3D Gesture Recognition through RF Sensing

Alejandro Alanis†, Trang Thai∗, Gerald Dejean‡, Ran Gilad-Backrach‡, Dimitrios Lymberopoulos‡
†Tecnologico de Monterrey   ∗GE Global Research Center   ‡Microsoft Research
Monterey, Mexico   Schenectady, NY   Redmond, WA
eden.alanis@itesm.mx   trang.thai@ge.com   {dejean,rang,dlymper}@microsoft.com

ABSTRACT

Human interaction with devices is constrained to the surface of these devices through widely used touch sensors. In this work, we enable touchless interfaces that allow humans to interact with devices from a distance. Our approach is based on the design of a two-dimensional array of RF sensors specifically designed to detect the proximity of human body. Each sensor in the array acts as a near-field RF proximity sensor. When parts of the human body come to close proximity to the sensor, they slightly disturb its frequency response, allowing the detection of human fingers or hands. Since our approach is RF-based, it presents several distinct advantages over current sensing technologies which include the ability to work without line of sight, the ability to be easily embedded behind any type of surface, and the ability to scale to almost any size; all while operating at a similar power domain to current proximity sensing technologies. Using a prototype implementation and data collected through a user study, we demonstrate that the RF-array can detect the position and distance of a human hand located at a distance of up to 2 inches with higher than 75% accuracy.

1. INTRODUCTION

Several decades ago, computers were mainly large desktop devices with which human interaction took place through dedicated peripherals such as a keyboard and a mouse. With the rise of mobile computing, phones and tablets became the primary computing devices making touch sensors the new standard in human-computer interaction. However, as computing devices evolve and co-exist in multiple form factors, ranging from tiny wearable devices, such as watches and wearable sensors, to huge displays like TVs, the limitations of touch sensor as the primary input medium start to emerge. In the case of small screens, like the ones found on smartwatches, touching the screen is inherently difficult as the user's fingers can completely cover the screen, constantly obstructing the visibility of the actual content. On the other hand, interacting with touch sensors on large displays becomes cumbersome as the user needs to physically move his arm (i.e., large desktop screens) or even walk across the device (i.e., large displays like the ones provided by Perceptive Pixel) to interact with the displayed content.

To address these limitations, the research community has been actively exploring touch-less human-computer interaction techniques. To be able to support the wide range of computing form factors, though, any touch-less human-computer interaction technique must satisfy four basic requirements. First, it should be able to scale along with the size of the device to be incorporated. A viable approach should be easily incorporated in a small wearable device like a watch, and at the same time appropriately scale to orders of magnitude higher sizes to enable interactions with larger devices such as TV displays. Second, the power consumption of the touch-less interface circuitry should be kept minimal to enable long-term operation of battery operated devices such as watches and phones. Third, the distance range at which the interface operates should be able to appropriately scale with the form factor of the device. For instance, for a watch-like device interactions within a few centimeters are appropriate, but for larger devices such as computer monitors and TVs, this distance can be in the order of meters. Fourth, the touch-less interface circuitry should be easily embedded in the device (i.e., stacked behind the display panel) without requiring precious real-estate that is usually dedicated to large displays.

Kinect [14] and LeapMotion [15] are the most popular commercially available touch-less user interfaces today. The former uses an advanced depth sensing camera that can accurately capture the skeleton of the user and enable whole-body gestures. Even though very accurate, it suffers from size and power limitations that prevent easy embedding in smaller, battery operated devices such as watches and phones. LeapMotion, on the other hand, leverages low power infrared sensors to accurately determine the distance of your fingers from the device, and properly translate your finger motion into gestures. Even though accurate and low power, such a solution is inherently difficult to embed into smaller devices because of the line-of-sight requirement. Similarly to other optical solutions[3], the infrared sensors and the supporting circuitry cannot be embedded behind the display or within the device, as they need to directly observe the user's fingers/hands.

At the physical layer, there has been a lot of research on generating touch-less interactions based on capacitive sen-

sors [25]. In general, capacitive technologies require a user to touch an electrode in order to be sensed. However, there have been a few instances where a user does not have to be in contact with the electrode for sensing to take place [1], but this vertical sensing range is extremely small (well below $1cm$). In addition, with this approach only a single scalar capacitive change is recorded. Even though a single scalar capacitive change can be accurate enough to identify a finger tap on a screen, it does not provide enough resolution for accurately detecting more complex 3D gestures.

Recently, researchers at Disney Research proposed a new technology called Touche [20], that instead of recording a single scalar capacitive change, it analyzes capacitive change across a small frequency band. This higher resolution information can be used to detect how an electrode is contacted, and possibly be used for complex gesture recognition. Capacitive-based approaches to touchless interaction have two major limitations. First, in order to detect the presence of a human finger from a distance, the size of the electrode must be relatively large (approximately two orders of magnitude larger that the surface area of a human finger). Second, capacitive based sensors do not scale well to large screens due to the increased resistance over long electrodes.

In this work, we present a pure RF-based approach to real-time, touchless gesture recognition [17]. Our approach consists of a 2-dimensional array of RF sensors specifically designed to detect the proximity of human body (i.e., finger, hand, etc.). Each sensor in the array acts as a near-field RF-based proximity sensor, that has been designed to resonate at the 6-8GHz frequency range. When parts of the human body come to close proximity with the sensor they slightly disturb its frequency response. By monitoring variations in the 6-8GHz frequency response of the sensor, the detection of human fingers or hands is feasible. By combining multiple sensors into a 2-D array, we create a high resolution RF-based proximity sensing solution that can track the fingers and hands of a user over space and time, and therefore recognize gestures.

Smaller or larger form factors can be easily achieved by simply adding or removing RF sensors from the 2-dimensional array. In addition, the effective range and sensitivity of the proximity sensing simply depends on the size of the sensor used, and thus it can be easily adjusted. Since the proposed solution is RF-based, the sensor array can be easily embedded behind a display or any other surface while still enabling real-time gesture recognition. Every surface or object added between the user's hand and the sensor array manifests as a constant shift in the frequency response of each sensor in the array, and therefore can be easily taken into account during calibration.

Recognizing gestures in real-time on top of the two dimensional array of RF sensor poses two major challenges. First, simultaneously driving an array of sensors with such high frequency signals becomes a bottleneck. Simply driving the input and sensing the frequency response of all the sensors can require high computation and power resources. We address this problem by intelligently multiplexing the sensors at a high rate. In that way, only one sensor is active at any given time, significantly reducing the power and computation requirements. In addition, we demonstrate how we can efficiently generate 6-8GHz signals at a low cost and low power, by simply leveraging an MSP430 micro-controller, an RC filter, and a voltage controlled oscillator. When combined, these two techniques allow us to drive and sense the frequency response of a 2-dimensional high frequency sensor array at a power overhead that is comparable to that of a simple IR proximity sensor.

Second, reading the frequency response of a sensor at this wide frequency band (6-8GHz) requires significant amount of time. For each sensor element, multiple frequencies need to be set and then the response at those frequencies needs to be recorded. Given that the frequency response of all sensors in the array must be computed, this process can easily take several seconds, preventing any real-time gesture recognition. To address this problem, we perform feature selection analysis, and show that only a small set of frequency responses can achieve recognition rates similar the highest possible gesture recognition rate.

Using a prototype implementation of the 2-dimensional sensor array[1], and data collected across 10 users, we first show that we can accurately detect the position and distance of the user's hand above the sensor array. By properly selecting the parts of the frequency response that are the most important to distance and position classification, we show that real-time gesture recognition can be performed with accuracy higher than 75%.

The rest of the paper is organized as follows. Section 2 presents the basic principle of the single sensor cell, and Section 3 describes the 2 dimensional sensor array design and its supporting circuitry. Section 5 presents the experimental results of a real user study for classifying four different gestures types in real time. Section 6 provides an overview of the related work, and Section 7 concludes the paper.

2. RF SENSOR CELL DESIGN

Building a sensor cell for sensing the presence of human fingers requires the use of a cell that is robust to wear and tear, cheap to manufacture, low in profile, and can be made small for integration into mobile devices. In addition, interaction between a human finger and the sensor cell must be unobstructed by dielectric materials in between, small in vertical sensing range, simple to scan the response, and have a high enough resolution to distinguish uniqueness.

