Communication-Free Inter-Operator Interference Management in Shared Spectrum Small Cell Networks

Communication-Free Inter-Operator Interference Management in Shared Spectrum Small Cell Networks

Cengis Hasan
Department of ECE, Stevens Institute of Technology, USA
dhasan1@stevens.edu

Mahesh K. Marina
School of Informatics, The University of Edinburgh, UK
mahesh@ed.ac.uk

***Abstract*—Emergence of shared spectrum such as CBRS 3.5 GHz band promises to broaden the mobile operator ecosystem and lead to proliferation of small cell deployments. We consider the inter-operator interference problem that arises when multiple small cell networks access the shared spectrum. Towards this end, we take a novel communication-free approach that seeks implicit coordination between operators without explicit communication. The key idea is for each operator to sense the spectrum through its mobiles to be able to model the channel vacancy distribution and extrapolate it for the next epoch. We use reproducing kernel Hilbert space kernel embedding of channel vacancy and predict it by vector-valued regression. This predicted value is then relied on by each operator to perform independent but optimal channel assignment to its base stations taking traffic load into account. Via numerical results, we show that our approach, aided by the above channel vacancy forecasting, adapts the spectrum allocation over time as per the traffic demands and more crucially, yields as good as or better performance than a coordination based approach, even without accounting the overhead of the latter.**

***Index Terms*—Shared spectrum, multi-operator small cell networks, interference prediction and management, machine learning, kernel embedding of distributions**

I. Introduction

Mobile data traffic continues to grow rapidly and scaling the capacity of mobile networks to meet this demand is a key driver for the emerging 5G mobile networks. Making cells smaller and densely deploying them has historically been the biggest contributor to capacity scaling of cellular networks to the extent that they have been named after this concept [1]. However, further increases in cell densification and deployment of small cells are stifled by the current deployment model requiring access to licensed spectrum that is typically possessed by less than a handful of traditional mobile network operators in each country.

Recent regulatory developments in spectrum sharing below 6 GHz[1] that allow sharing of lightly used spectrum held by legacy or public-sector incumbents (e.g., radars and satellite earth stations) via tiered spectrum access models [2], [3] are lowering the barrier for new entrants to the mobile network operator ecosystem by significantly reducing the spectrum acquisition cost. This in turn promises proliferation of small cell deployments. Citizen Broadband Radio Service (CBRS) initiative in the US allowing the shared use of 3.5 GHz band (3500-3700 MHz) via a three-tier access model is a case in point [4]. In the CBRS model, the incumbents make up the top tier. Middle tier users access the spectrum in 10 MHz chunks (up to 70 MHz) using Priority Access Licenses obtained for a medium term (3 years) via auctions. Users in the lowest tier called General Authorized Access (GAA) can access (at least 80 MHz) of spectrum unused by the higher tier users for free. A separate cloud-based management entity called Spectrum Access System (SAS) orchestrated by the regulator ensures that when higher tier users need to use the spectrum they get interference protection from lower tier ones. See [4] for use cases that encourage new entrants to the operator ecosystem leveraging the CBRS style spectrum, which is also expected to rise in amount by an additional 500 MHz with the inclusion of 3.7-4.2 GHz band. Licensed shared access (LSA) model for spectrum sharing that is being promoted for some bands in Europe [5], especially in its dynamic form [6], is another such relevant development. Such shared spectrum use is also now feasible technology-wise with the emergence of MulteFire [7], which brings the high performance and seamless mobile access of LTE to operate solely in unlicensed or shared bands without requiring an anchor in the licensed spectrum.

In this paper, we consider the inter-operator interference management problem[2] that arises when multiple small cell network operators access shared spectrum (e.g., as GAA users in the CBRS 3.5 GHz band). If not carefully managed, interference between multiple operator networks can lead to the "tragedy of the commons" phenomenon, thereby undermining the benefits of shared spectrum, broadened operator ecosystem and dense small cell deployments. However, currently there does not exist a widely accepted approach for operator-level coordination for secondary use of shared spectrum. Even for time-sharing of a given channel, there is no such mechanism like Listen-Before-Talk (LBT) that is mandated as with unlicensed spectrum (e.g., 5 GHz bands used by Wi-Fi networks). While having the cloud-based/centralized SAS mediate access to shared spectrum for interference protection to incumbents and higher tier users is essential, using it to coordinate spectrum sharing among the same tier users (e.g., GAA users in CBRS) as suggested in [8], [9] limits dynamic and fine-grained spectrum use. Another approach would be to have operators exchange detailed spectrum usage information between operators via a coordination protocol (e.g., [10]) for interference management purposes. This latter approach incurs overhead for coordination and may also not be

[1]Most existing mobile/wireless communication systems operate below 6 GHz. Even early deployments of 5G mobile networks are expected to use below 6GHz spectrum.

[2]Note that intra-operator interference management is not an issue as the operator can internally coordinate the spectrum allocation among its entities (base stations and mobiles).

preferable as operators would typically be competitors to each other; moreover, it may be challenging to realize in practice especially if they do not share the underlying infrastructure. We further discuss related work later in Section VII.

Towards this end, *we take a novel communication-free (CF) approach that seeks implicit coordination between operators without explicit communication*. Our approach can be summarized as follows.

- (§III) Operators view the time as a sequence of epochs. In each epoch, every operator through its mobiles senses the spectrum to measure the vacancy of channels in the available spectrum over space. Note that the extent of vacancy of a channel is inversely related to the level of interference on it – low (high) vacancy implies high (low) interference.
- (§IV) The channel vacancy data so obtained from the preceding epochs is used for predicting channel vacancy distribution in the next epoch. Specifically, we use kernel embedding of channel vacancy statistics and learn the evolution of the channel vacancy distribution to the next epoch through an autoregressive (AR) process. This mechanism for predicting channel vacancy distribution constitutes the core contribution of the paper and to our knowledge has not been considered till date in the wireless communications context. Although we focus on the specific context of interference prediction in shared spectrum small cell networks, the proposed mechanism is more generally applicable.
- (§V) The channel vacancy distribution so predicted is used by each operator at the beginning of next epoch for optimally assigning channels to its base stations (BSs) while also taking traffic load into account. This is achieved via the formulation of the optimization problem for communication-free channel assignment for each operator (that leverages the forecasted channel vacancy distribution) and the solution of its Lagrangian relaxation with a sub-gradient descent method.
- (§VI) Via an extensive set of numerical results, we show that our approach forecasts channel vacancy with good accuracy. This in turn is shown to result in spectrum efficiency performance that is as good as or even better than using an ideal inter-operator coordination protocol (i.e., with zero coordination overhead) but also significantly better than channel assignments that do not rely on channel vacancy forecasting.

The next two sections describe the system model and formally state the problem being tackled.

II. System Model

We consider multiple operators, each with its set of small cell BSs deployed in the same environment. For each operator, we assume the presence of a **Central Controller (CC)** for intra-operator coordination purposes including for channel assignments to BSs as well as to interface with external entities responsible for controlling access to shared spectrum (e.g., SAS in the CBRS context, LSA repository). This is illustrated in Figure 1. Note that there is no explicit coordination between different operators in our model. Moreover, the time in our model is a sequence of epochs.

Shared Spectrum Access. For each operator, as mentioned above, the CC is the entity that interacts with the shared spectrum access management system like SAS and enables access to shared spectrum to the operator. In our model, operators take the role similar to GAA users in CBRS model and can in turn share the available shared spectrum amongst them. The total amount of such shared spectrum can vary over time depending on the activity of the higher tier users and each operator is informed of such changes in availability via its CC. We view the available shared spectrum for multi-operator sharing at any given point in time as a set of channels that we refer to as *Shared Access (SA)* channels.

Spectrum Sensing. We assume that mobile users continually collect interference data by sensing its level in every channel. The sensing can be fine-grained on a subcarrier basis. In that case, the mobile divides the channel into subcarriers and determines the subcarrier (channel) level vacancy depending on the level of interference in a subcarrier (and over all subcarriers for a channel) with respect to some energy detection threshold. If it is above the threshold, the mobile sets an indicator variable to 1, otherwise it assigns a value which is lower than 1. Then, it feeds back this data to its operator's CC through the BSs. By utilizing this data, CC forecasts the interference in the next epoch as detailed in Section IV and accordingly performs channel assignments at the beginning of next epoch, and repeats this process (Figure 1). This results in a stochastic optimization problem as it will become apparent shortly. We assume that sensing is performed periodically throughout an epoch. Such sensing could be implemented in real hardware, for example following a very efficient algorithm like the one proposed in [11], or by leveraging the capabilities of mobiles in the LTE context for signal strength measurement and reporting to the associated BS.

