Entropy Field Decomposition Based Outage Detection for Ultra-Dense Networks

Ahmad Asghar*, Hasan Farooq*, Haneya Naeem Qureshi*, Adnan Abu-Dayya†, Ali Imran.*,
*Department of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK 74135 USA
†QMIC, Qatar Science & Technology Park, Doha, Qatar

Abstract—**Ambitious quality of experience expectations from 5G mobile cellular networks have spurred the research towards ultra-dense heterogeneous networks (UDHNs). However, due to coverage limitations of millimeter wave cells and lack of coverage data in UDHNs, discovering coverage lapses in such 5G networks may become a major challenge. Recently, numerous studies have explored machine learning-based techniques to detect coverage holes and cell outages in legacy networks. Majority of these techniques are susceptible to noise in the coverage data and only characterize outages in the spatial domain. Thus, the temporal impact of an outage, i.e., the duration of its presence remains unidentified. In this paper, for the first time, we present an outage detection solution that characterizes outages in both space and time while also being robust to noise in the coverage data. We do so by employing entropy field decomposition (EFD) which is a combination of information field theory and entropy spectrum pathways theory. We demonstrate that compared to other techniques such as independent component analysis and k-means clustering, EFD returns accurate detection results for outage detection even in the presence of heavy shadowing in received signal strength data which makes it ideal for practical implementation in emerging mobile cellular networks.**

Heterogeneous Network, Self Healing, Outage Detection, Millimeter Wave, Information Field Theory, Entropy Spectrum Pathways.

I. INTRODUCTION

QUALITY of experience (QoE) enhancement compared to legacy mobile cellular networks is the primary drive of 5th Generation mobile cellular networks [1]. To enable this QoE enhancement, 5G networks will rely on a combination of factors including 10x more throughput, less than 1 ms latency, and 10x more battery life than 4th Generation mobile cellular networks [2]. To meet these expansive requirements, several solutions have been proposed [3], with network densification [4] and millimeter wave (mmWave) spectrum utilization [5] among the most popular. It has been demonstrated that a combination of network densification and mmWave spectrum deployment could potentially yield exponential increase in area spectral efficiency [6].

However, network densification and mmWave spectrum utilization are not without their own limitations. Ultra-dense heterogeneous networks (UDHNs) are prone to generating sparse network coverage information due to low user density per cell [7]. On the other hand, mmWave cells are subject to very high pathloss due to their operation in 30GHz - 300GHz band. One solution to reduce the impact of high pathloss in mmWave cells is the deployment of highly directional antennas with beam-widths as low as $7°$ [8]. However, this opens mmWave cell networks to the problem of very large coverage gaps that must be filled by additional antennas per cell compared to traditional macro cells or by employing umbrella macro cells.

The challenges above highlight the difficulties of ensuring reliable and omnipresent coverage in mmWave-UDHNs, especially considering the QoE requirements for 5G networks. Even without these challenges, ensuring coverage in mobile cellular networks requires continuous network performance testing and monitoring.

There are several causes of gaps in network coverage. Gaps in network coverage may be a result of poor network planning. The network planning stage involves the determination of many parameters, such as the optimal number of base stations, the best locations to install base stations, the types of base station optimal for each location, the configuration of parameters such as antenna height, number of sectors and sector orientation, tilt, power, frequency reuse pattern, capacity dimensioning (e.g. number of carriers or carrier components per sector) [9]. Erroneous configuration of these parameters during the planning stage results in gaps in network coverage. Gaps in network coverage may also arise when changes occur in the radio environment. This could be due to environmental factors, such as changes in weather conditions, vegetation, topology, or construction of new infrastructures (e.g., high rise buildings or bridges) which could result in degradation of network coverage due to blocking. Gaps in network coverage could also result from partial or complete outages. This could be due to hardware and software failures (e.g., radio board failure, channel processing implementation error etc.) and external failures such as power supply or network connectivity failures [10].

While some cell outages are detected by operations support system functions using performance counters or alarms, some outages may not be detected for hours or even days. These outages are often detected through continuous long term performance analysis or subscriber complaints [10]. Therefore, coverage gaps due to poor network configuration or changes in radio environment causes (also known as coverage holes) require mobile cellular network operators to invest heavily in regular network coverage testing, usually through drive-tests. However, the process is time and resource consuming while lacking comprehensiveness due to inaccessibility of a major portion of the network i.e., all areas other than paved roads. On the other hand, identifying and resolving cellular network outages, which are a consequence of software or

hardware failure of network entities and do not raise an explicit alarm in operation and maintenance, requires highly trained engineers parsing gigabytes of network health logs and network performance indicator data looking for outages. This requires manual analysis and may require unplanned site visits, which makes cell outage detection a costly task [10]. Given the continuous growth in cell density and increasing pressure to reduce operational costs [11], both of the above approaches are quickly becoming impracticable [12], [13].

A. Related Work

To address the coverage hole and outage detection problem, 3GPP has introduced the minimization of drive test (MDT) reports feature [14] that we use for baseline coverage data in our work. MDT reports are a solution to the challenges of periodic drive testing, high carbon footprint, and rising operational costs of mobile cellular networks. These reports consist of serving and neighboring cells identities, downlink received power levels, and channel quality measurements. The measurements are collected periodically by user equipment (UE) in both connected and idle modes and are reported back to the network along with their location tags. This offers network providers a more detailed view of network coverage, including indoor network coverage, compared to drive tests. MDT reports have been used for designing coverage hole [15]–[17], outage detection solutions [18]–[24], as well as other self-organizing network (SON) functions [25]–[28].

1) Coverage hole and outage detection in mobile cellular networks: Given the significance of the problem in network service management, in the last few years, cell outage detection has been studied extensively. Only a small subset of studies is discussed here that is representative of the larger discussion on state-of-the-art in outage detection research. For a more comprehensive review of outage detection along with outage diagnosis and compensation techniques, the reader is referred to a recent review paper [29].

Most prior studies on outage detection have focused on homogeneous macro cellular networks [20], [21], [23], [24], [30]–[34], as large number of users per cell in macro cell provide enough data for classic machine learning methods to be trained for coverage anomaly detection. For example, in [30] the authors employ k-nearest neighbor (kNN) [35] and local outlier factor (LOF) [36] techniques to detect coverage anomalies in macro cell environment. The authors use the two clustering techniques to separate cells into normal and anomalous based on their received power measurements from MDT data, and use expert analysis to determine the accuracy of anomaly detection. In [21] the authors compare LOF with one-class support vector machines (SVM) [37] using MDT data with different levels of shadowing and inter-site distances. The authors compare the Receiver Operating Characteristic curves of the two techniques and demonstrate that for the same shadowing and inter-site distance, one-class SVM outperforms LOF considerably. Another work [23] detect outages (or sleeping cells) in a smart-city LTE telecommunications infrastructure using machine learning, where network users are represented by IoT devices. The authors in [23] compare

performance of several machine learning algorithms, such as support vector machines and ensemble learners (bagged decision trees, random forests and extra trees), decision trees and naive bayes. However, the focus of [23] was not to investigate performance of these algorithms under varying propagation and shadowing conditions and remained confined to the received power being computed according to the empirical Cost-Hata propagation model only. Another work addressing the same problem of sleeping cell detection in macro cells is [24]. In this work, the authors use deep autoencoders and one class support vector machine using the parameters of RSRP, SINR, location and cell ID. However, the authors do not study the impact of varying propagation conditions or shadowing in this work and the standard deviation of shadowing is limited to 4 dB for line-of-sight scenarios and 6 dB for non-line-of-sight scenarios. Using RSRP and RSRQ measurements, authors in [20] propose a cell outage detection method using an autoencoder. The authors in [20] first train the autoencoder neural network by the measurement reports from mobile stations, and then compare the reconstruction error of a new measurement report with a specified decision threshold to differentiate between outage and normal class. However, this work [20] does not mention the shadowing level or propagation environment. Another MDT-based approach to investigate cell outage detection in heterogeneous is proposed in [38], where a dynamic affinity propagation algorithm is used. Outages in this work are generated by reducing the transmission power of a specific cell to represent the hardware failure of cell outage beforehand. However, the effect of shadowing on proposed cell outage detection algorithm is not the focus of this work. Moreover, these works [20], [21], [23], [24], [30]–[34], [38] address the outage detection problem in a homogeneous macro cell environment only.

Apart from the use of MDT data for outage detection, some studies such as [31]–[33] also propose to use cell level performance metrics such as uplink and downlink data rates, radio link failures, and handover failures to detect network outages. In [31], [32] the authors use SVM and auto-regressive integrated moving average (ARIMA) to identify network anomalies using network throughput data. The authors construct healthy network performance models using the two techniques and predict future cell performance data from those models. If there is a significant deviation between actual and predicted data, the algorithm determines that the cell is in outage. In [33] the authors use handover and radio link failure data to predict network outages in the network. The authors create a diffusion map of changes in user associations due to handover and call failure events. The solution then maps the user association changes to cell dominance areas in the diffusion map. Finally, the solution runs k-means clustering algorithm [39] to detect the cell with abnormal changes in user associations.

A key insight from the studies discussed above is that the availability of sufficient data for model training and outage prediction is not a major concern in homogeneous macro cell networks. However, the same does not hold true for heterogeneous networks, especially with very high cell density as it results in very small number of users per cell (¡ 2 UE/cell)

[7]. To address the training data sparsity challenge, the authors of [22] propose to generate cell coverage maps from sparse network coverage data by employing Grey prediction model [40] to interpolate coverage data between randomly distributed locations of users sending MDT reports. In [22] the authors use LOF and kNN to detect outages from the data generated using Grey prediction. Another work [18] addresses the training data sparsity challenge through Generative Adversarial Network (GAN). In this work, the authors present a method that is able to learn from imbalanced cell outage data in cellular networks, through GAN and Adaboost. Using RSRP and RSRQ dataset, first, GAN is utilized to change distribution of imbalanced dataset by generating more synthetic samples for minority class, and then Adaboost is used to classify the dataset. However, the focus of this work is detection of outages in a macro cell environment with one level of shadowing only.

