ScreenSense: Utilizing Communication Signals for Dynamic Finger Tracking for On-Screen Antennas

Satyam Prakash Gupta[1]
Supervisor(s): Dr. Xing Wang[1]**, Shun Zhuge**[1]
[1]EEMCS, Delft University of Technology, The Netherlands

21 June 2026

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Name of the student: Satyam Prakash Gupta
Final project course: CSE3000 Research Project
Thesis committee: Dr. Xing Wang, Shun Zhuge, Mark Neerincx

Abstract

Future 6G smartphones are proposed to embed transparent on-screen antenna arrays that use communication signals for passive finger tracking. Our research proposes two novel localisation methods that exploit the finger's electromagnetic backscattering response. Using model-generated time-series data, we simulate the spatiotemporal backscattering of a finger hovering above a transparent planar array at sub-terahertz frequencies. We compare a classical matched filter and subspace methods against our proposed approaches: a CNN-adapted matched filter (MF-CNN) and a multi-tone CNN position regressor (MT-CNN), alongside a near-field subspace baseline. The learned methods achieve sub-millimeter accuracy and remain robust across variations in signal-to-noise ratio, array size, dielectric properties, and hover height, with MT-CNN offering the best trade-off between accuracy and latency.

1 Introduction

Sixth-generation (6G) networks will introduce integrated sensing and communication (ISAC), requiring antennas to support passive sensing capabilities alongside conventional communication functionalities. Future smartphones are expected to integrate on-screen antenna architecture, embedding planar antenna arrays directly into the screen surface, however user-induced blockages degrade network throughput [11], in activities like playing mobile games, streaming, etc. Lightweight algorithms are required to estimate the 2D coordinates by exploiting the electromagnetic (EM) spatiotemporal variations induced by these blockages on array response.

Backscatter-based localisation has so far been simulated mostly in the far field, or by reducing near-field objects to stationary point scatterers. Learned subspace methods, for instance, model the target as a point source and estimate its range and angle in the near field [6, 1], while related work focuses on enhancing the backscattered return of such point targets [4]. A finger, however, is not a point: it is an extended, lossy dielectric body that simultaneously backscatters and blocks the field across many array elements, captures neither its size nor its dielectric interactions. The ScreenAnt platform moves closer to the present setting by demonstrating a transparent on-screen array and characterising the throughput lost to touch-induced blockage [21], but it treats that blockage as a communication impairment to be mitigated rather than a signal to be exploited, and offers no forward model of the finger's backscatter from which localisation data could be generated.

On the algorithmic side, localisation has relied either on subspace methods for point sources that emit their own waves, or on matched filtering without a target-specific template. In most cases targets are stationary. Model-based neural networks such as SubspaceNet learn a surrogate covariance to sharpen subspace direction-of-arrival estimation, recently extended to near-field range and angle [18, 6]; yet these methods assume a stochastic source observed over many snapshots, from which a signal subspace can be separated. A finger is instead a single deterministic backscatterer rather than a stochastic emitter, so the covariance it produces lacks the clear signal subspace these methods rely on, and it is extended rather than a point. Matched filtering, the optimal detector for a known signal in noise [19], sidesteps the subspace assumption but is classically posed for far-field, planar wavefronts and a fixed template, leaving it unadapted to the spherical wavefronts of the near-field and to the finger's perturbation-dependent backscatter signature.

To address these limitations, this research investigates: how accurately can the spatial backscattering response of a finger above a transparent on-screen planar antenna array be used to estimate its 2D position in an ISAC scenario? This is decomposed into two sub-questions. First, how can the backscattering response of a finger be modelled to generate synthetic time-series array response data? Second, how accurately can 2D localisation be inferred from this time-series data across varying array sizes and noise levels?

The report is structured as follows. Section 2 provides background research on 6G, transparent arrays, and our near field setup. Section 3 formulates the system model and localisation algorithms. Section 4 describes the experimental setup. Section 5 presents an analysis of localisation accuracy against existing baselines. Section 6 addresses ethical considerations and reproducibility. Section 7 discusses the limitations of the proposed approach, and Section 8 concludes the paper.

2 Background

6G networks are expected to operate across millimetre-wave (mmWave) and sub-terahertz bands, with frequencies from roughly 24 GHz upward identified as key spectrum for the simultaneous delivery of high data rates and fine-grained sensing [14][2]. At these fre-

quencies the wavelength falls to around 12 millimetres, which allows small antenna arrays, and sub-centimetre localisation.

Transparent antenna arrays are a promising route to combined communication and sensing on 6G-capable smartphones.[16][10] Embedding the array as a transparent layer beneath the screen, as in Figure 1, frees chassis space and places the elements on the unobstructed front face, reducing hand blockage. A 6G device meeting ISAC requirements could reuse such an on-screen array for sensing tasks such as finger localization, the scenario studied in this paper.

For a near-field setup, the formulation of electromagnetic wave propagation is spherical when objects are below the Fraunhofer distance. [9] The propagation of EM waves is treated spherically, with non-negotiable curvature. The Fraunhofer distance $d_F = 2L^2/\lambda$ equation, where L is the dimension of the array, and λ is the wavelength [12]. In our setup the lowest array size of 5 this distance is 85.68 mm, in the worst case, a finger is expected to hover at 10 mm (11.7%).

3 Methodology

3.1 System Model

This section proposes the system model which contains the antenna array configuration, scattering geometry of the finger, transmission of waves and the mutual coupling among antennas. The final signal is assumed to be mono-static, same antenna sends and measures response, and the finger is treated like a flat object.