Interaction among Operators. Interference on a particular channel is determined by channel assignment strategies of different operators and varying channel conditions, and this changes from epoch to epoch. Such a setting makes it challenging for an operator to detect the actions/strategies of other operators, which are "hidden" in the interference data. Therefore, we treat the interference data as *random* and aim to model/predict it for the subsequent epoch.

III. Formal Description

Operators function in a discrete-time setting. Time is viewed as a sequence of time *epochs*, each with *duration* T. There are n_e *number of epochs per day*. We define all the entities of a particular operator without loss of generality: set of BSs as B, set of mobile users as M, set of all *available SA channels* in time epoch t as $C(t)$. A channel c of *bandwidth* w_c is comprised of n_c *subcarriers*.

We assume that each mobile associates with one of the BSs and denote by M_j *the set of users associated with BS* j. For simplicity, we focus on the downlink case; uplink case can be accommodated similarly. We represent by L_j the *total traffic load* to BS j in terms of **bps**. The path loss model of transmission between BS j and mobile i is

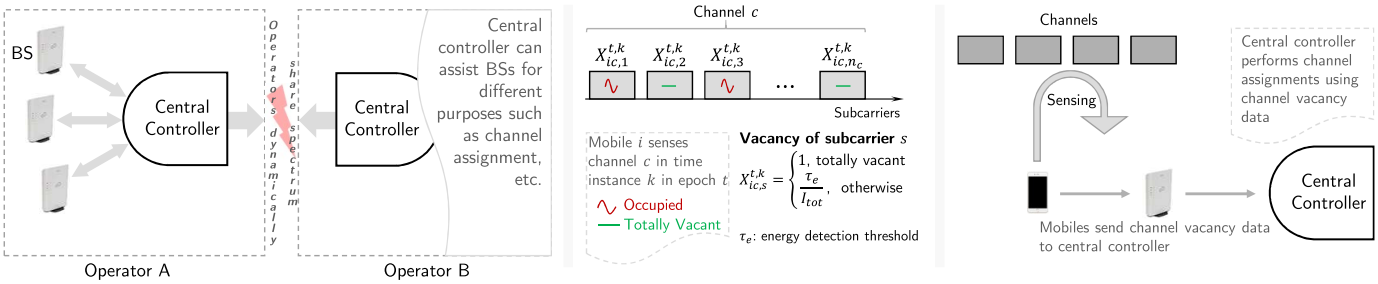


Fig. 1: Illustration of the system model, channel (subcarrier) vacancy measurement at the mobiles and reporting this info to the respective central controllers.

given by $P_{jc,s}g_{jic,s}d_{ji}^{-\alpha}$ where $P_{jc,s}$ and $g_{jic,s}$ are downlink transmission power per subcarrier s and constant path-loss factor (antenna, average channel attenuation) and channel fading coefficient in subcarrier s in channel c, respectively; d_{ji} is distance between j and i, and α is the path loss exponent.

A. Channel Vacancy

Within each epoch t, we assume that mobile i measures a *subcarrier vacancy matrix* given by $\mathbf{X}^t_{ic} = [X^{t,k}_{ic,s}] \in [0,1]^{n_c \times n_d}$ where s denotes a subcarrier in the channel c and k denotes a measurement instance in epoch t:

$$X^{t,k}_{ic,s} = \min\left[1, \frac{\tau_e}{I_{tot}}\right] = \begin{cases} 1, & \frac{\tau_e}{I_{tot}} \geq 1 \\ \frac{\tau_e}{I_{tot}}, & \text{otherwise} \end{cases} \quad (1)$$

meaning that *subcarrier vacancy* $X^{t,k}_{ic,s} \in [0,1]$ shows the ratio of *energy threshold* τ_e and total interference (I_{tot}) in the subcarrier. We set $X^{t,k}_{ic,s}$ to 1 if interference in the subcarrier is below τ_e which can be interpreted to mean that the subcarrier is *vacant*. Consider a particular mobile i and channel c in any one measurement time instance k in epoch t. We define *channel vacancy* in time instance k determined by that mobile as:

$$v^{t,k}_{ic} = \frac{1}{n_c}\sum_{s=1}^{n_c} X^{t,k}_{ic,s} \in [0,1], \quad (2)$$

which is basically the average of subcarrier vacancy over all subcarriers in the corresponding channel. CC uses the channel vacancy data collected from each of its mobiles for modeling/forecasting channel vacancy distribution which in turn is used for channel assignments to the operator's BSs based on the traffic load of their associated mobiles.

B. Spatial Map of Channel Vacancy

Consider a particular mobile i, located at ϕ_i in a given measurement time instance, that is sensing the vacancy level of each subcarrier s in every channel c determined by the effect of surrounding interferers at that time instance (Figure 2). For every measurement instance k in epoch t, CC creates a spatial map of channel vacancy based on data fed back by mobiles. The map guides CC to predict channel vacancy distribution at any location for the upcoming epoch. We model this map at any instance t as a Voronoi tessellation of mobile locations, $\phi^{t,k}_i$ for all $i \in M$, with each region colored based on its

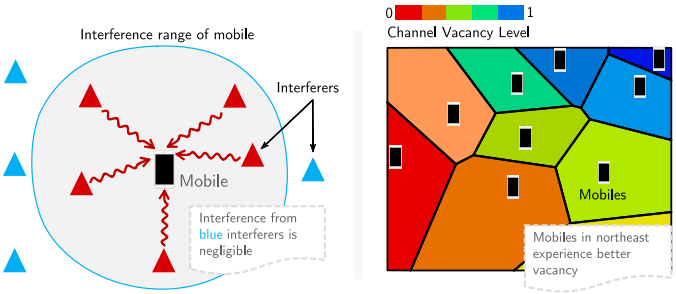


Fig. 2: Left: illustration of interference experienced by a mobile; Right: illustration of spatial channel vacancy level in a particular time instance – a Voronoi region shows a channel vacancy level.

measured (for the past) or predicted (for the future) channel vacancy level. To this end, two types of data are fed back by the mobile: experienced channel vacancy and mobile's location, i.e. $\{v^{t,k}_{ic}, \phi^{k,t}_i\}$. This way, in each epoch t, the CC gathers *channel vacancy vector* $[v^{t,1}_{ic}, v^{t,2}_{ic}, \ldots v^{t,n_d}_{ic}]$, where n_d denotes the number of channel measurement instances in each epoch for each mobile i about channel c, using data fed back by the mobiles. Accuracy of the map depends heavily on the number of mobiles and the distribution of their locations. Note that the map changes dynamically based on mobile movements, channel assignments of operators, etc.

Given the above, in our setting, the CC has to predict the channel vacancy distribution for the upcoming epoch t and for any mobile i using: (i) location of mobile i at the beginning of epoch, i.e., $\phi^{t,1}_i$; and (ii) the channel vacancy vector in the Voronoi regions covering location $\phi^{t,1}_i$ in the preceding Δ *days*. More clearly, assume that CC has access to sequences of data for $t' \in \{t-\Delta n_e, t-\Delta n_e+1, \ldots, t-n_e-1, t-n_e\}$ for all $i \in M$ and $c \in C(t)$. For every epoch t by going Δ days backward, CC builds up matrix $\mathbf{V}^{<t}_{ic} = [V^{<t,1}_{ic}, \ldots, V^{<t,\Delta}_{ic}] \in \mathbb{R}^{\Delta \times \Delta}$, where for $t' = t-(\Delta - z + 1)n_e$:

$$V^{<t,z}_{ic} = \left\{ v^{t',k}_{i'c}, \forall k = 1, \ldots, n_d \middle| \exists i' \in M, \left(\phi^{t,1}_i \in \mathcal{R}^{t',k}_{i'}\right) \right\}, \quad (3)$$

which includes all data in case any mobile i' passed region $\mathcal{R}^{t',k}_{i'}$ in epoch t'.

C. Sensing and Feedback Overhead

Time cost to sense and process vacancy data for a subcarrier is denoted by $\bar{T}_S$. Total time cost of doing it for all channels

TABLE I: Parameter definitions.

