Flexible and Spectrum Aware Radio Access through Measurements and Modelling in Cognitive Radio Systems

FARAMIR

Document Number D4.3

REM Prototype Implementation

Contractual date of delivery to the CEC:	29.02.2012
Actual date of delivery to the CEC:	29.02.2012
Project Number and Acronym:	248351 - FARAMIR
Editor:	Ioannis Dagres (IASA)
Authors:	Ioannis Dagres (IASA), Natalia Miliou (IASA), Andreas Polydoros (IASA), Marko Angjelicinoski (UKIM), Daniel Denkovski (UKIM), Valentin Rakovik (UKIM), Vladimir Atanasovski (UKIM), Liljana Gavrilovska (UKIM), Petri Mähönen (RWTH), Janne Riihijärvi (RWTH), Jaap van de Beek (HWSE), Sun Guolin (HWSE), Berna Sayrac (FT), Sebastien Grimoud (FT).
Participants:	IASA, UKIM, RWTH, HWSE, FT
Workpackage:	WP4
Security:	PU
Nature:	R
Version:	1.0
Total number of pages:	76

Abstract:

This deliverable presents the design, prototype implementation, and performance evaluation for REM's, including the structure of their individual elements, their information-exchange and storage needs and associated complexity, the computational needs and associated complexity and, finally, what is expected to come out of any such REM and with what quality. This quality assessed from three angles (theoretical/analytic, simulation, experiment), thus leading to a total design package for the subject matter.

Keywords: Radio Environmental Maps, Radio Interference Field Estimation, Detection, Localization, Neighbourhood mapping

Document Revision History

Version	*Date*	*Author*	*Summary of main changes*
0.1	19.1.2012	IASA	Document structure for circulation
0.2	23.1.2012	UKIM	Changes on the TOC structure
0.3	03.02.2012	UKIM	Input in chapters 2 and 4
0.4	03.02.2012	FT	Input in chapter 4
0.5	03.02.2012	HWSE	Input in chapter 4
0.6	15.02.2012	IASA	Input in chapters 3,4 and 5
0.7	23.02.2012	UKIM	Input in chapter 5
1.0	28.02.2012	Janne Riihijärvi (RWTH)	Final inputs and additions
1.0f	29.02.2012	Petri Mähönen (RWTH)	Coordinator review and approval

Table of Contents

1 Introduction

The final stage in the conceptual process of designing efficient and functional REM's in the context of the FARAMIR project consists of a detailed description, on the one hand, of the texture and management of the information flow for all elements of any such envisioned REM architecture (device-data management, information-storage management, processing/fusion-data management, user-data management), as well as of the quantitative characteristics of the actual process employed (the choice of the signal processing technique) for deriving and updating the stored information based on the raw measured data. The latter component (quality of processing) is addressed in three steps: first, broad and general-use tools have been developed and are herein described, tools which allows the incorporation of any technique for single/multiple-emitter location, transmit-power estimation, and detection (presence of the sources). Second, the tools allows for performance assessment for each desired combination of such a technique plus any chosen scenario (number and location of sensors and sources, transmit powers, etc.). This assessment is both analytic (via Cramer-Rao bounds) plus simulation plus experimental (measurements). For the latter case, any performed set of measurements can be an input into these tools. This extensive ability to address the three major avenues of assessment (analytic, simulation, processed measurements) allows for a side-by-side comparison and extraction of valuable conclusions. This tool is an exclusive creation within FARAMIR and can be made available and broadly used by all the partners and the community.

Since one of the fundamental figures of merit of any actual implementation is the complexity of the components of the architecture and of the processing (an important parameter in assessing both the suitability of specific platforms as well as the ability to update the map dynamically with real-time updates), the deliverable has expended effort in assessing the complexity of each critical component in terms of processing requirement and memory, particularly for those modules that are expected to carry the brunt of the computing effort.

It should be clear that an actual implementation of a particular REM in real practice and for a very specific scenario will involve considerations not fully covered here. Still, the current deliverable does present a conceptually full and documented vista of the design landscape for REM's, including the broad outlines of all their elements, their information-exchange and storage needs and associated complexity, the computational needs and associated complexity and, finally, what is expected to come out of any such REM and with what quality. This quality assessed from three angles (theoretical/analytic, simulation, experiment), thus leading to a total design package for the subject matter.

The deliverable is organized as follows: the specific nuances of the spectrum sensing, data collection, storage, processing and utilization, from a conceptual point of view are discussed in Chapter 2. The developed supporting tools for the assessment of the proposed techniques are described in Chapter 3. Chapter 4 outlines the requirements in term of algorithmic complexity and platform hardware for implementing the proposed techniques. Finally, in Chapter 5, we present the performance results from several experiments, conducted using FARAMIR’s REM platform.

2 REM prototype structure and data model

This section discusses the specific nuances of the spectrum sensing, data collection, data storage, data processing and data utilization, from a conceptual point of view. It complements the general notion of REM utilization elaborated in [1], [2] and focuses on the data management aspects in terms of the prototype's architectural components. Figure 1 presents the global overview of the REM prototype architecture comprising the basic information on the REM related entities and their structure.

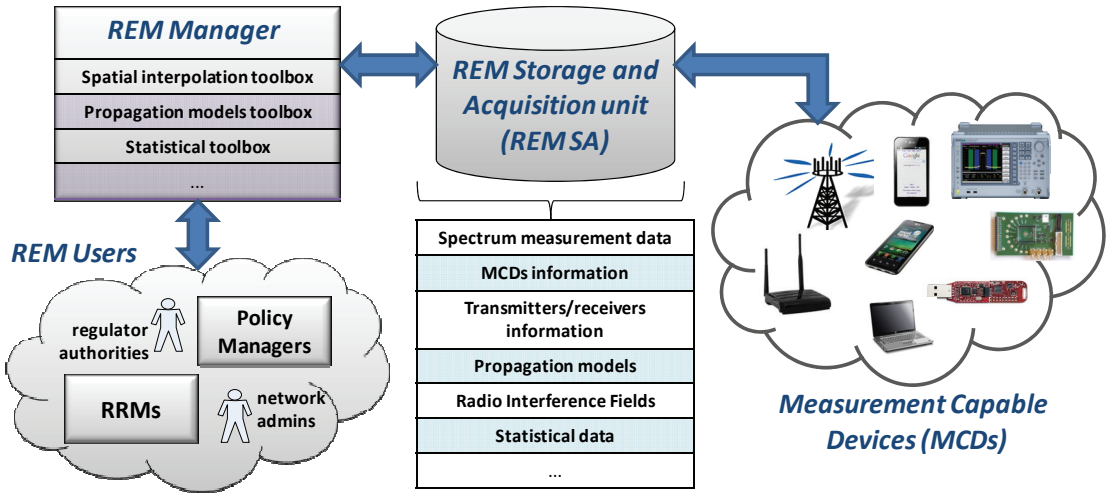

Figure 1: REM prototype architecture.

In the following we provide a short overview of especially the data management aspects of the prototype. Full description of all the developed components will be provided in the upcoming deliverable D6.2.

2.1 MCD data management

The REM prototype architecture should be able to integrate data from different types of spectrum sensing devices, considering their limitations, taking advantage of the diversity. The Measurement Capable Devices (MCDs) report their measurements to the REM-SA and provide options for remote reconfiguration and measurements querying. The REM prototype architecture envisions and enables the usage of the measurement capabilities of existing infrastructure components such as base stations, mobile terminals etc., or dedicated sensor networks. The different types of MCDs should be seamlessly integrated into the REM architecture in order to support heterogeneous MCD deployments. The following MCD aspects are tightly related with the REM data model and data management and should be carefully considered:

- **Sensing features.** Different MCDs types possess different sensing and detection capabilities. This makes the data management process diverse and more complex to

manage, but it can result in increased detection performance. The MCDs should be mutually calibrated and synchronized in order to get spatial observation data which is consistent and fusible.

- **Sensing techniques.** While the energy detection technique can provide data for RIF and propagation characteristics estimation, as well as localization, the engagement of more complex feature detection techniques can improve the detection performances and the assessment of the spectrum usage and provide capabilities to perform signal classification.
- **Geo-location capabilities.** In order to provide reliable measurement results and data that can be further used for REM processing, the MCDs must provide their exact location and time. The accuracy of the location and time information is as much important as the observed data precision for a reliable REM data derivation.

2.2 REM-SA data management

The REM-SA is the main storage of the measurement data. It should be able to store measurement data coming from different types of MCDs, associating the observation data to the type of the device and a sensing configuration used to perform the spectrum measurements. The REM-SA should constantly keep track of the active MCDs in the network and their current configurations, dynamically controlling the measurement process. Therefore, the REM-SA should be able to on-demand query the MCDs to make reconfigurations and perform specific measurements. The raw measurement data stored in the REM-SA can be dynamically used by the REM-Manager for the calculation of REMs, which are returned back to the REM-SA for storage and possible subsequent usage. Besides the raw spectrum measurement data coming from the MCDs, in terms of the data management and storage aspect the REM-SA should take into consideration the following information types:

- **Transmitter information.** This information type is related to the information concerning the type, location, transmit power, operating frequency and service area of the transmitter. Table 1 presents the transmitter information data storage model in terms of its sub-types and their elements.

Table 1: Transmitter information.

Sub-Type	Element	Comment
Transmitter Type	Incumbent or secondary	The transmitter type information can be important for many scenarios where a distinction between the transmitter types is needed.
	Type of incumbent/secondary	
Transmitter location	Based on coordinates	This type of location information can be used in cases where the exact location

		of the transmitter is needed, e.g. path-loss estimation. The transmitter locations can be estimated or statically input in the REM SA storage.
	Based on regions	This location information denotes the regions in which the transmitter is (deterministically or probably) located. Can be useful in scenarios like femto-cell deployment and TVWS.
Transmit power	Maximal transmit power	Useful for calculation of service areas and propagation models.
Operating frequency/bandwidth	Central frequency	The elements give insight of the spectral characteristics of the transmitter. Important for path-loss estimation. Can be utilized by variety of REM-User processes e.g. white space calculation, secondary spectrum sharing etc.
	Bandwidth	
Service area	Type of metric (RSS, SINR, bitrate, etc.)	Important for the case of TVWS scenarios.
	Result (polygon, pixel image)	

- **RIF information.** The RIF information type comprises the information concerning the radio interference fields (RIFs). Table 2 denotes the RIF information data storage model in terms of its sub-types and their elements.

Table 2: RIF information.

Sub-Type	Element	Comment
RIFs	Pixel image based	The pixel image representation requires the *dimension* and *resolution* of the estimated RIF. The radial basis function representation requires the *type of basis function, number of basis functions*, *corresponding*

	Radial basis function based	*function arguments*. Regardless of the representation type, the RIF estimation can be exploited by variety of processing tools, such as propagation models fitting, coverage areas estimation, detection of coverage holes, etc.
Operating frequency/bandwidth	Central frequency	The operating frequency gives the insight about the location of the RIF in spectral domain.
	Bandwidth	
Type of metric	RSS, SINR, bitrate, etc.	The RIFs can focus on different types of metrics, such as the distribution of the cumulative interference (derived based on RSS measurements), the spatial distribution of SINR (based on SINR measurements, e.g. by terminals, etc.)

- **MCD information.** This information type relates to the information on the MCDs, i.e. their measurement types, measurement configurations, detection methods and location. Table 3 denotes the MCD information data storage model.

Table 3: MCD information.

Sub-Type	Element	Comment
Measurement type	Sampling type (RSS, I-Q)	The diverse MCD capabilities can be exploited in different scenarios. For example, feature detection can be utilized for transmitter type identification, while energy based detection can be used for RIF and propagation estimation
	Detector type (RMS, Max Hold, Mean Hold, Min Hold)	
	Detection methods (energy, feature etc.)	
Measurement configurations	Resolution bandwidth	
	Sampling frequency	
	Operational bandwidth	
Location	Coordinates	The location of the MCD provides crucial information for process like RIF and propagation model estimation.

- **Propagation information.** The Propagation information type is consisted of information regarding the propagation model, environment and area dimensions. Table 4 denotes the Propagation information data storage model.

Table 4: MCD information.

Sub-Type	Element	Comment
Propagation model	Type of propagation model (arguments of the model)	Carries the information about the model type and its parameters.
	Area dimensions (latitude/longitude)	Every propagation model should be associated with an area of interest defined as a polygon with given longitudinal and latitudinal coordinates. This notion defines the spatial applicability of the given model.

- **Spectrum usage information.** The Spectrum usage information type should store data concerning the activity in a specific spectrum band in a given area of interest. Table 5 denotes the Propagation information data storage model.

Table 5: Spectrum usage information.

Sub-Type	Element	Comment
Spectrum activity	Duty cycle	The category includes the duty cycle and the distribution of the received signal from different transmitters at a given point of interest. This information can be used for localization purposes as well as for improved handover management and user density estimation.
	RSS distribution	
	ON/OFF activity patterns	
Area size	Area dimensions (latitude/longitude)	Every spectrum usage model should be associated with an area of interest. This notion

		defines the spatial applicability of the given information type.
Frequency band /channel	Central frequency	This information is necessary, since the spectrum activity should be associated with the frequency band.
	Bandwidth	

In order to increase its flexibility and responsiveness, the REM-SA should comprise *hierarchical storage architecture*. This approach implies that specific sub-sets/instances of the REM information would be stored in different layers of the REM-SA by exploiting the process of *data replication*. For example, consider the RIF information data management in a case of a femto-cell scenario, where the REM-SA is deployed at the HeNB and eNB level. The HeNB REM-SA will store the RIF information regarding its surrounding environment only, while the RIF information at the eNB REM-SA level would utilize the data stored in all of its underlain REM-SAs (i.e. HeNB REM-SAs).

2.3 REM-Manager data management

The REM-Manager is responsible for the processing and fusion of the raw measurement data into an accurate REM, i.e. an estimation of the spatial RIF, estimation of the transmitters' locations, various spatio-temporal statistics of the spectrum usage at different frequency bands, activity patterns, propagation model estimation, etc. The REM-Manager can perform real-time processing of the spectrum measurement data as well as processing on historical data stored in the REM-SA. The processed data can be stored in the REM-SA for further analysis, or reported to the REM-User which can utilize it for different purposes. This yields a modular REM-Manager development supporting different spatial interpolation, source localization and statistical analysis toolboxes. This sub-section describes the REM-Manager from a conceptual point of view focusing on its structure and features from a data modeling and information flow aspect. The main feature of the REM-Manager is its modular organization which enables easy management of the processing toolboxes (i.e. addition and removal of the toolboxes) and scalability. In terms of the data management aspect the REM-Manager architecture should be consisted of the following processing toolboxes in order to facilitate the envisioned processing tasks:

- **RIF toolbox.** The main task of this toolbox is to estimate the RIF in a given area of interest based on the stored MCD measurements in the REM-SA. The RIF estimation process is performed by spatial interpolation techniques and can be represented with pixel images or a set of radial basis functions. The pixel matrix representation delineates the result in a standard approach where the values of the pixels are stored in a matrix. The radial basis function representation delineates the RIF estimation based on a set of radial basis function and the values of its arguments. This toolbox should comprise multiple interpolation techniques and utilize a given one based on the scenario and REM-User requirements while targeting a given tradeoff between the precision of the technique and its computational complexity.

- **Location estimation toolbox.** The location estimation toolbox focuses on the position estimation of a transmitter/signal source based on the stored MCD measurements as well as the stored RIF estimations performed by the RIF toolbox. The toolbox should support the localization of both incumbent and secondary transmitters in order to provide rich environmental information to the REM-User. Additionally, the toolbox should have the capability to localize the transmitter based on its coordinates as well as based on the region in which the transmitter is located.
- **Propagation model toolbox.** The propagation model toolbox aims to estimate the path-loss in a given region of interest based on approaches elaborated in [2] as well as based on approaches that exploit the stored MCD measurements and stored RIF estimations. Additionally the toolbox should support the specification of the area of interest and support the usage of terrain data (for outdoor environments) or floor-plans (for indoor environments) for the calculation of the propagation model and parameters.
- **Service area toolbox.** The main task of the toolbox is to compute the service area of a given transmitter based on RSS or SINR measurements/estimates or mapping the measurements/estimates into higher layer metrics like Quality of Service metrics. The service area computational process underlies on a set of procedures that exploit the RIFs, transmitter configurations and propagation models. The service area results can be represented as polygons or binary pixel images. A typical case of the service area usage is the TVWS scenario, where the PU (TV) users' protection is of outmost importance.
- **Spectrum usage toolbox.** The focus of the spectrum usage toolbox is to calculate and present the activity levels of the transmitters based on variety of statistical methods such as semi Markov On-Off models and duty cycles. Moreover, it aims to calculate the distribution of the received signal strength from different transmitters at a given arbitrary point of interest. The spectrum usage toolbox and its features can be exploited by intelligent REM-Users (network entities), capable of utilizing the past behavior of the environment, in order to improve the network performance (e.g. Cognitive Resource Manager (CRM), Policy Manager (PM), etc.).

2.4 REM-User data management

The potential benefits of the REM data utilization can be seen through the high number of applications scenarios focusing on: REM facilitated frequency/power allocation, REM assisted spectrum access/(re-)usage, network optimizations through REM based service areas evaluation, etc. Besides the benefits on the aspect of operators and cognitive users, the REM data is a necessity for the regulator bodies to evaluate and track the spectrum usage, localize unauthorized transmissions and evaluate the conveyed frequency planning strategies.

In compliance with the previous discussion, there are several different types of REM users, such as Radio Resource Managers, Policy Managers, operator network administrators, regulator and public dedicated bodies etc. The following paragraphs explain the required REM data information types and the REM information flows with respect to the different types of REM users.

- **RIF information.** The RIFs can be a valuable asset to each type of the REM data users. They present a global overview of the spatial radio field and the propagation characteristics of the environment. All of the mentioned REM users can take advantage on this type of information, i.e. Radio Resource and Policy Manager can use this information to calculate the summary interference and locate congested areas in space and frequency, as well as spectrum holes and underutilized frequency/space areas. Using this information, these automated resource management entities can evaluate the spectrum efficiency and, hence, convey an improved frequency and power allocation. The RIF data can be useful for the network planning purposes of the operators, since it provides a rough estimate on their coverage areas and can help localizing coverage holes and equipment dysfunctions/malfunctions etc. The regulator bodies can use the RIF information to follow the spectrum usage and evaluate the spectrum planning.
- **Transmitter information.** The information on the active transmitters in the inspected area, their locations and used configurations, can be also valuable information to all REM users. The REM Manager utilizing the transmitter localization toolbox can dynamically localize and track transmitters, or detect the activity of a certain static transmitter (base station) etc. The information on the density of transmitters in a specific area, their configurations in terms of frequency, power and medium access schemes/parameters, can be an input to the RRM and PM to automatically evaluate the current spectrum sharing and resource allocation strategies and convey new if beneficiary for the network and the users. The PM and RRM entities can also use the transmitter localization REM feature to detect emergency situations such as the appearance and presence of a primary user. The operators can use the information on the user densities to better plan their resources, enhance the network resources in overloaded areas, or offload resources from underutilized areas. In addition, the regulator bodies can use this type of information to localize and track unauthorized users and transmissions.
- **Propagation information.** The propagation information is also crucial for the resource management problems, especially to the power and frequency allocation. Based on the estimated propagation losses and the estimated receiving SINR, the appropriate *MCS* can be selected by the RRM. The propagation losses can be also useful in situations when it is required to predict the caused interference to neighbor (primary) users, and limit the secondary user transmit power properly. The operators can use this information for planning purposes since the appropriate propagation model and parameters are vital for the accurate coverage/service areas estimation.
- **Spectrum usage information.** The spectrum usage information is very important information type especially, from the DSA perspective of view. The information on the available and underutilized channels, the duty cycles of the channels usage, mean ON/OFF periods, can be utilized by a Policy Manager to create/change spectrum policies allocating channels to different types of cognitive users. The spectrum usage information can be furthermore used by the PM to evaluate existing policies, as well as to learn and to enhance the policies backtracking the actions and the respective outcomes. Similarly to

the previous REM data types, the operators and regulators can benefit in terms of their respective network and spectrum planning activities.

- **MCD information.** The MCD information type is also important for the REM User. Using the information on the current configurations on the MCDs and their capabilities, the REM Users can require some specific measurements that can be beneficiary for the planning and the evaluation of their actions.

3 Supporting tools

The HW characteristics/constraints of the FARAMIR REM prototype allow compatible techniques to be implemented and assessed through real measurements. Two support tools have been developed in order to evaluate the performance of the proposed, compatible or not, techniques. The first one targets the assessment of fusion techniques for passive localization and RIFE and the second the assessment of emitter detection techniques. Both tools support techniques proposed within FARAMIR (see [3]) and also other proposed in the literature. They support also all the needed functionalities in order to evaluate their performance using a simulated environment as well as offline measurements taken from the FARAMIR REM prototype. They also support the computation and display of the respective theoretical bounds. The functionalities of these tools will be described in this chapter.

3.1 Fusion tool

This is a tool that was created in order to yield guidance to the experimentation work for passive localization and RIFE. It supports all four kinds of localization techniques, based on RSS, TOA, TDOA, and AOA. The GUI of this tool is displayed in Figure 2. It consists of a main axis area that displays visually the placement of the main players, the sensors and the emitters, and a number of panels that groups associated functionalities. Those functionalities will be described at the following sub-sections.