3.1.1 Array Configuration

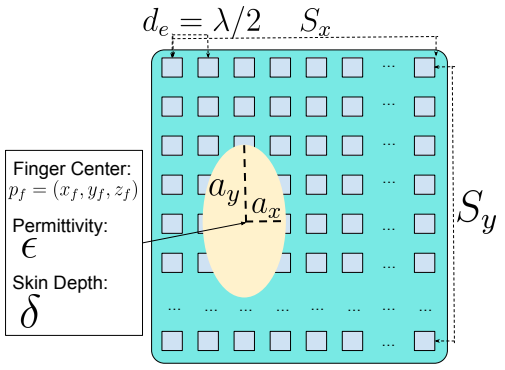

Figure 1: Visual Depiction of the Antennas Physical Configuration

The system consists of a $S_x \times S_y$ uniform planar array (UPA) of transparent ITO antenna elements integrated onto a 6G device screen, operating at $f_0 = 28\,\text{GHz}$ ($\lambda = c/f_0$), with half-wavelength element spacing $d_e = \lambda/2$ and total aperture $L = (S_x - 1)d_e$. Following [21], the array is centered at the origin with element coordinates

$$s_x = d_e\big(-\tfrac{S_x-1}{2} + \bar{s}_x\big)\,, \quad s_y = d_e\Big(\tfrac{S_y-1}{2} - \bar{s}_y\Big)\,, \quad (1)$$

where $\bar{s}_x = \text{mod}(s-1, S_x)+1$ and $\bar{s}_y = \lfloor (s-1)/S_x \rfloor + 1$ for $s \in \mathcal{S} = \{1, \ldots, S\}$. A hovering finger at height z_f satisfies $z_f \ll d_F = 2L^2/\lambda$, placing it in the radiative near-field and requiring spherical wavefront modelling [12].

3.1.2 Near-Field Scattering Geometry

The finger's instantaneous position is $\mathbf{p}_f(t) = (x_f(t), y_f(t), z_f)$, where lateral coordinates evolve over time and z_f is the nominal hover height. The round-trip distance from the array to element s is

$$r_s(t) = \sqrt{(x_f(t) - s_x)^2 + (y_f(t) - s_y)^2 + z_f^2}\,. \quad (2)$$

The finger's physical extent is modeled as a 2D Gaussian footprint $\rho_s(\mathbf{p}_f) \in (0, 1]$ with half-widths a_x and a_y, quantifying the fractional overlap between the finger and element s.

$$\rho_s(t) = \exp\left(-\frac{(s_x - x_f(t))^2}{2a_x^2} - \frac{(s_y - y_f(t))^2}{2a_y^2}\right), \quad (3)$$

The near-field steering vector $\mathbf{a}(t) \in \mathbb{C}^S$ has element s:

$$[\mathbf{a}(t)]_s = \sigma_f\, \rho_s(\mathbf{p}_f(t))\, \frac{e^{-j2kr_s(t)}}{r_s^2(t)}\,, \quad (4)$$

where $k = 2\pi/\lambda$, the factor e^{-j2kr_s} accounts for round-trip phase accumulation, r_s^{-2} reflects two-way amplitude decay, and σ_f is the backscatter factor.

3.1.3 Complex Transmission Matrix

Interaction of the 28 GHz signal with biological tissue is characterised by a complex permittivity $\tilde{\varepsilon}$ from the Gabriel dielectric model [20], which yields a skin depth δ and a maximum phase shift $\phi_{\max}$ through the finger. Element s experiences a complex transmission coefficient $\gamma_s = \beta_s e^{j\phi_s}$, where

$$\beta_s = \beta_{\min}^{\rho_s}, \qquad \phi_s = \phi_{\max}\, \rho_s\,, \quad (5)$$

with minimum transmission $\beta_{\min}$ under full coverage. These are collected into the diagonal blockage matrix

$$\mathbf{\Gamma}(t) = \mathrm{diag}(\gamma_1(t), \ldots, \gamma_S(t)). \tag{6}$$

3.1.4 Mutual Coupling

Electromagnetic coupling between array elements is modelled via the free-space scalar Green's function [1]:

$$[\mathbf{C}]_{ss'} = \begin{cases} 1 & s = s' \\ \kappa \, \dfrac{e^{-jkd_{ss'}}}{kd_{ss'}} & s \neq s' \end{cases} \tag{7}$$

where $d_{s,s'}$ is the inter-element distance from (1). κ is the coupling strength set to 0.3. At the compact spacings of an ITO screen array the off-diagonal terms are non-negligible, and $\mathbf{C}$ must be included in the forward model to correctly reproduce the observed spatial signature.

3.1.5 Received Signal Model

Combining the steering vector, the transmission matrix, and the coupling model, the received signal vector at time t is

$$\mathbf{y}(t) = \mathbf{C}\big[\mathbf{\Gamma}(t) \odot \mathbf{a}(t)\big] + \mathbf{n}(t), \qquad \mathbf{y}(t) \in \mathbb{C}^S, \tag{8}$$

where $\odot$ is the Hadamard product and $\mathbf{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ is additive white Gaussian noise. The signal component is entirely deterministic: at each snapshot the spatial pattern across elements is a function of $\mathbf{p}_f(t)$ alone, a property exploited directly by the localisation methods in the next section.

3.2 Localisation Algorithms

All methods estimate the finger position $\mathbf{p}_f$ from the received signal $\mathbf{y}(t)$ of (8). We develop two learned localisers: matched-filter convolution neural network (MF-CNN), multi-tone convolution neural network (MT-CNN) alongside the matched filter (MF), Multiple Signal Classification (MUSIC) and NF-Subspace[6] serve as comparison. Analytical time complexity is provided alongside inference time in table 4.

3.2.1 MF-CNN

Rationale: MF-CNN uses the same algorithm as the classical matched filter, paired with learned estimates of the physical parameters of the object it is detecting. The matched filter compares the snapshot $\mathbf{y}$ against a dictionary of position templates $\mathbf{t}(\mathbf{q})$ and takes the strongest correlation [19, 17]:

$$\hat{\mathbf{q}} = \arg\max_{\mathbf{q}} \frac{|\mathbf{t}(\mathbf{q})^{\mathsf{H}} \mathbf{y}|^2}{\|\mathbf{t}(\mathbf{q})\|^2}. \tag{9}$$

A fixed filter sets these templates at nominal tissue values and degrades when the finger's properties or hover height differ. Our contribution is to infer the template-shaping parameters from the data and rebuild the templates for the finger actually present.

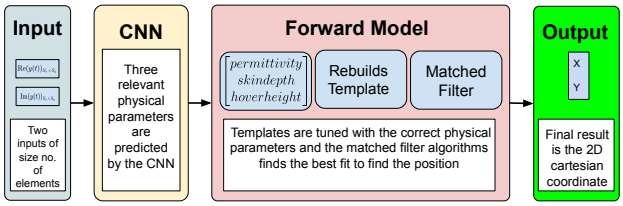

Figure 2: The snapshot image is mapped to template parameters $\hat{\boldsymbol{\eta}} = (\hat{\beta}_{\min}, \hat{\phi}_{\max}, \hat{z})$, which rebuild the matched-filter template.