Parameter	Definition
$C(t)$	available channels in epoch t
w_c, n_c	bandwidth and number of subcarriers of channel c
L_j	total traffic load to BS j in terms of **bps**
τ_e	energy threshold
$v_{ic}^{t,k}$	vacancy of channel c at mobile i at time instance k in epoch t
$\phi_i^{t,k}$	location of mobile i at time instance k in epoch t
n_d	number of channel measurement instances in each epoch
n_e	number of epochs per day
Δ	number of preceding days whose data is used for prediction
$\mu_{\mathcal{D}}$	kernel mean embedding of channel vacancy distribution $\mathcal{D}$
$\kappa(\cdot,\cdot)$	kernel function
ρ	order of AR model
λ	smoothing parameter in least-squares functional
$\hat{R}_{jic}$	extrapolated raw throughput from BS j to mobile i in channel c

can be given by $\bar{T}_S \sum_{c\in C(t)} n_c$. Besides, there is a time cost of feeding back channel vacancy and location data to the BS which are denoted by $\bar{T}_F^V$ and $\bar{T}_F^L$. Thus, total channel vacancy feedback time cost is given by $\bar{T}_F^V|C(t)|$. Note that both $\bar{T}_F^V$ and $\bar{T}_L^V$ depend heavily on block size (number of bits) of channel vacancy, location data and feedback transmission rate. Location data is not needed to be fed back always unless mobile moves. If mobile does not move during the epoch, then feedback cost will be zero. Otherwise, mobile will send its location data to its associated BS with cost $\bar{T}_L^V$. This overall cost limits n_d and given by:

$$n_d \leq \left\lfloor \frac{T}{\bar{T}_S \sum_{c\in C(t)} n_c + \bar{T}_F^V |C(t)| + \bar{T}_F^L} \right\rfloor. \tag{4}$$

n_d needs to be chosen as high as possible in order to capture the statistical properties of channel vacancy. Current state-of-the-art [11] enables very efficient and fast sensing. An energy detector is very quick, having the maximum delay of around 0.6 ns. A mobile device needs to collect several energy readings per sensed channel, e.g. for 1000 energy readings, it takes 0.05 μs with 20 MHz ADC. Consider a channel with bandwidth 20 MHz and 1200 subcarriers; so the time needed to sense a subcarrier is about $0.05/1200$ $\mu s \approx 4.17$ ns if number of energy readings per sensing is 1000. Errors in sensing and feedback can occur but they fall outside the scope of this paper.

D. Statistics of Channel Vacancy Data

There are several phenomena that change the statistics of channel vacancy data:

- **Channel assignment**: operators may change channel assignments to optimize the service to their respective mobiles.
- **Dynamic channel conditions**: wireless transmission channel may vary within an epoch.
- **Mobility**: user nodes may move.
- **Power control**: downlink transmission power may be adapted again to better serve users.

Probability distribution of channel vacancy contains all these dimensions. However it would be impossible to distinguish all these aspects from channel vacancy data. Therefore, we use a holistic framework that views this data as a source to predict its distribution for the upcoming epoch.

Definition 1 (Channel Vacancy Probability)**.** *For any threshold τ_e, it can be given by*

$$\mathbb{P}[v_{ic} \leq \chi] = \mathbb{P}\left[\frac{1}{n_c}\sum_{s=1}^{n_c} X_{ic,s} \leq \chi\right]. \tag{5}$$

Theorem 1. *(From central limit theorem): For large values of n_c, i.e, when $n_c \to \infty$, probability density of channel vacancy is given by normal distribution:*

$$\mathcal{D}_{ic} \to \mathsf{pdf}_{v_{ic}}(\chi) = \frac{1}{\sqrt{2\pi\sigma_{v_{ic}}^2}} \exp\left(-\frac{(\chi-\bar{v}_{ic})^2}{\sigma_{v_{ic}}^2}\right) \tag{6}$$

where expected value and variance of v_{ic} denoted by $\bar{v}_{ic}$ and $\sigma_{v_{ic}}^2$ are given by

$$\bar{v}_{ic} = \frac{1}{n_c}\sum_{s=1}^{n_c} \bar{X}_{ic,s} \text{ and } \sigma_{v_{ic}}^2 = \frac{1}{n_c^2}\sum_{s=1}^{n_c} \sigma_{X_{ic,s}}^2, \tag{7}$$

respectively.

Proof. See [24]. □

IV. Reproducing Kernel Hilbert Space Embedding of Channel Vacancy Data and AR Models

Channel vacancy levels are sampled i.i.d. from the respective *distributions*, $\mathcal{D}_{ic}^{t-\Delta n_e}, \mathcal{D}_{ic}^{t-(\Delta-1)n_e}, \ldots, \mathcal{D}_{ic}^{t-n_e}$. We aim to construct a distribution $\hat{\mathcal{D}}_{ic}^t$ to be as close as possible to the so far unobserved $\mathcal{D}_{ic}^t$. Then, we calculate extrapolated expected value of channel vacancy $\mathbb{E}[\mathbf{v}_{ic}^t]$ for the upcoming epoch t. We use *vector-valued regression* for learning how the (embedded) distribution of channel vacancy evolves from one epoch to the next epoch. In any epoch t, distribution $\mathcal{D}_{ic}^t$ is determined by vacancy probabilities of subcarriers $\tilde{p}_{ic,s}$. Note that $\tilde{p}_{ic,s}$ is affected by several factors: number of interferers (which is in turn dependent on the channel assignment strategies of operators), power per subcarrier, distance, channel fading coefficient, and other physical phenomena. The expected value and variance of a distribution randomly fluctuates in every epoch. So our aim is to infer any underlying *probabilistic* relation or similarity between epochs. From previous section, we have that channel vacancy follows Gaussian distribution (when n_c is high). We can expect subsequent distribution may also be Gaussian. However, the mean of channel vacancy would be very chaotic due to the non-stationary environment with multiple operators adapting their channel assignments reacting to each others' actions. So, we need a technique which does not require any prior knowledge about distributions.

A *reproducing kernel Hilbert space* (RKHS) $\mathscr{H}$ of functions on $\mathscr{Z}$ with kernel κ where $\mathscr{H}$ is a Hilbert space of functions $\mathscr{Z} \to \mathbb{R}$ with dot product $\langle\cdot,\cdot\rangle_{\mathscr{H}}$, satisfying the reproducing property:

$$\langle f(\cdot), \kappa(v,\cdot)\rangle_{\mathscr{H}} = f(v) \text{ and } \langle \kappa(v,\cdot), \kappa(v',\cdot)\rangle_{\mathscr{H}} = \kappa(v,v'), \tag{8}$$

where $v, v' \in \mathscr{Z}$. The notation $f(\cdot)$ refers to the function itself in the abstract. The linear map from a function f on $\mathscr{Z}$ to its value at v can be seen as an inner product. $\kappa(v,v')$ can be

interpreted as a *non-linear similarity measure* between v and v'. In the following, we introduce and solve three different autoregressive (AR) models related to channel vacancy data and its expected value.

A. AR Model of Kernel Mean Embedding of Channel Vacancy (KME)

For any $\mathcal{D}$, kernel mean embedding of channel vacancy v' is given by the following mappings:

$$\mu_{\mathcal{D}}[v'] := \mathbb{E}_{v\sim\mathcal{D}}[\kappa(v,v')] \quad \text{and}$$
$$\hat{\mu}_{\mathcal{D}}[v'] := \frac{1}{n_d}\sum_{k=1}^{n_d}\kappa(v^k,v') \quad \text{(empirical estimation)}$$

Whenever sufficient condition $\mathbb{E}_v[\kappa(v,v')] < \infty$ is met, then reproducing property imposes $\langle\hat{\mu}_{\mathcal{D}}, f\rangle_{\mathscr{H}} = \frac{1}{n_d}\sum_{k=1}^{n_d} f(v^k)$. We refer to [14], [15] for further reading. Without loss of generality, we consider a particular mobile and channel at the beginning of an epoch t with channel vacancy vector[3] V^z for $z = 1,\dots,\Delta$. We consider that channel vacancy distributions between time epochs can be approximated by an AR process of kernel embedding means, i.e., $\mu_{z+1} = \sum_{r=1}^{\rho}\mathbf{\Lambda}_r\hat{\mu}_{z-r+1} + \epsilon_z$[4] for some operator $\mathbf{\Lambda}_r : \mathscr{H}\to\mathscr{H}$, for all $r = 1,\dots,\rho$ such that ϵ_z for all $z = 1,\dots,\Delta-1$ are independent zero-mean random variables. Any operator $\mathbf{\Lambda}_r \in \mathscr{L}$ is a *linear* operator where $\mathscr{L}$ defines a space of *linear operators* as defined in [16]. Finding which operator is suitable is the main question here, and we fortunately are able to learn it by solving following least-squares functional with smoothing parameter $\lambda > 0$:

$$\min_{\mathbf{\Lambda}_1,\dots,\mathbf{\Lambda}_\rho}\sum_{z=\rho}^{\Delta-1}\left\|\hat{\mu}_{z+1} - \sum_{r=1}^{\rho}\mathbf{\Lambda}_r\hat{\mu}_{z-r+1}\right\|_{\mathscr{H}}^2 + \lambda\sum_{r=1}^{\rho}\|\mathbf{\Lambda}_r\|_{\mathscr{L}}^2. \quad (9)$$

