ARGOS: Leveraging Visual Priors for Scalable Wireless Navigation in Dynamic Environments

Arko Datta†, Tharaneeshwaran V U‡, Aravindh Sriram Kumar A G†, Ayon Chakraborty†
†Sensing and Networked Systems Engineering (SeNSE) Lab, IIT Madras
‡National Institute of Technology Puducherry, India

***Abstract*—Performance of wireless navigation systems degrade sharply in industrial environments dominated by metallic clutter and heavy multipath. The primary cause lies in how anchors are placed or selected. Methods that succeed in open spaces fail under severe non-line-of-sight (nLoS) and frequent layout changes, for instance, induced by moving forklifts or shifting shelves. We introduce Argos, a multimodal wireless digital twin that fuses visual and RF information to optimize anchor selection. Visual imagery reconstructs the 3D layout, updated continuously via existing surveillance cameras, while RF measurements capture material-specific attenuation and reflections. We show that layout priors alone are insufficient; combining them with the required RF optics yields a material-aware channel model that predicts range errors under severe nLoS. Argos adapts proactively to environmental changes, without the requirement of repeated RF calibration or retraining, sustaining sub-meter localization accuracy in dynamic scenes. We validate Argos in a $120\,m^2$ factory testbed spanning over 700 locations, where the digital twin is built from $5K$ RGB images and $0.4M$ UWB CIR samples. To our knowledge, this is the first system to exploit visual priors for adaptive orchestration of wireless navigation infrastructure.**

***Index Terms*—Wireless Digital Twin, Ultra Wideband, Positioning and Navigation, 3D reconstruction**

I. Introduction

State-of-the-art industrial environments such as factories, warehouses and construction sites rely on robust navigation infrastructures to ensure safety [1], enable continuous monitoring and support efficient operations [2]. Such systems must not only achieve high positioning accuracy [3] at scale, but also remain resilient to occlusions and adapt to dynamic, evolving layouts [4]. Radio Frequency (RF)-based navigation [5], [6] offers a lightweight and flexible alternative, requiring less computation, scaling to larger areas at lower deployment cost and being far less impacted by strict visual line-of-sight (LoS) compared to their vision-based counterparts.

Anchor Selection as a Primitive. In practice, RF navigation relies on fixed anchors with which tracked nodes perform ranging to estimate their respective distances [7]. For reliable localization, two conditions must be met. First, the ranging overhead must be minimized to keep update latency low, making it essential to restrict measurements to a selective subset of anchors. Second, the chosen anchors must yield consistently accurate range estimates. Non-line-of-sight (nLoS) effects caused by clutter or layout changes often corrupt such estimates [8]. Hence, not all anchors contribute equally to localization accuracy. Moreover, in large and dynamic scenes, both coverage and ranging quality evolve continuously. A static anchor choice is therefore ineffective - anchors that are initially reliable may degrade considerably in performance as surrounding layout changes. Anchor selection emerges as a *core localization primitive* [9], and refining it is essential for robust RF navigation in dynamic environments.

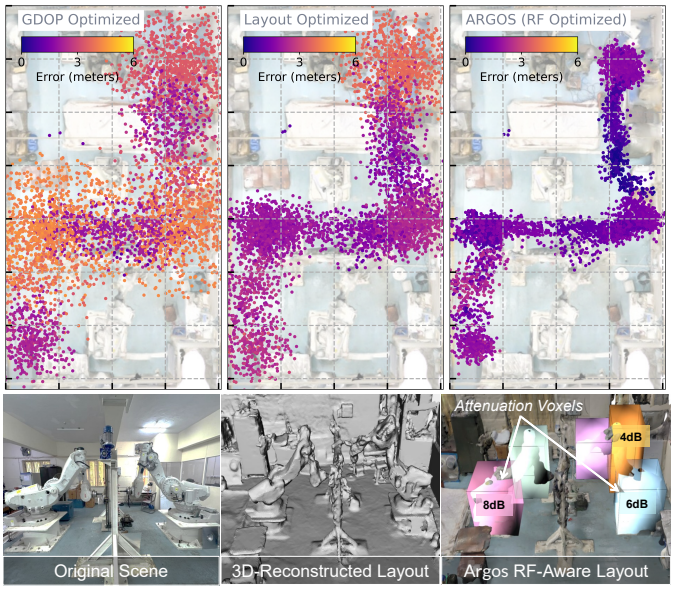

Fig. 1: A $\approx 120\,m^2$ section of a factory floor with dense metallic clutter and heavy equipment. Five UWB anchors are deployed to evaluate anchor deployment strategies under severe multipath and nLoS. *Top-Left*: GDoP-only deployment, ignoring structural details. *Top-Middle*: Incorporating 3D layout maps improves LoS coverage, reducing median error by over 5 m. *Top-Right*: Layout alone ignores material-specific attenuation; Argos adds material-aware priors for a further 2–3 m gain, achieving sub-meter localization even under extreme NLoS (bottom-right shows segmentation into attenuation voxels).

In this paper, we present Argos, an RF (UWB[1]) anchor selection framework designed to handle the dynamics, extreme nLoS and mobile clutter, typical of industrial environments. To motivate our approach, we highlight the key factors that govern anchor placement and, in turn, determine localization accuracy (see fig. 1). Most existing systems, including commercial ones [11], optimize anchor placement by minimizing geometric dilution of precision (GDoP [12]). While effective in open spaces, this fails completely in cluttered or nLoS-heavy environments. Extensions that integrate structural geometry of the environment improve LoS coverage [13], but layout alone

[1]We use Ultra-Wideband (UWB) as the reference RF system, given its balanced trade-off between range resolution and nLoS penetration. However, our formulation is generic and can extend to other technologies such as Wi-Fi FTM [10], provided the available bandwidth is sufficient to capture the channel's multipath characteristics with adequate granularity.

is insufficient. For instance, a large drywall may present a significant visual nLoS but remain largely transparent to RF, whereas a small metallic forklift can severely distort wireless multipath and disrupt ranging. This highlights the importance of *RF optics*: knowledge of material-dependent attenuation and reflections allows anchor selection strategies that can achieve sub-meter localization even under harsh nLoS conditions.

While static factors such as geometry and material properties are important, the greater challenge lies in dynamics. Anchors that are reliable at one moment may lose effectiveness instantly as, say, forklifts move or shelves shift, introducing new nLoS constraints where even a single obstruction can disrupt multiple links. Coping with this would require continuous knowledge of both layout changes and wireless channel states - information that is extremely challenging to obtain in real time. Frequent RF calibration to track such dynamics is equally impractical. The key challenge, therefore, is to maintain accurate channel knowledge and adapt anchor deployment at scale, without incurring unsustainable calibration overhead.

Problem Statement. With respect to the challenges discussed above, we consider the problem of robust anchor selection for RF-based localization in dynamic, cluttered environments. Specifically: *(a) Given synchronized multi-view visual observations along with RF measurements, can we estimate a digital twin that inherently embeds RF optics? (b) Can ray-traced channel predictions be used to select an anchor subset, under latency and coverage constraints, that minimizes a node's expected ranging error? (c) Can we enable fast, incremental updates as the scene evolves?*

Our Approach. Argos combines visual priors with RF data to build a wireless digital twin that adapts to environmental dynamics. Visual imagery is used for 3D reconstruction to capture the layout. These reconstructions can be updated at scale as changes are registered by existing infrastructure, e.g., through surveillance cameras. RF measurements, taken jointly with such visual imagery, are then used to estimate the RF attenuation, which are critical for accurately modeling wireless multipath. In the following, we elaborate on the core capabilities that enable Argos to realize such a wireless digital twin.

Channel Synthesis with Visual Priors. The first stage of Argos is to construct a digital twin that couples scene geometry with RF-relevant properties. Pure RF-based reconstruction is impractical at scale, as per-voxel tomography demands prohibitive measurements. Instead, Argos exploits visual priors to recover the 3D layout, reducing the problem size by orders of magnitude. Sparse RF measurements then enrich this layout with material-dependent attenuation and reflections, yielding an RF-aware representation of the environment. This twin enables accurate synthesis of channel impulse responses (CIRs) for any anchor–receiver pair, forming a practical basis for analyzing ranging errors and guiding anchor selection.

Modeling Range Error. With the digital twin calibrated, Argos models how *Time-of-Arrival* (ToA) estimation degrades under nLoS. In such conditions, the direct path is often weakened or obscured by stronger reflections, leading detectors to misidentify a multipath component as the first arrival path. To quantify this effect, we introduce the *expected first-path misdetection* (Δ_{FMD}) as our metric for ranging reliability. By generating CIRs for any anchor–receiver link, the twin predicts the spatial distribution of Δ_{FMD}, transforming it into not only a channel predictor but also an error predictor, proactively anticipating where anchors are likely to misreport ranges.