Architecture: A convolutional network reads the snapshot as a two-channel $S_x \times S_y$ image (real and imaginary parts) and regresses the effective minimum transmission $\hat{\beta}_{\min}$, maximum phase shift $\hat{\phi}_{\max}$, and hover height $\hat{z}$, the quantities set by the finger's permittivity, skin depth, and distance from the screen. Because these parameters describe the finger and not its location, the network pools over the array before regressing them. The predicted physical parameters rebuild the template bank, after which localisation proceeds.

Training: The network is trained on simulated snapshots spanning a wide range of SNRs. This noise-aware training helps keep the parameter estimates usable in practice: the spatial structure the network reads is buried at low SNR, and exposing it to degraded inputs during training stops the rebuilt templates from drifting when the signal is weak.

3.2.2 MT-CNN

Rationale: MT-CNN utilizes 3 different tones in multi-tone waveform modulation to exploit diverse frequencies for localisation. Probing the finger with several closely spaced tones rather than one gives each snapshot greater discriminative information, as every tone returns a slightly different view of the same backscatter, and combining them makes the position more distinctive. This borrows from chirp-based sensing such

as Soli [13], which sweeps frequency. We approximate that sweep with a three discrete tones, 27, 28 and 29 GHz. Unlike the matched-filter methods, MT-CNN needs no template dictionary or grid search; it maps the multi-tone snapshot straight to coordinates.

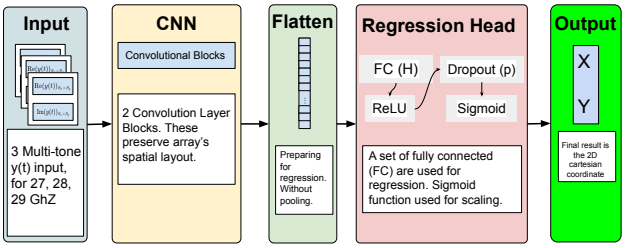

Figure 3: The multi-tone snapshot cube is mapped directly to the finger position $(\hat{x}, \hat{y})$ by a convolutional regressor.

Architecture: The MT-CNN has two convolutional layers followed by two fully connected layers. The input is the multi-tone snapshot stacked as a $2T$-channel $S_x \times S_y$ image (T tones, real and imaginary parts), preserving the array's spatial layout. The first convolutional layer (3×3, 32 filters) extracts local features of the backscatter footprint, and the second (3×3, 64 filters) combines them into higher-level spatial patterns. The feature maps are flattened without pooling, keeping the absolute position information that pooling would discard [8] and a 128-unit fully connected layer compresses them before the final layer regresses the coordinates $(\hat{x}, \hat{y})$ through a sigmoid scaled to the screen bounds.

Training: MT-CNN is trained directly on simulated trajectories, with noise added across a wide range of SNRs and the finger's physical parameters varied during training. This augmentation teaches it to localise even when the spatial pattern is partly buried or the tissue differs from nominal, so it degrades gracefully rather than collapsing at low SNR. A separate model is trained for each array size.

3.2.3 NF-Subspace

NF-Subspace is an external near-field adaptation of MUSIC, built in the AI-Subspace paper [6]. It is run unmodified on its own point-source signal under its own favourable conditions, ignoring the finger's footprint, blockage, and coupling. It is a best-case generic reference, not a like-for-like competitor; the gap to it measures the value of modeling the finger's true signature.

4 Experimental Setup

All data is model-generated from the forward model of (8); no measured data is used. The tissue and scattering constants of that model, derived at $f_0 = 28\,\text{GHz}$, are listed in Table 1.

Symbol	Description	Value
$\bar{\varepsilon}$	complex permittivity (A.2)	$17.50 - j9.63$
δ	skin depth (A.4)	1.53 mm
$\phi_{\max}$	max. phase shift (A.5)	2.99 rad
$\beta_{\min}$	min. transmission (A.6)	0.135
σ_f	backscatter factor (A.1)	0.772

Table 1: Derived physical constants of the forward model.

Noise will be analysed using signal-to-noise ratio (SNR). SNR is defined as $\text{SNR} = 10\log_{10}(\bar{P}_s/\sigma_n^2)$ [3], where $\bar{P}_s$ is the mean noiseless signal power of the configuration under test; the noise variance σ_n^2 is set per configuration so that the stated dB value is the true SNR of that signal. The SNR is swept over $\{-5, 0, 5, \ldots, 30\}$ dB; alongside signal noise, physical parameter perturbations are added to better reflect realistic testing conditions.

Two perturbation regimes are studied. In the *single-parameter* sweeps, one quantity is varied while the others stay nominal: skin permittivity $\varepsilon_r \in [8, 33]$, effective depth $d \in [0.5, 5]$ mm, and per-step height jitter standard deviation $\sigma_z \in [0, 8]$ mm. In the *combined* study a single scalar perturbation level $\ell \in [0, 1]$ scales the spread of all three simultaneously, drawn once per gesture for the tissue parameters and per snapshot for height:

$$\varepsilon_r \sim \mathcal{N}(17.5,\, \ell\,\sigma_\varepsilon), \tag{10}$$

$$d \sim \mathcal{N}(1.53\,\text{mm},\, \ell\,\sigma_d), \tag{11}$$

$$z_f(t) = z_f + \mathcal{N}(0,\, \ell\,\sigma_z^{\max}), \tag{12}$$

with $\sigma_\varepsilon = 8$, $\sigma_d = 1.5$ mm, $\sigma_z^{\max} = 6$ mm, and the draws clipped to physically admissible ranges. Level $\ell = 0$ matches the templates exactly; $\ell = 1$ represents heavy real-world mismatch.

Localisation accuracy is the root-mean-square error (RMSE) between estimated and true lateral positions over a trajectory, in millimetres. The lower, the better.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left\|\hat{\mathbf{p}}(t) - \mathbf{p}_f(t)\right\|^2}, \tag{13}$$

averaged over the test trajectories. All methods are compared at a single *realistic operating point* (SNR = 20 dB, $\ell = 0.5$), reporting the mean and standard deviation over repeated random perturbation draws.

5 Results

5.1 Robustness to Physical Parameters

Physical parameter perturbations are applied to three parameters: permittivity, skin depth, and hovering height. The error is measured, as motivated in 4.

