Beyond 5G: Reliable Extreme Mobility Management

Yuanjie Li[1*],Qianru Li[2], Zhehui Zhang[2], Ghufran Baig[3], Lili Qiu[3], Songwu Lu[2]

[1]Hewlett Packard Labs, [2] University of California, Los Angeles (UCLA), [3]University of Texas at Austin

ABSTRACT

Extreme mobility has become a norm rather than an exception. However, 4G/5G mobility management is not always reliable in extreme mobility, with non-negligible failures and policy conflicts. The root cause is that, existing mobility management is primarily based on *wireless signal strength*. While reasonable in static and low mobility, it is vulnerable to dramatic wireless dynamics from extreme mobility in triggering, decision, and execution. We devise REM, Reliable Extreme Mobility management for 4G, 5G, and beyond. REM shifts to *movement-based* mobility management in the *delay-Doppler domain*. Its signaling overlay relaxes feedback via cross-band estimation, simplifies policies with provable conflict freedom, and stabilizes signaling via scheduling-based OTFS modulation. Our evaluation with operational high-speed rail datasets shows that, REM reduces failures comparable to static and low mobility, with low signaling and latency cost.

CCS CONCEPTS

• **Networks → Mobile networks**; **Wireless access networks**; **Network reliability**; **Network protocol design**.

KEYWORDS

Mobile network, beyond 5G, extreme mobility management, reliability, policy conflicts, delay-Doppler domain, cross-band estimation.

1 INTRODUCTION

We have witnessed a boom in various extreme mobility scenarios, such as the high-speed rails, vehicle-to-everything, drones, and many more. Compared to traditional static and low-mobility scenarios, extreme mobility involve much faster client movement speed (up to 350km/h [1]) in the outdoor environment. Many extreme mobility scenarios need *always-on* Internet access anywhere, anytime. Today, a common solution is the mobile network, such as 4G, 5G

---

and beyond. It is the largest wireless infrastructure that offers wide-area mobility management for network access. It has served billions of users, and will hopefully serve trillions of Internet-of-Things.

This work starts with a simple question: *Is 4G/5G reliable for extreme mobility?* While the existing mobile network has been successful in supporting billions of mobile users, most users are moving slowly or static. With significantly faster client speed and 5G radios under higher frequency (e.g. sub-6GHz and above-20GHz millimeter waves), it is open to question whether existing mobility management design is still a good fit for extreme mobility.

Unfortunately, the answer is negative in reality. Our empirical study of 4G LTE over high-speed rails unveils that, the mobility events are more frequent and vulnerable. On average, the handovers between base stations occur every 11–20s. Different from static or low mobility scenarios, handover failure and policy conflicts arise with alarming frequency: The network failure ratio ranges between 5.2% and 12.5% depending on the train speed, and the policy conflicts occur every 194–1090s. Both challenge the functionality of mobile networks and amplify the failures, delays, transient oscillations, and persistent loops. While the results are from 4G LTE, we believe 5G will face similar challenges with its same mobility management design as 4G LTE, adoption of millimeter waves, and denser small cell deployments with more frequent handovers.

We show that, the fundamental cause of unreliable 4G/5G in extreme mobility is its *wireless signal strength-based design.* 4G/5G mobility takes wireless signal strength as input, relies on the client-side feedback to trigger, and decides the target based on policies. While reasonable in static and low mobility, this design is sensitive to dramatic wireless dynamics from the Doppler shift in extreme mobility. Such dynamics propagate to all phases of mobility management and cause slow feedback in triggering, missed good candidate cells in decision, and unreliable signaling in execution. Our empirical study further shows that, operators have tried to mitigate failures with proactive policies. However, their methods amplify the policy conflicts and eventually offset their failure mitigation.

We propose REM, **R**eliable **E**xtreme **M**obility management for 4G, 5G and beyond. Our key insight is *the client movement is more robust and predictable than wireless signal strength, thus suitable to drive mobility management*. So REM shifts to *movement-based* mobility management. REM is a signaling overlay in the *delay-Doppler domain*, which extracts client movement and multi-path profile with the recently proposed orthogonal time-frequency space (OTFS) modulation [3]. To relax the client-side feedback, REM devises a novel cross-based estimation to parallelize measurements. This is achieved by extending OTFS with singular value decomposition (SVD). REM further simplifies the policy with provable conflict freedom, and stabilizes the signaling with a novel scheduling-based OTFS. REM is backward compatible with 4G/5G in static and low mobility, without changing their designs or data transfers.

We prototype REM in commodity software-defined radio and evaluate it with high-speed rails datasets and 4G/5G standard channel

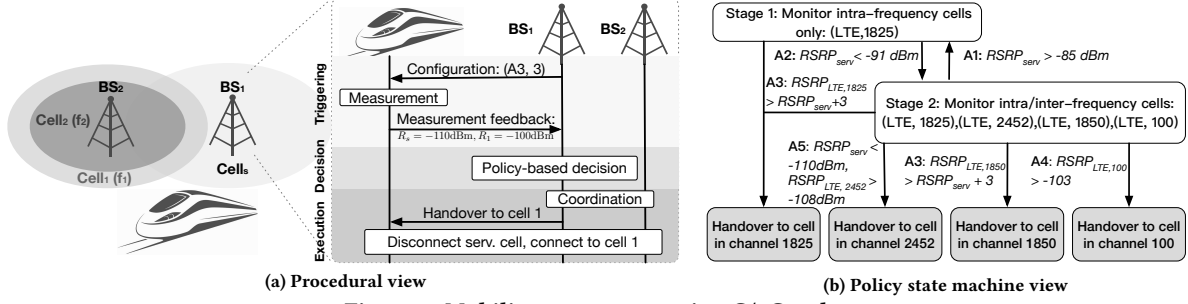(a) Procedural view   (b) Policy state machine view

Figure 1: Mobility management in 4G/5G today.

models. Compared to solutions today, REM eliminates policy conflicts, reduces failures by up to an order of magnitude (0.9×–12.7× depending on client speed). Even in extreme mobility, REM achieves comparable failure ratios to static and low mobility scenarios. Meanwhile, REM retains marginal overhead of signaling traffic and latency without hurting data transfer.

In summary, this work makes three main contributions:

(1) We conduct an empirical study for the network reliability in extreme mobility (§3). With the datasets from the Chinese high-speed rails, we unveil various causes of the failures and policy conflicts in all phases of the mobility management;

(2) We design REM, the first movement-based reliable extreme mobility management for 4G, 5G and beyond (§4–5). As a signaling overlay in the delay-Doppler domain, REM devises SVD-based cross-band estimation to relax the feedback, simplifies the policy for provable conflict freedom, and stabilizes the signaling with scheduling-based OTFS modulation;

(3) We prototype REM using software-defined radio (§6), and systematically evaluate REM's network failure reduction, policy conflict resolution, and system overhead (§7).

REM's artifacts are available at http://metro.cs.ucla.edu/REM.html.

2 MOBILITY MANAGEMENT TODAY

We introduce the 4G/5G mobility management today, and the general relationship between wireless and mobility.

4G/5G mobility management: To enable ubiquitous network access, 4G/5G deploys base stations to cover different areas. Each base station may run multiple cells under various frequency bands (using separate antennas) with different coverage and performance. As a client leaves the one cell's coverage, it will be migrated to another one (called *handover*) to retain its network access.

Figure 1a depicts 4G/5G handover [4, 5]. It has three phases. In the *triggering* phase, the serving cell configures a client to measure neighbor cells' signal strengths[1] with standard triggering criteria in Table 1. Upon receiving the client's feedback that meets the criteria, the serving cell moves to the *decision* phase. It runs its local policy to decide if more feedback is needed, if handover should start, and which cells to migrate to. It may also reconfigure the device for more feedback. After a handover decision, it moves to the *execution* phase, by coordinating with the target cell and sending handover command to the client. The client will disconnect from the serving cell, and connect to the target.

Table 1: Wireless triggering criteria in 4G/5G [4, 5]

Event	Criteria	Explanation
A1	$R_s > \Delta_{A1}$	Serving cell becomes better than a threshold
A2	$R_s < \Delta_{A2}$	Serving cell becomes worse than a threshold
A3 (A6)	$R_n > R_s + \Delta_{A3}$	Neighbor cell becomes offset better than serving cell
A4 (B1)	$R_n > \Delta_{A4}$	Neighbor cell becomes better than a threshold
A5 (B2)	$R_s < \Delta^1_{A5}, R_n > \Delta^2_{A5}$	Serving cell becomes worse than a threshold, and neighbor cell becomes better than a threshold

Wireless-mobility interplay: The wireless quality and client mobility mutually impact each other. On one hand, the wireless quality will decide the target cell for the mobile client (Figure 1a). On the other hand, as the client moves, the underlying signal propagation paths change accordingly and result in wireless dynamics (i.e., multi-path fading). The movement also incurs Doppler frequency shift, thus inter-carrier interference between cells and channel quality degradation. In 4G/5G OFDM/OFDMA[2], the channel remains approximately invariant in a very short duration $T_c \propto 1/v_{max}$ [6], where T_c is the coherence time and $v_{max} \propto vf/c$ is the maximum Doppler frequency, v is client movement speed and c is light speed. In static and low-mobility scenarios, the Doppler effect's impact is reasonably marginal (e.g., $T_c \approx 20$ms for a vehicle at 60km/h under 900MHz 4G LTE band). But in extreme mobility, a fast-moving client (e.g., 200–350km/h in high-speed rails) under higher carrier frequency (e.g., mmWave) will experience fundamentally more dramatic channel dynamics ($T_c \approx 1$ms as quantified in §3.1).

3 UNRELIABLE EXTREME MOBILITY

The 4G/5G mobility management is fundamentally a *wireless signal strength-based* design: It takes *wireless signal strength* as the main input, relies on client-side *wireless feedback* to trigger, and selects the target cell based on *wireless-driven policies*. While reasonable in static and low mobility, such design is sensitive to wireless dynamics in extreme mobility, and raises non-negligible network failures and policy conflicts in all phases of mobility management. We detail each phase (§3.1–3.3), analyze 5G's impact (§3.4), and define the problem (§3.5).

An overview of extreme mobility in reality: Table 2 compares two LTE datasets from high-speed rails (HSR, one from [7] and another from us) with our low mobility dataset (all detailed in §7). We make four high-level observations:

(1) Frequent handovers in extreme mobility: On average, a client on HSR experiences a handover every 20.4s, 19.3s, and 11.3s at <200km/h, 200–300km/h and 300–350km/h, respectively. Handover is more frequent as the train moves faster.

---

[1]In 4G/5G, the signal strengths can be RSRP, RSRQ or RSSI [4, 5].

[2]We use "OFDM" and "OFDMA" interchangeably since this paper focuses on wireless channel (not resource allocation), so they are equivalent.

Table 2: Network reliability in extreme mobility

		low mobility	high-speed rails (China)		
	Speed (km/h)	0 − 100	100 − 200	200 − 300	300 − 350
	Avg. handover interval	50.2 s	20.4 s	19.3 s	11.3 s
Failures (§3)	Total network failure ratio	4.3% (100%)	5.2% (100%)	10.6% (100%)	12.5% (100%)
	Feedback delay/loss (§3.1)	0.78% (18.0%)	1.7% (33.3%)	4.9% (46.3%)	6.9% (55.2%)
	Missed cell (§3.2)	1.8% (42.0%)	0.6% (11.1%)	0.4% (3.7%)	0.8% (6.4%)
	Handover cmd. loss (§3.3)	0.61% (14.0%)	1.1% (22.2%)	3.3% (31.5%)	2.4% (19.2%)
	Coverage holes	1.1% (26.0%)	1.7% (33.3%)	2.0% (18.5%)	2.4% (19.2%)
Conflicts (§3.2)	Avg. loop frequency	5,284.1s	410.1s	1,090.0s	194.6s
	Avg. # handovers/loop	2.2	3.9	3.0	3.3
	Avg. disruptions per loop	0.34 s	0.33 s	0.55 s	0.34 s
	Intra-frequency loops	0%	88.9%	100%	55.9%
	Inter-frequency loops	100%	11.1%	0%	44.1%

Table 3: Two-cell policy conflicts in HSR datasets.

| Conflicts	Type	Beijing-Taiyuan	Beijing-Shanghai [7]
A3-A4	Inter-frequency	4 (2.4%)	316 (23.6%)
A3-A5	Inter-frequency	1 (0.6%)	24 (1.8%)
A4-A4	Inter-frequency	2 (1.2%)	200 (14.9%)
A4-A5	Inter-frequency	5 (3.0%)	49 (3.7%)
A5-A5	Inter-frequency	0	2 (0.1%)
A3-A3	**Intra-frequency**	**155 (92.8%)**	**749 (55.9%)**

(2) Non-negligible failures in extreme mobility: Different from static or low mobility, the client suffers from frequent network failures in extreme mobility. To detect the network failures from mobility, we extract the handover events from LTE signaling messages, and check if the client successfully connects to the target cell for each handover. If not, the client loses radio connectivity and network access. We then compute the percentage of these failures out of all handover events. Table 1 shows the failure becomes more frequent with faster speed, from 5.2% at <200km/h to 12.5% at 300–350km/h.
(3) Diverse failure causes: For each failure event in extreme mobility, we check its nearby wireless signal strength, signaling messages, and configurations in the LTE datasets to analyze its causes. Table 1 shows the failures arise from triggering (§3.1), decision (§3.2), and execution (§3.3). They can also *unavoidably* occur in a no-coverage area (e.g., caves). In LTE today, failures from coverage holes are not dominant (19.2%–33.3%). So we focus on failures *with* coverage.
(4) Policy conflicts from failures: To mitigate these failures, operators adopt proactive handover policies[3]. However, such practice incurs frequent policy conflicts (every 194.6–1090.0s on average) and voids operators' failure mitigation efforts (§3.2).