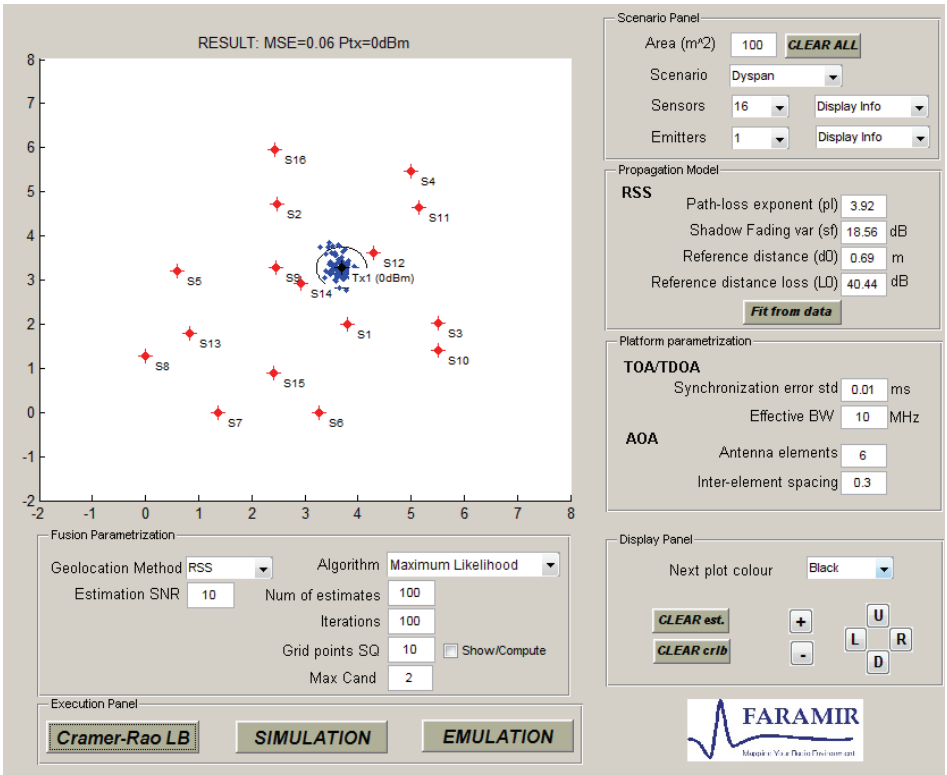

Figure 2: Fusion component GUI.

3.1.1 Scenario definition

The scenario panel groups the needed functionalities for a user to define a specific scenario. The scenario parameters describe the sensor deployment and the active emitters. This parameterization can be a custom one, where the user must provide all the needed information, or a specific scenario, that can be chosen from a list of predefined scenarios. The predefined scenarios are based on real experimentation, and thus supporting real measurements, conducted utilizing the FARAMIR platform. These offline measurements enable the field based experimental capability of this tool. The parameterization of the predefined scenarios can be used as baselines in order generate new scenarios. Up to now, the supported scenarios are based on RSS measurements. This means that the TOA, TDOA and AOA methods can be evaluated only at simulated environment.

3.1.2 Propagation model

This panel contains the parameterization of the propagation model. A classic log-distance model for the received power is used, described by the following equation:

$$P_r^i = P_{tx} - (L_0 + 10a\log(d_i / d_0) + n_{sf}) \quad (1)$$

Here L_0 and d_0 are the reference distance and power loss, a is the path loss exponent, n_{sf} is the log-normal shadow fading component. P_r^i is the received energy (in dB) of the i-th sensor due an emitter of transmit power P_{tx} at distance d_i.

The parameter values can be set either manually or estimated based on the stored measurements of the selected scenario. The estimation is based on an Ordinary Least Squares (OLS) fit of the data to the model described in (1). This fit is described by the following expression:

$$\begin{bmatrix} L_0 \\ a \end{bmatrix} = \left(H^T H\right)^{-1} H^T B \quad (2)$$

Here

$$H = \begin{bmatrix} -1 & -10\log_{10}(d_1 / d_0) \\ \vdots & \vdots \\ -1 & -10\log_{10}(d_i / d_0) \\ \vdots & \vdots \\ -1 & -10\log_{10}(d_{N_s} / d_0) \end{bmatrix} \text{ and } B = \begin{bmatrix} P_r^1 \\ \vdots \\ P_r^i \\ \vdots \\ P_r^{N_s} \end{bmatrix} \quad (3)$$

The reference distance is selected based on the sensor that is closest to the emitter. Compatible scenarios are the ones where only one emitter is active. The shadow-fading variance is estimated based on the error between the predicted and the real measurements.

3.1.3 Platform parameterization

In this panel the user must provide some basic platform characteristics that related to the performance of some passive localization methods. Related to the TOA/TDOA methods, the synchronization error of the sensors and the effective bandwidth of the emitter signal must be given. The synchronization error models the capability of the sensors to make synchronized measurements. Assuming that a synchronization protocol or a dedicated hardware is responsible for this, it is fair to model the remaining error as zero-mean Gaussian. The variance of this error in microsecond scale models this synchronization error. The effective bandwidth of the emitter signal is a parameter that can be used for characterizing the best achievable performance of a TOA estimator. Let the received signal at the *i*-th sensor is modeled as

$$r_i = A_i s(t - t_i) + \upsilon_i(t) \tag{4}$$

where t_i is the propagation delay. It has been shown in [5] that when the effective bandwidth of the signal and the SNRs at the sensors are sufficiently large, the error on estimating t_i can be approximated as zero mean Gaussian with variance

$$\sigma_i^2 = \frac{1}{8\pi^2 \beta^2 R_i} \tag{5}$$

where the effective BW β is defined as

$$\beta = \frac{\int f^2 |S(f)|^2 \, df}{\int |S(f)|^2 \, df} \tag{6}$$

and the reception SNR is defined as

$$R_i = \frac{\int |A_i s(t)|^2 \, dt}{N_0} \tag{7}$$

where A_i is the channel gain, υ_i the additive noise and $S(f)$ the furrier transform of $s(t)$.

Another parameter that is crucial to the estimation noise is the synchronization error. A simple model is adopted, zero mean Gaussian pdf with standard deviation given in microsecond. This is an idealized model but it is commonly used to evaluate requirements in synchronization accuracy. The simulator just adds the two above sources of errors.

The number of antenna elements that the sensors support and the inter-elements spacing are also two parameters that are used to compute the CRLB for the estimation of the AOA. All these

parameters are used to abstract the ranging parameter estimation procedure by generating a random error to true parameters at each sensor, based on the computed CRLB. Let ρ_i be the received angle at the *i*-th sensor. Ususally a function of this angle is measured,

$$\hat{\omega}_i = \omega_b(\rho_i) + \delta_i \tag{8}$$

where δ_i is Gaussian, zero mean with

$$\mathrm{var}(\delta_i) = \left(2R_i \frac{da_i^H(\omega_i)}{d\omega_i} \frac{da_i(\omega_i)}{d\omega_i} \right)^{-1} \tag{9}$$

with a_i the steering vector. If a uniform linear array (ULA) is used with Δ the normalized distance between two adjacent elements, then the steering vector is $a_i(\omega_i) = (1, e^{j\omega_i}, ..., e^{j(K-1)\omega_i})$ where $\omega_i = 2\pi\Delta\cos(\rho_i)$, then

$$\mathrm{var}(\delta_i) = \frac{3}{K(K+1)(2K+1)R_i} \tag{10}$$

This noise variance is used at the simulator, using the input SNR, number of antenna element and inter-element spacing input values.

Based on the platform description we use the CRLB of the ranging parameters estimation to create an abstraction layer at the sensor level. This enables the assessment of various fusion techniques with low-computational requirements.

3.1.4 Execution/Fusion parameterization panels

The execution panel contains the functionalities for computing the localization CRLB and for performing estimation based on simulation as well as experimentation. The CRLB computation uses the statistical description of the ranging parameter estimates (using the CRLB based abstraction at the sensor level) and displays the Root Minimum Mean Square localization Error ($2\times\mathrm{RMMSE}$). In the current version the *LS* and *ML* estimation are supported for all localization techniques as described in [3]. They are the two representatives of poor-and high performance algorithms with low and high computational requirements respectively. The *LS* algorithm assumes perfect parameter estimation at the sensors. It computes the solution of an over-determined system of equations that describes the measurements (over-determined assuming large number of sensors) by the use of the pseudo inverse matrix. The *ML* solution assumes perfect statistical knowledge of the propagation model (in the RSS case) and the ranging parameter estimates (in all cases). Build-in Matlab functions for non-convex optimization are used to find the *ML* solution and perfect initial point is used. The SNR input is related to the ranging parameter estimation performance which is in turn related to the input system SNR and the integration time used for the estimation. For the case of RSS three additional algorithms are supported. One is for single emitter localization based on joint propagation model and *Tx*-power estimation [6] using a rectangular

grid of candidate locations. This was the first algorithm implemented to the REM prototype, and it was successfully used at the IEEE DySPAN 2011 demonstration, for example. The tool provides the functionality to control the density of the rectangular grid and the coverage area of interest. The other two algorithms are for multiple source localization and transmit power estimation. The first one uses *LS* estimation of the transmit power on all candidate points and iterative elimination of candidate locations up to a predefined number of maximum active emitters. The second one computes the *ML* set of transmit power of all candidate locations and has been detailed analyzed in [3]. Both are using either a rectangular grid of candidate locations or a set of arbitrary candidate locations. These two options represent two different scenarios. One where no prior knowledge exists about the number and the positions of active emitters (the rectangular grid), and one where prior knowledge (possibly from a REM) is used to characterize the potential locations and maximum number of the active emitters.

3.2 Emitter detection tool

This tool is created in order to be used as guidance to the assessment and experimentation work for the emitter detection techniques. It supports a large number of detection techniques, those proposed within FARAMIR and others proposed in the literature. The GUI of this tool is displayed in Figure 3.

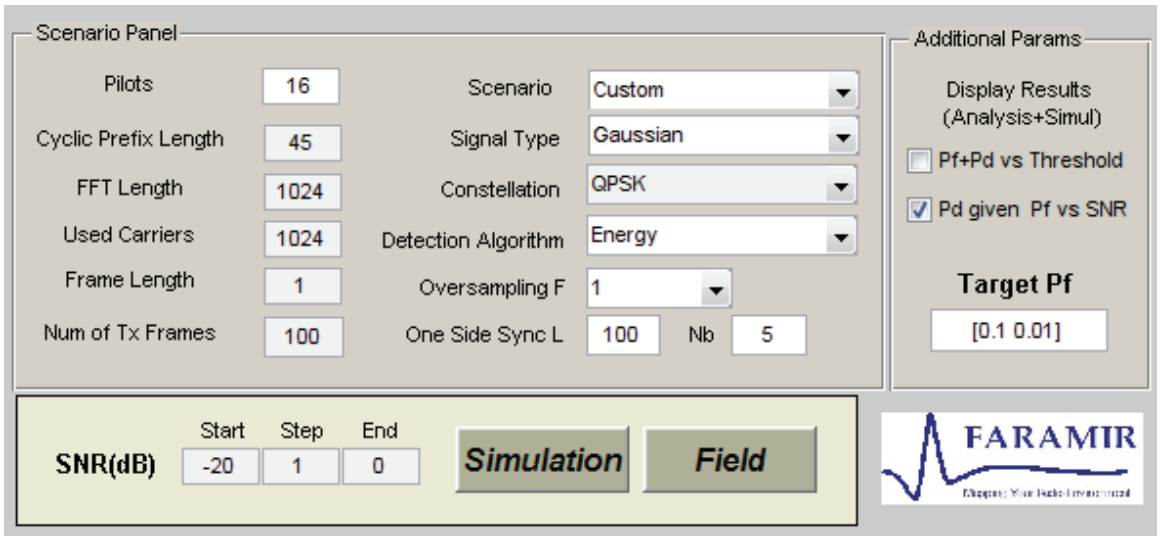

Figure 3: Emitter detection tool GUI

It is targeted on OFDM based tx-signal modulation formats, but it can easily extended support other formats too. The emitter signal can be either a modulated OFDM pulse with fully parameterized characteristics or a Gaussian 'OFDM' like pulse. This functionality enables to test the Gaussianity assumption, usually followed when analyzing the performance of proposed tests. On the receiver side, it supports a large list of detection algorithms, energy based, correlation based and cyclo-correlation based. It also supports the respective theoretical bounds (if exists). The

oversampling of the received signal is also a parameter. Choosing a non-custom scenario enables the Field based performance assessment by using offline stored measurement.

4 Algorithmic complexity and platform requirements

The algorithmic complexity and platform requirements of all techniques proposed in D4.3 ([3]) for REM generation will be commented in this chapter. A description on the needed data representation and storage requirements, the computational complexity and the HW requirements will be provided.

4.1 Direct RIFE

4.1.1 Spatial interpolation techniques

This part focuses on the evaluation of the computational complexity and the platform requirements of a set of spatial interpolated techniques, proposed as the basic RIF generation methods in the REM prototype.

4.1.1.1 Overview of the spatial interpolation techniques

The spatial interpolation techniques in the focus of the REM prototype were explained in details in deliverable D4.2 [3]. This subsection briefly elaborates on the main processing steps in these methods for completeness. These processing steps will be in the focus of computational complexity evaluation in subsection 4.1.2.

4.1.1.1.1 *Classic IDW method*

The classical IDW method [3] uses all the available observed spatial field measurements to perform the interpolation. In the case when there are N observed locations, the following interpolation function is used:

$$\hat{P}^r(x,y) = \sum_{i=1}^{N} d_i^{-d_{\exp}} P_i^r \Big/ \sum_{i=1}^{N} d_i^{-d_{\exp}} \ . \tag{11}$$

where P_i^r is the measured power at the point (x_i, y_i), d_i is the Euclidean distance between the referred point and the interpolation point (x, y) and $d_{\exp}$ is the distance exponent.

4.1.1.1.2 *IDW based Modified Shepard's Method*

The IDW modified Shepard's method [8] uses local nodal functions, representing two-dimensional surfaces fitted around each observation location. These nodal functions can take different analytical forms (may be quadratic, linear, constant etc.) and the appropriate form can be selected for the specific type of inspected phenomena and environment surroundings. The presented formulas focus on the MSM version having a fixed N_q and N_w parameters, instead of fix radii of influence R_q and R_w.

The estimation of the nodal functions parameters for an observed data point (x_k, y_k) is performed using a *Weighted Least Squares (WLS)* fitting based on the measured values of its N_q nearest neighbors:

$$\beta_k = (X_k^T W_k X_k)^{-1} X_k^T W_k y_k \,. \tag{12}$$

Here, $\beta_{k\,[p\text{x}1]}$ represents the vector of fitted nodal function parameters and $X_{k\,[N_q\text{x}p]}$ is the matrix of input parameters. The vector $y_{k\,[N_q\text{x}1]}$ is the output vector containing the residual power values ΔP_i ($= P_i - P_k$), $i \in D_k$, at the respective N_q neighboring points. D_k represents the circle region around the point (x_k, y_k) defined by its nearest N_q neighbor observation points, bounded with radius R_q^k. In the case of a linear nodal function ($p = 2$), the input matrix $X_{k\,[N_q\text{x}2]}$ contains the residual $\Delta x (= x_i - x_k)$, $\Delta y (= y_i - y_k)$ coordinates of the N_q neighbors, while in the case of a quadratic nodal function form ($p = 5$), $X_{k\,[N_q\text{x}5]}$ is constituted of the values $(\Delta x, \Delta y, (\Delta x)^2, (\Delta y)^2, \Delta x \Delta y)$ for each of the N_q nearest observation points. The weighting matrix W_k is a diagonal matrix containing the assigned weights w_i respective to each neighboring point:

$$w_i\,(x_k, y_k) = \left[\left(R_q^k - d_{ik}\right) \Big/ R_q^k d_{ik}\right]^{d_{\mathbf{exp}}}, \quad i \in D_k \,, \tag{13}$$

where d_{ik} is the distance between points (x_i, y_i) and (x_k, y_k), and $d_{\mathbf{exp}}$ is the distance exponent.

Finally, the interpolated value of the received power at an arbitrary spatial point (x, y) is calculated using the following function:

$$\hat{P}^r(x, y) = \sum_{k \in D} W_k(x, y) Q_k(x, y) \Big/ \sum_{k \in D} W_k(x, y) \,, \tag{14}$$

where W_k is the weight assigned to the k^{th} neighboring point at the location (x, y), while the region D represents the region around the point (x, y) containing its nearest N_w neighbor observation points. The weights W_k are once again calculated using the same formula as in eq. (13), replacing R_q with R_w. Q_k is the output of the nodal function of the data point k at the location (x, y) calculated as:

$$Q_k(x, y) = \beta_k{}^T \begin{bmatrix} x - x_k \\ y - y_k \end{bmatrix} + P_k \,, \tag{15}$$

for the case of a linear nodal function type. When quadratic function is used eq. (15) has a slightly different formulation.

4.1.1.1.3 *Gradient plus Inverse Distance Squared interpolation method*

The GIDS method [9] uses the same apparatus as the MSM method. However, instead of calculating a local multivariate function for each of the observation points, this method fits a global multivariate function around each unobserved location (x, y) of interest. The multivariate function can also take different analytical forms (may be quadratic, linear, constant etc.). However, oppositely on MSM, the GIDS interpolation performs an *Ordinary Least Squares (OLS)* regression, using the N_w closest neighboring locations to calculate the multivariate function coefficients:

$$\beta = (X^T X)^{-1} X^T y \,. \tag{16}$$

The notations here have the same meaning and form as in MSM (subsection 4.1.1.1.2), using N_w instead $\text{of } N_q$. For the case of a linear function model the predicted power at point (x, y) is calculated to be

$$\hat{P}^r(x,y) = \sum_{k\in D}\left(\frac{1}{{d_k}^2}\beta^T\begin{bmatrix}x-x_k\\ y-y_k\end{bmatrix}+P_k\right)\Bigg/\sum_{k\in D}\frac{1}{{d_k}^2}, \tag{7}$$

where the region D represents the region around the point (x, y) consisting its N_w nearest neighbors.

4.1.1.1.4 *Ordinary Kriging*

The most widely used geostatical interpolations are based on the *Kriging* method [10]. These are essentially optimal linear interpolation techniques in the sense of having minimum root-mean-squared-error. The Kriging based methods, specifically, use semivariograms to estimate the radio field and require stationarity of the measurement data in order to obtain a consistent estimation. The semivariograms express the statistical dependence of the field, which in the case of isotropic and intrinsically stationary field is only dependent on the distance ($\gamma(h)$). The semivariogram model and parameters are a necessity for the Kriging estimation. Various types of theoretical semivariogram models exist (spherical, gaussian, exponential, matern, etc.) and their parameters can be empirically evaluated by means of using *ML* or *LS* types of estimations. In the case of N spatial observations, an ordinary Kriging estimate is calculated using the following formula:

$$\hat{P}^r(x,y) = \sum_{i=1}^{N}\lambda_i P^r(x_i, y_i), \tag{17}$$

where the λ_i weights are calculated (using the known/estimated semivariogram model) as a subject to a constrained linear optimization, and their solution can be found as

$$\begin{bmatrix}\lambda_1\\ \lambda_2\\ \cdots\\ \lambda_N\\ m/2\end{bmatrix} = \begin{bmatrix}\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} & 1\\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} & 1\\ & & & & \\ \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN} & 1\\ 1 & 1 & & 1 & 0\end{bmatrix}^{-1}\begin{bmatrix}\gamma_{01}\\ \gamma_{02}\\ \\ \gamma_{0N}\\ 1\end{bmatrix}, \tag{18}$$

where γ_{ij} is the semivariance at a distance same as the one between points represented as i and j, while γ_{0j} is the semivariance at the distance equal to the distance between the point of interest (x, y) and the j^{th} point.

4.1.1.1.5 *Compressed Sensing*

In the RIFE estimation approach as it was outline in D4.2 [3], we cast the field estimation problem in the form of a reconstruction problem, which we will solve via techniques proposed in the recent body of work named "compressed sensing". The deterministic signal model of interest here is

$$P_r(j) = \sum_t d_{rt}^{-\alpha} P_t(i) \tag{19}$$

where the received power is modelled as a function of the transmitted power and the distance between the transmitter and the receiver. We then assume that we observe *samples* of the radio field and investigate to what extent we can reconstruct the field from these samples. In particular we will be interested in the number and the nature of these samples, how many samples do we need for proper reconstruction and under which conditions can these samples be captured.

We will be interested in the identification of any radio transmitters operating in the region. In particular, a radio transmitter located in gridpoint *i* will and emitting power with $P_t(i)$ will cause a radio field that can be sensed with appropriately tuned radio receivers, the sensors. A radio sensor located in another gridpoint, *j*, measures the radio field in terms of received power, $P_r(j)$. Figure 4 illustrates example positions of transmitters and sensors in the grid. We rewrite (19) as

$$P_r(j) = \Psi_{ij} P_t(i) \qquad \text{where} \quad \Psi_{ij} = d^{-\alpha} \tag{20}$$

denotes how the received power $P_r(j)$ in gridpoint *j* is related to the transmitted power $P_t(i)$ in gridpoint *i,* and where α is, as before, a known pathloss exponent and *d* is the (Euclidean) distance between gridpoint *i* and gridpoint *j*. Allowing each gridpoint to contain both a transmitter and a sensor, the power transfer (Eq. 3.25 of [3]) turns into a matrix/vector equation

$$\mathbf{p}_r = \mathbf{\Psi} \mathbf{p}_t, \tag{21}$$

where $\mathbf{p}_t$ is the *Nx1* vector of transmitted powers, $\mathbf{p}_r$ is the *Nx1* vector of received powers and $\mathbf{\Psi}$ is the channel power transfer matrix

$$\mathbf{\Psi} = \left[\Psi_{ij}\right]. \tag{22}$$

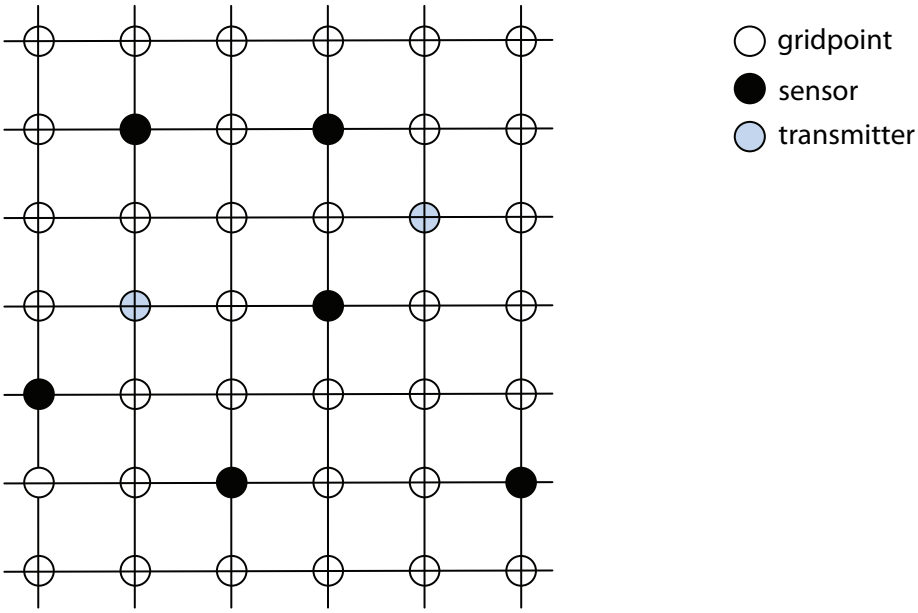

Figure 4: Example of a grid along with transmitter and sensor positions.