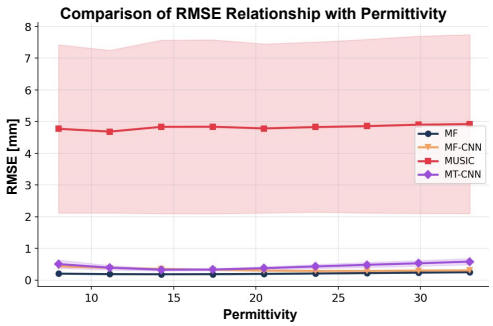

Figure 4: Performance of our models against permittivity variations [8, 33]

Permittivity: Under permittivity variation every method except MUSIC stays sub-millimetre, and here the gap between learned and classical methods nearly closes: MF is in fact the most accurate (0.18–0.25 mm with negligible spread), with MF-CNN and MT-CNN close behind (0.29–0.58 mm), while MUSIC again fails (≈4.8 mm, large variance). We analyzed the array response data, as shown in appendix B.1, where changing the permittivity showed no changes in the spatial response, only the amplitude of the back-scatter. Since amplitudes are normalized relatively, the changes in permittivity show very little variation in accuracy.

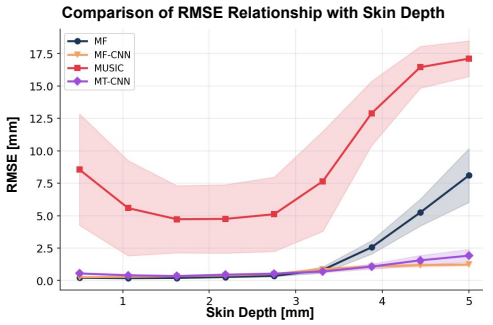

Figure 5: Performance of our models against skin depth variations [0.5, 5]mm

Skin Depth: Skin depth produces the clearest separation between the two families (Fig. 5). Up to roughly 3 mm all non-MUSIC methods are sub-millimetre and MF is the most accurate, but beyond this point MF collapses, its RMSE rising from 0.35 mm at 2.75 mm to 8.10 mm at 5 mm, while the learned methods stay bounded, MF-CNN never exceeding 1.22 mm and MT-CNN 1.91 mm.

A larger skin depth lets the field penetrate further, so the finger attenuates the signal less and imprints a weaker, shallower signature that a template fixed at the nominal contrast no longer matches, whereas MF-CNN estimates the reduced transmission directly and MT-CNN has been trained across such conditions. One likely reason of the MF collapse is that the response changes shape. We observed that at $\delta = 4.5$ mm the array response develops two lobes as shown in Fig. 11. With two near-equal matches the filter cannot decide between them, and even a stationary finger is localised alternately to either lobe (Fig. 12); because the lobes sit either side of the truth, the mean stays close while the variance grows, which is consistent with the steep rise in RMSE at large skin depth.

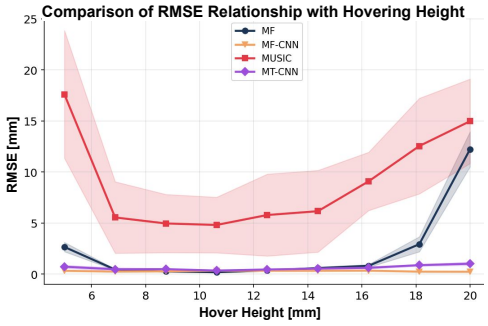

Figure 6: Performance of our models against hover height variations [5, 20]mm

Hovering Height: As hover height varies, the learned methods MF-CNN and MT-CNN remain far more robust than the classical MF and MUSIC algorithms, pairing lower RMSE with smaller standard deviation. MF-CNN is almost flat across the whole range (0.24–0.35 mm) because it estimates the hover height explicitly and rebuilds its template at the inferred distance, whereas the fixed-template MF is accurate only near the nominal 10 mm and degrades sharply at both extremes (2.67 mm at 5 mm, 12.23 mm at 20 mm). The effect is physical: as the finger rises its backscattered signal weakens and the contrast of its circular imprint on the array fades, as showing in Fig. 13, so a template fixed at the nominal height progressively mismatches. The CNN, in both approaches, appears to exploit the spatial features.

5.2 Localisation Performance on Simulated Trajectories

When all three dimensions are perturbed simultaneously (level $\ell = 0.5$, SNR $= 25\,\text{dB}$), the five methods separate clearly across every trajectory type, as summarized in Table 2.

Method	RMSE [mm]	Std [mm]
MF	0.24	0.20
MF-CNN	0.31	0.11
MUSIC	8.71	3.47
MT-CNN	0.43	0.10
NF-Subspace	3.80	1.59

Table 2: 2D localisation RMSE over 6 trajectories, linear, sinusoidal and 4 random.

Fig 7 overlays the predicted tracks of MT-CNN, MF-CNN and the NF-Subspace baseline on the ground-truth path, which traces a sinusoidal figure-of-eight.

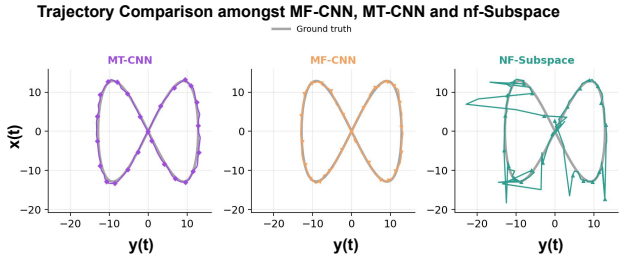

Figure 7: Sinusoidal trajectory trace of MF-CNN, MT-CNN and NF-Subspace.

Three methods reach sub-millimetre accuracy: MF (0.24 mm), MF-CNN (0.31 mm) and MT-CNN (0.43 mm). MF attains the lowest mean error but the largest spread of the three (± 0.20 mm), whereas MF-CNN (± 0.11 mm) and MT-CNN (± 0.10 mm) are markedly more consistent, trading a little accuracy for stability. The subspace methods trail by an order of magnitude, NF-Subspace at 3.80 mm and MUSIC, the worst overall, at 8.71 mm, and with far larger variance.

5.3 Robustness against Noise & SNR Analysis

Fig 8 sweeps the SNR from -5 to 30 dB at a fixed perturbation level $\ell = 0.5$. The MT-CNN is relatively the most accurate and stable method across the entire range, as it shows robustness at low-SNR levels, its error rises gracefully from 0.65 mm at 30 dB to 6.65 mm at -5 dB.

