Movement-Based Reliable Mobility Management for Beyond 5G Cellular Networks

Zhehui Zhang, Yuanjie Li, *Senior Member, IEEE*, Qianru Li, Jinghao Zhao, Ghufran Baig, Lili Qiu, *Fellow, IEEE*, and Songwu Lu, *Fellow, IEEE*

***Abstract*— Extreme mobility becomes a norm rather than an exception with emergent high-speed rails, drones, industrial IoT, and many more. 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 beyond 5G cellular networks while maintaining backward compatibility to 4G/5G. 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. REM reduces the network failures by up to an order of magnitude, eliminates policy conflicts, and improves application performance by 31.8% - 88.3% compared to legacy 4G/5G.**

***Index Terms*— Mobile network, beyond 5G, extreme mobility management, reliability, policy conflicts, delay-Doppler domain.**

I. Introduction

MOBILE users want *anywhere*, *anytime* network services. Even when the user is moving at a high speed, the user expects negligible service disruption. The demand for network reliability is more pressing given emerging delay-sensitive applications, e.g., mobile VR/AR. Such use scenarios demand *always-on* network service even under extreme mobility, such as the high-speed rails (35,000 km routes for over 1 billion passengers by 2019, with the speed up to 350 km/h [2]), vehicle-to-everything (e.g., autonomous driving and fleet management, with 6 million vehicles by 2022 [3]), drones, and many more.

In this work, we examine the network reliability under extreme mobility scenarios. We start with a key question: *Is 4G/5G reliable for delay-sensitive applications under extreme mobility?* While the existing mobile network has been successful in supporting wide-area mobility management, most users are moving slowly or static. It is open to question whether existing mobility management design meets reliability demands under extreme mobility for two reasons. On the one hand, the client is moving faster in high-speed rails, vehicles, industrial IoT, etc., at a speed of up to 500km/h. On the other hand, the upcoming 5G is adopting high-frequency radios (sub-6GHz and above-20GHz mmWave) for fast data transfer.

Unfortunately, the answer is negative in reality. Our empirical study of 4G LTE over high-speed rails unveils that, the handovers are more frequent and unreliable. On average, the handovers between base stations occur every 11–20s. 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. This leads to significant user-perceived disruptions for interactive delay-sensitive applications like mobile VR/AR. We also verified with a 5G dataset that the mobility management in 5G faces similar challenges as 4G LTE. In 5G, given the denser cell deployment, handovers happen even more frequently (50.2s to 41.9s) upon an even slower speed.

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 that *the client movement is more robust and predictable than wireless signal strength, thus suitable to drive mobility management*. Thus, REM shifts to *movement-based* mobility management. REM

Manuscript received 20 December 2021; revised 19 May 2022; accepted 27 June 2022; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor C. F. Chiasserini. Date of publication 1 August 2022; date of current version 16 February 2023. This work was supported in part by the NSF CNS-1910150, NSF CNS-2008026, NSF CNS-2008824, and NSF CNS-2107037. The work of Yuanjie Li was supported by the National Key Research and Development Plan of China under Grant 2018YFB1800301 and in part by the National Natural Science Foundation of China under Grant 62132009. An earlier version of this paper appeared in ACM SIGCOMM 2020 [DOI: https://doi.org/10.1145/3387514.3405873]. *(Corresponding author: Yuanjie Li.)*

Zhehui Zhang, Qianru Li, Jinghao Zhao, and Songwu Lu are with the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA 90095 USA.

Yuanjie Li is with the Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100190, China (e-mail: liyuanjie08@gmail.com).

Ghufran Baig is with Google Inc., Mountain View, CA 94043 USA.

Lili Qiu is with the Department of Computer Sciences, The University of Texas at Austin, Austin, TX 78712 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TNET.2022.3190788, provided by the authors.

Digital Object Identifier 10.1109/TNET.2022.3190788

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 [4]. 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 both low mobility datasets and high-speed rails datasets. Compared to solutions today, REM eliminates policy conflicts, and 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. We also evaluate REM's benefits for real-world applications, including mobile VR/AR and data transfer. Our experiments show that REM reduces disruption for mobile VR applications by 47.1%, improves object recognition accuracy by 88.3% for mobile AR, and reduces stalls by 31.8%-46.2%. 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 on the network reliability in extreme mobility (§III). 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 (§IV–V). 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, and systematically evaluate REM's network failure reduction, policy conflict resolution, benefits to applications, and system overhead (§VI).

This paper is organized as follows. We introduce the background of 4G/5G mobility management and the challenges under extreme mobility in §II. In §III, we present our findings on the limitations of mobility management in operation today based on extensive datasets. Then we present the key ideas in §IV and the design in §V. We show the implementation and evaluation in §VI. We evaluate the benefits of REM for emerging applications in §VII. Then, we discuss the remaining questions on REM in §VIII and compare REM with related work in §IX. Finally, §X concludes the paper.

II. Mobility Management Primer

We introduce the 4G/5G mobility management today, and the challenges under extreme mobility.

A. 4G/5G Mobility Management

To support *anywhere, anytime* network services, 4G/5G deploys base stations densely for clients to access. Each base station may run multiple cells, each under various frequencies (using separate antennas) with different coverage and bandwidth. The network provides fine-grained mobility via inter-cell handover. As a client leaves the one cell's coverage, it will be migrated to another one (called *handover*). Figure 1 shows there are two base stations named BS_1 and BS_2, and BS_2 supports two cells with different coverage. Note the client might handover between two cells at the same base station or at different base stations.

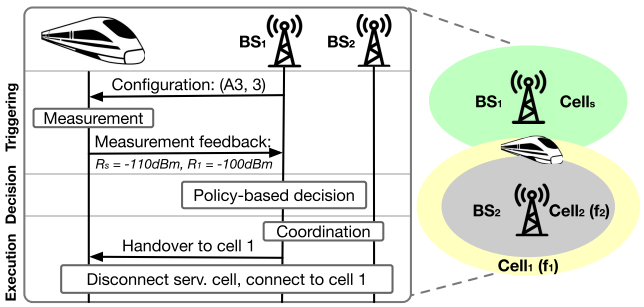


Fig. 1. Mobility management in 4G/5G today.

TABLE I

Triggering Criteria Based on Wireless Signal Strength in 4G/5G. R_s and R_n Are Signal Strength Indicators. Δ_{Ax} Represents the Threshold Parameter. [5], [6]

Event	Criteria	Explanation
A1	$R_s > \Delta_{A1}$	Serv. cell is better than a threshold
A2	$R_s < \Delta_{A2}$	Serv. cell is worse than a threshold
A3	$R_n > R_s + \Delta_{A3}$	Neigh. cell is offset better than the serv.
A4	$R_n > \Delta_{A4}$	Neigh. cell is better than Δ_{A4}
A5	$R_s < \Delta^1_{A5}$, $R_n > \Delta^2_{A5}$	Serv. cell is worse than Δ^1_{A5} and neigh. cell is better than Δ^2_{A5}

The 4G/5G mobility management is based on wireless signal strength: It takes client-perceived per-cell *wireless signal quality* as the main input, relies on client-side *channel feedback* to trigger, and decides the target cell based on local *policies*. Figure 1 depicts 4G/5G handover [5], [6]. The mobility management has three phases. In the *triggering* phase, the serving cell configures a client to measure specific frequency bands. The measured bands are selected based on whether there are neighboring cells located at that band. The client will measure the given bands and examine neighbor cells' signal strengths[1] with standard triggering criteria in Table I. The client will report the feedback if any criteria are satisfied. Upon receiving the client's feedback, the serving cell moves to the *decision* phase. It runs its local policy (detailed in §III-B) to decide if more feedback is needed, or if handover should start. If a handover decision is made, the serving cell notifies the device to *execute* handover to establish a connection with the target cell.

B. Challenges Under Extreme Mobility

Current 4G/5G mobility management relies on wireless signal strength indicators. However, extreme mobility challenges the reliability of such indicators due to increased

[1]In 4G/5G, the signal strengths can be RSRP, RSRQ or RSSI [5], [6].

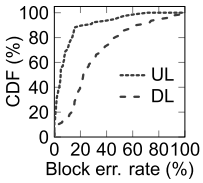


Fig. 2. BLER.

wireless dynamics and Doppler interference. As the client moves, the 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 and channel quality degradation. In 4G/5G OFDM/OFDMA,[2] the channel remains approximately invariant in a very short duration $T_c \propto 1/\nu_{max}$ [7], where T_c is the coherence time and $\nu_{max} \propto vf/c$ 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., $T_c \approx$ 20ms 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 ($T_c \approx$ 1ms as quantified in §III-A).

III. Unreliable Extreme Mobility

A fundamental problem for the mobility management today is that, it is based on *indirect wireless signal strength*, rather than *direct client movement*. While acceptable in static and low-speed mobility, such design is unreliable for extreme mobility with various network failures, delays, and persistent policy conflicts in practice. We conduct an empirical analysis to quantify the reliability deficiencies at different phases. Table II presents our results on two LTE datasets from high-speed rails and our driving dataset (all detailed in §VI).

We find that the failures arise from triggering (§III-A), decision (§III-B), and execution (§III-C). To mitigate these failures, operators adopt proactive handover policies to trigger handover promptly. We infer a serving cell's handover policy by following [8] to model the policy as a state machine, and infer it using the LTE signaling messages. We find that the proactive policy incurs non-negligible policy conflicts (every 194.6–1090.0s on average), which causes persistent loops and voids operators' failure mitigation efforts (§III-B). We next elaborate on them, analyze their root causes, and validate them with large-scale datasets. Our findings motivate the design of a reliable mobility management scheme under dramatic wireless dynamics under extreme mobility.