Adapting to Dynamics. Argos avoids recomputing the entire reconstruction whenever the environment changes. Instead, it segments the scene into attenuation-aware voxels linked to objects or structures. When objects move, lightweight visual updates shift the corresponding voxels within the digital twin, while only the affected multipath rays are re-traced. This incremental process keeps the twin aligned with the evolving layout, continuously refining attenuation parameters without costly RF recalibration. As a result, Argos remains accurate and adaptive, turning scene dynamics into an opportunity for proactive anchor selection rather than a disruption.

To the best of our knowledge, this is the first work to leverage visual priors for improving wireless localization infrastructure. We make the following contributions:

- We introduce Argos, a multimodal wireless digital twin that fuses jointly acquired visual and RF information to optimize anchor selection for navigation in cluttered industrial environments. By modeling both scene geometry and RF attenuation, Argos achieves sub-meter localization even under severe nLoS conditions.
- We demonstrate that Argos proactively adapts to environmental dynamics, updating within 5 s without repeated RF recalibration or model retraining. Visual priors confine updates to affected regions, enabling fast, selective adaptation that maintains accuracy and resilience in real time.
- We validate Argos in real industrial environments with heavy clutter, covering over 700 discrete locations across $120\,m^2$. The digital twin is constructed from $5K$ RGB images and $0.4\,M$ UWB CIR samples collected on-site.
- *Artifact Contributions.* We have open-sourced the data and code artifacts related to Argos implementation, available via https://sense.cse.iitm.ac.in/argos/.

II. Background and Challenges

A. Related Works

Anchor Deployment. Early strategies placed anchors based on arena geometry, aiming to minimize multilateration uncertainty through metrics such as GDoP [12]. These methods work in open spaces but quickly degrade under nLoS. Layout-based optimizations extended this idea by maximizing LoS coverage for a given scene, while meta-heuristic algorithms such as particle swarm optimization (PSO) and genetic algorithms (GA) [14], [15] were used to solve the combinatorial anchor placement problem. Some works further modeled range variance from multipath profiles and derived Cramer-Rao bounds [16] to estimate limits on localization accuracy.

Channel Modeling. Beyond anchor placement, data-driven models have attempted predicting channel quality from sparse measurements. Early interpolation methods such as IDW [17] and Kriging [18], later extended with deep variants [19], worked well in static settings. Recent neural approaches like U-Net [20], attention models [21] and generative schemes [22], [23], [24], [25] produced richer maps but remained data-hungry and brittle to new layouts. Efforts in radio tomographic imaging (RTI) reconstructed spatial attenuation maps from multi-link measurements. Early RSSI-based RTI gave only coarse fields [26], [27], while CIR-based methods sharpened detail by resolving multipath [28], [29], [30]. Such advances point toward 3D layout as a strong prior for channel synthesis.

Channel Synthesis with Visual Priors. Layout has emerged as a strong prior for channel prediction. Ray-tracing tools like Sionna RT [31] can synthesize channels from geometry, but only when material properties are known. Neural surrogates such as NeRF2 [13], NeWRF [32], or WiNeRT [33], [34] improve fidelity. However, in dynamic configurations, they need retraining, making the models heavy and difficult to adapt under dynamics. Meanwhile hybrids like RFCanvas [35] make layouts editable yet still ignore RF properties such as attenuation and reflection.

Recent Gaussian splatting methods [36] offer faster, real-time rendering compared to NeRF. However, the resultant reconstruction is a singular mesh without any segmentation assigned to individual objects, thereby lacking semantic capabilities. Visual pipelines like SfM + MVS [37], NeRF [38], Gaussian splatting or segmentation methods like SA3D [39] and SAGD [40] provide accurate layouts and semantics, but have not been exploited for large-scale RTI or for material-aware reconstructions that adapt in dynamic scenes.

B. Revisiting RF Anchor Selection Strategy

Accurate localization critically depends on the careful selection of anchors, since anchors positioned under occlusion not only reduce spatial coverage but also degrade the quality of ranging measurements. A seemingly straightforward solution is to deploy a uniform, dense grid of anchors. However, this strategy fails to scale, as the resulting ranging overhead grows disproportionately with the number of devices and anchors.

Scalability Issues. In Fig. 2, we demonstrate that simply increasing the number of anchors in a ToA-based localization setup does not guarantee improved accuracy. With twelve anchors uniformly deployed in a grid, Fig.2 (*left*) highlights the multilateration error at a single location when ranging to different anchor subsets. The error decreases as the number of anchors increases beyond five, but then worsens sharply when more anchors are included. This demonstrates how erroneous ranges can dominate multilateration and motivates the need for robust anchor selection to ensure good coverage. Second, more anchors directly increase ranging latency (≈25 ms per range) and reduce the feasible update rate. For a fixed update rate, fewer anchors allow each tag to collect more stable ranges, improving accuracy. This effect is also highlighted in

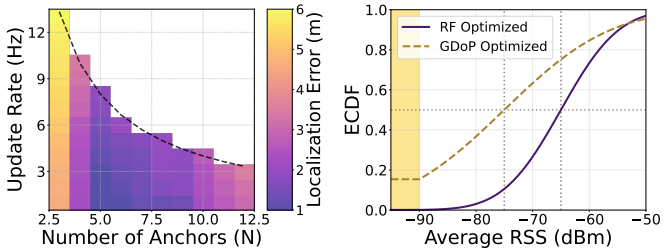

Fig. 2: *Left*: Adding anchors initially improves accuracy, but random selection with too many anchors degrades performance while also lowering the maximum update rate. *Right*: With 5 anchors, the best placement yields ≈10 dB higher median SNR and full coverage, versus ≈80% in the average case. Links below −90 dBm are treated as outages for reliable ranging.

fig. 2(*left*). Third, we record signal strength (RSS) measurements for each anchor in our testbed and report the CDF across two types of anchor configurations - GDoP optimized and the best configuration showcasing a scope for improvement. Fig. 2(*right*) uses five anchors, which provides a reasonable tradeoff between coverage, accuracy and latency. Locations with average RSS below −90 dBm are marked *out of coverage*. The best anchor configuration achieves about 10 dB higher median RSS and full coverage, whereas GDoP optimized anchor placement provides only ≈80% coverage.

TDoA alleviates some latency since anchors broadcast simultaneously, but suffers from synchronization overhead and amplified error propagation. In our experiments, using identical anchor configurations and coverage, TDoA was 2–2.5 m less accurate compared ToA, particularly in zones with heavy nLoS or poor RSS. This is because hyperbolic constraints amplify range-difference errors and synchronization offsets. Relying on wireless synchronization in our cluttered testbed made the problem worse: heavy NLoS often blocked synchronization frames, causing spurious updates and even system failures.

C. Anatomy of a Range Error

In ToA-based localization, the accuracy of the estimated range critically depends on how precisely the signal arrival time is extracted from the receiver's Channel Impulse Response (CIR). The CIR is represented as a discrete sequence of amplitude taps $\{A_i\}, i = 0, 1, \ldots, K-1$, where each tap denotes the channel gain contributed by either the direct path (LoS) or a multipath reflection (nLoS). The taps are sampled at intervals of $\Delta\tau$, determined by the signal bandwidth, which sets the temporal resolution of the CIR. Note that the signal bandwidth sets a fundamental quantization limit on path-delay estimation, and thus on ranging accuracy.

A leading-edge detector (LDE) identifies the earliest CIR tap above an amplitude threshold[41] Θ, i.e., $\hat{j} = \min\{i : A_i > \Theta\}$, with Θ tuned to the observed noise floor. Under ideal LoS, $\hat{j}$ matches the true first path j_{FP}. However, in cluttered environments the attenuated $A_{j_{\text{FP}}}$ may be missed or buried in noise, causing the detector to incorrectly lock onto a stronger reflected component $A_{\hat{j}}$ with $\hat{j} > j_{\text{FP}}$. We define this index offset as the *First-Path Misprediction Degree*,

$$\Delta_{\text{FMD}} = |\hat{j} - j_{\text{FP}}| = |\text{LDE}_{\Theta}(\text{CIR}) - j_{\text{FP}}| \qquad (1)$$

$\mathrm{LDE}_\Theta(\bullet)$ denotes the first-path index predicted by the LDE for a threshold Θ. A nonzero Δ_{FMD} directly translates into a bias in the ToA estimate that further scales with the tap resolution $\Delta\tau$. In protocols such as two-way ranging (TWR) [41], such bias compounds across the 3–4 ToA exchanges that constitute the round-trip measurement. Importantly, Δ_{FMD} is *not* symmetric across a link, for instance, its value is often larger when clutter is concentrated near the receiver, so these effects do not fully cancel out in ToA-based exchanges. In fading or heavy nLoS environments, weak direct paths are frequently overshadowed by stronger reflections [42], leading to systematically large Δ_{FMD} values (see, fig. 4 for more details). Fig. 3(*left*) illustrates this phenomenon, where the true first path A_{FP} and the estimated first path $\hat{A}_{\mathrm{FP}}$ are marked. The amplitude distributions of these taps overlap significantly, explaining why the LDE often defers detection to a delayed path, mispredicting the first path. Fig. 3(*right*) shows the ROC curves for the LDE, highlighting the trade-off between correct detection and false alarms as Θ varies. Under moderate NLoS, the detector still retains reasonable discriminability, but under heavy nLoS with poor SNR, the ROC curve collapses, indicating frequent failures and large Δ_{FMD}, which in turn translates into severe ranging bias.