In contrast to the plethora of studies on cell-wide outage detection, only a few studies have addressed the problem of coverage hole detection [15]–[17]. In [15] the authors argue that MDT report data does not necessarily represent a complete network coverage picture. Therefore, they apply Bayesian prediction framework on a real network MDT report data to predict missing data and generate radio network coverage maps. These maps are then used to predict the presence of coverage holes by estimating their likelihood based on neighboring pixel. The algorithm in [15] is extended in [16] to use four neighboring pixels, instead of one, to construct the network coverage environment map. Moreover, the authors use Bayesian kriging and interference cartography to generate these maps, which improves the prediction accuracy. In [17] the authors propose to use a combination of radio link failures along with a piece-wise deterministic coverage model to predict the boundaries of coverage and, consequently, coverage holes. The use of deterministic coverage model in [17] eliminates the uncertainty in coverage estimation due to shadowing.

Despite advancements in coverage hole and outage detection described above, there exist several key issues that need to be addressed for such solutions to be applicable to mmWave UDHNs:

a) Sensitivity to shadowing: Most of the existing cell outage detection solutions are highly sensitive to shadowing. This is a key observation made in [21] where authors investigate the impact of degree of shadow fading on the accuracy of several outage detection algorithms. The authors have shown that as the standard deviation of shadowing increases, machine learning-based outage detection models become less accurate ultimately becoming analogous to a coin toss. Similar conclusions are made in [19], where the authors propose a Long-Short Term Memory based femtocell outage detection solution in a multi-tier network. The authors in [19] conclude that when the shadow fading standard deviation is increased from 4 dB to 8 dB, the overall detection accuracy decreases significantly, which implies that outage detection task becomes more difficult under harder shadowing conditions. Moreover, the study in [41], which uses a hidden markov model based outage detection algorithm, also concludes that the detection accuracy degrades with higher shadowing standard deviation.

The same is true for accuracy of coverage hole detection solutions which rely on neighboring pixel data to predict the existence of a coverage hole [15], [16]. Since heavy shadowing will be a feature of mmWave-UDHNs, a practical solution for coverage hole and cell outage detection in such networks must be robust to the effects of shadowing.

b) Inclusion of Spatio-Temporal Domains: Network coverage is susceptible to randomness due to shadowing in both spatial and temporal domains. However, prior studies, such as the ones discussed above, only consider coverage information in terms of spatial snapshots taken at certain time instants. This implies that the solutions dependent on this information run at each time instant with no information cascading to subsequent time snapshots. This leads to the possibility of instantaneous coverage holes or outages triggering outage compensation algorithms if they occur at the time when the snapshot of the coverage is being built, or coverage holes or outages being missed out between the snapshots. Even the solutions based on time-averaged models such as [31], [32], [42] do not offer the temporal depth to deal with these instantaneous coverage fluctuations. Similarly, while some studies employ temporal coverage data to identify network anomalies [43]–[46], they do not explicate the complete spatial impact of these anomalies on the subscribers or their QoE. Given these issues, any coverage hole or outage detection solution must consider both spatial as well as temporal domains in the detection process.

c) Sensitivity to data distribution model: Many prior studies discussed above that leverage tools such as k-means, LOF, SVM, Bayesian kriging, or Bayesian prediction frameworks build on specific distribution that data and noise are assumed to follow. Many of these solutions assume data distribution to be either Gaussian, exponential or some variation of the two. However, acute dynamics of mobile environment, particularly those associated with mmWave UDHN can make these assumptions invalid. Therefore, an outage detection solution that does not rely on a-priori assumptions about distribution of data is highly desirable, particularly to be applicable for mmWave UDHN.

d) Coverage Hole and Outage Detection in mmWave UDHNs: Recent study [47] shows that likelihood of outages increases with cell density as well as complexity of the cell hardware which will be the case in mmWave-UDHNs. While some outage detection solutions have been proposed to incorporate heterogeneous network topology, to the best of our knowledge, a coverage hole or outage detection solution that explicitly targets mmWave-UDHNs while addressing idiosyncrasies of such networks does not exist.

B. Proposed Approach and Contributions

In this paper a new approach for anomaly detection is proposed, one that is based on entropy field decomposition (EFD), an algorithm first introduced by Frank and Galinsky [48]. EFD has previously been used successfully for brain activity mode detection in biomedical engineering [49] and the study of severe weather phenomenon [50]. However, EFD has not been applied in the context of wireless communication networks. Results and observations by applying EFD-based outage detection algorithm in the context of wireless

communication networks show that such a solution can overcome several limitations of existing solutions. Existing outage detection solutions mostly apply statistical learning that are not robust to shadowing. Their performance is sensitive to data and noise distributions and they do not take coverage profiles over a range of time as inputs to the outage detection algorithms. Moreover, they mainly focus on outage detection in macro cell environments and not in heterogeneous ultra-dense environments. The analysis and observations from this study show that an EFD-based approach offers a promising solution to detect outages in such unconsidered scenarios.

The motivation for leveraging EFD to solve cell outage and coverage hole detection problem stems from its ability to identify the flow of information in data over both space and time by combining information field theory [51] and entropy spectrum pathways theory [52]. Since outages in a cellular network vary both spatially and temporally, an EFD based approach to detect outages can be a promising solution. Furthermore, EFD is independent of baseline data model/distribution and it actively suppresses the effects of noise in activity mode detection process. Since mmWave-UDHN environments are marked by heavy shadowing with varying data distributions, it makes coverage data in such environments noisy. Therefore, an EFD based solution with the ability to suppress the effect of noise can be a favorable solution for coverage hole and outage detection in mmWave-UDHN environments.

The key contributions of this paper can be summarized as follows:

- We present an EFD-based coverage hole and outage detection solution that is independent of the signal propagation model. This means that the solution can seamlessly be integrated into other autonomous solutions regardless of underlying coverage model.
- The proposed solution can identify and overcome the impact of shadowing in the coverage data over both space and time. This allows the solution to detect lapses in coverage even in the event of heavy signal dispersion and shadowing. We demonstrate this by comparing the accuracy of the solution in presence of different shadowing levels.
- The proposed solution can incorporate data from both spatial and temporal domains. The spatio-temporal characterization of coverage holes and cell outages is a unique feature of this solution that, to the best of the authors' knowledge, has not been achieved so far.
- The proposed solution can detect outages in a realistic heterogeneous deployment of small and mmWave cells generated using a real-network planning software. Moreover, using the key insights derived from performance comparison of EFD-based solution with independent component analysis and k-means clustering based coverage hole and outage detection solutions, selection of an appropriate outage detection solution based on radio propagation environment conditions can be done.

The rest of the paper is organized as follows: Section II describes the system model, Section III presents the proposed EFD solution, Section IV presents the comparative performance analysis. Practical deployment of proposed solutions is also discussed in Section IV. Finally, Section V presents the conclusions of this study.

II. SYSTEM MODEL

We consider single-cell connectivity for the sake of simplicity. However, the proposed solution can be extended to dual connected systems without any major modifications since the aim is to identify outages and coverage holes, whether they are at control plane or user plane level. We also assume that the user association changes immediately after the outage which is a reasonable assumption for UDHNs with overlapping cell coverage.

A. Network Coverage

For the purpose of this study, we assume that a user is in outage when the downlink received power of that user from its associated cell $P_{r,u}^c$ falls below a threshold $P_{r_{out}}^{th}$ i.e., :

$$\text{Outage} := P_{r,u}^c \leq P_{r_{out}}^{th} \tag{1}$$

and in a coverage hole when $P_{r,u}^c$ falls below a threshold $P_{r_{ch}}^{th}$ i.e., :

$$\text{Coverage Hole} := P_{r,u}^c \leq P_{r_{ch}}^{th} \tag{2}$$

where $P_{r_{ch}}^{th} > P_{r_{out}}^{th}$. The rest of the discussion in this and subsequent section will focus on outage detection but it lends itself directly to coverage hole detection as well without any changes.

To calculate $P_{r,u}^c$, we use the standard exponential pathloss model, the log of which can be written as:

$$P_{r,u_{dBm}}^c = f(P_t^c, G_u, G_u^c, b, d_u^c, \beta) + \epsilon_u^c \tag{3}$$

where P_t^c is the transmit power of cell c, G_u is the gain of user equipment, G_u^c is the channel gain of cell c, b is the pathloss constant and depends on the clutter, ϵ_u^c is the shadowing at the location of user u from cell c and usually assumed to be log-normally distributed, d_u^c is the distance of subscriber u from cell c, and β is the pathloss exponent. Assuming that for a particular user u in cell c, G_u, G_u^c, b, d_u^c and β remain constant, we can re-write (3) as:

$$P_{r,u_{dBm}}^c = f(P_t^c) + \epsilon \tag{4}$$

Note that these parameters are assumed constant for each distinct user only. These parameters vary for different users according to the location of users and the propagation environment. Moreover, in our simulations, the function in (3) captures the propagation environment features through an advanced ray-tracing based realistic simulator that has realistic path loss, shadowing and geographical information. The system model is further described in detail in Section IV. Each subset $\hat{P}_t$ of the set of cell transmit powers P_t will result in a different set of received powers P_r. As such, we can define the likelihood of receiving a set of downlink received powers P_r given some set of transmit powers $\hat{P}_t$ as:

$$p(P_r) = \int p(P_r|\hat{P}_t)p(\hat{P}_t)d\hat{P}_t \tag{5}$$

In the event of a cell outage, the loss of transmission from the affected cell will result in a unique set of downlink received powers $\boldsymbol{P_{r_{out}}}$. Given this set of received powers, we can find the set of transmit powers including the affected cell transmit power using Baye's rule as:

$$p(\hat{\boldsymbol{P_t}}|\boldsymbol{P_{r_{out}}}) = \frac{p(\boldsymbol{P_{r_{out}}}, \hat{\boldsymbol{P_t}})}{p(\boldsymbol{P_{r_{out}}})} \qquad (6)$$