3.1 Triggering: Slow, Unreliable Feedback

4G/5G relies on client-side feedback to trigger handovers (§2). Such feedback tracks client-perceived *wireless quality* of cells based on standard criteria (Table 1). In extreme mobility, such wireless signal strength-based feedback can be sluggish and cause failures. It faces the fundamental dilemma between *exploration* (more measurements for proper decision) and *exploitation* (timely triggering for handover). This causes two reliability issues:
● **Slow feedback:** To avoid failures, the client should deliver feedback *before* it leaves serving cell's coverage. But existing feedback is slow for two reasons: *(1) Head-of-line blocking:* To decide an appropriate target cell, the client should detect *all* cells that meet the criteria. For wireless signal strength-based feedback, the

(a) Slow feedback

(b) Block errors in signaling loss
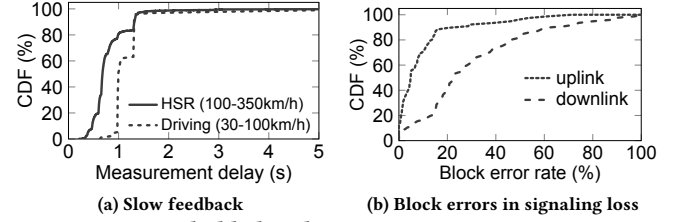
Figure 2: Unreliable handover triggering & execution.

client has to measure each cell *sequentially*, thus delaying later cells. Reducing the cells to measure can mitigate this delay, but at the risk of missing available cells (thus failures). *(2) Transient loop mitigation:* Instantaneous wireless measurement is dynamic and causes transient oscillations between base stations. To mitigate it, 4G/5G mandates the client to report a cell only if its criteria holds for a configurable triggering interval [4, 5][4]. This delays feedback with late handovers. Moreover, wireless quality may have changed *before* measurements, thus causing sluggish feedback and misleading triggering. Shortening the triggering interval may help, but causes more transient loops and signaling.
● **Lost feedback:** With dramatic wireless dynamics, the feedback is prone to loss/corruption in delivery. Such loss can be amplified by feedback delay: The client may have left serving cell's coverage before measurement, thus losing more feedback.
Validation: Table 2 shows 33.3–55.2% failures in HSR are from feedback delay/loss. The loss is mostly caused by errors: Figure 2b shows 9.9% block error rate before the loss, which implies the feedback is corrupted in delivery. For the feedback delay, Figure 2a shows a client on HSR takes 800ms on average to generate feedback from different bands, during which it has moved 44.6–78.0m (200–350km/h) along the rails and is thus too late for a viable handover. Moreover, the operator configures 40–80 ms the triggering interval for cells under same frequency as serving cell's (*intra-frequency cells*), and 128, 160, 256, 320 or 640 ms for others (*inter-frequency cells*). These are 2 orders of magnitude longer than 4G/5G OFDM coherence time $T_c \approx c/fv \in [1.16ms, 6.18ms]$ (§2) given $f \in [874.2, 2665]$ MHz and $v \in [200, 350]$ km/h from our datasets. Note operators have shortened triggering interval for faster feedback than low mobility (mostly 640ms in our dataset), but at the cost of more transient loops and signaling.
Opportunity: Shared physical multipath It is possible to accelerate feedback *without* reducing the cells to be explored. In reality, a base station usually operates multiple cells under different bands to improve the radio coverage and performance. Our dataset shows 53.4% of cells share the same base station with another cell[5]. These cells' signals traverse the same paths from the base station to the client, thus experiencing similar channels. In §5, we will use this to relax the exploration-exploitation dilemma for reliable feedback.

3.2 Decision: Complex, Conflicting Policy

4G/5G handover decisions are policy-driven by design. To accommodate diverse demands (good radio coverage, fast data speed, load balancing, failure mitigation, etc), each cell can customize its local

---

[3]We follow [8] to model a serving cell's handover policy as a state machine, and infer it using the LTE signaling messages and configurations from the serving cell. Our inference is coherent with the policy from real 4G/5G vendors and operators [9, 10].

[4]This configurable triggering interval is named as TimerToTrigger in 4G/5G.
[5]This is obtained by grouping the globally unique base station IDs from LTE cells' identifiers called ECIs [11].
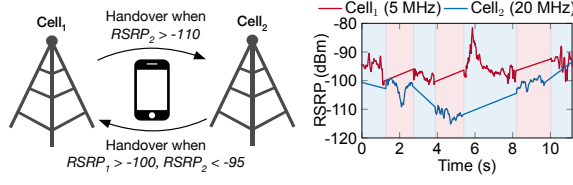
(a) Conflicts in load balancing   (b) Real handover traces
Figure 3: Policy conflicts from load balancing in HSR.

(a) Failure-induced conflicts   (b) Real handover traces
Figure 4: Failure-induced policy conflicts in HSR.

policies with configurable criteria in Table 1. Figure 1b exemplifies a typical policy inferred from our HSR dataset[3]. Such policy is tightly coupled with wireless feedback (§3.1). It is too complicated for extreme mobility, and suffers from two deficiencies:

• **Multi-stage policy:**   To tackle heterogeneous cells, most operators adopt *multi-stage* handover policies as exemplified in Figure 1b. The neighbor cells under the same frequency as serving cell's are measured and chosen first. Only if no intra-frequency cells are available, the policy will consider inter-frequency cells via measurement reconfiguration. The reason is to reduce inter-frequency measurements, which consumes more radio resource and slows down the data transfer[6]. But if the client moves fast, this policy can miss candidate cells without sending its feedback to the serving cell. Even if no intra-frequency cells exist, extra round trips (A2→reconfiguration→inter-frequency feedback) are needed for inter-frequency cells, during which the client may have missed the opportunity for handover and lost network access. The fundamental dilemma is that, inter-frequency measurements force existing policies to balance the spectral efficiency and decision delay.

• **Policy conflicts in extreme mobility:**   It has been shown that [14, 15], policies among cells can have conflicts and cause *persistent loops*. Figure 3a exemplifies a conflict from our dataset. Cell 1 and 2 have different bandwidths (5MHz v.s. 20MHz). For fast data speed, cell 1 moves a client to cell 2 if cell 2's signal strength $RSRP_2 > −110$dBm. But cell 2 adopts a different policy: It migrates a client to cell 1 if it is weak ($RSRP_2 < −95$dBm) and cell 2 is strong ($RSRP_1 > −100$dBm). Both policies can be *simultaneously* satisfied if $RSRP_1 > −100$dBm and $RSRP_2 \in (−119dBm, −95dBm)$. Then the client oscillates between cell 1 and 2 (8 handovers within 15s in Figure 3b). Such loop accumulates handover costs, disrupts client's service *and* incurs signaling storm for network.

Surprisingly, we note policy conflicts are amplified in extreme mobility, because of operators' desire for mitigating failures! This differs from [14, 15] that focus on static scenarios, and has been frequently observed in our dataset (detailed in validation below). As shown in §3.1, a fast-moving client may miss the cells and lose service due to slow feedback and decisions. To mitigate it, the operators adopt proactive policies in Figure 4a, by running handovers *before* neighbor cell is better than serving cell's. However, this raises conflicts if neighbor cells use the same policy. Such policy will not mitigate failures; the client will move back with loops.

Validation: Our empirical study confirms both problems. First, multi-stage policy can miss inter-frequency cells and induce handover failures. It accounts for 3.7%–11.1% failures in HSR (Table 2). Even so, operators still prefer multi-stage policy due to its low
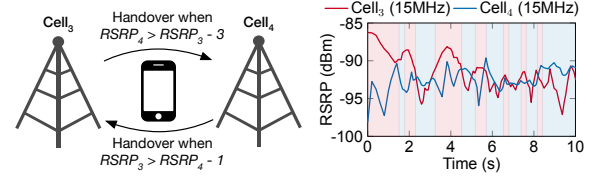
spectral waste. Without multi-stage policy, our dataset shows MeasurementGap in HSR would consume 38.3%–61.7% spectrum in inter-frequency measurements (depending on cell configurations).

Second, policy conflicts exist with alarming frequency in extreme mobility. Table 3 summarizes two-cell conflicts from our dataset. Note policy conflicts can also happen with >2 cells, so this result is a lower bound of conflicts in reality. On average, two-cell policy conflicts occur every 194.6–1090s in high-speed rails (3.8×–26.2× more than low mobility), each incurring 3.0–3.9 handovers on average. Surprisingly, intra-frequency policy conflicts (A3-A3) are much more than static or low-speed mobility [14, 15], and dominate the policy conflicts in extreme mobility (55.9%–100%). To trigger handovers early with less failures, the operators configure a proactive policy among cells (Figure 4a with $\Delta_{A3} < 0$). Such policy causes oscillations and voids the efforts of failure mitigation.

3.3   Execution: Unreliable Signaling

4G/5G can also fail if the serving cell cannot deliver handover command to the client. Similar to feedback loss in §3.1, such unreliable signaling mainly arises from the wireless dynamics in extreme mobility. It can also come from failure propagation of slow feedback in triggering (§3.1) and multi-stage policy in decision (§3.2).

Validation: Table 2 shows 19.2%–31.5% of network failures arises from the handover command loss. We detect these failures by observing successful delivery of feedback that can trigger handovers based on inferred policy (e.g., Figure 1b), but no handover command from serving cell until the client loses network access. We also observe high physical-layer block errors when such failure occurs. Figure 2b shows block error rate within 5 seconds before network failures. The average block error rate is 30.3% for downlink (handover command) and 9.9% for uplink (measurement feedback). This implies the signaling is corrupted during the delivery, thus failing to execute the handovers and losing network access.

3.4   Implications for 5G

The emergent 5G standards [5, 13, 16] offer various new features that 4G LTE lacks, such as the dense small cells, new radio bands (sub-6GHz and above-20GHz), renovated physical layer design, and advanced signaling protocols. Since 2019, 5G has been under active testing and deployment on the high-speed rails [17, 18]. While our empirical results in §3.1–§3.3 are from 4G LTE, we note reliable extreme mobility in 5G will be even more challenging because (1) 5G handovers [5] follow the same design as 4G [4]; (2) 5G adopts small dense cells under high carrier frequency, which incurs more frequent handovers that are more prone to Doppler shifts (§2) and failures; (3) while 5G refines its physical layers (e.g., Polar code and more reference signals [13]) to improve the reliability, they are still based on OFDM and suffers from similar issues.

---

[6]To measure an inter-frequency cell, a client should synchronize to it and measure its signal strength. The serving cell pre-allocates MeasurementGaps [12, 13] for this, during which the client cannot send/receive data.
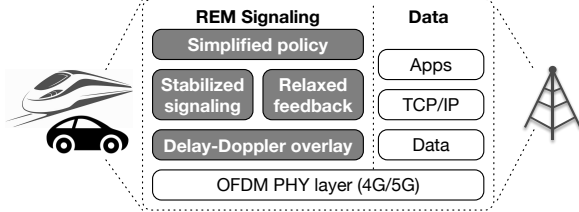
Figure 5: REM overview.

3.5 Problem Statement

This work aims at reliable extreme mobility management in 4G, 5G and beyond. We seek a solution with significantly less network failures, verifiable conflict-free policies, and negligible latency/signaling/spectral overhead. The solution should be reliable with dramatic wireless dynamics in extreme mobility, during which it may experience errors, delays, and failures in all phases of mobility management. The solution should be backward-compatible with existing OFDM-based 4G/5G (especially data transfer) in static and low mobility, and retain flexible policy for the operators.

4 INTUITIONS BEHIND REM

We devise REM, **R**eliable **E**xtreme **M**obility management to achieve all the goals in §3.5. Our key insight is that, extreme mobility is unreliable because of *wireless signal strength-based* management today. In extreme mobility, wireless signal strength is unreliable with Doppler shift and multipath fading (§2). This propagates failures to all phases of mobility management, i.e., sluggish feedback in triggering (§3.1), policy conflicts in decision (§3.2), and signaling loss/error in execution (§3.3). To achieve reliable extreme mobility, a fundamental solution is to shift to more dependable criteria.

Therefore, REM shifts from indirect wireless signal strength-based to direct *movement-based* mobility. Intuitively, the client movement decides its physical multi-paths and Doppler effect for each cell, thus impacting the wireless quality. Compared to wireless with short coherence and dramatic dynamics (§2), the client movement is slower and predicable by inertia, thus more reliable to drive the extreme mobility management. To this end, REM tracks the client movement in the *delay-Doppler domain*. With this knowledge, REM relaxes the feedback's exploration-exploitation dilemma in triggering phase, simplifies the policies in decision phase, and stabilizes the signaling traffic in execution phase.