Theorem 2. *For $r = 1,\dots,\rho$, solution of* (9) *is given by*

$$\hat{\mathbf{\Lambda}}_r = \sum_{r'=1}^{\rho}\Upsilon_{r,r'}\tilde{\mathbf{m}}_{r'} = \sum_{z,z'=\rho}^{\Delta-1}Q_{z,z'}\hat{\mu}_{z+1}\sum_{r'=1}^{\rho}\Upsilon_{r,r'}\hat{\mu}^{\top}_{z'-r'+1} \quad (10)$$

Proof. See [24]. □

If we apply learned operators $\hat{\mathbf{\Lambda}}$ to the last observed data from $\Delta-\rho+1$ to Δ, then the result is a prediction of kernel mean embedding in $\Delta+1$ with $\hat{\mu}_{\Delta+1} = \sum_{r=1}^{\rho}\hat{\mathbf{\Lambda}}_r\hat{\mu}_{\Delta-r+1}$ which we expect to approximate unknown $\mu_{\Delta+1}$. Note that $\hat{\mu}_{\Delta+1}$ can be further calculated by a weighted linear combination of observed distributions:

$$\hat{\mu}_{\Delta+1} = \sum_{z,z'=\rho}^{\Delta-1}Q_{z,z'}\hat{\mu}_{z+1}\sum_{r,r'=1}^{\rho}\Upsilon_{r,r'}\underbrace{\hat{\mu}^{\top}_{z'-r'+1}\hat{\mu}_{\Delta-r+1}}_{\langle\hat{\mu}_{z'-r'+1},\hat{\mu}_{\Delta-r+1}\rangle_{\mathscr{H}}}. \quad (11)$$

We can compute expected value of any function $f\in\mathscr{H}$ by using predicted $\hat{\mu}_{\Delta+1}$ with weighted linear combinations of f at channel vacancy data: $\hat{\mathbb{E}}[f(v)] = \langle\hat{\mu}_{\Delta+1}, f\rangle_{\mathscr{H}}$

$$\hat{\mathbb{E}}[f(v)] = \sum_{z,z'=\rho}^{\Delta-1}Q_{z,z'}\langle\hat{\mu}_{z+1},f\rangle_{\mathscr{H}}\sum_{r,r'=1}^{\rho}\Upsilon_{r,r'}\langle\hat{\mu}_{z'-r'+1},\hat{\mu}_{\Delta-r+1}\rangle_{\mathscr{H}}$$
$$= \sum_{z=\rho}^{\Delta-1}\sum_{k=1}^{n_d}\frac{b_z}{n_d}f(v^{k,z+1}) \quad (12)$$

with $b_z = \sum_{z'=\rho}^{\Delta-1}Q_{z,z'}\sum_{r,r'=1}^{\rho}\Upsilon_{r,r'}\langle\hat{\mu}_{z'-r'+1},\hat{\mu}_{\Delta-r+1}\rangle_{\mathscr{H}}$.

B. Kernel Embedding of AR Model of Channel Vacancy (KEC)

In this model, we consider that kernel embedding of channel vacancy which itself follows an AR model. We implement results from [17] Yule-Walker equations as in RKHS where expected values of channel vacancy are mapped from the input space to RKHS using nonlinear transformation. Define nonlinear map $\varphi(\cdot)$ from input space to RKHS $\mathscr{H}$. So, each channel vacancy data v^z is mapped to its corresponding $\varphi(v^z)$. AR process in RKHS $\mathscr{H}$ is given by $\varphi(v^{z+1}) = \sum_{r=1}^{\rho}\ell_r\varphi(v^{z-r+1}) + \varphi(\epsilon_z)$ assuming that $\varphi(\epsilon_z)$ is uncorrelated with $\varphi(v^{z-r+1})$, for all r. By applying Yule-Walker equations, one can find $\hat{\ell}_1,\dots,\hat{\ell}_\rho$, which are calculated by solving following problem: for $r = 1,\dots,\rho$,

$$\mathbb{E}[\tilde{\kappa}(v^z,v^{z-r+1})] = \sum_{r'=1}^{\rho}\ell_{r'}\mathbb{E}[\tilde{\kappa}(v^{z-r'+1},v^{z-r+1})] \quad (13)$$

where $\tilde{\kappa}(\cdot,\cdot)$ is *centered version* of kernel $\kappa(\cdot,\cdot)$ and its expectation is defined with

$$\mathbb{E}[\tilde{\kappa}(v^z,v^{z'})] = \frac{1}{n_d^2\Delta}\Bigg[\sum_{k,k'=1}^{n_d}\kappa(v^{k,z},v^{k',z'}) - \sum_{z''=1}^{\Delta}\sum_{k,k'=1}^{n_d}\kappa(v^{k,z},v^{k',z''})$$
$$-\sum_{z''=1}^{\Delta}\sum_{k,k'=1}^{n_d}\kappa(v^{k,z'},v^{k',z''}) + \frac{1}{\Delta}\sum_{z'',z'''=1}^{\Delta}\sum_{k,k'=1}^{n_d}\kappa(v^{k,z''},v^{k',z'''})\Bigg]. \quad (14)$$

In [17], solution of (13) is shown to be $\hat{\ell}_r = \sum_{r'=1}^{\rho}[W^{-1}]_{r,r'}\mathbb{E}[\tilde{\kappa}(v^{\Delta},v^{\Delta-r})]$, where $[W^{-1}]_{r,r'}\in\mathbf{W}^{-1}$, and $\mathbf{W}\in\mathbb{R}^{\rho\times\rho}$ with entries $W_{r,r'} = \mathbb{E}[\tilde{\kappa}(v^{\Delta},v^{\Delta+r-r'})]$. In this work, we consider only radial kernel which enables to extrapolate the expected value of channel vacancy for epoch t by solving following fixed-point equation: $\sum_{r=1}^{\rho}\hat{\ell}_r\mathbb{E}_v[(v - v^{\Delta-r+1})\kappa(v,v^{\Delta-r+1})] = 0$.

Lemma 1. *Gaussian kernel with variance σ_κ^2 results in fixed-point equation $\sum_{r=1}^{\rho}\hat{\ell}_r\mathcal{J}_r(v^{\Delta-r+1}) = 0$ where $\mathcal{J}_r(v^{\Delta-r+1}) = \mathbb{E}_v[(v - v^{\Delta-r+1})\kappa(v,v^{\Delta-r+1})]$ is given by*

$$\frac{\sigma_\kappa^3(v^{\Delta-r+1}-\bar{v})}{2(\sigma_v^2+\sigma_\kappa^2)^{3/2}}e^{-\frac{(v^{\Delta-r+1}-\bar{v})^2}{2(\sigma_v^2+\sigma_\kappa^2)}}\left(\operatorname{erf}\left(\frac{(v^{\Delta-r+1}-1)\sigma_v}{\sqrt{2\sigma_\kappa^2(\sigma_v^2+\sigma_\kappa^2)}}+\frac{(\bar{v}-1)\sigma_\kappa}{\sqrt{2\sigma_v^2(\sigma_v^2+\sigma_\kappa^2)}}\right)\right.$$
$$\left.-\operatorname{erf}\left(\frac{\sigma_v v^{\Delta-r+1}}{\sqrt{2s_r^2(\sigma_v^2+\sigma_\kappa^2)}}+\frac{\bar{v}\sigma_\kappa}{\sqrt{2\sigma_v^2(\sigma_v^2+\sigma_\kappa^2)}}\right)\right)$$
$$+\frac{\sigma_v\sigma_\kappa^2}{\sqrt{2\pi}(\sigma_v^2+\sigma_\kappa^2)}\left(e^{-\frac{(\bar{v}-1)^2}{2\sigma_v^2}-\frac{(v^{\Delta-r+1}-1)^2}{2\sigma_\kappa^2}} - e^{-\frac{\bar{v}^2}{2\sigma_v^2}-\frac{(v^{\Delta-r+1})^2}{2\sigma_\kappa^2}}\right) \quad (15)$$

[3] Note that V^z is same as $V^{<t,z}$ – we omit $<t$ for brevity.

