontinuity in the medium. The impedance changes are characterized in the S11 parameter in a certain way [19] and can be measured by a VNA. The discrimination of hand poses is defined by the antenna complex impedance change at a predetermined frequency due the self-resonance shift and performance change caused by alterations in the coupled virtual ground (i.e., hand pose). These principles are well understood, allowing us to run simulations using commercial software, which we describe in Section 3.3.

3 BACKGROUND EXPERIMENTS

At the start of our development, several crucial design parameters were unknown: antenna topology, frequency range, and worn location. To inform our implementation, we ran a series of progressive experiments utilizing physical measurements and software simulation.

3.1 Test Apparatus

As a test platform, we used a NanoVNA V2 [58] connected to a MacBook Pro "13 (2020) over USB. We could programmatically control the VNA with serial commands from a Python application we wrote for experimentation. Different antennas can be attached to port 1 using a standard SubMiniature A (SMA) connector with a 50 Ω-impedance. The antenna feed line is connected to the signal pin of the SMA connector, while the ground-side line is soldered to the FR-4 copper clad laminate and the SMA's ground pin. An elastic velcro strap was used to affix the apparatus to users' arms. All antennas were mounted to a 6mm acrylic sheet with cutouts for the wrist band to loop through. For all experiments in this section, the VNA was configured to record S11 parameters, specifically the return loss magnitude and phase shift of a 50MHz sweep (51 data points each), centered at each antenna's resonant frequency.

3.2 Test Procedure

We selected three exemplary hand poses for all of our experiments: neutral, thumb-to-index pinch, and fist (seen in Figure 9). With our human participants, we had them perform the three hand poses five times each, during which three frames of data were recorded. We selected two evaluation metrics to gauge progress: (1) what is the magnitude and difference between different hand pose signals, and (2) how well does a machine learning classifier distinguish between

the three hand poses using such signal. In our Figures, we visualize the difference of signals from one hand pose stimuli to the other and then calculate the average of the difference. To capture ground truth 3D hand keypoints for evaluation, we use MediaPipe Hands [70] and a webcam operating 30 cm below a user's hands (which provides 21 hand 3D keypoints). As a first machine learning evaluation metric, we use SciPy's ExtraTreesRegressor (default parameters) to predict the relative 3D position of 21 hand keypoints given data from our EtherPose band (with results reported as mean per-joint 3D position error). We use the Mano library [44] to generate an animated, 3D hand mesh (Figures 9 and 10; see also Video Figure). As a second machine learning evaluation metric, we train and test a three-class ExtraTreesClassifier (default parameters) on the three discrete hand gestures noted above (with results reported as accuracy percentage or confusion matrices).

3.3 Simulation Software and Method

We also ran a simulation campaign designed to complement our real-world experiments. This process was done with 3D full electro-magnetic (EM) models in CST Microwave Studio [53], a commercial electromagnetic analysis suite. CST uses finite element method, finite integration technique and transmission line matrix method, and is suitable for electrically large or small, low or high Q radiated structures. The simulation models were designed based on the antenna topology, dimensions, hand pose, and material properties adopted in the background experiments. We use a commercially available EM phantom for the hand and arm with material proper-ties provided by SPEAG [51]. In short, EM phantoms are models of the human body that accurately reproduce the effect of the body on electromagnetic radiation.

While we endeavored to create simulations as faithful to real life as possible, there were several limitations. For instance, the antenna manufacturing process and construction varied slightly. There are also inevitable differences between the dielectric constant of complex human tissues vs. a simplified EM phantom. That being said, the largest discrepancy between simulation and measurements was with the fist hand pose. This dissimilarity was chiefly due to the fact our EM phantom had limited articulation and could not be equivalently posed. However, despite the relative compromise between measurements with human tissue and simulations with EM phantoms, we found there was strong correlation in antenna measurements, boosting confidence in our prototype designs and the theoretical principles that underpinned their operation.

3.4 Antenna Topology

The first and most fundamental design parameter we explored was antenna topology. This impacts not only antenna resonant frequency, but its radiation pattern, both of which have significant implications for coupling with a user's hand. While frequencies between 300-2450Mhz strongly couple to the human body [3, 20, 35], antennas that operate in this range are generally too large to be integrated into a worn device, adding another design constraint.

We identified four antennas topologies of interest (Figure 3): basic monopole, cloverleaf, pagoda [5] and folded cylindrical helix (FCH) [66]. All four types produce a toroidal radiation field in-plane with the arm, which envelops the volume where the hand operates.
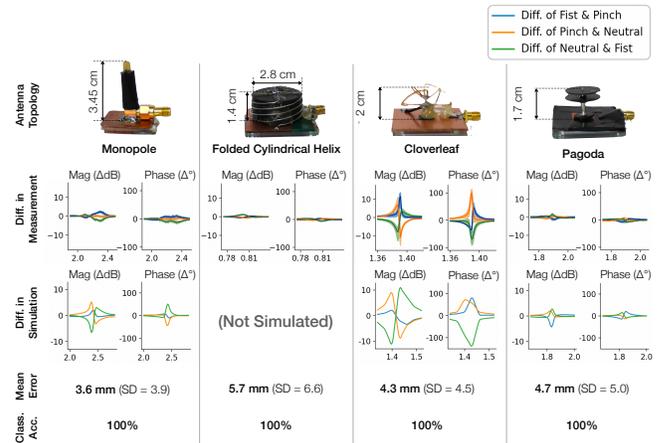

Figure 3: Results from measurements and simulations across four antenna topologies.

We purchased off-the-shelf 5GHz monopole, 5.8GHz cloverleaf, and 5.8GHz pagoda antennas, to which we added ground planes (made from FR-4 copper clad laminate) that shifted their resonant frequencies to 2.25, 1.38, and 1.80 GHz, respectively. We fabricated our own folded cylindrical helix antenna with a resonant frequency of 800MHz based on instructions in Xue et al. [66]

With these four antenna designs, we proceeded to run real world experiments to see how our three exemplary hand poses altered the antennas self-resonance, and therefore its characteristic impedance. We also ran matching software simulations for all but the folded cylindrical helix antenna. For all of these experiments – real and simulated – we held constant the antenna position: centered on the arm, just below the wrist crease, which we define as the "front" position (Figure 5).