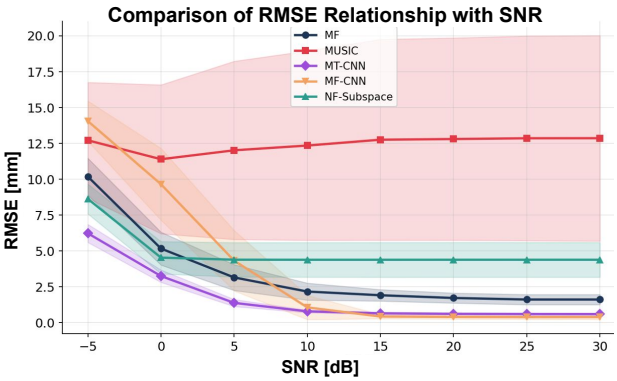

Figure 8: RMSE for all methods at a fixed perturbation level $\ell = 0.5$ as the signal-to-noise ratio is swept.

Interpreting this result, in MT-CNNs network design, the multiple tones supply independent looks at the same scene, and training across the full noise range teaches the network to exploit them when the signal is weak. The classic MF performs an order of magnitude worse that MT-CNN at low-SNR, and at high-SNR is still slightly worse. This can be interpreted as MF performing poorly due to the perturbation level, and below 5 dB, the error climbs steeply (6.43 mm at 0 dB, 11.47 mm at -5 dB). MF-CNN shares a similar trend but is slightly better, as at lower perturbation levels because it can correct for the physical parameter perturbations. NF-Subspace (baseline) is also relatively most stable across SNR but never accurate, holding ≈ 4 mm down to 5 dB and 8.38 mm at -5 dB, while MUSIC remains poor throughout ($\approx$ 18–20 mm).

5.4 Localisation Performance to Array Size

Scaling the number of elements, or array size, tests whether additional elements improve localisation. The element spacing is fixed at $\lambda/2$, elements are arranged in a square array, hence adding more elements increases the aperture. Array dimensions are kept in line with the Fraunhofer distance discussed in 2, to ensure near-field setup.

Method	25	49	100	196
MF	1.36 ± 0.22	1.71 ± 0.45	1.91 ± 0.24	1.88 ± 0.25
MF-CNN	0.30 ± 0.04	0.38 ± 0.10	0.39 ± 0.28	0.42 ± 0.30
MUSIC	5.93 ± 2.02	12.15 ± 3.17	18.22 ± 4.83	25.79 ± 7.54
MT-CNN	0.60 ± 0.04	0.52 ± 0.02	0.64 ± 0.05	0.70 ± 0.14
NF-Subspace	1.49 ± 0.49	4.42 ± 1.09	5.26 ± 1.05	10.89 ± 1.80

Table 3: 2D localisation RMSE versus array size (mean $\pm$ std, mm), evaluated at perturbation level 0.5 and SNR = 25 dB. Highlighted are subspace methods.

The results show that every method loses accuracy as the array grows, but the rate separates the sets of algorithms. The subspace methods degrade steeply, MUSIC rises from 5.93 mm at 25 elements to 25.79 mm at 196, and NF-Subspace from 1.49 mm to 10.89 mm. By contrast, MF, MF-CNN and MT-CNN rise only slightly: MF from 1.36 to 1.91 mm, MT-CNN within 0.52–0.70 mm, and MF-CNN from 0.30 to 0.42 mm.

5.5 Complexity and Computation

Empirical computation cost is measured by counting the Multiply-Accumulate (MAC) operations and the inference time, similar to the AI-Subspace paper [6]. Alongside the MAC count, an asymptotic analysis is performed using Big-Oh notation, in terms of the number of array elements S, the number of points G^2 in the spatial search grid, and the number of tones T.

Method	MACs	Asymptotic	Time [µs]
MF	78,400	$\mathcal{O}(SG^2)$	72.64
MF-CNN	1,018,368	$\mathcal{O}(SG^2)$	6608.78
MUSIC	4,111,345	$\mathcal{O}(S^3 + G^2S^2)$	947.82
MT-CNN	1,389,504	$\mathcal{O}(T * S)$	619.22
NF-Subspace	2,631,118	$\mathcal{O}(S^3 + G^2S^2)$	11334.62

Table 4: Empirical per-snapshot cost at the 7×7 array: analytical MAC count, asymptotic order, and measured inference time.

The classical estimators are dominated by the grid search. The matched filter correlates each of the G^2 candidate templates against the S-element snapshot, giving $\mathcal{O}(SG^2)$, and MF-CNN shares this cost, adding only a fixed-size network that rebuilds the templates from learned parameters. MUSIC additionally eigendecomposes the $S \times S$ covariance and projects every grid template onto the noise subspace, giving $\mathcal{O}(S^3 + G^2S^2)$. MT-CNN is the only grid-free method, as it performs one linear pass through its network, $\mathcal{O}(T*S)$, with the multi-tone input adding a factor T only at the first layer.

The measured costs follow these orders but also expose overheads the MAC count does not capture. MF is cheapest on both metrics because its templates are precomputed. MT-CNN, being grid-free, is the fastest of the learned methods and avoids any dependence on the spatial resolution G, so its advantage over the grid-based estimators widens as the screen and search grid grow.

6 Responsible Research

This research was conducted with a strong emphasis on transparency and ethical use of tools. The following subsections detail our efforts to ensure reproducibility, the validity of our synthetic evaluation, the ethical and privacy implications of the technology, and our responsible use of language models.

6.1 Reproducibility

All experiments, models, and datasets used in this research are fully documented and made publicly available to ensure transparency and reproducibility. We provide detailed instructions and code repositories to allow others to replicate our results and build upon our work. As mentioned in the experimental setup, the code base is made publicly available.[1].

6.2 Validity of Synthetic Evaluation

The experiments are simulation-driven, its conclusions are bounded by the generative model rather than by physical hardware. As discussed in section 7, this forms a natural limitation on the performance of the localisation methods, as the models have a simulation bias.

6.3 Usage of LLMs

Large language models (LLMs) were used solely to assist with rewriting and improving the clarity of the manuscript, with all content being carefully reviewed. All technical content, experimental design, and analysis were developed independently to maintain the integrity of the research.