A. Unreliable Feedback for Triggering

Reporting client-side feedback reliably to the serving cell is critical to trigger handovers (§II). Such feedback tracks the client-perceived *wireless quality* of cells based on standard criteria (Table I). In extreme mobility, such 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).

TABLE II
Network Reliability in Extreme Mobility

	driving	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
Tot. failure ratio	4.3%	5.2%	10.6%	12.5%
Feedback loss (§III-A)	0.78%	1.7%	4.9%	6.9%
Missed cell (§III-B)	1.8%	0.6%	0.4%	0.8%
HO cmd. loss (§III-C)	0.61%	1.1%	3.3%	2.4%
Coverage holes	1.1%	1.7%	2.0%	2.4%
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/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%

The dilemma results in two deficiencies of the current feedback scheme: *(1) Head-of-line blocking of measurement:* To decide an appropriate target cell, the client should detect *all* cells that meet the criteria. For wireless signal strength-based feedback, the 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 with extended triggering period:* Instantaneous wireless measurement is dynamic and causes transient oscillations between base stations. Direct reporting instantaneous wireless quality can trigger unnecessary handovers and transient oscillations between cells ("ping-pong loops"). To mitigate it, 4G/5G mandates the client to report a cell only if its criteria hold for a configurable triggering interval [5], [6].[3] This delays feedback with late handovers.

1) Reality Check: We gauge the impact of unreliable feedback by feedback delay and loss. To quantify the feedback delay and loss, we analyze the collected physical layer signaling message on measurement activities and link layer traces. Table II shows 33.3–55.2% of failures in HSR are from feedback delay/loss. The loss is mostly caused by errors: Figure 2 shows a 9.9% block error rate before the loss, which implies the feedback is corrupted in delivery. For the feedback delay, Figure 3 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 as the triggering interval for cells under the 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 $T_c \approx c/fv \in [1.16ms, 6.18ms]$ (§II) given $f \in [874.2, 2665]$ MHz and $v \in [200, 350]$ km/h from our datasets. Note operators have shortened the triggering interval for faster feedback than low mobility

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

[3]This configurable triggering interval is named as TimerToTrigger in 4G/5G.

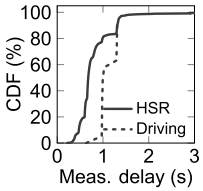
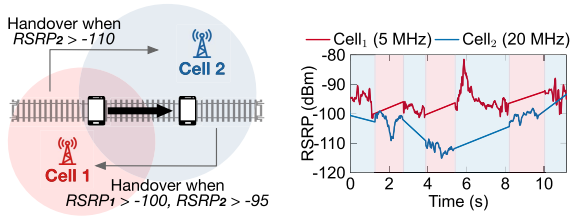
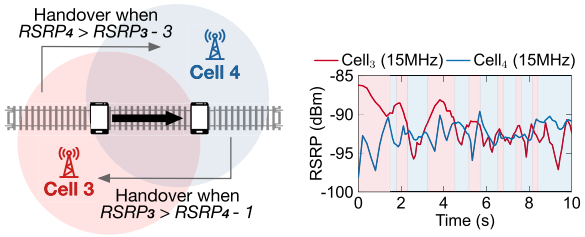


Fig. 3. Slow meas.



(a) Conflicts in load balancing (b) Real handover traces

Fig. 4. Policy conflicts from load balancing in HSR.



(a) Failure-induced conflicts (b) Real handover traces

Fig. 5. Failure-induced policy conflicts in HSR.

(mostly 640ms in our dataset), but at the cost of more transient loops and signaling.

2) Opportunity: Shared Physical Multipath: We find the opportunity of shared physical multipath to reduce delay due to sequential measurement and extended triggering period. In reality, several cells would share a base station to aggregate computation. These cells are under different bands to improve the radio coverage and carrier aggregation. Our dataset shows 53.4% of cells share the same base station with another cell.[4] These cells' signals traverse the same paths from the base station to the client, thus experiencing similar channels. This indicates that the measurement of multiple cells at the single base station could be accelerated without missing potential cells. In §V, we show how to relax the exploration-exploitation dilemma for reliable feedback.

B. Conflicting Policy for Decision

On receiving client-side measurements, the serving cell should decide the target cell. 4G/5G handover decisions are policy-driven design. To accommodate diverse demands (good radio coverage, fast data speed, load balancing, failure mitigation, etc.), each cell can customize its local policies with configurable criteria in Table I. Figure 6 exemplifies a typical policy inferred from our HSR dataset. Such policy is tightly coupled with wireless feedback (§III-A). The fundamental reason is that wireless feedback is unreliable in extreme mobility. Operators thus design multi-stage, complex, even conflicting polity to mitigate late handovers. However, these policies cause more failures and conflicts. Such policy suffers from two deficiencies:

• **Multi-stage policy:** Most operators adopt *multi-stage* handover policies as exemplified in Figure 6. At stage 1, cells under the same frequency as serving cell's are measured and chosen first. Only if the serving cell gets bad, the policy moves to stage 2 via measurement reconfiguration. The rationale is for high performance: Measuring the inter-frequency cells will force the client to temporarily switch its frequency bands and thus cannot send/receive data.[5] This scheme works well under low mobility. But if the client moves fast, this policy can miss potential inter-frequency cells. The extra round trips of reconfiguration and feedback are needed for inter-frequency cells, during which the client might have moved away.

• **Policy conflicts in extreme mobility:** It has been shown that [13], [14], policies among cells can have conflicts and cause *persistent loops*. Figure 4a exemplifies a conflict from our dataset. Cell 1 and 2 have different bandwidths (5MHz vs. 20MHz). For fast data speed, cell 1 moves a client to cell 2 if cell 2's signal strength $\text{RSRP}_2 > -110$dBm. But cell 2 adopts a different policy: It migrates a client to cell 1 if it is weak ($\text{RSRP}_2 < -95$dBm) and cell 2 is strong ($\text{RSRP}_1 > -100$dBm). Both policies can be *simultaneously* satisfied if $\text{RSRP}_1 > -100$dBm and $-110dBm < \text{RSRP}_2 < -95dBm$). Then the client oscillates between cell 1 and 2 (8 handovers within 15s in Figure 4b). Such loop disrupts client's service *and* incurs signaling storms for the network.

Surprisingly, we note policy conflicts are amplified in extreme mobility, because of operators' desire to mitigate failures! This differs from [13], [14] that focus on static scenarios, and has been frequently observed in our dataset (detailed in validation below). As shown in §III-A, a fast-moving client may miss the cells and lose service due to slow feedback and decisions. As shown in Figure 5a,[6] the device is leaving from cell 3 to cell 4. The operator designed the threshold to trigger handover even when cell 4's RSRP is 3dB worse than cell 3's RSRP. However, this raises conflicts if neighbor cells use the same policy. Such policy will not mitigate failures; the client will move back with loops.

1) Reality Check: Our empirical study confirms both problems. First, the multi-stage policy can miss inter-frequency cells and induce handover failures. We quantify missed cells by checking whether a handover failure is recovered to an inter-frequency cell which is not measured before. In such a case, the client ended up with connection loss and re-connected with that missed cell. We find that missed cell cases account for 3.7%–11.1% of failures in HSR (Table II).

[4]This is obtained by grouping the globally unique base station IDs from LTE cells' identifiers called ECIs [10].

[5]To measure an inter-frequency cell, a client should synchronize to it and measure its signal strength. The serving cell pre-allocates Measurement-Gaps [11], [12] for this, during which the client cannot send/receive data.

[6]The relative location of railway and cells are conjectured from device measured signal strength.

TABLE III
MOBILITY IN 5G

Conflicts	Type	BJ-Taiyuan	BJ-Shanghai [9]
A3-A4	Inter-freq	4 (2.4%)	316 (23.6%)
A3-A5	Inter-freq	1 (0.6%)	24 (1.8%)
A4-A4	Inter-freq	2 (1.2%)	200 (14.9%)
A4-A5	Inter-freq	5 (3.0%)	49 (3.7%)
A5-A5	Inter-freq	0	2 (0.1%)
A3-A3	Intra-freq	**155 (92.8%)**	**749 (55.9%)**

TABLE IV
TWO-CELL POLICY CONFLICTS IN HSR DATASETS

Speed (km/h)	0 – 40
Avg. handover interval	41.9 s
Total failure ratio	2.7%
Feedback delay/loss	1.02%
Missed cell	0.32%
Handover cmd. loss	0.7%
Coverage holes	0.67%

We also examined why operators prefer the multi-stage policy. If the client stays at the stage to measure inter-frequency cells continuously, such measurements consume 38.3%–61.7% spectrum in inter-frequency measurements (depending on cell configurations).

Policy conflicts exist with alarming frequency in extreme mobility. We quantify all conflicts between two cells by inferring and checking handover policy from the dataset. Table IV 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 [13], [14], and dominate the policy conflicts in extreme mobility (55.9%–100%). To trigger handovers early with fewer failures, the operators configure a proactive policy among cells (Figure 5a with $\Delta_{A3} < 0$). Such policy causes oscillations and voids the efforts of failure mitigation.

