Efficient Identification of a Multi-Objective Pareto Front on a Wireless Experimentation Facility

Michael Tetemke Mehari, Eli De Poorter, Ivo Couckuyt, Dirk Deschrijver, *Senior Member, IEEE*,
Günter Vermeeren, David Plets, Wout Joseph, *Senior Member, IEEE*,
Luc Martens, *Member, IEEE*, Tom Dhaene, and Ingrid Moerman

Abstract—Wireless systems often need to optimize multiple conflicting objectives (low delay, high reliability, and low cost), which are difficult to fulfill simultaneously. In such cases, the wireless system exhibits multiple optimal operation points, referred to as the optimal Pareto front (OPF). However, due to the large number of parameter settings to be evaluated and the time-consuming nature of performing wireless experiments, it is typically not possible to identify the OPF by exhaustively evaluating all possible settings. Instead, for many use cases, an approximation is good enough. To this end, this paper applies a multi-objective surrogate-based optimization (MOSBO) toolbox to efficiently optimize wireless systems and approximate the OPF using a limited number of iterations. Moreover, a real Wi-Fi conferencing scenario is optimized that has two conflicting objectives (exposure and audio quality) and four configurable parameters (Tx-Power, Tx-Rate, Codec Bit-Rate, and Codec Frame-Length). The benefits of using the MOSBO approach for such a network problem is demonstrated by approximating the OPF using 94 iterations instead of requiring the exploration of 6528 different parameter combinations, while still dominating 96.58% of the complete design space.

Index Terms—Multi-objective Pareto optimization, MOSBO, exposure, audio quality, wireless systems and experimentation.

I. INTRODUCTION

T HE introduction of wireless systems, replacing the legacy wired systems, created a wide range of opportunities. Nowadays, industrial environments are equipped with wireless sensors in places where it is difficult to put wired connections for temperature and pressure readings. It is also very cheap to deploy cellular networks these days rather than installing expensive copper wirings at customers' premises. Such application areas are pushing the need for wireless systems which otherwise are difficult to be realized by wired systems or impractical in some areas.

On the other hand, the opportunities that were envisioned in wireless systems are challenged by the need to optimize multiple conflicting objectives. Wireless surveillance systems, for example, target the highest video quality whilst utilizing a minimal bandwidth. Wireless conferencing systems also strive for realizing the best audio quality but simultaneously aim to limit their wireless exposure. Another example is Automated Guided Vehicles (AGVs) working in a large factory hallway. While robustness is typically a first priority, communication between AGVs needs to be secure which also demands higher network utilization during frequent roaming between access points. Such wireless systems, showing conflicting objectives, exhibit multiple non-dominated operating points (i.e. parameter settings) which in literature is referred to as the Optimum Pareto Front (OPF) [1], [2]. The OPF is a set of performance objectives which cannot be further improved by any other parameter combinations without affecting at least one objective. For designers of wireless systems, it is typically not possible to identify the OPF because it requires an exhaustive search of the parameter space which tends to be time intensive and sometimes impossible. Instead, an approximation is good enough. To this end, an Approximate Pareto Front (APF) [2] is calculated using a number of multi-objective optimizers such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II [3]), the S-Metric Selection Evolutionary MultiObjective Algorithm (SMS-EMOA [4]) and the Strength Pareto Evolutionary Algorithm 2 (SPEA2 [5]).

Even though all of these are targeted at multi-objective optimization, they are typically not designed to minimize the number of iterations. In fact, most evolutionary based multiobjective variants (i.e. NSGA-II, SPEA2, SMS-EMOA, ...) inherently require a large number of iterations to identify the APF. On the other hand, evaluating real-life wireless systems is time-intensive since each experiment requires resource deployment, configuration, execution and evaluation of the wireless experiment. For example, when using the Orbit Management Framework (OMF) for experimentation control, an experiment having N wireless nodes adds an average delay of 5.17*N ms on a single message orchestration [6]. As such, most designers of wireless systems aim to approximate the OPF using a limited number of experiments, specially when relying on real-life deployments rather than simula-

tions. As a solution, this paper considers a Multi Objective Surrogate-Based Optimization (MOSBO) approach [7]. MOSBO makes use of Kriging models and Hypervolume based Probability of Improvement (HV-PoI) to economize experiment runs and approximate the OPF using a limited number of experiments.

The paper introduces the following novel contributions.