Figure 3 provides an overview of these results. The second row shows all 15 trials (5 repeats × 3 frames) for each of the hand-pose pairs (subtracted from one another to highlight the difference) as performed by a real user. The darker colored single lines indicate the average of the lighter colored 15 lines rendered in the background. The third row in the Figure 3 is the difference of the simulated antenna S11 data across each pose (one curve for each pair). While there are differences between the simulated and real-world data, the main result is apparent. The average changes in S11 magnitude and phase shift were 0.52 dB and 3.60° for monopole, 0.25 dB and 1.85° for FCH, 1.93 dB and 12.11° for cloverleaf, 0.36 dB and 2.43° for pagoda. The max changes in S11 of the cloverleaf antenna were 11.55 dB and 99.55°, indicating a 3.8x S11 magnitude difference caused by hand pose stimuli. All antennas change their impedance characteristic in response to the different hand poses, though per-haps most dramatically with the cloverleaf antenna, with almost 50db radiated at its peak frequency.

To test how the measured signal impacted machine learning accuracy, we used our five rounds of real hand pose data to train and test (leave-one-round-out cross validation) a continuous hand model (see Test Procedure section above). We can compare our model's pose predictions against the MediaPipe-captured ground truth and compute mean per-joint position error (MPJPE), which
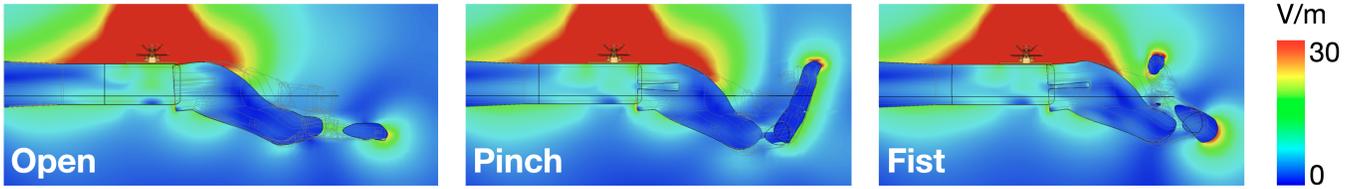
Figure 4: Simulation of electric field distribution for our cloverleaf antenna in the front position across three exemplary poses.
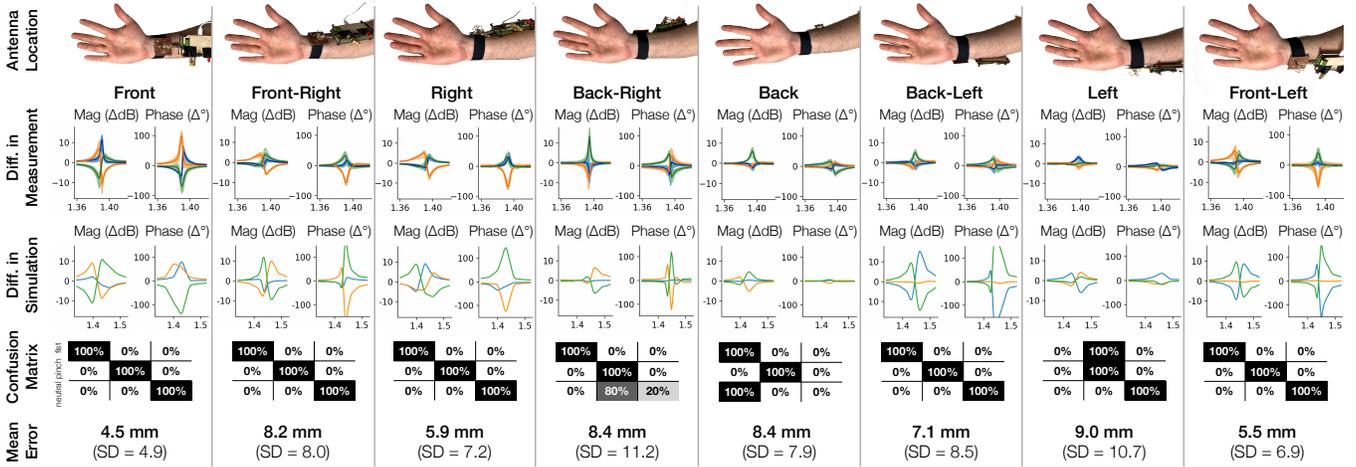

Figure 5: Results from measurements and simulations for our cloverleaf antenna placed at eight positions on the wrist. Color key provided in Figure 3; the blue, orange and green lines indicate the signal difference between fist and pinch, pinch and neutral, and neutral and fist poses, respectively.

we report in the "Mean Error" row of Figure 3. We also trained a classification model that simply predicts the three discrete poses. However, all four antennas had 100% classification accuracy, and so we do not provide confusion matrices in this particular figure ("Class. Acc." row in Figure 6).

All antenna designs showed promise, and were able to accurately predict pose, especially in the discrete pose classification task. The monopole antenna performed best in our machine learning evaluation, but its inherently tall profile was a significant detractor. Balancing accuracy and feasibility, we decided to move forward with our cloverleaf antenna, which performed second best in our machine learning results, demonstrated the most salient differences in its S11 data, and offered a compact geometry.

To better understand how the cloverleaf antenna's electric field was being altered by the three hand poses, we simulated and rendered the electric field distribution alongside the phantom. These sagittal cross-sectional electric field distributions at the antenna resonant frequency can be seen in Figure 4. We can see that in the open hand pose, the electric field distribution at the fingers is reduced, with an small concentration at the middle finger tip. In the pinch pose, the electric field is more evident along the length of the fingers, with an even higher concentration at the finger tips. Finally, in the fist pose, a different electric field distribution is observed with high concentration at the thumb knuckle and index finger tip. These electric field simulation results support the hypotheses that

changes to antenna impedance characteristics are resultant from the coupled ground plane's (i.e., forearm and hand) morphology changes. If the extended coupled ground plane created by the forearm and hand would be electrically infinite (i.e., larger than several wavelengths), changes in the electric field distribution would not be noticed or appear in the antenna characteristic impedance.

3.5 Antenna Location

With our antenna topology selected, we next moved to study the impact of body location on antenna signal. As before, we used a combination of software simulation and real-world measurements (using the same apparatus and procedure as above). Holding other parameters constant, we tested eight body placements: front, front-right, right, back-right, back, back-left, left, and front-left. Figure 5 shows the real-world and simulated S11 plots across the eight positions, along with classification and MPJPE. Coincidentally, the front position (which we utilized in our first experiment) performed best, with 100% classification accuracy for the three poses and the lowest joint error (4.5 mm). The S11 plots also showed the most expressivity in response to the three hand poses, which had the largest signal change of 1.93 dB and 12.11° on average. Both simulation results and measurements show a trend of decreasing signal difference as the antenna moves from the front to the rear for all hand-pose pairs. For these reasons, we decided to proceed with the front position in our design.