To meet all the requirements of sensing the proximity of a human finger, a topology consisting of a two-port, half-wavelength coupled bandpass filter with a resonator patch loaded on top was selected. Figure 1 shows a model of the sensor cell. On the bottom of the sensor lies a ground plane.

---

[1]A video showing the sensor array in action can be seen in `http://sdrv.ms/J3LHGh`
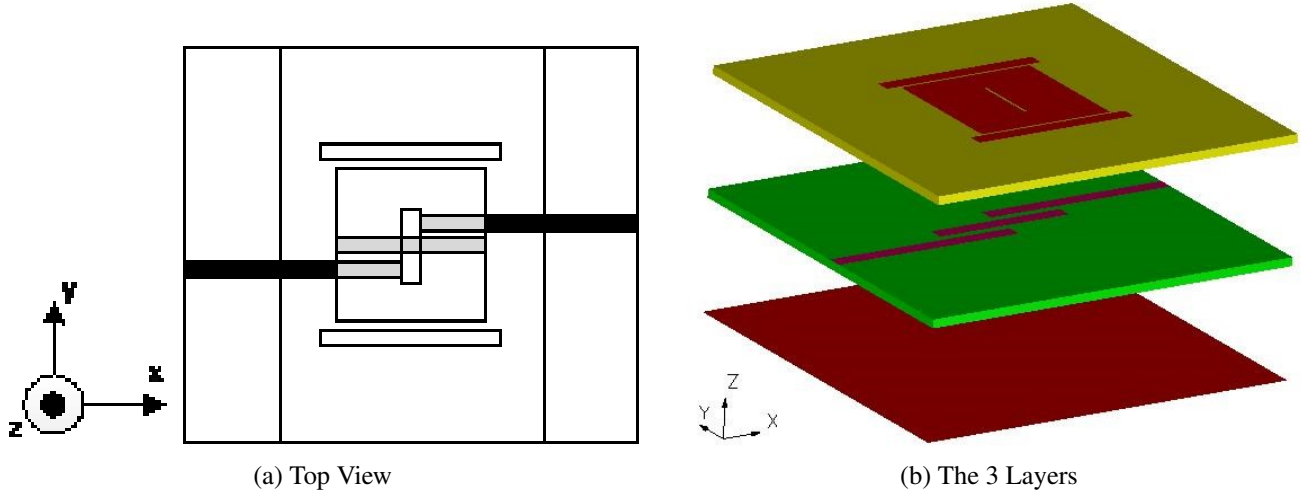
(a) Top View

(b) The 3 Layers

Figure 1: A schematic view of the single sensor cell. The bottom layer is the ground plane. The two-port, half-wavelength coupled bandpass filter is the middle layer, and the resonator path sits on top. The size of the resonator patch of the cell is 15 mm x 15 mm, and the total thickness is 1.6 mm.

The filter is in the middle, and the resonator patch is on the top. The frequency response of the bandpass filter is between 6-8 GHz.

This sensor works by having the user's finger interact with the resonator patch of the cell, while the spectral response is recorded through a bandpass filter. The operational principle of the single sensor lies on couple-line filters, a detailed explanation of which can be found in [7]. At a high level, the sensor cell in Figure 1 works as follows. Each of the two ports acts as an input and output respectively. At the input port (right most line in the middle layer in Figure 1) a sine wave signal with a frequency in the range of 6-8GHz is applied. The application of the high frequency signal excites current in the input port, that leads to capacitive coupling between the input port and the middle line. This in turn leads to capacitive coupling between the middle line and the output port (left most line in the middle layer in Figure 1). This coupling of energy allows signals at frequencies within a certain band to be transmitted, while other outside of the given band to be suppressed. The length of the middle coupling line is a half-wavelength long, and in turn, determines the frequency of operation for the filter. By placing a resonator patch above the filter, energy from the filter couples to the resonator patch. This has two effects. First, the patch creates a second bandpass response around 7.8 GHz, and secondly, the patch radiates a small zone of electromagnetic fields above its surface. This zone of electromagnetic energy establishes the vertical sensing range for the sensor cell. Therefore, placing a human finger above the cell alters the frequency response, thus, creating a unique spectral signature of vertical placement of the human finger above the cell.

In practice, a sine wave at the frequency range of 6-8GHz is applied to one of the ports of the sensor cell. Due to the
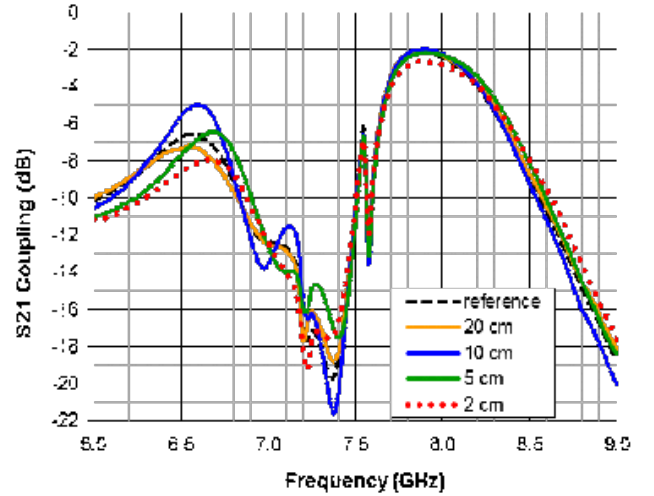

Figure 2: Transmission response for a human finger placed 2, 5, 10, and 20 mm above a single sensor cell as well as the case when no finger is present (denoted as reference).

capacitive coupling between the 3 lines in the middle layer shown in Figure 1, a frequency response can be automatically recorded at the other port of the single sensor cell. When a human finger comes in close proximity to the sensor cell, the electromagnetic field radiated from the resonator patch on the top layer in Figure 1 is disturbed resulting into a change to the frequency response at the output port. By exciting the input port with sine waves of different frequencies within the 6-8GHz range, multiple transmission responses can be recorded to better characterize the distance, position, and size of the finger placed close to the sensor cell.

To better illustrate the proposed sensor's operation, Figure 2 shows the transmission response of the sensor across frequencies when a finger is placed 2, 5, 10, and 20 mm above the sensor cell, as well as in the case where no finger is present (denoted as reference). It is clear that each placement of the finger creates a unique frequency response. It is extremely important to note that the purpose of this research is to design a proximity sensor to sense different positions of a human finger in close proximity to the sensor. Using a small band of frequencies, as opposed to a single frequency, provides more detailed characterization of the finger placement, enabling better distinction across multiple distances. For example, the frequency response at $d = 5$mm and 20mm are the same at 6.6 GHz, but if we analyze a band of frequencies around 6.6 GHz, it is evident that the responses at $d = 5$mm and 20mm are clearly unique to each other. This feature is essential in high resolution sensing, and the key to enabling more complex 3D gestures.

2.1 Design Considerations

This sensor design offers several benefits. It is a circuit that uses only passive distributed printed components, so the cell itself is not a source of power consumption. It is very cheap to manufacture, low in profile, and provides a small but practical vertical sensing range. Another reason for choosing this sensor cell topology is due to the properties of bandpass filters. Bandpass filters allow a higher amount of energy to be transmitted through from one port to another; thus, the change in received power versus transmitted power can be more easily detected.

Ideally, the frequency of operation of the sensor cell should be low. This would enable low-cost, low power circuity to easily interface with the input and output ports of the sensor cell. Unfortunately, the frequency of operation of the sensor cell is fundamentally limited by the size of the surface area of the object being measured (in this case, the human finger). If a sensor cell is larger than the finger's surface area, the finger does not significantly affect the interrogation zone of the cell, and thus it cannot be reliably detected. As a result, the size of the sensor cell has to be similar to the surface area of an average human finger.

However, the operating frequency of the cell is limited by its size. As shown in Figure 1, the length of the middle line in the middle layer of the sensor determines the operating frequency of the bandpass filter. The longer this line is, the lower its operating frequency. Given that the size of the cell needs to be smaller than the surface area of a typical human finger, the operating frequency of the sensor cell is constrained to the 6-8GHz range.