C. Unreliable Signaling for Execution

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

1) Reality Check: Table II shows 19.2%–31.5% of network failures arise 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 6), but no handover command from the serving cell until the client loses network access. We also observe high physical-layer block errors when such failure occurs. Figure 2 shows the 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.

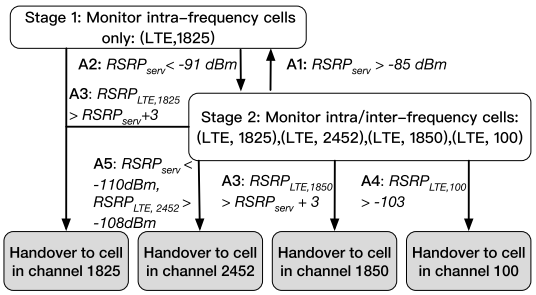


Fig. 6. Multi-stage policy state machine view.

D. Applicability in 5G

5G standards [6], [12], [15] introduced many new technologies to improve reliability, such as new radio physical design, cloud-native deployment. However, the mobility management mechanism stays the same. As the management is based on wireless signal strength, it still follows the three phases, triggering, decision, and execution (§II). The feedback and signaling rely on OFDM-based transmission, which is unreliable for making handover decisions. The decision policy tightly couples with signaling reliability, motivating operators to adopt complex, even conflicting decision logic.

There are even more challenges for the mobility management in 5G. First, the density of cells increase in 5G, which implies more frequent handovers. Especially in the non-standalone mode, each device is connected with a 4G cell and a 5G cell, thus the device will experience more cell switching. Recent research on 5G mobility [16] conforms to our statement that 5G handover is more frequent. Second, 5G adopts mmWave bands (e.g., 29GHz), where the Doppler spread is more severe. The 5G also added new numerology with a shorter slot time [12], which is more susceptible to Doppler spread caused issues.

Although the mobility mechanism does not change, one might wonder whether the infrastructure change in 5G brings more benefits to optimize mobility performance. The main infrastructure update from 4G to 5G is the adoption of cloud-native deployment. The control plane operations can be moved closer to the RAN to speed up the decision. However, as revealed in §III-A, triggering delay is the main bottleneck. The measurement and reporting are still constrained by the round-trip between the device and base station. As the unstable OFDM-based modulation remains unchanged, the problems with signaling and policy remain.

1) Reality Check: We also study 5G mobility management in an empirical study with 3866 handover samples. We run 45-hour driving experiments with 5G phones with low mobility. The experiments are performed under AT&T, one of the

largest wireless operator in the US. During the experiment, the devices keep active connection by downloading files and sending heartbeat to the servers. The driving speed ranges from 10-40 km/h to keep 5G connections. We collect signaling events with MobileInsight [8] to check handover events.

We find that mobility reliability becomes even become worse in 5G since handover will be more frequent and the issues persist. As shown in Table III. The average handover interval is 41.9s, which is even smaller than 50.2s in 4G with 0-100 km/h driving speed. The failure ratio due to feedback delay/loss and handover command loss are 1.02% and 0.7%, which are even higher than 0.78% and 0.61% in 4G. This proves that the unreliability of signaling under mobility persists. The failure ratio due to missed cell and coverage holes dropped in 5G, which is benefited from the denser deployment. In conclusion, the signaling-strength-based mobility management in 5G is still susceptible to signaling loss and unreliable channel feedback. The policy-based decision is expedited due to denser deployment but the handover frequency increases at the same time.

IV. REM Key Ideas

We devise REM, **R**eliable **E**xtreme **M**obility management to achieve the following goals:

- **Excellent reliability** which mitigates signaling loss, channel feedback delay, and policy conflicts;
- **Retain policy flexibility** for the network operators;
- **Backward compatibility** with existing OFDM-based designs and static/low-speed mobility, without hurting wireless data performance.

Our key idea is that, extreme mobility is unreliable because of the *wireless signal strength-based* design, which is susceptible to Doppler shift and multipath fading (§II). REM shifts from indirect wireless signal strength-based to direct *movement-based* mobility. Compared to wireless with short coherence and dramatic dynamics, the client movement is slower and thus more reliable to drive extreme mobility management. REM tracks the client movement in the *delay-Doppler domain* to defeat against the unreliability of wireless dynamics. Benefiting from the reliable movement-based design, REM relaxes the feedback's exploration-exploitation dilemma in the triggering phase, offers conflict-free policies in the decision phase, and stabilizes the signaling in the execution phase.

A. Channel Model in Delay-Doppler Domain

We now introduce the channel model in delay-Doppler domain and why it is the rescue to mobility management. The channel in the *delay-Doppler domain* is represented as [17]:

$$h(\tau,\nu)=\sum_{p=1}^{P} h_p\delta(\tau-\tau_p)\delta(\nu-\nu_p), \tag{1}$$

where P is the number of propagation paths, h_p, τ_p, ν_p are the gain, propagation delay and Doppler frequency shift associated with p-th path, and δ is the Dirac delta function. Figure 8a shows a channel with 3 paths. The delay-Doppler form reflects the *multi-path geometry* between cell and client in movement.

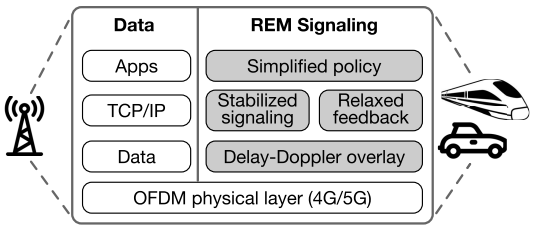


Fig. 7. REM overview.

The delay-Doppler representation decomposes channel parameters into two types: frequency dependent and frequency independent. Note h_p depends on the propagation medium, τ_p depends on p-th path's length, and $\nu_p \propto \frac{vf}{c}$ is the only parameter that depends on the carrier frequency f. As we find in §III-A, many cells are co-located at the same base station. For co-located cells, they operate on different frequencies but share the same physical paths to the device. Due to the shared physical path, the h_p and τ_p are the same. If we are able to estimate the only remaining parameter ν_p, we can estimate a cell's channel without measuring it.

The delay-Doppler domain facilitates mobility management due to its stability. Compared to time-frequency domain representation $H(t,f)$, delay-Doppler representation $h(\tau,\nu)$ is more stable benefited from the low variance of path delay and Doppler shift [4], [18], [19]. By representing and transferring signals in the delay-Doppler domain, the system will exploit the full time-frequency diversity, and therefore experience more stable channels and less loss/corruption.

B. REM Roadmap

With stabilized signals in the delay-Doppler domain, REM devises the mobility management accordingly. REM builds a signal overlay with the recently proposed OTFS modulation [4]. On top of the signal overlap, REM further extends OTFS to refine all phases of mobility management. REM proposes the OTFS-based overlay for co-existence with OFDM, relaxes reliance on feedback with cross-channel estimation, and simplifies mobility policy to guarantee conflict-free handovers. Figure 7 overviews REM's main components.

- **Delay-Doppler signaling overlay (§V-A):** REM places the signaling traffic and reference signals in a 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 triggering (§III-A) and execution (§III-C), and exposes movement information to later phases.
- **Relaxed reliance on feedback (§V-B):** To mitigate the failures from slow and unreliable feedback (§III-A), REM devises cross-band estimation in the delay-Doppler domain. This approach accelerates the feedback without reducing the cells to be explored, and facilitates earlier handovers with fewer failures.
- **Simplified, conflict-free policy (§V-C):** To eliminate policy conflicts and failures from missed cells (§III-B), 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 conflict-freedom.

V. The REM Design

A. Delay-Doppler Signaling Overlay

REM designs the mobility management with delay-Doppler domain channel representation. So REM requires its signaling traffic (e.g., measurement feedback, handover commands, reference signals) sent with the delay-Doppler channel. As the backward compatibility requirement indicates, changing existing 4G/5G designs or affecting OFDM-based data transfer is not desired. To this end, REM leverages recent advances in OTFS in the delay-Doppler domainREM integrates an OTFS-based signaling overlay atop OFDMand enables the co-existence.

1) OTFS-Based Overlay: OTFS is a modulation in the delay-Doppler domain. Intuitively, OTFS multiplexes information symbols across all the available carrier frequencies and time slots, aiming to directly capture the underlying multi-path geometry. Symbols experience all the diverse paths of the channel and exhibit less variance. OTFS is suited for the time and frequency selective fading channel, less vulnerable to errors, and more robust to Doppler spread.