3.6 Secondary Antenna Location

After selecting the front position to host a cloverleaf antenna, it was apparent there was sufficient space on the backside of the wrist to host a second antenna. However, as the two antenna will interact with one another, it was not as straightforward as selecting the next-best-performing antenna from the previous study. To give more confidence and clarity in our selection, we ran a final set of simulations and real world measurements. This time, all conditions included a frontal cloverleaf antenna, and we tested a second cloverleaf antenna in five possible positions (left, back-left, back, back-right, and right). Front-left and front-right positions were not possible, as the antennas could not physically fit side-by-side on our wristband.

Figure 6 shows these results, which were created using the same data capture procedure and simulation method as above. Among the five positions, the combination of front & back-right performed best (followed closely behind by front & back). For the secondary antenna position, right and back-right induced distinctive phase shifts (right: 2.31 dB, back-right: 10.39°) and magnitude (right: 1.14 dB, back-right: 13.07°) on average compared to other positions (magnitude: 0.66 dB, phase: 4.25°). For the front antenna position, the signal with the back-right antenna (magnitude: 0.94 dB, phase: 5.33°) was slightly larger than the one with the right antenna (magnitude: 0.72 dB, phase: 3.99°). The machine learning accuracy was essentially unchanged from the prior experiment using only a single front antenna. However, we decided to move ahead with a two-antenna design, as we saw in pilot testing that it could prove useful in capturing some hand configurations in an expanded pose set.

4 IMPLEMENTATION

Informed by our background studies, we then built a proof-of-concept, self-contained implementation. We now describe the hardware and software that comprises this prototype.

4.1 Hardware

Our wristband (Figures 1 and 8) features two cloverleaf antennas with ground planes, located in the front and back-right positions, as identified in our background studies. The design and dimensions of these antennas can be seen in Figure 3, cloverleaf. Similar to our testing apparatus, the antennas are mounted to 6mm thick acrylic with cutouts that allow the elastic strap to loop through, permitting flexible antenna placement for a variety of wrist diameters. While we tried to make the two antennas identical, there were nonetheless small construction differences. The resonant frequencies of the front and back-right antennas on the wrist are 1.33 and 1.39 GHz, respectively, while the magnitude of S11 at resonant frequencies are -30 and -42 dB, respectively.

To increase physical robustness for our later user studies, we laser cut horseshoe-shaped wood shims to support the cloverleaf antennas. Each antenna is attached to its own dedicated NanoVNA v2 [58] with a rigid SMA connector. Both VNAs connect to a single Raspberry Pi Zero 2 W over USB, which also provides power. The Raspberry Pi runs our software (described in the next section), with machine learning results streamed over WiFi.

At maximum sensing rate and streaming data over wifi, our band consumes 4.5W. In other words, our bands' 16Wh LiPo battery provides around 3.5 hours of runtime. Our device cost around $250 to make, with the two VNAs ($98 each), Raspberry Pi Zero 2 W ($15), and battery ($6) dominating the bill of materials. We stress that our design is a proof-of-concept, not optimized for size, aesthetics or manufacturing cost. There are several avenues towards miniaturization. For instance, a RF multiplexer would allow for a single VNA to utilize two or more antennas, rather than having duplicate VNAs. Further, a more advanced VNA could utilize the two antennas to measure S12 and S21 parameters, which could provide new and useful data for pose estimation. If research on single-chip VNAs [8, 9, 14] comes to fruition, it would allow for dramatic miniaturization in the future.

4.2 Software & Featurization

On the Raspberry Pi, our software communicates with the two VNAs over USB serial. To initialize itself, each VNA is programmed to measure the return loss magnitude ±20Mhz centered at 1.38 GHz in 21 steps. The peaks (antenna resonant frequencies) are detected and each VNA re-centers itself on this value, most often a small shift, but one that is useful to maximize sensitivity. The VNAs are then configured to sense this frequency range continuously in a alternating fashion (to avoid interfering with one another), such that only one VNA is transmitting and measuring at a time. Each VNA measures return loss magnitude (21 data points) and phase shift (21 data points). As we have two VNAs, a single complete frame of data contains 84 total data points. The data capturing takes approximately 410ms, resulting in a frame rate of 2.4Hz.

Then, for each of the 4 sets of values (two return loss magnitude arrays and two phase shift arrays), we take the first derivative (20 features × 4), find the index of the peak (1 feature × 4), as well as the mean, min, max, and standard deviation (4 values × 4). This produces an additional 100 features.

Lastly, using each VNA's 21 magnitude values and 21 phase shift values, we compute the impedance at each frequency, resulting in 84 additional features (21 real and 21 imaginary components × 2 VNAs). On this, we similarly compute the first derivative (20 features × 4), mean and standard deviation (2 features × 4). This produces another 172 features, for a grand total of 356 (84+100+172) features.

4.3 Input Modalities & Machine Learning

Drawing inspiration from the literature, we selected three hand input modalities of interest. First and most general is 3D hand pose, previously demonstrated in systems such as Digits [29], Back-Hand-Pose [63], FingerTrak [23], but never with an RF approach. Systems such as WristWhirl [17] and RotoWrist [46] demonstrated 2DOF wrist angle input, which we selected as a second input modality. And finally, we were intrigued by fine-grained finger input (sometimes called "micro-gestures") seen in systems such as Soli [36], Yu et al. [69], Paradiso et al. [41], and Sluÿters et al. [48]. Each of these input modalities required a different model and training pipeline.