In the physical reality of our above problem, typically, there are only few transmitters actively transmitting (relative to the grid size *N*), that is, only few entries of the power vector $\mathbf{p}_t$ are non-zero. This motivates the exploitation and investigation of the mentioned recent advances in compressed sensing.

4.1.1.1.6 *Gaussian-based Kriging interpolation*

The Gaussian-based Kriging interpolation, proposed for the in-band femtocell optimization scenario, uses the derivation framework described in the Appendix of deliverable D4.2 [3]. In short, it calculates the conditional mean of the radio field at prediction/interpolation samples ($\mathbf{y}_1$) conditioned on the known measured samples ($\mathbf{y}$), using the following:

$$\mathbb{E}_{|\mathbf{y}}(\mathbf{y}_1) = \mathbb{E}[p(\mathbf{y}_1|\mathbf{y})] = \mathbb{E}[\iint_{\phi,\tau} \left[\int_{\theta} p(\mathbf{y}_1|\phi,\theta,\tau,\mathbf{y}) p(\theta|\phi,\tau,\mathbf{y})\right] p(\phi,\tau|\mathbf{y})] \quad (23)$$

where $\mathbb{E}$ stands for statistical expectation, $p(.)$ is the probability density function, ϕ is the shadowing correlation distance, $1/\theta$ is the shadowing variance and τ is the additive white gaussian noise variance multiplied by θ.

As indicated in D4.2 ([3]), the inner integral is a Student distribution with the following mean:

$$\mathbb{E}_{|\phi,\tau,\mathbf{y}}(\mathbf{y}_1) = \left(X_1 - S_{1y}S_{yy}^{-1}X\right)H''^{-1}H'b' + \left(S_{1y}S_{yy}^{-1} + \left(X_1 - S_{1y}S_{yy}^{-1}X\right)H''^{-1}X^T S_{yy}^{-1}\right)\mathbf{y} \quad (24)$$

and the posterior distribution of the model parameters ϕ and τ, $p(\phi,\tau|\mathbf{y})$, is given by the following expression:

$$p(\phi,\tau|\mathbf{y}) = p(\phi,\tau)K_1(\mathbf{y})(2\pi)^{-\frac{N}{2}}|S^{-1}|^{\frac{1}{2}}|H|^{\frac{1}{2}}\left(\frac{\nu q}{2}\right)^{\frac{\nu}{2}}\Gamma^{-1}(\frac{\nu}{2})\Gamma(\frac{\nu+N}{2})|H+X^TS^{-1}X|^{-\frac{1}{2}}\times$$

$$\left(\frac{\nu q+\mathbf{y}^TS^{-1}(\mathbf{y}-X\mathbf{b})-(\mathbf{b}^TH+\mathbf{y}^TS^{-1}X)(H+X^TS^{-1}X)^{-1}X^TS^{-1}(\mathbf{y}-X\mathbf{b})}{2}\right)^{-\frac{\nu+N}{2}} \quad (25)$$

where the reader is referred to D4.2 ([3]) for notational explanation.

4.1.2 Algorithmic complexity

This subsection focuses on the evaluation of the time complexity of the spatial interpolation methods briefly explained in subsection 4.1.1.1. The methods' computational complexity is evaluated in terms of their underlying tasks, as well as their overall computational complexity.

4.1.2.1 General matrix operations complexity

This part focuses on the complexity of the basic operations required by the spatial interpolation RIF estimation methods. Generally, the calculations in each of the methods include matrix operations such as matrix multiplication and inverse matrices calculation. Namely, the standard matrix multiplication method of non-square matrices A_{lxm} and B_{mxn} yields a computational complexity of $O(lmn)$, while the standard matrix inverse calculation (Gauss–Jordan elimination [10], Gaussian elimination [11], LU decomposition [12] etc.) of a square matrix C_{nxn} implies a computational complexity of $O(n^3)$. There exist even more efficient matrix multiplication and inverse calculation methods which result in time complexities of $O(n^{2.807})$ [13] and $O(n^{2.376})$ [14]. The parallel processing can yield even lower complexity of the matrix operations [21]. The matrix operations efficiency is closely related to the overall operation of all the RIF estimation techniques in the focus of this Section.

4.1.2.2 Complexity of the classic IDW method

Since the classic IDW method (subsection 4.1.1.1.1) uses all available N data points to make a single estimation, in the case of M interpolation points the time complexity of the method is linearly dependent on these both factors, resulting in a asymptotic time complexity of $O(MN)$ (see Table 6).

Table 6: Classic IDW complexity.

Classic IDW complexity				
#	**Processing step**	**Methods used**	**Asymptotic complexity** $M,N->\infty, M>N$	**No. of executions**
1	**Interpolation**	/	O(*N*)	*M,* for each interpolation point
2	**Overall**	/	***O*(*MN*)**	1

4.1.2.3 Complexity of the Modified Shepard's Method

Compared to the classic IDW method, the MSM (subsection 4.1.1.1.2) does not take all *N* observation points into account when performing the estimation about a single interpolation point, and it considers only the nearest N_w neighbors. The parameters N_w and N_q are generally smaller than the number of observation points, significantly in the asymptotic case ($N,M->\infty$).

However, the dominant influence on the time complexity of the MSM method, depending on the number of observation points *N* is the used Nearest Neighbor Search algorithm. If an inefficient algorithm is used the MSM method can even be more complex than the classic IDW method. This is due to the requirements that the NN search should be performed for each observation and interpolation point.

The WLS fitting (Eq. (16)) has an asymptotic complexity of $O(N_q p^2)$. This processing step takes three matrix multiplications tasks and one matrix inverse task. The highest and dominant complexity comes from the product of the first three matrices $X_k^T W_k X_k$, since N_q is usually higher than *p*. Namely, the MSM method typically uses a linear ($p=2$) or a quadratic model ($p=5$) for the nodal functions. It is important to note that the weighting matrix has a non-significant influence on the time complexity as it is a diagonal matrix and the multiplication can be performed in parallel. Therefore, the WLS problem has the same complexity as the OLS problem.

Table 7 presents the complexity of each of the processing steps and the overall complexity of the MSM method when the used NN search algorithm is KD-tree [22], and the standard methods for matrix operations [10][11][12]. The overall asymptotic time complexity in this case is $O(M\log N)$, which is significantly lower than the classic IDW method.

Table 7: Modified Shepard's Method complexity.

Modified Shepard's Method complexity				
#	**Processing step**	**Methods used**	**Asymptotic complexity** $M,N \to \infty, M>N$	**No. of executions**
1	**Nearest Neighbor Search**	KD-tree	$O(\log N)$	N, for each observation point
2	**WLS nodal function fitting**	Standard matrix multiplication and inverse calculation	$O(2N_q p^2+p^3+N_q p) \approx O(N_q p^2)$	N, for each observation point
3	**Nearest Neighbor Search**	KD-tree	$O(\log N)$	M, for each interpolation point
4	**Interpolation**	/	$O(N_w)$	M, for each interpolation point
5	**Overall**	/	$O[M(N_w+\log N) + N(N_q p^2+\log N)] \approx$ $\boldsymbol{O(M \log N)}$	1

Table 8: Gradient plus Inverse Distance Squared complexity.

GIDS complexity				
#	**Processing step**	**Methods used**	**Asymptotic complexity** $M,N \to \infty, M>N$	**No. of executions**
1	**Nearest Neighbor Search**	KD-tree	$O(\log N)$	M, for each interpolation point
2	**OLS function fitting**	Standard matrix multiplication and inverse calculation	$O(2N_w p^2+p^3+N_w p) \approx O(N_w p^2)$	M, for each interpolation point

3	**Interpolation**	/	$O(N_w)$	*M,* for each interpolation point
4	**Overall**	/	$O[M(N_w+N_w p^2+\log N)]$ ≈ ***O*(*M*log*N*)**	1

4.1.2.4 Complexity of the Gradient plus Inverse Distance Squared Method

The GIDS method, briefly explained in subsection 4.1.1.1.3, also uses only N_w neighbor observations when calculating a single estimation. Similarly as the MSM method, the source of the highest complexity of the GIDS interpolation is the NN search technique. Additionally, here the OLS fitting is performed for each of the interpolation points *M*, which is a main disadvantage compared to the MSM method because the number of interpolation points *M* is generally larger than the number of observation points.

Table 8 elaborates on the complexity of each processing task and the overall complexity of the GIDS interpolation. KD-tree is again used a as an NN search algorithm along with the standard methods for matrix operations. It is important to note that although the GIDS method calculates an OLS fit for each interpolation point, the overall asymptotic time complexity is the same as the MSM method *O*(*M*log*N*).

4.1.2.5 Complexity of the Ordinary Kriging method

The Ordinary Kriging method is the optimal interpolation method since it minimizes the mean square error (subsection 4.1.1.1.4). However, the method requires the knowledge of the semivariogram model and its parameters. The estimation of the semivariogram parameters requires a significant number of input observations for a reliable estimation, which increases the computational complexity of the Kriging method in overall. The computational complexity of the semivariogram fitting, despite the number of observations *N*, highly depends on the selected model and the number of parameters *p*, as well as the fitting technique, i.e. Least Squares, Maximum Likelihood etc. Another calculation which is a preprocessing step of the Kriging method and requires a high computational complexity is the semivariance matrix inversion. These steps can be avoided if the semivariogram and the inverse semivariance matrices are already known for the inspected radio environment.

However, the constrained linear optimization problem, i.e. the Kriging weights estimation is the crucial processing step significantly affecting the time complexity of the Kriging method in overall. It requires the multiplication of high dimension matrices (when the input sample size *N* is high) and this processing is performed for each interpolation point.

The complexity analyses presented on Table 9 focuses on the Least Squares semivariogram fitting and standard methods for matrix multiplication and inverse calculation. Even with the pre-

processing steps related to the semivariance, the Kriging method asymptotic time complexity is governed by the constrained linear optimization, which is equal to $O(MN^2)$. There is a lot of work on reducing the computational complexity of the Kriging based methods [23][24][25][26], most of it being focused on reducing the complexity through sparsification of the (co)variance matrix.

Table 9: Ordinary Kriging complexity.

Ordinary Kriging complexity				
#	**Processing step**	**Methods used**	**Asymptotic complexity** $M,N -> \infty, M>N$	**No. of executions**
1	**Semivariogram fitting**	Least Squares	$O(2Np^2+p^3+Np) \approx$ $O(Np^2)$	1, before Kriging interpolation; 0, if known and not empirically evaluated
2	**Semivariance matrix inverse**	Standard matrix inverse calculation	$O(N^3)$	1, before Kriging interpolation; 0, if known and not empirically evaluated
3	**Constrained linear optimization (kriging weights calculation)**	Standard matrix multiplication	$O(N^2)$	*M,* for each interpolation point
4	**Interpolation**	/	$O(N)$	*M,* for each interpolation point
5	**Overall**	/	$O[M(N^2+N)+N^3+Np^2] \approx$ $\boldsymbol{O(MN^2)}$	1

4.1.2.6 Complexity of the compressed sensing method

4.1.2.6.1 *Processing cost in the sensors (compressive sampling)*

In the "compressed sensing" approach, the terminology "samples" ("measurements") of ***y*** relates to the *linear projections* of a vector ***y*** onto base vectors $\mathbf{\Phi}_m$. Collection of *M* such projections (introducing the matrix whose *M* rows are the *M* projection vectors $\mathbf{\Phi}_m$) then produces the measurement vector $\mathbf{z} = \mathbf{\Phi y}$. Applied to our signal model, the observed samples are then

$$\mathbf{z} = \mathbf{\Phi \Psi p}_t, \tag{26}$$

where $\mathbf{\Phi}$ is an *MxN* measurement matrix. As a final step we specify the measurement matrix $\mathbf{\Phi}$ in accordance with two realistic categories:

Static sensors. A first choice of the measurement matrix is based on the physical fact that sensors sample the radio field in one particular location. The measurement matrix $\mathbf{\Phi}$ associated with this choice has each of its rows contain one single non-zero entry in the particular column whose index is associated with the gridpoint where this sensor is located. All other entries are zero. In other words, the projection vectors are trivial unit vectors.

Static sensors may be integrated with base stations nodes of mobile networks or, alternatively, consist of dedicated sensor nodes that are of a wall-mounted or lamp-post-mounted type.

Mobile sensors. Exploiting our assumption that the unknown transmitters do not move and do not change their powers in time, hence the radio field is static in time, we will allow sensors to sample the radio field in different locations *at different time instants*. While moving across the region of interest, mobile sensors can sense the field in various locations and then apply the projection of these measurements onto a compressing vector, a row in $\mathbf{\Phi}$, whose entries are zero except for those locations at which the mobile sensor has been sensing the field. These locations are compressed with a short signature sequence. We must make a careful distinction between a *measurement* and a *sample*. One mobile sensor produces a series of field *measurements* in various locations as it moves from one point in the grid to another. By projecting these measurements onto a row in $\mathbf{\Phi}$, it then generates one *sample*, an entry of the observation vector ***z***.

Mobile sensors are typically integrated with mobile phones or other consumer devices that people carry with them and move around naturally in spatial patterns that are typiclly beyond the direct control of the network operator. The random nature of these patterns fit well with the principles of compressed sensing where random measurements appear to generate high performance reconstruction.

In both scenarios a number of aspects relate to the implementation complexity (both the memory requirements and the processor in the sensor).

$$z_i = \mathbf{\varphi}_i \mathbf{\Psi p}_t, \tag{27}$$

where z_i is the scalar sample generated by the *i*th sensor and $\boldsymbol{\varphi}_i$ is the projection vector employed by the *i*th sensor to compress the measurements. The matrix-vector multplication $\boldsymbol{\Psi}\mathbf{p}_t$ is done impliciotly by the radio channel - the result of this multiplication is the vector of received powers measurable in each point of the grid. The very process of sensing/measuring and storing the received power in a certain frequency band by carried out by the sensing device thus yields entries of this vector.

Regardless of the kind of sensor (static or mobile) the process of compressing is then a very low-complexity process. For *static sensors* the projection vector $\boldsymbol{\varphi}_i$ consists of a vector with only a single unit entry and is hence transparent in the processing. For mobile sensors, the compression essentially consists of adding the various power measurements performed at various locations into one scalar number. Along with this scalar sample, the mobile sensor also will have to store the locations of each of the measurements.

4.1.2.6.2 *Signalling cost*

Signalling load varies between the two scenarios (static and mobile sensors). Table 10 illustrates the payload of the reporting of one sample. Whereas the locations of the static sensors can safely be assumed known by the network (and hence the matrix $\boldsymbol{\Phi}$), mobile sensors have to inform the network about the actual locations reflected by the scalar measurement value. This is a significant difference between the two scenarios in terms of cost. The signalling overhead in the network associated with the sensing is a key concern in the literature and in any practical realization of this kind of distributed sensing concepts. It is thus critical for this mode of operation that efficient compression techniques for the reporting of the locations are developed.

As a means of increasing the signalling efficiency, many samples may also be *collected* in the sensor during a period of time and reported to the network in batches. This is likely to reduce signalling overhead due to coding gains but may not be feasible when samples need to be rapidly avalable at the network.

Table 10: Signalling content per reporting instant when
F frequency bands are scanned and *M* locations are being used by the mobile sensors.

Static sensors	**Mobile sensors**
F scalar sample values per report	*F* scalar sample values per report *M* positions per report (xy-coordinates)

4.1.2.6.3 *Processing cost in the network (reconstruction)*

The network receives samples from a group of spatiallly distributed sensors and collects these samples in the vector $\mathbf{z}$.

In the network, the vector $\mathbf{p}_t$ is then reconstructed from the measurements $\mathbf{z}$. The theory of compressed sensing has at the core of its proven results that solutions to

$$\min\left\|\mathbf{p}_t\right\|_1 \text{ subject to } \mathbf{z}=\mathbf{\Phi\Psi p}_t, \tag{28}$$

(where $\left\|\mathbf{x}\right\|_1=\sum_i\left|x_i\right|$), under specific conditions for the matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, are the sparsest existing solutions, and hence can be used to perfectly reconstruct $\mathbf{p}_t$ from the limited number of measurements in ***z,*** in a reasonably low-complexity fashion. One such solution (known as *basis pursuit*) is based on recasting (28) as a linear program and then using the standard primal-dual algorithm for linear programming (whose solution procedure is based on the classical Newton method). Hence the problem can be solved in polynomial time. In fact, this complexity argument in relation to finding sparse solutions is one of the key results of teh compressive sensing theory.

In absolute terms, the computation of the solution (28) may still need considerable processing power. In a system perspective, computational complexity is concentrated at the network side where typically the resources are most abundent, while the sensors are released from most of the algorithmic complicity.

4.1.2.7 Complexity of the Gaussian-based Kriging interpolation

Equations (24) and (25) which are necessary for computing the prediction of (23), are time consuming because of the inversion of the matrix **S**. Furthermore, assuming the priors for ϕ and τ which are discrete random variables, this inversion needs to be done for each pair of possible ϕ and τ values, $(\phi_k,\tau_l)\ \forall k,l$. The computation of matrix **S** from scratch is not efficient, and Gaussian elimination performs betters when **X** has few regressors. A more viable option, the one we have used in this study, is the computation of the eigen-decomposition of **S** whose practical computational complexity is $O(n^3)$. This saves computational time because it has to be done only for one value of τ. Denoting $\mathbf{V_{k,l}}$ and $\mathbf{D_{k,l}}$ respectively the eigenvectors and eigenvalues for the couple $\mathbf{S}(\phi_k,\tau_l)$, we can update the eigen-decomposition as follows:

$$\mathbf{V}_{k,l,2}=\mathbf{V}_{k,l,1}$$
$$\mathbf{D}_{k,l,2}=\mathbf{D}_{k,l,1}+(\tau_{l,2}-\tau_{l,1})\mathbf{I}$$

Note that in theory, it is possible to reduce the computational complexity of the matrix inversion to $O(n^{2.807})$ [13] and $O(n^{2.376})$ [14], i.e. to the same order of complexity as matrix multiplication. However, these algorithms provide ε-approximate solutions for some predetermined accuracy levels [15] and have been reported as not being practically useful for double precision floating point computations on matrices with dimensions higher than 10^3.

4.1.3 Platform requirements

The spatial interpolation techniques use the input from a limited number of sensors to perform the estimation in the unobserved locations of the RIF. The different spatial interpolation techniques have different behaviors with respect to the number of MCDs, the MCDs geometry etc. This subsection targets the platform requirements of the spatial interpolation techniques in terms of the measurement data requirements, the number and the placement of the sensors.

4.1.3.1 Observation data requirements

In general, the data required by the spatial interpolation techniques for their proper operation should be prepared in the appropriate manner:

- **the observation data should be provided in appropriate format – (*x, y, P, t*)**. Each MCD should provide the received power readout (RSS) associated with the location (*x*, *y*) and the time *t* of gathering.
- **the MCDs should be mutually calibrated**. This is a necessary requirement since any measurement data inconsistencies would result in a biased interpolation error. The calibration process is especially a challenge when the data is collected from a heterogeneous set of MCDs. Even when the MCDs are calibrated, the different sensitivity (noise) levels result in higher interpolation errors when the lower bound of the dynamic range of the field is near the noise floor of the devices.
- **the MCDs should be mutually synchronized.** For the most accurate RIF estimation the MCDs need to be mutually synchronized. This can be a strict requirement since the dynamism of the radio environment field in some scenarios may result in very inconsistent RSS readouts. In these cases, time-domain RSS averaging might help coping with the dynamism.

In terms of the observation data requirements, the different techniques might experience significantly different behavior when dealing with measurements inconsistencies of the aforementioned types. While the IDW based approaches (subsections 4.1.1.1.1, 4.1.1.1.2, 4.1.1.1.3) are more robust to "outliers", the Kriging method (subsection 4.1.1.1.4), especially the semivariogram fitting might be very sensitive to erroneous RSS readouts. This can lead to incorrect estimation of the semivariogram parameters, and the semivariogram inaccuracy propagates into high Kriging interpolation errors.