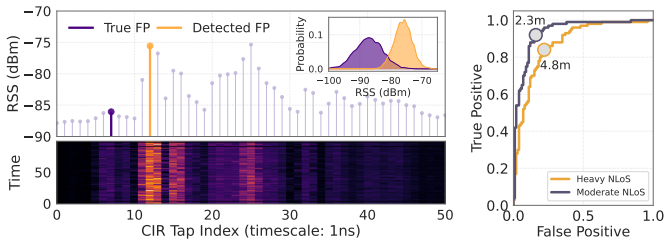

Fig. 3: *Left*: Sample UWB CIR showing the true first path A_{FP} and the detected path $\hat{A}_{\mathrm{FP}}$. Overlapping amplitudes often cause the detector to lock onto a later multipath component, introducing nonzero Δ_{FMD}. *Right*: ROC curves of first-path detection for varying thresholds Θ under moderate and heavy NLoS. Severe NLoS shifts the curve, increasing misprediction probability and bias.

D. How Accurately Can We Estimate Range Errors?

Proactive prediction of range errors is critical for optimizing anchor placement and ensuring reliable localization. Unlike reactive designs or traditional GDoP-based strategies that ignore layout and signal quality, our focus is on *effective coverage*. As illustrated in Fig. 4, Δ_{FMD} remains low when RSS exceeds -60 dBm, but increases sharply below this threshold, indicating reduced accuracy and coverage. In the following, we identify key challenges that guide our design of a proactive, channel-aware localization system.

■ **Challenge 1. Reliable Estimation of the UWB Channel.** The key challenge in range-based localization is predicting whether the true ToA or first path in the CIR is detectable or masked by noise and reflections. This requires modeling multipath and material attenuation. Purely geometric layouts are insufficient, as similar structures can yield very different channels depending on surrounding material properties. For example, a drywall may introduce only mild attenuation, whereas a metallic shelf of similar dimensions can dominate multipath and severely bias the first-path estimates (see fig. 4 (*right*)).

Approach. Argos fuses 3D visual reconstructions with sparse RF data to build an attenuation-aware *wireless digital twin*. Instead of photorealism, the reconstruction targets just enough structural and material detail to predict Δ_{FMD} accurately. The key challenge is choosing the right granularity of reconstruction. Too much detail adds computational overhead with little gain, while too little fails to capture multipath reliably. The goal is an on-demand, RF-aware channel map that enables reliable first-path detection and ToA estimation for an anchor.

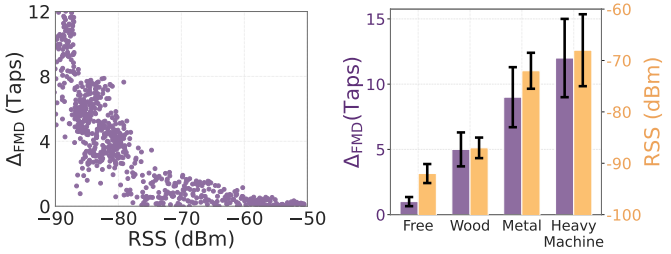

Fig. 4: *Left:* Δ_{FMD} has a strong dependence on RSS, beyond -60 dBm range errors are negligible. *Right:* For a fixed anchor–tag distance, material properties heavily influence propagation loss as well as Δ_{FMD}. For metal and heavy machinery blockages, the first path is non-existent due to complete blockage of LoS path.

■ **Challenge 2. Predicting Δ_{FMD} from Channel Maps.** Once an RF-aware channel map is available, the next challenge is to predict how it influences range estimation. While prior work has largely focused on estimating coverage or classifying CIRs as LoS or NLoS (often using supervised learning [8]), such coarse labels fail to explain how range errors manifest. What truly matters is not just whether a path exists, but whether the direct path is detectable and distinguishable from potentially stronger reflected components. To select anchors robustly, we must estimate Δ_{FMD} quantitatively across space for each anchor–receiver pair.

Approach. In Argos, we model the full CIR profile based on the predicted channel map, capturing both geometric multipath and stochastic fading. By simulating CIR amplitude distributions and applying LDE logic, we estimate the probability that the first path falls below threshold and is preempted by a later, stronger tap, yielding a non-zero Δ_{FMD}.

■ **Challenge 3. Scaling with Layout Dynamics.** In realistic deployments, the RF environment is rarely static. For instance, forklifts move, pallets and shelves are rearranged and temporary obstructions appear or disappear. These dynamics continuously alter coverage and channel maps, making explicit RF recalibration or full 3D reconstructions impractical in real time. The challenge is scalability: keeping the digital twin updated without reprocessing the entire layout.

Approach. Instead of repeatedly reconstructing the full 3D scene, we model the environment as discrete, movable objects with associated RF attenuation profiles. Position changes, detected via existing surveillance, trigger updates only to

affected regions of the digital twin. With known or estimated attenuation, ray-based coverage can be recomputed efficiently, enabling lightweight updates while maintaining accurate RSS and Δ_{FMD} estimates over time.

III. Design of The Argos System

At its core, Argos builds an RF attenuation-aware digital twin to anticipate channel behavior at scale. Since exhaustive RF-*only* calibration is infeasible in large, evolving scenes, Argos adopts a multimodal approach. A visual pipeline reconstructs a voxelated 3D model of the scene from RGB imagery, which is then segmented to isolate large objects that affect the wireless channel. Next, this voxel map is refined by a tomographic imaging module that assigns voxel-wise attenuation coefficients, using the RF inputs. With object-level attenuation captured, Argos employs ray tracing to predict multipath characteristics for candidate anchor-receiver links and translates these into errors in ToA estimates or Δ_{FMD}. Such translation enables proactive assessment of localization accuracy and guides anchor selection. Finally, to maintain robustness as the scene evolves, Argos avoids repeated full volumetric reconstructions. Instead, it leverages existing surveillance streams to detect object movements and updates the digital twin incrementally: moved objects are translated to new voxel positions and their attenuation properties reapplied. This lightweight update mechanism keeps channel estimates accurate at scale without the need for expensive recalibration. These lightweight updates ensure scalability while keeping channel estimates accurate without costly recalibration. The overall schematic of Argos is shown in Fig. 5, and the following subsections expand on each component in detail.

A. Scaling Channel Synthesis with Visual Priors

In our setting, RF-only tomography is infeasible: even moderate voxel resolutions yield millions of attenuation unknowns, and the measurement volume needed to solve them is prohibitive. Argos overcomes this by leveraging visual priors to reconstruct the scene layout, cutting the dimensionality of the tomography problem by nearly four orders of magnitude and making RF-aware reconstruction tractable at scale. The first stage of Argos is to recover the physical layout of the scene from joint $\{\mathsf{Visual}, \mathsf{RF}\}$ captures. At each capture location we record an RGB image together with CIR spectrograms from multiple anchors. We assume that captures during the bootstrapping phase are of higher fidelity and collected in a more orchestrated or controlled manner than those obtained during runtime deployment.

■ **3D Layout Reconstruction.** The visual stream is passed onto the 3D reconstruction pipeline. We begin with COLMAP [43], a widely used structure-from-motion (SfM) tool. Given multiple overlapping images of the scene, SfM estimates the position and orientation of each corresponding camera location by tracking how visual features move across frames. These camera locations are also used to spatially tag the corresponding CIR spectrograms, linking visual viewpoints with RF measurements for subsequent RTI processing. Further it outputs a sparse 3D point cloud that captures the rough geometry of the scene but too crude for estimating our channel models. To enrich this geometry, we run COLMAP's Multi-View Stereo (MVS) module. MVS infers pixel-level depth from multiple views and fuses the results into a dense point cloud, providing a more complete reconstruction of scene surfaces. However, this cloud is still just a collection of unconnected samples and can be noisy or incomplete in regions with poor texture or occlusions. To make the dense points continuous and efficient to render, we use *Gaussian Splatting*, a recent technique that represents the scene as 3D Gaussian ellipsoids instead of millions of discrete points. Each Gaussian encodes position, orientation and when 'splatted' onto the image plane they blend into smooth renderings. We adopt *MVSGaussian* (MVSG), a recent variant of 3DGS [36] tailored to operate directly on MVS outputs.

■ **Semantic Segmentation and Voxelization.** The multi-view point cloud obtained from SfM reconstruction is first segmented at the object level. Classical clustering methods such as RANSAC or DBSCAN were tested but proved unreliable in cluttered industrial layouts. Instead, we adopt SAGD, a recent boundary-enhanced extension of the Segment Anything Model (SAM). SAGD begins by generating 2D semantic masks from SAM in each camera view, propagates them across multiple views and enforces consistency to stitch a complete 3D segmentation. This results in a semantically segmented 3D scene with objects collected in the set $\mathcal{O}$, where each object is uniquely identified by $o_i \in \mathcal{O}$.