4.3.1 Continuous 3D Hand Pose. For this model, we use SciPy ExtraTreesRegressor (default parameters, 100 estimators) to predict the relative 3D position of 21 hand keypoints. As an input vector,
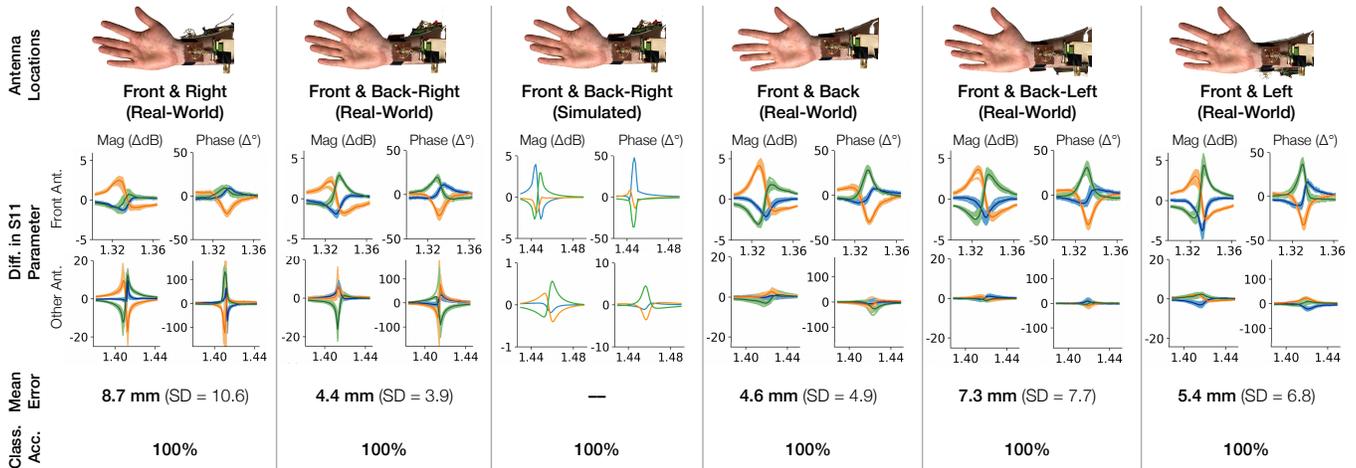
Figure 6: Results from measurements and simulations for our cloverleaf antenna placed in the front location, and a second cloverleaf antenna placed at one of five other positions on the wrist. Color key provided in Figure 3; the blue, orange and green lines indicate the signal difference between fist and pinch, pinch and neutral, and neutral and fist poses, respectively.
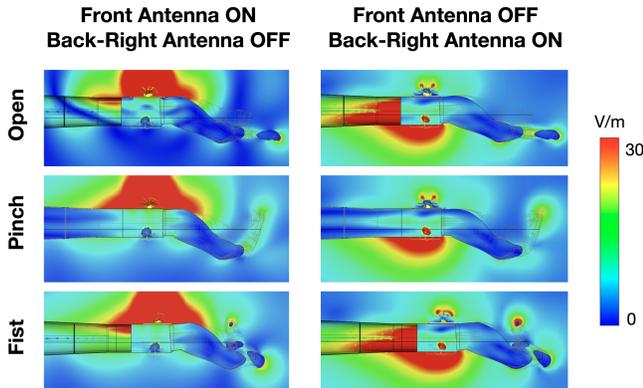

Figure 7: Simulated electric field intensity distribution at resonant frequency. One antenna is placed in the front position and the other is placed in the back-right position.

Figure 8: A labeled view of our proof-of-concept hardware laid flat. See Figure 1 for a photo of the device being worn.

we take the last three frames of featurized data. Same as the Test Procedure described in Section 3, we capture ground truth 3D hand keypoints using MediaPipe Hands [70] and a webcam operating 30cm below a user's hands. To produce an animated 3D hand mesh, we use the Mano Library [44] (seen in Figures 1 and 9, bottom row, as well as our Video Figure). Given joint coordinates from the machine learning, we used the inverse kinematics solver with Levenberg–Marquardt algorithm to obtain right-hand Mano model parameters for visualization.

4.3.2 Continuous 2DOF Wrist Angle. To capture training data for wrist angle, we use the same setup as 3D hand pose (MediaPipe Hands + webcam). As a proxy for 2DOF wrist angle, we compute the palm normal using MediaPipe's wrist, index_finger_mcp, and pinky_mcp keypoints. We use an ExtraTreesRegressor model (default parameters, 300 estimators) to predict the wrist left/right and ulnar/radial flexions (Figure 10, bottom row). As with our hand
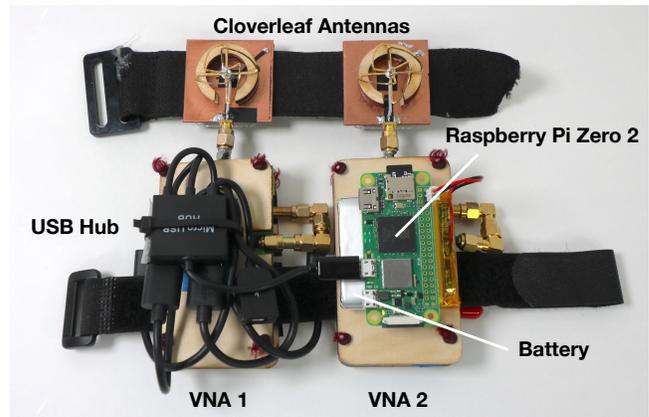
pose model, we use the most recent three frames of featured data from our band as the input vector.

4.3.3 Micro-Gestures. As one example of micro-gesture input, we track the thumb's position relative to the other four fingers, held together and acting like an trackpad. We trained our model (Extra-TreesRegessor, default parameters, 300 estimators) on discrete hand locations presented visually on a computer monitor. Once trained on this grid of data, our model can interpolate to provide continuous tracking. For instance, one can use their index finger as a slider control (Figure 11), clutched if desired. In our user study, described next, we selected a subset of horizontal and vertical positions.

5 USER STUDY
To evaluate the performance of our hand-sensing wristband and its three input modalities, we recruited nine participants (mean age
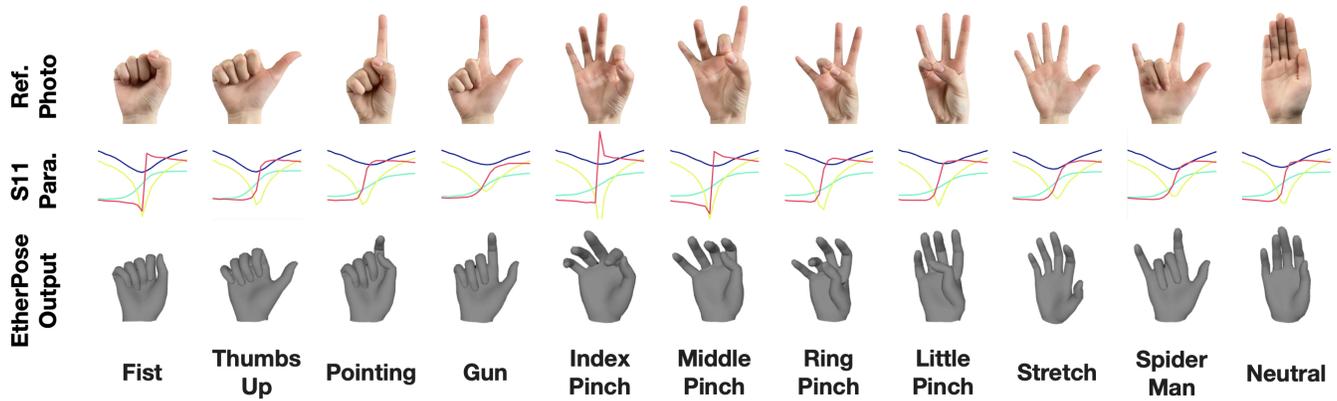
Figure 9: The eleven hand poses requested in our user study (top row). Importantly, these were only used as hand pose "destinations", with training and testing data collected continuously for all intermediate pose states, which are much more varied. The middle row shows an example S11 parameter sparkline captured in each pose, while the bottom row shows example hand pose output from our pipeline.