4.1.3.2 Requirements on MCDs geometry and number

The different spatial interpolation approaches experience different behaviors respective to the MCDs placement, i.e. whether the MCDs are distributed in a gridded manner or scattered randomly in the area of interest. The number of spatial measurements (MCDs number) is also a significant factor in the spatial interpolation performances. The results presented in deliverable D4.2 [3], subsection 3.3, implicate that:

- The spatial interpolation methods in focus generally tend to perform better when the MCDs are places in a regular grid for both, the indoor and outdoor scenarios.

- Out of all three inspected methods, the Ordinary Kriging, the IDW MSM method and the GIDS method, the Kriging method generally provides the best results for all scenarios when the number of observations is larger.
- When dealing with limited number of spatial observations the IDW MSM interpolation method outperforms the Kriging interpolation method. This is due to the fact that the Kriging based interpolations require the correlation structure of the data has to be either known or estimated. The semivariogram estimation is difficult to do reliably if only a limited number of observations is available, and can lead into unreliable results and lack of robustness.

4.1.3.3 Example: Gaussian-based Kriging Interpolation for in-band femtocell

The platform requirements of the Gaussian-based Kriging interpolation method used for the in-band femtocell optimization scenario basically consist of the signaling and storage cost of constructing, maintaining, updating and using REMs. The architecture work outlined in deliverable D2.4 [1] for the in-band femtocell optimization scenario, where we have given a detailed description of the measurement reporting of the MCDs to the REM Acquisition on the air interface, can be used to calculate this cost.

Calculation of the signaling cost of measurement reporting involves the message size for each *MeasurementReport* message and the required frequency of transmitting the *MeasurementReport* messages. The former one is given by the detailed architectural work presented in deliverable D2.2 ([4]). It is a function of the number of reported neighboring cells *NCells*, which is around 700 bits/message. The second one depends on requirements on REM quality, which in turn depends on factors like how many measurement samples are needed per unit area and per unit time, and what is the area of the intended geographical zone. The required number of measurement samples per unit area depends on: 1) the spatial characteristics of the environment (notably the *correlation distance*, which is defined as the minimum distance above which the measured signals are considered as uncorrelated), 2) the spatial signal model and 3) spatial interpolation technique used to predict the "unmeasured" samples. What is meant by the correlation distance is the medium-scale correlation distance of the signal (related to the shadowing component but not to the fast fading). As for the required number of measurements per unit time, it depends on the coherence time of the radio environment, which means the time interval above which the signal samples are uncorrelated.

Below, we will provide typical values for these parameters and come up with a rough signaling overhead analysis:

According to [16], 50% correlation distance for urban micro environments is 5m. Considering the proximity between the respective scales of microcells and femtocells (transmission power, coverage areas, deployment etc.) and without loss of generality, we adopt this value as a representative value of the correlation distance of the femtocell environment.

Previous studies have demonstrated that a uniform spatial sampling distance up to 4-5 times of the correlation distance results in sufficiently high-quality REMs [17][18]. On the other hand, it has

been shown in [19] that using 25% of the uniformly distributed samples together with a Gaussian field assumption and the Kriging interpolation technique yields good results. Combining these, we can obtain the required number of measurement samples per unit area as $\frac{4 \text{ samples}}{(5 \text{ m})(5 \text{ m})} * (\%25) = \frac{1 \text{ sample}}{25 \text{ m}^2}$. The required total number of measurement samples for an area *A* (in m^2) is then equal to $\frac{A}{25}$.

The coherence time can be calculated as follows [20]:

$$T_c = \frac{0.423\,\lambda}{\mathrm{v}} = \frac{0.423}{v}\frac{c}{f} = \frac{1.269\times 10^8 \frac{m}{s}}{vf}$$

where *v* is the average speed of the objects in the radio environment (in m/s), *f* is the carrier frequency of the radio signals (in Hz) and *c*=3 x 10^8 m/s is the speed of light. The LTE frequency bands in Europe will be around three carrier frequencies: 2.6 GHz, 800 MHz and 1800 MHz. Then the coherence time T_c can be expressed as a function of the three LTE carrier frequencies and the average speed *v*:

$$T_c^{2.6GHz} = \frac{1.269\times 10^8}{vf} = \frac{0.05}{v}$$
$$T_c^{800MHz} = \frac{1.269\times 10^8}{vf} = \frac{0.16}{v}$$
$$T_c^{1800MHz} = \frac{1.269\times 10^8}{vf} = \frac{0.07}{v}$$

For example, considering that the average speed of objects in a static indoor environment is of the order of centimeters per second, $T_c^{2.6GHz} \cong 5\,s$, $T_c^{800MHz} \cong 16\,s$ and $T_c^{1800MHz} \cong 7\,s$. In other words, measurements that are separated in time by more than 5-15 s are uncorrelated; meaning that the REM has to be updated each 5-15 s.

Combining this with the total number of measurements required for a REM that covers an area of *A* m^2, which is equal to $\frac{A}{25}$, the total number of required measurements per second varies between $\frac{A}{25}\times\frac{1}{15} = \frac{A}{375}$ and $\frac{A}{25}\times\frac{1}{5} = \frac{A}{125}$.

Finally, knowing that each measurement sample takes 81+77**NCells* bits, the total number of bits/s required to build the REM varies between $s_{low} = \frac{A(81+77*NCells)}{375}$ and $s_{high} = \frac{A(81+77*NCells)}{125}$.

Tables 11 and 12 depict different cases where REM coverage zone and the number of reported neighboring cells take different values, for lower and higher update rates respectively.

Table 11: Measurement reporting message size for lower values of update rate.

s_{low} (bps)	NCells=0	NCells=1	NCells=2	NCells=3	NCells=4	NCells=5	NCells=6	NCells=7	NCells=8
A=25m^2	6	11	16	21	26	32	37	42	47
A=50m^2	11	22	32	42	52	63	73	83	93
A=75m^2	17	32	47	63	78	94	109	124	140
A=100m^2	22	43	63	84	104	125	145	166	186
A=125m^2	27	53	79	104	130	156	181	207	233
A=150m^2	33	64	94	125	156	187	218	248	279
A=175m^2	38	74	110	146	182	218	254	290	326
A=200m^2	44	85	126	167	208	249	290	331	372
A=225m^2	49	95	141	188	234	280	326	372	419
A=250m^2	54	106	157	208	260	311	362	414	465

Table 12: Measurement reporting message size for higher values of update rate.

S_{high} (bps)	NCells=0	NCells=1	NCells=2	NCells=3	NCells=4	NCells=5	NCells=6	NCells=7	NCells=8
A=25m2	*17*	*32*	*47*	*63*	*78*	*94*	*109*	*124*	*140*
A=50m2	*33*	*64*	*94*	*125*	*156*	*187*	*218*	*248*	*279*
A=75m2	*49*	*95*	*141*	*188*	*234*	*280*	*326*	*372*	*419*
A=100m2	*65*	*127*	*188*	*250*	*312*	*373*	*435*	*496*	*558*
A=125m2	*81*	*158*	*235*	*312*	*389*	*466*	*543*	*620*	*697*
A=150m2	*98*	*190*	*282*	*375*	*467*	*560*	*652*	*744*	*837*
A=175m2	*114*	*222*	*329*	*437*	*545*	*653*	*761*	*868*	*976*
A=200m2	*130*	*253*	*376*	*500*	*623*	*746*	*869*	*992*	*1116*
A=225m2	*146*	*285*	*423*	*562*	*701*	*839*	*978*	*1116*	*1255*
A=250m2	*162*	*316*	*470*	*624*	*778*	*932*	*1086*	*1240*	*1394*

As a typical femtocell example, if we consider a residential and/or professional environment that covers up to 3 femtocells (*NCells*=2) with a coverage zone of around 200 m^2, the number of bits required for measurement reporting for REM construction purposes will be between 126 and 376 bits per second.

4.2 Fusion techniques for localization

What we call "fusion methods" here include all these techniques that combine the different measurements/estimates from all sensors in order to localize the active sources. Several fusion techniques have been introduced in the literature; these techniques depend on what signal parameters are measured at the sensors. As already mentioned, the most common signal parameters are the time of arrival, time difference of arrival, angle of arrival and receive signal strength. Two main categories of such techniques are the probabilistic and non-probabilistic. This categorization is based on the existence or proper use of soft information from the sensors. A small overview of basic fusion techniques was given in D4.2 ([3]) for localizing *single* sources for both categories. The algorithmic as well as the platform complexity are described herein.

4.2.1 Algorithmic complexity

4.2.1.1 Non probabilistic LS based fusion

The generic LS solution of the corresponding over-determined system of equations by the use of the pseudo-inverse matrix is given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{b}$$

The construction of Matrix H depends on the used localization method (see [3]). The dimension of H is $N_s - 1 \times p$, where $p = 2$ for TOA, AOA and RSS while $p = 3$ for TDOA. This results in asymptotic complexity of $\sim O(pN_s^2)$ for matrix multiplications and $O(p^3)$ for the inverse calculation.

4.2.1.2 Probabilistic *ML* fusion

In the general case, we can model the measurements of the sensors as

$$\boldsymbol{\beta} = \theta(\mathbf{x}) + \boldsymbol{\varepsilon}$$

where $\theta(\cdot)$ is a non-linear function of the position of the source, described at D4.2 ([3]) for all four types of measurements, and $\boldsymbol{\varepsilon}$ is the additive noise of the measurements assumed to be a zero mean Gaussian vector with a covariance matrix $\mathbf{S} = diag\{\sigma_1^2, ..., \sigma_N^2\}$. The *ML* estimator of the location of the transmitter is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left([\theta(\mathbf{x}) - \boldsymbol{\beta}]^T \mathbf{S}^{-1} [\theta(\mathbf{x}) - \boldsymbol{\beta}] \right)$$
$$= \arg\min_{\mathbf{x}} \left(\sum_{i=1}^{N} \frac{\left(\theta(\mathbf{x}) - \boldsymbol{\beta}\right)^2}{\sigma_i^2} \right)$$

This non-convex optimization problem has several drawbacks. One is that it may contain several local minima and saddle points, therefore by local optimization methods, the final solution highly depends on the starting point provided and there is no guarantee that the final solution converges to the global minimum. Another is that the search for the global minimum is very difficult and time-consuming.

In order to tackle the problem with cost function local minima, a positioning algorithm based on projection on convex sets (POCS) [35] was proposed some years ago by Blatt and Hero in [36][37]. The algorithm is both robust to local minima in the objective function, of low complexity, and has a computation which is possible to distribute over the nodes in, e.g., a sensor network. In [36],[37], distance estimates based on RSS measurements was considered, but the algorithm is applicable in any positioning application where sensor-source distance estimates are available. This method, referred to as circular POCS, was shown to perform well in the case of a source node located inside the sensor convex hull, defined by the outer perimeter nodes in the network, but experiences problems in locating sources outside of this convex hull, and requires knowledge of channel parameters such as path-loss and transmission power. Another way to circumvent the non-convexity of conventional *ML* estimator is the semi-definite programing (SDP) relaxation technique. The basic idea behind the technique is to convert the non-convex quadratic distance constraints into linear constraints by introducing a relaxation to remove the quadratic term in the formulation. Similar relaxations were developed for solving other distance geometry problems, see, e.g., Alfakih et al. [38] and Laurent [39].

4.2.2 Platform requirements

In order to determine the required number of sensors for target localization accuracy we make the following assumptions:

- All sensors estimate the ranging parameters of interest with performance equal to that of the CRLB.
- There is always a specific coverage area around an emitter, where only sensors within this area can estimate the ranging parameters of interest.
- A rectangular grid for the positions of the sensor deployment network is always assumed

The coverage area is determined by a range parameter (measured in dB) that models the maximum propagation loss that allows ranging parameter estimation by the sensors.

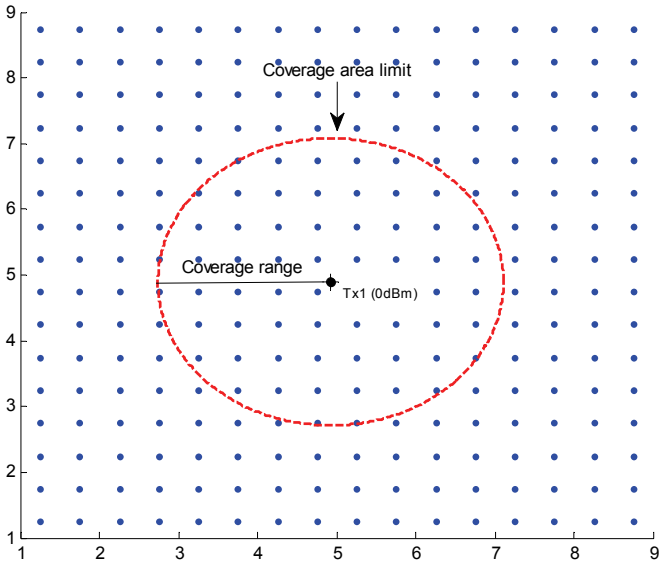

Figure 5: Sensor density computation framework.

Beginning with a very dense sensor network, we gradually decrease the density and measure the lower bound on localization accuracy until no sensor lies within the coverage area. Figure 5 displays a snapshot of this procedure for a specific network density. The following figures depict the results for RSS based localization for various scenario settings. The coverage area the path-loss exponent (α) and the shadowing variance (σ_f^2) choices cover indoor as well as outdoor scenarios.

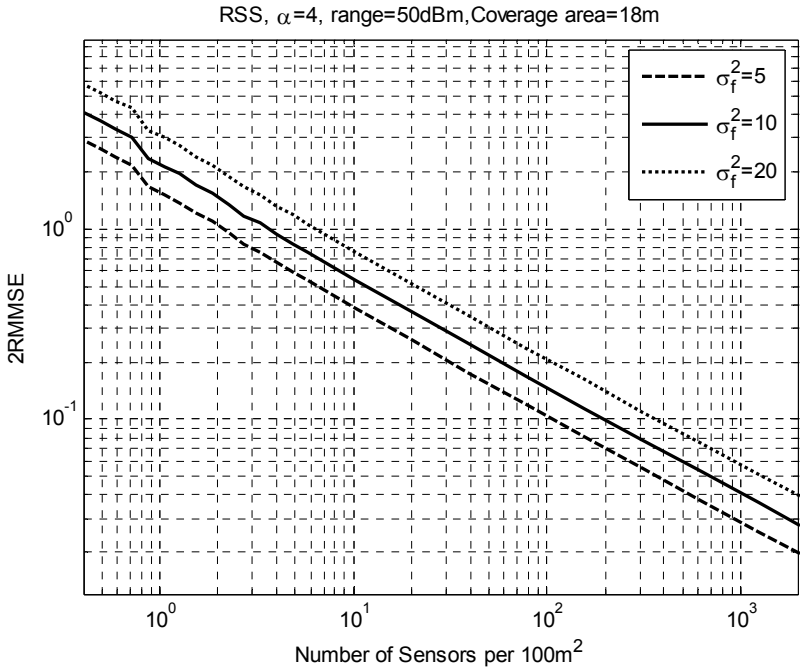

Figure 6: Localization accurancy vs. #Sensors (1).

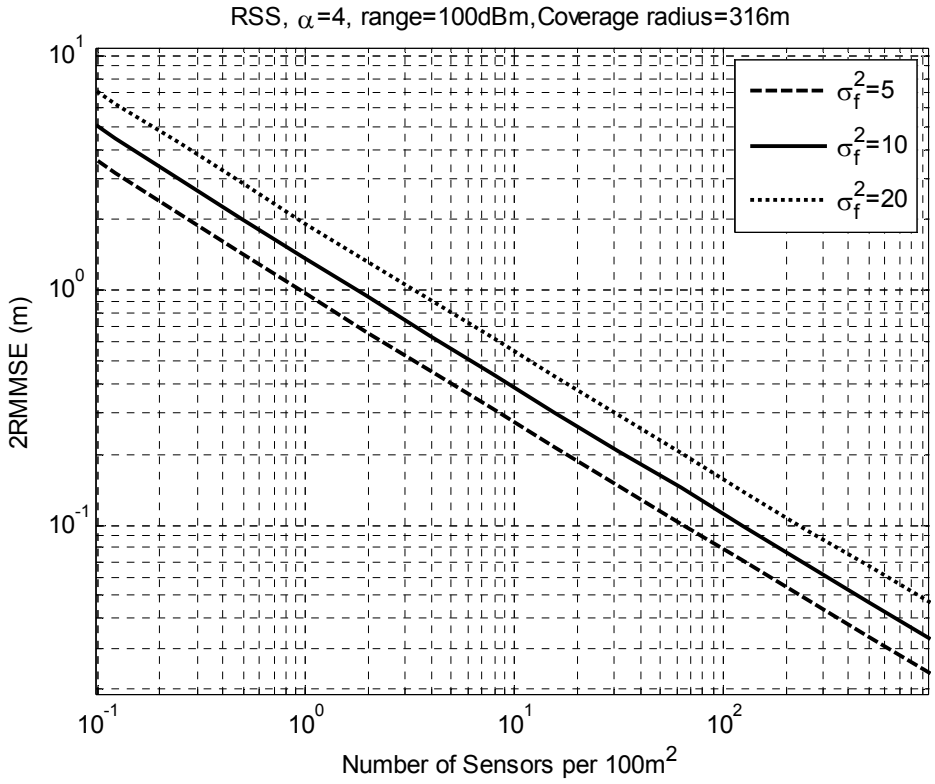

Figure 7: Localization accurancy vs. #Sensors (2).

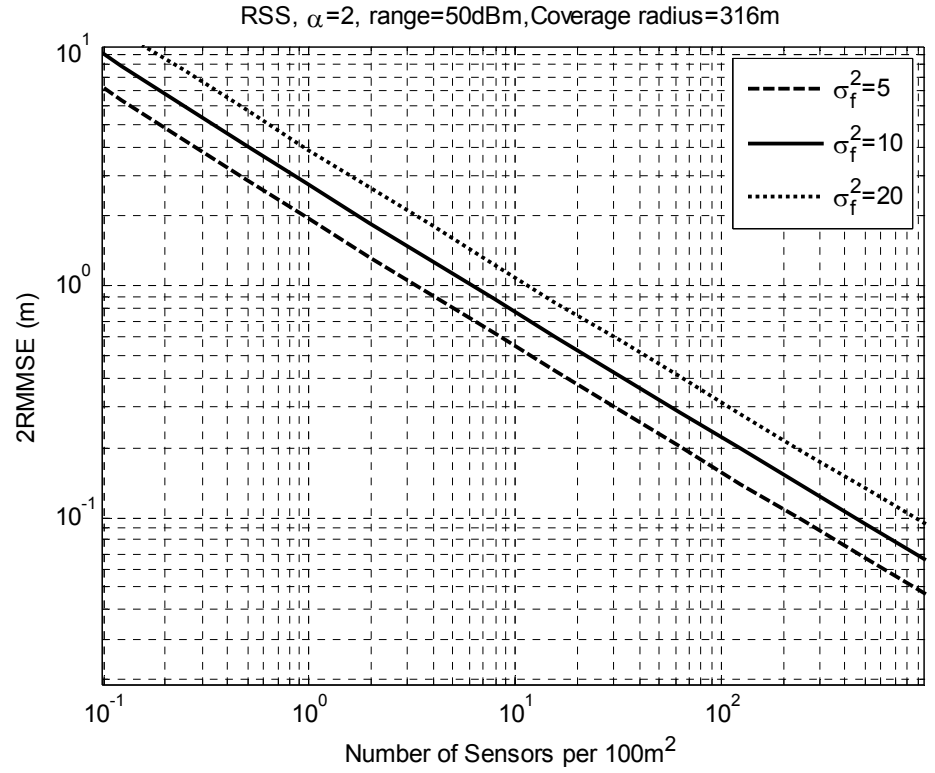

Figure 8: Localization accurancy vs. #Sensors (3).

4.3 Noise-Robust source detection

In D4.2 ([3]), a new framework for Constant-False-Alarm-Rate (CFAR) feature-based detection was proposed. The focus was on (but are conceptually not limited to) spatial signal features that arise due to the reception of the signal at different points in space. These features allow detection (*first*, and that is the novelty) as well as subsequent localization of unknown sources based on an estimation pre-processor. The proposed family of statistical tests is based on the appropriate processing of the received samples in order to create successive and statistically independent estimates of signal features that have meaning if and only if a signal is present. By first quantizing the feature range and subsequently estimating the most probable value for the feature under consideration, a histogram is created which signifies the relative probabilities for the different (discrete) sub-regions on the value of the feature. Based on an analytical description of this histogram, a test is created on the existence or not of the signal, as well as the value of the feature (as an immediate fridge benefit). The whole procedure takes therefore two steps: the first where successive estimates of a feature are made *as if the signal exists* and a second where these estimates are employed in order to decide on the binary hypothesis of the signal existence, as well as the feature value.

4.3.1 Algorithmic complexity

The following test was proposed for the joint detection/estimation problem: Define

$$T = \max\left(\text{hist}(\mathbf{y})\right) \lessgtr \lambda \tag{29}$$

the test for the detection task, and $\hat{c} = \text{index}(T)$ for the estimation, where $\mathbf{y} = [y_1, y_2, ..., y_{N_b}]$ is the total vector of observations, $\text{hist}(\cdot)$ is the histogram operator, namely a counter for the number of occurrences for each of the N_L events, $\max(\cdot)$ is an operator that provides the value of the largest number of the input vector and $\text{index}(\cdot)$ is its index. Other tests could also be devised with possibly better performance, but the currently proposed test is simple and performs almost optimally in the parameter regions of interest.