To apply RTI, the reconstructed scene is voxelized into a global set $\{v_g\}$, partitioned into free voxels $\{v_g^{free}\}$ and object voxels grouped by segmented objects $o_i \in \mathcal{O}$, where each object is represented by $\{v_{g_j}^{o_i}\}$ with g_j and o_i denoting voxel and object identifiers, respectively. We introduce two optimizations to shrink the voxel-wise RTI unknowns. First, all free-space voxels share a single average attenuation coefficient α_{free}, collapsing $\{v_g^{free}\}$ into a single variable. Second, voxels of the same object o_i are tied under common material parameters, for instance, attenuation α_{o_i} and boundary/reflection β_{o_i}. This reduces the dimensionality from per-voxel variables to $\{\alpha_{free}\} \cup \{\alpha_{o_i}, \beta_{o_i}\}_{o_i \in \mathcal{O}}$, and naturally accommodates dynamics, since moving an object only reassigns its voxel set without increasing the parameter count. A surface mesh is generated per object for ray tracing, while the voxel grid remains the computational backbone for tomography and incremental updates.

Voxel Resolution. In practice, voxel resolution is chosen by balancing fidelity and tractability: it must be fine enough to match the spatial resolution implied by CIR taps (centimeter scale at sub-nanosecond bandwidths) yet coarse enough to keep the number of unknowns feasible. This layout-aware representation thus provides the crucial bridge to the RTI stage, where attenuation values are solved at the object level rather than per-voxel, enabling both scalability and adaptability.

■ **Making the Layout RF-Aware.** The voxel map from visual reconstruction encodes geometry but lacks essential RF-optical

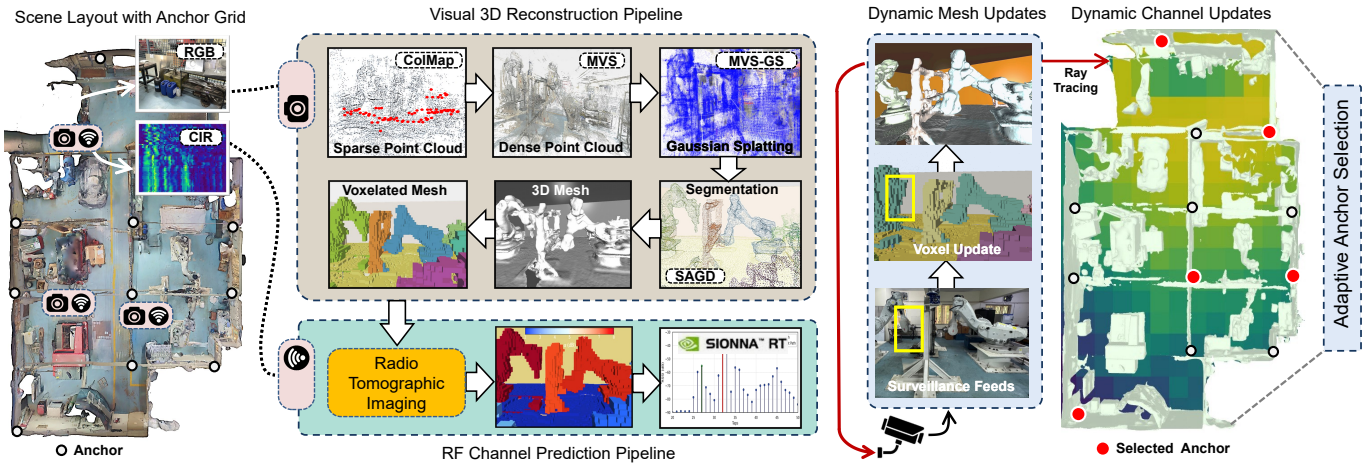

Fig. 5: Overall schematic of the Argos framework. The visual reconstructions and the dynamic updates run in realtime. Some demonstrations of the various functional stages of the framework are available here: https://sense.cse.iitm.ac.in/argos/

properties such as attenuation and reflection. To make it RF-aware, we define RF properties to individual objects in the scene. To achieve this, we perform tomographic estimation using CIRs collected together with the visual imagery during bootstrapping. At this stage, all anchors (denoted $\mathcal{A}$) are sequentially activated as transmitters so that CIRs can be recorded at multiple tagged receiver locations, maximizing geometric diversity and improving reconstruction quality. Unlike RSSI-based tomography, which compresses the channel into a single aggregate path loss, our approach leverages individual multipath components. Because the scene layout is known, we can ray-trace candidate paths between a transmitter $a \in \mathcal{A}$ and each capture location, while CIRs provide the corresponding measured delays and amplitudes. Each multipath component is then associated with a ray, and the residual loss along that ray is attributed to two object-specific parameters: (i) a per-voxel attenuation coefficient α_o, capturing signal decay through object o, and (ii) a boundary-loss coefficient β_o, capturing reflections and scattering at object interfaces. For a given anchor–receiver pair, the path loss of ray r_j is modeled as

$$PL_{r_j} = \sum_{o \in \mathcal{O}} L_{r_j,o}\, \alpha_o \; + \; \sum_{o \in \mathcal{O}} I_{r_j,o}\, \beta_o \tag{2}$$

where $L_{r_j,o}$ counts the voxels of o traversed by r_j and $I_{r_j,o}$ counts the boundary interactions. Stacking terms across all rays yields two matrices: L, the voxel-traversal matrix, and I, the boundary-interaction matrix. Intuitively, each row of L encodes *which voxels a ray passes through and how many*, while I encodes *where it reflects or refracts*. Concatenating them forms the projection matrix $\mathcal{M} = [L \quad I]$, and the unknown parameter vector $x = [\alpha^\top \; \beta^\top]^\top$ contains per-object attenuation and boundary-loss coefficients. The tomography problem thus reduces to solving $PL = \mathcal{M}x$, where PL is the vector of measured path losses over all anchor–receiver rays.

We solve for x in a regularized least-squares sense,

$$\hat{x} = \arg\min_{x} \; \|PL - \mathcal{M}x\|_2^2 + \lambda \|x\|_2^2, \tag{3}$$

whose closed-form solution $(\mathcal{M}^\top \mathcal{M} + \lambda I)^{-1} \mathcal{M}^\top PL$ provides stable attenuation estimates even with limited or noisy CIR data. Although the inverse problem is formally ill-posed, activating all anchors during bootstrap increases the diversity of ray geometries, improving the rank and conditioning of $\mathcal{M}$. This module drives the digital twin towards RF-awareness, rather than assigning accurate material values at voxel-level. Combined with visual priors (geometry and object grouping) and multipath-resolved CIRs, this makes the estimation considerably less human-dependent on material-association in creating such twins. This method is also more stable and faster than conventional RSSI-based tomography which operates on a voxel-level granularity. In this way, multipath-aware RTI upgrades the voxel map into a material-aware digital twin, enabling accurate channel prediction from sparse measurements.

Channel Synthesis and CIR Generation. With the voxelated mesh now enriched with estimated material properties, Argos can realistically predict the CIR for any anchor–receiver link. We employ NVIDIA Sionna RT to emulate multipath propagation over this calibrated digital twin, ensuring high-fidelity channel synthesis. Since the same CIR measurements were used to calibrate the attenuation image, the sim-to-real gap is minimized, allowing predictions to track physical measurements more closely.

B. ToA Error Guided Anchor Selection

With the material-property-calibrated digital twin, Argos can generate CIRs for any anchor-receiver link. However the problem of sim-to-real gap in channel modeling still persists due to approximations in geometry when simulating ray-traces. We address this by fusing the collected ground truth channel model with the channel model synthesized by Sionna. Using the simulated channel model together with local fading statistics, we predict how LDE_Θ (eqn. 1) detects the first path under nLoS conditions. This helps us estimate how Δ_{FMD} is spatially distributed, which can guide anchor selection: choosing the subset that offers wide coverage, low misdetection (Δ_{FMD}) and acceptable update rates.

■ **Modeling ToA Estimation Error.** We model the Channel Impulse Response and consequently the Δ_{FMD} map of the entire arena using $\mathsf{Sionna-RT}$ fused with real collected CIRs. Let X denote the random tap index reported by LDE_Θ. $\Pr(X = j)$ corresponds to the probability of the event that tap j is the first to exceed the threshold Θ, while all preceding taps remain below it. This can be expressed as,

$$\Pr(X = j) = \Big(\prod_{\tau=0}^{j-1} [1 - q_i(\Theta)] \Big) \, q_j(\Theta) \tag{4}$$

Here, $q_\tau(\Theta) = \Pr(A_\tau > \Theta)$ is the probability that tap τ is detectable, or, the tap amplitude is above the detection threshold Θ. The product term calculates how likely all taps before j lie below Θ, while the tap j is the first such tap to lie above Θ. With a delay spread of K taps, the expectation of the $\mathsf{LDE}'\mathsf{s}$ output is then $\mathbb{E}[X] = \sum_{j=0}^{K-1} j \, \Pr(X = j)$. Since Δ_{FMD} is the difference between real and reported first-path taps, the relation for the expected Δ_{FMD}, $\mathbb{E}[\Delta_{\mathsf{FMD}}] = |j_{\mathsf{FP}} - \mathbb{E}[X]|$ hence follows.