- An architecture for integrating a MOSBO optimizer in realistic conditions for approximating the OPF of complex wireless systems using a limited number of experiments
- The introduction of advanced optimization objectives, including (i) an advanced calculation of end-to-end audio quality Mean Opinion Score (MOS) and (ii) an exposure Specific Absorption Rate (SAR) metric for Wi-Fi traces.
- An analysis of the influence of different MOSBO design criteria, such as the stopping criteria and initial sample size, on the overall performance of the system.
- An experimental validation of the overall multi-objective optimization through a large-scale Wi-Fi conferencing system using an experimental testbed facility.

The remainder of this paper is structured as follows. Section II presents a literature survey of multi-objective optimization solutions in wireless systems. Next, the MOSBO optimizer is presented in Section III. Section IV introduces a Wi-Fi conferencing scenario which is used to validate the suitability of the MOSBO optimizer. The results from the Wi-Fi conferencing experiments are discussed in Section V. Finally, conclusion and future work are presented in Section VI.

II. RELATED WORK

During the validation section, a Wi-Fi conferencing scenario will be considered that aims to optimize two conflicting objectives: improving audio quality (e.g. using MOS score) while reducing the transmission exposure (expressed as SAR). This section will discuss aspects related to the conflicting objectives that will be investigated as well as the state of the art in multi-objective optimization of wireless networks.

A. Electromagnetic Exposure

Due to the increased use of wireless technologies, an increasing amount of attention is given to the impact of electromagnetic radiation on the human body. In exposure assessment, one distinguishes between (i) compliance testing and (ii) realistic exposure assessment. The former evaluates if worst-case exposure situations comply with exposure limits and is either uplink (transmission from a wireless end device to a base station) or downlink (transmission from a base station down to a wireless end device) focused. The exposure metrics used in compliance testing are incident electric and magnetic field, incident power density, SAR, etc. The latter evaluates the exposure of a person under realistic exposure conditions which is of interest in epidemiological studies and often combines both uplink and downlink exposures together. For analyzing and optimizing wireless deployments, the realistic exposure assessment is important [8], especially when co-optimizing the exposure with other performance criteria. In [9], the authors

address this concern by creating a network planning tool to jointly optimize transmission exposure and coverage. Furthermore, the author in [10] decomposes exposure into uplink and downlink and applied a joint minimization to lower the total human exposure dose. A more realistic approach by using a wireless testbed is proposed by [8]. The paper evaluated the Exposure Index of an LTE data scenario implemented in a real urban area (part of the 7th district of Paris) and quantifies the total exposure of a population in the area.

B. Audio Quality

In contrast to methods for calculating the exposure in Wi-Fi networks, the field of calculating audio quality objective is rather mature and sufficient research has already been carried out. One way of improving the audio quality in streaming applications is by allowing dynamic source rate adaptation. The authors in [11] presented this concept such that by using RTCP receiver reports to understand the Network condition, the bandwidth of the source audio is adjusted and the quality of the receiver is improved with continuous delivery and lower packet loss. On another level, a method for optimizing the audio quality of a VOIP application (using MOS score) in a WiMAX network is presented in [12]. Furthermore, the author in [13] extended this work towards multiple wireless technologies by including Wi-Fi and LTE along with WiMAX and optimized three different audio codecs while searching for the best audio quality. In [14], Quality of Experience (QoE) of an audio in wireless networks is guaranteed by jointly optimizing application layer and lower layer networking parameters. Looking into handover performance and the influence on audio quality, the authors in [15] have proposed the use of a MOS-based handover scheme over the traditional Received Signal Strength (RSS) based handover scheme. The numerical results show that the MOS-based handover scheme maintains high call quality and reduce the probabilities for both handover dropping and call dropping.

C. Multi-Objective Optimization in Wireless Networks

As discussed in the introduction, wireless networks typically exhibit a wide range of conflicting objectives. In the field of antenna design, for example, there exists a significant amount of work on Pareto front analysis and multi-objective optimization in relation to field/electrical and geometrical properties. The authors in [16] and [17] have used Response Surface Approximation (RSA) models and sequential domain patching to optimize the geometry of a compact DRA antenna and a planar monopole antenna respectively. Surrogate modelling tools are also applied in antenna design to cut down the time intensive operation [18], [19]. In wireless protocol stacks, however, the use of Pareto front analysis and multi-objective optimization is quite limited. As such, this section gives an overview of multi-objective optimization approaches that have previously been applied to wireless protocols.