Delay-Doppler domain: A wireless channel decides how radio signals from the sender propagates along multiple physical paths, and combines at the receiver. A time-varying channel can be characterized in multiple ways. 4G/5G measures its OFDM channel in the time-frequency domain: An OFDM channel is defined as a function of time and carrier frequency $H(t, f)$. Equivalently, we can represent the same channel in the *delay-Doppler domain* [19]:

$$h(\tau, v) = \sum_{p=1}^{P} h_p \delta(\tau - \tau_p)\delta(v - v_p) \qquad (1)$$

where P is the number of paths (direct, reflected, and scattered ones), h_p, τ_p, v_p are p-th path's complex attenuation, propagation delay (distance) and Doppler frequency shift, and δ is the Dirac delta function. Figure 6a exemplifies a channel with 3 paths. The delay-Doppler form reflects the *multi-path geometry* between cell

and client in movement. Given $h(\tau, v)$ and a sent signal $s(t)$, the received signal $r(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau, v)s(t - \tau)e^{j2\pi vt}d\tau dv$. The OFDM channel $H(t, f)$ and delay-Doppler channel $h(\tau, v)$ are related by

$$H(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau, v)e^{j2\pi(tv-f\tau)}d\tau dv = \sum_{p=1}^{P} h_p e^{j2\pi(tv_p-f\tau_p)}$$

Compared to $H(t, f)$, delay-Doppler representation $h(\tau, v)$ is more stable since its variance relates to slower path delay and Doppler change [3, 20, 21] (see Appendix A for an analysis).

Why delay-Doppler domain: The delay-Doppler domain unveils client movement and multi-path propagation $\{h_p, \tau_p, v_p\}$. Mobility management on top of it can benefit in all its phases:

• *Triggering: Relaxed reliance on feedback.* Movement-based feedback allows fast and reliable triggering with relaxed exploration-exploitation (more measurements v.s. timely triggering) tradeoff. Cells from the same base station share the physical propagation paths to the client. Instead of measuring all cells sequentially, the client only measures one cell and performs *cross-band estimation* to others from the same location. This accelerates the feedback without reducing the cells to be explored.

• *Decision: Simplified, conflict-free policy.* The decision policy in the delay-Doppler domain can be simplified for two reasons. First, by replacing the inter-frequency measurement with cross-band estimation, the tradeoff between decision latency and spectral efficiency is bypassed. This eliminates the need for multi-stage policy (§3.2). Second, it reduces configurations (A1, A2, A4, A5) for heterogeneous cells that share the multipath, thus reducing the conflicts.

• *Execution: Stabilized signaling.* Similar to 4G/5G OFDM, we can represent, modulate, and transfer signals in the delay-Doppler domain. Compared to OFDM, the delay-Doppler signal transfer is directly coupled with the slowly-varying multi-path evolution. So it will exploit the full time-frequency diversity, and therefore experience more stable channels and less loss/corruption. This mitigates failures from signaling/feedback loss or corruption.

REM roadmap: REM devises a signaling overlay in delay-Doppler domain with the recently proposed OTFS modulation [3]. REM further greatly extends OTFS to refine all phases of mobility management. Figure 5 overviews REM's main components.

• **Delay-Doppler signaling overlay (§5.1):** REM places the signaling traffic and reference signals in an delay-Doppler domain overlay. This overlay runs on top of existing OFDM, without changing 4G/5G designs or data traffic. It stabilizes the signaling in execution (§3.3), and exposes movement information to later phases.

• **Relaxed reliance on feedback (§5.2):** To mitigate the failures from slow and unreliable feedback (§3.1), REM devises cross-band estimation in the delay-Doppler domain. This approach accelerates the feedback without reducing the cells to be explored, and facilitate earlier handovers with less failures.

• **Simplified, conflict-less policy (§5.3):** To eliminate policy conflicts and failures from missed cells (§3.2), REM simplifies the policy in the delay-Doppler domain. It eliminates the multi-stage decision with cross-band estimation, reduces the configurations, and enables easy-to-satisfy conditions for the conflict-freedom.

5 THE REM DESIGN

We next elaborate each component in REM.

5.1 Delay-Doppler Signaling Overlay

REM runs its mobility management in delay-Doppler domain. To achieve so, REM should place its signaling traffic (e.g., measurement feedback, handover commands, reference signals) and modules (triggering, decision, execution) in this domain. We prefer to do so without changing existing 4G/5G designs or affecting OFDM-based data transfer. To this end, REM leverages recent advances in OTFS in delay-Doppler domain, builds a signaling overlay atop OFDM, extends OTFS with adaptive scheduling to enable the co-existence of OTFS signaling and OFDM data, and uses it to mitigates failures from signaling loss/corruption in execution (§3.3).

Delay-Doppler overlay with OTFS: OTFS is a modulation in the delay-Doppler domain. Intuitively, OTFS couples information with the multi-path geometry, modulates signals in the delay-Doppler domain, and multiplexes signals across all the available carrier frequencies and time slots. By exploiting full time-frequency diversity, signals enjoy similar channels with less variance, become robust to Doppler shifts and less vulnerable to loss and errors.

Figure 6a shows the OTFS modulation. It runs on top of OFDM. The OFDM time-frequency domain is discretized to a $M \times N$ grid (each being a 4G/5G radio resource element) by sampling time and frequency axes at intervals T and Δf[7], respectively. The modulated OFDM samples $X[n, m]$ are transmitted for a duration of NT and bandwidth of $M\Delta f$. Given a $M \times N$ time-frequency domain, the delay-Doppler domain is also a $M \times N$ grid ($\frac{k}{M\Delta f}, \frac{l}{NT}$), where $k = 0..M - 1, l = 0..N - 1$ where $\frac{1}{M\Delta f}$ and $\frac{1}{NT}$ are the quantization steps of path delay and Doppler frequency, respectively. The OTFS modulator arranges MN data symbols in the delay-Doppler grid, denoted as $x[k, l]$. It then converts $x[k, l]$ to $X[n, m]$ in OFDM using the discrete Symplectic Fourier transform (SFFT)

$$X[n, m] = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} x[k, l] e^{-j2\pi(\frac{mk}{M} - \frac{nl}{N})} \qquad \text{(SFFT)} \quad (2)$$

$$x[k, l] = \frac{1}{NM} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X[n, m] e^{j2\pi(\frac{mk}{M} - \frac{nl}{N})} \qquad \text{(ISFFT)} \quad (3)$$

The OFDM signal $X[n, m]$ is transmitted via legacy 4G/5G radio. The received signal $Y[n, m]$ is in the time-frequency domain. Then inverse SFFT (ISFFT) in (3) is applied to $Y[n, m]$ and yields $y[k, l]$ in the delay-Doppler domain. With channel noises, we have [3, 22]

$$y[k, l] = \frac{1}{NM} \sum_{k'=0}^{M-1} \sum_{l'=0}^{N-1} h_w(k'\Delta\tau, l'\Delta v) x[k - k', l - l'] + n[k, l] \quad (4)$$

where $h_w(\tau, v) = \int\int e^{-j2\pi\tau'v'} h(\tau', v') w(v - v', \tau - \tau') d\tau' dv'$ is the convolution of channel $h(\tau', v')$ and rectangular signal window: $w(\tau, v) = \sum_{c=0}^{N-1} \sum_{d=0}^{M-1} e^{-j2\pi(vcT - \tau d\Delta f)}$, $n(k, l) = ISSFT(N[n, m])$ is ISFFT of time-frequency noises. Compared to OFDM channel $H(t, f)$ with short coherence T_c, the delay-Doppler channel $h_w(\tau, v)$ is invariant of multi-path fading or inter-carrier interference from Doppler shift, thus more stable and reliable in a longer period.

Challenge: Coexistence with OFDM data   REM only adopts delay-Doppler domain for its signaling traffic. We are neutral to if
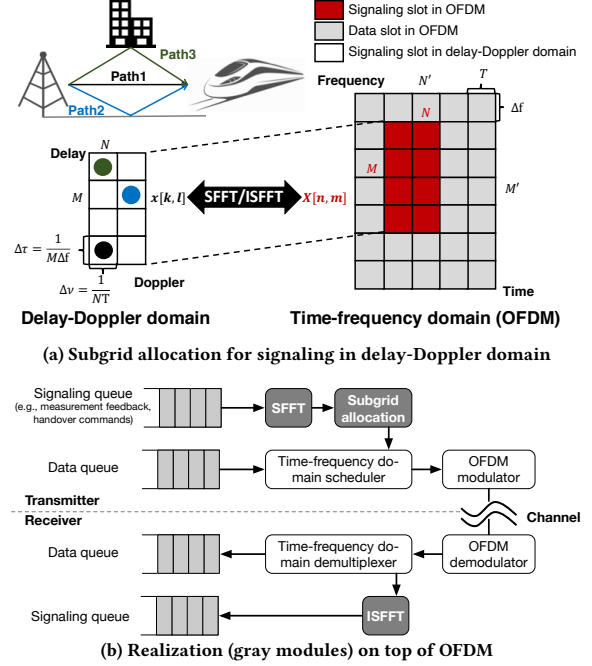
(a) Subgrid allocation for signaling in delay-Doppler domain

(b) Realization (gray modules) on top of OFDM

Figure 6: Signaling overlay in delay-Doppler domain.

data traffic should also use OTFS. While OTFS can help data combat Doppler shifts, it also incurs more data processing delays and may not be preferred by latency-sensitive scenarios. Instead, REM supports hybrid mode between OTFS-based signaling and OFDM/OTFS-based data. It offers flexibility for operators with both choices.

The challenge for this hybrid mode is that, to function correctly, OTFS requires a continuous $M \times N$ OFDM grid. But in 4G/5G, the signaling and data traffic are multiplexed in the OFDM grid. In case data still uses OFDM, the signaling traffic may span on *disjoint* OFDM slots, and cannot run OTFS directly. A possible solution is to define separated data and signaling grids, which however may waste the radio resource and needs 4G/5G physical layer redesign.

Our solution: Scheduling-based OTFS   To address this, we note *the 4G/5G signaling traffic is always prioritized in scheduling and delivery by design* [4, 5]. Before successful signaling procedures, the data traffic may not be correctly delivered or processed. So given pending signaling traffic, the base station will always schedule the radio resource and deliver the signaling traffic first, regardless of if any data is waiting. REM leverages this readily-available feature to allocate a *sub-grid* for OTFS-based signaling traffic first. It decouples OTFS-based signaling and OFDM-based data for co-existance, without changing the 4G/5G design or adding delay/spectral cost.

Figure 6b illustrates REM's ultimate signaling overlay. At the transmitter (base station for downlink and client for uplink), the overlay modulates the signaling traffic and reference signals with SFFT, and forwards them to the signaling radio bearer for traffic scheduling. Given the signaling traffic, the scheduler will always process them first by design. To ensure the applicability of OTFS, REM adapts the scheduler to guarantee that, all signaling traffic is always placed in a $M \times N$ subgrid of the 4G/5G resource grid ($M \le M', N \le N'$). On receiving these signaling, the receiver demodulates them in OFDM, runs REM's overlay to further demodulate in OTFS, and then forwards to upper layer for further mobility actions.
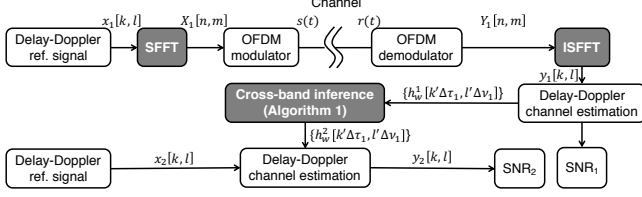
Figure 7: REM's cross-band channel estimation. Gray boxes are additional modules to OFDM today.

Overhead for signaling: REM adds the SFFT/ISFFT to pre/post-process the signaling traffic, with the complexity of $O(MNlog(MN))$. Such complexity is similar to 4G/5G uplink's SC-FDMA on top of OFDM (with additional fast Fourier transform). No additional delays, spectral waste or other overhead is incurred for the data traffic.

5.2 Relaxed Reliance on Feedback

With the delay-Doppler overlay, REM relaxes the handover's reliance on the feedback for fast and satisfactory triggering (§3.1). To achieve so, the key is to relax the unique dilemma in extreme mobility, between *exploration* of more measurements for satisfactory triggering and *exploitation* for fast triggering. We observe that, cells from the same base station share the multi-paths to the client and thus similar channels in the delay-Doppler domain (§3.1). To this end, REM devises *cross-band estimation* to parallelize the feedback: It measures *one only* cell per base station, extracts the multi-path profile from this measurement, maps it to other cells from the same base station, and estimates these cells' qualities *without* measurements. This allows the serving cell to make decisions *without* waiting for all feedback and triggering intervals in §3.1.

Existing cross-band estimations: Cross-band estimation is recently proposed in [23–25] to save the channel feedback overhead. Existing solutions are designed in the time-frequency domain and primarily for static scenarios. The idea is to extract the multi-path profiles (path delay, attenuation, phase, etc) from one band's channel estimation, and map it to another band traversing the same paths. In the time-frequency domain, the channel $H(f, t)$ differs among frequency bands, and does not reveal path parameters. So [24, 25] estimate the multi-path profile with non-linear optimization or machine learning. Unfortunately, these approaches face two fundamental limitations in extreme mobility. First, they do not consider the Doppler effect in mobility. Second, their optimization and machine learning are too slow to track the fast-varying channel dynamics (§7.2). The hardware acceleration with GPU, FPGA, or multi-core CPU could help. But such hardware is too expensive for the resource and energy-constrained mobile devices.

REM's intuition: To overcome these limitations, REM generalizes and simplifies the cross-band estimation in the delay-Doppler domain. Compared to the time-frequency domain representation $H(t, f)$, the delay-Doppler domain representation $h(\tau, \nu)$ in Equation (1) directly unveils the multi-path profiles $\{h_p, \tau_p, \nu_p\}$ and is more feasible for cross-band estimation. Besides, $h(\tau, \nu)$ evolves slower than $H(t, f)$ (§4), thus reducing frequent feedback and facilitating shorter triggering interval. With the delay-Doppler domain, REM can tackle the Doppler shift in extreme mobility, and eliminates the optimization and machine learning in existing solutions.