The key modeling choice is how to evaluate $q_\tau(\Theta)$. Each tap amplitude A_τ can be physically well described by a $\mathsf{Nakagami}^{\mathsf{m}}$[44] distribution, $A_\tau \sim \mathsf{Nakagami}(m_\tau, \Omega_\tau)$, which is modeled for short-term RF fading. Here m_τ characterizes fading severity and $\Omega_\tau = \mathbb{E}[A_\tau^2]$ is the mean power. Since the Nakagami distribution does not yield a closed form for threshold exceedance, we approximate A_τ by a Gaussian fit with equivalent mean μ_τ and variance σ_τ^2. Hence $q_i(\Theta) \approx \mathsf{Q}\big((\Theta - \mu_i)/\sigma_i\big)$, where $\mathsf{Q}(x)$ denotes the area under the tail of the standard normal distribution from x to ∞. μ_τ are taken from the tap amplitudes of the ray-traced CIR, while the variances σ_τ^2 are interpolated from CIR measurements using inverse-distance weighting. Although estimating $\mathbb{E}[\Delta_{\mathsf{FMD}}]$ is ill-posed, however, visual priors, multipath CIRs for μ_i and anchor diversity at bootstrap together yield a stable solution.

Scalability. Argos caches the Δ_{FMD} in a lookup table (FMD cache) $\mathcal{F} : (L_{\mathrm{tx}}, L_{\mathrm{rx}}) \mapsto \Delta_{\mathsf{FMD}}$, so that anchor selection and scene updates can query range errors in constant time. Ray-tracing outputs are also stored as multipath profiles, enabling local updates when objects move without rerunning the full pipeline. In this way, Argos transforms the digital twin into not only a channel predictor but also an error predictor, providing the first use of visual priors to proactively model ToA errors in cluttered environments.

■ **Optimal Anchor Subset Selection.** In a ToA system, each anchor transmits in a separate TDMA slot. Adding anchors improves geometry and reduces range errors, but also lowers the update rate. Fewer anchors keep updates fast but risk poor coverage and higher range errors.

Let $\mathcal{L}$ be the set of candidate receiver locations, where each $L \in \mathcal{L}$ is assigned a prior weight w_L reflecting its likelihood of occupancy (e.g., from trajectory priors or deployment maps). For each location L, let $\mathcal{A}_L \subseteq \mathcal{A}$ denote the set of anchors that cover L. Each anchor $a \in \mathcal{A}_L$ contributes an expected misdetection error $\Delta^a_{\mathsf{FMD},\,\mathsf{L}}$, average error at L being, $\overline{\Delta}_{\mathsf{FMD},\,L} = \frac{1}{|\mathcal{A}_L|} \sum_{a \in \mathcal{A}_L} \Delta^a_{\mathsf{FMD},\,\mathsf{L}}$. We seek a selected anchor set $S \subseteq \mathcal{A}$ that satisfies three requirements: (i) *Coverage:* the weighted fraction of locations covered by at least one anchor in S must exceed a target threshold p. (ii) *Coverage density:* every covered location must be supported by at least four anchors, i.e., $|\mathcal{A}_L \cap S| \;\geq\; 4, \quad \forall L \in \mathcal{L}$. (iii) *Error minimization:* the weighted average misdetection across all locations should be minimized,

$$\min_{S \subseteq \mathcal{A}} \; \sum_{L \in \mathcal{L}} w_L \, \overline{\Delta}_{\mathsf{FMD},L} \tag{5}$$

The subset size is constrained by latency: since each anchor occupies a ToA slot, $|S| \leq N_{\max}$, where $N_{\max}$ is determined by the maximum desired update rate. The problem naturally reduces to a weighted set-cover: each anchor *covers* a subset of prior-weighted locations and contributes an associated error cost through its $\Delta^a_{\mathsf{FMD},L}$ values. We solve it using a greedy strategy, where at each step the candidate anchor a is evaluated by the additional prior-weighted locations it brings under coverage together with the average misdetection it introduces. The anchor offering the best tradeoff is added to S, and the process repeats until the *coverage*, *coverage density* and *latency* constraints are met. This greedy rule yields near-optimal subsets in practice while keeping computation scalable.

C. Adaptation to Dynamics

Anchor selections are not static: in cluttered environments, multipath can shift as objects move, altering ranging errors. A region once well served by a subset of anchors may later degrade if its multipath profile changes. To handle this, Argos leverages static surveillance cameras that continuously track object motion and trigger lightweight updates to the digital twin, avoiding costly full reconstructions. When an object $o_i \in \mathcal{O}$ moves or rotates, its displacement is represented by a rigid-body transform $(\Delta_{tr}, \Delta_{rot})$. This transform is applied to all voxels $\{v^{o_i}_{g_j}\}$ spanning the object, where g_j are their global IDs. The original voxels are reset to free space $\{v^{free}_g\}$, while the transformed set becomes $\{v^{o_i}_{g'}\}$ at the new locations. At the mesh level, the same transform shifts or rotates the object surface directly, keeping the geometric representation consistent. Currently Argos considers transforms in the objects identified during the 3D reconstruction phase, which can alter the multipath profile when moved. Transient objects like humans do not significantly affect the scene and hence not considered in the system.

Next, Argos identifies rays in the multipath profile $\mathcal{M}$, whose trajectories intersect voxels that changed their object labels. To speed up computation, *only* these affected rays are re-traced and updated for the corresponding transmitter–receiver pairs. The FMD cache $\mathcal{F}$ is refreshed in tandem, so error predictions remain accurate. Finally, the incremental updates also enrich the tomography. Initial estimation may be underdetermined, but as objects move, new ray equations are naturally added, while unchanged ones remain valid. This refines α_{o_i} and β_{o_i} over time, turning scene dynamics into additional calibration opportunities. In this way, Argos maintains a lightweight, scalable digital twin that evolves with the environment, supporting proactive anchor selection as well as live modeling of range errors and CIRs across the scene.

IV. Argos Testbed and Evaluation Results

A. Argos *Testbed Setup*

We evaluate Argos in a $14.5\,\text{m} \times 8\,\text{m}$ section of a fabrication facility, dense with heavy machinery, movable racks and metal fencing that create severe nLoS and multipath (fig. 6). Four ceiling-mounted surveillance cameras provide complete coverage, enabling coarse ($\approx$2 m) tracking of all objects.

Hardware Infrastructure. We deploy twelve Decawave DW1000 UWB transceivers as fixed anchors in a grid, operating in ToA mode. The anchors run at 3.9 GHz channel with 500 MHz bandwidth and 24 dB transmit gain. Receiver nodes perform TWR with selected anchors and extract CIRs from packets, timestamped and tagged with anchor IDs. The hardware reported CIR is sampled at 1 ns resolution, with the first path reported at a higher precision of 12.65 ps. However, for Δ_{FMD} estimation, the effective ground-truth resolution remains 1 ns. Each receiver node is powered and interfaced through a dedicated smartphone, which logs range and CIR spectrograms locally. The DW1000 uses an industry-standard LDE algorithm to timestamp the first path (§ II-C), adapting thresholds to the noise floor for robustness. Thresholds are set through the `LDE_CFG1` register, which applies separate scales to the noise floor and noise peaks. This setup suppresses false triggers while preserving attenuated nLoS paths. The smartphone also captures HD images synchronized with CIR data to form $\{\text{Visual}, \text{RF}\}$ tuples, which are uploaded over WiFi to a central controller (NVIDIA RTX3090 GPU system).

Software Pipelines. We use the central controller for running the software pipelines in real time. For visual reconstruction, we use ColMap[37], [43] with SfM to generate the initial point cloud. MVSGaussian[36] then densifies it using MVS and represents the scene with Gaussian splats. Finally, SAGS[45] applies semantic segmentation to produce a segmented point cloud. From this, we generate a mesh augmented with material properties estimated by RTI and export it in a format compatible with Sionna-RT. This enables CIR-level ray-tracing simulations with access to individual propagation paths.

Dataset. We collect ground truth along a 50 m marked trajectory (fig. 6*(right)*), covering about 700 locations. At each location, we record (a) multiple RGB images from different heights and directions, (b) CIR spectrograms from all anchors over 5s at about 100 Hz and (c) the true physical distance measurement of the receiver location from all twelve anchors. The dataset includes about 5K images and 0.4M CIR samples. Additionally, we generate synthetic channel traces for the same receiver locations via ray-tracing in Sionna-RT that are used for validating the accuracy of predicted CIR or Δ_{FMD}.