Throughput and outage probability, for example, are two conflicting objectives described in [20]. Maximizing the throughput of the first network counteracts the outage probability of the second network, which depends on the received

interference due to the first network activity. In their work, an analytical framework is used to simulate and find the OPF of the two wireless networks by applying a channel coding scheme. Another Pareto front optimization on intrusion detection accuracy of wireless sensor networks and memory consumption is considered by Stehlk et al. [21]. NSGA-II and SPEA2 optimizers were used to locate the APF of the design space composed of 25 million parameters. The inherent large design space demands the use of evolutionary algorithms since they work in sample batches of the population and their computational demand is relatively simple. Looking into a multi-user Orthogonal Frequency Division Multiplexing (OFDM) system, a Pareto front of carrier capacity and power consumption is constructed by using Multi-Objective Particle Swarm Optimizer (MOPSO) [22]. Even though they compare the performance of a modified MOPSO with NSGA-II, the exceptionally low computational power required by the MOPSO limits its usability for computational expensive wireless network problems. Looking into cellular networks, the authors in [23] optimized the throughput sum power of base and relay stations while guaranteeing an end to end Signal-to-Noise Ratio (SNR) metric. A more recent work on multi-layer parameter optimization of Wireless Sensor Networks is presented by [24]. The authors have collected an extensive list of performance metrics (i.e. energy, throughput, delay and loss) out of 7 multi-layer parameters (i.e. packet inter-arrival time and payload size from application layer, maximum queue size, number of transmission and retry delay from MAC layer, transmission power level and distance between nodes from PHY layer). In total, around 50 thousand parameter configurations were experimented within 6 month duration which shows how exhaustive searching can be a tedious and time consuming task.

In general, most of the wireless problems considered in literature use simulations with low computational complexity for approximating the OPF. However, for experimental and time intensive problems, the above solutions become impractical. To the best of our knowledge, this paper is the first to experimentally optimize and analyse the Pareto front of exposure and audio quality objectives from an experimental Wi-Fi conferencing set-up.

III. MOSBO

Surrogate-based Optimization (SBO) methods have proven themselves to be effective in solving complex optimization problems, and are increasingly being used in different fields [18], [25]–[27]. Unlike multi-objective evolutionary algorithms such as NSGA-II [3], SMS-EMOA [4] and SPEA2 [5], surrogate-based methods typically require very few experiment iterations to converge. This feature makes surrogate-based methods very attractive for solving optimization problems that require time-consuming wireless experimentation.

A. Efficient Multi-Objective Optimization (EMO)

The expected improvement and Probability of Improvement (PoI) criteria are widely used for single-objective
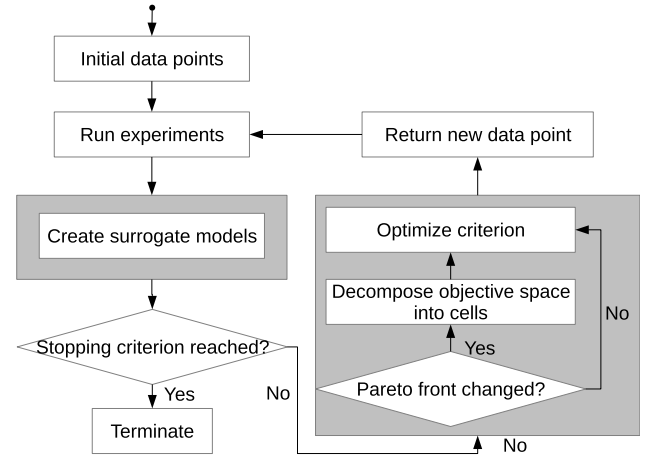

Fig. 1. Flow chart of the Efficient multi-objective Optimization (EMO) algorithm [7].

optimization such as electromagnetic and aerodynamic problems [19], [28]. Recently, multi-objective versions of these criteria are increasingly being used to solve complex multi-objective problems [7], [29]. While they have been used in SBO schemes, due to the computational requirements, their applicability in practice has been limited to problems of 2 objectives. The recently introduced EMO algorithm [7] provides an efficient computation method and can be applied to problems up to 7 objectives.