[4] We omit $\mathcal{D}$ to simplify presentation, i.e. μ_z refers to $\hat{\mu}_{\mathcal{D}^z}$.

Proof. From Theorem 1, we have that channel vacancy has normal distribution with mean $\bar{v}$ and variance σ_v. Then, we have

$$\mathcal{J}_r(v^{\Delta-r+1}) = \int_0^1 \frac{v-v^{\Delta-r+1}}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-v^{\Delta-r+1})^2}{2\sigma_\kappa^2}\right) \exp\left(-\frac{(v-\bar{v})^2}{2\sigma_v^2}\right) dv$$

of which solution can be shown to be eq. (15). □

Empirical Estimation: fixed-point equation can have the following form:

$$\hat{v}^{k,t} = \frac{\sum_{r=1}^{\rho} \hat{\ell}_r \kappa(\hat{v}^{k,t}, v^{k,\Delta-r+1}) v^{k,\Delta-r+1}}{\sum_{r=1}^{\rho} \hat{\ell}_r \kappa(\hat{v}^{k,t}, v^{k,\Delta-r+1})}, \quad (k = 1, \ldots, n_d).$$

Then, we calculate empirical mean of channel vacancy forecast by $\hat{\bar{v}}^t = \frac{1}{n_d} \sum_{k=1}^{n_d} \hat{v}^{k,t}$.

C. AR Model of Kernel Embedding of Channel Vacancy Expected Value (KEV)

We model the expected value of channel vacancy as a AR model in RKHS $\mathscr{H}$. Consider that we only embed expected value of channel vacancy in RKHS. Then, we have $\kappa(\bar{v}, \cdot)$ which maps expected value of channel vacancy $\bar{v}$ to RKHS $\mathscr{H}$. AR process of such a mapping is given by $\kappa(\bar{v}^{z+1}, \cdot) = \sum_{r=1}^{\rho} \ell_r \kappa(\bar{v}^{z-r+1}, \cdot) + \epsilon_z$. Least-squares formulation to calculate parameters $\ell_1, \ldots, \ell_\rho$:

$$\min_{\ell_1,\ldots,\ell_\rho} \sum_{z=\rho}^{\Delta-1} \left\| \kappa(\bar{v}^{z+1}, \cdot) - \sum_{r=1}^{\rho} \ell_r \kappa(\bar{v}^{z-r+1}, \cdot) \right\|_{\mathscr{H}}^2 + \lambda \sum_{r=1}^{\rho} \|\ell_r\|_{\mathscr{L}}^2, \tag{16}$$

where $\ell_r \in \mathscr{L}$, for all r, are vector regressors that can be calculated by $\hat{\ell}_r = \sum_{z,z'=\rho}^{\Delta-1} \tilde{Q}_{z,z'} \bar{v}^{z+1} \sum_{r'=1}^{\rho} [\tilde{M}^{-1}]_{r,r'} \bar{v}^{z'-r'+1}$, for all $r = 1, \ldots, \rho$, where $\tilde{Q}_{z,z'} \in \tilde{\mathbf{Q}} = (\tilde{\mathbf{K}} + \lambda \mathbf{I})^{-1}$, and entries $\tilde{\mathbf{K}} \in \mathbb{R}^{(\Delta-1)\times(\Delta-1)}$ are given by $K_{z,z'} = \kappa(\bar{v}^z, \bar{v}^{z'})$; $[\tilde{M}^{-1}]_{r,r'} \in \tilde{\mathbf{M}}^{-1}$ where $\tilde{\mathbf{M}} \in \mathbb{R}^{\rho\times\rho}$ with entries

$$\tilde{M}_{r,r'} = \sum_{z,z'=\rho}^{\Delta-1} \tilde{Q}_{z,z'} \kappa(\bar{v}^{z-r'+1}, \bar{v}^{z'-r+1}).$$

Note that this is a direct result from Theorem (2) where instead of calculating kernel mean embedding $\hat{\mu}$, we only need to calculate kernel value $\kappa(\cdot, \cdot)$ of expected values of channel vacancies.

Remark 1. *Note that KME and KEV are methods in which we first do kernel embedding of some data and then use AR model in RKHS. On the other hand, KEC defines AR model of channel vacancy data in input space, then embeds it to RKHS which results in another AR model. In KEC, error term in input space is also embedded to RKHS. In KME and KEV, error is defined in RKHS itself.*

Remark 2. *Although we focus on interference prediction in multi-operator shared spectrum small-cell networks, kernel embedding technique can be applied in any wireless communications context where interference prediction is of interest.*

V. Channel Assignment

In this section, we study channel assignment for shared spectrum small cell networks. Without loss of generality, we define problem parameters of a particular operator. We examine two cases:

- *Proposed Communication-Free (CF) Scheme* in which channel assignments are simultaneously made by each operator at the beginning of each epoch. Extrapolation of channel vacancy to the next epoch is essential in this case.
- *Inter-Operator Coordination Protocol (CP) based approach*: With this approach, operators asynchronously perform assignments at the beginning of the epoch. Once an operator performs channel assignment for the current epoch, it informs other operators. Essentially each operator performs its assignments taking into account the known assignments from other operators. This reflects the approach taken in [10]. However we consider an idealized version assuming that zero coordination overhead is incurred.

A. Proposed Communication-Free (CF) Scheme

Maximum bandwidth that can be allocated to a mobile $i \in M_j$ can be given by $\sum_{c\in C(t)} w_c/|M_j|$; moreover, we are able to calculate average interference using extrapolated channel vacancy data, for example, any mobile i in channel c shall experience $\tau_e/\bar{v}_{ic}$ interference, where note that if $\bar{v}_{ic} = 1$ then, interference will be equal to τ_e which shall be considered to be negligible. Assignment variable is given by

$$\mathsf{x}_{cjj'} = \begin{cases} 1, & \text{channel } c \text{ is assigned to BS } j \text{ and } j', \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

where $\mathsf{x}_{cjj} \equiv \mathsf{x}_{cj}$ basically means that channel c is assigned to BS j. Besides, if channel c is assigned to BSs j and j', then we have $\mathsf{x}_{cjj'} = \mathsf{x}_{cj}\mathsf{x}_{cj'}$ which can be linearized with $\mathsf{x}_{cjj'} \leq \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B$, $\mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \leq 1, \quad \forall c \in C(t), \forall j, j' \in B$. We use Jensen's inequality to define a lower bound for throughput which requires to know only expected value of channel vacancy for calculating channel assignments. We define *extrapolated average raw throughput*[5] of mobile i in channel c as following:

$$\begin{aligned} \hat{R}_{jic}(\mathbf{x}) &= \frac{w_c}{|M_j|} \log\left(1 + \frac{P_{jc}\bar{g}_{jic}d_{ji}^{-\alpha}}{\tau_e/\hat{\bar{v}}_{ic} + N_0}\right) x_{jc} \\ &\geq \hat{\mathbb{E}}_{v_{ic}}\left[\frac{w_c}{|M_j|} \log\left(1 + \frac{P_{jc}\bar{g}_{jic}d_{ji}^{-\alpha}}{\tau_e/v_{ic} + N_0}\right) \mathsf{x}_{jc}\right] \end{aligned} \tag{18}$$

where $P_{jc} = \frac{1}{n_c}\sum_{s=1}^{n_c} P_{jc,s}$, $\bar{I}_{jic} = \frac{1}{n_c}\sum_{s=1}^{n_c} I_{jic,s}$ is average interference across subcarriers resulting from BS j to mobile i. Note that intra-operator interference is hidden in estimated $\tau_e/\hat{\bar{v}}_{ic}$. Moreover, we may set an upper bound I_{max} (e.g. $I_{max} = \tau_e$) for intra-operator interference that a mobile receives, i.e., $\sum_{j'\in B\setminus j} \bar{I}_{j'ic}\mathsf{x}_{cjj'} \leq I_{max}, \quad \forall c \in C(t)$. Lower bound of raw throughput is chosen to be the constraint related

[5]Whenever interference is observed, *observed average raw throughput* is calculated by $R_{jic}(\mathbf{x}) = \frac{w_c}{|M_j|} \log\left(1 + \frac{P_{jc}\bar{g}_{jic}d_{ji}^{-\alpha}}{\bar{I}_{ic}+N_0}\right) \mathsf{x}_{cj}$

to the traffic accumulated in a BS. For every BS j and traffic load L_j, we have constraint

$$\sum_{c\in C(t)}\sum_{i\in M_j}\hat{R}_{jic}(\mathbf{x}) \geq L_j, \quad \forall j \in B, \tag{19}$$

ensuring that traffic demand is satisfied. Traffic demand is assumed to be known at the beginning of the epoch.