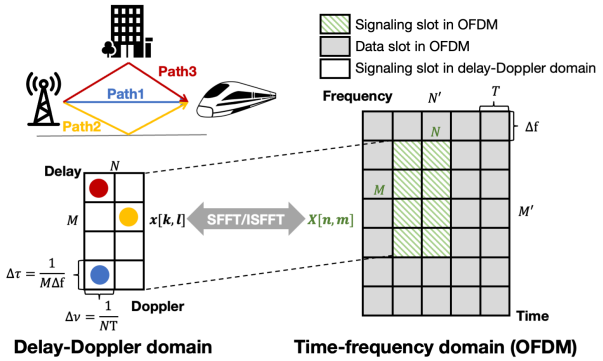
Figure 8a shows the OTFS modulation. It runs on top of OFDM. The OFDM time-frequency domain is discretized to a $M \times N$ grid (each being a 4G/5G radio resource element) by sampling time and frequency axes at intervals T and Δf,[7] respectively. Given a $M \times N$ time-frequency domain, the delay-Doppler domain is also a $M \times N$ grid $(\frac{k}{M\Delta f}, \frac{l}{NT})$, where $k = 0..M-1, l = 0..N-1$. Note that $\frac{1}{M\Delta f}$ 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 inverse Symplectic Fourier transform (ISFFT): 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 SFFT is applied to $Y[n,m]$ and yields $y[k,l]$ in the delay-Doppler domain. With channel noises, we have [4], [20]

$$y[k,l] = \frac{1}{NM} \sum_{k'=0}^{M-1} \sum_{l'=0}^{N-1} h_w(k'\Delta\tau, l'\Delta\nu) \times x[k-k', l-l'] + n[k,l] \quad (2)$$

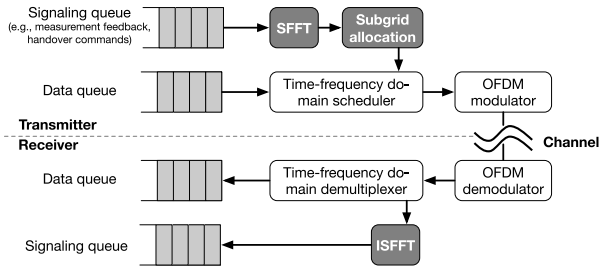
where $h_w(\tau,\nu) = \int\int e^{-j2\pi\tau'\nu'} h(\tau',\nu') w(\nu-\nu', \tau-\tau') d\tau' d\nu'$ is the convolution of channel $h(\tau',\nu')$ and rectangular signal window: $w(\tau,\nu) = \sum_{c=0}^{N-1}\sum_{d=0}^{M-1} e^{-j2\pi(\nu cT - \tau d\Delta f)}$, $n(k,l) = ISSFT(N[n,m])$ is ISFFT of time-frequency noises.

As shown in Figure 8, OTFS signal overlap runs on top of OFDM and co-exists with OFDM data. REM supports the hybrid mode for backward compatibility and efficiency. We thus leave the choice of whether data traffic should use OTFS to the operator. OTFS for data would also reduce Doppler shifts for faster data speed [4], [9], but at the cost of more per-data processing delays. Instead of mandating if OTFS should be used for data, we leave this decision to operators and future designs, and offer a universal solution for *both* choices.

[7]In 4G OFDM, $T = 66.7\mu s$, $\Delta f = 15$KHz [11]. In 5G OFDM, T can be $4.2, 8.3, 16.7, 33.3$ or $66.7\mu s$ and Δf can be 15, 30, 60, 120 or 240KHz [12].



(a) Subgrid allocation for signaling in delay-Doppler domain



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

Fig. 8. Coexistence of OTFS signaling (delay-Doppler domain) and OFDM data (time-frequency domain). The cell dynamically allocates $M \times N$ resource elements mapped to a OTFS sub-grid in OFDM's $M' \times N'$ grid.

Note the co-existence of OFDM and OTFS requires that OTFS symbols occupy a continuous $M \times N$ OFDM grid. However, if we adopt the legacy radio resource scheduling scheme in 4G/5G, the signaling and data traffic are multiplexed in the OFDM grid, which means the signaling traffic is scheduled with *disjoint* OFDM slots. Under such a case, the OTFS cannot run. The naive solution is to reserve a dedicated sub-grid for OTFS-based signal traffic. But the volume of signal traffic varies. Reserving resources based on the maximum volume cause waste when there is less signal traffic.

2) Design of the Scheduler: We design a scheduler that allocates OTFS-based subgrid dynamically. Our scheduler is based on the insight that *the 4G/5G signaling traffic is always prioritized in scheduling and delivery by design* [5], [6]. This is because without successful signaling procedures to configure the protocol stack, the data traffic may not be correctly delivered or processed. Due to the signaling traffic's functional importance, the 4G/5G scheduler always schedules the radio resource for the signaling traffic first over data traffic. Thus, the signaling traffic can always use OTFS, while the data traffic could still use OFDM. We leverage this readily-available feature to decouple the scheduling of OTFS-based signal traffic and OFDM-based data traffic.

Figure 8b shows the design of REM's OTFS signaling overlay. It inserts an OTFS overlay between 4G/5G OFDM radio and signaling layers (i.e., radio resource control [5], [6]). At the transmitter side, REM modulates the signaling traffic with SFFT, allocates a subgrid, and sends it to the scheduler. The scheduler prioritizes signal traffic, including feedback

and handover commands for mobility management and reference signals for cell measurements. It places all signaling in a separate subgrid from OFDM symbols, ensuring the orthogonality between signaling and data. At the receiver side, REM demodulates OFDM symbols, and then run ISFFT to decode signaling traffic. Although REM adds the SFFT/ISFFT to pre/post-process the signaling traffic, the complexity is similar to 4G/5G uplink's SC-FDMA on top of OFDM (with additional fast Fourier transform).

3) Scheduling of M, N: M, N can be small depending on signaling traffic volume. If data uses OTFS, this problem will disappear since $M = M', N = N'$. The grid size of OTFS is the same as the OFDM. If data does not use OTFS, the required grids for signaling depend on signaling volume. To improve it, the scheduler may allocate more guard resource elements to enlarge M', N', at the cost of resource waste. Such tradeoff is similar to 5G's additional DMRS for high mobility today [12], but is better than DMRS since it helps stabilize the signaling channel. Note the maximum captured Doppler $M\Delta\tau = 1/\Delta f$ depends on the subcarrier spacing Δf only, while the maximum captured delay $N\Delta\nu = 1/T$ depends on the symbol length. Therein, the selection of M, N only affects the quantization steps but not the bound of captured delay and Doppler shift.

4) Applicability to 5G: The design of signal overlay is applicable to 5G since 5G still relies on OFDM-based reference signals.There are two main differences, flexible subcarrier spacing setting and dynamic reference signal allocation [12]. The 5G cells could configure subcarrier spacing from 15 KHz to 240 KHz instead of using the fixed one in 4G. The selection of subcarrier spacing affects quantization steps of delay and Doppler in a similar way as the selection of M, N. In order to accommodate different numerology, the scheduling algorithm assigns M, N based on current numerology Δf and signaling traffic. In 5G, Synchronization Signal Blocks (SSB) are equivalent to the reference signals in LTE. These SSBs are scheduled with more flexible patterns to improve spectrum efficiency. With dynamic SSB allocation, REM can be applied with more compatibility.

B. Relaxed Reliance on Feedback

With the delay-Doppler overlay, REM relaxes the handover's reliance on the feedback for fast and satisfactory triggering (§III-A). 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 (§III-A). 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.

REM's feedback reliance is inspired by the recent advances of cross-band estimation in the static scenarios in the time-frequency domain [21]–[23]. 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, this is realized with non-linear optimization [22] or machine learning [23]. While feasible in the static scenarios, these existing solutions face two challenges 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 (§VI-B).

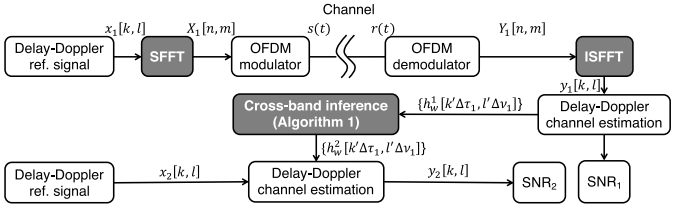


Fig. 9. REM's cross-band channel estimation. Gray boxes are additional modules to OFDM today.

1) REM's Intuition: To overcome these limitations, REM generalizes and simplifies the cross-band estimation in the delay-Doppler domain. This becomes possible with the recently proposed OTFS modulation. REM leverages two insights with delay-Doppler domain representation. First, the delay-Doppler representation shows the multi-path profile explicitly. The cross-band channel estimation only needs to consider the transformation from one frequency to another frequency sharing the same physical paths. Second, REM leverages the decomposition of the multi-path profile into frequency dependent and independent components.

As indicated in the channel representation $h(\tau, \nu)$, two propagation paths could vary in three dimensions, amplitude, delay and Doppler. The delay and attenuation are frequency-dependent. For each physical path, the delay and attenuation are persistent regardless of the operating frequency. The delay is decided by the distance of the path between the cell and the device. The attenuation is decided by both the distance and fading effects. This is the reason why two cells only differ in Doppler shifts in the delay-Doppler domain.

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 $\nu_p^1/\nu_p^2 = f_1/f_2$ (§II). 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.

2) REM's Cross-Band Estimation: We next elaborate on REM's cross-band estimation. With the signaling overlay (§V-A), REM multiplexes 4G/5G's reference signals[8] in the

[8]The cell-specific reference signals in 4G LTE, and CSI-RS in 5G NR [12]. Both are decoupled from demodulation reference signals for data transfer.

delay-Doppler domain (Figure 9). REM first estimate the delay-Doppler channel $\{h_w(k\Delta\tau, l\Delta\nu)\}_{k,l}$ based on expected and received reference signals $(y(k,l), x(k,l))$ in (2) by applying standard channel estimation [24].