A flowchart of the EMO algorithm is shown in Figure 1. The algorithm begins with the generation of an initial set of experiments X corresponding to different settings $\mathbf{x}$ of the network parameters. These initial configurations are executed on the testbed, in order to evaluate corresponding values of the objectives $f_j(\mathbf{x})$, for $j = 1 \ldots m$. Each objective function $f_j(\mathbf{x})$ quantifies a QoS performance characteristic and is approximated by a Kriging surrogate model. Based on the models, useful criteria can be constructed to identify new configurations of network parameters that likely improve the currently identified Pareto set $\mathcal{P}$ of Pareto-optimal solutions. As such, these criteria are used to define a new experiment (i.e., a point) in the parameter space, which is again executed on the testbed to evaluate the expensive QoS objective functions $f_j(\mathbf{x})$. The models are then updated with this new information and this process is repeated in an iterative manner until a predefined stopping criterion is met.

This paper adopts the hypervolume-based PoI criterion. It is important to note that the computation of these criteria requires a prediction of the modeling uncertainty. Hence, the choice of surrogate model is limited to those which can provide the uncertainty of the prediction (such as e.g. Kriging).

B. Kriging

Kriging models are very popular in the optimization of complex systems [30]. This is partly due to the fact that Kriging models provide the mean and prediction variance which can be exploited by statistical sampling criteria. Their popularity also stems from the fact that many implementations are widely available [31]–[33].

Assume that a set of n samples $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ in d dimensions having the target values $\mathbf{y} = (y_1, \ldots, y_n)'$ is given.
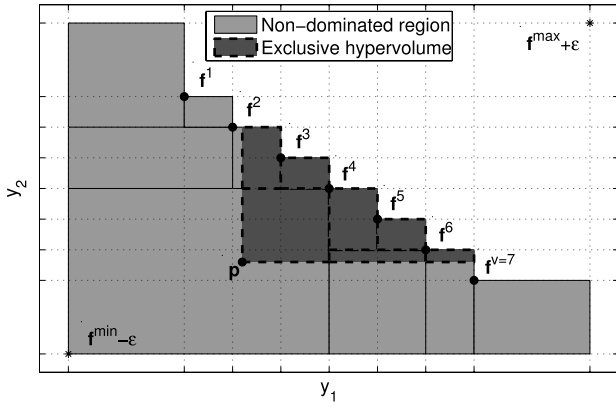
Fig. 2. A Pareto set for two objectives consisting of Pareto points f^i, for $i = 1 \ldots v$. f^{min} and f^{max} denote the ideal and anti-ideal point respectively [7]. The exclusive hypervolume expands the dominated region of the Pareto front (white region enclosed between f^{max} and the Pareto front) which at the same time shrinks the non-dominated region.

The prediction mean and prediction variance of Kriging are then derived, respectively, as,

$$\hat{y}(\mathbf{x}) = \alpha + r(\mathbf{x}) \cdot \Psi^{-1} \cdot (\mathbf{y} - \mathbf{1}\alpha) \tag{1}$$

$$s^2(\mathbf{x}) = \sigma^2 \left(1 - r(\mathbf{x})\Psi^{-1}r(\mathbf{x})^\top + \frac{(1 - \mathbf{1}^\top \Psi^{-1}r(\mathbf{x})^\top)}{\mathbf{1}^\top \Psi^{-1}\mathbf{1}} \right) \tag{2}$$

where $\mathbf{1}$ is a vector of ones, α is the coefficient of the constant regression function, determined by Generalized Least Squares (GLS), $r(\mathbf{x})$ is a $1 \times n$ vector of correlations between the point $\mathbf{x}$ and the samples X, and $\sigma^2 = \frac{1}{n}(\mathbf{y} - \mathbf{1}\alpha)^\top \Psi^{-1}(\mathbf{y} - \mathbf{1}\alpha)$ is the variance.

Ψ is a $n \times n$ correlation matrix of the samples X,

$$\Psi = \begin{pmatrix} \psi(\mathbf{x}_1, \mathbf{x}_1) & \ldots & \psi(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ \psi(\mathbf{x}_n, \mathbf{x}_1) & \ldots & \psi(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix},$$

with ψ being the correlation function. The correlation function greatly affects the accuracy of the Kriging model and in this paper the Matérn correlation function [34] with $v = \frac{3}{2}$ is used,

$$\psi(\mathbf{x}_a, \mathbf{x}_b)_{v=\frac{3}{2}}^{Matérn} = \left(1 + \sqrt{3}l\right) \exp\left(-\sqrt{3}l\right),$$

with $l = \sqrt{\sum_{i=1}^{d} \theta_i (x_a^i - x_b^i)^2}$. The hyperparameters θ are identified using Maximum Likelihood Estimation (MLE).