24.1) for a 90 minute study. We now describe our data collection procedure and results.

5.1 Procedure

After a brief introduction to the study, participants were fitted with our prototype wristband on their right wrist, continuing when they felt it was comfortable and secure. Our study was completed in two phases with two sessions in each. At the start of each session, the band completed its initialization process, centering on each antennas' resonant frequencies. During data collection, the user was seated in front of a computer monitor, which provided visual instructions.

For hand pose, we use eleven common poses drawn from prior work (Figure 9). To this set, we added four wrist flexions (i.e., angles;
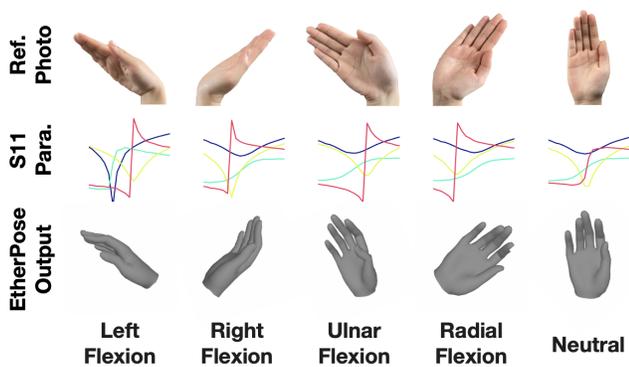

Figure 10: The five wrist flexions requested in our user study (top row). Note that these were only used as wrist angle "destinations", with training and testing data collected continuously for all intermediate pose states. The middle row shows an example S11 parameter sparkline captured in each pose, while the bottom row shows example hand pose output from our pipeline.

Figure 10), with the neutral pose acting as 0° left/right flexion and 0° ulnar/radial flexion. Rather than looking at discrete gesture classification, we follow the continuous hand pose evaluation procedure outlined in FingerTrak [23]. More specifically, users were prompted to slowly match their hand to the pose requested on the computer display. We continuously capture ground truth hand pose using MediaPipe and a webcam located to the side of the hand. When a new frame of data arrives from our prototype (at 2.4 FPS), the most recent MediaPipe hand keypoints are recorded along side the data. Deviating from FingerTrak's procedure, we do not request the user return to the neutral position between requested poses, as this allows us to capture more interesting and diverse intermediate pose states. A single round of data collection consisted of a random ordering of the fifteen hand poses. We collected eight rounds of hand pose data in this fashion, which formed one session.

After a small break, we then repeated the whole process, collecting another session of eight rounds of data, but with the armband covered with two layers of 100% cotton t-shirt fabric, simulating a sleeve of a medium weight garment. Completing all sixteen rounds (over two sessions) of hand pose data collection took approximately 60 minutes and yielded approximately 2000 hand pose instances per participant (19,170 instances across all 9 participants).

After a small break, participants proceeded to a microgesture study. In this procedure, the participants held their non-thumb fingers together in a flat manner, acting like a trackpad for the thumb. We found that neither a Leap Motion Controller or MediaPipe Hands was sufficiently accurate in this task to provide a ground truth. Instead, we collected discrete touch positions visually requested on a computer monitor, with the experimenter manually pressing a key to capture a trial at that instant in time. To account for variation in participant hand size, we use a normalized unit grid as labels for the requested touch positions.

One round of data collection consisted of touching the thumb to one of six positions on the grouped fingers arranged in a T-shape (one axis along the index finger, and the other axis running down the center of all fingers; see Figure 14 for an illustration).

Figure 11: In this micro-gesture example sequence, a user can use their thumb and index finger like a slider. Laptop included to illustrate the captured signal and slider state.

Five data frames were captured in each touch position, the order of which was randomly requested. One round consisted of all six positions, and one session consisted of eight rounds. As with our hand pose procedure, an identical second session was completed by participants, but with EtherPose covered in two layers of cotton fabric. Completing two sessions of eight rounds of microgesture data collection took approximately 30 minutes. All combined, this produced 8 rounds × 2 sessions × 6 positions × 5 data frames × 9 participants = 4320 trials.

5.2 Results: Effect of Cloth Covering

As noted in the above procedure, half of our study data was collected with the EtherPose band uncovered, while the other half was collected when the band was covered with two layers of 100% cotton t-shirt fabric, simulating a medium weight sleeve.

We first ran our hand pose, wrist angle and microgesture results separating these two conditions, expecting at least some performance degradation when the fabric covered the antennas. However, and encouragingly, we found no performance difference: Hand pose MPJPE when uncovered: 11.32 mm (SD=7.60) vs. covered: 11.66 mm (SD=7.80); Mean wrist angular error when uncovered: 5.60° (SD=1.30) vs. covered: 6.04° (SD=1.37); Micro-gesture mean 2D position error uncovered 12.5 mm (SD=5.1) vs. covered: 11.7 mm (SD=4.3). Thus we simply combine session data for the following results sections, as this is more indicative of real-world use.

5.3 Results: Continuous Hand Pose

As a user's hand is unique to them, we first trained per-participant hand pose models using 12 of a user's 16 rounds of data (drawn from the two sessions of eight rounds). For testing, we use all combinations of four sequential rounds (i.e., rounds 1/2/3/4, rounds 2/3/4/5, etc.), yielding 7 train/test combinations, the results of which are averaged together. Across our nine participants, we found a mean per-joint position error (MPJPE) of 11.57 mm (SD=7.57). These results are broken out by hand joint in Figure 12. This is comparable in performance to the 12.0 mm positional error achieved by Finger-Track [23], which employs four wrist-borne thermal cameras.

5.4 Results: Continuous Wrist Rotation

Following the same train/test procedure as our hand pose analysis, we found a mean wrist angular error of 5.87° (SE=0.06) across our nine participants. Broken out, we see a mean left/right flexion error

of 5.36° (SE=0.056) and a mean ulnar/radial flexion error of 6.37° (SE=0.065). Figure 13 provides a breakdown of error by participant, though there is no significant effect.

5.5 Results: Combined Model

Using the same train/test procedure, we also evaluated a combined hand pose & wrist angle prediction model (i.e., a single model outputs both hand pose and wrist angle given EtherPose data). We found a hand pose MPJPE of 11.58 mm (SD=7.57) and a wrist angle error of 6.16° (SE=0.063). This performance is almost identical to our prior results.