REM leverages the path-sharing feature of two cells from the same base station. Note channel estimation in (2) has

$$\frac{1}{MN}h_w(k\Delta\tau, l\Delta\nu) = \sum_{p=1}^{P} \frac{\Gamma(k\Delta\tau, \tau_p)}{M} \cdot h_p e^{-j2\pi\tau_p\nu_p} \cdot \frac{\Phi(l'\Delta\nu, \nu_p)}{N} \quad (3)$$

where $\Gamma(k\Delta\tau, \tau_p) = \sum_{m=0}^{M-1} e^{j2\pi(k\Delta\tau-\tau_p)m\Delta f}$, $\Phi(l\Delta\nu, \nu_p) = \sum_{n=0}^{N-1} e^{-j2\pi(l\Delta\nu-\nu_p)nT}$. To decouple frequency-dependent and frequency-independent terms, we can rewrite it in a matrix form:

$$\mathbf{H} = \mathbf{\Gamma P \Phi} \quad (4)$$

where $\mathbf{H} \in \mathbb{C}^{M\times N}$ is the channel matrix from (2): $H(k,l) = \frac{1}{MN}h_w(k\Delta\tau, l\Delta\nu)$. Note $\mathbf{H}$ can be represented by

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

We have $\mathbf{\Gamma} \in \mathbb{C}^{M\times P}$ to denote the *frequency-independent* path delay spread matrix $\mathbf{\Gamma} = \frac{1}{M}\Gamma(k\Delta\tau, \tau_p)$. $\mathbf{\Phi} \in \mathbb{C}^{P\times N}$ is the *frequency-dependent* path Doppler spread matrix with $\Phi(p,l) = \frac{\Phi(l\Delta\nu,\nu_p)e^{-j(\theta_p+2\pi\tau_P\nu_P)}}{N}$, where θ_p is the frequency-independent path phase: $h_p = |h_p|e^{-j\theta_p}$. Note that $\mathbf{P} \in \mathbb{R}_{\geq 0}^{P\times P}$ is a diagonal matrix that represents the *frequency-independent* attenuation matrix with $\mathbf{P}(p,p) = |h_p|$. After estimating the cell 1's channel matrix $\mathbf{H_1}$, we can decompose it as $\mathbf{H_1} = \mathbf{\Gamma P \Phi_1}$. We note that the path delay $\mathbf{\Gamma}$ and attenuation $\mathbf{P}$ are *frequency-independent*, while the frequency-dependent Doppler shift $\mathbf{\Phi_2}$ can be derived from $\mathbf{\Phi_1}$ given $\frac{\nu_p^1}{\nu_p^2} = \frac{f_1}{f_2}$. Then we can obtain cell 2's channel $\mathbf{H_2} = \mathbf{\Gamma P \Phi_2}$. Such derivation does not requires optimization problem and thus saves computation complexity.

3) Delay-Doppler Decomposition: 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). 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 fact, we can prove their relation as follows (proved in Appendix A of supplementary materials):

Theorem 1 (Cross-Band Estimation With SVD): A delay-Doppler decomposition $\mathbf{H} = \mathbf{\Gamma P \Phi}$ is 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.

Both conditions in Theorem 1 are not hard to satisfy with reasonable M and N. For condition (i), it has been observed that the multi-path are sparse and limited in common scenarios [25]–[27]. Even one smallest 4G/5G physical resource block with $M = 12, N = 14$ can support up to 12 paths, which is sufficient for standardized multi-path models in 4G (7 paths for EPA, and 9 paths for EVA/ETU [28]) and 5G (12 paths for TDL-A/B/C [29]). For condition (ii), it holds if the operator chooses a larger subgrid, e.g., under a common 40ms triggering interval and a 20MHz channel (§III-A), $(M,N) = (1200, 560)$, the distance corresponding to the delay tap is $\Delta\tau c \approx 15$m and the Doppler tap is $\Delta\nu \approx 25$Hz. 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) and the Doppler is approximately multiple times of 25Hz (typically 1150Hz) [30].

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 multi-path profile $\{h_p, \tau_p, \nu_p^2\}_{p=1}^{P_{max}}$ (line 2–8) based on the derivations in Appendix B of supplementary materials. 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.

Under low mobility scenarios, the naive SVD decomposition fails to adapt to the constraint of Doppler spread. We thus design a heuristic-based algorithm to adapt to low mobility scenarios. Doppler spread is limited by the device velocity. Based on Doppler formula, the Doppler spread τ_p is modeled by $\frac{v_s}{c}f$, which is bounded by the maximum value for velocity v_s and carrier frequency f. Thus we can heuristically derive the upper bound of Doppler spread ν_p under the driving case, considering the velocity limit v_s^{max} and f. With such restrictions, we further examine the validity of each inferred path by the estimated ν_p after Algorithm 1. The client could use a static v_s^{max} or change v_s^{max} adaptively.

4) Sparsity for Estimation: We further leverage the sparsity of multi-path propagation to improve estimation accuracy and efficiency. To improve the estimation accuracy, we show that under the single path model, the SVD decomposition holds under no assumptions on Doppler and delay. To improve efficiency, the sparsity of propagation paths serves as a constraint to filter significant paths among possible $min(M,N)$ paths.

Propagation path sparsity in high-speed rails originates from the environmental setting and operator deployments. Cellular networks for high-speed rails require dedicated antennas to serve fast-moving users [30]. The dedicated antenna guarantees that the line-of-sight (LOS) propagation path usually exists. The existence of the LOS path means a dominant path presents higher amplitude [31]. Thus a single path model captures the channel quality. This is why the single path model is adopted by the 3GPP high-speed-train propagation model [28].

We first relax Theorem 1's condition on delay and Doppler with the insight of propagation path sparsity. It relaxes the condition of SVD decomposition because the sparsity of propagation indicates the sparsity of the channel matrix. Theorem 1 sets constraints on path delay and Doppler since the decomposed delay spread matrix $\mathbf{\Gamma}$ and Doppler spread matrix $\mathbf{\Phi}$ have to be semi-unitary. If the delay and Doppler are not integrals, the fractional delay and Doppler cause coupling between paths, destroying the unitarity of the matrix. However, there is no coupling between path when there is a single path.

We prove that when there is a single path, SVD decomposition holds. Theorem 2 shows that the SVD decomposition holds under any delay and Doppler if the number of physical paths is one, as shown in Appendix C of supplementary materials. To summarize, the proof relies on the insight that there is no coupling between paths. Theorem 1 relies the conditions on delay τ_p and Doppler ν_p to eliminate the coupling between paths. After we show that the delay spread matrix $\mathbf{\Gamma}$ and Doppler spread matrix $\mathbf{\Phi}$ are unit vectors, the decomposition is thus valid.

Theorem 2 (Single Path Decomposition): The compact SVD decomposition for a delay-Doppler channel matrix $\mathbf{H} = \mathbf{U\Sigma V}$ is equivalent to decomposition $\mathbf{H} = \mathbf{\Gamma P \Phi}$ with $\mathbf{U} = \mathbf{\Gamma}, \mathbf{\Sigma} = \mathbf{P}, \mathbf{V} = \mathbf{G}$ when $P = 1$.

The condition of Theorem 2 holds when there is a single path. Theorem 2 indicates the correlation between the path amplitude and the decomposed singular value. Even when there are more paths, the dominant path corresponds to the major component of SVD decomposition. During our evaluation in §VI-B, we find out that the decomposition error is small even when there are many propagation paths. This is because the dominant path is decomposed correctly.

5) Avoid Over-Fitting: We adapt Algorithm 1 to filter out weak propagation paths. This is done by setting a threshold on the number of inferred paths to leverage path diversity. In Algorithm 1 line No. 2, we constrain the number of paths as P_{max}. This design will not affect Theorem 1's validity since we perform path pruning after SVD decomposition. Specifically, we first rank all potential paths by their amplitude and select the P_{max} strongest ones. In practice, there are two ways to decide the threshold. The operators can decide the number of paths empirically based on 3GPP standards. For example, the number of paths is 5 for driving scenarios in the reference propagation model.

Another way to decide the threshold is to compare the estimated path loss to the strongest path. Inspired by the correlation between the path amplitude and the decomposed singular value, we rank the decomposed components based on the singular value. The largest singular value is mapped to the strongest path, which dominates estimation accuracy. Other weaker paths can be affected by interference and noises. We thus omit the paths that are lower than the strongest one. The offset to filter paths depends on operators' experiences.

Constraining the number of paths improves the efficiency as well. If $P_{max} = MIN(M, N)$, the optimization is reduced to the naive estimation. With $(M, N) = (1200, 560)$, there are 560 paths derived from SVD decomposition. In such a case, setting $P_{max} = 9$ reduces the computation of path by 62 times.

Algorithm 1 REM's Cross-Band Channel Estimation

Require: Band 1's channel matrix $\mathbf{H_1}$, $H_1(k,l) = h_w^1(k\Delta\tau, l\Delta\nu)$ from (2)

Ensure: Band 2's channel matrix $\mathbf{H_2}$

1: Decompose $\mathbf{H_1} = \mathbf{\Gamma P \Phi_1}$ using SVD matrix factorization;
2: **for** each path $p = 1, 2, \ldots P_{max}$ **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: $\mathbf{H_2} \leftarrow \mathbf{\Gamma P \Phi_2}$;

The idea of constraining the number of paths adheres to the nature of SVD to reduce data dimension. Since SVD extracts the path loss with the singular values in $\mathbf{P}$, the algorithm can easily rank the diagonal matrix and filter out weak ones.

6) Defeat Against Channel Noise: The noises impact channel estimation accuracy and indirectly affect cross-band estimation. REM is robust to noises since it runs in the delay-Doppler domain. According to (2), 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 a more robust channel estimation for h_w. Current OFDM-based channel estimation has shown that the low-rank estimator with SVD decomposition can describe the channel well without being affected by channel noise. Our evaluation in §VI experimentally proves that SVD decomposition performs well under various channel noises.