1) Hypervolume-Based Probability of Improvement: In a multi-objective setting the improvement I over the current Pareto set $\mathcal{P}$ can be defined in several ways. The hypervolume metric (or $\mathcal{S}$-metric) [35] is often used to evaluate the goodness of the Pareto set. The hypervolume indicator $\mathcal{H}(\mathcal{P})$ denotes the volume of the region in the objective space dominated by the Pareto set $\mathcal{P}$, bounded by a reference point $f^{max} + \varepsilon$, where f^{max} denotes the anti-ideal point.

A better Pareto set has a higher corresponding hypervolume $\mathcal{H}(\mathcal{P})$. The contributing hypervolume $\mathcal{H}_{contr}(\mathbf{p}, \mathcal{P})$ of a Pareto set $\mathcal{P}$ relative to a point $\mathbf{p}$ (see Fig. 2) is defined as,

$$\mathcal{H}_{contr}(\mathbf{p}, \mathcal{P}) = \mathcal{H}(\mathcal{P} \cup \mathbf{p}) - \mathcal{H}(\mathcal{P}), \tag{3}$$
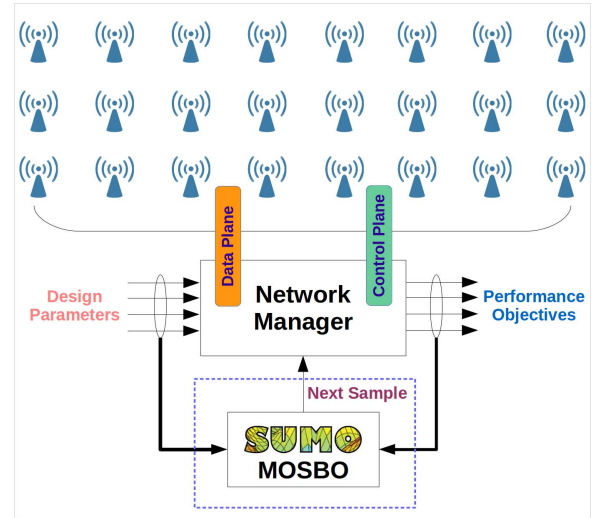

Fig. 3. Generalized architecture of a wireless network showing the wireless nodes, the network manager and the associated data/control planes.

$\mathcal{H}_{contr}$ measures the contribution (or improvement) offered by the point $\mathbf{p}$ over the Pareto set $\mathcal{P}$ and can be used to define a scalar improvement function I as,

$$I(\mathbf{p}, \mathcal{P}) = \begin{cases} \mathcal{H}_{contr}(\mathbf{p}, \mathcal{P}) : \mathbf{p} \text{ is not dominated by } \mathcal{P} \\ 0 \qquad\qquad : \text{otherwise.} \end{cases} \tag{4}$$

Let $y_j = f_j(\mathbf{x})$, $\hat{y}_j(\mathbf{x})$ be the prediction mean, and $s_j^2(\mathbf{x})$ be the prediction variance of a given surrogate model associated with the jth objective, then a Gaussian probability density function ϕ_j with mean $\hat{y}_j(\mathbf{x})$ and variance $s_j^2(\mathbf{x})$ is defined as

$$\phi_j[y_j] \triangleq \phi_j\big[y_j; \hat{y}_j(\mathbf{x}), s_j^2(\mathbf{x})\big], \tag{5}$$

In this paper, $\mathcal{H}_{contr}$ is used as the hypervolume contribution for I to compute the hypervolume-based probability of improvement (PoI) [7]. The hypervolume-based PoI can be written as the product of the improvement function $I(\hat{y}, \mathcal{P})$ and the multi-objective PoI $P[I]$,

$$P[I] = \int_{\mathbf{y} \in A} \prod_{j=1}^{m} \phi_j[y_j] dy_j, \tag{6}$$

$$P_{hv}[I] = I(\hat{\mathbf{y}}, \mathcal{P}) \cdot P[I], \tag{7}$$

where $\hat{\mathbf{y}} = (\hat{y}_1(\mathbf{x}), \ldots, \hat{y}_m(\mathbf{x}))$ is a vector containing the prediction models of each objective function for a point $\mathbf{x}$. The integration area A of $P[I]$ corresponds to the non-dominated region. The reader is referred to [7] for further details.