III. ENTROPY FIELD DECOMPOSITION

For a deterministic system with a fixed signal propagation model and no shadowing, the estimation of (6) is simply a question of going through all the subsets $\hat{\boldsymbol{P_t}}$ and calculating the resulting sets $\boldsymbol{P_r}$. However, for a system with random variations in the signal, this estimation becomes more complex. Furthermore, if these random variations affect the system both spatially and temporally, as is the case in real mobile cellular networks, obtaining the conditional probabilities in (6) becomes intractable. However, one method of obtaining an estimate of these probabilities which has been explored in [51] is to use information field theory which represents the probability distributions in terms of an information field. In our case, we must first represent the transmit power or signal data as an information field such that:

$$P_t(x_l, y_l, t_l) \equiv P_t(\zeta_l) = \int \vartheta\delta(\zeta - \zeta_l)d\vartheta \qquad (7)$$

where $\zeta_l = x_l, y_l, t_l$ represents the transformation of spatial coordinates $x_i, y_i, i = 1, ..., NM$ and temporal coordinate $t_j, j = 1, ..., O$ as a point on the information field ϑ. The key descriptor of an information field is the Hamiltonian $\mathcal{H}$ which corresponds to the total energy of the field [53] and is defined as:

$$\mathcal{H}(\boldsymbol{P_r}, \vartheta) = -\ln p(\boldsymbol{P_r}, \vartheta) \qquad (8)$$

Using the above transformations, (6) can be re-written as:

$$p(\vartheta|\boldsymbol{P_r}) = \frac{e^{\mathcal{H}(\boldsymbol{P_r}, \vartheta)}}{\mathcal{Z}(\boldsymbol{P_r})} \qquad (9)$$

where $\mathcal{Z}(\boldsymbol{P_r}) = \int e^{\mathcal{H}(\boldsymbol{P_r}, \vartheta)}d\vartheta$ is called the partition function. Since the spatio-temporal received powers in a real network are not independent of each other, we consider $P_t(\zeta_l)$ as an interacting field [51] whose Hamiltonian can be derived through Taylor series expansion of (8) as given in [48]:

$$\mathcal{H}(\boldsymbol{P_r}, \vartheta) = \mathcal{H}_0 + \frac{1}{2}\vartheta^\dagger \mathcal{D}^{-1}\vartheta - \boldsymbol{j}^\dagger\vartheta +$$

$$\sum_{n=1}^{\infty}\frac{1}{n!}\int ... \int V_{\zeta_1...\zeta_n}^{(n)}\vartheta(\zeta_1)...\vartheta(\zeta_n)d\zeta_1...d\zeta_n \qquad (10)$$

where $\mathcal{D}$ matrix is the information propagator, the vector $\boldsymbol{j}$ is the information source, the $(.)^\dagger$ notation represents the adjoint of a matrix, and $\mathcal{H}_0$ is the free energy Hamiltonian [53] which can be obtained by integrating the joint probability $p(\boldsymbol{P_r}, \vartheta)$ over $\boldsymbol{P_r}$ and ϑ. Since $\mathcal{H}_0$ is a consequence of an interaction-less field, it simply acts as a scaling factor for an interacting field. Also, since we assume that the received power at each point is not independent of other points due to shadowing and fading effects, we can safely ignore $\mathcal{H}_0$ for this study. The terms $V_{\zeta_1...\zeta_n}^{(n)}$ represent the interactions of up to n field components and are integrated over each coordinate. The matrix $\mathcal{D}$ and vector $\boldsymbol{j}$ can be obtained by using the free

theory formalism for a Gaussian signal [51] and are given as:

$$\mathcal{D} = \left[\boldsymbol{\sigma_{\hat{P}_t}^2}^{-1} + f(\hat{\boldsymbol{P_t}})^\dagger\mathcal{N}^{-1}f(\hat{\boldsymbol{P_t}})\right]^{-1} \qquad (11a)$$

$$\boldsymbol{j} = f(\hat{\boldsymbol{P_t}})^\dagger\mathcal{N}^{-1}\boldsymbol{P_r} \qquad (11b)$$

where $\boldsymbol{\sigma_{\hat{P}_t}^2} = \langle\hat{\boldsymbol{P_t}}\hat{\boldsymbol{P_t}}^\dagger\rangle$ is the covariance of the transmit powers and $\mathcal{N}$ is the covariance of noise in the data.

The interaction terms $V_{\zeta_1...\zeta_n}^{(n)}$ can be obtained using entropy spectrum pathways theory which ranks the optimal pathways within a disordered lattice according to their path entropy [52]. To construct the entropy pathways, we must obtain a coupling matrix $\mathcal{Q}$ of points on the information field lattice. This can be done by generating an adjacency matrix $\mathcal{A}_{ij}$ of spatio-temporal points in the dataset $\boldsymbol{P_r}$ and using the transformation $\zeta_l = x_l, y_l, t_l$ to obtain the components of $\mathcal{Q}$ matrix as follows:

$$\mathcal{Q}(\zeta_i, \zeta_j) = P_r(i)P_r(j)\mathcal{A}_{ij} \qquad (12)$$

It is important to highlight here that the $\mathcal{Q}$ matrix can be used to represent any relationship between two or more points in the network regardless of the signal propagation model and the distributions of shadowing and fading. This is a major advantage compared to other techniques such as Bayesian classification that rely on some underlying assumptions regarding data and noise distributions which can lead to very high misclassification error if the actual distribution differs from the assumed one.

The information field can be reconstructed via entropy spectrum pathways that allow the representation of the field in terms of the eigenmodes of $\mathcal{Q}$ using Fourier expansion. In mathematical terms, field components are given as:

$$\vartheta(\zeta_l) = \sum_{k}^{K}[a_k\varphi^{(k)}\zeta_l + a_k^*\varphi^{*(k)}\zeta_l] \qquad (13)$$

where $\varphi^{(k)}$ is the kth eigenmode, a_k is the mode amplitude of kth eigenmode and the $*$ operator refers to the conjugate of a number, while K is the number of significant eigenmodes considered for field transformation. A key insight here is that by only considering the most important eigenmodes and keeping K to a reasonably small value compared to the total number of eigenmodes, we can obtain a decent estimate of the information field, thus reducing the problem complexity significantly. To test the importance of an eigenmode, we can compare the corresponding eigenvalue λ_k with the determinant of the noise covariance matrix $\mathcal{N}$.

Using the above information, we can now obtain the transformed information Hamiltonian $\mathcal{H}(\boldsymbol{P_r}, \boldsymbol{a_k})$ by:

$$\mathcal{H}(\boldsymbol{P_r}, \boldsymbol{a_k}) = \frac{1}{2}\boldsymbol{a_k}^\dagger\boldsymbol{\Lambda}\boldsymbol{a_k} - \boldsymbol{j_k}^\dagger\boldsymbol{a_k} +$$

$$\sum_{n=1}^{\infty}\frac{1}{n!}\sum_{k_1}^{K}...\sum_{k_n}^{K}\tilde{V}_{k_1...k_n}^{(n)}a_{k_1}...a_{k_n} \qquad (14)$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathcal{Q}$, $\tilde{V}_{k_1...k_n}^{(n)}$ represent the interaction terms of the eigenmodes, and $\boldsymbol{j_k}$ is the amplitude of the kth eigenmode in expansion of

the information source j:

$$j_k = \int j\varphi^{(k)}d\zeta \qquad (15)$$

To calculate the values of mode amplitudes, the Hamiltonian is minimized with respect to the field and the field is replaced with its transformation in terms of its eigenmodes which gives [48]:

$$\Lambda a_k = j_k - \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k_1}^{K} ... \sum_{k_n}^{K} \tilde{V}_{kk_1...k_n}^{(n)} a_{k_1}...a_{k_n} \qquad (16)$$

If the field interaction terms $V_{\zeta_1...\zeta_n}^{(n)}$ are defined as powers of the coupling matrix $\mathcal{Q}$ such that:

$$V_{\zeta_1...\zeta_n}^{(n)} = \frac{\alpha^{(n)}}{n} \sum_{p=1}^{n} \prod_{\substack{m=1 \\ m \neq n}}^{n} \mathcal{Q}_{\zeta_p}\mathcal{Q}_{\zeta_m} \qquad (17)$$

then the mode interaction terms $\tilde{V}_{k_1...k_n}^{(n)}$ are obtained by:

$$\tilde{V}_{k_1...k_n}^{(n)} = \frac{\alpha^{(n)}}{n} \sum_{p=1}^{n} \left(\frac{1}{\lambda_{k_p}} \prod_{m=1}^{n} \lambda_{k_m} \right) \int \left(\prod_{r=1}^{n} \varphi^{(k_r)}(\zeta) \right) d\zeta \qquad (18)$$

where coefficients $\alpha^{(n)}$ should be chosen sufficiently small to ensure the convergence of (18). Using the formulation presented above, we can obtain the posterior probability of the field as defined in (9). The decomposition of the information field into its modes provides a reflection of the field into its independent modes.

A. Outage Detection Using Entropy Field Decomposition

The result of applying EFD to the coverage data is an entropy field which identifies how information flows across the spatio-temporal domains. In order to make this resulting field output useful for outage detection in MCNs, algorithm 1 presents our proposed outage detection and localization solution.