7) 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 [22], [23] that rely on optimization or machine learning, thus suitable to track the fast-varying channel in extreme mobility.

C. Simplified, Conflict-Free Policy

REM last simplifies the handover policy for high reliability and verifiable correctness (§III-B). 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.

1) REM's Simplification Approach: REM simplifies policy today with cross-band estimation in delay-Doppler domain in §V-B. Figure 10 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 needed events. With SNR as the coherent metric, cells are directly comparable based on information theory as SNR explicitly indicates cell

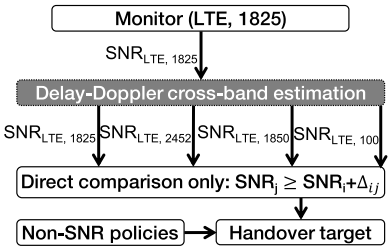


Fig. 10. REM's policy simplification for Figure 6.

capacity $C = B\log(\text{SNR}+1)$ (B is the bandwidth). Note SNR should always be evaluated in handover, regardless of other metrics to be used. Otherwise, "blind handovers" will *always* happen with loops [13], and lose network access if the target cell's coverage is weak;

(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 §V-B. 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 the next step (but still with the same conflict-freedom guarantees below).

(3) Simplify policy with minimal events. REM replace all indirect comparison events (A1, A2, A4, A5) and their thresholds with direct comparison (A3). By removing the multi-stage decision, A1/A2 events are removed. For other events, REM replaces them with A3. To remove A5 used for indirect signal strength comparison between cells, REM uses direct A3 comparison with $\Delta_{A3} = \Delta^2_{A5} - \Delta^1_{A5}$. To remove A4, there are two cases in extreme mobility. First, due to the multi-stage policy, most A4 events occur after A2 is triggered. They are equivalent to A5 with $\Delta^1_{A5} = \Delta_{A2}, \Delta^2_{A5} = \Delta_{A4}$ and replaced by A3 with the above procedure. Second, for load balancing or adding capacity [8], [32], [33], a small amount of A4 events are directly triggered without A2 (§III-B). 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 $C = Blog(SNR+1)$, where Δ_{A3} decides capacity difference.

2) REM's Conflict-Freedom Guarantees: Compared to today's policies in §III-B, REM eliminates most events except A3. This leads to fewer conflicts between events, and simpler conflict resolutions than [13], [14]. We start with the policy with delay-Doppler SNR only. We obtain the following result (proved in Appendix D of supplementary materials):

Theorem 3 (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, $\Delta^{i\to j}_{A3} + \Delta^{j\to i}_{A3} \geq 0$.

Theorem 3 shows that, two-cell threshold coordination is necessary *and* sufficient condition for policy conflict freedom. Compared to the conflict freedom conditions today [13], [14], Theorem 3 is much simpler with fewer events and less threshold coordination between cells. Violation of Theorem 3 happens in extreme mobility when the operator tries proactive handovers to mitigate failures (§III-B). With REM, operators do not need this since REM has mitigated most failures.

REM retains flexibility for operators by supporting non-SNR metrics, such as priorities, traffic load, and access control. We prove Theorem 4 in Appendix E of supplementary materials that with coordinated SNR events, Theorem 3 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.

Theorem 4 (Conflict-Freedom in General): For any settings of non-SNR metrics in REM, satisfying Theorem 3 still guarantees loop-freedom.

VI. Evaluation

We implement and evaluate REM's reliability in extreme mobility (§VII), and its efficiency and overhead of its key components (§VI-B). The details of implementation are presented in Appendix F of supplementary materials.

Experimental Setup: To approximate real extreme mobility, we run trace-driven emulations over a USRP-based testbed. The details of mobility traces are presented in Appendix F of supplementary materials. Our testbed consists of servers running OAI [34] cellular protocol stack and the USRPs as clients and base stations. The servers run OAI [34] cellular protocol stack. We have USRP B210/N210 to test with real channels, which are connected to servers with Intel Xeon CPU E5-2420 v2 and 16GB memory. To emulate operational settings, we configure the testbed based on the above datasets. Specifically, we extract protocol configurations and mobility policies for each cell from the dataset and test with various settings. To compare REM with legacy 4G/5G, we replay the mobility traces from our datasets and evaluate if REM can prevent failures under the same settings. Note we run the USRPs under the unlicensed 2412/2432MHz band instead of licensed ones to comply with FCC regulations.

A. Overall Reliability in Extreme Mobility

We compare REM and legacy LTE on failure ratios $\eta = \frac{K_{LTE}}{K}$ and reduction $\epsilon = \frac{K_{LTE}-K_{\text{REM}}}{K_{\text{REM}}}$, 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 §V-A–§V-B, we assume REM can prevent them only if it reduces the BLER to 0. This under-estimates REM's failure reduction since signaling may be delivered with non-zero BLER. For failures from missing cells in the multi-stage policy in §V-C, the client will eventually reconnect to a missed candidate cell if its SNR is better than the old cell. We use missed cell's SNR to check whether REM guarantees successful handover before the client loses connection. Since SNR is not collected in Beijing-Shanghai dataset, we do not assess REM's failure reduction for missing cells and thus underestimate its effectiveness. Table V shows REM's reduction of network failures and policy conflicts.

1) Overall Reliability Improvement: Table V shows REM reduces the overall failures and conflicts in both HSR datasets at all train speeds. In Beijing-Shanghai route, REM

TABLE V
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%	0.03%	25.0×	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	∞

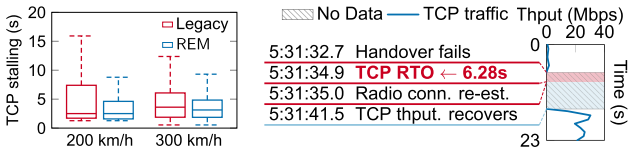
reduces existing LTE's failure ratio by 1.2× (5.2%→2.4%) at 100-200km/h, 3.0× (10.6%→2.6%) at 200–300km/h, and 2.6× (12.5%→3.5%) at 300-350km/h. In Beijing-Taiyuan route at 200–300km/h, REM the failure ratio by 0.9× (8.1% →4.2%). In all cases, REM achieves comparable failure ratios to static and low-speed mobility (e.g., driving in Table II). Note that these reductions consider the unavoidable failures from coverage holes. 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.

2) Failure Reduction in Triggering: With the stabilized signaling (§V-A), 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 (§V-B).

3) Failure/Conflict Reduction in Decision: By eliminating the multi-stage policy, REM mitigates the failures from missed inter-frequency cells (3× reduction in Beijing-Taiyuan dataset). With the 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 §V-C, 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 (§VI-B).

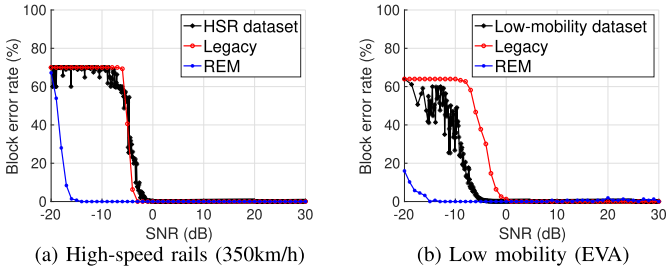
4) Failure Reduction in Execution: REM reduces its failures to 0–0.4%. Our dataset shows many handover commands in OFDM-based LTE are corrupted/lost with acceptable SNR ([−5dB, 0dB]). Instead, REM explores the full frequency-time diversity in delay-Doppler domain to mitigate the signaling errors/corruptions.

5) Benefits for Data Transfer: 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. We replay the iperf's TCP data transfer in the tcpdump traces and quantify their TCP performance with/without 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 11a shows REM's TCP stalling time reduction. With less failures, REM reduces the average TCP stalling from 7.9s to 4.2s (46.8% reduction) at 200km/h, and from 6.6s to 4.5s (31.8% reduction)



(a) TCP stalling time (b) TCP data transfer in failures

Fig. 11. REM's benefit for TCP.



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

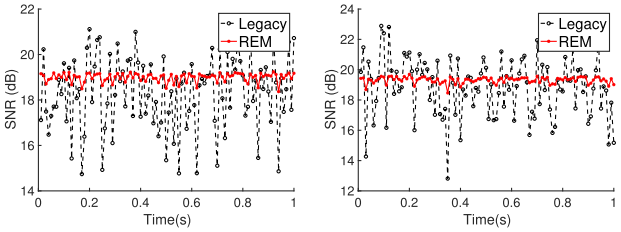
Fig. 12. REM's error reduction for signaling.

at 300km/h. Note that TCP stalling time is usually longer than the network failures because of its retransmission timeout (RTO). This is exemplified in Figure 11b: When a 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.

B. Efficiency and Overhead

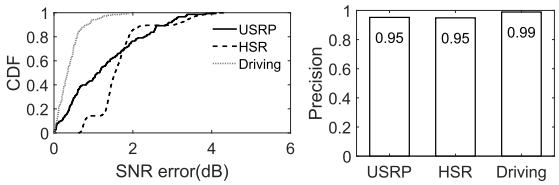
1) Stabilized Signaling in Delay-Doppler Domain (§V-A): We first examine how delay-Doppler domain reduces signaling errors/loss. We replay our datasets with same signaling message length and SNR, and evaluate their BLER in a 4G/5G subframe ($M = 12, N = 14$ for 1ms [11], [12]) in standard reference multipath models for high-speed train and driving [35], [36]. Figure 12 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 13 compares REM and legacy LTE's SNR in the same setting. In OFDM, slots in different carrier frequency and time experience different channel gains $H(f, t)$ and thus diverse SNRs. Instead, REM



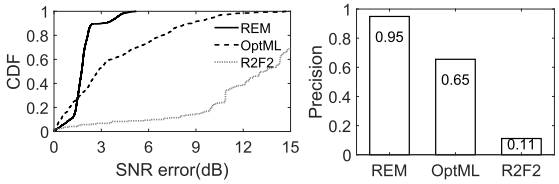
(a) High-speed rails (350km/h) (b) Low mobility (EVA)

Fig. 13. Stabilized delay-doppler domain.