C. Integration of MOSBO in Network Architectures

Even though the MOSBO optimizer has been quite popular for optimizing electromagnetic and antenna designs, it has not yet been used in the scope of wireless networking. This section discusses how the MOSBO optimizer can be integrated in wireless network applications.

A general wireless network is shown in Figure 3, and consists of individual wireless nodes, a network manager and a control and data plane. The network manager utilizes the data plane to send or receive data to and from the
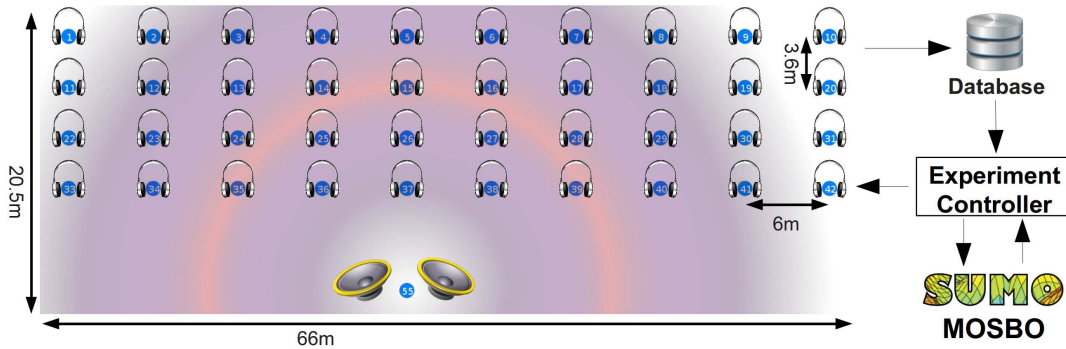
Fig. 4. Top level view of the Wi-Fi conferencing set-up mapped onto the wireless testbed. Listener nodes are located on the first 4 rows (nodes 1-20, 22-31 and 33-42) and the speaker node is located on the bottom center (node 55).

wireless devices. In addition, there are a number of design parameters and performance metrics that can be configured or retrieved on the devices of the wireless network using the control plane, typically using SNMP or similar network management protocols. In traditional networks, configuration parameters are often configured using safe (but non-optimal) settings, or configured based on predictions from non-realistic simulations. To find optimal settings, usually an exhaustive search of all design parameters is required. In contrast, the network manager can make use of optimization tools so that the time consuming part of the exhaustive search can be traded off with a minor quality degradation. To this end, MOSBO can be used in wireless networks to locate the optimum design settings within a relatively short amount of time. In order to do so, the MOSBO optimizer is connected to the network manager as shown in the Figure 3 (shown inside the broken rectangle). In the learning/exploration phase, the MOSBO optimizer builds a surrogate model out of the design parameters and their performance objectives. After that, MOSBO starts optimizing the system from the constructed model and provides new design parameters to be configured on the wireless network. This way, the MOSBO optimizer is integrated in wireless networks and near-optimum performance is realized within a relatively short duration of time.

IV. EXPERIMENTAL VALIDATION

To evaluate the suitability of the MOSBO approach in wireless problems, a Wi-Fi conferencing scenario is validated. This section outlines the scenario, the input parameters and the performance objectives of the Wi-Fi conferencing scenario.

A. Experiment Scenario

In a multilingual conferencing session, a speaker's voice is translated into different languages and streamed to listeners. Such an application is typically used in inter-country meetings where different people use different languages to communicate each other with the help of translators. Usually such a conferencing system relies on a wired network and scalability is often a challenge or building a new system is time consuming. The counter part, a Wi-Fi conferencing system, is used in this paper since it addresses the aforementioned challenges. Compared to the multilingual audio conferencing system, a Wi-Fi conferencing system broadcasts the translated

TABLE I
EXPERIMENT RESOURCE DESCRIPTION: HARDWARE COMPONENTS AND SOFTWARE TOOLS

#	Resource	Description
1	Wi-Fi nodes	ZOTAC NM10-A-E
2	Wi-Fi chipset	Atheros Sparklan WPEA-110N/E/11n mini PCI
3	Wi-Fi driver	ath9k
4	OS	Ubuntu 14.04 LTS
5	kernel	Linux 3.13.0-33-generic
6	Optimizer	MOSBO

TABLE II
INPUT PARAMETERS OF THE WI-FI CONFERENCING EXPERIMENT. THE DESIGN SPACE CONSISTS OF 6528 ($32 \times 3 \times 4 \times 17$) ELEMENTS

#	Input parameters	Range
1	Opus codec Bit-Rate	[6400, 7200, 8000, ..., 31200] bps
2	Opus codec Frame-Length	[20, 40, 60] msec
3	Wi-Fi Tx-Rate	[6, 12, 18, 24] Mbps
4	Wi-Fi Tx-Power	[0, 1, 2, 3, ..., 16] dBm

audio stream via a Wi-Fi channel and the listeners pick and play it through a wireless headset.