3. RF-BASED 3D GESTURE SENSING

The single sensor cell described in the previous section was designed to sense the vertical proximity of a human finger as it approaches the interrogation zone of the cell. Note that the cell itself only offers a 1-dimension interaction with
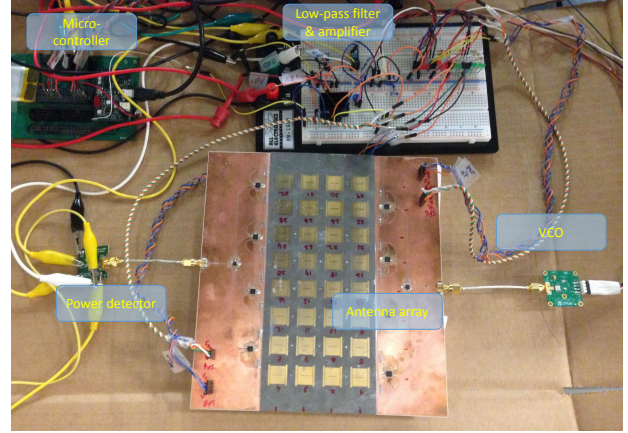

Figure 4: The prototype implementation of the two-dimensional RF array with the supporting circuitry.

a human finger. Whole hand gesture recognition, however, requires the ability to sense the whole hand and its position in the 3D space. To enable this type of 3D interaction, we leverage the sensor cell as a building block to construct a two-dimensional array. In particular, we combine 32 sensor cells into a single 4×8 matrix as shown in Figure 3. The size of the RF array is slightly larger than that of a typical human hand. As the human hand hovers over the board, each sensor cell senses a different part of the hand. As the hand moves to the left and right, the interrogation zone of different subsets of sensor cells is disturbed, allowing us to track the hand in the two dimensional space. In addition, as the hand moves along the z axis, the distance from the sensors changes, resulting in a change in the frequency response of these sensors (Figure 2). By monitoring these changes in each cell's frequency response, the RF matrix can track the hand in three dimensions.

However, combining 32 sensor cells into a single two-dimensional array to form a high resolution RF-sensor poses several challenges. First, each of the sensors needs to be properly excited with multiple frequencies in the range of 6-8Ghz, and the corresponding frequency response of the sensor must be recorded. Exciting the sensor cells with such high frequency signals at a low power and low cost is challenging. Second, to enable real-time gesture recognition, all 32 sensors should be excited separately and all 32 frequency responses must be read faster than the user performs the gesture.

To address these challenges, we have designed a system-level architecture around the RF array that is comprised of 4 distinct components: a low power micro-controller, a sensor cell driver, a switching network, and a power detector.

The low power micro-controller, a TI MSP430 CPU[12], was chosen as it has a low power envelope while supporting the types of IO needed, and providing the necessary com-
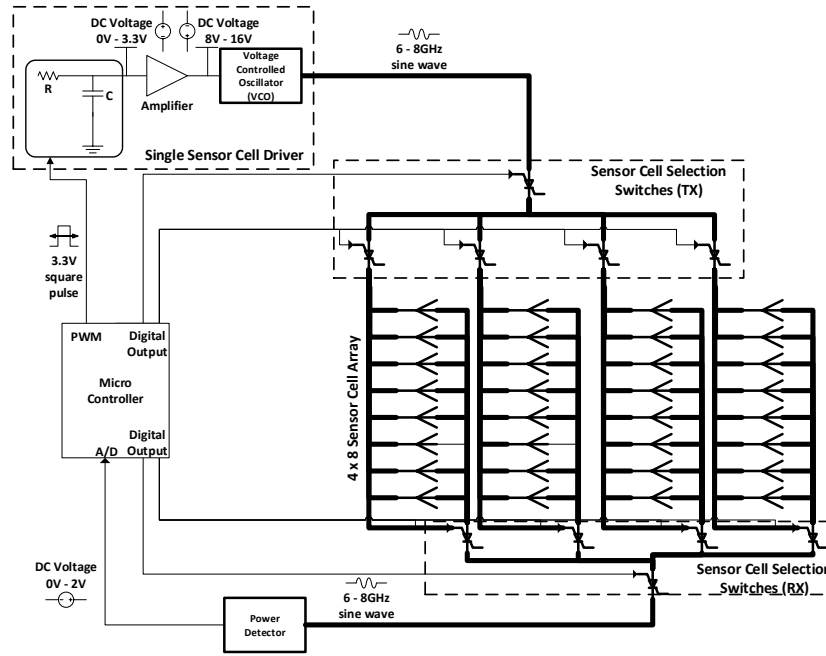
Figure 3: A schematic diagram of the setup configuration. The bolded components and connectors represent the high frequency (RF) components of the circuit. There is a DC coupling capacitor on every connection which is not presented here. The unbolded components and connectors are the low frequency and control components.

puting power. The micro-controller is able to configure a two-layer network of switches to select an individual sensor cell from the array. In that way, at any given time, only a single sensor cell is active. Even though more time consuming, this approach is necessary to constraint power consumption, as simultaneously driving all 32 sensors with high frequency signals comes with a high power overhead.

Therefore, the micro-controller selects a a sensor cell and configures the sensor cell driver to generate a high frequency signal in the 6-8GHz range and use it to drive the selected sensor cell. The sensor cell driver contains the necessary circuitry to enable a low power microprocessor running at 16MHz to generate 6-8Ghz sine waveforms. When the high frequency signal at the input of the sensor cell has been activated, a power detector is used at the output port of the sensor cell to record the frequency response of the selected sensor cell at the specified frequency. The power detector [22] converts the frequency response of the sensor cell into a DC voltage[2] that is directly sampled by the embedded A/D converter on the micro-controller. This process is repeated for multiple frequencies spanning the 6-8Ghz range, and for each of the 32 sensor cells in the RF array.

Once a complete scan of the RF array has been completed,

the hosting device uses an inference algorithm to compute the position of the hand, if any. By tracking the position and distance of the hand from the board continuously, the RF array enables real-time 3D gesture sensing.

Figure 5 shows the the data acquired from a complete scan of the sensor matrix under 3 conditions. The red curve shows the measurements when there is no object in the vicinity of the sensor matrix, the blue shows the data when a hand is present over the upper part of the board at a distance of 0.5-inches. The green curve shows the data when a hand is covering the lower part of the board at a distance of 0.5-inches. While the signal follows similar patterns in all conditions, some deviations are apparent and allow recognition of the different gestures. To better demonstrate this, Figure 6(a) shows the response of a single cell under the same conditions. This plot is a zoom in on the response of the second from the left cell on the lowest row in Figure 5. The measurements from this cell are very similar when the board is clear and when a hand is covering its upper parts. However, when a hand is covering the lower parts of the board, the measurements deviate in a way that allows recognition. Similarly, Figure 6(b) is a zoom in on the upper right cell of the matrix. In this figure we see larger deviation when the hand covers the upper part of the board.

By closely observing Figure 5, two imperfections of the prototype board we built become apparent. First, note that the frequency response of the different RF-cells is very dif-