Specifically, consider two cells from the same base station. Given cell 1's channel estimation $\{h_w^1(k\Delta\tau, l\Delta\nu)\}_{k,l}$, REM estimates cell 2's

channel $\{h_w^2(k\Delta\tau, l\Delta\nu)\}_{k,l}$ *without* measuring it. To do so, REM first extracts multi-path profile $\{h_p, \tau_p, \nu_p^1\}$ from cell 1 $\{h_w^1(k\Delta\tau, l\Delta\nu)\}_{k,l}$. Note that the path delays τ_p and attenuations h_p are *frequency-independent*, thus identical for cell 1 and 2. The Doppler shifts of cell 1 ν_p^1 and cell 2 ν_p^2 are *frequency-dependent* and $\nu_p^1 \neq \nu_p^2$. But they are correlated by $\nu_p^1/\nu_p^2 = f_1/f_2$ (§2). So with cell 1's multi-path profile, we can estimate cell 2 by reusing $\{h_p, \tau_p\}$ and deriving $\{\nu_p^2\}$ from ν_p^1.

REM's cross-band estimation: REM first estimates cell 1's channel in the delay-Doppler domain. With its signaling overlay (§5.1), REM reuses 4G/5G's reference signals[8] but pre/post-process them in the delay-Doppler domain (Figure 7). By comparing received and constant sent reference signal $(y(k, l), x(k, l))$, we can estimate the delay-Doppler channel $\{h_w(k\Delta\tau, l\Delta\nu)\}_{k,l}$ by applying standard channel estimation [26] to OTFS's input-output relation in (4).

Now consider two cells from the same base station. Given cell 1's channel estimation $\{h_w^1(k\Delta\tau, l\Delta\nu)\}_{k,l}$, REM estimates cell 2's channel $\{h_w^2(k\Delta\tau, l\Delta\nu)\}_{k,l}$. We note channel estimation in (4) has

$$\frac{1}{MN}h_w(k\Delta\tau, l\Delta\nu) = \sum_{p=1}^{P} \frac{\Gamma(k\Delta\tau, \tau_p)}{M} \cdot h_p e^{-j2\pi\tau_p\nu_p} \cdot \frac{\Phi(l'\Delta\nu, \nu_p)}{N}$$

(5)

where we have $\Gamma(k\Delta\tau, \tau_p) = \sum_{d=0}^{M-1} e^{j2\pi(k\Delta\tau-\tau_p)d\Delta f}$, $\Phi(l\Delta\nu, \nu_p) = \sum_{c=0}^{N-1} e^{-j2\pi(l\Delta\nu-\nu_p)cT}$. We can rewrite it in a matrix form:

$$\mathbf{H} = \mathbf{\Gamma P \Phi} \quad (6)$$

where $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the channel estimation matrix from (4): $H(k, l) = \frac{1}{MN}h_w(k\Delta\tau, l\Delta\nu)$.

$$\mathbf{H} = \frac{1}{MN}\begin{bmatrix} h_w(0,0) & \cdots & h_w(0,(N-1)\Delta\nu) \\ h_w(\Delta\tau,0) & \cdots & h_w(\Delta\tau,(N-1)\Delta\nu) \\ \cdots & \cdots & \cdots \\ h_w((M-1)\Delta\tau,0) & \cdots & h_w((M-1)\Delta\tau,(N-1)\Delta\nu) \end{bmatrix}$$

$\mathbf{\Gamma} \in \mathbb{C}^{M \times P}$ is the *frequency-independent* path delay spread matrix from Equation 5: $\Gamma(k, p) = \frac{\Gamma(k\Delta\tau, \tau_p)}{M}$,

$$\mathbf{\Gamma} = \frac{1}{M}\begin{bmatrix} \Gamma(0,\tau_1) & \Gamma(0,\tau_2) & \cdots & \Gamma(0,\tau_P) \\ \Gamma(\Delta\tau,\tau_1) & \Gamma(\Delta\tau,\tau_2) & \cdots & \Gamma(\Delta\tau,\tau_P) \\ \cdots & \cdots & \cdots & \cdots \\ \Gamma((M-1)\Delta\tau,\tau_1) & \Gamma((M-1)\Delta\tau,\tau_2) & \cdots & \Gamma((M-1)\Delta\tau,\tau_P) \end{bmatrix}$$

$\mathbf{\Phi} \in \mathbb{C}^{P \times N}$ is the *frequency-dependent* path Doppler spread matrix with $\Phi(p, l) = \frac{\Phi(l\Delta\nu, \nu_p)e^{-j(\theta_p+2\pi\tau_p\nu_p)}}{N}$, θ_p is the frequency-independent path phase: $h_p = |h_p|e^{-j\theta_p}$.

$$\mathbf{\Phi} = \frac{1}{N}\begin{bmatrix} \Phi(0,\nu_1)e^{-j(\theta_1+2\pi\tau_1\nu_1)} & \cdots & \Phi((N-1)\Delta\nu,\nu_1)e^{-j(\theta_1+2\pi\tau_1\nu_1)} \\ \Phi(0,\nu_2)e^{-j(\theta_2+2\pi\tau_2\nu_2)} & \cdots & \Phi((N-1)\Delta\nu,\nu_2)e^{-j(\theta_2+2\pi\tau_2\nu_2)} \\ \cdots & \cdots & \cdots \\ \Phi(0,\nu_P)e^{-j(\theta_P+2\pi\tau_P\nu_P)} & \cdots & \Phi((N-1)\Delta\nu,\nu_P)e^{-j(\theta_P+2\pi\tau_P\nu_P)} \end{bmatrix}$$

and $\mathbf{P} \in \mathbb{R}_{\geq 0}^{P \times P}$ is the *multi-path attenuation* diagonal matrix:

$$\mathbf{P} = \begin{bmatrix} |h_1| & 0 & \cdots & 0 \\ 0 & |h_2| & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & |h_P| \end{bmatrix}$$

Given the cell 1's channel estimation matrix $\mathbf{H_1}$, if we can decompose it as $\mathbf{H_1} = \mathbf{\Gamma P \Phi_1}$, then the frequency-independent path delay $\mathbf{\Gamma}$ and attenuation $\mathbf{P}$ can be directly reused by cell 2, while the frequency-dependent Doppler shift $\mathbf{\Phi_2}$ can be derived from $\mathbf{\Phi_1}$ since $\frac{\nu_p^1}{\nu_p^2} = \frac{f_1}{f_2}$. Then we can obtain cell 2's channel $\mathbf{H_2} = \mathbf{\Gamma P \Phi_2}$.

---

[8]The cell-specific reference signals in 4G LTE, and CSI-RS in 5G NR [13]. Both are decoupled from demodulation reference signals for data transfer.

So how to decompose the delay-Doppler channel matrix $\mathbf{H}_1 = \mathbf{\Gamma P \Phi_1}$? It turns out that, such decomposition can be approximated by the classical singular value decomposition (SVD) [27]. Recall that SVD can factorize *any* matrix $\mathbf{H} \in \mathbb{C}^{M \times N}$ into two unitary matrices and a diagonal matrix: $\mathbf{H} = \mathbf{U \Sigma V}$, where $\mathbf{U} \in \mathbb{C}^{M \times M}$ is a unitary matrix with $\mathbf{UU}^* = \mathbf{I_M}$, $\mathbf{V} \in \mathbb{C}^{N \times N}$ is a unitary matrix with $\mathbf{VV}^* = \mathbf{I_N}$, and $\mathbf{\Sigma} \in \mathbb{R}_{\geq 0}^{M \times N}$ is a diagonal matrix with non-negative real numbers on the diagonal (i.e., singular values). Intuitively, SVD factorizes a matrix into two orthonormal bases $\mathbf{U}$ (for each row) and $\mathbf{V}$ (for each column), and attenuation $\mathbf{\Sigma}$. In practice, to reduce matrix dimensionality, SVD typically keeps the major singular values ("principle components") and truncate negligible ones. In this way, SVD approximates a matrix as $\mathbf{H} \approx \mathbf{U' \Sigma' V'}$ where $\mathbf{U'} \in \mathbb{C}^{M \times P}$, $\mathbf{\Sigma'} \in \mathbb{C}^{P \times P}$, and $\mathbf{V'} \in \mathbb{C}^{P \times N}$ with smaller matrix dimension $P \leq \min(M, N)$. This form is the same as our delay-Doppler channel decomposition $\mathbf{H} = \mathbf{\Gamma P \Phi}$. In fact, we can prove their relation as follows (proved in Appendix B):

Theorem 1 (Cross-band estimation with SVD). *A delay-Doppler decomposition* $\mathbf{H} = \mathbf{\Gamma P \Phi}$ *is also a singular value decomposition if (i) the number of physical paths* $P \leq \min(M, N)$*; and (ii) for any two paths* $p \neq p'$*, we always have* $\tau_p - \tau_{p'} = k\Delta\tau$ *and* $\nu_p - \nu_{p'} = l\Delta\nu$ *for some non-zero integer* k, l*.*

In reality, we note condition (i) almost always holds. It has been shown the real 4G/5G channels have sparse multi-paths[28–30][9]. Condition (ii) also approximately holds in reality: With 40ms triggering interval for a 20MHz channel (§3.1), $(M, N) = (1200, 560)$ and the wavelength is $c/f \approx 15$m. In the high-speed rails, the line-of-sight distance between the base station and the train is approximately multiple times of 15m (typically between 80m and 550m [33]). The non-line-of-sight reflection/scattering propagation paths are even longer. So such (M, N) results in fine-grained delay/Doppler sampling $(\Delta\tau, \Delta\nu)$ and approximates condition (ii).

Algorithm 1 shows REM's cross-band estimation via SVD. Given cell 1's channel estimation matrix $\mathbf{H}_1$, we run SVD and use it as an approximation of $\mathbf{H}_1 = \mathbf{\Gamma P \Phi_1}$ (line 1). Note cell 1's $\mathbf{\Gamma P}$ is frequency-independent and can be reused by cell 2. To estimate cell 2, we need to infer $\mathbf{\Phi_2}$ from $\mathbf{\Phi_1}$. To this end, Algorithm 1 estimates multipath profile $\{h_p, \tau_p, \nu_p^2\}_{p=1}^{P_{max}}$ (line 2–8) based on the derivations in Appendix C. Then Algorithm 1 re-constructs $\mathbf{\Phi_2}$ and estimates cell 2 as $\mathbf{H}_2 = \mathbf{\Gamma P \Phi_2}$. Algorithm 1 supports multi-antenna systems such as MIMO and beamforming, by running it on each antenna.

Complexity: REM's runs SFFT/ISFFT to process the reference signals and Algorithm 1 for cross-band estimation. Both have polynomial complexity: The SFFT/ISFFT complexity is $O(MN \log MN)$, and Algorithm 1's complexity is $O(\min(M, N) \max(M, N)^2)$. It is faster than [24, 25] that rely on optimization or machine learning, thus suitable to track the fast-varying channel in extreme mobility.

The impact of channel noises: The noises impacts channel estimation accuracy and indirectly affects cross-band estimation. REM is robust to noises since it runs in the delay-Doppler domain. According to (4), the noise in the time-frequency domain $N[n, m]$ is smoothed to $n[k, l]$ in the delay-Doppler domain via IFFT. For typical 4G/5G noises, this results in more robust channel estimation

---

Algorithm 1 REM's cross-band channel estimation

Input: Band 1's channel estimation matrix $\mathbf{H}_1$, $H_1(k, l) = h_w^1(k\Delta\tau, l\Delta\nu)$ from (4)
Output: Band 2's channel estimation matrix $\mathbf{H}_2$
1: Decompose $\mathbf{H}_1 = \mathbf{\Gamma P \Phi_1}$ using SVD matrix factorization;
2: **for** each path $p = 1, 2, \dots \min(M, N)$ **do**
3:     For any $\forall l, l' \neq l \in [0, N-1]$ and $\forall k, k' \neq k \in [0, M-1]$;
4:     $\nu_p^1 \leftarrow e^{-j2\pi \nu_p^1 T} = \frac{1}{N(N-1)} \sum_{l, l'} \frac{\Phi_1(p, l) - \Phi_1(p, l')}{\Phi_1(p, l) e^{j2\pi l \Delta\nu T} - \Phi_1(p, l') e^{j2\pi l' \Delta\nu T}}$;
5:     $\tau_p \leftarrow e^{j2\pi \tau_p \Delta f} = \frac{1}{M(M-1)} \sum_{k, k'} \frac{\Gamma(k, p) - \Gamma(k', p)}{\Gamma(k, p) e^{-j2\pi k \Delta\tau \Delta f} - \Gamma(k', p) e^{-j2\pi k' \Delta\tau \Delta f}}$;
6:     $\nu_p^2 \leftarrow \nu_p^1 \frac{f_2}{f_1}$;   ▷ *Transfer to band 2's Doppler frequency*
7:     $e^{-j\theta_p} \leftarrow \frac{1}{N} \sum_l \frac{\Phi(p, l) N}{\Phi(l\Delta\nu, \nu_p) e^{-j2\pi \tau_p \nu_p}}$;
8: **end for**
9: Compute $\mathbf{\Phi_2}$ with $\{h_p, \tau_p, \nu_p^2\}_p$;
10: $\mathbf{H}_2 \leftarrow \mathbf{\Gamma P \Phi_2}$;

---

for h_w and thus decomposition. REM may be less robust if the OFDM noises are carefully crafted (e.g., spamming attack), so that the channel estimation is inaccurate. Both OFDM and REM would be affected then, and REM is no worse than OFDM in terms of reliability.