B. *Experiments and Performance Evaluation*

Fig. 6 shows snapshots of the reconstructed mesh generated from images. This segmented 3D scene serves as the preliminary mesh toward building an RF-aware digital twin of the environment. The digital twin is voxelized at 30 cm resolution ($1\,\text{ns} \times c$) and the coverage threshold is set to -90 dBm.

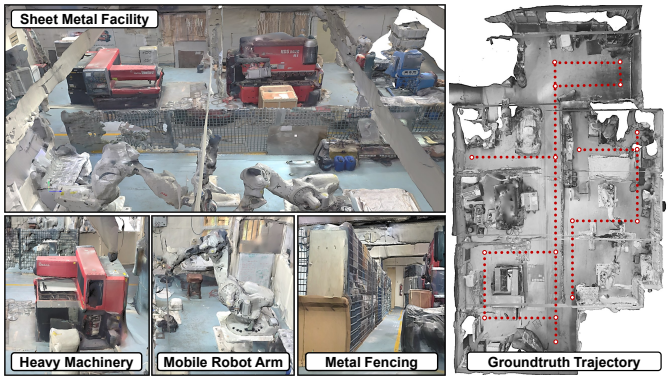

Fig. 6: 3D reconstructed views of the testbed with the ground-truth trajectory (red dotted line) marked along which imagery and RF data were collected.

■ **Overall Localization Accuracy**. We evaluate Argos 's accuracy by comparing anchor configurations from three techniques: a GDoP-based baseline [12], Argos with layout-only information, i.e., no RF attenuation details (Argos_{LO}) and the complete pipeline. We perform anchor selection with budgets of 3, 5 and 8 anchors from the grid of twelve fixed anchors.

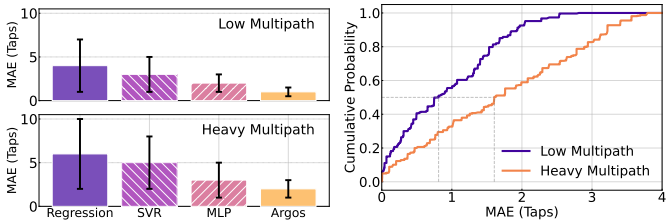

Fig. 7: *Top:* In low multipath, regression, SVR and MLP yield MAEs of $\approx$4, 3 and 2 taps, while Argos achieves near-perfect estimation. Under heavy multipath, errors rise to $\approx$6, 5 and 3 taps, with Argos remains minimally affected. Argos also shows lower variance or lesser uncertainity overall. *Bottom:* ECDF of MAE between measured and estimated Δ_{FMD} via $\text{Sionna} - \text{RT}$. Deviations remain within 4 taps, medians being 1 and 2 taps under low and heavy multipath respectively.

(a) *Δ_{FMD} Model Accuracy*: Accurate prediction of Δ_{FMD} is central to optimizing anchor selection. A single tap misprediction results in $\approx$30 cm of ToA estimation error. While learning-based methods have been explored to classify LoS/nLoS or predict range errors from CIR spectrograms, it is challenging to generalize such models across large areas and they often overfit to small regions. Under scene dynamics or heavy multipath, their predictions become highly unreliable compared to our ray-tracing–based analytical model. In contrast, Argos remains reliable, with MAEs ($|\Delta^{True}_{\text{FMD}} - \Delta^{Pred}_{\text{FMD}}|$) of $\approx$1–2 taps even under heavy multipath. ML baselines like polynomial regression, support vector regression (SVR) and multi-layer perceptron (MLP) are evaluated and average 3–7 taps of MAE. Further, their accuracy degrade sharply with increasing multipath or on introducing scene dynamics (see, fig. 7 for details). The ground-truth first path (j_{FP}) is derived directly from the physical range measurement.

(b) *Improvement Over Layout-Based Optimization*: We evaluate the added value of incorporating RF information on top of the layout mesh, comparing Argos_{LO} with Argos. As demon-

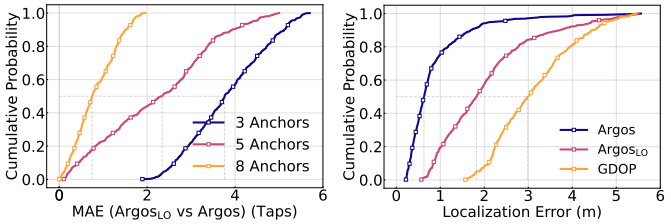

Fig. 8: *Left:* Argos achieves up to a 4 tap median reduction in Δ_{FMD} over Argos$_{\mathsf{LO}}$. *Right:* Under a 5-anchor budget, Argos yields a median localization error of 0.6 m, compared to 1.8 m (Argos$_{\mathsf{LO}}$) and 2.97 m (GDoP).

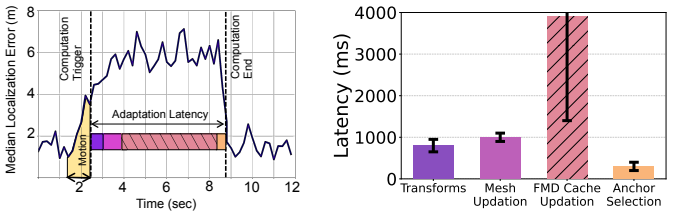

Fig. 9: *Left:* Time-series of localization error under 6% voxel churn. Error rises during object movement and is restored to sub-meter once the cache is updated. *Right:* Approximate breakdown of the update pipeline: transform calculation (800ms), mesh update (1s), Δ_{FMD} cache recomputation (4s), and anchor reselection (300ms).

strated in fig. 8, RF awareness yields clear benefits. Argos reduces median Δ_{FMD} by up to 4 taps under sparse anchor budgets. Even as anchor density increases and overall errors decrease, it continues to maintain a measurable advantage.

(c) *Baseline Comparisons*: Beyond Δ_{FMD}, we examine the compounded localization error, which accumulates over multiple TWR exchanges while partially correcting for range bias. Since range errors vary across anchors, their combination drives the overall localization error distribution. We show in fig. 8 (*right*) that with 5 anchors, incorporating RF awareness enables Argos to achieve sub-meter median error in harsh multipath, outperforming Argos$_{\mathsf{LO}}$ and GDoP-based optimizations.

■ **Performance in Dynamic Environments.** We evaluate Argos 's robustness to scene dynamics by introducing controlled churn in the arena and tracking localization stability. Starting from the baseline object layout (Fig. 10*(left)*), Argos builds an Δ_{FMD} cache, i.e., $\mathcal{F} : (L_{\mathrm{tx}}, L_{\mathrm{rx}}) \mapsto \Delta_{\mathsf{FMD}}$, using a 30 cm voxel resolution and a -90 dBm coverage threshold. This enables anchor selection and scene updates to query expected range errors in constant time. Localization errors are then measured at the same ground-truth locations while the layout is churned, with changes quantified by the fraction of voxels that differ between the initial and updated maps. We test churn levels from 2% (e.g., moving a single shelf) up to 10% (relocating all shelves and robotic arms). Larger-scale dynamics are emulated in SIONNA-RT, which we have shown to be reliable for Δ_{FMD} predictions. It is to be noted that Argos does not wait for large changes in the scene. On detecting small, visually distinguishable changes, it calculates the transform of the individual objects, recalculates the scene, and updates the Δ_{FMD} cache and anchors. Pipelining of tasks can lead to further speedups: for instance, mesh updation and the channel modeling-anchor selection loop can be treated as two separate stages to achieve parallelization.

(a) *Adaptation Latency*. Fig. 9 shows the effect of 6% voxel churn as a result of object movement in the scene. The median error shoots to ≈6 m, then recovers within ≈5 s after the update pipeline completes. We analyze recovery time by breaking it into four stages: voxel transform computation, mesh update, Δ_{FMD} cache recomputation, and anchor reselection. The Δ_{FMD} cache stage dominates (60–70% of total latency) due to ray-tracing overhead, but can be reduced by parallelizing ray-tracing across anchors.

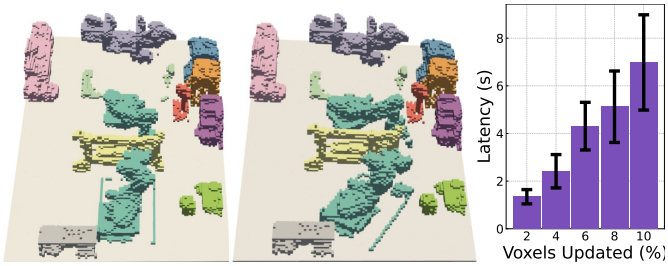

Fig. 10: Effect of layout dynamics on recomputation latency. *Left:* Original voxel map. *Middle:* Changed map with displaced robotic arms (green) – 6% voxel change. *Right:* Δ_{FMD} cache recomputation latency grows with voxel churn: ≈2 s for 2% updates and ≈7 s for 10%.