7 Discussion

The frameworks in subsection 3.1 are limited by the system model, in that it has only been built on a single

[1]The full source code, along with instructions, is available at: `https://github.com/praksatyam/screensense`

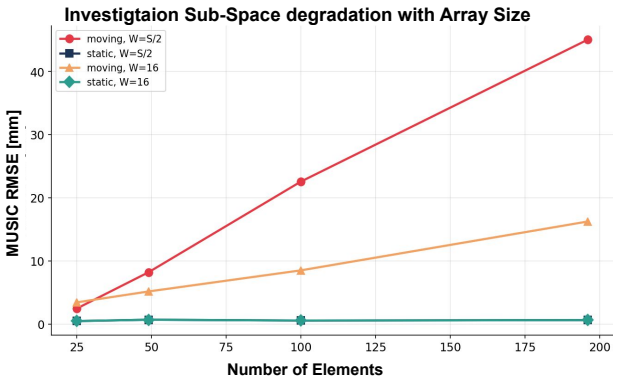

Figure 9: Isolating MUSIC's degradation with array size. With a static finger (squares, diamonds).

finger at a time, and is parameterized for a single type of object. Real interactions involve multiple contact zones, and the localisation models can be extended by implementing clustering algorithms, like K-Means. Isolating each cluster and running the inference from the algorithms could address this problem. Furthermore, generalizing the system model to other objects is also possible, using common and more free-physical parameters and using a similar method to MF-CNN will extend it.

The faster degradation of the subspace methods, in the array size experiment in subsection 5.4, is driven mainly by how they handle moving targets, and is amplified as the covariance window grows with the array. MUSIC separates a signal subspace from a noise subspace by eigen decomposing the spatial covariance, a separation that is only clean when the signal energy concentrates in as many dominant eigenvalues as there are sources; a stationary point source gives a rank-one covariance and a clear eigenvalue gap.[3] Two effects erode this as the array grows. The window holds $W = \lfloor S/2 \rfloor < S$ snapshots, so the covariance is rank-deficient and increasingly ill-conditioned with S, the regime in which the sample covariance is no longer a reliable estimate of the true one [15]. At the same time the finger moves during the window, so its response no longer occupies a single eigen-direction.

To identify which effect dominates, we repeat the array-size sweep for MUSIC under a 2×2 control: the finger is either static or moving, and the covariance window is either $W = \lfloor S/2 \rfloor$, which scales with the array, or fixed at $W = 16$. A fixed window removes the growth of the rank deficiency; a static finger removes the motion.

The static configurations remain at sub-millimetre error across every array size and are indistinguishable from one another, which shows that the rank-deficient covariance is not itself the problem: a stationary source is rank one whatever the window length. The error appears only once the finger moves, and its magnitude follows the window rather than the number of elements, the two moving curves cross exactly where their windows cross ($\lfloor S/2 \rfloor = 16$ near $S \approx 32$). A larger window spans a longer arc of the trajectory, spreading the deterministic response over more eigen-directions, and a larger aperture compounds this by resolving finer near-field cells [7]. More elements therefore cannot supply a signal subspace that a single moving source does not contain.

Synthetic data, using the same forward model, is used to generate the signal, and the matched filter uses the same model for prediction explaining the strong result in section 5. Although we perturb the physical parameters, to stress this match, the generator is shared between the data and the model-based estimators, which likely flatters their accuracy relative to real measurements where the model is only approximate. The learned methods are less exposed to this, as they are trained rather than analytically matched.

The framework has not been validated against a physical array or measured signals, and has not been run on resource-constrained computers. The required 28 GHz transparent-array hardware was unavailable. As a result the reported accuracies should be read as an optimistic bound that real measurements are needed to confirm.

8 Conclusion and Future Work

This work modeled the deterministic near-field backscatter of a finger hovering above a transparent 28 GHz antenna array and used it to generate synthetic time-series for 2D finger localisation, against which classical and learned estimators were compared. The backscattering response can be captured by a physically grounded deterministic model parametrized by the finger's dielectric properties and near-field geometry, and from the resulting time-series 2D position can be inferred to sub-millimetre accuracy. This accuracy is sustained across array sizes and down to low SNR only by the learned methods: MT-CNN is the strongest overall, pairing the lowest error with a low, grid-free latency, while MF-CNN is the most robust to physical variation. The classical subspace methods, by contrast, fail on this deterministic single-snapshot signal and worsen as the array grows. Within the simulated set-

ting considered in this work, learning-based localisation provides the most accurate and practical approach.

To extend this project further integration testing with real sensors, with resource constrained devices, and integrating with communication protocols is required. When using real sensors the accuracy levels are exptected to be magnitudes worse compared to computer-simulated environments. Before performing inference with resource-constrained devices the models must be quantized, to meet memory constraints. Communication protocols expected to be used with 6G protocols are expected to use OFDM, under wideband scenarios [2], and this is yet to be tested under this setup.

A Appendix A - Derivation Extension

A.1 Backscattering Factor

The *radar cross section* (RCS) σ_f quantifies the power scattered by the finger back toward the array, relative to the power incident on it. For a dielectric sphere with complex permittivity $\tilde{\varepsilon}$ in a uniform incident field, the induced polarisation $\mathbf{P}$ is given by the Clausius-Mossotti relation:

$$\mathbf{P} = \varepsilon_0 \frac{\tilde{\varepsilon} - 1}{\tilde{\varepsilon} + 2} \mathbf{E}_{inc}, \tag{14}$$

where the factor $(\tilde{\varepsilon}-1)/(\tilde{\varepsilon}+2)$ is the Clausius-Mossotti polarisability. The $+2$ in the denominator arises from the *depolarisation field* — the dipole moment induced in the sphere generates its own internal field that partially opposes the applied field, a self-consistent effect. In the Rayleigh scattering limit (object size $\ll \lambda$), the scattered power is proportional to $|\mathbf{P}|^2$, giving:

$$\boxed{\sigma_f(\tilde{\varepsilon}) = \sigma_0 \left| \frac{\tilde{\varepsilon} - 1}{\tilde{\varepsilon} + 2} \right|^2} \tag{15}$$

Substituting $\tilde{\varepsilon} = 17.50 - j9.63$ and setting $\sigma_0 = 1$:

$$\frac{\tilde{\varepsilon} - 1}{\tilde{\varepsilon} + 2} = \frac{16.50 - j9.63}{19.50 - j9.63}, \tag{16}$$

$$\sigma_f = \left| \frac{16.50 - j9.63}{19.50 - j9.63} \right|^2 = 0.772. \tag{17}$$

A.2 Complex Permittivity

The finger is modelled as a lossy dielectric with frequency-dependent electromagnetic properties. At 28 GHz, biological tissue (skin) is described by the Gabriel et al. dielectric model [5]:

$$\tilde{\varepsilon} = \varepsilon_r - j \frac{\sigma_t}{\omega \varepsilon_0}, \tag{18}$$

where $\varepsilon_r = 17.5$ is the relative permittivity (energy storage) and $\sigma_t = 15.0\,\mathrm{S\,m^{-1}}$ is the conductivity (energy absorption). Substituting at $f_0 = 28\,\mathrm{GHz}$:

$$\begin{aligned} \tilde{\varepsilon} &= 17.50 - j \frac{15.0}{2\pi \times 28 \times 10^9 \times 8.854 \times 10^{-12}} \\ &= 17.50 - j9.63 \end{aligned} \tag{19}$$

Physical interpretation.