---

[2]As an alternative, it is possible to mix the signal from the sensor with the signal from the VCO to get a DC level signal without using a power detector. We chose to use the power detector to simplify the experimentation process.
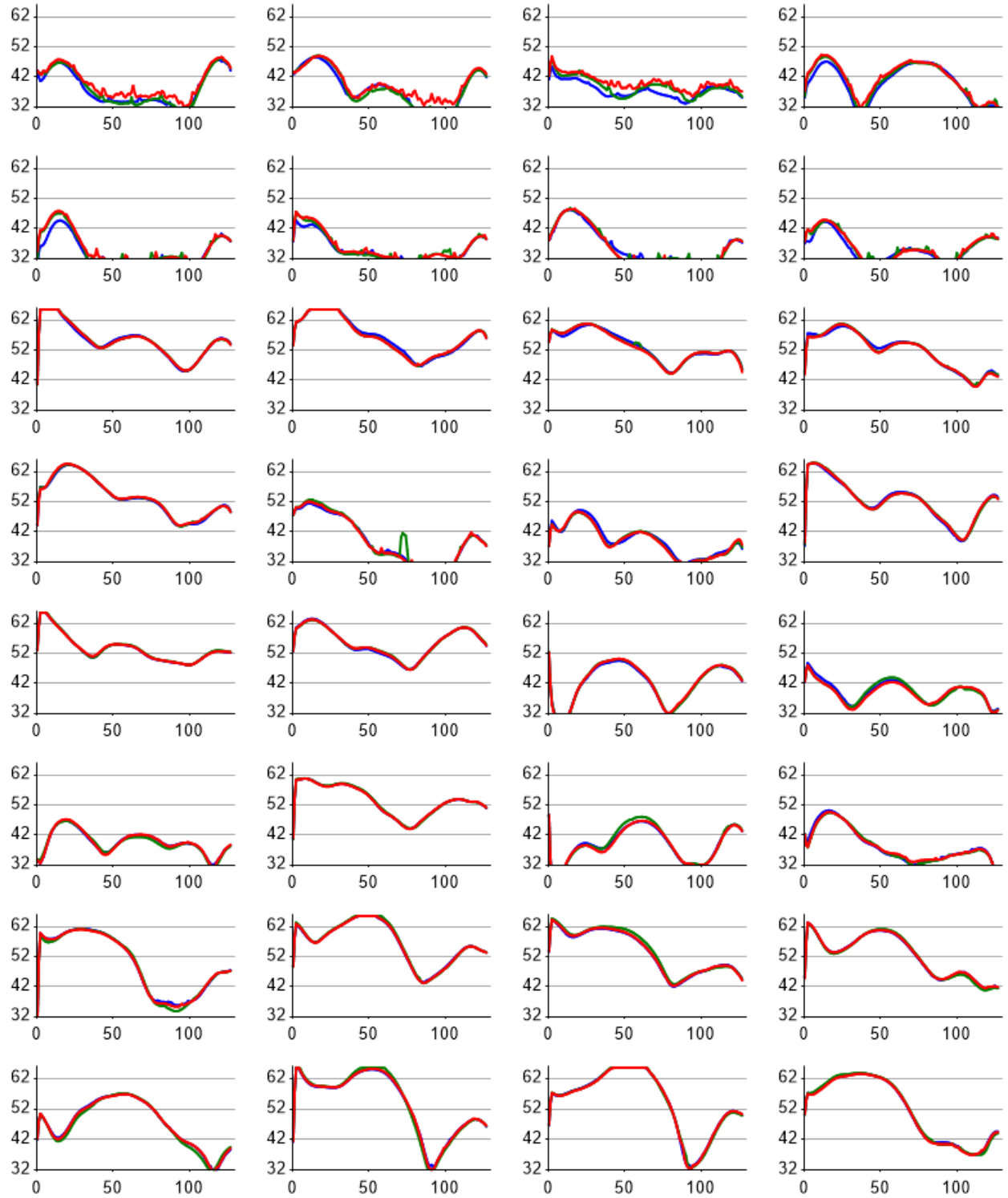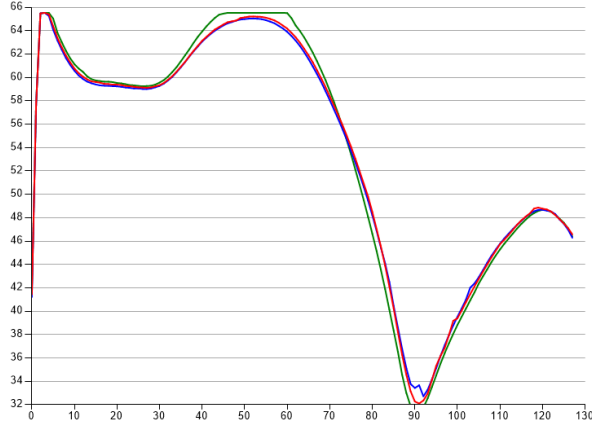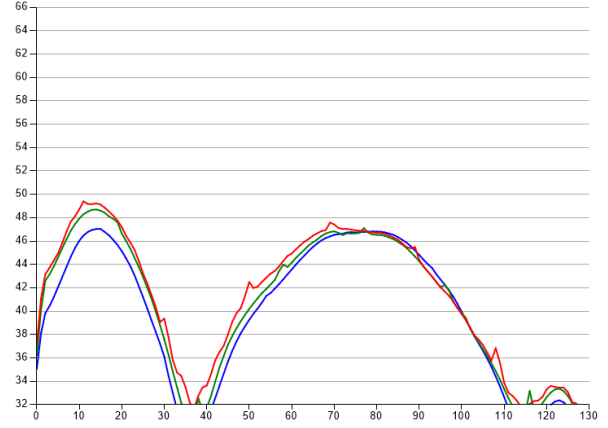
Figure 5: The frequency response of the 32 cells under 3 conditions. The red curve represents the case in which there is no object in the proximity of the RF matrix, the blue is when a hand is on the upper part of the board and the green is when a hand is on the lower part of the board. The X-axis is the value of the PWM and hence represents the frequency and the Y-Axis is the power measurement as measured by the A/D converter.

| (a) Cell 29 | (b) Cell 3 |

Figure 6: The frequency response of two sensor cells from Figure 5. (a) sensor cell 29 in the lower part of the board. (b) sensor cell 3 in the upper part of the board. For both cells, the frequency response under 3 conditions is shown. The red curve represents the case where there is no object in the proximity of the RF array, the blue is when a hand is on the upper part of the array and the green is when a hand is on the lower part of the array. The X-axis is the value of the PWM and hence represents the frequency and the Y-Axis is the power measurement as measured by the A/D converter. The cell in the lower part of the board (cell 29) reacts to the hand above it while the cell on the upper part of the board (cell 3) reacts to the hand above its part of the board.

ferent from one another. This is due to the fact that the prototype was built out of two substrate boards that were manually glued together. This was necessary since each board is made of a different material as described in Section 2. However, the slight misalignments create noticeable affects when working in such high frequencies. It is also noticeable that the eight upper cells are reading lower values and have higher noise levels. Since these cells were connected to the same SP8T switches, this is likely to be due to imperfections in the manual soldering process. Both of these imperfections can be addressed when production-quality construction of the RF array is leveraged.

In the next sections we describe the design of the sensor cell driver and the RF array in more detail.

3.1 Single Sensor Cell Driver

The purpose of the sensor cell driver is to enable the low power microprocessor running at 16MHz to generate 6-8Ghz sine waveforms. To enable this, we leverage a voltage controlled oscillator (VCO) chip [19]. The VCO is able to generate high frequency signals based on a DC input voltage in the range of 7-9V. The higher the voltage is, the higher the frequency of the signal will be.

However, the MSP430 processor's voltage level is at 3.3V, significantly lower than the 7-9V input range of the VCO. In addition, even within the 0-3.3V operating voltage range of the micro-controller, there is no way to generate DC voltage levels at finer levels of granularity (i.e., 1V, 1.2V, 1.4V etc.). The ability to generate multiple constant DC levels is necessary to excite the sensor cells with multiple frequencies

using the VCO.

We address these limitations in two steps. First, we feed the Pulse Width Modulation (PWM) output of the micro-controller into a low pass filter (R-C circuit) to generate fine-grained DC voltage levels. The PWM output of the micro-controller is a square pulse with an adjustable duty cycle. By properly adjusting the duty cycle, we control the amount of time that the PWM pin is at 3.3V. While at 3.3V, the capacitor of the R-C circuit is charged to a non-zero voltage level. While the PWM output is at 0V, the output of the R-C filter is not 0V, but at the voltage level that the charged capacitor is able to maintain. Note that as the capacitor discharges over time, the output voltage of the R-C filter also decreases. However, as the PWM output oscillates between 0V and 3.3V, the capacitor is automatically re-charged, being able to maintain a constant DC voltage at the output of the R-C circuit. The DC voltage depends on the duty cycle of the PWM pulse. By properly setting the duty cycle, different DC voltage levels are generated at the output of the R-C filter.

The PWM signal has a frequency of 124KHz (16MHz divided by 135). In order to obtain the needed voltage levels, the duty cycle is scanned in the range of $1/135 \ldots 128/135$. In that way, the frequency response of each sensor cell is recorded for 128 different frequencies.