Figure 4 shows the Wi-Fi conferencing set-up which is composed of a speaker node transmitting an audio signal on Wi-Fi channel 1 (2412 MHz center frequency), 40 listener nodes receiving the audio signal, a central database collecting the measurement data, a MOSBO optimizer optimizing design parameters and an Experiment Controller (EC) orchestrating the experiment. The hardware components and software tools used in the Wi-Fi conferencing set-up are shown in Table I.

In order to realize the Wi-Fi conferencing set-up, the *iMinds w-iLab.t* wireless testbed [36] is used. The *iMinds w-iLab.t* wireless testbed is equipped with heterogeneous devices such as embedded PCs (having Wi-Fi, Zigbee and Bluetooth technologies), Long Term Evolution (LTE) femtocells/UE dongles, advanced spectrum sensing devices (i.e. Universal Software Radio Platform (USRP), IMEC Sensing Engines, and Wireless open Access Research Platform (WARP) boards) and roomba robots to facilitate mobility experiments.

B. Input parameters

While transmitting the audio signal, the speaker node can dynamically adapt four configurable parameters (i.e. Wi-Fi Tx-Power, Wi-Fi Tx-Rate, Codec Bit-Rate and Codec Frame-Length). Table II provides the description and ranges of each
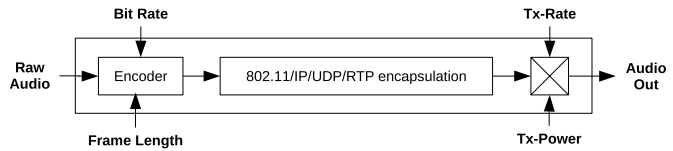
Fig. 5. Audio transmit path of the Wi-Fi conferencing experiment. Four configurable parameters are available: codec bit rate, codec frame length, Wi-Fi Tx-rate and Wi-Fi Tx-power.

input parameter. The speaker's audio transmit path, indicating all input parameters, is shown in Figure 5. A raw audio file is given to an encoder unit which outputs compressed audio frames of a given bit rate and frame length using an opus compression format [37]. Opus is a highly versatile audio codec providing a wide range of bit rates and frame sizes which work from narrow band to full band audio frequencies. It also has a very low latency compared to other audio codecs which makes it attractive for real-time applications. Afterwards, the opus encoded frame is encapsulated, rate and power adjusted before sent over the air.

C. Performance Objectives

The scenario is optimized towards two objectives: increasing the audio quality and reducing the transmission exposure. The main reason for selecting the two objectives is the conflicting influence each configurable setting has on their performance: all settings from Table II influence both audio quality and exposure objectives. Increasing the codec Bit-Rate of an audio signal, for example, will improve the audio quality but negatively impacts the exposure since more packets need to be transmitted. Similar conflicting influences are present for the Tx-Rate, codec Frame-Length and Tx-Power parameters. Although it is easy to predict that all settings from Table II will influence both objectives, without performing the experiments it is not possible to predict which of these settings will have the largest influence on each of the objectives and which combination of settings will result in optimal performances. This situation is very typical for many wireless systems: engineers typically have a-priori domain knowledge about expected impacts of parameter settings, but this knowledge does not suffice to identify the exact trade-offs and optimal settings. The next sections describe in more detail how each of the objectives are calculated.

1) Audio Quality: As a quantitative measure to evaluate audio quality, MOS scores are often used. A MOS score represent audio quality using a scale of 1 to 5 [38], with 5 being the best and 1 being the worst. In wireless networks, the audio quality typically degrades due to (i) the encoder settings and (ii) the transmission of the audio over the lossy medium. Most MOS score calculations do not differentiate between these two different influences. As such, a new formulation of the MOS score that takes into account both the degradation from the encoder settings and the network settings is shown in Figure 6. In this formulation, the audio quality (MOS) of the Wi-Fi conferencing scenario is calculated twice: first after the encoder unit and later after the wireless transmission.