4.3.1.1 Application to cyclic prefix based OFDM detection

Let N_{cp} be the cyclic prefix length, N_{FFT} the FFT length, and $N_{S-total}$ the total number of OFDM symbols received. We assume the signal bandwidth and the above parameters known. The needed processing is depicted in Figure 9. Since the starting time of a symbol is unknown, the ambiguity range is $N_{cp} + N_{sym}$, equal to the previously defined N_L. For each of the N_L possible events, the following two vectors are defined: $\mathbf{r}_{1,i} = [..., y(i + k(N_{cp} + N_{sym})), ...]$ and $\mathbf{r}_{2,i} = \mathbf{r}_{1,i+N_{cp}}$, with $k \in (0, ..N_{S-total})$; $i \in (1, ..N_{cp} + N_{sym})$. Classic estimation of the symbol starting time involves the inner product of $\mathbf{r}_{1,i}$ with $\mathbf{r}_{2,i}$ for all possible candidate values of i. In low SNR, it has been shown [34][31] that it suffices to use the real part of the inner product, i.e.

$$\hat{t} = \arg(\max(real(\mathbf{r}_{1,i}^H \mathbf{r}_{2,i}))) \tag{30}$$

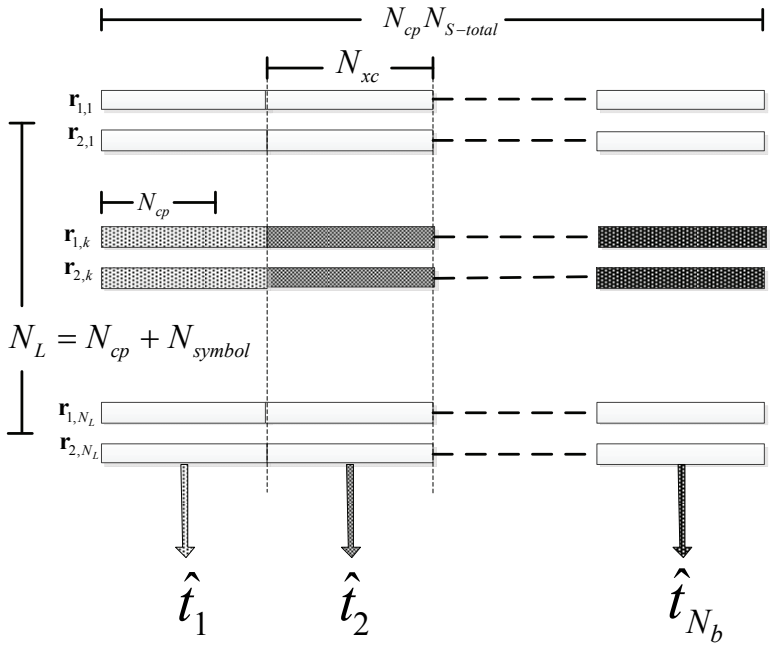

Figure 9: Proposed processing for CP-based detection example.

As shown in Figure 9, N_{xc} is the number of samples employed here for each estimate ($1 \le N_{xc} \le N_s$), where N_s is the total number of received symbols. The constraint imposed by the finite number of samples is described by the following inequality $N_b N_{xc} \le N_s$. The asymptotic processing complexity of the proposed test is only $O(N_L \cdot N_s)$. $O(N_L)$ is the complexity of the maximum and histogram operations.

4.3.1.2 Application to TDOA based source detection

Here the proposed framework was adopted as a test for detecting a source via TDOA estimation. The envisioned scenario assumes multiple pairs of sensors doing an identical test, and then properly combining the individual TDOA estimates for a final statistic. The focus here is on the intermediate step of individual TDOA estimations, not the final geometric transformation of these multiple estimates to localization. The complexity of this last step is described at previous section.

Let r_1, r_2 be the two received sequences by two spatially distinct sensors, both of which either observe noise only or signal plus noise. The signal, if present, is modeled as a white process (extension to colored signals is straightforward). Then, employing classic cross-correlation-based estimation as shown in (30), the lag corresponding to the maximum value is the candidate TDOA estimate.

Figure 10 demonstrates how statistically independent TDOA estimates can be generated. Clearly, at least two estimates are needed for non-trivial tests. Let N_L be the range of values of TDOA, N_s be the total number of samples, N_b be the number of TDOA estimates, and N_{xc} be the number of

samples used for producing each estimate. These parameters are constrained by the inequality $N_b N_{xc} + N_L \leq N_s$.

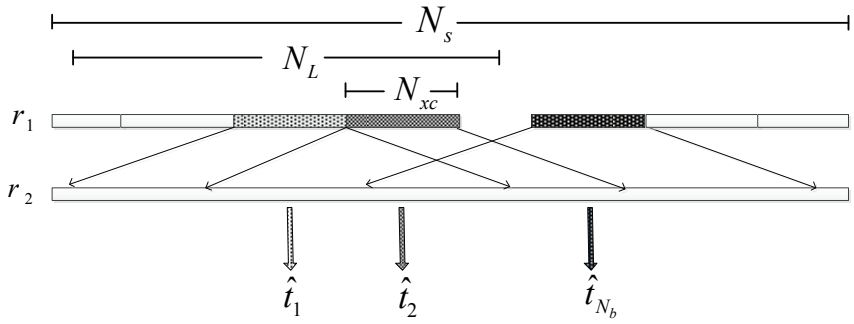

Figure 10: Proposed processing for TDOA-based detection example.

The asymptotic processing complexity of the proposed test is again $O(N_s \cdot N_L)$. $O(N_L)$ is the complexity of the maximum and histogram operations. In this example the complexity is dictated by N_L, the range of possible TDOA values. This range value depends on the geographic region of interest and the relative positions of the two sensors involved in this computation. In all cases, the maximum TDOA value is directly related to the distance between the two sensors under consideration, and the total number of samples is related to the needed localization accuracy.

4.3.2 Platform requirements

The platform requirements are not directly related with the test but the overall context under which this test is been used. For example, in the case of OFDM based detection, only the computational complexity must be met for each sensor. If used for TDOA based detection then the platform must meet the requirements of a TDOA based localization system. This means that the sensors must be able to be synchronized and also send a large number of data to the reference sensor that computes the cross-correlation. If used in an AOA based context, not shown herein, the platform requirements must meet the requirements of a AOA based localization system. Next section comments about the requirements of such a system.

4.4 Angle-of-Arrival estimation in dense multi-path environments

In order to overcome the limitations that the traditional approaches impose on the number of impinging propagation paths that can be spatially separated, a two-stage, *signal specific* approach was developed in [32]. The first stage of this approach estimated the Overall Channel Impulse Response (OCHIR), which includes the effects of the propagation channel plus those of transmit and receive pulse shaping filters. In the second stage, the OCHIR estimates were converted to TOA/AOA estimates using two-dimensional versions of traditional array processing algorithms. Both subspace- as well as *ML*-based algorithms could be used in this second stage. The advantages of this broad approach are: (a) it can be used to spatially separate propagation paths whose number may exceed the number of array elements (provided that there is enough temporal

separation between the paths), and (b) it allows for the joint TOA and DOA estimation of each propagation path.

In [33] we addressed the extension of the aforementioned two-stage approach to environments where training data may not be available and the collection records may be very small (thus not enabling correlation-matrix construction). We introduced a reduced-complexity, *LS*-based conversion algorithm for the second stage which first estimates the PTS and then uses these PTS and OCHIR estimates to estimate the path-DOAs. The present conversion algorithm is applicable to both static as well as non-static (time-varying) propagation environments and can provide reliable estimation performance even when a single observation data block is available. The performance results included here show that the DOA estimation performance of this procedure is close to the Cramer-Rao Lower Bound (CRLB) in a variety of challenging propagation scenarios.

4.4.1 Algorithmic complexity

The two-step procedure of the proposed algorithm is depicted in Figure 11. For the OCHIR estimation, a PSP-LMS algorithm has been proposed in D4.2 [3]. The memory of the trellis diagram is the main parameter that determines the complexity of this step.

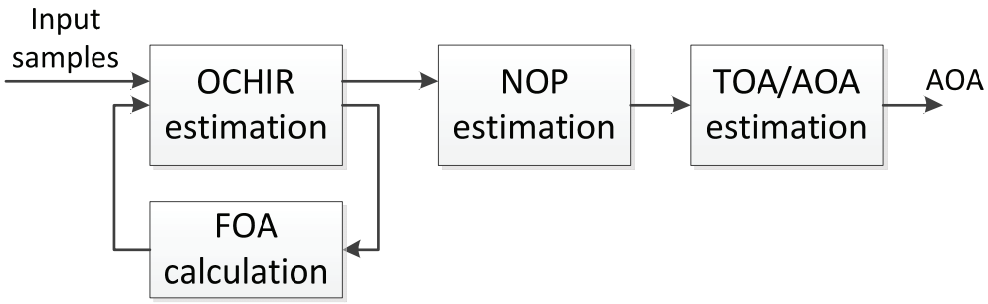

Figure 11: Block diagram of the proposed algorithm for AOA estimation.

If we denote with L_s the memory of the channel at the oversampled received signal, then the adequate trellis memory must be greater or equal to $M_o + L_s - 1$, where M_o is the modulation order. For the update of the state-transition metric the computational complexity is $O(M \cdot N_s \cdot L_c)$ where: N is the observation length (in number of symbols), N_s is the number of samples per symbol (assumed integer), and L_c is the memory of the OCHIR (in number of symbols) and M is the array-element length. The OCHIR estimation is repeated L_{FOA} times, for FOA estimation. The number of path estimation (NOPE) requires: a) the estimation of the sample-correlation matrix ~ O($N_e \; N_{oc}^2$) b) the computation of the Eigen-values of the sample-correlation matrix $O(N_{oc}^3)$, where $N_{oc} = M \cdot N_s \cdot L_s$ is the number of estimated OCHIR coefficients and c) the one-dimensional

minimization of the chosen metric (based on an information theoretic criterion) $\sim O(N_i N_{oc})$, where $N_i < N_{oc}$ is the maximum number of paths.

For the conversion of TOA/AOA estimates, two algorithms have been proposed. The Joint TOA/DOA estimation is the most complex one and requires the implementation of a *3P*-dimensional minimization procedure for the minimization of 2.27 (see D4.2). The complexity for the computation of an output of this function is $O(N_{oc}^3)$. The overall a procedure is computationally expensive, especially when the number of physical propagation paths is large. A reduced-complexity TOA/AOA conversion algorithm was also proposed which first estimates the path TOAs and then uses these TOA estimates to estimate the path-AOAs. This algorithm requires a P-dimensional minimization procedure for TOA estimation and P two-dimensional minimization procedures for AOA estimation and, therefore it is easier to implement than the Joint TOA/AOA conversion algorithm. Table I summarizes the complexity of the proposed solution.

4.4.2 Platform Requirements

At a sensor level, the hardware requirements for AOA estimation involve mainly the support of an antenna array and the processing capabilities to implement the algorithm. The number of antennas as well as the processing power of a sensor determines the quality of the AOA estimate. The multipath environment requires a large number of antenna elements for path separation. In addition, the sort time separation of the multipath components (< symbol time) requires oversampling and thus processing at very high rates. For the above reasons the total complexity of the AOA estimation part is extremely high. Higher than the fusion part where all AOA estimates are combined for the final estimation of the location. This last step requires a non-linear optimization of a non-convex function.

Since the sensor cost is very high, the number of needed sensors targeting specific localization accuracy is a critical parameter for the determination of the deployment cost. Two indicative scenarios are displayed at Figure 12 and Figure 13, one representative for indoor and one for outdoor. The SNR range is the parameter that determines (together with the path-loss) the coverage range of the emitter, in terms of the ability of a sensor to estimate the AOA. At Figure 12, the coverage is 46m while at Figure 13 is 316m. A rectangular grid of sensor deployment is assumed, where all sensors within the coverage range are able to measure the AOA.

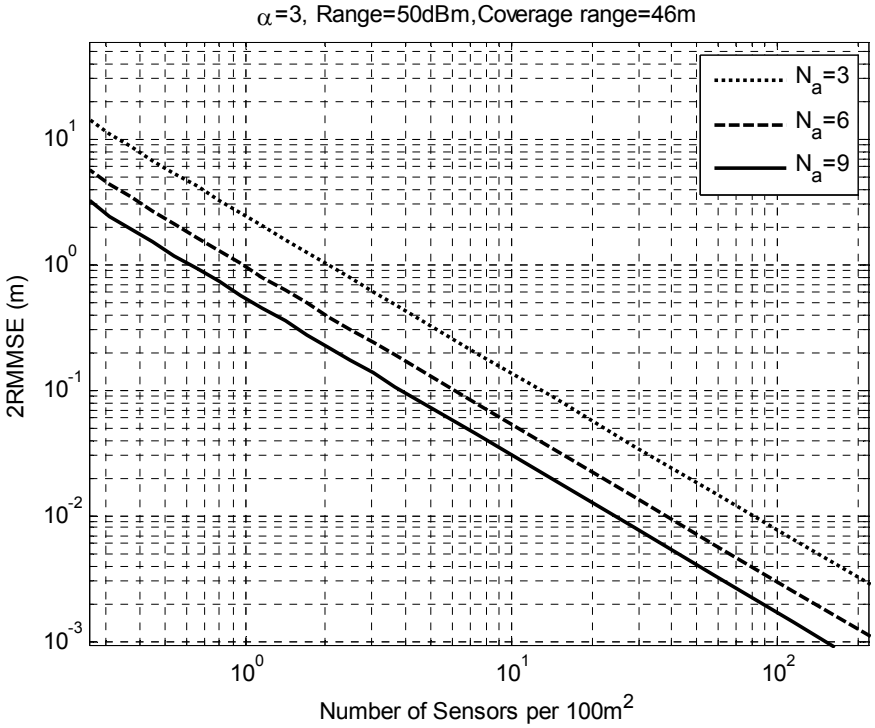

Figure 12: Sensor Density for the indoor scenario.

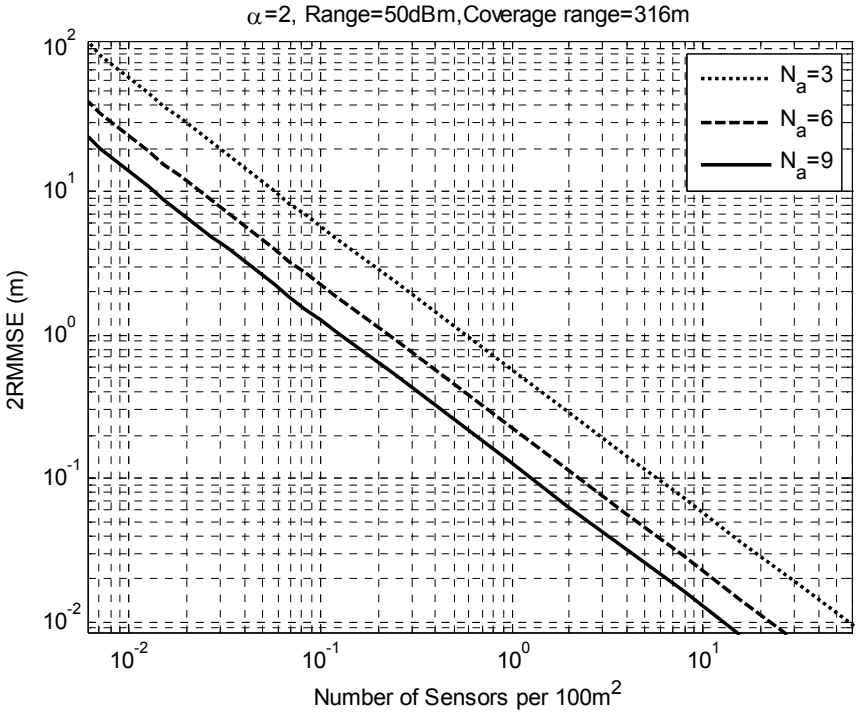

Figure 13: Sensor density for the outdoor scenario.

4.5 Joint transmit power estimation and localization

An algorithm for joint transmit power estimation and localization of an unknown number of transmit emitters in a log-normal fading environment based on power measurements has been proposed (see [3]). In particular, it is considered that sensors of known number measuring power are placed on known locations on a geographical area of interest. The observations $P_r^1,\ldots,P_r^N$ are connected with the unknown emitters via a known probabilistic law, namely:

$$P_r^k = \sum_{i=1}^{S} \frac{1}{d_{ik}^{\alpha}} \mathrm{H}_{ik} P_t^i + \sigma_k^2$$

Here, σ_k^2 models the variance of the additive zero-mean Gaussian thermal noise corresponding to the k-th sensor and P_t^i is the unknown transmitted power corresponding to the *i*-th among all S unknown transmitting emitters. d_{ik} models the distance between the *i*-th possibly transmitting emitter and the kth sensor; α is the path-loss exponent. The shadow fading component H_{ik} is modeled as a log-normally distributed r.v., generated by exponentiating a zero-mean and σ^2-variance Gaussian random variable, henceforth referred to as $\ln N(0,\sigma^2)$. Note that σ^2 is assumed to be known and identical for all emitter-sensor pairs. The shadow-fading components are modeled as uncorrelated (see [27]) since it is known that spatial correlation drops quickly at small distances (a few meters).

The lack of knowledge for the emitter location has been tackled by quantizing the geographical area of interest and henceforth we model the received power measurements as a concatenation of power contributions stemming from a set of candidate transmitting emitter placed on a hypothetical regular geographical grid. Therefore S henceforth denotes the hypothesized number of emitters and P_t^i the unknown transmitted power from the *i*-th hypothesized emitter, to be estimated.

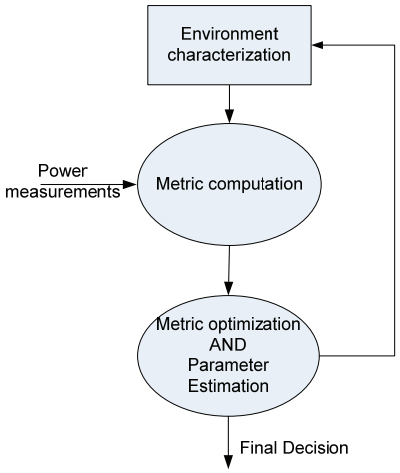

Figure 14: A schematic representation of the algorithm.

The computation of the joint probability density characterizing the observations $P_r^1,...,P_r^N$ is a cumbersome problem. In [28] a new analytic characterization for it was obtained. The key idea behind it is that for a particular combination of emitters transmitting some power, the joint distribution of the power received at all sensors is a multiplication of the power distribution at each sensor. The distribution of the power at each sensor can be then described separately.

For each $k \in \{1,...,N\}$ $P_r^k - \sigma_k^2$ can be approximated by a log-normally distributed random variable $\ln N(\mu_{p_k}, \beta_{p_k})$, where

$$\mu_{p_k} = \ln\left(\sum_{i=1}^{S} \frac{P_t^i}{(d_{ik})^a}\right) + \frac{\sigma^2}{2} - \frac{\beta_{p_k}}{2}$$

$$\beta_{p_k} = \ln\left((e^{\sigma^2} - 1)\frac{\sum_{i=1}^{S}\left(\frac{P_t^i}{(d_{ik})^\alpha}\right)^2}{\left(\sum_{i=1}^{S}\frac{P_t^i}{(d_{ik})^\alpha}\right)^2} + 1\right)$$

The joint probability law of the power received by all sensors is then given by

$$(P_r^1 - \sigma_1^2,...,P_r^N - \sigma_N^2) \square \prod_{k=1}^{N} \ln N(\mu_{p_k}, \beta_{p_k})$$

The goal is to maximize this multiplicative law for all possible combinations of $P_t^1,...,P_t^s$. This yields a non-convex optimization problem that can only be solved numerically. To this end, global optimization techniques, and in particular simulated annealing ([29]), are employed. More specifically, the optimization problem to be solved is:

$$\min \sum_k \frac{1}{2}\ln\beta_{p_\kappa} + \sum_k \frac{(\ln P_r^k - \mu_{p_k})^2}{2\beta_{p_\kappa}},$$

where μ_{p_k} and β_{p_κ} are defined above. The minimization above is over the values of power on each of the S candidate locations and these parameters are involved in the expressions of μ_{p_k} and β_{p_κ}

A schematic representation of the above algorithm is provided in Figure 14.

The algorithm can be briefly described by the following steps:

Step 1: Obtain environment characterization (α, σ)

Step 2: Obtain power measurements

Step 3: Compute metric

Step 4: Optimize metric result (combining also other system specifications, eg. source inter-distance and power upper/lower bounds)

Step 5: IF accuracy specification is not met AND time has not elapsed GO TO Step 1, ELSE GO TO Step 6

Step 6: Provide localization and power estimation

The termination of the above algorithm, ie. Step 6, requires particular attention and explanation. In fact, since at various problem settings constraints and side information could be provided this determines the stage where the algorithmic solution is final. Common forms of side information is an upper and/or lower bound for the number of active emitters, a lower bound for the distance between two active emitters etc. In this case the soft estimates of Step 4 are processed accordingly and Step5 decides whether the answer is final or another iteration is required.

4.5.1 Algorithmic complexity

The optimization to be performed in Step 4 is essentially a non-convex optimization problem that can be solved numerically. The complexity of this step is characterized by the grid size. For a kxk-size grid the complexity is exponential with respect to k. In order to solve this problem global optimization tools can be employed, in which case the complexity is essentially characterized by the complexity of the particular such chosen tool. In the simulation investigation presented in [28] particularly, simulated annealing ([29]), is employed. This is a probabilistic method for finding the global optimum of a function that might posses several local optima. Specifically, for simulated annealing it has been shown that the complexity can be in some cases bounded by a polynomial. For a detailed analysis on that see [30] and references therein. Alternatively, other tools such as genetic algorithms can be employed.

In a system where the above algorithm is employed sequentially in time and changes in the channel and the emitters are relatively slow, knowledge can be accumulated with time and facilitate the solution of the optimization Step 4. In such cases, eg. when the initial conditions for the optimization are sufficiently good, the well-known convex optimization tools can prove sufficient and complexity can then be reduced.