Note that the DC voltages produced this way are still within the 0V-3.3V range, which is less than the voltage input range of the VCO. An amplifier [21] is used to raise the voltage levels, so that they are compatible with the VCO input.
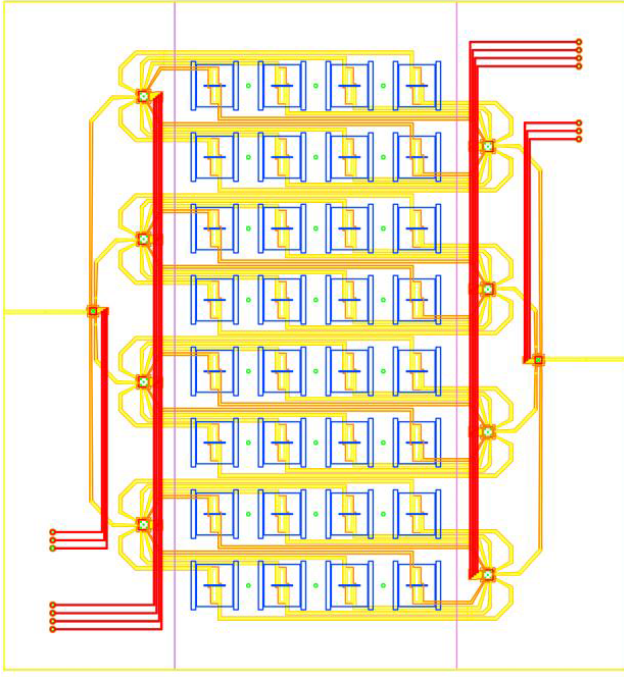
7

Figure 7: Illustration of the 4 x 8 sensor board with integrated switches. The different colors represent the different layers. The blue is the top layer, the yellow is the middle layer and the red is the bottom layer. The total size of the board is 210 mm x 195mm. The row-to-row separation between the cells is about 6.5 mm, while the column-to-column separation is about 7.2 mm.

3.2 RF Array

Preserving the same requirements outlined in the description of the single sensor cell, a 4 x 8 array of single cells is designed onto a single multilayer board to create a new 3D interactive sensor layer. Figure 7 shows an illustration of this new design. The arrangement of utilizing a two-port, half-wavelength coupled bandpass filter with a resonator patch is maintained for each of the 32 sensor cells. Besides the 32 sensors, the board is equipped with 10 RF switches, additional RF lines and DC control lines to control the switches, and DC bias lines to power the switches. In order to maintain a compact design that is simple to implement and cheap to manufacture, the ground plane and the bandpass filter exist on the same layer in a coplanar waveguide configuration. All of these lines have 50Ω impedances. This implementation allows the use of only two substrate boards to realize this sensor design. The use of 32 single sensor cells in this configuration produces 32 transmission responses that can be analyzed for unique signatures in the presence of a human hand similar to the responses of Figure 5.

In this multi-sensor design, it is strongly desired to minimize the number of RF lines on the board to decrease potential cross coupling of signals and reduce power consumption.

Therefore, switches are implemented in this design to facilitate the transmission and reception of signals. Of the 10 switches used in the sensor board, eight of them are GaAs MESFET single-pole eight-throw (SP8T) switches [10], and two are GaAs MESFET single-pole four-throw (SP4T) switches [9].

Figure 8(a) displays a simplified and enlarged block diagram of the interaction between the SP8T switches and a representative set of 8 sensor cells. A pair of SP8T switches (one for the transmission side of the filter and one to control the receive side) is used to control the RF signals of the eight sensor cells. The RF lines of each SP8T switch are voltage-controlled by a 3-to-8 decoder. Each SP8T switch contains one single-pole RF line, eight RF throw lines (connected to each sensor cell), three voltage-controlled lines to control the eight RF throw lines, and a DC bias line. The RF lines of the SP8T switch are bidirectional; therefore, for the Tx switches, the single-pole RF line operates as an input line, and the eight RF throw lines serve as output lines. When referring to the Rx switches, this operation is reversed. All of the DC lines (bias line and control lines) are on the backside of the board to preserve good isolation between RF and DC signals. Replicating this process four times governs how the signals of the 4 x 8 matrix are controlled.

Figure 8(b) illustrates a block diagram of the interaction between the SP4T and SP8T switches. Here, each of the single-pole RF lines of the Tx SP8T switch is connected to a throw line of the Tx SP4T switch. The same connections exist on the receive side. The RF lines of each SP4T switch are voltage-controlled by a 2-to-4 decoder. Each SP4T switch contains one single-pole RF line, four RF throw lines (each connected to an SP8T switch), two voltage-controlled lines to control the four RF throw lines, and a DC bias line. The RF lines of the SP4T switch are also bidirectional; hence, for the transmit side, the single-pole RF line operates as an input line, and the four RF throw lines serve as output lines. When referring to the Rx switches, this operation is reversed.

4. TIMING AND POWER ANALYSIS

In this section we analyze the timing of the sensors, that is, the rate in which it completes a scan, and its power requirements.

4.1 Timing

Each scan of the RF matrix requires the micro-controller to set 128 different frequencies, sample the frequency response for each frequency through the embedded ADC, and repeat this process for all 32 sensor cells. Table 1 shows the time it takes for each of these steps to complete in our implementation. Simply recording the frequency response of a sensor cell at a given frequency takes $520\mu s$, with the micro-controller's ADC being responsible for the majority of this time ($450\mu s$). Scanning all 128 frequencies for all 32 sensor cells, takes $2.2s$.

Our experimental setup was tuned towards verifying that the measurements of the sensor allow recognizing the prox-
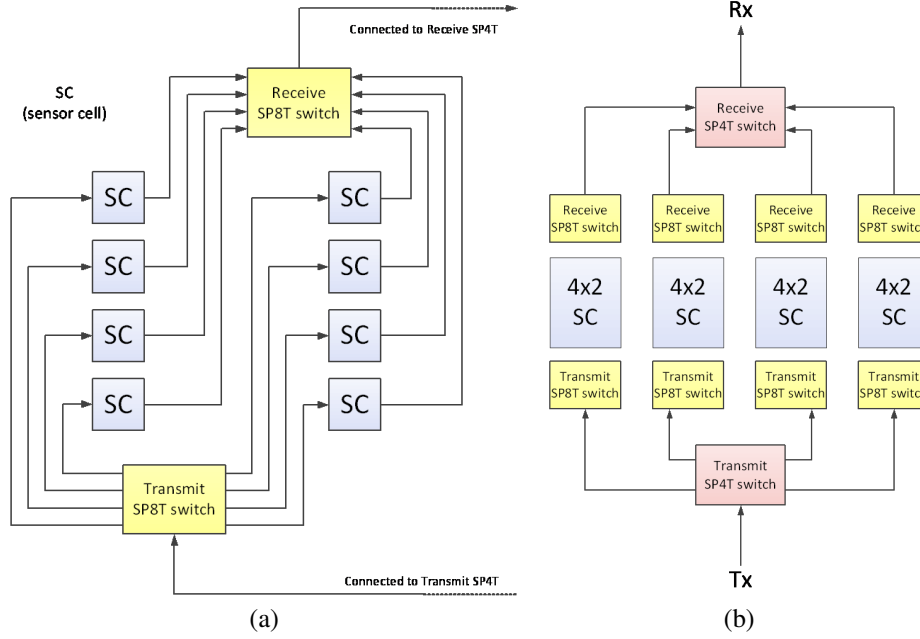
Figure 8: Block diagram showing: (a) interaction between SP8T switches and sensor cells, (b) Block diagram showing interaction between SP8T and SP4T switches

imity of an object. Therefore, we did not attempt to optimize the timing of our measurement equipment. However, to turn this sensor into a useful human-computer interface, the scan rate has to be improved by at least two orders of magnitude. Therefore, we provide an analysis of method to gain this speedup that according to our analysis can reduce the time to complete a scan to less than 6 ms.