We are interested in minimizing total spectrum usage while satisfying average traffic load requirements of BSs, i.e.

CHANNEL ASSIGNMENT PROBLEM

$$\begin{aligned} Z_{\text{IP}} := \min_{\mathbf{x}} & \sum_{c\in C(t)}\sum_{j\in B} w_c \mathsf{x}_{cj} \text{ subject to} \\ & \sum_{c\in C(t)}\sum_{i\in M_j}\hat{R}_{jic}(\mathbf{x}) \geq L_j, \quad \forall j \in B \\ & \sum_{j'\in B\setminus j}\bar{I}_{j'ic}\mathsf{x}_{cjj'} \leq I_{max}, \quad \forall c \in C(t), \forall j \in B, \forall i \in M_j \\ & \mathsf{x}_{cjj'} \leq \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B \\ & \mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \leq 1, \quad \forall c \in C(t), \forall j, j' \in B \end{aligned} \tag{20}$$

The reason of our choice to minimize total spectrum usage is with the aim that operators coexist fairly in unlicensed spectrum. It is well known that selfishness degrades spectrum efficiency in unlicensed/shared spectrum settings. Constraints in (19) may not be satisfied always, thus, we shall look at Lagrangian relaxation of the optimization problem which is given by

LAGRANGIAN RELAXATION

$$\begin{aligned} Z(\eta) := \min_{\mathbf{x}} & \sum_{c\in C(t)}\sum_{j\in B} w_c \mathsf{x}_{cj} + \sum_{j\in B}\eta_j\Gamma_j(\mathbf{x}) \text{ subject to} \\ & \sum_{j'\in B\setminus j}\bar{I}_{j'ic}x_{cjj'} \leq I_{max}, \quad \forall c \in C(t), \forall j \in B, \forall i \in M_j \\ & \mathsf{x}_{cjj'} \leq \mathsf{x}_{cj}, \quad \forall c \in C(t), \forall j, j' \in B \\ & \mathsf{x}_{cj} + \mathsf{x}_{cj'} - \mathsf{x}_{cjj'} \leq 1, \quad \forall c \in C(t), \forall j, j' \in B \end{aligned} \tag{21}$$

where Γ_j and $\eta_j \geq 0$, $\forall j \in B$ are gradients and Lagrangian multipliers, respectively: $\Gamma_j(\mathbf{x}) = L_j - \sum_{c\in C(t)}\sum_{i\in M_j}\hat{R}_{jic}(\mathbf{x})$.

Subgradient Optimization: We utilize subgradient optimization method to find "good" values of Lagrangian multipliers. Let us denote by δ^l the *step-size* in iteration l given by $\delta^l = \xi^l(Z(\eta^l) - Z^*)/\sum_{j\in B}\Gamma_j^2(\tilde{\mathbf{x}}^l)$ where $\tilde{\mathbf{x}}^l$ denote the assignment variables that solve optimally Lagrangian relaxed problem (21) in iteration l; $\xi^{l+1} = \frac{1}{2}\xi^l$ if Z did not increase in last Q iterations, otherwise $\xi^{l+1} = \xi^l$ with parameters $n_{max} > 1$, and $0 < \xi^0 \leq 2$; and we have $Z^* = \min_{0\leq l^*\leq l} Z(\eta^{l^*}) - (1+l)^{-1}$. Pseudo-code is given in Algorithm 1.

Algorithm 1 OPTIMAL CHANNEL ASSIGNMENT

Input: $\{\bar{I}_{jic}\}_{\forall j,i,c}$, $\{L_j, M_j\}_{\forall j}$, n_{iter}
Initialization: Choose starting values $\eta^0 = (\eta_1^0, \eta_2^0, \ldots, \eta_{|B|}^0)$
Extrapolate $\{\hat{\bar{v}}_{ic}\}_{\forall i,c}$ using KEM, KEC or KEV
if Z_{IP} not exists **then**
  while $l \leq n_{\text{iter}}$ **do**
    Compute $Z(\eta^l)$ and $\{\tilde{\mathsf{x}}_{cj}^l\}_{\forall c,j}$ from (21)
    Compute subgradients $\Gamma_j(\tilde{\mathbf{x}}^l), \forall j \in B$
    if $\Gamma_j^l = 0, \forall j \in B$ **then**
      STOP, because the optimal value has been found
    end if
    Compute δ^l
    Compute $\eta_j^{l+1} = \max(0, \eta_j^l + \delta^l\Gamma_j^l)$, $\forall j \in B$
    $l = l + 1$
  end while
end if

B. Benchmark: Channel Assignments with an Inter-operator Coordination Protocol (CP)

For evaluating the performance of our above described communication-free scheme, we consider a protocol which involves explicit coordination among operators via communication between them. Specifically, (i) operators inform each other about their current channel assignments, (ii) utilizing this information and current state of channel interference, they perform channel assignments asynchronously as follows: each operator perform assignments and broadcast this to opponent operators. By this way, they make effort not to interfere heavily with each other. Such a protocol shall provide a fair comparison with our proposed communication-free proactive channel assignment scheme. The obvious downside of this baseline scheme is the overhead associated with coordination in terms of delay and communication overhead but we overlook this drawback by not accounting this overhead.

Algorithm 2 CHANNEL ASSIGNMENT WITH COORDINATION

for every operator sequentially **do**
  $C \leftarrow$ select channels not used by others
  while Traffic load is not satisfied **do**
    $C \leftarrow C \cup$ a channel used by others
    Perform channel assignments using Algorithm 1
    if all channels selected **then**
      Exit from while-block
    end if
  end while
  Inform other operators about used channels
end for

VI. NUMERICAL RESULTS

In this section, we do numerical simulations for better understanding interference prediction error performance of our different AR models (§IV) and comparison of the proposed Communication Free (CF) optimal channel assignment with the baseline coordination protocol (CP) approach. We consider that available channels do not change over time. In every Monte Carlo iteration, we generate random locations of mobiles, and assume that mobiles are associated with their nearest BS. Unless otherwise stated, we set system and channel parameters as following: $P_s = 100/n_c$ mW, for every channel c and subcarrier s; Path loss exponent $\alpha = 3$, Rayleigh channel with variance 1; $\tau_e = -63/n_c$ dBm, $n_c = 1200$ subcarriers per every channel; Parameters in Algorithm 1: $n_{\text{iter}} = 100$, $n_{max} = 5$, $\xi^0 = 2$.

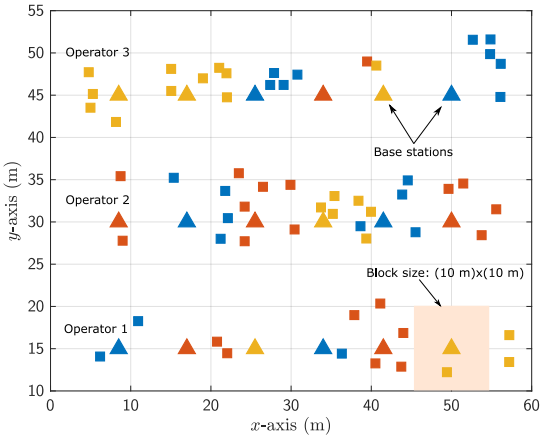


Fig. 3: Example scenario: Multi-operator deployment.

Locations of BSs and Mobiles: In every simulation, we divide area into grids, and consider that a mobile is somewhere within a "block" (e.g., part of a building) and the corresponding BS at the center of that block (Figure 3). We randomly generate locations of mobiles located in the vicinity of its associated BS. We consider a (60 m)×(60 m) area with 3 operators, each has deployed 6 small BSs. Every operator serves 20 mobiles. So, totally there are 18 BSs and 60 mobiles in that area.

Epochs: We calculate mean of considered variables using Monte Carlo method on hourly basis ($n_e = 24$ epochs per day) by collecting data going back to 50 days, although we find that benefit beyond 20 days in the past is marginal (see below).