The algorithm takes user-cell association $\mathbb{U}_c$ information from the MDT data and the network coverage data when no outage is present in the network $P_{r_{norm}}$, as well as real-time spatio-temporal coverage data $P_{r_{RT}}$. The real-time data is processed using EFD which outputs the data points where high information flows are detected. In simple terms, the EFD algorithm gives the boundary between outage and non-outage effected areas. The points with high energy i.e., the points at the boundary of the outage are passed to a localization module which identifies the cell with which the high energy points are associated. Once the degraded cell is identified, it is passed as an output which can be fed to an outage diagnosis and compensation algorithm.

B. Complexity Analysis of EFD-based Outage and Coverage Hole Detection

The complexity of algorithm 1 depends on two key factors: 1) the size of input matrices $P_{r_{norm}}$ and $P_{r_{RT}}$ and 2) the duration of time over which outage detection is being done. These factor define the size of adjacency matrix $\mathcal{A}$ which in

---

Algorithm 1 Outage Detection Using EFD

Require: $P_t, P_{r_{norm}}, P_{r_{RT}}, \mathbb{U}_c$
Ensure: $\vartheta(\zeta_l), c$ in outage

1: Calculate $\mathcal{Q}$ from (12) using $P_{r_{RT}}$
2: Obtain eigenvalues and eigenvectors of $\mathcal{Q}$
3: Calculate entropy field ϑ using the eigenvalues and eigenvectors of $\mathcal{Q}$ from (13)
4: **for** $l \in 1, ..., NMO$ **do**
5:      **if** $\vartheta_l > 0$ **then**
6:          $\{u(x_{out}, y_{out})\} = \{u(x_{out}, y_{out})\} + u(x_l, y_l)$
7:      **end if**
8: **end for**
9: **if** $\{u(x_{out}, y_{out})\} = \phi$ **then**
10:      continue
11: **else**
12:      **for** $u(x_{out}, y_{out}) \in \{u(x_{out}, y_{out})\}$ **do** $\{c_{out}\} = \{c_{out}\} + \{c = \arg\max_{\forall c \in \mathbb{C}} P_{r,u(x_{out},y_{out})norm}^c\}$
13:      **end for**
14: **end if**

---

turn defines the computation required to estimate the matrix $\mathcal{Q}$. Let $N = M$ and $L = N^2 * O$, then $\mathcal{A}$ and $\mathcal{Q}$ are LxL matrices. Calculating $\mathcal{A}$ is then $\mathcal{O}(L^2)$ complex. $\mathcal{Q}$ matrix calculation only requires us to replace estimate $\mathcal{Q}_{ij}$ values at points where $\mathcal{A}_{ij} \neq 0$ which is $\mathcal{O}(log(L))$ complex. The eigenvalue decomposition of $\mathcal{Q}$ is $\mathcal{O}(L^3)$ complex [54]. The reconstruction of $\vartheta(\zeta_l)$ is $\mathcal{O}(K)$ complex based on (13). Finally, identifying points with non-zero entropy is a simple search action which is also $\mathcal{O}(log(L))$ complex. Thus, the complete algorithm is $\mathcal{O}(L^3 + L^2 + 2 * log(L))$ complex. Fig. 1 shows the complexity of the EFD-based outage and coverage hole detection algorithm when $N = M$ and $O = 2$ which confirms the estimated complexity.
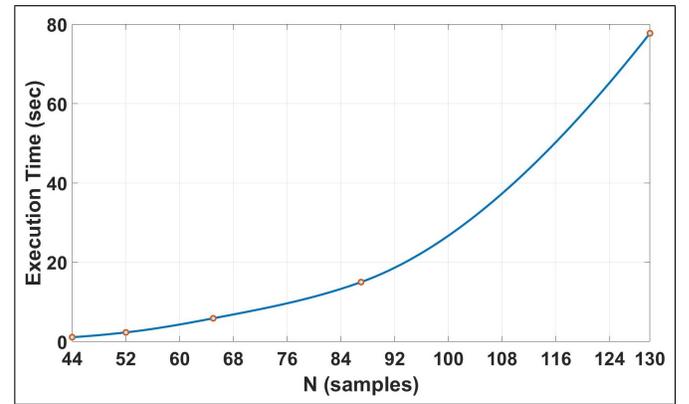
Fig. 1: Complexity of EFD-based outage and coverage hole detection algorithm

C. Other techniques for outage detection

To evaluate the performance of EFD-based outage detection, we compare it with two other techniques namely 1) independent component analysis (ICA) [55] and 2) k-means clustering [39]. Both these techniques are briefly described here before the discussion on simulation details and results.

1) Independent Component Analysis: Just like EFD, ICA is an unsupervised clustering technique and is highly useful in real datasets where the noise distribution in the source signals may not always be Gaussian. ICA finds clusters in a dataset in the form of source signals which can be used to reconstruct the received signals such that:

$$P_r = A * P_t \tag{19}$$

where P_r is the received signal vector, P_t are the source signal vectors and A is the weight matrix for signal reconstruction.

There are several methods of estimating P_t and A including projection pursuit, infomax, maximum likelihood estimation [56], and reconstruction ICA [57]. For the purpose of this study, we have employed reconstruction ICA.

2) k-means Clustering: k-means clustering is another commonly used unsupervised clustering technique. The k-means clustering algorithm splits the data into k separate clusters based on their distance from randomly distributed means. The most common implementation of k-means clustering is also referred to as Lloyd's algorithm [58] which iteratively updates the values of the k different means based on the distance between the means and data points. The algorithm begins by assigning each data point P_{r_p} to a set of points $S_i^{(t)}$ which are closest to the mean μ_i in terms of Euclidean distance such that:

$$S_i^{(t)} = \{P_{r_p} : ||P_{r_p} - \mu_i^{(t)}||^2 \leq ||P_{r_p} - \mu_j^{(t)}||^2 \forall j, 1 \leq j \leq k\} \tag{20}$$

The means are updated based on the new sets or clusters such that:

$$\mu_i^{(t+1)} = \frac{1}{S_i^{(t)}} \sum_{P_{r_p} \in S_i^{(t)}} P_{r_p} \tag{21}$$

The algorithm continues to update the means until it converges or some stopping criteria for distance is met.

IV. SIMULATION AND RESULTS

In this section, we analyze the performance of the proposed EFD based solution for detection of coverage holes and network outages in spatial and temporal dimensions. We present results for different outage scenarios, as well as different levels of shadowing.

A. Simulation Setup

To obtain coverage data, we simulate an ultra-dense network of mmWave cells and small cells which will be typical in a 260m x 260m area of downtown New York City. The network layout of this area showing site deployment is shown in Fig. 2. The coverage data is generated using Atoll Radio Planning Software [59]. The geographical information of the network scenario is shown in Fig. 3. The geographical data comprising of geo raster data and geo vector data was used to depict a realistic scenario. The raster data gives a grid-based representation of the terrain with a defined resolution. The raster files we used are DTM (Digital Terrain Model) representing the elevation of the ground over sea level, clutter classes representing the type of terrain (land cover or land

Fig. 2: Network layout showing sites deployment.

use) and clutter heights (also called a digital height model) representing individual heights (altitude of clutter over the DTM, for example, building heights). Each pixel of a clutter class file contains a code which corresponds to a certain type of ground use or cover. In the clutter height file, a height is given for each point on the map. The geo vector data models the buildings and their height, in the form of one or several ArcView SHP files. A 5m resolution was sufficient to capture the propagation characteristics for our study. A lesser resolution was providing redundant information and a using a greater resolution led to missing information. All of these geo files were incorporated into our model to represent a realistic scenario. Different clutter types have different shadowing effects. Therefore, in our simulations, each clutter type has different standard deviation per frequency representing its shadowing characteristics. The range of shadowing standard deviation for different clutter classes is given in Table 1.

The small cells (red) are configured to operate in 2GHz frequency spectrum with omni-directional antennas, while the mmWave cells (green) are configured to operate in the 28GHz spectrum with directional antennas. Antenna specifications for mmWave cells were obtained from [60]. Among the different ranges of mm-wave antenna gains [61], [62], we chose an antenna gain towards the lower end to show the performance of the proposed algorithm under more stringent conditions. This is because when an outage is simulated by switching off (reducing transmission power of) a cell with high gain antenna, it is easier to detect that outage as compared to the case when an outage is simulated by switching off (reducing transmission power of) a cell with low gain antenna. Note that since the proposed algorithm does not have to rely on details of beamforming, we do not implement a beamforming model. We only model realistic static antenna pattern.

The path loss model was chosen to be aster propagation model [63], rather than empirical or semi-empirical path loss models, that are based on measurements in a specific environment and limited in their ability to capture idiosyncrasies of various propagation environments. In contrast, we employ

(a) Digital height model.

(b) Digital land use map.