5.3 Simplified, Conflict-Free Policy

REM last simplifies the handover policy for high reliability and verifiable correctness (§3.2). Our goal is to: (1) avoid multi-stage policy whenever possible, without missing cells or delaying handovers; and (2) eliminate policy conflicts in extreme mobility. Meanwhile, REM still retains flexibility for operators to customize their policies.

Extreme mobility policy in delay-Doppler domain: Compared to the complex policy today, extreme mobility policy in delay-Doppler domain can be simplified for three reasons:

(1) *Bypassed the latency-spectral efficiency tradeoff:* As shown in §3.2, multi-stage policy is common today to balance the spectral efficiency and decision latency for inter-frequency cells. This is mostly unnecessary with REM's cross-band estimation in §5.2. Inter-frequency cells can be inferred from intra-frequency cells at the location, *without* extra round trips or allocating radio resource.

(2) *Coherent, stable decision metric:* Delay-Doppler domain enables more stable channel and signal-noise-ratio (SNR), and makes SNR-based handover feasible[10]. This benefits policy simplification with less events (Table 1). In signal strength-based 4G/5G mobility, A4 is used for load balancing and A5 is for indirect signal strength comparison between heterogeneous cells (§3.2). These events are not "must-haves" if SNR is used, since SNRs between cells are directly comparable and decide the capacity $C = B \log(\text{SNR} + 1)$ (B is the bandwidth) based on information theory.

(3) *Reduced demand for proactive policies:* In extreme mobility, the policy conflicts are amplified by operators' demand for proactive failure mitigation (§3.2). In delay-Doppler domain, this demand can be satisfied by REM instead (§5.1–§5.2), thus eliminating the need for conflict-prone proactive policies.

REM's simplification approach: Figure 8 exemplifies how REM simplifies an extreme mobility policy today in four steps:

(1) *Replace received signal strength with delay-Doppler SNR.* This helps stabilize the input and simplifies events needed. Note SNR should always be evaluated in handover, regardless of other metrics to be used. Otherwise, "blind handovers" will *always* happen with loops [14], and lose network access if target cell's coverage is weak;

---

[9]In 4G/5G, even the smallest OFDM resource block has $M = 12$, $N = 14$ and thus can support up to 12 paths. This suffices for standard reference multi-path models in 4G (7–9 paths depending on the scenario [31]) and 5G (12 paths [32]).

[10]In theory, 4G/5G OFDM could also use SNR for handover. But this is rare (if not non-existent) since OFDM SNR fluctuates rapidly and causes frequent oscillations (§3.1). Instead, 4G/5G decides handover using stabler signal strength [4, 5, 8–10, 34].
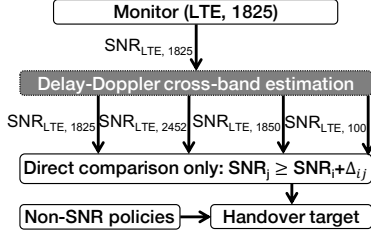
Figure 8: REM's policy simplification for Figure 1b.

(2) Replace multi-stage policy with cross-band estimation. If inter-frequency cells are co-located with intra-frequency ones, REM replaces A1/A2-based multi-stage policy with cross-band estimation in §5.2. This avoids missing cells and bypasses the tradeoff between latency and spectral cost for inter-frequency cells. Otherwise, REM retains the multi-stage policy and moves to next step (but still with the same conflict-freedom guarantees below).

(3) Remove unnecessary events in policy. By removing the multi-stage decision, A1/A2 events are removed. For other events, REM replaces them with A3. For each A5 event today, REM replaces it with an equivalent A3 with $\Delta_{A3} = \Delta_{A5}^2 - \Delta_{A5}^1$, since A5 $R_s < \Delta_{A5}^1, R_n > \Delta_{A5}^2$ implies $R_n > R_s + \Delta_{A5}^2 - \Delta_{A5}^1$. To remove A4, there are two cases in extreme mobility. First, due to multi-stage policy, most A4 events occur after A2 is triggered. They are equivalent to A5 with $\Delta_{A5}^1 = \Delta_{A2}, \Delta_{A5}^2 = \Delta_{A4}$ and replaced by A3 with above procedure. Second, for load balancing or adding capacity [8, 10, 34], a small amount of A4 events are directly triggered without A2 (§3.2). They can also be replaced by A3: The serving cell can equally find a cell with less load or more capacity using A3 comparison on $C = B\log(SNR+1)$, where Δ_{A3} decides capacity difference. Afterwards, REM only regulates A3 for conflict freedom as detailed below.

(4) Retain remaining policies: A cell may decide handovers based on other metrics, such as priorities, traffic load, and access control. REM keeps them without changes, and retains flexibility for operators.

REM's simple conflict-freedom guarantees: Compared to today's policies in §3.2, REM eliminates most events except A3. This leads to less conflicts between events, and simpler conflict resolutions than [14, 15]. We start with the policy with delay-Doppler SNR only. We obtain the following result (proved in Appendix D):

Theorem 2 (Conflict-freedom with delay-Doppler SNR only). *When only delay-Doppler SNR is used in* REM*'s simplfied policy, no persistent loops will occur if and only if between any two cells c_i and c_j that cover the same area, $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} \geq 0$.*

Theorem 2 shows that, two-cell threshold coordination is necessary *and* sufficient condition for policy conflict freedom. Compared to the conflict freedom conditions today [14, 15], Theorem 2 is much simpler with less events and threshold coordination between cells. Violation of Theorem 2 happens in extreme mobility when operator tries proactive handovers to mitigate failures (§3.2). With REM, operators do not need this since REM has mitigated most failures.

We next show that, even with other criteria (preferences, load balancing, access control, etc), Theorem 2 is still sufficient (though not necessary) for conflict freedom.

Theorem 3 (Conflict-freedom in general). *For any settings of non-SNR metrics in* REM*, satisfying Theorem 2 still guarantees loop-freedom.*

Theorem 3 is proved in Appendix E. Intuitively, with coordinated SNR events, Theorem 2 ensures handovers between cells will not be simultaneously satisfied. Regardless of other policies, this condition suffices for conflict freedom. This simplifies the policy configurations with provable conflict freedom.

6 IMPLEMENTATION

We implement REM on Ettus USRP software-defined radio running OpenAirInterface [35] software cellular stack, with one emulating a client and another as a base station. REM is realized as a signaling overlay between LTE physical layer and radio resource control (RRC) protocol [4, 5] in the client and base station. Our implementation is backward compatible: If the client *or* base station does not support REM, both disable REM overlay and rollback to 4G/5G.

• **Delay-Doppler signaling overlay (§5.1):** We realize it on both the client and base station. In 4G/5G, the pending signaling messages are queued in the signaling radio bearer (SRB) at radio link control (RLC) layer [36, 37], We first estimate how many slots (thus subgrid size) they need by volume. Then we run OTFS modulation for them, and then forward them to medium access control (MAC) layer [38, 39]. We further adapt MAC's traffic scheduler to always place all signaling messages in a subgrid in OFDM to meet the OTFS requirement. All data traffic will not be affected since they are handled by the data radio bearers (DRBs) in RLC and scheduled with lower priority in MAC layer.

• **Relaxed reliance on feedback (§5.2):** The base station reuses 4G/5G reference signals and modulate them with OTFS. For the client, it first groups cells by their physical base stations based on the global cell identifiers ECI in 4G LTE [11] and NCGI in 5G NR [16].. Then it chooses one cell per base station to measure (intra-frequency cell if any, otherwise inter-frequency cell), estimate its delay-Doppler channel with standard procedure [26], runs Algorithm 1 to estimate other cells from the same base station, and reports them to the serving cell.

• **Simplified, conflict-free policy (§5.3):** The base station configures the client to measure all intra/inter-frequency cells' with A3 that meet Theorem 2 and 3, and disable other events (thus no multi-stage decision). The non-SNR policies (e.g., preferences and load balancing) remain unchanged.

7 EVALUATION

We evaluate REM's reliability in extreme mobility (§7.1), and its efficiency and overhead of its key components (§7.2).

Experimental setup: To approximate real extreme mobility, we run trace-driven emulations over USRP-based testbed.

• *Extreme mobility dataset:* Table 4 summarizes our datasets, including **(1) Fine-grained HSR dataset:** We collected it over Chinese high-speed rails in 07/2019–08/2019. We have tested a 1,136 km rail route at 200–300km/h between Beijing and Taiyuan, China. We run a Skype video call in Xiaomi MI 8 phone using China Telecom, and collect the full-stack 4G LTE signaling messages (PHY, MAC, RLC, RRC) using MobileInsight [8]. **(2) Coarse-grained HSR dataset:** We used an open dataset from [7] for larger-scale evaluations. This dataset is collected when the mobile client runs continuous downlink data transfer via TCP-based iperf over Beijing-Shanghai HSR route at 200/300/350 km/h. It includes 357.9 GB data by traveling

Table 4: Overview of extreme mobility datasets

		Low mobility	High-speed rails (China)	
		Los Angeles (Fine-grained)	Beijing-Taiyuan (Fine-grained)	Beijing-Shanghai [7] (Coarse-grained)
	Movement speed	0–100km/h	200–300km/h	200–350km/h
	Route distance	619 km	1,136 km	51,367 km
	Mobile operators	AT&T, T-Mobile, Verizon, Sprint	China Telecom	China Mobile, China Telecom
	# Signaling messages	46,814	49,781	601,720
Wireless	Carrier frequency	731.5–2648.6MHz	874.2–2120MHz	1835–2665MHz
	Bandwidth	5, 10, 20MHz	5, 10, 15, 20MHz	5, 10, 15, 20MHz
	Channel metrics for OFDM	SNR, BLER, CQI, MCS, RSRP, RSRQ	SNR, BLER, CQI, MCS, RSRP, RSRQ	RSRP, RSRQ
	RSRP range (dBm)	[−136, −44]	[−134, −59]	[−140, −60]
	SNR range (dB)	[−20, 30]	[−20, 30]	N/A (not collected)
Mobility	# Cells (base stations)	932 (503)	1,281 (878)	3,139 (1,735)
	# Feedback	4,023	3,588	81,575
	# Policy configurations	2,771	3,783	38,646
	# Handovers	1,157	2,030	23,779

51,367 km on the trains. Different from the fine-grained one, this dataset only has RRC messages, thus missing fine-grained OFDM channel information. Together with the LTE signaling messages, it also collects the tcpdump packet traces from the mobile client and server. **(3) Low mobility dataset:** It is our baseline. Since 02/2017, we have collected it with MobileInsight, by driving on highways in Los Angeles with AT&T, T-Mobile, Verizon, and Sprint.

• *Testbed:* Our testbed is based on §6. It consists of USRP B210/N210 as client and base stations, which connected to servers with Intel Xeon CPU E5-2420 v2 and 16GB memory. The servers run OAI [35] cellular protocol stack. To approximate operational settings, we configure the testbed's radio power, protocol configurations and mobility policies based on above datasets. We run USRP under the unlicensed 2412/2432MHz band instead of lincensed ones. To compare REM with legacy design, we replay our datasets and evaluate if REM can prevent failures in same settings.

Ethics: This work does not raise any ethical issues.

7.1 Overall Reliability in Extreme Mobility

We evaluate REM's reduction of network failures and policy conflicts in extreme mobility. To compare REM with legacy mobility management, we replay our datasets in Table 4, and evaluate how many failures/conflicts in Table 2 are reduced by REM. For each handover from our datasets, we extract its feedback and handover command , and infer its corresponding policies with the same approach in §3. Based on them, we configure our testbed with same policies, and adapt base stations' runtime transmission power of reference signals with same dynamics of signal strengths (RSRPs) and SNRs in datasets. We repeat this setup with/without REM overlay, and examine if this handover will succeed. To assess REM's benefits for end-to-end applications, we also replay the iperf's TCP data transfer in the tcpdump traces if the coarse-grained HSR dataset is used, and quantify their TCP performance with/without REM.

We compare REM and legacy LTE on failure ratios $\eta = \frac{K_{LTE}}{K}$ and reduction $\epsilon = \frac{K_{LTE}-K_{REM}}{K_{REM}}$, where K is total handover counts, and K_{LTE} (K_{REM}) is the total handover failure counts in LTE (REM). Since the failures occur randomly with wireless dynamics, we assess REM's *worst-case* failure reduction as a lower bound. For failures from signaling loss/corruption in §5.1–§5.2, we assume REM can prevent them only if it reduces the error rate to 0. This underestimates REM's failure reduction since signaling may be delivered with non-zero block error rate. For failures from missing cells in multi-stage policy in §5.3, the client will eventually reconnect to a missed candidate cell if its SNR is better than old cell (before which

(a) TCP stalling time   (b) TCP data transfer in failures (from dataset)
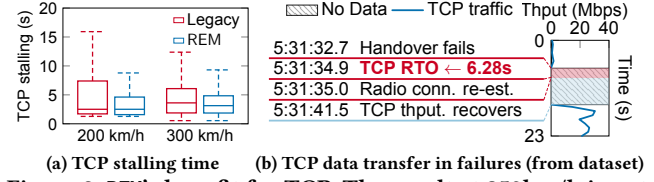
Figure 9: REM's benefit for TCP. The result at 350km/h is not shown since its LTE signaling messages and TCP traces were not simultaneously collected and evaluated.

the client has no service). We use this to detect if a cell is available but missed. Since SNR is not collected in Beijing-Shanghai dataset, we do not assess REM's failure reduction for missing cell and thus under-estimates its effectiveness.. Table 5 shows REM's reduction of network failures and policy conflicts, and Figure 9 shows REM's benefits for TCP and applications.