5.6 Results: Micro-Gestures

We also trained per-user models on 14 out of 16 rounds of a user's micro-gesture data, testing on the participants' 15 and 16th holdout rounds (i.e., 8-fold cross validation; results averaged). To account for different participant hand sizes, study data was collected on a unit grid, as noted above. To "reproject" this into real world units, which are more interpretable, we used an average human hand size with a unit scalar of 23 mm. Across our nine participants, we found an average 2D positional error (i.e., in the plane of the four fingers) of 12.1 mm (SD=9.9). The distribution of touch trials can be seen in Figure 14, along with 2σ error ellipses.
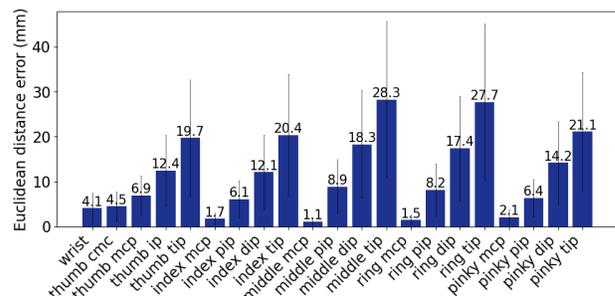

Figure 12: EtherPose's mean per-joint position error (MPJPE) for 21 hand keypoints. The error bar indicates standard deviation.
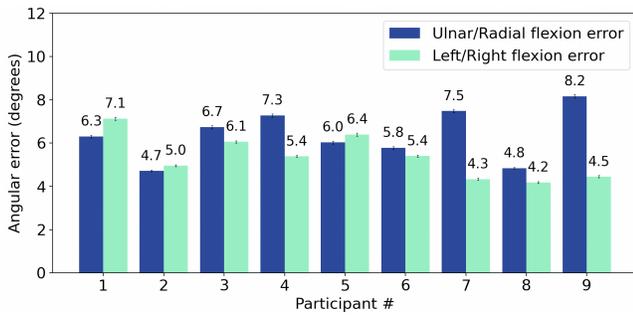
Figure 13: EtherPose's mean angular error in our wrist angle study broken out by participant, left/right, and ulnar/radial flexions. The error bar indicates standard error.

Figure 14: Plot of all microgesture trials collected in our user study, normalized by hand size. Error ellipses are 2σ.

6 LIMITATIONS & FUTURE WORK

While EtherPose shows promise as a technique, there are several significant drawbacks that we must highlight, but which also illuminate avenues for future work.

Perhaps the greatest weakness of our present system is the inability to work without calibration across worn sessions and users. This is because even small changes in the worn location, or hand shape/size can have a significant impact on our antennas' impedance characteristic. For the foreseeable future, we envision users having to complete at least a basic calibration when re-wearing the band. This process is not uncommon in worn hand pose research systems, and even commercial systems such as the Myo Armband [4] and Noitom Hi5 VR Glove (IMU) [16] require a per-worn-session calibration. Another potential way to achieve generalizability would be a big data approach, building a training corpus from many users, worn locations, body poses, environmental conditions, etc., which has proven successful in deep learning computer vision approaches. It may also be possible to generate synthetic S11 training data using simulations, not unlike those we utilized throughout our development process. Such an approach could also lessen the need for user-specific calibration data and a faster out-of-the box experience. Finally, EtherPose could prove useful in a multimodal sensing approach, used in conjunction with techniques like EMG or EIT, with each technique's individual strengths combining to enable superior cross-user/session robustness.

We also note that when the arms are operating in front of the user, such as in a VR experience, our band works well. However, when the arm gets too close to the user's body (or any conductive object, such as a steel door), the antennas begin to couple to the torso and their impedance characteristic changes. Less severe is metal jewelry, such as rings, as they are already part of the ground plane. Though the noise caused by these external components starts to occur when they enter the limited range of 2 wavelengths, it may limit user's movements, so we would note that this is an important and inherent limitation of our technique. The only potential solution is to employ more directional antennas emitting towards the hands, instead of radiating outwards, a topic we hope to explore in future work.

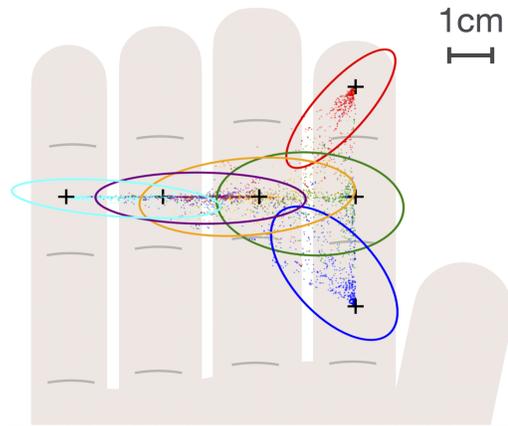In the same vein, we are also interested in experimenting with different antenna topologies (including mixed topologies on one band) and greater number of antennas. Of course, more compact antennas than our current cloverleaf design would also be preferable. We are particularly interested in exploring flexible PCB antennas, enabling truly thin form factors that could be integrated into the strap of a smartwatch (perhaps with an ASIC in the watch body interfacing with a totally passive band).

While our prototype exceeded our expectations in terms of accuracy, we note that when pose estimation fails, it tends to be catastrophically wrong (i.e., a totally incorrect pose), which is problematic for interactive use (i.e., users may be forgiving to small errors, but not big ones that interrupt tasks). It may be that improved machine learning models and more training data could mitigate this issue. In terms of framerate, our current prototype's 2.4 FPS is slow – good enough for a proof of concept, but not good enough for fluid interactive use. The system framerate would have to be improved in a commercial implementation.

In terms of a path to miniaturization, we must also acknowledge that our current band, built from two VNAs and a Raspberry Pi Zero 2, would require significant engineering efforts to miniaturize into a consumer product. The miniaturization would almost certainly require a custom RF chip to meet the strict space and power requirements of mobile devices. However, it can be done, as we have seen with the incredible advances in cellular modem chips and the emergence of single-chip VNAs [8, 9, 14]. The antennas would also require miniaturization, and there are several possible routes. The main goal in the current embodiment was not to optimize far-field antenna figure of merit, such as total efficiency, antenna aperture or bandwidth. Instead, the design goal was to identify an antenna topology that produces higher variation in self-resonance frequency associated with coupled finite ground planes. Thus, some additional antenna candidates can be considered, such as the cavity-Backed dipole, planar Inverted L (PILA), and aperture-fed microstrip antenna, which can be as small as a few millimeters. Also, some helical antennas (spherical, folded cylindrical, and disk-loaded) [6, 65] have enough radiation with a narrow bandwidth, and some optimizations are possible to decrease size (e.g., number of helix turns, wire radius, dielectric loading).