Parameters of Kernel-based Extrapolation: For performing one step-ahead prediction, we use a sliding window of size $\Delta/2 + 1$ to estimate ρ. The first half part of the data $\Delta/2$ is used to compute ρ, and the last sample for performing one-step ahead prediction for different values of $\rho < \Delta/2 - 1$. We select the value of ρ which offers the lowest mean square error. The kernel used in extrapolating channel vacancy is Gaussian kernel $\kappa(v, v') = \exp(-|v - v'|^2/2\sigma_\kappa^2)$ where kernel variance σ_κ^2 is a critical parameter that determines the performance of the kernel, and we set it to $\sigma_\kappa^2 = 5$. We tried different kernel variances, and picked the aforementioned one because of its better performance. On the other hand, we found out that it is reasonable choosing $\Delta = 20$ (the values in the range $\Delta > 20$ did not significantly change the performance).

Cost of Feedback: Duration of an epoch is $T = 60$ seconds. For every epoch, we set $n_d = 1000$ examples. We assume that there is no any loss or error when channel vacancy and location data is fed back to CC.

Traffic Load: For each mobile, we target 60 Mbps throughput; thus, if there are $|M_j|$ mobiles associated with BS j then, total traffic load becomes $L_j = 60|M_j|$ Mbps.

A. Error Performance of Kernel-based Extrapolation

For any epoch t, mean square error is given by

$$\text{MSE}(t) = \frac{1}{50} \sum_{c \in C(t)} \sum_{i \in M} \sum_{D=1}^{50} |\bar{v}_{ci}(t - 24D) - \hat{\bar{v}}_{ci}(t - 24D)|^2$$

which captures the fact that t and D show the hour and day, respectively; further, we have $\bar{\text{MSE}} = \frac{1}{24} \sum_{t=1}^{24} \text{MSE}(t)$.

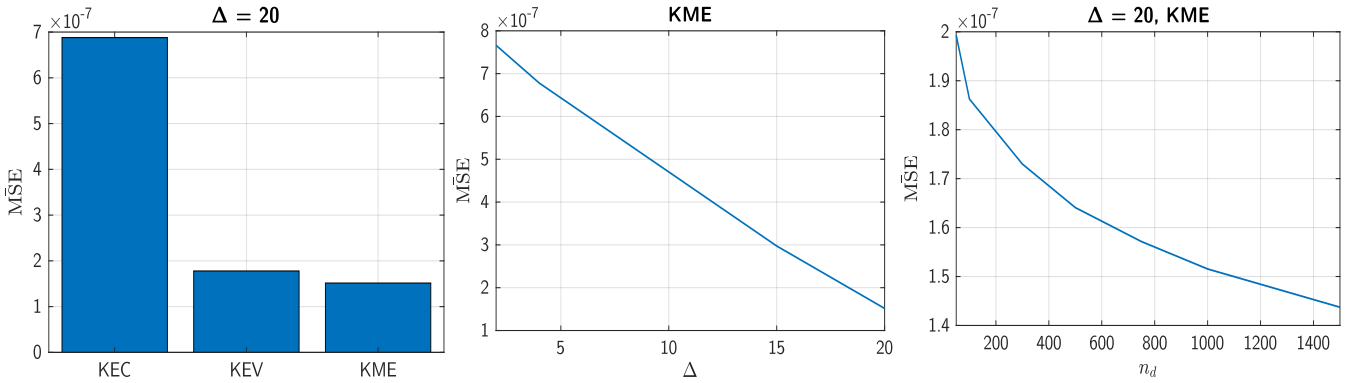
In Figure 4(a), we compare $\bar{\text{MSE}}$ performance of KEC, KEV and KME for $\Delta = 20$. KME has the best performance since it embeds all channel vacancy data while KEV embeds only its expected value. KEC performs worse since the embedded error to RKHS may cause degradation of performance. Figure 4(b) plots $\bar{\text{MSE}}$ performance of KME with respect to increasing values of Δ. It shows obvious advantage of using more data for extrapolation. On the other hand, $\Delta > 20$ did not provide better performance in terms of spectrum efficiency which is examined in the sequel. Figure 4(c) shows $\bar{\text{MSE}}$ performance of KME when n_d is increased. The result is not surprising to show that high volume of collected data is expected to demonstrate better error performance.

B. Spectrum Efficiency Performance

In Figure 5, we depict mean spectrum efficiency of three operators per epoch and its daily mean over 24 epochs. For no forecast case, we utilize last observed data in previous day. In Figure 6, empirical CDF of total allocated spectrum for 50-day data is shown for different extrapolation techniques and compared with no forecast and CP cases. Due to the high error with no forecasting implies that CF optimization makes the algorithm to minimize spectrum usage very inefficiently. The cost of such a minimization is seen in Figure 5 in which spectrum efficiency becomes lowest among others. There is also a direct relation of error performance of extrapolation and spectrum efficiency. While KME has the best spectrum efficiency (even better than CP), KEC has the worst as it is also valid in error performance (look at Figure 4(a)). KME and no-forecast provide about 7.5 bps/Hz and 6 bps/Hz mean spectrum efficiency, respectively. It implies that for a 120 MHz spectrum bandwidth, KME and no-forecast shall manage 900 Mbps and 720 Mbps traffic, respectively. Moreover, in the considered scenario, there are 20 mobiles per operator; so, traffic per mobile in case of KME and no-forecast is about 45 Mbps and 36 Mbps, respectively. CDF of spectrum allocation of CP is very similar to KME except for Operator 2.

VII. Related Work

Various aspects of inter-operator spectrum sharing have received attention in the literature. [18] surveys the work on licensed spectrum sharing. In [19], the interaction of operators is modeled as a repeated non-cooperative game, and the utilities of the game are chosen to provide useful properties. Static and dynamic sharing is studied with cooperation and punishment states where once an operator deviates from the cooperation state, the punishment is everlasting, so that the operator suffers a net loss in revenue. [20] considers a two-stage game for investment and competition when spectrum is shared between a primary and secondary operators. The authors demonstrate that



(a) $\bar{\text{MSE}}$ comparison of different learning techniques.

(b) $\bar{\text{MSE}}$ with respect to Δ.

(c) $\bar{\text{MSE}}$ with respect to n_d.

Fig. 4: $\bar{\text{MSE}}$ performance.

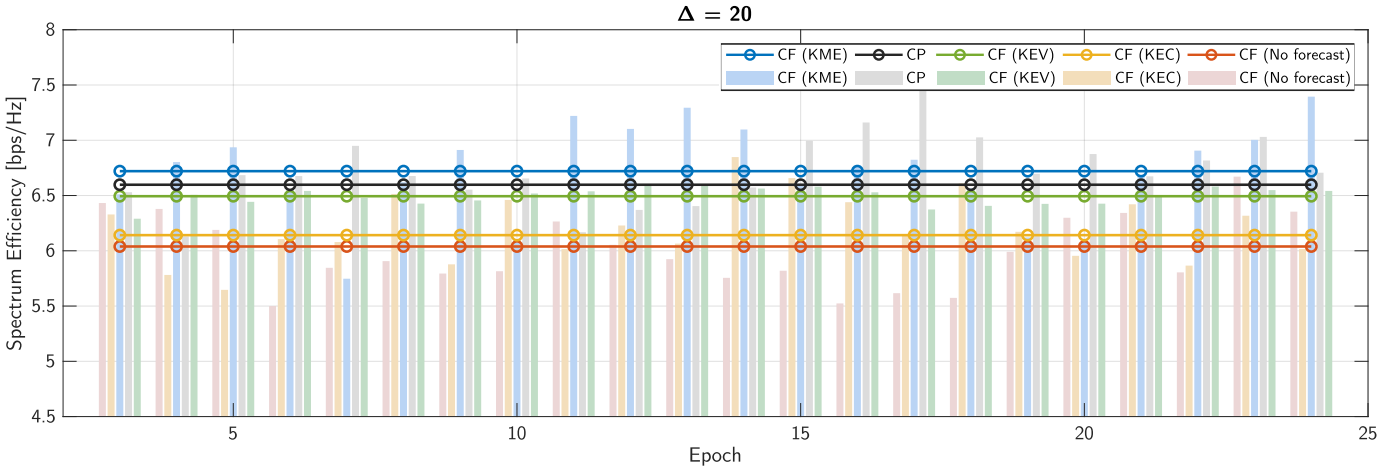


Fig. 5: (Bar) Mean spectrum efficiency per epoch. Mean has been calculated using a 50-day data of spectrum efficiency. Epochs are in hour-granularity. (Circle marked lines) Daily mean spectrum efficiency. CP = Coordination protocol, CF = Communication-free.

with their model only one secondary operator will invest and compete with the primary. In [12], two operator coexistence is studied where the first operator has only the macro-cell network architecture and does not possess small-cell network deployment, and the second operator has only the small-cell network architecture. Macro-cell network is benefited through offloading services offered by small-cell network, allowing macro-cell network to satisfy its users' capacity demands.