We propose two main tools to gain the desired speedup. The first tool is a more careful use of the micro-controller and its abilities. In our experiments, we have used the 16 bits sigma-delta analog to digital converter that the MSP430 offers. This converter uses interrupts to indicate the readiness of a new sample and was responsible to most of the delay in the scanning process. As Table1 shows, the ADC was responsible to 456 μs out of the 520 μs that each frequency reading consumed, or in other words 88% of the time. However, the same micro-controller has a 12 bits ADC which is capable of sampling at up to 200 KS/s that is 5 μs per sample [12].

The second most time consuming action in each sample is the time it takes the R-C circuit to stabilize every time a new frequency is dialed. We can lower the penalty for this by choosing to first tune a frequency and then scan all the sensor cells. Therefore, under this scan order, every measurement takes 5.24 μs and additional 50 μs every time a frequency is changed. As a result, measuring all 128 frequencies on 32 cells can be done in less than 28 ms which amounts to 35 scans per second.

Even this number can be further improved. In Section 5.3 we show that it is not necessary to scan all 128 frequen-

Table 1: Delays for the different parts of the RF-sensor scanning process

Stage	Time
Setting the PWM duty cycle	Negligible
R-C circuit stabilization	50 μs
VCO stabilization	90 ns
Switching delay	150 ns
Power Detector stabilization	90 ns
Reading a data sample from ADC	456 μs
Time to read one frequency response per cell	520 μs
Time to scan all 128 frequencies per cell	66.56 ms
Time to read one frequency across all 32 cells	16 ms
Time to scan all 128 frequencies for all 32 cells	2.2 s

cies. Instead, we leverage feature selection techniques to show that the accuracy of the recognition is not significantly harmed when not all frequencies and cells are scanned. Reducing the number of frequencies scans results in a super-linear improvement in the timing. For example, if we consider scanning only 32 different frequencies, not only do we get a 4x speedup due to the lower number of measurements required, but we can also increase the frequency of the PWM signal and hence reduce the stabilization time of the R-C circuit to about 12.5 μs. Therefore, in this setting, a complete scan can take as little as 5.8 ms or 173 scans per second.

(a) Hand on the left side of the board.　　(b) Hand on the middle of the board.　　(c) Hand on the right side of the board.
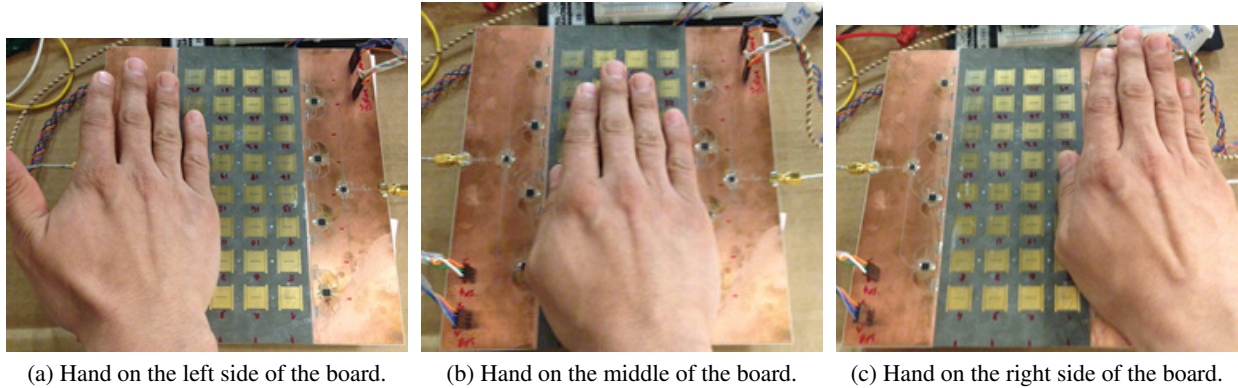
Figure 9: The 3 different positions considered for gesture classification. At each position, data was recorded while the subject's hand was at 5 different distances from the board. From completely touching the board, all the way to 2-inches distance from the board in 0.5inches increments.

4.2 Power consumption

Table 2 shows the current consumption of the different components of the RF sensor's supporting circuitry. When added up, the current dissipation of the proposed approach is approximately $115mA$[3]. Note, however, that the power detector requirement is an artifact of our prototype implementation and could be eliminated. Instead, the signal at the output of the sensor cell could be directly mixed with the signal at the input of the cell to provide a DC voltage representing the transmission response of the sensor cell at the current frequency. In this case, the overall dissipation of the RF array would be approximately $75mA$.

In comparison, a typical IR proximity sensor, one of the most widely used proximity sensors in mobile devices, has a typical current dissipation of $30mA$. However, a single IR proximity sensor can only perform 3D gesture recognition when the human hand is at a minimum distance from the sensor. When the hand is very close to the sensor, it completely covers its surface, and any movement of the hand cannot be reliably detected for 3D gesture recognition. In this case, an array of multiple IR proximity sensors is required. When two or three IR proximity sensors are combined the current dissipation becomes approximately $60ma$ and $90mA$ respectively, which is comparable to the current dissipation of the proposed RF sensor. When a power detector is not utilized, as explained above, the power consumption of the proposed RF sensor can be lower. As a result, the proposed RF-based 3D gesture sensing solution can operate at a similar power domain to traditional approaches while eliminating some of their bottlenecks such as line-of-sight requirements.

Table 2: Power consumption of the different components.

Component	Power
Amplifier	4.3mA
VCO	53mA
Switches	16mA
Power Detector	41.6mA

5. EVALUATION

A user study with 10 adult subjects (7 men and 3 women) was conducted to evaluate the performance of the proposed RF-based gesture recognition approach. The goal of the user study was to characterize the ability of the sensor array to accurately and quickly recognize gestures at different positions and distances. Each subject was asked to place his hand at three different positions over the board as shown in Figure 9. For each position, data was collected when the user's hand was at 5 different distances from the board; from completely touching the board, all the way to 2-inches distance from the board in 0.5-inches increments. For every position and distance from the board, 10 complete scans of the sensor array were collected to capture the temporal variability of the RF signals. Overall, 150 scans were recorded for each user. During each scan, the frequency response of all 32 sensors on the board was measured on 128 different frequencies at the range of 6-8GHz which corresponds to a resolution of about 16MHz. Therefore, each complete scan consists of 4096 measurments.

The 4096 measurments at each position and distance combination become the set of features used for training machine learning models to classify the distance and position of a human hand above the board. In this work, we train Random Forest Trees [2] using the WEKA machine learning toolkit [8]. We perform a *leave-one-user-out* evaluation, where one user is left out as the test dataset and the remain-
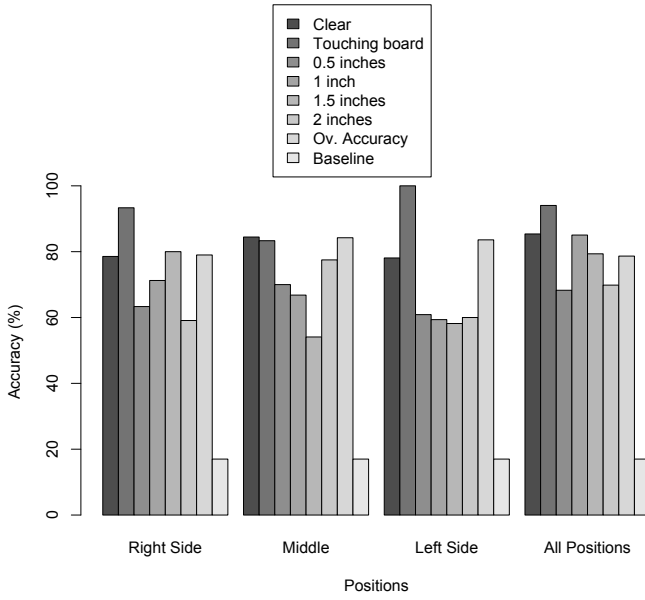
---

[3]In our calculations the current dissipation of the CPU is not taken into account as any approach to the problem would require one to process the data.

Figure 10: Distance classification accuracy.

Figure 11: Position classification accuracy.

ing users are used as the training dataset. This process is repeated, leaving each one of the users out at a time and performance results are averaged at the end.