- $\mathrm{Re}(\tilde{\varepsilon}) = 17.5$: the tissue stores electromagnetic energy and the material's polarisation partially follows the oscillating field.
- $\mathrm{Im}(\tilde{\varepsilon}) = -9.63$: the tissue absorbs electromagnetic energy and current flows that dissipate power as heat. This is responsible for the rapid attenuation of mmWave signals in tissue.
- Loss tangent $\tan\delta = |\mathrm{Im}(\tilde{\varepsilon})|/\,\mathrm{Re}(\tilde{\varepsilon}) = 0.55$: the tissue is highly lossy and the absorbed power is more than half the stored power per cycle.

A.3 Complex Refractive Index

The complex refractive index governs how EM waves propagate *inside* the tissue:

$$\tilde{n} = \sqrt{\tilde{\varepsilon}} = n_r - jn_i = 4.329 - j1.112, \tag{20}$$

where:

- $n_r = \mathrm{Re}(\tilde{n}) = 4.329$: the wave's phase velocity inside tissue is c/n_r, the wave slows down and accumulates extra phase relative to free space.
- $n_i = |\mathrm{Im}(\tilde{n})| = 1.112$: the wave's amplitude decays exponentially inside tissue.

A.4 Skin Depth

The skin depth δ is the depth at which the wave amplitude decays to $1/e$ of its surface value:

$$\delta = \frac{c}{\omega\, n_i} = \frac{3\times 10^8}{2\pi \times 28\times 10^9 \times 1.112} = 1.53\,\mathrm{mm}. \tag{21}$$

Reasoning for using $d_f = \delta$. The finger is physically 15 mm thick. However, the EM wave does not penetrate the full finger, it is substantially absorbed within the skin depth. Using the full finger thickness $d_f = 15$ mm would give $\beta_{\min} = e^{-2kn_i\cdot 0.015} \approx 0$, implying total signal blocking, which is physically incorrect for a wave that grazes the surface. Using the skin depth $d_f = \delta = 1.53$ mm as the effective interaction depth correctly models the wave interacting only with the surface layer of tissue before being absorbed or scattered. This gives:

$$\begin{aligned}\beta_{\min} = e^{-2kn_i\delta} &= e^{-2\times 587\times 1.112\times 0.00153}\\ &= e^{-2.0} = 0.135.\end{aligned} \tag{22}$$

A.5 Phase Shift

The wave passing through tissue of refractive index n_r accumulates additional phase compared to free-space propagation. Over the effective interaction depth $d_f = \delta$:

$$\begin{aligned}\phi_{\max} &= k(n_r - 1)\,\delta\\ &= 587\times(4.329-1)\times 0.00153\\ &= 2.99\ \mathrm{rad} = 171.5^\circ.\end{aligned} \tag{23}$$

The -1 subtracts the free-space phase that would have accumulated without the finger. The per-element phase shift, weighted by the footprint, is:

$$\phi_s(t) = \rho_s(t)\,\phi_{\max}. \tag{24}$$

A.6 Amplitude Attenuation

The attenuation at element s depends on how much of the finger overlaps with it, interpolated between no finger ($\beta = 1$) and maximum finger coverage ($\beta = \beta_{\min}$):

$$\beta_s(t) = 1 - \rho_s(t)\big(1 - \beta_{\min}\big), \tag{25}$$

where $\beta_{\min} = e^{-2kn_i\delta} = 0.135$ (from equation 22). When $\rho_s = 0$ (no finger), $\beta_s = 1$ (no attenuation). When $\rho_s = 1$ (fully under finger), $\beta_s = \beta_{\min}$ (maximum attenuation, retaining 13.5% of amplitude).

B Appendix B - Physical Parameter Perturbation Data

All values follow the same setup as defined in Table 4, and physical parameters are changed to extremes, to show the impact on the array response.

B.1 Permittivity

Permittivity is changed to shows array response at 6 and 30. Ceteris Paribus.

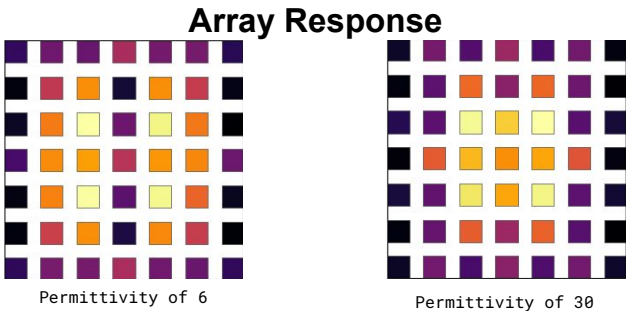

Figure 10: Visualization of the permittivity impact on array response

B.2 Skin Depth

Skin Depth is changed to shows skin depth impact of 3mm to 4.5mm depth. Ceteris Paribus.

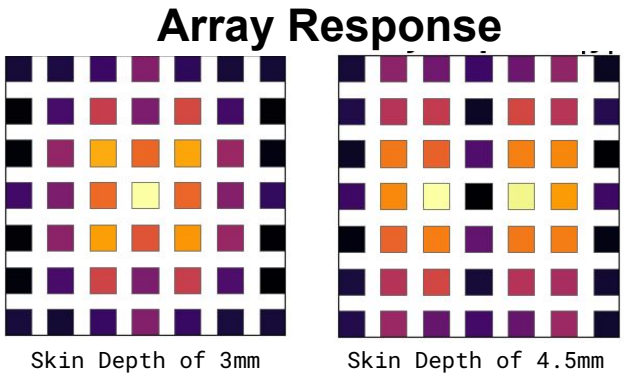

Figure 11: Visualization of the skin depth impact on array response

The position of the finger is kept stationary in the center of the array, the MF-implementation shows an oscillating behavior.