(b) *Effect of Object Movement*. Δ_{FMD} cache recomputation latency increases with object movements or voxel churns, as more rays must be retraced (fig. 10). Non-cache overheads are negligible; with about 6% change, recomputation takes ≈4s and with 10% nearly 7s. In practice, large spikes in voxel churns are relatively uncommon – factory dynamics evolve gradually over time, e.g.,forklifts or shelves move. Since such changes unfold over tens of seconds to minutes, far slower than Argos 's 5–7s recomputation cycle, the system maintains continuous operation, showing transient error spikes followed by recovery. Performance can be further improved through IMU fallback, localized updates, and parallelized ray-tracing, which are not yet implemented in the current version of Argos.

■ **Scene Reconstruction.** We analyze how 3D reconstruction fidelity affects localization accuracy in Argos. First, we assess semantic segmentation quality by measuring segmentation accuracy. Second, we vary voxel resolution from 10 cm to 100 cm in steps of 10 cm and run Argos with a five-anchor budget at a -90 dBm coverage threshold. To capture material effects, both RF-aware and layout-only (Argos$_{\mathsf{LO}}$) digital twins are evaluated.

(a) *Required Amount of Granularity*. Fig. 11 (*top*) shows that voxel resolution strongly impacts localization accuracy. Coarse voxels (100 cm, PSNR ≈8 dB) produce a median error of ≈6 m, while finer voxels (10 cm, PSNR ≈22 dB) achieve sub-meter accuracy. Beyond PSNR ≈14 dB (40 cm), accuracy gains diminish while Δ_{FMD} cache recomputation cost rises steeply. We adopt 30 cm voxels that balance sub-meter accuracy with practical update latency. Notably, this inflection point coincides with the CIR time resolution of 1 ns (30 cm).

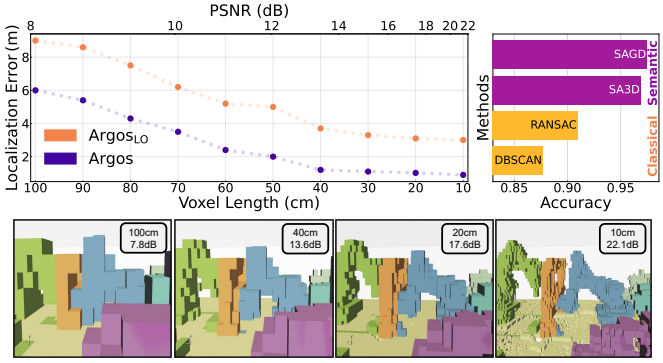

Fig. 11: Impact of 3D reconstruction fidelity. *Top, Left:* Median localization error decreases with finer voxelization as PSNR improves, but accuracy gains saturate beyond 40 cm voxels (≈14 dB). Across all resolutions, Argos outperforms Argos$_{LO}$. *Top, Right:* Segmentation accuracy across methods. *Bottom:* Example reconstructions with 10, 20, 40 and 100 cm voxel resolutions.

(b) Comments on Segmentation. We manually segment 47 objects, ranging from shelves, forklifts and robotic arms to barrels, wooden partitions and large scrap plates, whose dimension exceeds our voxel length. Groundtruth segments are created using the Blender tool [46]. Fig. 11 (*top, right*) reports accuracy of different methods. Classical approaches such as DBSCAN (87%) and RANSAC (91%) often merge adjacent objects, especially in clutter. Semantic methods perform better: SA3D and SAGD achieve 98–99%, with SAGD yielding the cleanest boundaries by mitigating splatting artifacts.

■ **Discussion on Coverage.** Good Δ_{FMD} is ineffective if large regions remain uncovered. We study the RF coverage with under two settings: (i) varying RSS thresholds from -92 to -50 dBm with five anchors, and (ii) varying anchor count (3–12) at a fixed -90 dBm threshold.

(a) Effect of Coverage Threshold. Higher RSS thresholds restrict coverage to strong links, lowering Δ_{FMD} but leaving uncovered patches. Lower thresholds expand coverage by including weak links, which raises Δ_{FMD}. A mid-level threshold offers the best trade-off, ensuring reasonable coverage while keeping mismatch distances manageable (see, fig. 12(*top-left*)).

(b) Effect of Anchor Count. Increasing the anchor budget improves coverage and reduces error, with the most significant gains between three and five anchors. Beyond five, improvements become marginal, while additional anchors increase ToA-TWR cycle time due to longer TDMA schedules (see figs. 2 and 12(*top-right*)).

V. Conclusion

In this work, we introduced Argos, a multimodal wireless digital twin that fuses jointly captured visual and RF data to optimize anchor selection in cluttered industrial environments. By combining geometric and attenuation cues, Argos yields a material-aware scene representation that remains robust across diverse multipath conditions. A central capability of Argos is its proactive use of visual priors to isolate changes, enabling fast, selective updates without repeated RF recalibration or model retraining. This design makes the system lightweight, scalable and resilient to real-world scene dynamics. Argos

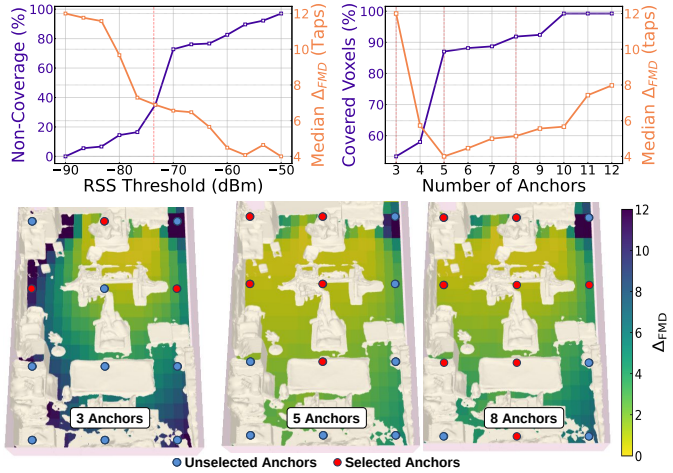

Fig. 12: *Top-left:* Effect of RSS threshold with 5 anchors. At -74 dBm, coverage drops to ≈40% uncovered while median Δ_{FMD} improves to 7 taps, giving a practical balance. *Top-right:* Effect of anchor count at -90 dBm. Largest gains occur from 4 to 5 anchors; beyond 5, coverage gains and Δ_{FMD} reduction are marginal while ToA-TWR overhead grows. *Bottom:* Δ_{FMD} heatmaps for representative anchor placements.

achieves median localization errors as low as 0.6 m, improving accuracy by up to state-of-the-art 3× over baselines. Moreover, it sustains sub-meter accuracy even under heavy multipath and scene dynamics, with update times as low as 5s.

Currently, Argos is limited to objects present during the initial scene reconstruction. Newly introduced objects, such as a new forklift moving into the scene, cannot yet be incorporated automatically into Argos created digital twin. Extending the system with automatic mesh generation and material association for previously unseen objects is part of future work. Argos is designed for cluttered, dynamic environments and has been evaluated under such conditions; however, its performance depends on the quality and quantity of data collected during the bootstrapping phase, which directly affects the fidelity of the digital twin (fig. 11). Finally, scalability is constrained by the number of transmitters and receivers, as each update requires recomputing channel models for all affected transmitter–receiver pairs.

We also plan to extend Argos beyond localization. By coupling visual priors with RF-aware channel synthesis, Argos establishes a general mechanism for constructing and maintaining RF-augmented digital twins of real environments. The ability to synthesize CIRs at scale enables a broader class of CIR-driven sensing and inference tasks. As a result, Argos moves beyond addressing a single localization infrastructure problem and provides a foundation for RF-aware digital twins that can support a wider range of integrated sensing and communication applications.

Acknowledgements

This research work is partially supported by FedEx Smart Center CSR Grant at IIT Madras (Project No. SB25261147CSFEDX008970) and ANRF Advanced Research Grant (Project No. ANRF/ARG/2025/005719/ENS).

REFERENCES

[1] Occupational Safety and Health Administration, U.S. Department of Labor, "Forklift Accidents in USA," https://www.osha.gov/ords/imis/AccidentSearch.search?acc_keyword=%22Forklift%22&keyword_list=on, [Accessed 02-10-2025].

[2] Fei Dai, Feng Li, Jianhua Ma and Qiuli Chen, "Intelligent Warehouse Based on Radio Frequency Identification and Recurrent Neural Network," 2025.

[3] Yeawon You, JinYi Yoon, Dayeon Kang, Jeewoon Kim and HyungJune Lee, "CollageMap: Tailoring Generative Fingerprint Map via Obstacle-Aware Adaptation for Site-Survey-Free Indoor Localization," in *IEEE PerCom*, 2025.

[4] M. B. Luca Barbieri and M. Nicoli, "UWB Localization in a Smart Factory: Augmentation Methods and Experimental Assessment," *IEEE Trans. on Instrumentation and Measurement*, 2021.

[5] Bertrand Perrat, Francisco Zampella, Miltiadis Chrysopoulos, Zhuo Wang and Firas Alsehly, "Optimization-based wi-fi radiomap construction for multifloor indoor positioning," in *2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2024.