Overall reliability improvement: Table 5 shows REM reduces the overall failures and conflicts in both HSR datasets at all train speeds. In Beijing-Shanghai route, REM reduces existing LTE's failure ratio by 1.2× (5.2%→2.4%) at 100-200km/h, 3.0× (10.6%→2.6%) at 200−300km/h, and 2.6× (12.5%→3.5%) at 300-350km/h. In Beijing-Taiyuan route at 200−300km/h, REM the failure ratio by 0.9× (8.1% →4.2%). In all cases, REM achieves comparable failure ratios to static and low-speed mobility (e.g., driving in Table 2). Note all these failure ratios include the *unavoidable* failures from coverage holes, which can only be avoided with better coverage. Without coverage holes, REM achieves negligible failures (0.6%−1.1%) and failure reductions (3.9×−12.7×) by up to one order of magnitude.

Failure reduction in triggering: With the stabilized signaling (§5.1), REM reduces the feedback-induced failures to be negligible (0.1%−0.2%). Note failure reductions in decision and execution can also be indirectly related to faster feedback with cross-band estimation (§5.2). We currently classify them to later phases and are working on more accurate breakdown.

Failure/conflict reduction in decision: By eliminating the multi-stage policy, REM mitigates the failures from missed interfrequency cells (3× reduction in Beijing-Taiyuan dataset). With coarse-grained dataset, we cannot evaluate this benefit in Beijing-Shanghai route since no SNRs were collected by that dataset. So REM's failure reduction is under-estimated in this dataset. Moreover, with the simplified policy in §5.3, REM eliminates policy conflicts in all scenarios. While this also eliminates operators' proactive policies that try to prevent failures, such elimination will not negatively affect the failure mitigation with REM's failure reduction (§7.2).
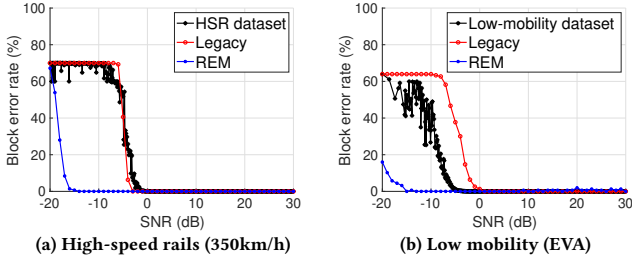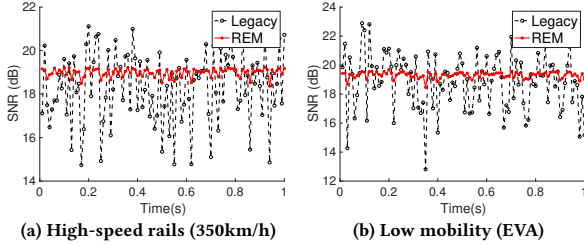
Failure reduction in execution: REM reduces its failures to 0– 0.4%. Our dataset shows many handover commands in OFDM-based LTE are corrupted/lost with acceptable SNR ([−5dB, 0dB]). Instead, REM explores the full frequency-time diversity in delay-Doppler domain to mitigates the signaling errors/corruptions.

On coverage holes: REM cannot reduce failures from coverage holes. After years of operation, HSRs have been mostly covered with more cells (thus <3.5% failures). Without coverage holes, REM achieves negligible failures (0.7%−1.1% depending on train speed) and more failure reductions (3.9×−12.7×).

Benefits for applications.   We last assess how REM benefits TCP and application data transfer. We define the TCP stalling time as the duration that a TCP connection cannot transfer data. With

Table 5: Reduction of failures and policy conflicts in high-speed rails (LGC=Legacy)
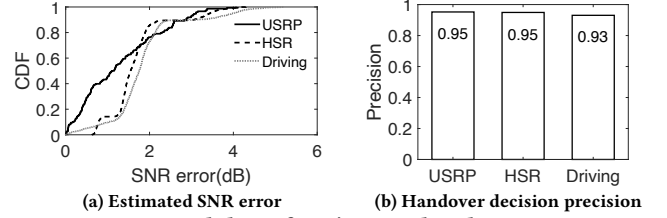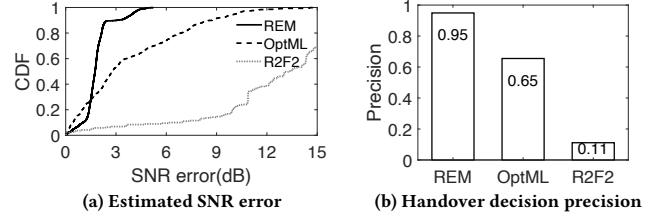
| | | Low mobility | | | Beijing-Taiyuan | | | Beijing-Shanghai | | | | | | | | |
| | | 0 − 100km/h | | | 200 − 300km/h | | | 100 − 200km/h | | | 200 − 300km/h | | | 300 − 350km/h | | |
		LGC	REM	ϵ	LGC	REM	ϵ	LGC	REM	ϵ	LGC	REM	ϵ	LGC	REM	ϵ
Failure	Total failure ratio η	4.3%	3.0%	0.43×	8.1%	4.2%	**0.9×**	5.2%	2.4%	**1.2×**	10.6%	2.63%	**3.0×**	12.5%	3.5%	**2.6×**
	Failure w/o coverage hole	3.2%	1.9%	0.68×	4.6%	0.7%	**5.6×**	3.4%	0.7%	**3.9×**	8.6%	0.63%	**12.7×**	10.1%	1.1%	**8.2×**
	Feedback delay/loss	0.78%	0.05%	14.6×	2.4%	0.1%	23×	1.7%	0.1%	16×	4.9%	0.2%	23.5×	6.9%	0.23%	29.0×
	Missed cell	1.8%	-	-	0.8%	0.2%	3×	0.6%	-	-	0.4%	-	-	0.8%	-	-
	Handover cmd. loss	0.61%	0.04%	14.2×	1.4%	0.4%	2.5×	1.1%	0	∞	3.3%	0.03%	109×	2.4%	0.03%	79.0×
	Coverage holes	1.1%	1.1%	0	3.5%	3.5%	0	1.7%	1.7%	0	2.0%	2.0%	0	2.4%	2.4%	0
Conflict	Total HO in conflicts	0.95%	0	∞	33.2%	0	∞	19.3%	0	∞	5.5%	0	∞	19.1%	0	∞
	Intra-frequency conflicts	0	0	0	31.2%	0	∞	18.2%	0	∞	5.5%	0	∞	12.7%	0	∞
	Inter-frequency conflicts	0.95%	0	∞	2.0%	0	∞	1.1%	0	∞	0	0	∞	6.4%	0	∞

(a) High-speed rails (350km/h)   (b) Low mobility (EVA)

Figure 10: REM's error reduction for signaling

(a) High-speed rails (350km/h)   (b) Low mobility (EVA)

Figure 11: Stabilized delay-Doppler domain.

(a) Estimated SNR error   (b) Handover decision precision

Figure 12: Viability of REM's cross-band estimation.

(a) Estimated SNR error   (b) Handover decision precision

Figure 13: Cross-band estimation with the HSR dataset.

the network failures, the radio connectivity is down and TCP data transfer is blocked. We replay the LTE signaling messages and packet traces in this dataset, and assess the TCP stalling time in legacy LTE and REM. Note in the coarse-grained HSR dataset, the iperf application at the client and server continuously generate data. So the TCP stalling will not be caused by the idle application or connection. Figure 9a shows REM's TCP stalling time reduction. With less failures, REM reduces the average TCP stalling from 7.9s to 4.2s at 200km/h, and from 6.6s to 4.5s at 300km/h. Note TCP stalling time is usually longer than the network failures because of its retransmission timeout (RTO). This is exemplified in Figure 9b: When network failure occurs, the TCP congestion control aggressively increases RTO for backoff, thus significantly delaying the data transfer. By reducing the failures in extreme mobility, REM mitigates such scenarios and benefits the applications' data transfer.

7.2 Efficiency and Overhead

Stabilized signaling in delay-Doppler domain (§5.1): We first examine how delay-Doppler domain helps reduces signaling errors/loss. We replay our datasets in Table 4 with same signaling message length and SNR, and evaluate their block error rate in a 4G/5G subframe ($M = 12, N = 14$ for 1ms [12, 13]) in standard reference multipath models for high-speed train and driving [40, 41]. Figure 10 confirms REM reduces errors by exploiting time-frequency diversity. This mitigates failures from signaling loss/corruption.

Besides less errors, delay-Doppler domain also facilitates more stable channels and SNRs. Figure 11 compares REM and legacy LTE's SNR in the same setting above. In OFDM, slots in different carrier frequency and time experience different channel gains $H(f, t)$ and thus diverse SNRs. Instead, REM adopts OTFS to spread signaling traffic across the entire time-frequency grid, explores the full frequency/time diversity and results in stable channel gains $h_w(\tau, v)$ for all slots in the grid (Equation 4). This results in more stable SNRs, facilitates SNR-based policy in REM and less transient loops.

Relaxed feedback (§5.2): We first explore whether REM retains accurate handover decisions by replacing directly measurements with cross-band estimation. With our dataset, we extract all handovers' measurements and triggering events/thresholds, run REM's cross-band estimation to estimate the target cell if it's co-located with another one, compare the estimated cell quality with the direct measurement, and evaluate whether REM's cell estimation can trigger the same events for handover. Figure 12 shows that, REM can achieve ≤2dB estimation errors for ≥90% measurements, and correctly triggers ≥90% handovers. To improve the correct triggering of handovers with cross-band estimation, the operator can further fine-tune its event thresholds (Table 1) to tolerate estimation errors.

We further compare REM's accuracy with R2F2 [24] and OptML [25], the state-of-the-art cross-band estimations. Note that R2F2 and OptML require to configure the maximum number of paths to be explored, which will affect their estimation accuracy. For fair comparison, we empirically find their optimal configuration (6 paths for both R2F2 and OptML), and show the results under this setting. Moreover, to train the OptML model, we randomly choose
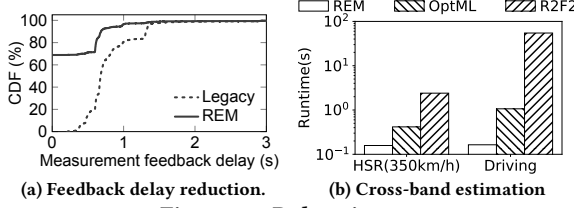
Yuanjie Li, Qianru Li, Zhehui Zhang, Ghufran Baig, Lili Qiu, Songwu Lu

(a) Feedback delay reduction.   **(b) Cross-band estimation**
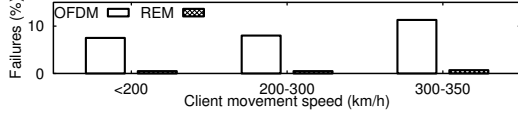
Figure 14: Delays in REM.

Figure 15: Failures without aggressive policies.

80% data from the HSR dataset, and use the remaining 20% data to test OptML. Figure 13 REM achieves 86.8% lower mean SNR error than R2F2, and 51.9% lower mean SNR error than OptML in the high-speed rail scenario. As explained in §5.2, this is because REM explicitly tackles the Doppler effect in extreme mobility.

We last quantify REM's acceleration for the feedback. For each saved measurement in above experiment, REM reduces its measurement durations (including the triggering interval in §3.1) and round-trips of sending this feedback (totally T_1). Meanwhile, REM incurs extra delay due to its runtime of cross-band estimation T_2, so the feedback latency savings is $T_1 - T_2$. Figure 14a shows REM reduces the average feedback latency from 802.5 ms to 242.4 ms. We also compare REM's runtime T_2 with state-of-the-arts under 4G/5G reference multi-path channels without Doppler (unsupported by R2F2/OptML). Figure 14b shows REM outperforms both, without optimization or machine learning. In the HSR, REM saves the runtime from 2.4s (416.3ms) in R2F2 (OptML) to 158.1ms, thus 14× (1.6×) reduction. While it is possible to accelerate R2F2 and OptML with advanced hardware (e.g., FPGA and GPU), such solution is too expensive for the resource and energy-constrained mobile devices.

Simplified, conflict-free policy (§5.3):   As shown in Table 5, REM's simplified policy provably prevent conflicts. Since operators adopt these conflict-prone policies for proactive failure mitigation (§3.2), one may wonder if eliminating the conflicts will cause more failures. We show REM prevents this situation. For all the conflict-prone handover events in our dataset, we follow Theorem 2 and 3 to update thresholds, and repeat the evaluation in §7.1 to evaluate if more failures will happen in REM. Figure 15 compares the failures (without coverage holes) after REM fixes conflicts. It shows that REM still retains negligible failures, since it prevents late handovers with faster feedback and signaling loss/corruption with delay-Doppler OTFS modulation. Both ensure operators do not need to rush the handovers when channel quality is still satisfactory.

8 DISCUSSION

Coverage holes and implementation issues: REM currently only mitigates failures with cell radio coverage. Otherwise, no network services exist and no solutions can prevent failures unless the coverage hole is fixed. Besides, the failures from the client/network implementation bugs is also beyond REM's scope.