7 CONCLUSION

We have presented our work on EtherPose, a self-contained armband featuring two compact cloverleaf antennas. The impedance characteristics of these antennas change as a user alters their hand pose. We capture this effect by measuring the S11 parameter (i.e., frequency-dependent reflected power) of two strategically-placed antennas. In our user study, we show that our system can provide real-time hand pose estimation with a mean per-joint position error (MPJPE) of 11.57 mm. We also studied 2DOF wrist rotation, which had a mean angular error of 5.87°, and 2D micro-gesture tracking, which was accurate to within 12.1mm on average.

ACKNOWLEDGMENTS

REFERENCES

[1] Shahzad Ahmed, Karam Dad Kallu, Sarfaraz Ahmed, and Sung Ho Cho. 2021. Hand gestures recognition using radar sensors for human-computer-interaction: A review. *Remote Sensing* 13, 3 (2021), 527.
[2] Riku Arakawa, Azumi Maekawa, Zendai Kashino, and Masahiko Inami. 2020. Hand with Sensing Sphere: Body-Centered Spatial Interactions with a Hand-Worn Spherical Camera. In *Symposium on Spatial User Interaction*. 1–10.
[3] Md Taslim Arefin, Mohammad Hanif Ali, and AKM Fazlul Haque. 2017. Wireless body area network: An overview and various applications. *Journal of Computer and Communications* 5, 7 (2017), 53–64.
[4] Myo EMG armband. 2022. https://developerblog.myo.com/
[5] Maarten Baert. 2022. Pagoda Antenna. https://www.maartenbaert.be/quadcopters/antennas/pagoda-antenna/
[6] Steven R Best. 2004. The radiation properties of electrically small folded spherical helix antennas. *IEEE Transactions on antennas and propagation* 52, 4 (2004), 953–960.
[7] Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. 2015. Cyclopsring: Enabling whole-hand and context-aware interactions through a fisheye ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 549–556.
[8] Hyunchul Chung, Qian Ma, Mustafa Sayginer, and Gabriel M. Rebeiz. 2017. A 0.01–26 GHz single-chip SiGe reflectometer for two-port vector network analyzers. In *2017 IEEE MTT-S International Microwave Symposium (IMS)*. 1259–1261. https://doi.org/10.1109/MWSYM.2017.8058835
[9] Hyunchul Chung, Qian Ma, Mustafa Sayginer, and Gabriel M Rebeiz. 2020. A Packaged 0.01–26-GHz single-chip SiGe reflectometer for two-port vector network analyzers. *IEEE Transactions on Microwave Theory and Techniques* 68, 5 (2020), 1794–1808.
[10] Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R Smith, Matthew S Reynolds, Desney S Tan, and Shwetak N Patel. 2012. An ultra-low-power human body motion sensor using static electric field sensing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 99–102.
[11] Helen Cooper, Brian Holt, and Richard Bowden. 2011. Sign language recognition. In *Visual analysis of humans*. Springer, 539–562.
[12] T DeFanti and DJ Sandin. 1977. Sayre Glove Final Project Report. *US NEA R60-34-163 Final Project Report* (1977).
[13] Artem Dementyev and Joseph A Paradiso. 2014. WristFlex: low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 161–166.
[14] Analog Device. 2021. *10 MHz to 20 GHz, Integrated Vector Network Analyzer Front-End*. https://www.analog.com/media/en/technical-documentation/data-sheets/adl5960.pdf
[15] Rui Fukui, Masahiko Watanabe, Tomoaki Gyota, Masamichi Shimosaka, and Tomomasa Sato. 2011. Hand shape classification with a wrist contour sensor: development of a prototype device. In *Proceedings of the 13th international conference on Ubiquitous computing*. 311–314.

[16] Hi5 VR Glove. 2022. https://www.noitom.com/hi5-vr-glove
[17] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. Wristwhirl: One-handed continuous smartwatch input using wrist gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 861–872.
[18] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-tip gesture recognition using pyroelectric infrared sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 553–563.
[19] Anitha Govind. 2022. Antenna Impedance Matching–Simplified. *Abracon LLC* (2022), 1–6.
[20] Peter S Hall and Yang Hao. 2006. Antennas and propagation for body centric communications. In *2006 First European Conference on Antennas and Propagation*. IEEE, 1–7.
[21] Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, et al. 2020. MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 87–1.
[22] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 453–462.
[23] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D hand pose tracking by deep learning hand silhouettes captured by miniature thermal cameras on wrist. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–24.
[24] Yasha Iravantchi, Mayank Goel, and Chris Harrison. 2019. BeamBand: Hand gesture sensing with ultrasonic beamforming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–10.
[25] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. Interferi: Gesture sensing using on-body acoustic interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[26] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I Hong. 2018. Towards wearable everyday body-frame tracking using passive RFIDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–23.
[27] Pyeong-Gook Jung, Gukchan Lim, Seonghyok Kim, and Kyoungchul Kong. 2015. A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors. *IEEE Transactions on Industrial Informatics* 11, 2 (2015), 485–494.
[28] Wolf Kienzle, Eric Whitmire, Chris Rittaler, and Hrvoje Benko. 2021. Electroring: Subtle pinch and touch detection with a ring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
[29] David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 167–176.
[30] Daehwa Kim, Keunwoo Park, and Geehyuk Lee. 2021. AtaTouch: Robust Finger Pinch Detection for a VR Controller Using RF Return Loss. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.
[31] Gierad Laput and Chris Harrison. 2019. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[32] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 321–333.
[33] Jinha Lee, Alex Olwal, Hiroshi Ishii, and Cati Boulanger. 2013. SpaceTop: integrating 2D and spatial 3D interactions in a see-through desktop environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 189–192.
[34] Cheng Li and Kris M Kitani. 2013. Pixel-level hand detection in ego-centric videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3570–3577.
[35] Yang Li and Youngwook Kim. 2016. Classification of human activities using variation in impedance of single on-body antenna. *IEEE Antennas and Wireless Propagation Letters* 16 (2016), 541–544.
[36] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–19.
[37] Yang Liu, Chengdong Lin, and Zhenjiang Li. 2021. WR-Hand: Wearable Armband Can Track User's Hand. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–27.
[38] Yilin Liu, Shijia Zhang, and Mahanth Gowda. 2021. NeuroPose: 3D Hand Pose Tracking using EMG Wearables. In *Proceedings of the Web Conference 2021*. 1471–1482.
[39] Salma Mirhadi, Mohammad Soleimani, and Ali Abdolali. 2012. Analysis of finite ground plane effects on antenna performance using discrete Green's function. In *2012 15 International Symposium on Antenna Technology and Applied Electromagnetics*. 1–3. https://doi.org/10.1109/ANTEM.2012.6262332

[40] Joseph Paradiso, Craig Abler, Kai-yuh Hsiao, and Matthew Reynolds. 1997. The magic carpet: physical sensing for immersive environments. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*. 277–278.