[6] Anurag Pallaprolu, Belal Korany, and Yasamin Mostofi, "Wiffract: A New Foundation for RF Imaging Via Edge Tracing," in *ACM Mobicom 2022*.

[7] Fazeelat Mazhar, Muhammad Gufran Khan and Benny Sällberg, "Precise Indoor Positioning Using UWB: A Review of Methods, Algorithms and Implementations," *Wireless Personal Communications*, 2017.

[8] Mohammadmoradi Heydariaan, Gnawali Hessam and Omprakash, "Toward Standard Non-Line-of-Sight Benchmarking of Ultra-Wideband Radio-Based Localization," in *CPSBench*, 2018.

[9] Alireza Ansaripour, Milad Heydariaan, Omprakash Gnawali and Kyungki Kim, "ViPER: Vehicle Pose Estimation using Ultra-WideBand Radios," in *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2020.

[10] Wenpeng Wang, Fateme Nikseresht, Viswajith G. Rajan, Jiechao Gao and Bradford Campbell, "Enabling Ubiquitous Occupancy Detection in Smart Buildings: A WiFi FTM-Based Approach," in *DCOSS-IoT*, 2023.

[11] WIPELOT, "Forklift Tracking and Fleet Management RTLS System," https://wipelot.io/forklift-tracking-and-fleet-management-rtls-system, [Accessed 02-10-2025].

[12] Y. Ding, D. Shen, K. Pham and G. Chen, "Optimal Placements for Minimum GDOP With Consideration on the Elevations of Access Nodes," *IEEE Trans. on Instrumentation and Measurement*, 2025.

[13] Xiaopeng Zhao, Zhenlin An, Qingrui Pan and Lei Yang, "NeRF2: Neural Radio-Frequency Radiance Fields," in *ACM MobiCom*, 2023.

[14] Lei Zhang, Kan Jiao, Wei He and Xinheng Wang, "Anchor Deployment Optimization for Range-Based Indoor Positioning Systems in Non-Line-of-Sight Environment," *IEEE Sensors Journal*, 2024.

[15] Antti Saikko, Jukka Talvitie, Joonas Säe, Juho Pirskanen and Mikko Valkama, "Positioning and Tracking in DECT-2020 NR With Proactive Anchor Selection for Range, Angle, and RSS Measurements," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, 2025.

[16] Marko Angjelichinoski, Daniel Denkovski, Vladimir Atanasovski and Liljana Gavrilovska, "Cramér–Rao Lower Bounds of RSS-Based Localization With Anchor Position Uncertainty," *IEEE Trans. on Information Theory*, 2015.

[17] Sheng-Po Kuo and Yu-Chee Tseng, "Discriminant Minimization Search for Large-Scale RF-Based Localization Systems," *IEEE Trans. on Mobile Computing*, 2011.

[18] Chao Liu, Aroland Kiring, Naveed Salman, Lyudmila Mihaylova and Inaki Esnaola, "A Kriging algorithm for location fingerprinting based on received signal strength," in *Sensor Data Fusion (SDF)*, 2015.

[19] Gabriel Appleby, Linfeng Liu and Li-Ping Liu, "Kriging Convolutional Networks," in *AAAI*, 2020.

[20] Ron Levie, Çağkan Yapar, Gitta Kutyniok and Giuseppe Caire, "RadioUNet: Fast Radio Map Estimation With Convolutional Neural Networks," *IEEE Trans. on Wireless Communications*, 2021.

[21] Dianxin Luan and John S Thompson, "Channelformer: Attention Based Neural Solution for Wireless Channel Estimation and Effective Online Training," *IEEE Trans. on Wireless Communications*, 2023.

[22] Songyang Zhang, Achintha Wijesinghe and Zhi Ding, "RME-GAN: A Learning Framework for Radio Map Estimation Based on Conditional Generative Adversarial Network," *IEEE Internet of Things Journal*.

[23] Usman Mahmood Khan, and Raghav H Venkatnarayan and Muhammad Shahzad, "RFMap: Generating indoor maps using RF signals," 2020.

[24] Xiucheng Wang, Keda Tao, Nan Cheng, Zhisheng Yin, Zan Li, Yuan Zhang and Xuemin Shen, "RadioDiff: An Effective Generative Diffusion Model for Sampling-Free Dynamic Radio Map Construction," *IEEE Trans. on Cognitive Communications and Networking*, 2025.

[25] Le Zhao, Zesong Fei, Xinyi Wang, Jihao Luo andZhong Zheng, "3D-RadioDiff: An Altitude-Conditioned Diffusion Model for 3D Radio Map Construction," *IEEE Wireless Communications Letters*, 2025.

[26] Joey Wilson and Neal Patwari, "Radio Tomographic Imaging with Wireless Networks," *IEEE Trans. on Mobile Computing*, 2010.

[27] Ossi Kaltiokallio, Riku Jäntti and Neal Patwari, "ARTI: An Adaptive Radio Tomographicorekondy2023winert Imaging System," *IEEE Trans. on Vehicular Technology*, 2017.

[28] Jingru Wei, Yongtao Ma and Yuxiang Han, "A New Method for Enhancing the Robustness of Close-Range Targets Imaging Based on UWB," *IEEE Sensors Journal*, 2025.

[29] B. K. C. R. Karanam and Y. Mostofi, "Magnitude-Based Angle-of-Arrival Estimation, Localization, and Target Tracking," in *Proc. of ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2018.

[30] B. Korany, C. R. Karanam and Y. Mostofi, "Adaptive Near-Field Imaging with Robotic Arrays," in *Proc. of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2018.

[31] Jakob Hoydis, Faycal Ait Aoudia, Sebastian Cammerer, Merlin Nimier-David, Nikolaus Binder, Guillermo Marcus and Alexander Keller, "Sionna RT: Differentiable Ray Tracing for Radio Propagation Modeling," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023.

[32] Haofan Lu, Christopher Vattheuer, Baharan Mirzasoleiman and Omid Abari, "NeWRF: A Deep Learning Framework for Wireless Radiation Field Reconstruction and Channel Prediction," *arXiv preprint arXiv:2403.03241*, 2024.

[33] Pratik Kumar Orekondy, Shreya Kadambi, Hao Ye, Joseph Soriaga Aras Behboodi, "WiNeRT: Towards Neural Ray Tracing for Wireless Channel Modelling and Differentiable Simulations," in *ICLR*, 2023.

[34] Jakob Hoydis, Fayçal Aït Aoudia, Sebastian Cammerer, Florian Euchner, Merlin Nimier-David, Brink Ten, Keller Stephan and Alexander, "Learning Radio Environments by Differentiable Ray Tracing," *IEEE Trans. on Machine Learning in Communications and Networking*, 2024.

[35] Xingyu Chen, Zihao Feng, Ke Sun, Kun Qian and Xinyu Zhang, "RFCanvas: Modeling RF Channel by Fusing Visual Priors and Few-shot RF Measurements," in *ACM SenSys*, 2024.

[36] Tianqi Liu , Guangcong Wang, Shoukang Hu, Liao Shen, Xinyi Ye, Yuhang Zang, Zhiguo Cao, Wei Li aZiwei Liu, "MVSGaussian: Fast Generalizable Gaussian Splatting Reconstruction from Multi-View Stereo," 2024.

[37] Schonberger,Frahm Johannes L and Jan-Michael, "Structure-From-Motion Revisited," in *CVPR*, 2016.

[38] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," 2021.

[39] Jiazhong Cen, Zanwei Zhou, Jiemin Fang, Wei Shen, Lingxi Xie, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and others, "Segment anything in 3d with nerfs," 2023.

[40] Xu Hu, Yuxi Wang, Lue Fan, Junsong Fan, Junran Peng, Zhen Lei, Qing Li, and Zhaoxiang Zhang, "SAGD: Boundary-Enhanced Segment Anything in 3D Gaussian via Gaussian Decomposition," *arXiv preprint arXiv:2401.17857*, 2024.

[41] Decawave Ltd., *DW1000 User Manual*. [Online]. Available: https://www.sunnywale.com/uploadfile/2021/1230/DW1000%20User%20Manual_Awin.pdf

[42] Caleb Phillips, Douglas Sicker and Dirk Grunwald, "A Survey of Wireless Path Loss Prediction and Coverage Mapping Methods," *IEEE Communications Surveys & Tutorials*, 2013.

[43] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys and Jan-Michael Frahm, "Pixelwise View Selection for Unstructured Multi-View Stereo," in *ECCV*, 2016.

[44] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in *Statistical methods in radio wave propagation*, 1960.

[45] Hu Xu, Wang Yuxi, Fan Lue,Fan Junsong, Junran Peng, Zhen Lei, Qing Li and Zhaoxiang Zhang, "Semantic Anything in 3D Gaussians," *arXiv preprint arXiv:2401.17857*, 2024.

[46] *Blender - a 3D modelling and rendering package*. [Online]. Available: http://www.blender.org