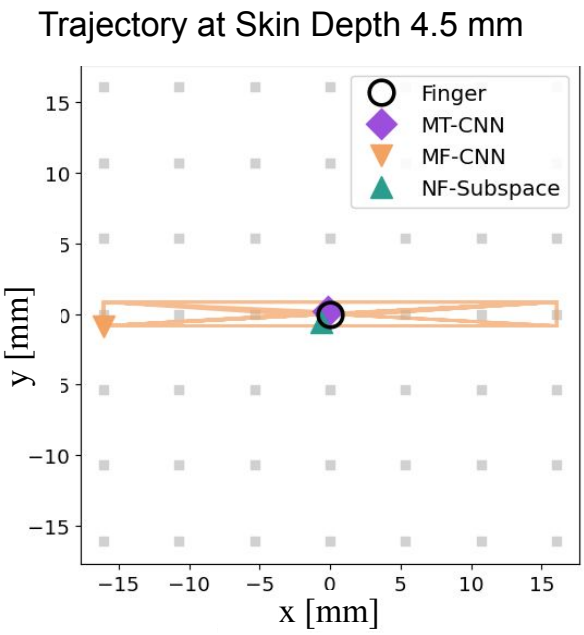

Figure 12: Oscillating behavior shown by MF implementations.

B.3 Hover Height

Hovering height of finger is changed from 6.525 mm to 18.35 mm. Ceteris Paribus.

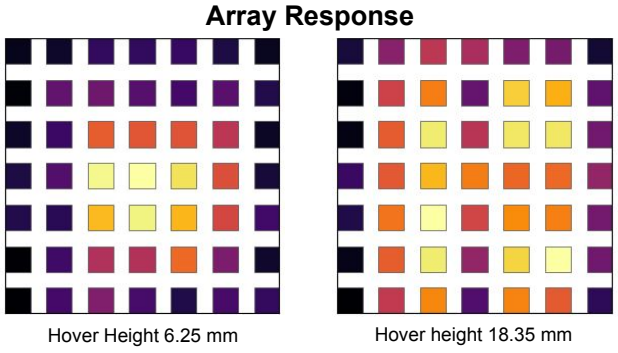

Figure 13: Visualization of the hovering height impact on array response

References

[1] Navneet Agrawal, Ehsan Tohidi, Renato L. G. Cavalcante, and Sławomir Stańczak. Towards bridging the gap between near and far-field characterizations of the wireless channel. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2024.

[2] Emil Bjornson, Ferdi Kara, Nikolaos Kolomvakis, Alva Kosasih, Parisa Ramezani, and Murat Babek Salman. Enabling 6g performance in the upper mid-band by transitioning from massive to gigantic mimo. *IEEE ComSoc*, 2025.

[3] A. Bruce Carlson and Paul B. Crilly. *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*. McGraw-Hill, 5th edition, 2010.

[4] Hsi-Tseng Chou, Wen-Jin Gao, Jianhua Zhou, Baiqiang You, and Xian-Hui He. Enhancing electromagnetic backscattering responses for target detection in the near zone of near-field-focused phased array antennas. *IEEE*, 2020.

[5] S. Gabriel, R. W. Lau, and C. Gabriel. The dielectric properties of biological tissues: Iii. parametric models for the dielectric spectrum of tissues. *King's College*, 1996.

[6] Arad Gast, Luc Le Magoarou, and Nir Shlezinger. Near field localization via ai-aided subspace methods. *IEEE*, 2026.

[7] Alex B. Gershman, Ljubiša Stanković, and Vladimir Katkovnik. Sensor array signal tracking using a data-driven window approach. *Signal Processing*, 80(12):2507–2515, 2000.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[9] Joseph W. Goodman. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

[10] Wonbin Hong, Jaehyun Choi, Dongpil Park, Myun soo Kim, Chisang You, Doochan Jung, and Junho Park. mmwave 5g nr cellular handset prototype featuring optically invisible beamforming antenna-on-display. *IEEE Communications Magazine*, 58(8):54–60, 2020.

[11] Huan-Chu Huang, Jie Wu, Shuang Cui, and Dua-Chyrh Chang. Antenna-on-display (aod) for wireless mobile devices: Retrospect and prospect. *PIER (Progress in Electronic Research)*, 2025.

[12] Alva Kosasih, Özlem Tuğfe Demir, Nikolaos Kolomvakis, and Emil Bjornson. Spatial frequencies and degrees of freedom: Their roles in near-field communications. *IEEE Signal Processing Magazine*, 2025.

[13] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM*, 2016.

[14] Fan Liu, Yuanhao Cui, Christos Masouros, Jie Xu, Tony Xiao Han, Yonina C. Eldar, and Stefano Buzzi. Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. *IEEE J. Sel. Areas Commun.*, 40(6):1728–1767, 2022.

[15] Xavier Mestre and Miguel Ángel Lagunas. Modified subspace algorithms for DoA estimation with large arrays. *IEEE Transactions on Signal Processing*, 2008.

[16] Junho Park, Seung Yoon Lee, Jongmin Kim, Dongpil Park, Woo Choi, and Wonbin Hong. An optically invisible antenna-on-display concept for millimeter-wave 5g cellular devices. *IEEE Transactions on Antennas and Propagation*, 2019.

[17] Sebastian Paul, Fabian Schwartau, Markus Krueckumier, Reinhard Caspary, Carsten Monka-Ewe, Jeorg Schoebel, and Wolfgana Kowalsky. A systematic comparison of near-field beamforming and fourier-based backward-wave holographic imaging. *IEEE Antenna and Propagation*, 2021.

[18] Dor H. Shmuel, Julian P. Merkofer, Guy Revach, Ruud J. G. van Sloun, and Nir Shlezinger. Subspacenet: Deep learning-aided subspace methods for doa estimation. *IEEE*, 2024.

[19] George L. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):311–329, 1960.

[20] Ting Wu, Theodore S. Rappaport, and Christopher M. Collins. The human body and millimeter-wave wireless communication systems: Interactions and implications. *IEEE*, 2015.

[21] Shun Zhuge and Qing Wang. Screenant: Transparent on-screen antennas for 6g. *IEEE*, 2025.