On data speed: While primarily for reliability, REM also benefits data performance in general for three reasons. First, REM reduces failures and policy conflicts, thus avoiding serve performance downgrade. Second, REM's cross-band estimation saves MeasurementGap

for inter-frequency cells, thus offering more spectrum for data transfer. Last but not the least, if data also uses OTFS, REM's SNR-based policy also selects the cell with high capacity $C = B \log(\text{SNR} + 1)$. Theorem 2 and 3 still hold by replacing SNR with capacity.

Implications on IoT and edge: REM helps them simplify their application-layer operations. With REM, the IoT/edge will have a more stable network condition. This facilitates predictive solutions for IoT/edge to improve the quality-of-experiences (e.g., in virtual reality [42]) and saves signaling overhead (e.g., in massive IoT).

9 RELATED WORK

Reliable and fast mobility management has been an active topic for years. Most efforts follow the wireless signal strength-based design in today and explore how to refine its signaling procedures [43, 44], handover decision [42, 45], transport-layer data speed in mobility [7, 46], policy conflicts [14, 15], to name a few. Instead, REM revisits the foundations of wireless signal strength-based design, unveils diverse network failures and policy conflicts below the IP layer, and proposes a shift to movement-based reliable extreme mobility.

REM is inspired by prior efforts for refining wireless robustness, and generalizes them to mobility. It follows similar design philosophy to geographical routing [47–49], but in a different scenario in mobility management. REM leverages the delay-Doppler domain from the radar community and recent advances in OTFS modulation [3, 22, 50]. But REM moves beyond wireless modulation and generalizes to mobility management. REM's relaxed feedback in §5.2 extends the cross-band estimation in [24, 25] to mobility scenarios, and simplifies the estimation in the delay-Doppler domain.

10 CONCLUSION

Extreme mobility has become popular with various emergent high-speed mobility scenarios (rails, vehicles, drones, etc) and high-frequency radios (e.g., mmWave). Unfortunately, we show 4G/5G is not well prepared to support them. The fundamental problem is that, 4G/5G's *wireless signal strength-based* design is vulnerable to dramatic wireless dynamics in extreme mobility. We thus devise REM, a *movement-based* mobility management in delay-Doppler domain. REM relaxes the feedback with cross-band estimation, simplifies the policy for provable conflict-freedom, and stabilizes the critical signaling traffic scheduling-based OTFS modulation.

REM is an initial step toward movement-based mobile network design and management. Its core philosophy is *client movement is more robust and predictable than wireless*, thus suitable to drive mobility management in extreme mobility. Beyond reliability, this idea can be generalized to broader scopes such as channel prediction, wireless performance optimization, geographical routing, and delay-Doppler based localization. More client movement insights can be explored in the future, such as the predictive client trajectory (e.g., in rails and satellites), explicit sheer geometric modeling, and historical base station measurements. We hope REM could stimulate more efforts toward predictable, robust mobile networks.

REFERENCES

[1] Wikipedia. High-speed rail in China. https://en.wikipedia.org/wiki/High-speed_rail_in_China, 2019.

[2] Bloomberg. Vehicle-to-Everything Market Communications Ecosystem 2030 Account for a Market Worth $1.2 Billion by 2022. https://www.bloomberg.com/press-releases/2019-09-23/vehicle-to-everything-v2x-market-communications-ecosystem-2030-account-for-a-market-worth-1-2-billion-by-2022, 2019.

[3] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank. Orthogonal time frequency space modulation. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, March 2017.

[4] 3GPP. TS36.331: Radio Resource Control (RRC), Mar. 2015.

[5] 3GPP. TS36.331: 5G NR: Radio Resource Control (RRC), Jun. 2019.

[6] Wikipedia. Coherence time (communications systems). https://en.wikipedia.org/wiki/Coherence_time_(communications_systems), 2019.

[7] Jing Wang, Yufan Zheng, Yunzhe Ni, Chenren Xu, Feng Qian, Wangyang Li, Wantong Jiang, Yihua Cheng, Zhuo Cheng, Yuanjie Li, et al. An active-passive measurement study of tcp performance over lte on high-speed rails. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16. ACM, 2019.

[8] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. Mobileinsight: Extracting and analyzing cellular network information on smartphones. In *The 22nd ACM Annual International Conference on Mobile Computing and Networking (MobiCom'16)*, New York, USA, October 2016.

[9] ZTE Handover Description. https://tinyurl.com/o8enuz9.

[10] Huawei. Intra-RAT Mobility Management in Connected Mode Feature Parameter Description. https://www.honorcup.ru/upload/iblock/164/6.pdf, 2016.

[11] 3GPP. TS36.300: E-UTRA and E-UTRAN; Overall description; Stage 2, 2011.

[12] 3GPP. TS36.211: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, 2017.

[13] 3GPP. TS38.211: 5G NR; Physical channels and modulation, Jun. 2019.

[14] Yuanjie Li, Haotian Deng, Jiayao Li, Chunyi Peng, and Songwu Lu. Instability in distributed mobility management: Revisiting configuration management in 3g/4g mobile networks. In *The 42nd ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'16)*, Antibes Juan-les-Pins, France, June 2016.

[15] Zengwen Yuan, Qianru Li, Yuanjie Li, Songwu Lu, Chunyi Peng, and George Varghese. Resolving policy conflicts in multi-carrier cellular access. In *The 24th ACM Annual International Conference on Mobile Computing and Networking (MobiCom'18)*, New Delhi, India, October 2018.

[16] 3GPP. TS38.300: 5G NR: Overall description; Stage-2, Jan. 2020.

[17] ZTE and China Telecom: 5G network test on a high speed train. IEEE ComSoc Technology Blog. https://techblog.comsoc.org/2019/11/30/zte-and-china-telecom-5g-network-test-on-a-high-speed-train-uplink-enhancement-fast-verification/, Nov 2019.

[18] China's First 5G-Covered High-Speed Railway Switches On. Caixin Global. https://www.caixinglobal.com/2020-01-13/chinas-first-5g-covered-high-speed-railway-switches-on-101503588.html, Jan 2020.

[19] Philip Bello. Characterization of randomly time-variant linear channels. *IEEE transactions on Communications Systems*, 11(4):360–393, 1963.

[20] Anton Monk, Ronny Hadani, Michail Tsatsanis, and Shlomo Rakib. Otfs-orthogonal time frequency space. *arXiv preprint arXiv:1608.02993*, 2016.

[21] Ronny Hadani and Anton Monk. Otfs: A new generation of modulation addressing the challenges of 5g. *arXiv preprint arXiv:1802.02623*, 2018.

[22] Patchava Raviteja, Khoa T Phan, Yi Hong, and Emanuele Viterbo. Interference cancellation and iterative detection for orthogonal time frequency space modulation. *IEEE Transactions on Wireless Communications*, 17(10):6501–6515, 2018.

[23] Kaltenberger, Florian and Jiang, Haiyong and Guillaud, Maxime and Knopp, Raymond. Relative Channel Reciprocity Calibration in MIMO/TDD Systems. In *2010 Future Network & Mobile Summit*, pages 1–10. IEEE, 2010.

[24] Vasisht, Deepak and Kumar, Swarun and Rahul, Hariharan and Katabi, Dina. Eliminating Channel Feedback in Next-Generation Cellular Networks. In *Proceedings of the 2016 ACM SIGCOMM Conference*, pages 398–411. ACM, 2016.

[25] Bakshi, Arjun and Mao, Yifan and Srinivasan, Kannan and Parthasarathy, Srinivasan. Fast and Efficient Cross Band Channel Prediction Using Machine Learning. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom)*, page 37. ACM, 2019.

[26] MATLAB. Channel Estimation. https://www.mathworks.com/help/lte/ug/channel-estimation.html, 2018.

[27] Singular value decomposition (SVD). Wkipedia. https://en.wikipedia.org/wiki/Singular_value_decomposition, 2019.

[28] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. WiDeo: Fine-grained Device-free Motion Tracing using RF Backscatter. In *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI'15)*, pages 189–204, 2015.

[29] Teng Wei, Anfu Zhou, and Xinyu Zhang. Facilitating Robust 60GHz Network Deployment by Sensing Ambient Reflectors. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI'17)*, pages 213–226, 2017.

[30] Nicolai Czink, Markus Herdin, Hüseyin Özcelik, and Ernst Bonek. Number of multipath clusters in indoor mimo propagation environments. *Electronics letters*, 40(23):1498–1499, 2004.

[31] 3GPP. TS36.141: Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Conformance Testing, Oct. 2019.

[32] 3GPP. NR; Base Station (BS) Radio Transmission and Reception, Oct. 2019.

[33] Yanchao Tang. The research on lte coverage solutions on high-speed railway. *Designing Techniques of Posts and Telecommunications*, 12:20–23, 2014.

[34] Haotian Deng, Chunyi Peng, Ans Fida, Jiayi Meng, and Y Charlie Hu. Mobility support in cellular networks: A measurement study on its configurations and implications. In *Proceedings of the Internet Measurement Conference 2018*, pages 147–160. ACM, 2018.

[35] OpenAirInterface. https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/home, 2019.

[36] 3GPP. TS38.322: Technical Specification Group Radio Access Network; NR; Packet Data Convergence Protocol (PDCP) specification, Jun. 2017.

[37] 3GPP. TS36.322: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification, Sep. 2012.

[38] 3GPP. TS38.321: 5G NR; Medium Access Control (MAC) protocol specification, Jun. 2019.

[39] 3GPP. TS36.321: Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification, Mar. 2014.

[40] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception, Jul. 2017.

[41] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception, Oct. 2019.

[42] Zhaowei Tan, Yuanjie Li, Qianru Li, Zhehui Zhang, Zhehan Li, and Songwu Lu. Enabling Mobile VR in LTE Networks: How Close Are We? In *ACM SIGMETRICS*, 2018.

[43] Zafar Ayyub Qazi, Melvin Walls, Aurojit Panda, Vyas Sekar, Sylvia Ratnasamy, and Scott Shenker. A high performance packet core for next generation cellular networks. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, pages 348–361. ACM, 2017.

[44] Yuanjie Li, Zengwen Yuan, and Chunyi Peng. A control-plane perspective on reducing data access latency in lte networks. In *ACM Mobicom*, Snowbird, Utah, USA, October 2017.

[45] Shichang Xu, Ashkan Nikravesh, and Z Morley Mao. Leveraging context-triggered measurements to characterize lte handover performance. In *International Conference on Passive and Active Network Measurement*, pages 3–17. Springer, 2018.

[46] Li Li, Ke Xu, Tong Li, Kai Zheng, Chunyi Peng, Dan Wang, Xiangxiang Wang, Meng Shen, and Rashid Mijumbi. A measurement study on multi-path tcp with multiple cellular carriers on high speed rails. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 161–175. ACM, 2018.

[47] Brad Karp and Hsiang-Tsung Kung. Gpsr: Greedy perimeter stateless routing for wireless networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 243–254. ACM, 2000.

[48] Ananth Rao, Sylvia Ratnasamy, Christos Papadimitriou, Scott Shenker, and Ion Stoica. Geographic routing without location information. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 96–108. ACM, 2003.

[49] Young-Jin Kim, Ramesh Govindan, Brad Karp, and Scott Shenker. Geographic routing made practical. In *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, pages 217–230. USENIX Association, 2005.

[50] Patchava Raviteja, Khoa T Phan, and Yi Hong. Embedded pilot-aided channel estimation for otfs in delay–doppler channels. *IEEE Transactions on Vehicular Technology*, 68(5):4906–4917, 2019.

Appendices are supporting material that has not been peer-reviewed

A STABLE DELAY-DOPPLER CHANNEL

The variance of delay-Doppler channel $h_w(\tau, v)$ over time $\frac{\partial h_w(\tau,v)}{\partial t} = \frac{\partial h_w(\tau,v)}{\partial \tau}\frac{\partial \tau}{\partial t} + \frac{\partial h_w(\tau,v)}{\partial v}\frac{\partial v}{\partial t}$ relates to the path delay and Doppler variance. The path delay $\tau = \frac{d}{c} \propto \frac{vt}{c}$ (d is path length, a is client acceleration), so its change $\frac{\partial \tau}{\partial t} \propto \frac{v+at}{c} \to 0$ since $v \ll c$ even under extreme client movement (e.g., 10^{-7} for v=500km/h). The Doppler change $\frac{\partial v}{\partial t} \propto \frac{\partial (fv/c)}{\partial t} = \frac{f}{c}a$ relates to the client's acceleration a and is negligible unless the client speeds up or down (infrequent in high-speed rails). Therefore, $h(\tau, v)$ remains constant in a much longer duration than $H(t, f)$ (whose coherence time $T_c \propto \frac{1}{v_{max}}$).

B PROOF OF THEOREM 1

PROOF. We prove that when $P \leq \min(M, N)$ and $\tau_p - \tau_{p'} = k\Delta\tau$ and $v_p - v_{p'} = l\Delta v$ for any p, p', the delay-Doppler decomposition $\mathbf{H} = \mathbf{\Gamma P\Phi}$ results in unitary matrices $\mathbf{\Gamma}$ and $\mathbf{\Phi}$ and $M \times N$ diagonal matrix $\mathbf{P}$, thus being a SVD decomposition. Given $P \leq \min(M, N)$ paths, we can always insert "virtual paths" (with 0 attenuation) and expand $\mathbf{P}$ as a $M \times N$ diagonal, non-negative matrix as follows[11]:

$$\mathbf{P} = \begin{bmatrix} |h_1| & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & |h_2| & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & |h_p| & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

This is equivalent to a $\min(M, N)$-path channel with $|h_p| = 0$ when $p > P$. Therefore, we only need to prove Theorem 1 always holds when $P = \min(M, N)$, and $P < \min(M, N)$ will also hold with this expansion. The following proof focuses on $M < N$ so that $P = M$; $M > N$ follows the similar proof.