4.5.2 Platform requirements

The algorithm previously described requires limited platform components in order to be implemented. In particular, since the estimation is based on power measurements, it essentially requires sensors measuring power. It also requires awareness of location for these sensors, relative to some fixed point inside or outside the geographical area of interest. Since the location of the emitter(s) is(are) not known, the pairwise distance between each hypothesized point in the area and each sensor is critical to the performance of the algorithm. The number of sensors is also an important parameter and is related to the expected accuracy in solution. In particular, in [28] it is investigated numerically and argued that the number of required sensors is O(S) and increasing it beyond this level does not result to important increase in performance.

5 Experimental evaluation

5.1 Introduction

Performance results based on experimental trials will be displayed in this chapter. Since only compatible techniques with the current HW capabilities of the FARAMIR prototype can be demonstrated, the RSS based techniques are under investigation. The measurements have been collected by the use of FARAMIR REM platform and involve indoor scenarios only. A number of MCDs are placed at a specific area and a number of emitters are used to generate a field in a controlled manner. The results from two different scenarios will be demonstrated, scenario A, based on the DySPAN 2011 demonstration and scenario B, based on field trials conducted at the UKIM premises.

> **Scenario A:** Four different kinds of MCDs were used, with a total of 16, in order to localize one emitter. The success of this experiment was visually demonstrated on the DySPAN 2011 conference demonstration session, where the emitter was allowed to be moved at different locations. Here we reuse the collected experimental data using also different algorithms, and compare the results with the theoretical bounds.
>
> **Scenario B:** Ten TI eZ430 RF2500 devices and three USRP2 devices was set up in an indoor environment. The devices were placed in a $25m^2$ area in a classroom having a large number of chairs and tables as obstacles and shadowers. The TI sensors were used as MCDs, while the USRP2 devices were employed as signal emitters generating 5 MHz wide OFDM signals. The experiments focused on the 2.4 GHz ISM band evaluation. Here the additional interest is to localize two of them, and predict their transmit power, and also to test the RIFE reliability when a small number of measurements is used.

The algorithms that we are going to use are the following:

- For one emitter localization the *ML* estimation will be used. The DySPAN version, proposed in [6], will be used as the 'practical' algorithm for implementation on the REM platform, while the full *ML* search (not based on a grid) will be used for comparison, as the highest-complexity, best-performance alternative. Another significant difference is that the former does not assume any knowledge about the propagation model.
- For the two-emitter case also two algorithms will be used. The *LS*-based algorithm will serve as the low-complexity 'practical' scheme, while the joint *ML* algorithm will serve as the high complexity best-performance alternative.
- For the RIFE problem the classical IDW, the modified classical IDW, the modified Shepard's method, referred as IDW, IDWM and MSM, respectively.

5.2 Passive localization and power estimation

With the help of the performance bounds and the simulation capability of the software tools we will try to:

a) Verify the experimental results with the theory
b) Identify some key issues that appear in real experiments and usually are not taken into account when testing algorithms through simulations.
c) Draw some guidelines, or even rules of thump for a successful sensor deployment to meet the needed QoS requirements.

The main difficulty in such efforts is the channel propagation model. We will use the simple log-distance model in our case. The simplicity of this model allows us to implement practical algorithms but on the same is the main source of error since it is not such an accurate model, especially for an indoor environment. The joint simulation/experimentation will allow us to know when the results that we are observing are caused by the adopted model.

5.2.1 Single emitter case

5.2.1.1 Scenario A

The measurements for this scenario are taken from the DySPAN 2011 demonstration setup. As already stated above, in this scenario we used 16 MCDs of various types. The positions of these sixteen sensors and a fixed position of one emitter are depicted in -a. The real measurements that we are going to use in this scenario are taken only for this fixed position of the emitter. Using the known position of the emitter we initially estimate the parameters of the log-distance propagation model by LS fitting and depict the results in Figure 15-b. The fitted parameters gave a path-loss of 3.92 and a log-normal shadow-fading variance of 18dB.

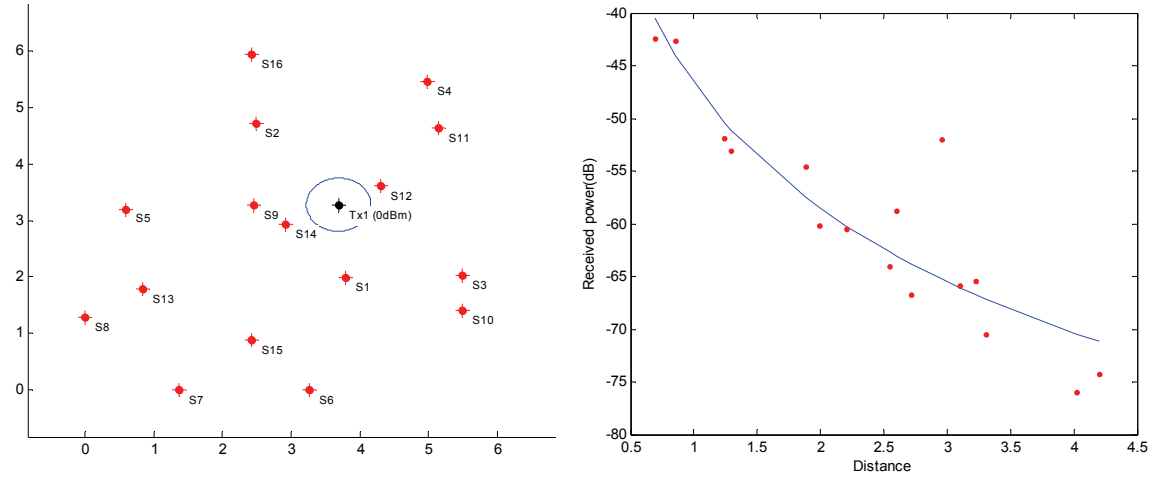

Figure 15: a) DySPAN Demo setting b) LS propagation model fit.

As it can be observed the model fits the data quite well. This can be also verified by the CRLB of localization error (MMSE~0.055). A blue circle that indicates the $2 \cdot \mathrm{RMMSE}$ based on this model is also plotted, that visually demonstrates the minimum localization error (95% of times falls within this cycle) that can be achieved. At Figure 16 the performance of the full *ML* search (ML1) is displayed for a simulated environment with the extracted propagation model and the emulated environment, using the stored RSS measurements. As it is expected, the simulated performance follows well the analysis, giving performance very close to the theoretical limit. The results from the real also fell into the predicted area. The exhibit a bias is expected, since what we really have by using the measurements is just one realization of the adopted model. The different estimates around this 'bias' are the result of the fast fading, that changing the measured power as the time evolves at a small fraction. This is not the case with the simulated data, where the different estimates are produced by generating different loss due to shadow fading per sensor, per sample. In both cases the Tx power estimation is very close to the real value.

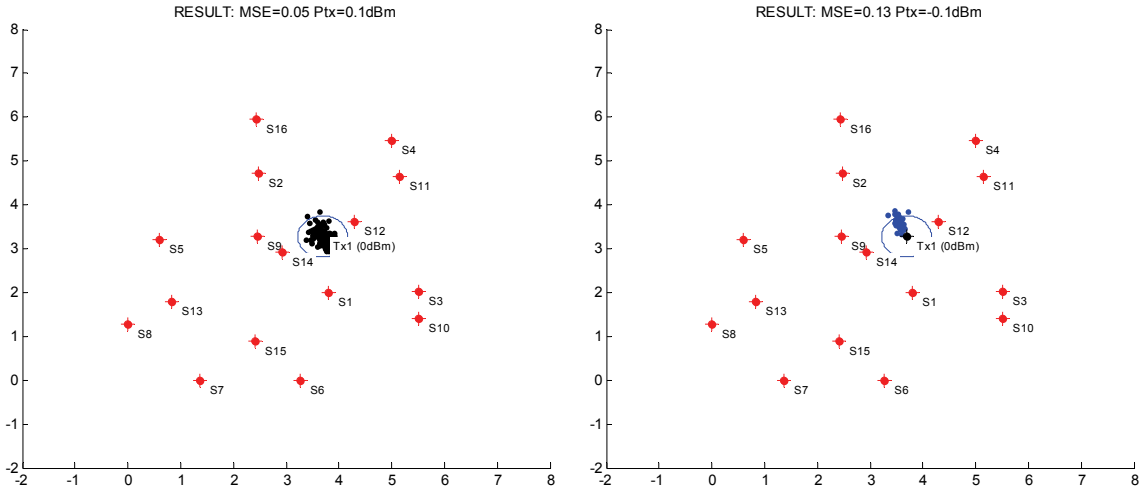

Figure 16: a) Simulation and b) Experimentation with full ML1 search.

The result of the *ML*-Grid search [6], implemented and used at the IEEE DySPAN 2011 demonstration setup is displayed at Figure 17 for the simulated and the emulated environment. The green dots represent the rectangular Grid of search that the algorithm is using. It has 15^2 points for the displayed area, resulting at an inner point distance ~0.5 meters. The blue points indicate that these Grid points represent a number of solutions. A small, but acceptable degradation on the performance is observed, due to the limitation putted by the Grid, and the fact this algorithm does not assume prior knowledge of the propagation model.

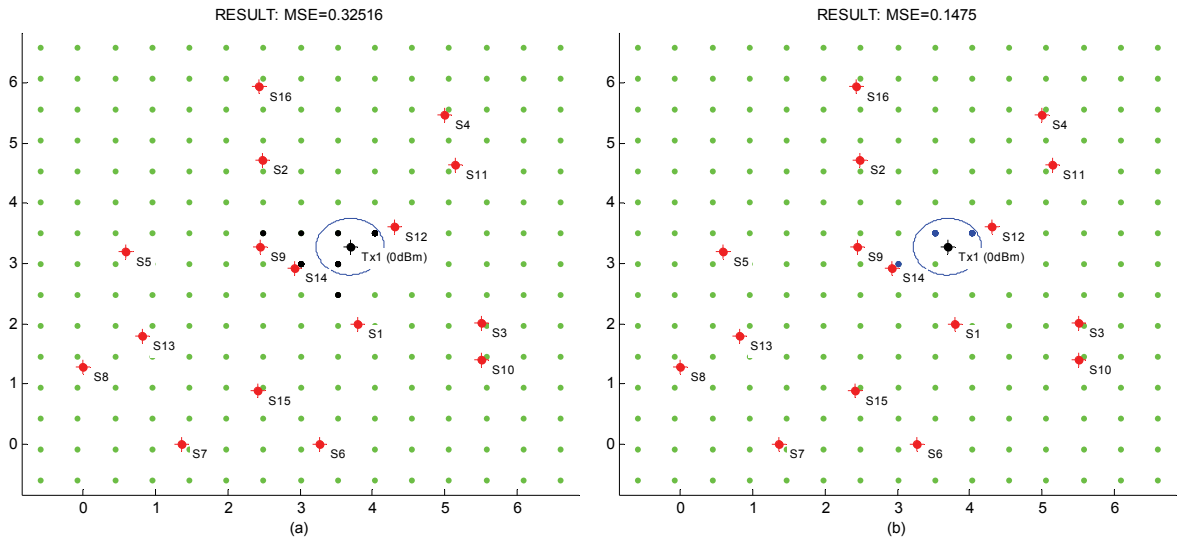

Figure 17: a) Simulation and b) Experimentation with Grid based ML2 search.

The performance results are good and in agreement with theory, but one can argue that the density of the nodes is too high. We will test the robustness of this scenario to the density of the nodes by turning – off some of the sensors (i.e. not take into account their measurements). At Figure 18 and Figure 19 the same results as previous are displayed, but now four of the closest sensors were turned-off. For the case of ML1, the degradation of the performance follows the expected one, as we can compare from the simulation and experimental results of Figure 18.

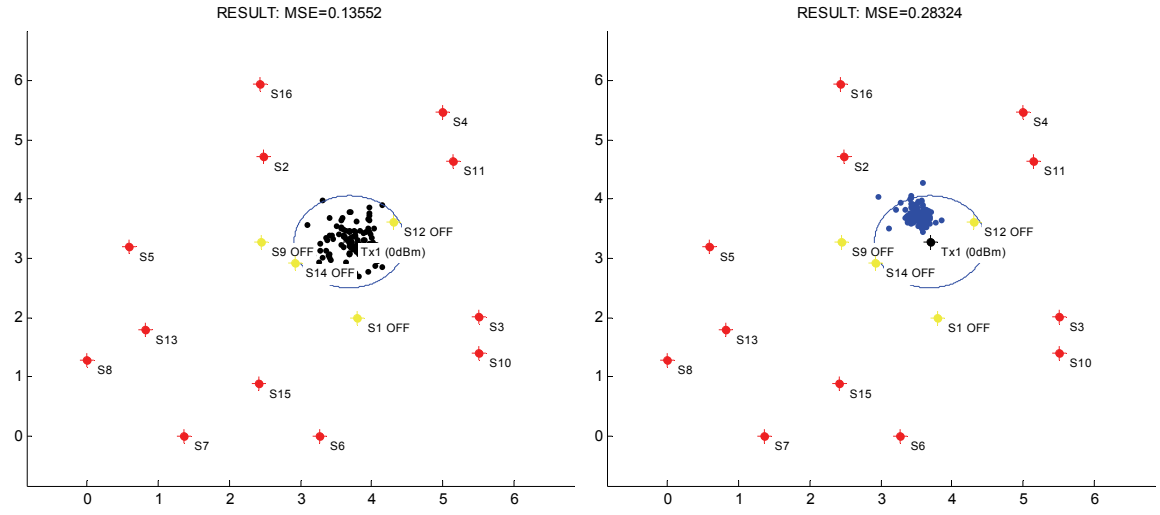

Figure 18: a) Simulation and b) Experimentation with ML1.

At the ML2 case, the performance degradation is very high. The reduction of the number of sensors has a higher impact on this algorithm, since it has to estimate more parameters than the ML1.

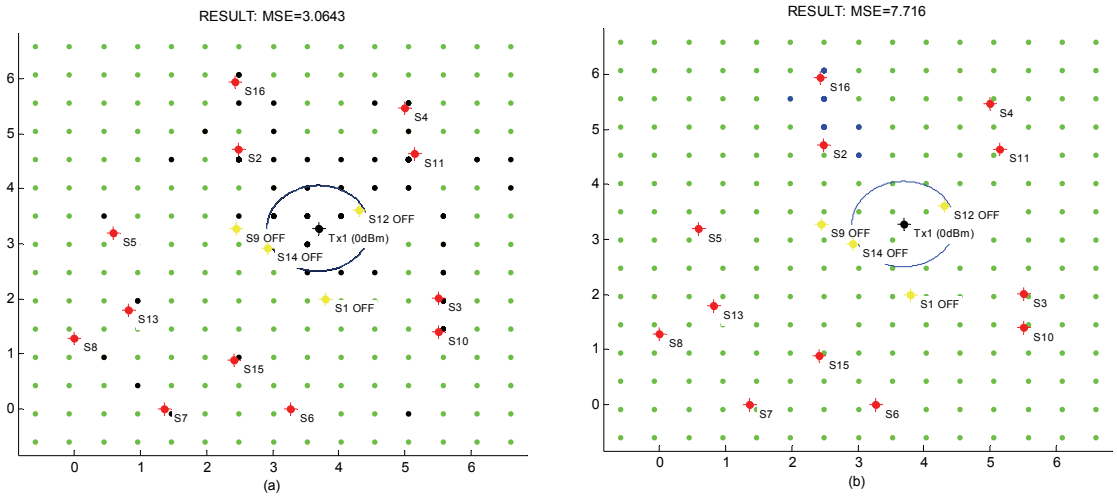

Figure 19: a) Simulation and b) Experimentation with ML2.

At the next three figures, we demonstrate the same results, but now we only leave 5 active sensors. The ML1 algorithm still follows the theory, since it uses a good estimate of the propagation model.

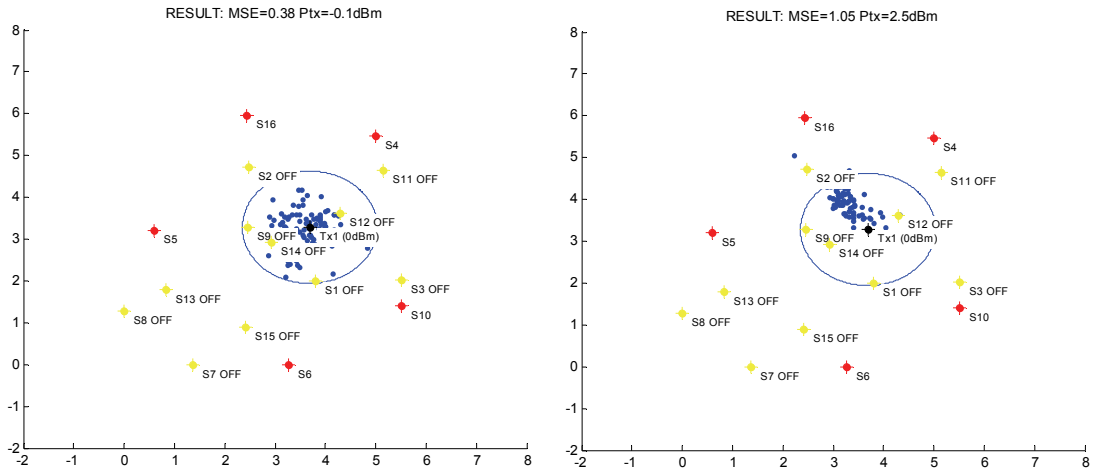

Figure 20: a) Simulation and b) Experimentation with ML1.

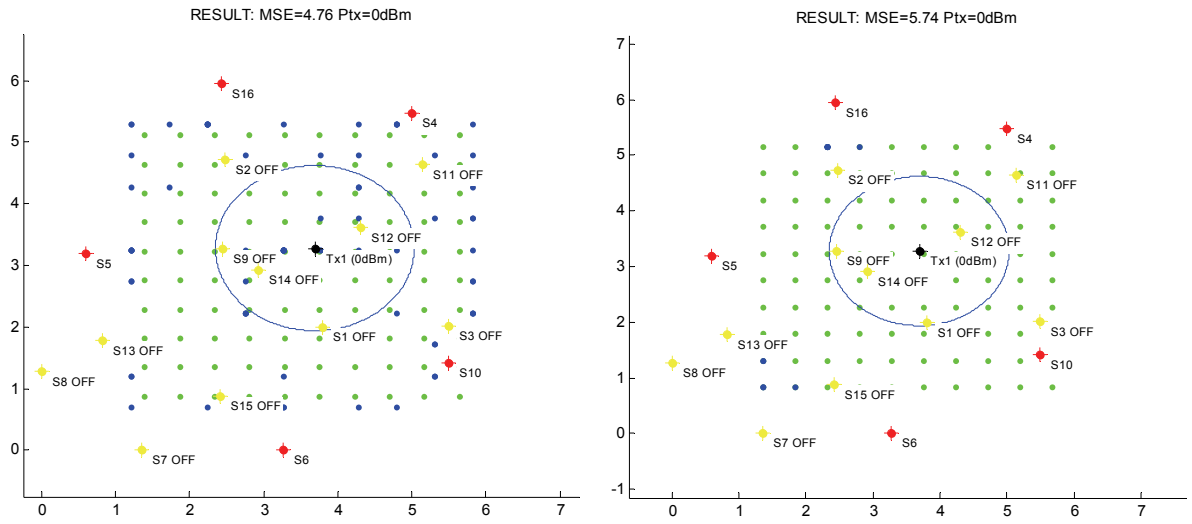

Figure 21: a) Simulation and b) Experimentation with ML2

ML2 totally loses the target as shown in Figure 10. In this case we have also decrease the coverage area of the Grid.

The benefit of having a good propagation model when the density of the sensors is small has been depicted by the preceding simulation/experimental results. ML1 exhibits excellent performance with respect to the theoretical bounds, but there is something that needs to be commented. The propagation model was estimated by the use of the same emitter, thus it can model the area around the correct position with high accuracy. In practical scenarios, the propagation model will be estimated by the use of several emitters (probably the MCDs, playing the role of an emitter), and this model will describe, on average, the propagation characteristics of the entire area of interest. The ML2 algorithm exhibits high sensitivity to the number of sensors, but it does not take into account any prior knowledge. At the next scenario we will see the consequences of the different model fit based on the position used, as it supports multiple emitters.

5.2.1.2 Scenario B

These measurements were taken at UKIM premises, using 10 MCD's, and three potential transmitters. The Tx-power level of each transmitter was selected between four values, 0, -10, -20 and -30dBm.

Figure 22 depicts the performance of the ML1, ML2 and LS based localization algorithms for an emitter positioned in the center of the sensor deployment area. The displayed results use the propagation model that was estimated by using the same place emitter position, called the 'local' model. A more realistic scenario is to estimate the propagation model by using measurements from different positions in the area of interest. The results of this procedure are depicted in Figure 23.