[8] describes an end-to-end architecture for CBRS spectrum and it does not focus on spectrum sharing. But as part of their architecture, the authors mention coordination of secondary spectrum sharing among operators by the SAS, which as we stated at the outset limits the ability to perform dynamic and fine-grained spectrum sharing. Similarly in the work of [9], operators coordinate through the common Global Spectrum Controller (GSC). [10] is an example of an alternative approach for our problem. Here the authors propose a inter-operator communication protocol that involves exchange of spectrum usage of an operator with others although channel state information itself is not disclosed. We compare this approach with our approach and show that we can get similar performance without having to communicate with other operators. Our approach takes a radically different and new approach from the aforementioned work by not requiring communication between operators. Moreover, as the identities of "interferers" cannot be gleaned from the sensed interference information, inter-operator interaction cannot be modeled as a multi-agent game. We therefore take the interference forecasting approach and have the operators independently determine their channel assignments based on that information.

In [16], the order of AR model is one; in this work, we take into account the general case and solve corresponding least squares problem. [21] considers AR model in tensor product reproducing kernel Hilbert space. Authors show its performance against other AR models. From a forecasting viewpoint, the authors in [23] use the classical time-series method of Holt-Winters for forecasting traffic of a slice in the 5G network slicing context to guide future admission control policies. In contrast, the forecasting problem we tackle is significantly harder as in our setting each operator needs

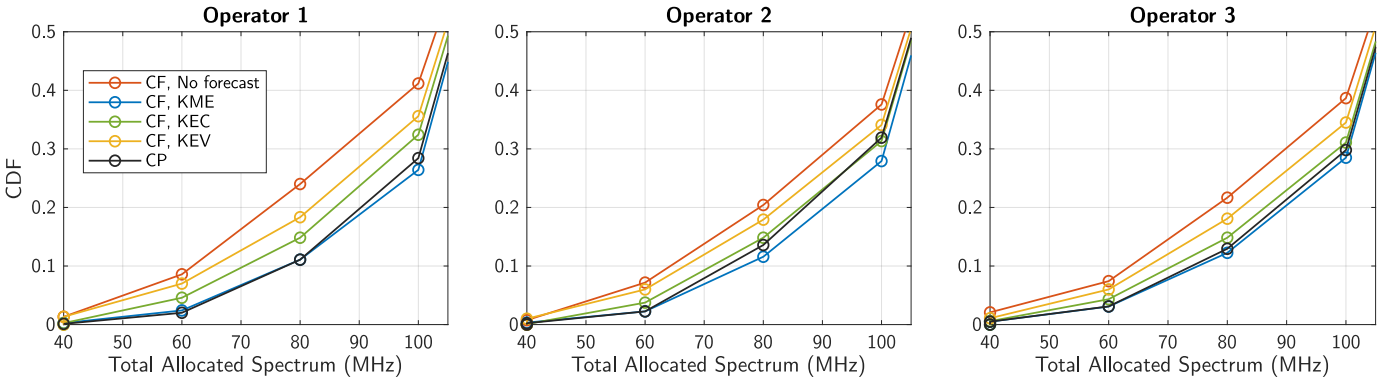


Fig. 6: CDF of total allocated spectrum. 50-day data has been used to generate empirical CDF.

to have an accurate notion of interference that is actually the consequence of continually changing actions (channel assignments) of other operators as well as other factors (time-varying channel conditions, user mobility, power control, etc.).

VIII. Conclusions

We have studied inter-operator spectrum sharing problem in shared spectrum small cell networks for which we devise a communication-free proactive optimal channel assignment scheme. As a key enabler of this scheme and core contribution of this work, we analyzed channel vacancy data which is defined over inter-operator interference, and by using kernel mean embedding technique, applied vector-valued regression for predicting time-varying probability distribution of channel vacancy and its expected value using various autoregressive (AR) models. We have compared our communication free scheme with coordination protocol and the case that does not involve any forecasting. Through extensive simulations, we showed that communication-free scheme performs at least as good or better than an idealized coordination protocol (with zero coordination overhead) and always outperforms the no forecasting alternative.

Acknowledgment

This work was supported in part by The Leverhulme Trust. We would like to thank Uğur Özdemir and the anonymous reviewers for their valuable comments and suggestions.

References

[1] J. Zander and P. Mahonen, "Riding the data tsunami in the cloud: Myths and challenges in future wireless access," *IEEE Communications*, Mar 2013.

[2] FCC, "FCC Rule Making on 3.5 GHz Band/Citizens broadband radio service," Apr 2015.

[3] Ofcom, "3.8 GHz to 4.2 GHz Band: Opportunities for innovation," *White Paper*, Apr 2016.

[4] K. Mun, "CBRS: New shared spectrum enables flexible indoor and outdoor mobile solutions and new business models," *White Paper*, Mar 2017.

[5] M. Matinmikko et al., "Spectrum sharing using licensed shared access: The concept and its workflow for LTE-ADVANCED networks," *IEEE Wireless Communications*, Apr 2014.

[6] S. Rathinakumar and M. K. Marina, "GAVEL: Strategy-proof ascending bid auction for dynamic licensed shared access," *ACM MobiHoc*, 2016.

[7] MulteFire Alliance, "MulteFire release 1.0 technical paper," *White Paper*, Jan 2017.

[8] C. W. Kim, J. Ryoo and M. M. Buddhikot, "Design and implementation of an end-to-end architecture for 3.5 GHz shared spectrum," *IEEE DySPAN*, 2015.

[9] M. G. Kibria, G. P. Villardi, K. Ishizu, F. Kojima and H. Yano, "Resource allocation in shared spectrum access communications for operators with diverse service requirements," *EURASIP J. Adv. Signal Process.*, 2016.

[10] B. Singh et al., "Coordination protocol for inter-operator spectrum sharing in co-primary 5G small cell networks," *IEEE Communications*, Jul 2015.

[11] S. Yoon, L. E. Li, S. C. Liew, R. R. Choudhury, I. Rhee and K. Tan, "QuickSense: Fast and energy-efficient channel sensing for dynamic spectrum access networks," *IEEE INFOCOM*, 2013.

[12] M. G. Kibria, G. P. Villardi, K. Nguyen, K. Ishizu and F. Kojima, "Heterogeneous networks in shared spectrum access communications," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 145-158, Jan. 2017.

[13] Nokia, "Nokia applies spectrum innovation to create private mobile networks and expand opportunities for connected vertical industries," *Press Release*, Feb 2017.

[14] A. Smola, A. Gretton, L. Song, B. Schölkopf, "A Hilbert space embedding for distributions," *in Hutter M., Servedio R.A., Takimoto E. (eds) Algorithmic Learning Theory. ALT 2007. Lecture Notes in Computer Science*, vol 4754. Springer, Berlin, Heidelberg

[15] L. Song, K. Fukumizu and A. Gretton, "Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric enference in graphical models," *in IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 98-111, July 2013.

[16] C. H. Lampert, "Predicting the future behavior of a time-varying probability distribution," *IEEE CVPR 2015*, 2015.

[17] M. Kallas, P. Honeine, C. Francis, H. Amoud, "Kernel autoregressive models using Yule-Walker equations," *Signal Processing*, vol. 93, no. 11, pp. 3053-3061, 2013.

[18] R. H. Tehrani, S. Vahid, D. Triantafyllopoulou, H. Lee and K. Moessner, "Licensed spectrum sharing schemes for mobile operators: A survey and outlook," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2591-2623, Fourthquarter 2016.

[19] F. Teng, D. Guo and M. L. Honig, "Sharing of unlicensed spectrum by strategic operators," in *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 3, pp. 668-679, March 2017.

[20] C. Liu, S. Fu and R. A. Berry, "Investing in shared spectrum," in *IEEE DySPAN 2017*, 2017.

[21] E. A. Valencia, M. A. lvarez, "Short-term time series prediction using Hilbert space embeddings of autoregressive processes," *Neurocomputing*, vol. 266, pp. 595-605, 2017.

[22] H. Lian, "Nonlinear functional models for functional responses in reproducing kernel hilbert spaces," *The Canadian Journal of Statistics / La Revue Canadienne De Statistique*, vol. 35, no. 4, 2007, pp. 597-606.

[23] V. Sciancalepore *et al.*, "Mobile traffic forecasting for maximizing 5G network resource utilization," *IEEE INFOCOM*, 2017.

[24] https://www.dropbox.com/sh/60pfe47lhq6m3ga/AABOcs-52ksec50FC7rcWla1a?dl=0