(c) Digital terrain model.
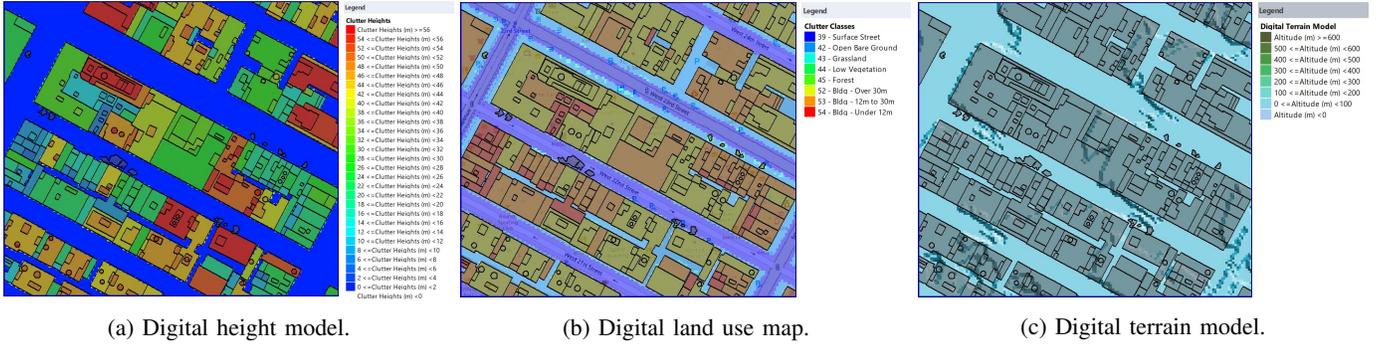
Fig. 3: System model geographical information.

mmWave and standard Aster propagation model which use advanced ray tracing for mmWave and small cell pathloss calculation while taking factors such as clutter information, atmospheric absorption, and frequency-selective fading into account. It also incorporates vertical diffraction over roof-tops, horizontal diffraction/reflection based on ray-launching and ray-tracing calculation on raster data as well as on vector building data. Moreover, it has the support of automatic calibration using continuous wave to further calibrate the model. All these features of aster propagation model enabled the modeling of a realistic network scenario. 5G mm-wave range bands have supported channel bandwidths of 50 MHz, 100 MHz, 200 MHz and 400 MHz [64]. In our simulations, we use 100 MHz bandwidth. We do not consider the impact of varying channel bandwidths as bandwidth will not significantly impact the kind of outages considered in this study that is focused on ultra-dense heterogeneous environment. This is because the effect of channel bandwidth will be significant in a high load scenario only and by having a multi-carrier ultra-dense heterogeneous environment, we ensure that high load scenario will not be encountered. However, the EFD based outage detection algorithm can similarly be applied in different channel bandwidth settings as well. We use cell transmit power as the input signal for information field generation; however, any other coverage actuation parameter such as transmitter antenna tilts or transmitter antenna gains can be used just as easily if the relationship between the parameter and output data is known. To simulate outages we de-activate the transmitters of Small Site 1 and (mmWave) Site 1 shown in Fig. 2. In this work, we are not concerned with the cause of outage or root cause analysis, rather, we focus on detection of outage when it occurs, regardless of its reason of occurrence. Rest of the details of simulation parameters are given in Table I.

Simulation data for each scenario and shadowing level were collected at 12 separate time stamps. The first 6 timestamps correspond to a 'no outage' scenario where the only difference in received power would be due to shadowing. The final 6 timestamps correspond to an outage in either a small cell or a mmWave cell where the impact of both shadowing and outage is evident. To compare the results of the three algorithms in the manuscript, each 'no outage' time stamp at time t was compared with an outage time stamp at time $t + 6$. The resulting 6 comparisons were averaged to present the

TABLE I: Parameter Settings for Simulation

System Parameters	Value
Number of Sites	mmWave: 7; Small: 4
Transmission Frequency	mmWave: 28GHz; Small: 2 GHz
Transmission Bandwidth	mmWave: 100 MHz; Small: 20 MHz
Transmit Power	mmWave: 30 dBm, Small: 20 dBm
Antenna Tilt	mmWave: 7°; Small: 0°
Antenna Gain	mmWave: 18 dBi; Small: 5.7 dBi
Site Placement	Random
Shadowing	Street: 0 dB - 10 dB
	Open Space: 0 dB - 10 dB
	Grassland: 2 dB - 12 dB
	Low Vegetation: 4 dB - 14 dB
	Building $\geq$ 30m: 8 dB - 18 dB
	Building 12m - 30m: 10 dB - 20 dB
	Building ¡ 12m: 12 dB - 22 dB

performance charts in the manuscript. In this way, not only the outage can be detected in spatial dimension, but also the time of occurrence of outage and the duration of the outage can be determined by running the algorithm in real-time over a period of time. More discussion on practical implementation of this solution is presented in Section IV-B-4. The idea is to understand how the algorithm performs at different levels of shadowing to identify outages by comparing pre-outage and post-outage received power levels.

B. Results

The results presented below compare proposed EFD-based outage detection solution with ICA and k-means clustering based outage detection algorithms. Simulations were carried out for a range of shadowing levels to assess the efficacy of EFD in mitigating the effects of noise in the data. Furthermore, we present results for 2 different scenarios: 1) no outage, and 2) outage to investigate the ability of EFD to distinguish between coverage holes and outages. Note that though the algorithms for EFD, ICA and k-means clustering use data from all three spatio-temporal dimensions, the presented results for baseline coverage data are averaged over time. To compare the results of EFD, k-means and ICA, we present two snapshots, one for 0 dB open area shadowing and one for 10 dB open area shadowing. We also compare the true positive detection rate and false positive detection rate for each algorithm with open area shadowing ranging from 0 dB to 10 dB which means that we are benchmarking the performance of each
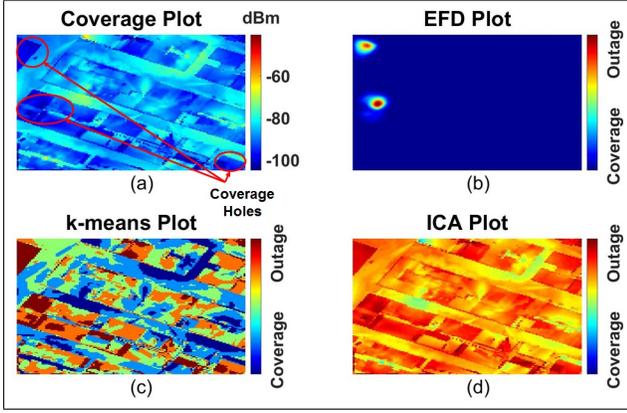
Fig. 4: Coverage hole detection using EFD, k-means clustering and ICA with 0 dB open space shadowing
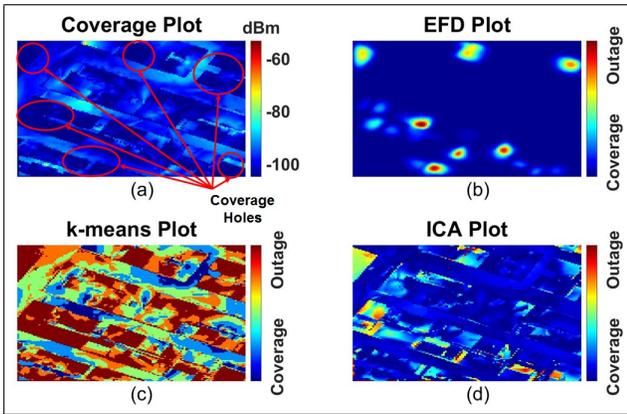
Fig. 5: Coverage hole detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

algorithm in a range of scenarios in a dense urban setting. It is important to highlight here that as we change the open area shadowing, shadowing for other clutter classes such as different building heights is also changed by the same amount. We ran Monte Carlo simulations for the detection algorithms for 100 iterations and the results shown for true positive rate and false positive rate are average over all the simulations.

1) Impact of Shadowing on Coverage Hole Detection: For this study, we have defined coverage holes as points where $P_r < -105$ dBm. Since all three algorithms are unsupervised, this threshold can be set to any value without the loss of generality.

Figs. 4 and 5 show coverage holes in the network and the results for the three coverage hole detection algorithms when open area shadowing is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of reference signal received power (RSRP), Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

While we can visually observe the efficacy of each algorithm from Figs. 4 and 5, a comprehensive evaluation of the three algorithms requires more analytical comparison. To this end, Fig. 6 presents the True Positive detection Rate (TPR) of the three algorithms when open area shadowing in the network
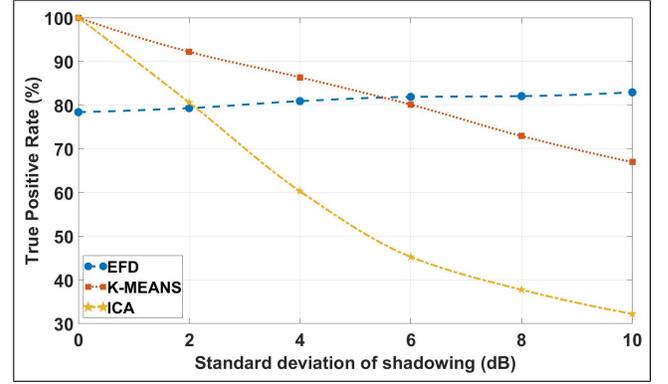
Fig. 6: True positive rate of coverage hole detection over varying shadowing with EFD, k-means clustering and ICA
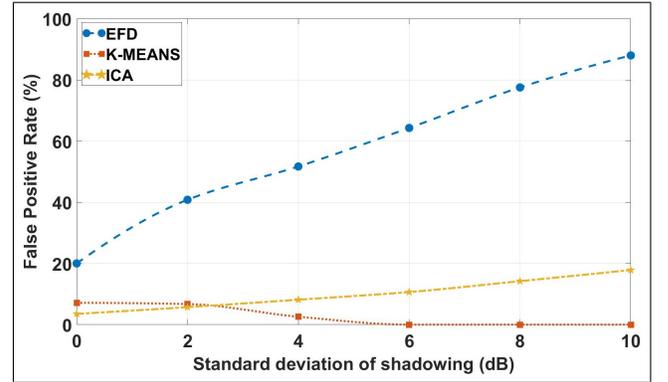
Fig. 7: Flase positive rate of coverage hole detection over varying shadowing with EFD, k-means clustering and ICA

is varied between 0 dB and 10 dB. We can see that k-means clustering performs the best of all three algorithms in terms of TPR for low shadowing levels. However, as the level of shadowing increases, EFD begins to become more dominant as the difference between indoor and outdoor coverage becomes starker. On the other hand, TPR for k-means clustering begins to fall as shadowing increases because the predicted cluster mean values start getting closer to each other resulting in some coverage holes being miss classified.