5.1 Distance Classification

Using the complete set of features, we first train models to classify the distance at which the human hand is from the board. To do so, we use 6 classes. Five of them represent the different distances from the board (0-inches to 2-inches with 0.5-inches increments), and the last one represents the case where the board is clear in the sense that the human hand is not hovering above the board. Using these classes, we first train models for each of the 3 positions of the human hand separately (Figure 9). Then we combine the data across all hand positions to train a single distance classification model. In all cases, Random Forest Tree models were trained with 10 trees.

Figure 10 shows the distance classification accuracy achieved by the different models. Overall, distance of the human hand can be classified with an accuracy that is higher than 78%. Since we considered 6 different classes, a random predictor would have an accuracy of only 17% which is approximately 5 times lower than the accuracy achieved by the random forest tree models. This indicates the ability of the RF-based sensor cells to accurately capture the presence and proximity of the human hand in their frequency responses.

When considering the accuracy within each individual class representing a distance from the board, more information about the capabilities and limitations of the RF array are revealed. Clearly, the highest accuracy is achieved when the human hand completely touches the board. This is expected,
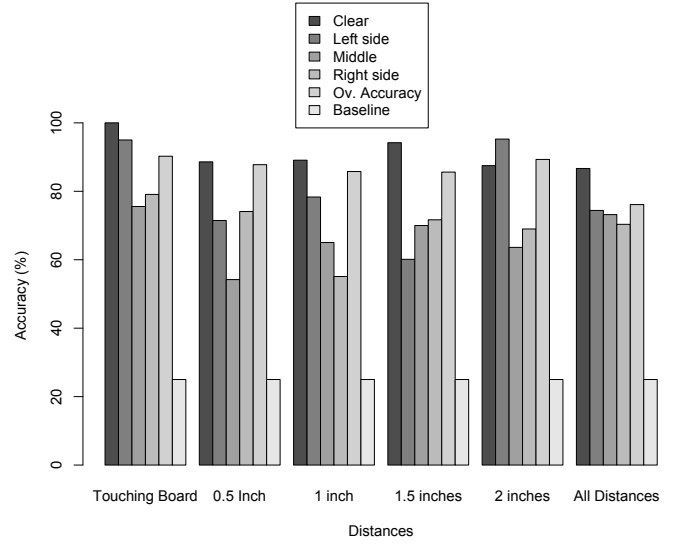
as when the human hand touches the board, the impact on the frequency response of the sensor is maximized resulting into more distinguishing frequency responses. As the human hand distance from the board increases, the impact of the human hand on the frequency response of the sensor cells is reduced, and the classification accuracy decays accordingly.

Note that the classification accuracy for the clear board (no hand above the board) is not as high as when the human hand touches the board. This is due to the temporal variations of the sensor cells' frequency responses that are caused by the manufacturing process of the current prototype. We believe that a production-quality sensor cell would minimize these temporal variations enabling higher classification accuracy. However, even under our current implementation imperfections, recognition accuracy remains higher than 78%.

When data across hand positions is used to train the distance classification model, the accuracy achieved is higher compared to the accuracies achieved for individual positions. At first, this is counter-intuitive as one would expect the combination of data across hand positions to have greater variability, and therefore lead to lower accuracy. However, the spatial placement of the sensor cells helps mitigate these variabilities, as different sensor cells react to the human hand at different positions. In addition, when data across positions is combined, a larger training dataset is available, helping the distance classification model to better capture data variations.

Figure 10 also shows that distance classification varies when different positions of the human hand are considered. These variations can be as high as 26%. Intuitively, distance classification accuracy should not depend on the position of the user's hand. However, as shown in Figure 5 the fre-
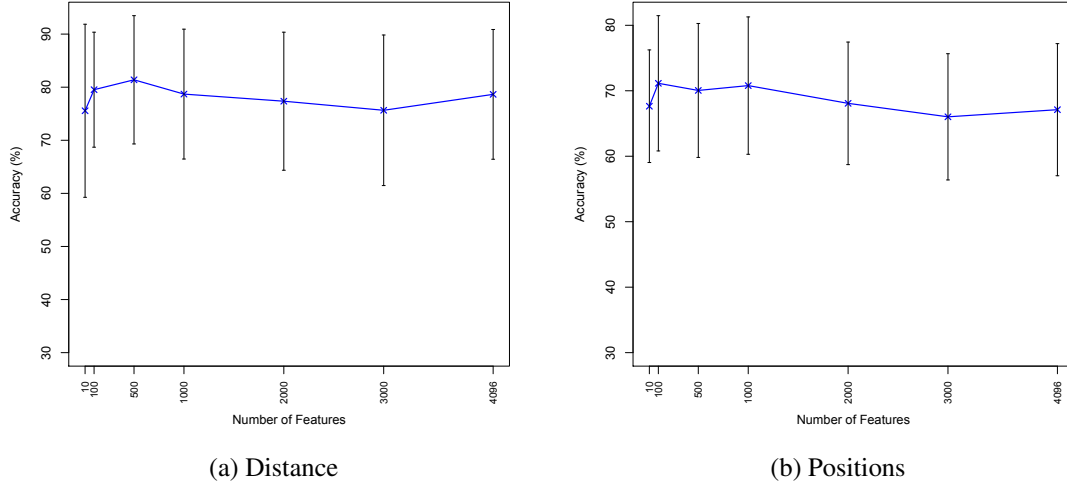
(a) Distance

(b) Positions

Figure 12: Accuracy achieved in both classification tasks, when different number of features is used.

quency response of each sensor cell is different due to the manual manufacturing process of the prototype. As a result, different sensor cells have different sensitivity characteristics leading to different areas of the board being able to recognize distances with a varying degree of success. Such imperfections are expected and can be addressed when production-level manufacturing processes are used.

5.2 Position Classification

The exact same process we used for distance classification is repeated in the case of position classification. This time random forest trees are trained to classify the position of the human hand on the board. Overall, four different classes are used indicating the three different positions if the user's hand and the case where the board is clear (no user hand is above the RF array). Accuracy is reported when the data recorded at specific distances are used, as well as when all data across distances are combined.

Figure 11 shows the performance of predicting the position of the user's hand across all possible combinations. Overall, an accuracy higher than 75%, and in most cases higher than 85% is achieved. Given that we train models against 4 classes, the achieved accuracy is approximately 3 times higher than the accuracy of a random predictor (25%). As a result, the spatial resolution that the RF array of sensor cells offers can accurately capture the position of the user's hand.

When considering the accuracy within each individual class representing a position of the user's hand, the imperfections of the manufacturing process of the prototype board can be seen. The highest prediction is achieved for the "Clear" class where there is no human hand above the RF array. However, the accuracy varies across positions with the "Middle" and "Right" position achieving the lowest accuracy with respect

to the rest of the classification labels. Again, this is attributed to the imperfections of the different sensor cells leveraged in the RF array (Figure 5).

Similarly to distance classification, position classification accuracy degrades as the distance of the user's hand from the board increases (Figure 11). When the data across all distances are combined to train a single position classification model, the overall accuracy is reduced to 76% from 85%. This reduction is expected given the variability in the recorded data that the placement of the user's hand across multiple distances introduces. However, even in this case, the classification model can achieve high classification accuracy.

5.3 Feature Set Size Reduction

In all of the experiments presented so far, all 4096 available features were used when training the random forest tree classification models. However, as discussed in Section 4.1, the larger the number of features we leverage the higher the time it takes to scan all sensor cells in the RF array, and therefore the slower gesture recognition is going to be. To study the impact of the number of features leveraged during training on the overall accuracy of the classification models, we repeat the leave-one-user-out cross-validation evaluation of the previous section, but this time while limiting the number of feature that the trained models can leverage. In particular, we re-compute the accuracy for both distance and position classification models when the random forest tree models are restricted to leverage any 10, 100, 500, 1000, 2000, 3000, or all 4096 features.

Figure 12 shows the distance and position classification accuracy as a function of the number of features used. When 500 and 100 features are used, the highest accuracy for classifying distances (79% and 82% respectively) and positions

(70% and 71% respectively) is achieved. Note that this corresponds to only using 12.2% and 2.44% of the overall number of features available. In addition, when only 10 features are leveraged, distance and classification accuracy becomes 67% and 75% respectively. As a result, even a very small number of features can provide high enough classification accuracies.