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

Fig. 14. Viability of REM's cross-band estimation.



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

Fig. 15. Cross-band estimation with HSR.

adopts OTFS to spread signaling traffic across the entire time-frequency grid, explores the full frequency/time diversity and results in stable channel gains $h_w(\tau, \nu)$ for all slots (Equation 2). This results in more stable SNRs, and facilitates SNR-based policy in REM and less transient loops.

2) Relaxed Feedback (§V-B): We evaluate two aspects. First, we quantify the accuracy of REM's cross-band estimation by estimated SNR errors. Figure 14 shows that, REM can achieve $\leq$2dB estimation errors for $\geq$90% measurements. Then we evaluate whether REM's cell estimation can trigger the same events for handover. With our dataset, we extract all handovers' measurements and triggering events/thresholds to estimate handover decision precision. Figure 14 shows that, REM can achieve $\leq$2dB estimation errors for $\geq$90% measurements, and correctly triggers $\geq$90% handovers. We also evaluate how heuristic-based approach improves the vanilla SVD decomposition, the average SNR error is reduced by 77.8% (1.85 dB to 0.41 dB). We further validate the impact of fractional Doppler when Theorem 1 does not hold. We evaluate how the fractional Doppler affects SNR error. As shown in Figure 16, the SNR error is below 2dB.

We further compare REM's accuracy with R2F2 [22] and OptML [23], 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 a fair comparison, we empirically find their optimal configuration (6 paths for both R2F2 and OptML), and show the results under this setting. Moreover,

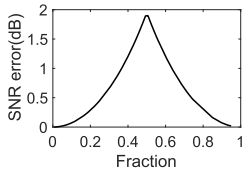
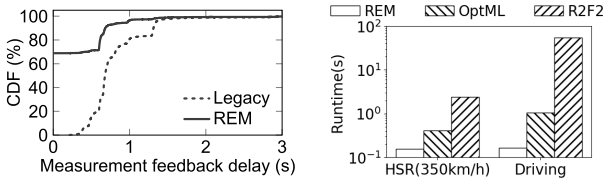


Fig. 16. Impact of fractional Doppler.



(a) Feedback delay reduction. (b) Cross-band estimation

Fig. 17. Delays in REM.

to train the OptML model, we randomly choose 80% data from the HSR dataset, and use the remaining 20% data to test OptML. Figure 15 shows 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 §V-B, 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, REM reduces its measurement duration (including the triggering interval in §III-A) and round-trips of feedback (totally T_1). Meanwhile, REM incurs extra runtime of cross-band estimation T_2, so the feedback latency savings is $T_1 - T_2$. Figure 17a 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 17b 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$\times$ (1.6$\times$) reduction. While it is possible to accelerate R2F2 and OptML with advanced hardware (e.g., FPGA and GPU), such a solution is too expensive for the resource and energy-constrained mobile devices.

3) Simplified, Conflict-Free Policy (§V-C): As shown in Table V, REM's simplified policy provably prevents conflicts. One may wonder if eliminating the conflicts will cause more failures. We show REM prevents this situation. For all the conflict-prone handover, we follow Theorem 3 and 4 to update thresholds, and repeat the evaluation in §VII to evaluate if more failures will happen. Figure 18 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.

VII. Benefits for Applications

How can REM benefit real applications with enhanced mobility and efficiency? In this section, we evaluate the performance improvement for emerging applications with stringent latency requirements like AR/VR.

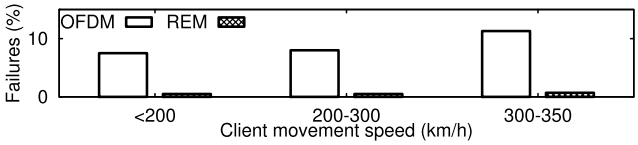
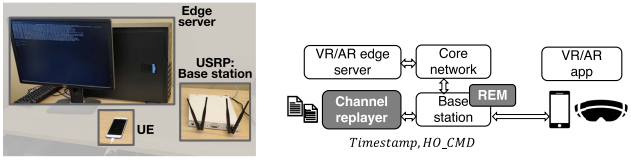


Fig. 18. Failures without aggressive policies.



(a) Testbed (b) Emulator architecture

Fig. 19. Testbed setup for edge-based VR/AR.

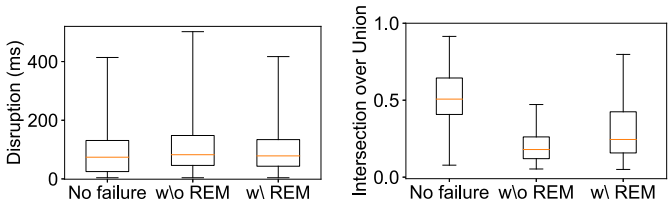
A. Experimental Setup

To evaluate the performance of mobile VR/AR applications, we use Pixel 2XL as the client and deploy the edge server with the testbed (§VI). The overall setup is shown in Figure 19. The client is connected with the base station within the coverage of the USRP, so handover will not be triggered without channel dynamics in the lab environment. Therefore, we replay HSR traces to emulate the scenario with and without REM. Specifically, we inject handover to the testbed and control the delay based on replayed traces.

We launch AR/VR demo applications to test performance. The following settings are consistent with either legacy mobility management or REM. We deploy the VR and AR engine at the edge server co-located with the core network. For the mobile VR application, we consider VR streaming of medium quality with the same setting as [37]. The client sends motion updates to the server, expecting to receive a streamed VR frame and render the view with the updated frame. We evaluate performance by checking whether the request frame is missing after the client renders an updated view. In our experiments, we let the client send periodic motion updates. For the mobile AR application, the client streams real-time video to the edge server for object detection. After receiving the identified location of recognized objects, the client will render the bounding box of the object on the current frame. If the object recognition result is delayed due to network failure, the rendered bounding box might not overlap with the ground truth bounding box. We use the same streaming content for Mobile VR and AR to guarantee the results are not affected by streaming content. We quantify the timeliness of the recognition result for evaluation.

B. Disruption Reduction for Mobile VR

We evaluate the disruption that a requested frame is missing when the user updates its view. Figure 20a shows REM reduces the median (95%ile) disruption from 82.5 (508.7) ms to 78.5 (418.4) ms for affected frames. We also evaluate the disruption under the static case. The median (95%ile) disruption is 74.0 (415.8) ms, which proves REM reduces the median disruption added by handover failure by 47.1%. Note not all frames experience disruption, we find that the percentage



(a) VR frame disruption (b) IoU for AR object recognition

Fig. 20. Performance of mobile VR/AR.

of affected frames is similar for the case with and without REM. REM outperforms legacy 4G/5G since it mitigates the disruption by reducing failure-caused disruption to normal handover latency.

C. Recognition Performance for Mobile AR

We evaluate the performance based on Intersection over Union (IOU), which is a common metric to evaluate whether the identified object bounding box matches with the ground truth in object detection and tracking [38]. Figure 20a shows shows REM improves the median (95%ile) IoU from 0.18 (0.43) to 0.24 (0.59) for affected frames. To quantify the overall performance, we take the IoU threshold as 0.25 as proposed in [38]. The ratio exceeding the threshold is 49.7% (88.3% improvements) with REM compared with 26.4% in the case without REM. We notice that REM's benefit is more significant under low-IoU samples. This is because REM reduces the probability of handover failure where IoU is low due to failure-caused disruption.

VIII. Discussion

A. Applicability of REM

REM is applicable to all modes of mobility and other application scenarios in general. REM is applicable to ALL mobility, not just the extreme case, as validated in our extensive evaluation. REM also applies to various application scenarios. The design of movement-based mobility management does not rely on a fixed trajectory. The inherent reason is that movement evolves much slower than wireless. As long as the client and the network both adopt OTFS-based signaling, REM could benefit all mobile scenarios, e.g., IoT, drones, etc.

B. Impact on Performance

While REM does not explicitly target performance, REM also benefits data performance. With reduced handover failures and policy conflicts, the clients are faced with less disruption. Besides, REM's cross-band estimation enables measurement without MeasurementGap for inter-frequency cells, thus saving more spectrum for data transfer. What's more, if data also uses OTFS, REM's SNR-based policy would select the cell with high capacity $C = B\log(\text{SNR}+1)$, thus improving the data speed. We note that dual connectivity could potentially leverage the SNR-based policy to choose the cells with the highest capacity to aggregate. We leave that as future work.

C. REM's Deployability

REM requires deployment at both the device side and the network side to support OTFS based communication.

In order to support OTFS, only a signal processing module (ISFFT/SFFT) needs to be implemented and all OFDM modules can be reused. The main challenge is how to incrementally roll out the OTFS based signaling. REM handled the co-existence of OFDM and OTFS with a dynamic scheduler. With the scheduler, the base station adaptively allocates resources to OTFS grids and OFDM grids, enabling incremental deployment of REM.