For a complete perspective, we also compare the False Positive detection Rate (FPR) of the three algorithms which is given in Fig. 7. We can see that EFD starts off with a relatively high FPR which continues to increase with shadowing. This is because EFD identifies the information boundary in the field which means areas near the edge of a coverage hole will also have non-zero entropy leading to false positives. As a result, as the number of coverage holes increases, so does the FPR for EFD. Conversely, the FPR for k-means decreases as shadowing increases. For k-means this is again because of reduced mean separation which leads to fewer false predictions. For ICA, the FPR increases because of its unsupervised nature which means it clusters some areas with $P_r > -105$ dBm with coverage holes. Since the overall P_r values fall dramatically with increase in shadowing standard deviation, the potential for misclassification obviously increases.

2) Impact of shadowing on Outage Detection: To compare the performance of the three algorithms in detecting outages
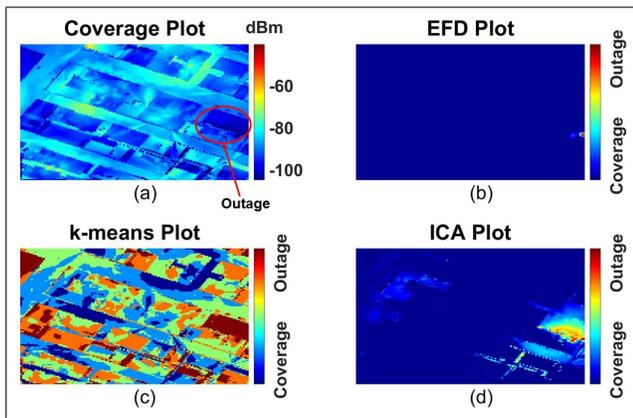
Fig. 8: Small cell outage detection using EFD, k-means clustering and ICA with 0 dB open space shadowing
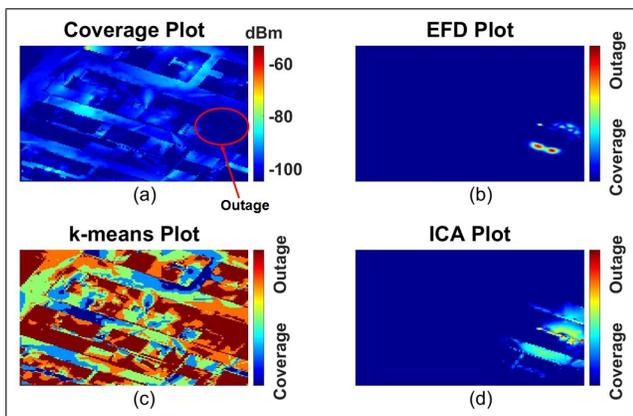
Fig. 9: Small cell outage detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

in the network regardless of network coverage or shadowing, we employ two different scenarios: 1) small cell outage and 2) mmWave cell outage. For this study, we have defined outage as points where user association u_c changes after a cell becomes inactive. Again, since all three algorithms are unsupervised, this outage criteria can be applied without the loss of generality.

a) Small Cell Outage Detection: Figs. 8 and 9 show the small cell outage and the results for the three algorithms when open area shadowing is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of reference signal received power (RSRP), Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

Fig. 8 and 9 present visual comparison of the impact of shadowing on the performance of the three algorithms. From Fig. 8., at 0 dB shadowing standard deviation, we observe that while EFD can detect outage, ICA is also performing well clearly bringing forth the area in outage. However, as shadowing level increases to 10dB standard deviation in Fig. 9, we can observe that EFD algorithm's performance improves significantly, while yielding better localization of outage compared to ICA. For a quantitative analysis of the performance of EFD, ICA and k-means algorithms in the presence of different
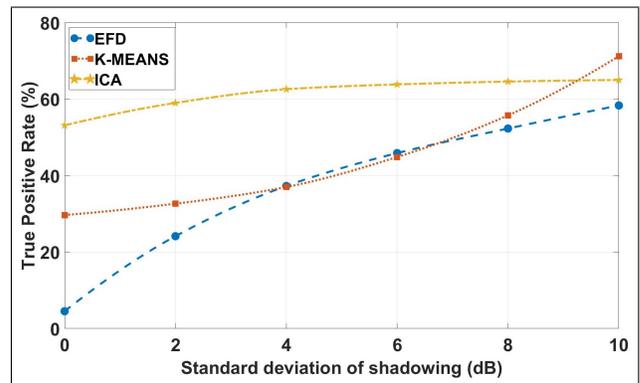
Fig. 10: True positive rate of small cell outage detection over varying shadowing with EFD, k-means clustering and ICA
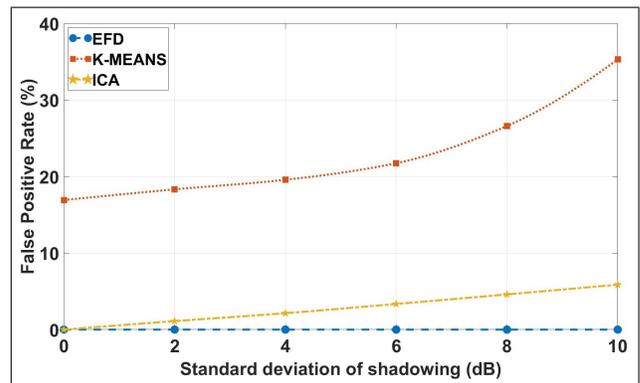
Fig. 11: False positive rate of small cell outage detection over varying shadowing with EFD, k-means clustering and ICA

shadowing levels, we compare the FPR and TPR of the three algorithms with varying levels of shadowing in Fig. 10 and Fig. 11.

From Fig. 10 we can see that ICA performs the best of all three algorithms in terms of TPR for lower levels of shadowing. However, as the level of shadowing increases, k-means begins to become the more dominant algorithm. On the other hand EFD begins with rather low value of TPR but improves as the level of shadowing increases. The increase in TPR for k-means clustering is due to the fact that k-means does not the have the capacity to separate outages from coverage holes resulting in all of the outage points and coverage holes being lumped together as shadowing increases. On the other hand, the increase in TPR for EFD stems from the fact that at low shadowing, the number of points associated with the outage effected small cell are very high but EFD only detects the source of the outage. However, with the increase in shadowing, the number of points associated with the outage affected cell grows smaller which means EFD is able to identify them with higher accuracy.

Conversely, a comparison of FPR from Fig. 11 shows that as the shadowing increases, the ability of k-means to distinguish outages from coverage holes clearly decreases. Same is true for ICA but to a lesser degree since ICA is better at separating source of variation in data than k-means. Compared to both ICA and k-means, EFD has zero FPR for all levels of shadowing due to its ability to extract the source
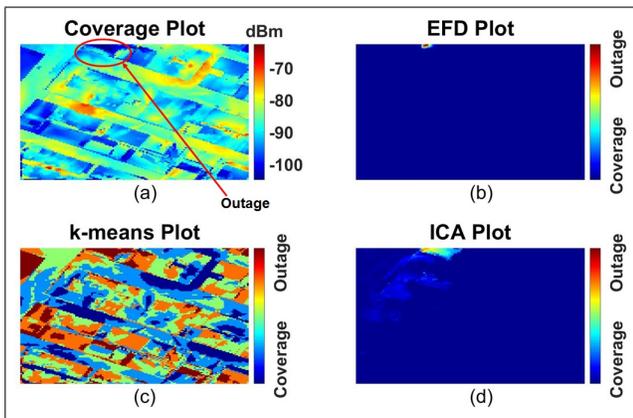
Fig. 12: mmWave cell outage detection using EFD, k-means clustering and ICA with 0 dB open space shadowing
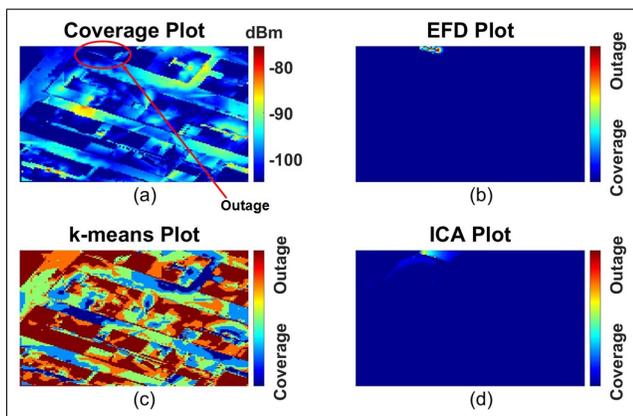
Fig. 13: mmWave cell outage detection using EFD, k-means clustering and ICA with 10 dB open space shadowing

of an anomaly very cleanly from the data regardless of the noise variations.

It is pertinent to note here that the trend for EFD TPR is increasing with increase in shadowing and when taken in context with its FPR, it clearly shows that EFD is the better algorithm for outage detection in high noise environments. Furthermore, since EFD takes both spatial and temporal data as input, not only does it identify the point where received power changes in space, it also detects where the received power changes in time. Therefore, in the event of an outage, it can precisely pinpoint the spot where the outage occurred.

b) mmWave Cell Outage Detection: Figs. 12 and 13 show the mmWave cell outage and the results for the three algorithms when open area shadowing is set to 0 dB and 10 dB respectively. Subfigs. (a) show network coverage in terms of reference signal received power (RSRP), Subfigs. (b) show the results for EFD algorithm, Subfigs. (c) show the results of k-means clustering, while Subfigs. (d) show the results for ICA algorithm.