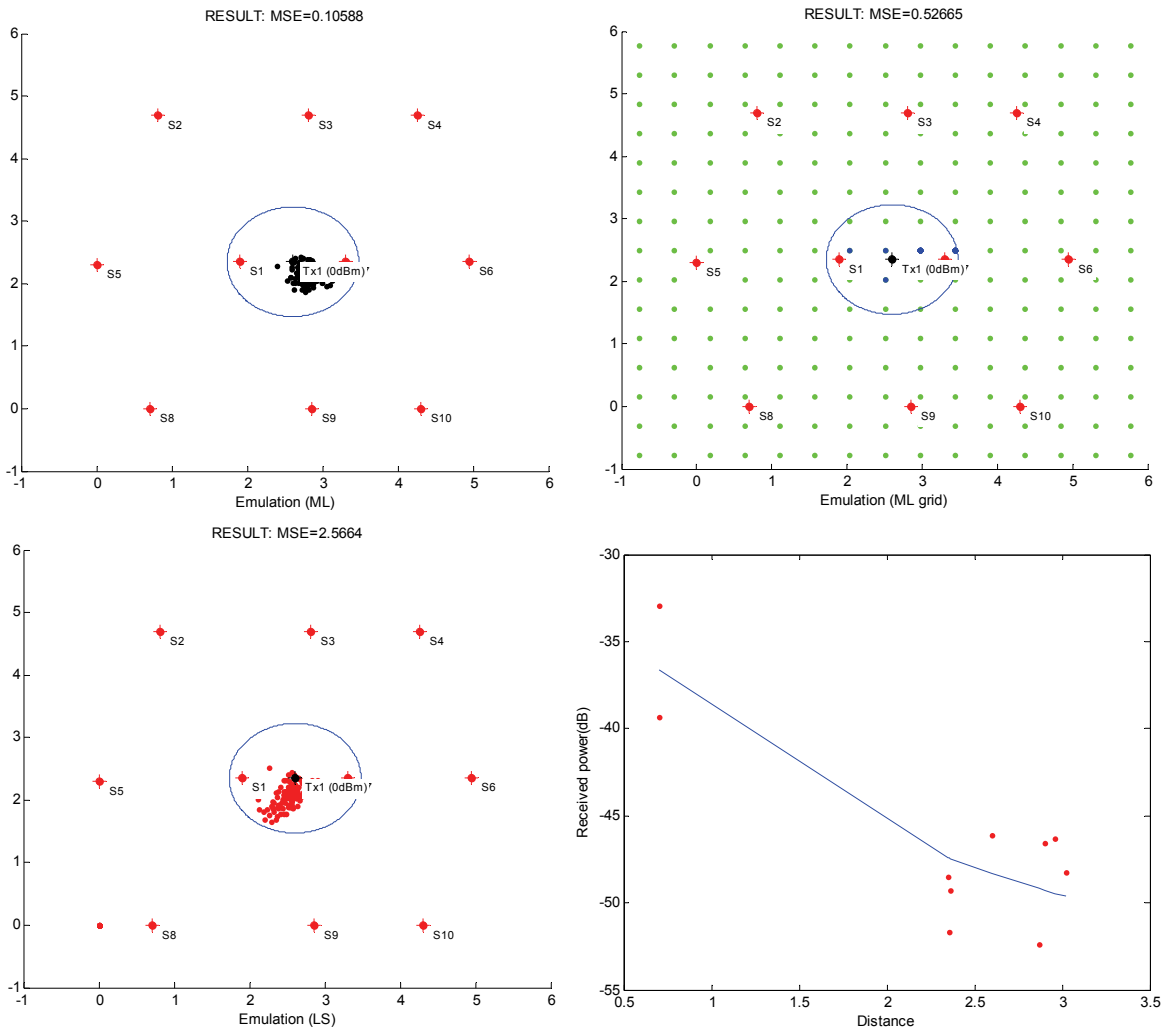

Figure 22: Performance of various localization techniques for Scenario B.

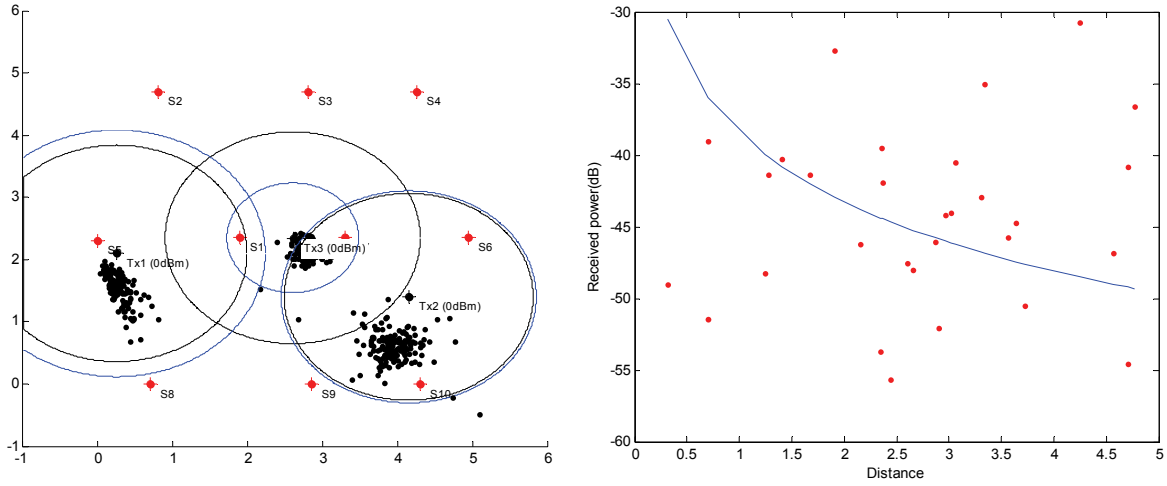

Figure 23: Performance of localization based on the "global" model.

All measurements from the three emitters are used to estimate the 'global' propagation model. We also depict the RMMSE for localizing each one of the emitters by using the 'global' model (black circles), as well as the local model (blue circles). It is clear that the localization accuracy is not so good as we approach the edge sensors. It is expected due to the geometry of the sensors as well as the propagation characteristics (those areas are close to walls). The estimates were made by using the global model. As we can see, the performance for the middle emitter (Tx3) is still best characterized by the local model. That is, even if the global model does not well describe the propagation characteristics, the performance seems to be robust to model errors. Table 13 summarizes the fitted model values for all the locals as well as the global model.

Table 13: Propagation Model fit.

Propagation Model	Path-loss	Shadow fading (dB)
Tx1 local	1.62	26
Tx2 local	1.5	18
Tx3 local	2.4	8
Global	1.6	19

5.2.2 Multiple emitter case

The analytic performance characterization of multiple emitter localization is a cumbersome problem, due to the form of the distribution that describes the received energy (see D4.2 [3]). In this paragraph we use the CRLB for one emitter to analyze the performance results we get at the

multiple emitter case. We use the scenario were Tx1 and Tx2 are active (see Figure 24). We represent with green emitters the additional candidate locations that our algorithm uses together with the real ones for detecting the emitters. So, in this case we always use candidate locations that include also the locations of the open emitters. This information could have been provided to us by the REM.

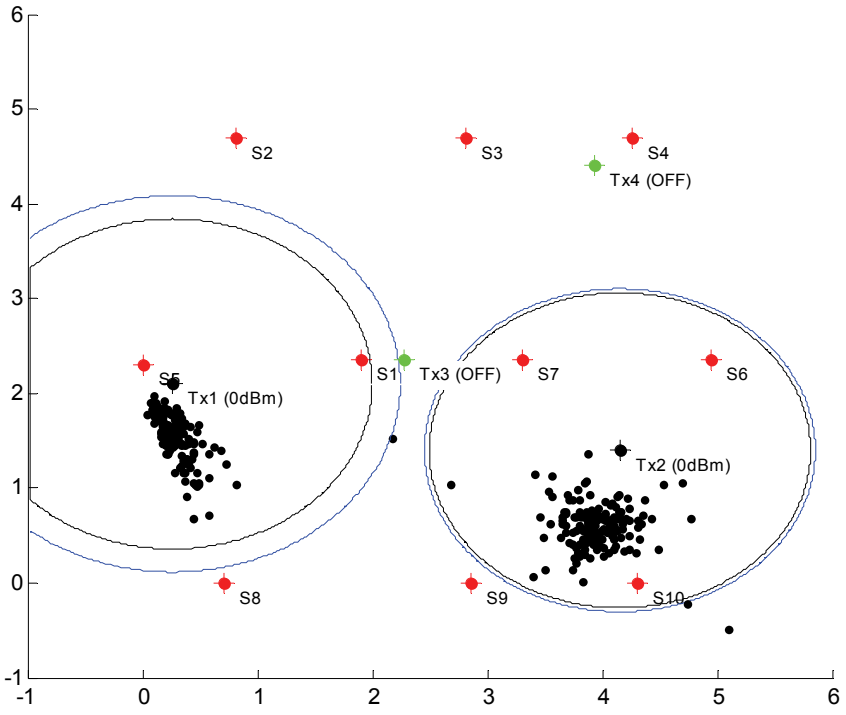

Figure 24: Scenario B, two active emitters, four candidate locations.

At the first set of experiments, we use only two candidate locations, as depicted in figure 14. The locations are chosen so as to be outside of the $2\times\mathrm{RMMSE}$ region, where 95% of time the ML estimate resides in this region (*at the single emitter case*). A *LS*-based algorithm is used with iterative candidate cancellation. The algorithm takes as input the maximum number of possible active emitters and gives as output the $N_{\max}$ best candidate locations and their estimated tx-power. Table 14, Table 15 and Table 16 display the performance results of the above setting, using the experimental data and using $N_{\max}=2,3,4$.

Table 14: $N_{\max}=2$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	100	100	0	0
Power(dBm)	-1	2.6000	NaN	NaN

Table 15: $N_{\max} = 3$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	100	100	46.6000	53.3000
Power(dBm)	-1	2.7000	-18.7000	-25.5000

Table 16: $N_{\max} = 4$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	100	100	100	100
Power(dBm)	-1	2.9000	-21.6000	-27.7000

The algorithm manages to provide as output the correct candidate locations 100% of the time, for all values of $N_{\max}$. If we look at the estimated values of Figure 24 we notice that the wrong candidate locations are always further away from the estimates than the correct candidates. This can be used as an explanation of the results. Table 17 Table 18 and Table 19 display the performance results when simulation is used following the global propagation model.

Table 17: $N_{\max} = 2$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	91.1000	87.4000	12.9000	8.6000
Power(dBm)	2.3000	2.2000	-0.8000	-6.3000

Table 18: $N_{\max} = 4$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	98.6000	96.5000	43.4000	61.5000
Power(dBm)	1.9000	1.7000	-4.7000	-10.1000

Table 19: $N_{\max} = 6$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4
Winning %	100	100	100	100
Power(dBm)	1.7000	1.9000	-8	-10.8000

The results are still good, but a leakage towards the wrong candidate locations can be observed. This is expected, since a large number of ML estimates are expected to be closer to the wrong two candidates. In the case were all four candidates are considered to be active (Table 19), a 10dB difference in tx-power is the result that separates the active from the non-active emitters. In the same case when using experimental data, this was around 20dB.

One must note here, that the experimental results just represent one realization of the simulated experiment. That is because, in the simulated case, the propagation channel (shadow fading) is random for each different realization, while at the experimental case is constant. The spreading that we observe at the experimental results is accounted to the fast fading and not the shadow fading.

At the second set of experiments we increase the number of candidate locations to eight, as seen in Figure 25. Locations inside the RMMSE circles are used as well as close to the estimated locations.

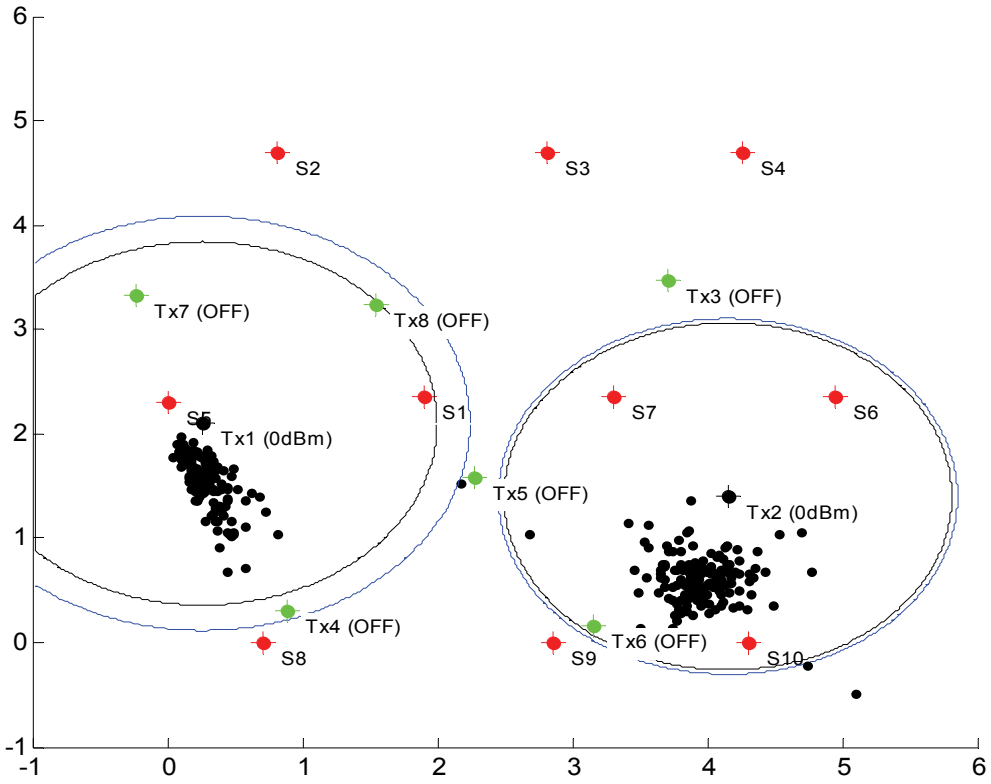

Figure 25: Scenario B, two active emitters, eight candidate locations (1).

The performance results of the experimental data are in agreement with what we 'expect' them to be by visual inspection of Figure 25. Tx4 and Tx6 are the wrong candidates that are close to the estimated positions, and thus are expected to be selected when the number of active emitters is greater than two. Indeed, as we see in Table 21, when $N_{\max} = 3$, they are 90% collectively chosen as the third correct candidate.

Table 20: $N_{\max} = 2$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	98	100	0	0.6000	1.3000	0	0	0
Power(dBm)	-1	2.5000	NaN	-2.9000	1.1000	NaN	NaN	NaN

Table 21: $N_{\max} = 3$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	100	0	51.3000	6	40.6000	0	2
Power(dBm)	-1.1000	2	NaN	-6.3000	-3.2000	-7.3000	NaN	-6.2000

Table 22: $N_{\max} = 5$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	100	5.3000	99.3000	43.3000	98	0	54
Power(dBm)	-1	2.1000	-13.5000	-7.6000	-13	-8.6000	NaN	-16

Table 23: $N_{\max} = 6$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	100	100	100	100	100	100	100
Power(dBm)	0.2000	4.6000	-Inf	-5.9000	-22.6000	-7.3000	-28.4000	5.3000

At the simulation case, Table 24~Table 27, it is evident that the performance degrades rapidly, since we place candidates close and inside the RMMSE region. As a rule of thumb we can use the CRLB of one emitter as a lower bound on multiple emitter separability.

Table 24: $N_{\max} = 2$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	70.4000	53.9000	18.8000	5.5000	20.1000	10.6000	14.4000	6.3000
Power(dBm)	3.1000	2.2000	2.2000	-0.1000	2.6000	-3.1000	4.3000	1.8000

Table 25: $N_{\max} = 3$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	87.7000	69.9000	28.5000	23.8000	26.6000	32.8000	18.1000	12.6000
Power(dBm)	2.2000	1.2000	0.2000	-5.7000	1.6000	-6.6000	1.6000	-0.6000

Table 26: $N_{\max} = 5$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	99.2000	87	53.6000	86.8000	36.6000	84.5000	32.5000	19.8000
Power(dBm)	1.9000	1.5000	-2.1000	-9.3000	0.4000	-10	-2	-3.6000

Table 27: $N_{\max} = 6$ (Simulation).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	100	100	100	100	100	100	100
Power(dBm)	2	2.5000	-1.4000	-8.7000	-0.4000	-9.4000	1.4000	2.3000

At the last set of experiments, we move two of the candidates into the area that the single emitter estimates fall. More precisely, we move Tx4 as close to this area as the true position (Tx1) lies and Tx6 in the middle of Tx2 estimates (see Figure 26).

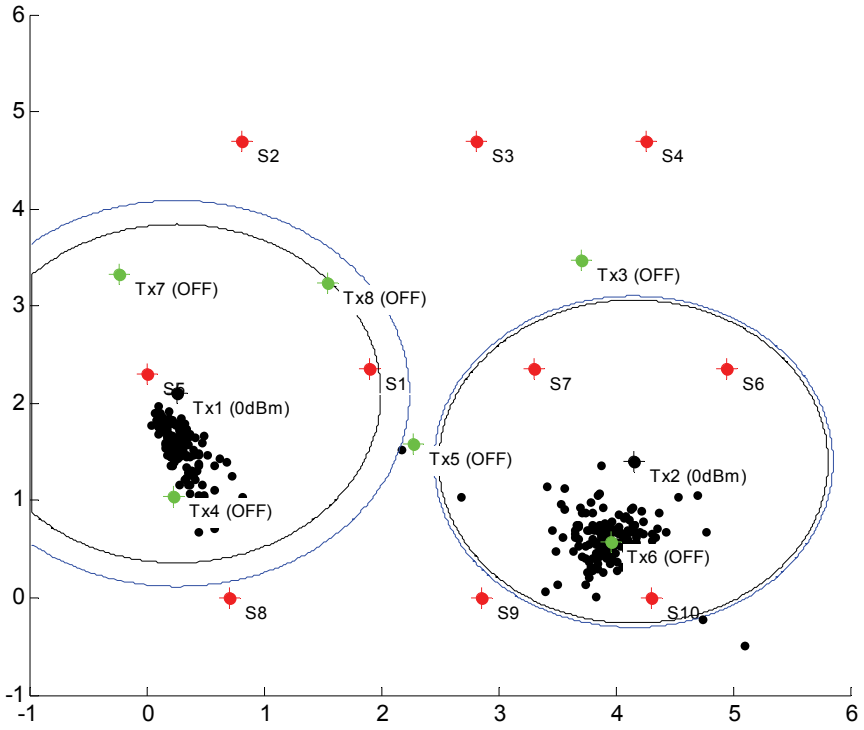

Figure 26: Scenario B, two active emitters, eight candidate locations (2).

As expected, for the case of Tx1 and Tx4, and when $N_{\max} = 2$ (Table 28), the two candidates share approximately equally the probability to be selected. Also for the case of Tx6, it is selected 98% of times, since we have put this candidate in the middle of Tx2 estimates.

Table 28 $N_{\max} = 2$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	66	1	1	30	0	98	0	0
Power(dBm)	-0.5000	1.4000	0.5000	3.1000	1.1000	1.3000	NaN	1.5000

Table 29 $N_{\max} = 3$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	96	5	19	72	3	100	0	3
Power(dBm)	-1.3000	-3.2000	-4.2000	-1.9000	-4.5000	1.2000	NaN	-3.5000

Table 30 $N_{\max} = 5$ (Experiment).

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	34	94	98	26	100	0	45
Power(dBm)	-1.3000	-9.1000	-6.8000	-3.1000	-10.1000	1.2000	NaN	-13.9000

Table 31: $N_{\max} = 6$

	Tx-1	Tx-2	Tx-3	Tx-4	Tx-5	Tx-6	Tx-7	Tx-8
Winning %	100	100	100	100	100	100	100	100
Power(dBm)	-0.6000	-8.6000	-4.9000	-1.1000	-12.5000	1.7000	-15.9000	1.4000

In conclusion, for this baseline scenario of two emitters, separation can be 'guaranteed', as long as the candidate locations are in a distance that is related to the RMMSE of the single emitter scenario. This does not mean that we get error location estimates, since the wrong candidate location could be indeed closer to the *ML* solution than the real source. It just signifies that our estimate on which emitter is active is wrong. So, in case where we are interested in estimating the location of the active emitters, then a dense grid will provide better results. The prior knowledge of the possible number of active emitters also plays significant role to the performance results. When the identification of the exact active emitters from a list of candidates is our target, then some criteria of the minimum distance between the candidates could be drawn by the use of the CRLB of single emitters.

5.3 Direct RIFE based on spatial interpolation

This section focuses on the performance evaluation of the IDW based spatial interpolations, when applied to the RIF estimation problem. The aim is to test the behavior of the spatial interpolation methods, detailed in the previous chapter, in terms of the RIF estimation reliability with a limited number of MCDs. Scenario B is used for the performance evaluation. Figure 27 illustrates the placement of the sensors and the signal sources. A total of 24 different scenarios (combinations) were tested, having each of the signal sources active as a single transmitter or in a pair of

transmitters, with three possible transmit power levels, i.e. -5, -15 and -25 dBm. Each inspected scenario was evaluated for 5 minutes to gain sufficient time domain statistics.

It is important to note that the radio environment during the measurements was not completely controlled, and some of the errors may originate from the outer interference.

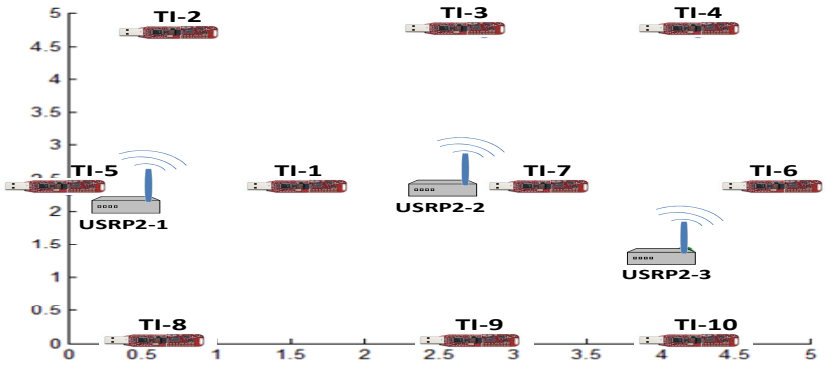

Figure 27: Evaluation scenario.