D. Compatibility

REM is backward compatible to both and benefits them too. It also reduces their failures (although less common than extreme mobility), saves the signaling overhead from excessive feedback, and eliminates their policy conflicts. REM is backward compatible since it only requires an optional overlap on top of OFDM in legacy 4G/5G. If either the client or network does not support REM, it can seamlessly rollback to legacy 4G/5G by disabling REM overlays.

IX. Related Work

Reliable and fast mobility management has been an active topic for years. Most efforts follow the wireless signal strength-based design today and explore how to refine its signaling procedures [39], [40], handover decision [37], [41], transport-layer data speed in mobility [9], [42], policy conflicts [13], [14], to name a few. These approaches are limited since they still follow the wireless-signal-strength based approach. There are also related works focused on how to improve the reliability of mobility on the device side [37], [43], [44]. However, these approaches are limited since the device has to follow the network's decision. They do not solve the fundamental problem underlying the instability of OFDM signals. 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 [45]–[47], but in a different scenario in mobility management. REM leverages the delay-Doppler domain from the radar community and recent advances in OTFS modulation [4], [20], [48]. But REM moves beyond wireless modulation and generalizes to mobility management. REM's relaxed feedback in §V-B extends the cross-band estimation in [22], [23] to mobility scenarios, and simplifies the estimation in the delay-Doppler domain.

X. 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 that 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.

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 geometric modeling, and historical base station measurements. We hope REM stimulates more efforts toward predictable, robust mobile networks.

References

[1] Y. Li, Q. Li, Z. Zhang, G. Baig, L. Qiu, and S. Lu, "Beyond 5G: Reliable extreme mobility management," in *Proc. Annu. Conf. ACM Special Interest Group Data Commun. Appl., Technol., Architectures, Protocols Comput. Commun.*, New York, NY, USA, Jul. 2020, pp. 344–358.

[2] (2019). Wikipedia. *High-Speed Rail in China*. [Online]. Available: https://en.wikipedia.org/wiki/High-speed_rail_in_China

[3] (2019). Bloomberg. *Vehicle-to-Everything Market Communications Ecosystem 2030 Account for a Market Worth $1.2 Billion by 2022*. [Online]. Available: 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

[4] R. Hadani *et al.*, "Orthogonal time frequency space modulation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[5] *Radio Resource Control (RRC)*, document TS36.331, 3GPP, Mar. 2015.

[6] *5G NR: Radio Resource Control (RRC)*, document TS38.331, 3GPP, Jun. 2019.

[7] (2019). Wikipedia. *Coherence Time (Communications Systems)*. [Online]. Available: https://en.wikipedia.org/wiki/Coherence_time_(communications_systems)%

[8] Y. Li, C. Peng, Z. Yuan, J. Li, H. Deng, and T. Wang, "Mobileinsight: Extracting and analyzing cellular network information on smartphones," in *22nd ACM Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, Oct. 2016, pp. 202–215.

[9] J. Wang *et al.*, "An active-passive measurement study of TCP performance over LTE on high-speed rails," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, Aug. 2019, pp. 1–16.

[10] *E-UTRA and E-UTRAN; Overall Description; Stage 2*, document TS36.300, 3GPP, 2011.

[11] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation*, document TS36.211, 3GPP, 2017.

[12] *5G NR; Physical Channels and Modulation*, document TS38.211, 3GPP, Jun. 2019.

[13] Y. Li, H. Deng, J. Li, C. Peng, and S. Lu, "Instability in distributed mobility management: Revisiting configuration management in 3G/4G mobile networks," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Modeling Comput. Sci.*, Antibes Juan-les-Pins, France, Jun. 2016, pp. 261–272.

[14] Z. Yuan, Q. Li, Y. Li, S. Lu, C. Peng, and G. Varghese, "Resolving policy conflicts in multi-carrier cellular access," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, New Delhi, India, Oct. 2018, pp. 147–162.

[15] *5G NR: Overall Description; Stage-2*, document TS38.300, 3GPP, Jan. 2020.

[16] A. Narayanan *et al.*, "A variegated look at 5G in the wild: Performance, power, and QoE implications," in *Proc. ACM SIGCOMM Conf.*, New York, NY, USA, Aug. 2021, pp. 610–625.

[17] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. CS-11, no. 4, pp. 360–393, Dec. 1963.

[18] A. Monk, R. Hadani, M. Tsatsanis, and S. Rakib, "OTFS–orthogonal time frequency space," 2016, *arXiv:1608.02993*.

[19] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," 2018, *arXiv:1802.02623*.

[20] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501–6515, Oct. 2018.

[21] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. Future Netw. Mobile Summit*, Jun. 2010, pp. 1–10.

[22] D. Vasisht, S. Kumar, H. Rahul, and D. Katabi, "Eliminating channel feedback in next-generation cellular networks," in *Proc. ACM SIGCOMM Conf.*, Aug. 2016, pp. 398–411.

[23] A. Bakshi, Y. Mao, K. Srinivasan, and S. Parthasarathy, "Fast and efficient cross band channel prediction using machine learning," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2019, p. 37.

[24] (2018). MATLAB. *Channel Estimation*. [Online]. Available: https://www.mathworks.com/help/lte/ug/channel-estimation.html

[25] K. Joshi, D. Bharadia, M. Kotaru, and S. Katti, "WiDeo: Fine-grained device-free motion tracing using RF backscatter," in *Proc. 12th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2015, pp. 189–204.

[26] T. Wei, A. Zhou, and X. Zhang, "Facilitating robust 60 GHz network deployment by sensing ambient reflectors," in *Proc. 14th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2017, pp. 213–226.

[27] N. Czink, M. Herdin, H. Ozcelik, and E. Bonek, "Number of multipath clusters in indoor MIMO propagation environments," *Electron. Lett.*, vol. 40, no. 23, pp. 1498–1499, Nov. 2004.

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

[29] *Base Station (BS) Radio Transmission and Reception*, document TS38.104: NR, 3GPP, Oct. 2019.

[30] Y. Tang, "The research on LTE coverage solutions on high-speed railway," *Designing Techn. Posts Telecommun.*, vol. 12, no. 1, pp. 20–23, Dec. 2014.

[31] C. X. Wang, A. Ghazal, B. Ai, P. Fan, and Y. Liu, "Channel measurements and models for high-speed train communication systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 974–987, 2015.

[32] H. Deng, C. Peng, A. Fida, J. Meng, and Y. C. Hu, "Mobility support in cellular networks: A measurement study on its configurations and implications," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 147–160.

[33] (2016). Huawei. *Intra-RAT Mobility Management in Connected Mode Feature Parameter Description*. [Online]. Available: https://www.honorcup.ru/upload/iblock/164/6.pdf

[34] (2019). OpenAirInterface. [Online]. Available: https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/home

[35] *Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Radio Transmission and Reception*, 3rd Generation Partnership Project (3GPP), Jul. 2017.

[36] *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception*, 3rd Generation Partnership Project (3GPP), Oct. 2019.

[37] Z. Tan, Y. Li, Q. Li, Z. Zhang, Z. Li, and S. Lu, "Supporting mobile VR in LTE networks: How close are we?" in *Proc. Abstr. ACM Int. Conf. Meas. Model. Comput. Syst.*, Jun. 2018, pp. 1–31.

[38] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a benchmark for multi-target tracking," 2015, *arXiv:1504.01942*.

[39] Z. A. Qazi, M. Walls, A. Panda, V. Sekar, S. Ratnasamy, and S. Shenker, "A high performance packet core for next generation cellular networks," in *Proc. Conf. ACM Special Interest Group Data Commun.*, Aug. 2017, pp. 348–361.

[40] Y. Li, Z. Yuan, and C. Peng, "A control-plane perspective on reducing data access latency in LTE networks," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, Snowbird, UT, USA, Oct. 2017, pp. 56–69.

[41] S. Xu, A. Nikravesh, and Z. M. Mao, "Leveraging context-triggered measurements to characterize LTE handover performance," in *Proc. Int. Conf. Passive Act. Netw. Meas.* Cham, Switzerland: Springer, 2019, pp. 3–17.

[42] L. Li *et al.*, "A measurement study on multi-path TCP with multiple cellular carriers on high speed rails," in *Proc. Conf. ACM Special Interest Group Data Commun.*, Aug. 2018, pp. 161–175.

[43] M. Asefi, J. W. Mark, and X. S. Shen, "A mobility-aware and quality-driven retransmission limit adaptation scheme for video streaming over VANETs," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1817–1827, May 2012.

[44] S. Shi, V. Gupta, and R. Jana, "Freedom: Fast recovery enhanced VR delivery over mobile networks," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2019, pp. 130–141.

[45] B. Karp and H.-T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. MobiCom*, 2000, pp. 243–254.

[46] A. Rao, C. Papadimitriou, S. Shenker, and I. Stoica, "Geographic routing without location information," in *Proc. 9th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2003, pp. 96–108.

[47] Y.-J. Kim, R. Govindan, B. Karp, and S. Shenker, "Geographic routing made practical," in *Proc. 2nd Conf. Symp. Netw. Syst. Design Implement.*, vol. 2, 2005, pp. 217–230.

[48] P. Raviteja, K. T. Phan, and Y. Hong, "Embedded pilot-aided channel estimation for OTFS in delay–Doppler channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4906–4917, May 2019.