From Fig. 14 we can see that ICA performs the best of all three algorithms in terms of TPR at low to mid levels of shadowing and its TPR remains stable with increasing shadowing. However, this increase needs to be seen in the context of its FPR given in Fig. 15 which also increases as the level of shadowing increases. This is again because at
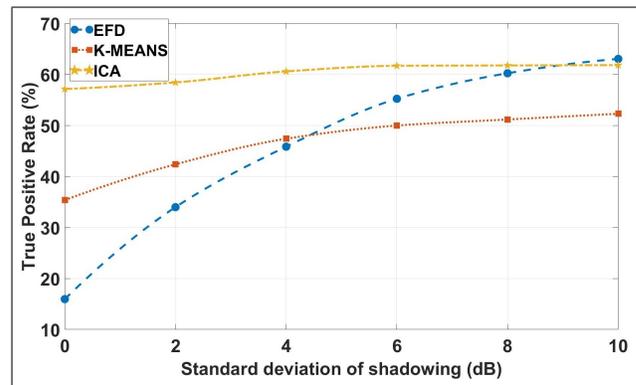
Fig. 14: True positive rate of mmWave cell outage detection over varying shadowing with EFD, k-means clustering and ICA
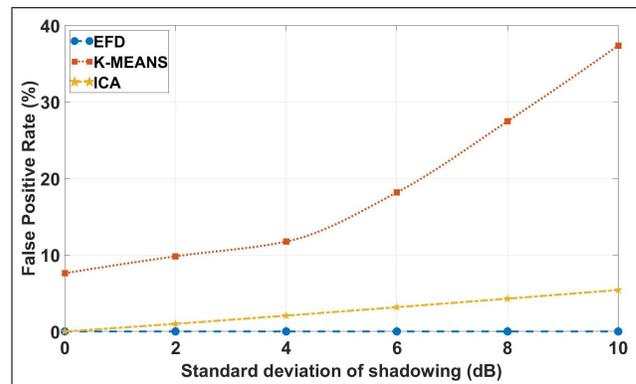
Fig. 15: Flase positive rate of mmWave cell outage detection over varying shadowing with EFD, k-means clustering and ICA

higher levels of shadowing, it becomes more difficult for the algorithm to separate the source of an outage from the source of a deep coverage hole.

In comparison, EFD again starts from a lower TPR and gradually improves becoming the best algorithm at 10dB open space shadowing. The reason for this is the same as before i.e., it is able to extract the source of the outage very well from the data and as shadowing increases the impact of outage becomes smaller compared to coverage holes.

For k-means clustering, the TPR in the case of a mmWave cell outage improves only gradually since there were not a lot of points associated with the cell to begin with. However, as mentioned previously, it still lumps coverage holes and outages together which means its FPR increase dramatically with increase in shadowing.

3) Key Insights from the Results: While the relative performance of EFD, ICA and k-means is obvious from the results presented above, some of the more implicit, yet key insights are given below:

- k-means clustering is effective for coverage hole detection at lower shadowing levels while EFD is more effective at identifying coverage holes at higher levels of shadowing.
- Both EFD and ICA algorithms are more effective at spatially extracting the source of an outage than they are at identifying the impact of the outage making them better

suited for identification of a cell in outage than k-means. In contrast, k-means is more suitable for identifying the impact of an outage.

- EFD is a highly effective choice for locating outages in spatio-temporal data since it can not only identify the source of an outage but also the time at which said outage occurred. This is of great value in real networks where outages can be very costly if allowed to continue for extended periods.

4) Practical Implementation of EFD-based Coverage Hole and Outage Detection Solution: The results presented in this section have served to highlight the power of EFD based solution at overcoming the challenges posed by shadowing and temporal variations towards coverage hole and outage detection solutions. These advantages of EFD make it a very attractive deployment proposition in future mmWave UDHNs.

For implementation in a practical setup, EFD offers several design options and flexibilities including tweaking the value of K to obtain more complex coverage hole and outage profiles. The algorithm is also agnostic to the distribution of noise in the data which also makes it an ideal choice in unpredictable propagation environments such as UDHNs in dense urban areas. Furthermore, the algorithm sensitivity to changes in information flow can also be controlled by the parameter α in (18). A higher value of α makes the algorithm more sensitive to information changes while a smaller value makes it more focused.

V. CONCLUSION

In this work, we have identified several key issues that state-of-the-art coverage hole and outage detection algorithms face i.e., they 1) only consider instantaneous network coverage profile, 2) are sensitive to shadowing, and 3) make assumption regarding data and noise distributions. To overcome these limitations, we have proposed a novel entropy field decomposition based solution which detects outages and coverage holes in spatio-temporal coverage data. We present results comparing the efficacy of entropy field decomposition with state-of-the-art outage detection methods including independent component analysis and k-means clustering in a dense urban mmWave-small cell UDHN over a range of shadowing values. Results show that entropy filed decomposition is a powerful tool in combating the effects of shadowing on coverage hole and outage detection while also demonstrating its efficiency at extracting spatio-temporal information flows from coverage data.

VI. ACKNOWLEDGMENT

REFERENCES

[1] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[2] L. Pierucci, "The quality of experience perspective toward 5G technology," *IEEE Wireless Communications*, vol. 22, no. 4, pp. 10–16, August 2015.

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[4] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, June 2011.

[5] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[6] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, October 2016.

[7] S. Chernov, M. Pechenizkiy, and T. Ristaniemi, "The influence of dataset size on the performance of cell outage detection approach in LTE-A networks," in *Proc. 10th International Conference on Information, Communications and Signal Processing (ICICS)*, Dec 2015, pp. 1–5.

[8] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends in dense urban mmWave cellular networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 3809–3814.

[9] A. Taufique, M. Jaber, A. Imran, Z. Dawy, and E. Yaacoub, "Planning Wireless Cellular Networks of Future: Outlook, Challenges and Opportunities." *IEEE Access*, vol. 5, pp. 4821–4845, 2017.

[10] M. Amirijoo, L. Jorguseski, T. Kurner, R. Litjens, M. Neuland, L. Schmelz, and U. Turke, "Cell outage management in LTE networks," in *IEEE International Symposium on Wireless Communication Systems*, 2009, pp. 600–604.

[11] A. Ericsson, "Ericsson mobility report: On the pulse of the Networked Society," Tech. Rep., 2015.

[12] R. K. Sahoo, M. S. Squillante, A. Sivasubramaniam, and Y. Zhang, "Failure data analysis of a large-scale heterogeneous server environment," in *International Conference on Dependable Systems and Networks, 2004*, 2004, pp. 772–781.

[13] F. Xing and W. Wang, "On the survivability of wireless ad hoc networks with node misbehaviors and failures," *IEEE Transactions on Dependable and Secure Computing*, vol. 7, no. 3, pp. 284–299, 2010.

[14] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of drive tests solution in 3GPP," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 28–36, June 2012.

[15] B. Sayrac, J. Riihijärvi, P. Mähönen, S. Ben Jemaa, E. Moulines, and S. Grimoud, "Improving Coverage Estimation for Cellular Networks with Spatial Bayesian Prediction Based on Measurements," in *Proc. ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, 2012, pp. 43–48.

[16] A. Galindo-Serrano, B. Sayrac, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," in *Proc. 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, May 2013, pp. 35–40.

[17] H. W. Liang, C. H. Ho, L. S. Chen, W. H. Chung, S. Y. Yuan, and S. Y. Kuo, "Coverage hole detection in cellular networks with deterministic propagation model," in *Proc. 2nd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, June 2016, pp. 1–6.

[18] T. Zhang, K. Zhu, and D. Niyato, "A generative adversarial learning-based approach for cell outage detection in self-organizing cellular networks," *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 171–174, 2019.

[19] H. T. Oğuz, A. Kalaycioğlu, and A. Akbulut, "Femtocell Outage Detection in Multi-Tiered Networks using LSTM," in *IEEE 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2019, pp. 1–5.

[20] P.-C. Lin, "Large-Scale and High-Dimensional Cell Outage Detection in 5G Self-Organizing Networks," in *IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2019, pp. 8–12.

[21] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A learning-based approach for autonomous outage detection and coverage optimization," *Transactions on Emerging Telecommunications Technologies*, vol. 27, no. 3, pp. 439–450, 2016.

[22] O. Onireti, A. Zoha, J. Moysen, A. Imran, L. Giupponi, M. A. Imran, and A. Abu-Dayya, "A Cell Outage Management Framework for Dense Heterogeneous Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2097–2113, April 2016.

[23] O. G. Manzanilla-Salazar, F. Malandra, H. Mellah, C. Wetté, and B. Sansò, "A Machine Learning Framework for Sleeping Cell Detection in a Smart-City IoT Telecommunications Infrastructure," *IEEE Access*, vol. 8, pp. 61 213–61 225, 2020.

[24] U. Masood, A. Asghar, A. Imran, and A. N. Mian, "Deep learning based detection of sleeping cells in next generation cellular networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 206–212.

[25] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog, "Minimization of drive tests in 3GPP release 11," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36–43, November 2012.

[26] F. Chernogorov and J. Puttonen, "User satisfaction classification for Minimization of Drive Tests QoS verification," in *Proc. IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2165–2169.

[27] J. Puttonen, J. Turkka, O. Alanen, and J. Kurjenniemi, "Coverage optimization for Minimization of Drive Tests in LTE with extended RLF reporting," in *Proc. IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sept 2010, pp. 1764–1768.

[28] H. N. Qureshi, A. Imran, and A. Abu-Dayya, "Enhanced MDT-Based Performance Estimation for AI Driven Optimization in Future Cellular Networks," *IEEE Access*, vol. 8, pp. 161 406–161 426, 2020.

[29] A. Asghar, H. Farooq, and A. Imran, "Self-Healing in Emerging Cellular Networks: Review, Challenges, and Research Directions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1682–1709, thirdquarter 2018.

[30] A. Zoha, A. Saeed, A. Imran, M. A. Imran, and A. Abu-Dayya, "A SON solution for sleeping cell detection using low-dimensional embedding of MDT measurements," in *Proc. IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, Sept 2014, pp. 1626–1630.