First consider the delay spread matrix $\mathbf{\Gamma}$. Note that

$$\Gamma(k, p) = \frac{1}{M} \sum_{d=0}^{M-1} (e^{j2\pi\Delta\tau\Delta f})^{kd} \cdot (e^{-j2\pi\Delta f})^{d\tau_p} = \sum_{d=0}^{M-1} \Gamma_1(k, d)\Gamma_2(d, p)$$

where $\Gamma_1(k, d) = \frac{1}{\sqrt{M}}(e^{j2\pi\Delta\tau\Delta f})^{kd}$ and $\Gamma_2(d, p) = \frac{1}{\sqrt{M}}(e^{-j2\pi\Delta f})^{d\tau_p}$. So we can factorize $\mathbf{\Gamma} = \mathbf{\Gamma_1\Gamma_2}$, where $\mathbf{\Gamma_1}, \mathbf{\Gamma_2} \in \mathbb{C}^{M \times M}$

$$\mathbf{\Gamma_1} = \frac{1}{\sqrt{M}} \begin{bmatrix} (e^{j2\pi\Delta\tau\Delta f})^{0\cdot 0} & \cdots & (e^{j2\pi\Delta\tau\Delta f})^{0\cdot(M-1)} \\ \cdots & (e^{j2\pi\Delta\tau\Delta f})^{kd} & \cdots \\ (e^{j2\pi\Delta\tau\Delta f})^{(M-1)\cdot 0} & \cdots & (e^{j2\pi\Delta\tau\Delta f})^{(M-1)\cdot(M-1)} \end{bmatrix}$$

$$\mathbf{\Gamma_2} = \frac{1}{\sqrt{M}} \begin{bmatrix} (e^{-j2\pi\Delta f})^{0\cdot\tau_1} & \cdots & (e^{-j2\pi\Delta f})^{0\cdot\tau_M} \\ \cdots & (e^{-j2\pi\Delta f})^{d\tau_p} & \cdots \\ (e^{-j2\pi\Delta f})^{(M-1)\cdot\tau_1} & \cdots & (e^{-j2\pi\Delta f})^{(M-1)\cdot\tau_M} \end{bmatrix}$$

since $P = \min(M, N) = M$. We show that both $\mathbf{\Gamma_1}$ and $\mathbf{\Gamma_2}$ are unitary, thus $\mathbf{\Gamma} = \mathbf{\Gamma_1\Gamma_2}$ being unitary. For $\mathbf{\Gamma_1}$, we have

$$\mathbf{\Gamma_1\Gamma_1^*}(k, k') = \sum_{d=0}^{M-1} \Gamma_1(k, d)\Gamma_1^*(d, k') = \frac{1}{M} \sum_{d=0}^{M-1} e^{j2\pi\Delta\tau\Delta\Gamma(k-k')d}$$

If $k = k'$, we have $\mathbf{\Gamma_1\Gamma_1^*}(k, k) = 1$, Otherwise

$$\mathbf{\Gamma_1\Gamma_1^*}(k, k') = \frac{1 - e^{j2\pi\Delta\tau\Delta\Gamma(k-k')M}}{1 - e^{j2\pi\Delta\tau\Delta\Gamma(k-k')}} = \frac{1 - e^{j2\pi(k-k')}}{1 - e^{j2\pi\Delta\tau\Delta\Gamma(k-k')}} = 0$$

since $\Delta\tau = \frac{1}{M\Delta f}$, so $\mathbf{\Gamma_1\Gamma_1^*} = I_M$ and $\mathbf{\Gamma_1^*\Gamma_1} = (\mathbf{\Gamma_1\Gamma_1^*})^* = I_M$ is unitary. For $\mathbf{\Gamma_2}$, we have

$$\mathbf{\Gamma_2^*\Gamma_2}(p, p') = \sum_{d=0}^{M-1} \Gamma_2^*(p, d)\Gamma_2(d, p') = \frac{1}{M} \sum_{d=0}^{M-1} e^{j2\pi\Delta\Gamma(\tau_p - \tau_{p'})d}$$

If $p = p'$, we have $\mathbf{\Gamma_2^*\Gamma_2}(p, p) = 1$. Otherwise

$$\mathbf{\Gamma_2^*\Gamma_2}(p, p') = \frac{1 - e^{j2\pi\Delta f M(\tau_p - \tau_{p'})}}{1 - e^{j2\pi\Delta\Gamma(\tau_p - \tau_{p'})}} = \frac{1 - e^{j2\pi k}}{1 - e^{j2\pi\Delta\Gamma(\tau_p - \tau_{p'})}} = 0$$

since $\tau_p - \tau_{p'} = k\Delta\tau$ for some integer k and $\Delta\tau = \frac{1}{M\Delta f}$. Therefore, $\mathbf{\Gamma_2^*\Gamma_2} = I_M$ and $\mathbf{\Gamma_2\Gamma_2^*} = (\mathbf{\Gamma_2^*\Gamma_2})^* = I_M$ are also unitary, and $\mathbf{\Gamma^*F} = \mathbf{\Gamma F^*} = \mathbf{\Gamma_1\Gamma_2\Gamma_2^*\Gamma_1^*} = I_M$ is unitary. Similarly we can prove $\mathbf{\Phi}$ is also unitary when $v_p - v_{p'} = l\Delta v$ for any p, p'. So $\mathbf{\Gamma}$, $\mathbf{P}$ and $\mathbf{\Phi}$ meets the definition in SVD, and $\mathbf{H} = \mathbf{\Gamma P\Phi}$ is a SVD decomposition. □

[11]This is how SVD is widely used for matrix dimensionality reduction.

C DERIVATION OF ALGORITHM 1

We detail how Algorithm 1 leverages SVD to estimate per-path delay-Doppler for cross-band estimation. Given band 1's channel estimation matrix $\mathbf{H_1}$, we run SVD and use it as an approximation of $\mathbf{H_1} = \mathbf{\Gamma P\Phi_1}$. Note that band 1's $\mathbf{\Gamma P}$ is frequency-independent and thus can be reused by another band. To estimate band 2's channel $\mathbf{H_2} = \mathbf{\Gamma P\Phi_2}$, we need to infer $\mathbf{\Phi_2}$ from $\mathbf{\Phi_1}$. To do so, note that

$$\Phi_1(l\Delta v_i, v_p^1) = \sum_{c=0}^{N-1} e^{j2\pi(l\Delta v_i - v_p^1)cT} = \frac{1 - e^{-j2\pi v_p^1 NT}}{1 - e^{j2\pi(l\Delta v_i - v_p^1)T}}, \forall l$$

$$\Gamma(k\Delta\tau, \tau_p) = \sum_{d=0}^{M-1} e^{-j2\pi(k\Delta\tau - \tau_p)d\Delta f} = \frac{1 - e^{j2\pi\tau_p M\Delta f}}{1 - e^{-j2\pi(k\Delta\tau - \tau_p)\Delta f}}, \forall k$$

So we have

$$\frac{\Phi_1(p, l)}{\Phi_1(p, l')} = \frac{1 - e^{j2\pi(l'\Delta v - v_p^1)T}}{1 - e^{j2\pi(l\Delta v - v_p^1)T}}, \frac{\Gamma(k, p)}{\Gamma(k', p)} = \frac{1 - e^{-j2\pi(k'\Delta\tau - \tau_p)\Delta f}}{1 - e^{-j2\pi(k\Delta\tau - \tau_p)\Delta f}}$$

for any (k, k') and (l, l'). Then we can extract

$$e^{-j2\pi v_p^1 T} = \frac{\Phi_1(p, l) - \Phi_1(p, l')}{\Phi_1(p, l)e^{j2\pi l\Delta vT} - \Phi_1(p, l')e^{j2\pi l'\Delta vT}}$$

$$e^{j2\pi\tau_p\Delta f} = \frac{\Gamma(k, p) - \Gamma(k', p)}{\Gamma(k, p)e^{-j2\pi k\Delta\tau\Delta f} - \Gamma(k', p)e^{-j2\pi k'\Delta\tau\Delta f}}$$

When the conditions in Theorem 1 was not strictly satisfied (mainly due to small (M, N) and thus imperfect sampling), SVD and above derivations are approximations of delay-Doppler estimation. For high accuracy, Algorithm 1 computes the average of above delays/Dopplers across all (k, k') and (l, l') (line 4–5). Then we can convert each path's Doppler $v_p^2 = \frac{f_2}{f_1}v_p^1$ for every path p (line 6). Now with $\{h_p, \tau_p, v_p^2\}_{p=1}^{P_{max}}$, Algorithm 1 follows the definitions in §5.2, construct $\mathbf{\Phi_2}$ and estimate cell 2 as $\mathbf{H_2} = \mathbf{\Gamma P\Phi_2}$ (line 9–10).

D PROOF OF THEOREM 2

PROOF. We first prove necessity. Note that REM's simplified policy only use direct wireless comparison (A3 in Figure 1). If no persistent loops occur, then no persistent loops between any two cells c_i and c_j in the same area. That means, the following two conditions will not happen simultaneously:

$$\begin{cases} \text{SNR}_j > \text{SNR}_i + \Delta_{A3}^{i \to j} & (c_i \to c_j) \\ \text{SNR}_i > \text{SNR}_j + \Delta_{A3}^{j \to i} & (c_j \to c_i) \end{cases} \quad (7)$$

If $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} < 0$, (7) can always simultaneously occur for any SNR_j and $\text{SNR}_i \in [\text{SNR}_j + \Delta_{A3}^{j \to i}, \text{SNR}_j - \Delta_{A3}^{i \to j}]$. This contradicts with loop-freedom. So we conclude $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} \geq 0, \forall i, j$.

We next prove sufficiency by recursion. If $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} \geq 0, \forall i, j$, (7) has asserted that no 2-cell persistent loops will occur for any $(\text{SNR}_i, \text{SNR}_j)$. Assume $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} \geq 0$ asserts any 1, 2, ..., $(n - 1)$-cell loop freedom among $c_1, c_2, \ldots, c_{n-1}$. Now consider n cells $c_1, c_2, \ldots, c_{n-1}, c_n$. Since $\Delta_{A3}^{i \to j} + \Delta_{A3}^{j \to i} \geq 0, \forall i, j$, any 1, 2, ..., $(n - 1)$-cell loop freedom still retains among these cells. Then consider if n-cell loop $c_1 \to c_2 \to \cdots \to c_n \to c_1$ can happen for some $(\text{SNR}_1, \text{SNR}_2, \ldots, \text{SNR}_n)$. To incur it, the following

conditions should be satisfied simultaneously

$$
\begin{cases}
\mathrm{SNR}_2 > \mathrm{SNR}_1 + \Delta_{A3}^{1\to2} & (c_1 \to c_2) \\
\mathrm{SNR}_3 > \mathrm{SNR}_2 + \Delta_{A3}^{2\to3} & (c_2 \to c_3) \\
\cdots & \cdots \\
\mathrm{SNR}_n > \mathrm{SNR}_{n-1} + \Delta_{A3}^{n-1\to n} & (c_{n-1} \to c_n) \\
\mathrm{SNR}_1 > \mathrm{SNR}_n + \Delta_{A3}^{n\to1} & (c_n \to c_1)
\end{cases}
\tag{8}
$$

summing up all conditions results in $\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} + \cdots + \Delta_{A3}^{n-1\to n} + \Delta_{n\to1} < 0$. But since $\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} \geq 0, \Delta_{A3}^{2\to3} + \Delta_{A3}^{3\to4} \geq 0, \ldots, \Delta_{A3}^{n-1\to n} + \Delta_{A3}^{n\to1} \geq 0, \Delta_{A3}^{n\to1} + \Delta_{A3}^{1\to2} \geq 0$, summing up them results in $2(\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} + \cdots + \Delta_{A3}^{n-1\to n} + \Delta_{n\to1}) \geq 0$ and thus contradiction. So we conclude that no n-cell loop will occur for any SNR settings, and conclude the sufficiency by recursion. □

E PROOF OF THEOREM 3

Proof. We prove it by contradiction. Assume $\Delta_{A3}^{i\to j} + \Delta_{A3}^{j\to i} \geq 0, \forall i, j$ but a persistent loop $c_1 \to c_2 \to \cdots \to c_n \to c_1$ happen for some $(\mathrm{SNR}_1, \mathrm{SNR}_2, \ldots, \mathrm{SNR}_n)$. Regardless of any other non-SNR policies between $c_1, c_2, \cdots, c_n$ and how they are evaluated, Equation (8) will still hold and result in $\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} + \cdots + \Delta_{A3}^{n-1\to n} + \Delta_{n\to1} < 0$. But since $\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} \geq 0, \Delta_{A3}^{2\to3} + \Delta_{A3}^{3\to4} \geq 0, \ldots, \Delta_{A3}^{n-1\to n} + \Delta_{A3}^{n\to1} \geq 0, \Delta_{A3}^{n\to1} + \Delta_{A3}^{1\to2} \geq 0$, summing up them results in $2(\Delta_{A3}^{1\to2} + \Delta_{A3}^{2\to3} + \cdots + \Delta_{A3}^{n-1\to n} + \Delta_{n\to1}) \geq 0$ and thus contradiction. □