By reducing the total number of features, the overall number of frequency responses we need to collect from each sensor cell is reduced. For instance, scanning a single frequency from a single sensor cell takes approximately $520\mu s$. Scanning only 10 frequencies, instead of 4096, would require only $166ms$, instead of 2.2 seconds (Table 1). In that way, the scan rate of the RF array can be improved by several orders of magnitude, while maintaining an acceptable recognition accuracy.

The experiments conducted show that the RF sensor matrix is delivering information about the position and the distance of a hand from the sensor. The accuracies reported are significantly higher than the baseline. Furthermore, we conjecture that significantly higher accuracies are achivable by more control on the production process of the board and by larger training dataset.

6. RELATED WORK

Research in proximity sensors has been performed for well over 50 years. Most of these sensors are derived from detecting the change of electromagnetic fields in the presence of a target [5, 13]. They fall into three main classes: optical proximity sensors, inductive proximity sensors, and capacitive proximity sensors. Much research has taken place on designing optical proximity sensors for detecting human presence [23]. These sensors have been embedded across many industries including mobile devices, gaming consoles, household appliances, and automotive applications. Although this technology requires line-of-sight interaction to function property (an attribute that may not be feasible for some scenarios), these sensors can be manufactured in a small form factor (where circuit space may be constrained) and provide great variation in range of detection.

Inductive proximity sensors have been utilized primarily the detection of metals in machinery or automotive equipment [11, 18] or for detection of cars in traffic. These sensors offer non-touch detection, but the target object has to be magnetic and induce a current in the sensor or sensor must be powered to detect materials that interact with the magnetic field. As a result, inductive sensing of humans is difﬁﾅcult to achieve.

Capacitive sensors have a wider range of applications from mobile computing devices [1] to liquid level sensing [6]. A major advantage of capacitive sensors is the ability to sense objects that are in physical contact with electrodes as well as those that are not in physical contact; this is extremely useful for interacting with devices using the human hand and fingers as input tools. Decades ago, a musical instru-

ment, called a theremin, was very popular due to this interaction [24]. An individual would stand in close proximity to two capacitive plates connected to the instrument. Each plate is loaded with an inductor to create an oscillator. Since our human body has conductive properties, it behaves as a grounded capacitor. As the distance between the hands varies, the capacitance of the oscillator changes. Therefore, when the individual moves his hands he is able to control the frequency and amplitude of the audio signal, which in turn, controls the pitch and the volume of the music.

Researchers at MIT expanded greatly on the idea of capacitive sensing by placing electrodes underneath the display of a portable computer [25]. In this approach, the electric field between two electrodes is disturbed by the presence of a human finger which facilitates a change in the capacitance. They call this technology electric field imaging. This approach produces a limited amount of information that can be obtained from a scalar capacitive change which may not be suitable for complex gesture recognition. In addition, this approach operates at low frequencies where the sensitivity of sensing around the electrode can be small.

Recently, researchers at Disney Research proposed a new technology called Touche [20]. This technology is a capacitive sensor that detects the physical contact of an electrode. It has been utilized in many instances such as with a utensil to correct eating posture and embedded in a couch to signal the presence of a human sitting down. This method expands on the scalar capacitance change of [25] by analyzing this change across a small frequency band which creates a signature profile that can change based on the way the electrode is contacted. Here, the electric fields are tightly conïﬁﾅned to the electrode direct contact is necessary for sensing to take place. At low frequencies, the fields of such electrodes are mostly quasi-static, and sensing is based on field disturbances. A fundamental limitation of such an approach is that the disturbance depends on the electrode size relative to the objects. Given a defined object size, the sensitivity and range can be improved with the electrode size, but it also increases the amount of noise in the system. Furthermore, a minimum size limitation on the electrode may be highly undesirable in mobile electronics applications [4, 16].

7. CONCLUSIONS

We have presented a new RF-based approach to touchless gesture recognition that enables more flexible and natural interaction to current alternatives. Being RF-based, the proposed solution presents several distinct advantages over current sensing technologies which include the ability to work without line of sight, the ability to be easily embedded behind any type of surface, and the ability to scale to almost any size; all while operating at a similar power domain to current proximity sensing technologies. Using an empirical user study we demonstrated that the proposed sensor can indeed accurately sense the proximity of the human hand in three dimensions with an accuracy that is higher than 75%.

Our current prototype implementation, though, suffers from inefficiencies that, if properly addressed, can increase the performance and potential of the approach. For example, although we have shown in our analysis that the scan rate can be improved by more than 2 orders of magnitude, our experimental setup achieved a low scan rate. Furthermore, we have discussed several of the manufacturing inefficiencies of the current prototype, and suggested that improving the manufacturing process and collecting more data will result into great improvements in terms of both the accuracy and the usability of the sensor. We plan to address several of these open topics in our future studies.

8. REFERENCES

[1] F. Aezinia, Y. Wang, and B. Bahreyni. Touchless capacitive sensor for hand gesture detection. In *Sensors, 2011 IEEE*, pages 546–549. IEEE, 2011.

[2] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[3] A. Butler, S. Izadi, and S. Hodges. Sidesight: multi-touch interaction around small devices. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 201–204. ACM, 2008.

[4] O. Camacho and E. Viramontes. Designing touch sensing electrodes: electrical considerations and recommended layout patterns. `http://www.freescale.com/files/sensors/doc/app\_note/AN3863.pdf`.

[5] J. L. Carr, C. C. Jobes, and J. Li. Development of a method to determine operator location using electromagnetic proximity detection. In *Robotic and Sensors Environments (ROSE), 2010 IEEE International Workshop on*, pages 1–6. IEEE, 2010.

[6] C.-T. Chiang and Y.-C. Huang. A semicylindrical capacitive sensor with interface circuit used for flow rate measurement. *Sensors Journal, IEEE*, 6(6):1564–1570, 2006.

[7] S. Cohn. Parallel-coupled transmission-line resonator filters. *IRE Trans. Microw. Theory Tech.*, 6(2):223–231, April 1958.

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1), 2009.

[9] Hittite. Sp4t positive control switch hmc345lp3. `http://www.hittite.com/products/view.html/view/HMC345LP3`.

[10] Hittite. Sp8t positive control switch hmc321lp4. `http://www.hittite.com/products/view.html/view/HMC321LP4`.

[11] B. Inc. inductive proximity sensors. `http://www.balluff.com/balluff/MUS/en/products/Inductive-Sensors.jsp`.

[12] T. Instrument. Low power cpu msp430f47187.

[13] K. Koibuchi, K. Sawa, T. Honma, T. Hayashi, K. Ueda, and H. Sasaki. Loss estimation and sensing property enhancement for eddy-current-type proximity sensor. *Magnetics, IEEE Transactions on*, 42(4):1447–1450, 2006.

[14] Microsoft. Kinect. `http://www.xbox.com/en-US/kinect`.

[15] L. Motion. Leap motion. `http://www.leapmotion.com`.

[16] B. Osoinach. Proximity capacitive sensor technology for touch sensing applications. *Freescale White Paper*, page 12, 2007.

[17] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):677–695, 1997.

[18] Pepperl+Fuchs. proximity sensors. `http://www.pepperl-fuchs.com/global/en/classid_142.htm`.

[19] RFMD. voltage controlled osciolator (vco) rfvc1802. `http://www.rfmd.com/CS/Documents/RFVC1802DS.pdf`.

[20] M. Sato, I. Poupyrev, and C. Harrison. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 483–492. ACM, 2012.

[21] L. Technology. Dual/quad rail-to-rail input and output precision op amps lt1632. `http://www.linear.com/product/LT1632`.

[22] L. Technology. Rms power detector ltc5582. `http://www.linear.com/product/LTC5582`.

[23] S. Tsuji, A. Kimoto, and E. Takahashi. A multifunction tactile and proximity sensing method by optical and electrical simultaneous measurement. 2012.

[24] M. Vennard. Leon theremin: The man and the music machine. *BBC News Magazine*, March 2012.

[25] T. G. Zimmerman, J. R. Smith, J. A. Paradiso, D. Allport, and N. Gershenfeld. Applying electric field sensing to human-computer interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 280–287. ACM Press/Addison-Wesley Publishing Co., 1995.