The IDW methods of interest are: the classical IDW, the modified classical IDW, the modified Shepard's method, referred as IDW, IDWM and MSM, respectively. An adaptive MSM (AMSM) approach, which used adaptive N_q and N_w parameters depending on the location of the interpolation point was also tested. Oppositely, the classical MSM interpolation used static parameters, proven to give the best results in overall for all interpolation points. An additional modification of the MSM approach was also evaluated. Namely, this approach used an inverse covariance matrix of the input values (x, y coordinates) for the WLS estimation of the nodal functions parameters, instead of using the proposed weights in (16). Respectively, the MSM and the AMSM interpolation taking this WLS approach are subsequently referred as MSMC and AMSMC. All MSM approaches used linear nodal function model.

The evaluation on each of the abovementioned methods focused on the error, the absolute error (AE) and the mean absolute error (MAE) per sensor. The calculations of the errors were made with a sensors exclusion approach, averaging over all combinations starting from nine (out of ten) active sensors down to only five active ones. The errors were calculated as residuals of the interpolated values and the values measured by the excluded sensors. Only the corner sensors with IDs 2, 4, 8 and 10 were not excluded to alleviate the border effects.

The obtained results are cumulative for all 24 inspected scenarios. They are presented as box plots, where the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considered as "outliers". The used whisker length is 1.5 times the interquartile range.

Figure 28 presents the comparison between the different IDW methods in terms of the MAE per sensor for the case of nine active sensors. In addition, the figure presents the impact of the distance exponent in the interpolation MAE. The results prove that the IDW methods generally

perform better when the distance exponent d_{exp} is equal to 1. The AMSMC method with d_{exp}=1 is the most reliable interpolation in overall, because it offers the lowest and the less variable MAE. It is important to note that the MSM based methods do not perform well when the used d_{exp} is equal to 2. The IDW and the IDWM methods have similar behavior, the latter experiencing slightly lower variance of MAE but having more "outliers".

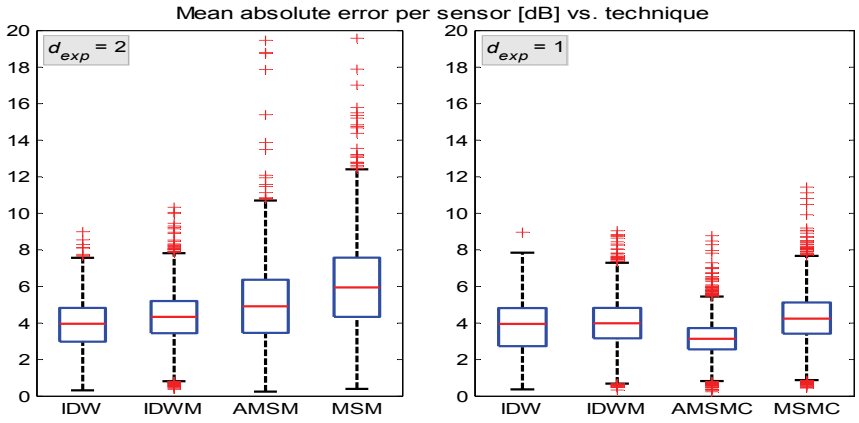

Figure 28: Comparison of IDW based methods in terms of MAE per sensor, impact of distance exponent for interpolation with 9 active sensors.

The following analyses focus only on the interpolation methods having d_{exp} equal to 1, due to the better interpolation performances. The focus of the MSM based interpolations remains on the MSMC and the AMSMC approaches, since they provide significantly better performances.

Figure 29 depicts the box plots presenting the MAE per sensor performances of the IDW interpolation methods of interest with respect to the number of active sensors in the interpolation. As verified by the testbed results, the AMSMC approach (with d_{exp}=1) provides the best MAE performances for each case, with the standard deviation of MAE per sensor ranging from 1.04 dB to 1.56 dB, for nine and five sensors, respectively. However, this interpolation is mostly affected by the decrease of the number of sensors. This is due to the fact that the optimization (adaptation) space of the N_q and N_w parameters is reduced with the decrease of the number of active sensors. While the classic IDW approaches are more robust to "outliers", the MSM approaches are not.

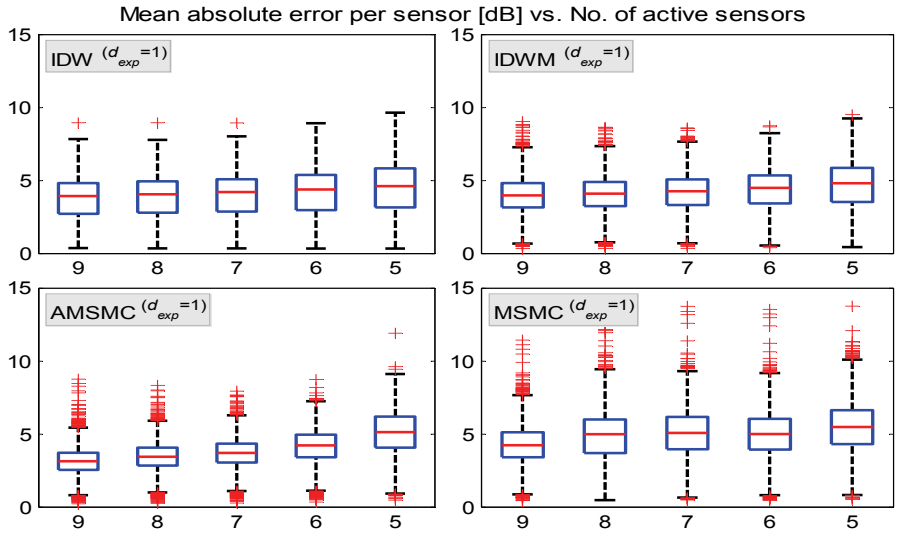

Figure 29: Impact of the number of active sensors to the MAE per sensor for the different IDW based methods.

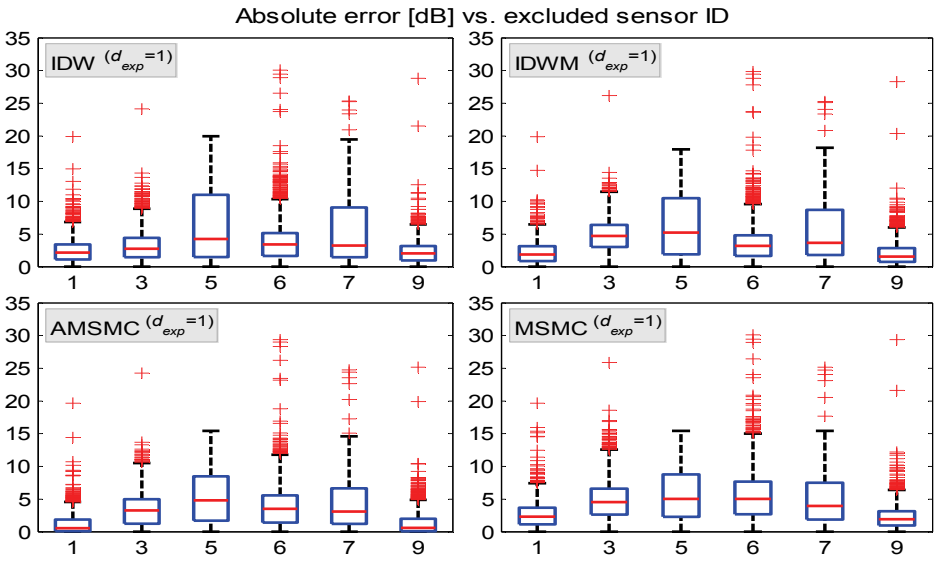

Figure 30: AEs at the excluded sensor ID locations for 9 active sensors.

Figure 30 presents the dependence of the absolute error on the location of the interpolation point. Namely, the AEs are evaluated at the locations of the excluded sensor, for the case of nine active sensors. The AMSMC interpolation approach again provides the best results for most of the interpolation points (excluded sensors positions). The results show that some of the locations suffer higher interpolation errors, i.e. locations of excluded sensors with IDs 5 and 7 are mostly affected. The standard deviation of the AE at these sensors positions for the AMSMC approach is 3.74 dB and 3.44 dB, while for the classic IDW the respective values are 5.03 dB and 4.53 dB. The

error has a negative bias at these locations for all tested IDW methods. This is logical, considering Fig. 2 and the positions of the signal sources – the sensors with IDs 5 and 7 are the closest ones, receiving the highest signal power. When these points are excluded and the dynamic range of the field is high (high power transmissions), the interpolation surface "down-fits" the extremes. Figure 31 serves to clarify the causes of this effect. It represents the propagation model estimate in the inspected area, obtained with LS fitting of the parameters in the simplified propagation model function [40]. The estimated standard deviation of the shadowing was 4.46 dB. Referring to the problem of the high "down-fits" nearby the signal sources, these results prove that these areas are the most critical, since the power drop at distance of 1m is around 30 dB (Fig. 6). The IDW based methods are unable to estimate this extreme in the lack of a nearby observation.

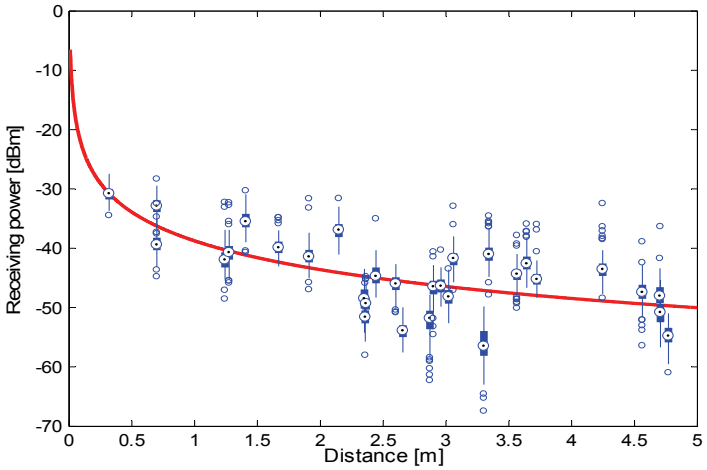

Figure 31: Propagation model estimate in the inspected area.

The results presented in this section serve to prove the applicability of the IDW based methods to the problem of spatial interpolation of RIFs with a low number of observations. Out of all methods, the AMSMC interpolation, with d_{exp} equal to 1 has proven to offer the lowest interpolation errors. However, the main concern of the IDW based spatial interpolation is the erroneous RIF estimation in nearby transmitter areas. A reasonable approach in these cases would be to use the measurements to perform an initial transmitter location and power estimation. This information can be fed to the IDW based interpolation to reliably synthesize the RIF.

6 Conclusions

The goal of this document was to describe the prototype implementation of all components that deal with the generation and evaluation of REMs.

The description started by presenting specific nuances of the spectrum sensing, data collection, data storage, data processing and data utilization, from a conceptual point of view that complements the general notion of REM utilization elaborated in [1][2]. Here the focus was on the data management aspects in terms of the prototype's architectural components.

A description of the developed supporting tools, needed for the generation and evaluation of REMs was also provided. Tools for accessing the quantitative characteristics of the actual process employed (the choice of the signal processing technique) for deriving and updating the stored information based on the raw measured data. They allowed the extension of our research efforts to techniques that are not compatible with the current version of the FARAMIR platform. These tools will be continuously updated, supporting new techniques and functionalities until the end of the project.

The algorithmic complexity and platform requirements of all techniques proposed for REM generation were also provided. Since one of the fundamental figures of merit of any actual implementation is the complexity of the components of the architecture and of the processing (an important parameter in assessing both the suitability of specific platforms as well as the ability to update the map dynamically with real-time updates), the deliverable has expended effort in assessing the complexity of each critical component in terms of processing requirement and memory, particularly for those modules that are expected to carry the brunt of the computing effort. A specific example of direct RIFE using the Gaussian-based Kriging interpolation method for an in-band femtocell scenario was also provided.

Finally, performance results based on field trials were provided. The measurements have been collected by the use of FARAMIR REM platform and involve indoor scenarios only. Techniques for single and multiple emitter localization, Tx-power estimation and RIFE were assessed and key issues that appear in real experiments were identified.

It should be clear that an actual implementation of a particular REM in real practice and for a very specific scenario will involve considerations not fully covered here. Still, the current and final (for this WP) deliverable does present a conceptually full and documented vista of the design landscape for REM's, including the broad outlines of all their elements, their information-exchange and storage needs and associated complexity, the computational needs and associated complexity and, finally, what is expected to come out of any such REM and with what quality.

Glossary and Definitions

Term	*Description*
3GPP	3rd Generation Partnership Project
AE	Absolute Error
AIC	Akaike Information Criterion
AOA	Angle Of Arrival
CFAR	Constant False Alarm Rate
CP	Cyclic Prefix
CRLB	Cramer-Rao Lower Bound Cyclic Prefix
CRM	Cognitive Resource Manager
DSA	Dynamic Spectrum Access
FOA	Frequency Of Arrival
GIDS	Gradient-plus-Inverse-Distance-Squared
HeNB	Home eNB
IDW	Inverse Distance Weighted
LMS	Least Mean Squares
LOS	Line-Of-Sight
LS	Least Squares
LSE	LS-Estimates
MAE	Mean Absolute Error
MCD	Measurement Capable Devices
MCS	Modulation and Coding Scheme
MDL	Minimum-Description Length
ML	Maximum Likelihood

Term	*Description*
MLE	Maximum Likelihood Estimator
MSM	Modified Shepard's Method
NOPE	Number Of Paths Estimation
OCHIR	Overall Channel Impulse Response
OLS	Ordinary Least Squares
OFDM	Orthogonal Frequency Division Multiplexing
PM	Policy Manager
POCS	Projection On Convex Sets
PSP	Per-Survivor Processing
PTS	Path Temporal Separation
REM	Radio Environmental Map
REM-SA	REM Storage and Acquisition unit
RIF	Radio-Interference Field
RIFE	Radio-Interference Field Estimation
RMSE	Root-Mean Squared Error
RRM	Radio Resource Management
RSS	Received Signal Strength
SDP	Semi-Definite Programing
SINR	Signal to Interference plus Noise Ratio
TDOA	Time Difference Of Arrival
TOA	Time Of Arrival
TVWS	TV White Spaces
White space	part(s) of spectrum allocated to a particular radio system (primary radio system) in particular location(s) that may be temporary unused by this primary radio system in some location(s) and thus allowed by

Term	*Description*
	radio regulations to be used by another radio system(s) (secondary radio system) on a temporary secondary basis without causing harmful interference to the primary radio system, where harmful interference and protection mechanisms are defined in the radio regulations

References

[1] FARAMIR Deliverable D2.4: "Final System Architecture", http://www.ict-faramir.eu/.

[2] FARAMIR Deliverable D4.1: " Radio Environmental Maps: Information Models and Reference Model ", http://www.ict-faramir.eu/.

[3] FARAMIR Deliverable D4.2: "Neighbourhood Mapping Technology and Algorithm Report", http://www.ict-faramir.eu/.

[4] FARAMIR Deliverable D2.2: "Scenario Definitions," http://www.ict-faramir.eu/.

[5] Qi Yihong, H. Kobayashi and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Trans. on Wireless Comm*, vol. 5, pp. 1536-1276, Mar. 2006.

[6] Richard K. Martin, Ryan W. Thomas, "Algorithms and bounds for estimating location, directionality, and environmental parameters of primary spectrum users," *IEEE Transactions on Wireless Comm*. vol 8, pp. 5692-5701, Nov. 2009.

[7] D. Shepard, "A two dimensional interpolation function for irregularly spaced data," *in proc. of the 23rd National Conf. of the Association for Computing Machinery,* Princeton, NJ, ACM, pp. 517–524, 1968.

[8] R. J. Renka, "Multivariate Interpolation of Large Sets of Scattered Data," *ACM Transactions on Mathematical Software*, vol. 14, no. 2, pp. 139-148, June, 1988.

[9] I. A. Nalder and R. W. Wein, "Spatial interpolation of climatic normals: Tests of a new methods in Canadian boreal forests," *Agricultiral and Forest Meteorology*, vol. 92, pp. 211-225, 1998.

[10] M. Sherman, "*Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*," Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, 2011.

[11] S. Athloen and R. McLaughlin, "Gauss-Jordan reduction: A brief history," American Mathematical Monthly 94, pp. 130-142, 1987.

[12] J.E. Gentle, "Gaussian Elimination," 3.1 in Numerical Linear Algebra for Applications in Statistics, Springer-Verlag, pp. 87-91, 1998.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "LU Decomposition and Its Applications," 2.3 in Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed., Cambridge University Press, pp. 34-42, 1992.

[14] V. Strassen, "Gaussian elimination is not optimal," *Numerische Mathematik 13*, pp. 354–356, 1969.

[15]D. Coppersmith and S. Winograd, ”Matrix multiplication via arithmetic progressions,” *J. Symbolic Computation 9*, pp. 251–280, 1990.

[16]Victor Y. Pan, Zhao Q. Chen, “The Complexity of the Matrix Eigenproblem,” STOC 1999: 507-516.

[17]D.S. Baum et.al., “An interim channel model for beyond-3g systems: extending the 3gpp spatial channel model (SCM),” Proc. VTC-Spring, 2005.

[18]S. Grimoud, B. Sayrac, S. Ben Jemaa and E. Moulines, “An Algorithm for Fast REM Construction,” Proc. CrownCom 2011.

[19]S. Grimoud, B. Sayrac, S. Ben Jemaa and E. Moulines, “Best Sensor Selection for an Iterative REM Construction,” accepted for publication in Proc.VTC-Fall 2011.

[20]A. Ben Hadj Alaya-Feki et.al., “Informed spectrum usage in cognitive radio networks: Interference Cartography (Invited Paper)”, Proc. IEEE CRNETS, PIMRC 2008.

[21]T.S. Rappaport, “Wireless Communications, Principles and Practice,” Prentice Hall, Uppler Saddle River, NJ, 2nd edition, 2002, chapter 4.

[22]S. Ohshima, K. Kise, T. Katagiri and T. Yuba, “Parallel Processing of Matrix Multiplication in a CPU and GPU Heterogeneous Environmen,” *in proc. VECPAR'2006*, pp. 305-318, 2006.

[23]J. L. Bentley, ”Multidimensional binary search trees used for associative searching,” *Communications of the ACM 18(9)*, pp. 509-517, 1975.

[24]B.V. Srinivasan, R. Duraiswami, R. Murtugudde, "Efficient kriging for real-time spatio-temporal interpolation,” *20th Conference on Probability and Statistics in the Atmospheric Sciences, American Meteorological Society*, Jan. 2010.

[25]N. Memarsadeghi, V. C. Raykar, R, Duraiswami, and D. M. Mount, “Efficient Kriging via Fast Matrix-Vector Products”, *in proc. IEEE Aerospace Conference*, pp. 1-7, 2008.

[26]S. Sakata, F. Ashidaa and M. Zakob, “An efficient algorithm for Kriging approximation and optimization with large-scale sampling data,” *Computer Methods in Applied Mechanics and Engineering*, Vol. 193, No. 3, pp. 385-404(20), Jan. 2004.

[27]R. Furrer, M. G. Genton and D. Nychka, “Covariance Tapering for Interpolation of Large Spatial Datasets,” *Journal of Computational and Graphical Statistics 15(3)*,pp. 502-523, 2006.

[28]M. Gudmundson, “Correlation Modle in Shadow Fading systems,” in *Electronic Letters, vol. 27,* Nov. 1991

[29]N.Miliou, A.Moustakas, A.Polydoros, “Interference Source Localization and Transmit Power Estimation Under Log-normal Shadowing,” in Proceedings of Eusipco 2011, Barcelona, Spain, August 2011.

[30] Kirkpatrick, S.; C. D. Gelatt, M. P. Vecchi (1983-05-13), "Optimization by Simulated Annealing," *Science*. New Series 220 (4598): 671–680. doi:10.1126/science.220.4598.67

[31] D. Bertsimas and J. N. Tsitsiklis, "Simulated Annealing", *Statistical Science*, Vol. 8, No. 1, 1993, pp. 10-15.

[32] C. Sachin, L. Jarmo, K. Visa, "Collaborative Autocorrelation-Based Spectrum Sensing of OFDM signals in Cognitive Radios," 42nd Annual Conference on Information Sciences and Systems 2008, Princeton, NJ, USA, Mar. 2008.

[33] M. C. Vanderveen, B. C. Ng, C. B. Papadias, and A. Paulraj, "Joint angle and delay estimation (JADE) for signals in multipath environments," in Proc. 30th Asilomar Conf. Circuits, Syst. Comput., Pacific Grove, CA, vol. 2, pp. 1250-1254, Nov. 1996.

[34] Prokopios Panagiotou, Andreas Polydoros, Phunsak Thiennviboon, "Multiple-path Direction-of-Arrival estimation for cognitive radio sensing", PIMRC 2010, pp. 791-796.

[35] C. Sachin, L. Jarmo, K. Visa, "Collaborative Autocorrelation-Based Spectrum Sensing of OFDM signals in Cognitive Radios," 42nd Annual Conference on Information Sciences and Systems 2008, Princeton, NJ, USA, Mar. 2008.

[36] Y. Censor and S. A. Zenios, "Parallel Optimization, Theory, Algorithms, and Applications," Oxford University press, 1997.

[37] D. Blatt and A. O. Hero, "Sensor network source localization via projection onto convex sets (POCS)," In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA, Mar. 2005.

[38] D. Blatt and A. O. Hero, "APOCS: a rapidly convergent source localization algorithm for sensor networks," In Proc. IEEE/SP 13th Workshop on Statistical Signal Processing, pp. 1214–1219, Jul. 2005.

[39] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving Euclidean distance matrix completion problems via semidefinite programming," Comput. Optim. Appl., vol 12, Nov. 1999.

[40] M. Laurent, "Matrix completion problems," in collection of Encyclopedia of Optimization, 2009.

[41] A. Goldsmith, "Wireless Communications", Cambridge University Press, 2005.