[41] Joe Paradiso, Nick Yu, and Che King Leo. 2022. Gesture-Sensing Radars Project. https://resenv.media.mit.edu/Radar/index.html

[42] Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. 2014. Sign language recognition using convolutional neural networks. In *European Conference on Computer Vision*. Springer, 572–578.

[43] Jun Rekimoto. 2001. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Proceedings Fifth International Symposium on Wearable Computers*. IEEE, 21–27.

[44] Javier Romero, Dimitrios Tzionas, and Michael J. Black. 2017. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 36, 6 (Nov. 2017).

[45] Julius Cosmo Romeo Rudolph, David Holman, Bruno De Araujo, Ricardo Jota, Daniel Wigdor, and Valkyrie Savage. 2022. Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–14.

[46] Farshid Salemi Parizi, Wolf Kienzle, Eric Whitmire, Aakar Gupta, and Hrvoje Benko. 2021. RotoWrist: Continuous Infrared Wrist Angle Tracking using a Wristband. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. 1–11.

[47] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 483–492.

[48] Arthur Sluÿters, Sébastien Lambot, and Jean Vanderdonckt. 2022. Hand Gesture Recognition for an Off-the-Shelf Radar by Electromagnetic Modeling and Inversion. In *27th International Conference on Intelligent User Interfaces*. 506–522.

[49] Joshua Smith, Tom White, Christopher Dodge, Joseph Paradiso, Neil Gershenfeld, and David Allport. 1998. Electric field sensing for graphical interfaces. *IEEE Computer Graphics and Applications* 18, 3 (1998), 54–60.

[50] Joshua Reynolds Smith. 1999. *Electric field imaging*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[51] SPEAG. 2022. https://speag.swiss

[52] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor. (2017), 12.

[53] CST Studio. 2022. https://www.3ds.com/products-services/simulia/products/cst-studio-suite/

[54] David J Sturman and David Zeltzer. 1994. A survey of glove-based input. *IEEE Computer graphics and Applications* 14, 1 (1994), 30–39.

[55] Wei Sun, Franklin Mingzhe Li, Congshu Huang, Zhenyu Lei, Benjamin Steeper, Songyun Tao, Feng Tian, and Cheng Zhang. 2021. ThumbTrak: Recognizing Micro-finger Poses Using a Ring with Proximity Sensing. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. ACM, Toulouse & Virtual France, 1–9. https://doi.org/10.1145/3447526.3472060

[56] D. Tayli and M. Gustafsson. 2016. Physical Bounds for Antennas Above a Ground Plane. *IEEE Antennas and Wireless Propagation Letters* 15 (2016), 1281–1284. https://doi.org/10.1109/LAWP.2015.2504795

[57] Hoang Truong, Shuo Zhang, Ufuk Muncuk, Phuc Nguyen, Nam Bui, Anh Nguyen, Qin Lv, Kaushik Chowdhury, Thang Dinh, and Tam Vu. 2018. Capband: Battery-free successive capacitance sensing wristband for hand gesture recognition. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. 54–67.

[58] NanoVNA V2. 2022. https://nanorfe.com

[59] Vicon. 2022. https://www.vicon.com/

[60] Hongyi Wen, Julian Ramos Rojas, and Anind K. Dey. 2016. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3847–3851. https://doi.org/10.1145/2858036.2858466

[61] Mathias Wilhelm, Daniel Krakowczyk, and Sahin Albayrak. 2020. PeriSense: ring-based multi-finger gesture interaction utilizing capacitive proximity sensing. *Sensors* 20, 14 (2020), 3990.

[62] Mathias Wilhelm, Daniel Krakowczyk, Frank Trollmann, and Sahin Albayrak. 2015. eRing: multiple finger gesture recognition with one ring using an electric field. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*. 1–6.

[63] Erwin Wu, Ye Yuan, Hui-Shyong Yeo, Aaron Quigley, Hideki Koike, and Kris M Kitani. 2020. Back-hand-pose: 3d hand pose estimation for a wrist-worn camera via dorsum deformation network. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 1147–1160.

[64] Bin Xu, Yang Li, and Youngwook Kim. 2017. Classification of finger movements based on reflection coefficient variations of a body-worn electrically small antenna. *IEEE Antennas and Wireless Propagation Letters* 16 (2017), 1812–1815.

[65] Dong Xue, Brian Garner, and Yang Li. 2016. Electrically-small folded cylindrical helix antenna for wireless body area networks. In *2016 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 1–4.

[66] Dong Xue, Brian A Garner, and Yang Li. 2017. On-body radiation of 3D-printed fold cylindrical helix (FCH) wearable antenna. In *2017 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 1–4.

[67] Yuki Yamato, Yutaro Suzuki, Kodai Sekimori, Buntarou Shizuki, and Shin Takahashi. 2020. Hand Gesture Interaction with a Low-Resolution Infrared Image Sensor on an Inner Wrist. (2020), 5.

[68] Hui-Shyong Yeo, Erwin Wu, Juyoung Lee, Aaron Quigley, and Hideki Koike. 2019. Opisthenar: Hand poses and finger tapping recognition by observing back of hand using embedded wrist camera. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 963–971.

[69] Myoungseok Yu, Narae Kim, Yunho Jung, and Seongjoo Lee. 2020. A frame detection method for real-time hand gesture recognition systems using CW-radar. *Sensors* 20, 8 (2020), 2321.

[70] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).

[71] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 167–173.

[72] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust touch detection for on-skin AR/VR interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 1151–1159.

[73] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing hand gesture recognition with high resolution electrical impedance tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 843–850.

[74] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. Skintrack: Using the body as an electrical waveguide for continuous finger tracking on the skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1491–1503.

[75] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. 2018. Through-wall human pose estimation using radio signals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7356–7365.

[76] Junhan Zhou, Yang Zhang, Gierad Laput, and Chris Harrison. 2016. AuraSense: enabling expressive around-smartwatch interactions with electric field sensing. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 81–86.