[31] G. Ciocarlie, U. Lindqvist, K. Nitz, S. Nováczki, and H. Sanneck, "On the feasibility of deploying cell anomaly detection in operational cellular networks," in *Proc. IEEE Network Operations and Management Symposium (NOMS)*, May 2014, pp. 1–6.

[32] G. F. Ciocarlie, U. Lindqvist, S. Nováczki, and H. Sanneck, "Detecting anomalies in cellular networks using an ensemble method," in *Proc. 9th International Conference on Network and Service Management (CNSM)*, Oct 2013, pp. 171–174.

[33] F. Chernogorov, J. Turkka, T. Ristaniemi, and A. Averbuch, "Detection of Sleeping Cells in LTE Networks Using Diffusion Maps," in *Proc. IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.

[34] B. Hussain, Q. Du, and P. Ren, "Deep learning-based big data-assisted anomaly detection in cellular networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[35] D. T. Larose, *k-Nearest Neighbor Algorithm*. John Wiley & Sons, Inc., 2005, pp. 90–106.

[36] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-based Local Outliers," *SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, May 2000.

[37] J. Weston and C. Watkins, "Multi-class support vector machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Tech. Rep., 1998.

[38] Y. Ma, J. Zhu, L. Liu, F. Lv, and Y. Wang, "A dynamic affinity propagation clustering algorithm based on MDT in self-healing heterogeneous networks," in *IEEE International Symposium on Communications and Information Technologies (ISCIT)*, 2016, pp. 318–323.

[39] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[40] T.-L. Tien, "A new grey prediction model FGM(1, 1)," *Mathematical and Computer Modelling*, vol. 49, no. 7, pp. 1416 – 1426, 2009.

[41] M. Alias, N. Saxena, and A. Roy, "Efficient cell outage detection in 5G HetNets using hidden Markov model," *IEEE Communications Letters*, vol. 20, no. 3, pp. 562–565, 2016.

[42] M. Mardani and G. B. Giannakis, "Estimating traffic and anomaly maps via network tomography," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1533–1547, 2016.

[43] I. Nevat, D. M. Divakaran, S. G. Nagarajan, P. Zhang, L. Su, L. L. Ko, and V. L. L. Thing, "Anomaly detection and attribution in networks with temporally correlated traffic," *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–14, 2017.

[44] Y. Xie and S. Z. Yu, "A large-scale hidden semi-markov model for anomaly detection on user browsing behaviors," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 54–65, 2009.

[45] I. C. Paschalidis and G. Smaragdakis, "Spatio-temporal network anomaly detection by assessing deviations of empirical measures," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 685–697, 2009.

[46] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "Characterizing and optimizing cellular network performance during crowded events," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1308–1321, 2016.

[47] H. Farooq, M. S. Parwez, and A. Imran, "Continuous Time Markov Chain Based Reliability Analysis for Future Cellular Networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[48] L. R. Frank and V. L. Galinsky, "Detecting spatio-temporal modes in multivariate data by entropy field decomposition," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 39, p. 395001, 2016.

[49] ——, "Dynamic Multiscale Modes of Resting State Brain Activity Detected by Entropy Field Decomposition," *Neural Computation*, vol. 28, no. 9, pp. 1769–1811, 2016, pMID: 27391678.

[50] L. R. Frank, V. L. Galinsky, L. Orf, and J. Wurman, "Dynamic multiscale modes of severe storm structure detected in mobile doppler radar data by entropy field decomposition," *J. Atmos. Sci.*, vol. 75, no. 3, pp. 709–730, Mar. 2018. [Online]. Available: https://doi.org/10.1175/JAS-D-17-0117.1

[51] T. A. Enßlin, M. Frommert, and F. S. Kitaura, "Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis," *Phys. Rev. D*, vol. 80, Nov 2009.

[52] L. R. Frank and V. L. Galinsky, "Information pathways in a disordered lattice," *Phys. Rev. E*, vol. 89, no. 3, p. 032142, Mar. 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.89.032142

[53] W. R. Hamilton, "On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function," *Philosophical transactions of the Royal Society of London*, vol. 124, pp. 247–308, 1834.

[54] V. Y. Pan and Z. Q. Chen, "The complexity of the matrix eigenproblem," in *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, ser. STOC '99, New York, NY, USA, 1999, pp. 507–516.

[55] T.-W. Lee, *Independent Component Analysis*. Springer US, 1998, pp. 27–66.

[56] J. V. Stone, *Independent component analysis: a tutorial introduction*. MIT press, 2004.

[57] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[58] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Marc.

[59] Forsk, "Atoll Overview," [Online], Available: http://www.forsk.com/atoll-overview. [Accessed: 12-Oct-2018].

[60] T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock, *Millimeter wave wireless communications*. Pearson Education, 2015.

[61] Taoglas, "KSF 410.A Datasheet ," [Online], Available: https://www.everythingrf.com/product-datasheet/741-318-ksf410-a. [Accessed: 24-Dec-2020].

[62] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[63] Forsk, "Aster Propagation Model," [Online], Available: http://www.forsk.com/aster-propagation-model. [Accessed: 12-Oct-2018].

[64] R. Dilli, "Analysis of 5G Wireless Systems in FR1 and FR2 Frequency Bands," in *IEEE International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2020, pp. 767–772.

Dr. Ahmad Asghar (S'17) received his B.Sc. degree in Electronics Engineering from Ghulam Ishaq Khan Institute of Science and Technology, Pakistan, in 2010 and the M.Sc. degree in Electrical Engineering from Lahore University of Management and Technology, Pakistan in 2014. He has a Ph.D. in Electrical and Computer Engineering from the University of Oklahoma, USA where he contributed to multiple NSF funded studies on 5th Generation Cellular Networks. His research work includes studies on Self-Healing and Self-Coordination of Self-Organizing Functions in Future Big-Data Empowered Cellular Networks using analytical and machine learning tools. He currently works at Microsoft Corporation as a Data&Applied Research Scientist.

Dr. Adnan Abu-Dayya (SM'14) Adnan Abu-Dayya received the Ph.D. degree in digital mobile communications (electrical engineering) from Queen's University, Kingston, ON, Canada, in 1992.,He worked with AT&T Wireless, Seattle, USA, for ten years, where he served in a number of senior management positions covering product innovations, emerging technologies, systems engineering, product realization, and intellectual property management. He is the Executive Director of Qatar Mobility Innovations Center, Doha, Qatar. He led the establishment of the Qatar Mobility Innovations Center in 2009. It is one of the first independent innovations institution in the Middle East focused on translating Research and Development and technology innovations into scalable digital platforms and solutions in the field of intelligent mobility and smart cities. He worked as a Senior Manager with the Advanced Technology Group, Nortel Networks, Canada, and as a Senior Consultant with the Communications Research Centre, Ottawa, ON, Canada. He has ten issued patents, and about hundred refereed publications. Dr. Abu-Dayya serves as the Chairman of the Advisory Board of the Electrical and Computer Engineering Department, Texas A&M University at Qatar, where he is a member of the Steering Committee of the Smart Grid Research Center

Haneya Naeem Qureshi (S'16) received her BS degree in Electrical Engineering from Lahore University of Management Sciences (LUMS), Pakistan, in 2016 and M.S. degree in Electrical and Computer Engineering from the University of Oklahoma, USA in 2017. She is currently a Ph.D. candidate in Electrical and Computer Engineering with the University of Oklahoma, USA working in the Artificial Intelligence (AI) for Networks Laboratory, where she is contributing to several NSF-funded projects. She is also an ORISE fellow working with the Center for Devices and Radiological Health, U.S Food and Drug Administration (FDA), Maryland, where she is evaluating the use of the 5th generation of mobile communication networks (5G) in medical devices. Her other current research interests include network automation and combination of machine learning and analytics for future cellular systems. She has also been engaged in system design of Unmanned Aerial Vehicles deployment, channel estimation and pilot contamination problem in Massive MIMO TDD systems.

Dr. Ali Imran (M'15, SM'17) is a Presidential Associate Professor in ECE and founding director of the Artificial Intelligence (AI) for Networks (AI4Networks) Research Center and TurboRAN Testbed for 5G and Beyond at the University of Oklahoma. His research interests include AI and its applications in wireless networks and healthcare. His work on these topics has resulted in several patents and over 100 peer-reviewed articles including some of the most influential papers in domain of wireless network automation. On these topics he has led numerous multinational projects, given invited talks/keynotes and tutorials at international forums and advised major public and private stakeholders and co-founded multiple start-ups. He holds a B.Sc. degree in electrical engineering from the University of Engineering and Technology Lahore, Pakistan, in 2005, and the M.Sc. degree (Hons.) in mobile and satellite communications and the Ph.D. degree from the University of Surrey, Guildford, U.K., in 2007 and 2011, respectively. He is an Associate Fellow of the Higher Education Academy, U.K. He is also a member of the Advisory Board to the Special Technical Community on Big Data, the IEEE Computer Society.

Dr. Hasan Farooq (S'14) received his B.Sc. degree in Electrical Engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2009 and the M.Sc. by Research degree in Information Technology from Universiti Teknologi PETRONAS, Malaysia in 2014 wherein his research focused on developing adhoc routing protocols for smart grids. He has a Ph.D. degree in Electrical and Computer Engineering from the University of Oklahoma, USA. His research area is Big Data empowered Proactive Self-Organizing Cellular Networks focusing on Intelligent Proactive Self-Optimization and Self-Healing in HetNets utilizing dexterous combination of machine learning tools, classical optimization techniques, stochastic analysis and data analytics. He has been involved in a multinational QSON project on Self Organizing Cellular Networks (SON) and is currently contributing to two NSF funded projects on 5G SON. He is the recipient of Internet Society (ISOC) First Time Fellowship Award towards Internet Engineering Task Force (IETF) 86th Meeting held in USA, 2013. He currently works at Ericsson as an AI&ML Research Scientist.