In an earlier work [39], the audio quality degradation over a wireless medium has been discussed. Starting from a reduced
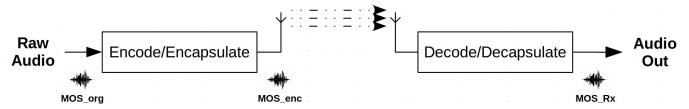

Fig. 6. MOS calculation flowchart. The audio quality degradation is calculated in two phases: once after the encoder unit and again after the wireless transmission.

TABLE III
AUDIO BANDWIDTH AND EFFECTIVE SAMPLE RATE
OF DIFFERENT AUDIO CLASSES

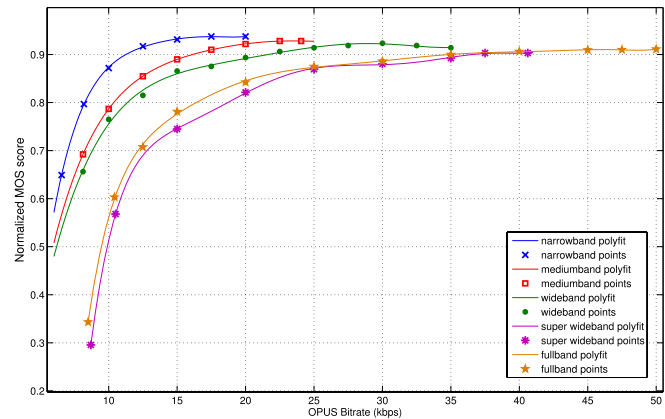| Audio class	Bandwidth	Sample rate
NB (narrowband)	4 kHz	8 kHz
MB (medium-band)	6 kHz	12 kHz
WB (wideband)	8 kHz	16 kHz
SWB (super-wideband)	12 kHz	24 kHz
FB (fullband)	24 kHz	48 kHz

Fig. 7. Normalized OPUS MOS scores as a function of bitrate for different audio classes.

audio quality after the encoder unit (MOS_enc), the audio is impacted by transmission latency, jitter and packet loss which results in the received audio quality (MOS_Rx).

$$MOS\_Rx = f(MOS\_enc, latency, jitter, packetLoss) \quad (8)$$

Whereas inside the encoder unit, a quality loss is introduced which is a function of the original audio quality, encoder bit rate, type of encoder and audio class used.

$$MOS\_enc = f(MOS\_orig, bitrate, type, class) \quad (9)$$

Unlike the calculation of MOS degradation over a wireless medium, most of the work done to characterize the encoder losses is through subjective tests by using humans evaluating the quality of an encoder output at different bitrates and for different audio samples [37], [40], [41], [42]. Since subjective testing is not possible in an automated wireless system, in this work the Perceptual Objective Listening Quality Assessment (POLQA) method [43] is applied. POLQA is a digital speech analyser model which compares the original and the degraded audio samples and calculates the perception difference using the tradition MOS scale. POLQA is also fully automated and as such is an ideal candidate for the Wi-Fi conferencing scenario. The authors of [44] have evaluated the POLQA estimator by applying speech samples to the opus encoder at different sample rates. Because the audio

$$MOS\_enc(MOS\_orig, bitrate, class) = MOS\_norm(bitrate, class) * (MOS\_orig - 1) + 1 \tag{10}$$

$$EI^{SAR} = \frac{1}{T} \sum_{t,p,e,r,c,l,pos}^{N_T, N_p, N_E, N_R, N_C, N_L, N_{pos}} f_{t,p,e,r,l,c,pos} \left[\sum_u^{N_U} \left(d^{UL}\overline{P}_{TX} \right) + d^{DL}\overline{S}_{inc} \right] \left[\frac{W}{kg} \right] \tag{11}$$

sample rate inherently impacts the audio bandwidth, this also means that the POLQA estimator has already been applied for different classes of audio as shown in Table III.

Since our goal is to calculate the MOS score after the encoder unit, we make the assumption that the reduced audio quality is independent of the original input quality but only on the encoder output bitrate. Therefore the different OPUS POLQA curves from [44] are normalized and the resulting output is shown in Figure 7. The performance curves in Figure 7 are not calculated from a single audio source but rather from 5 different audio speeches whose bandwidth utilization is according to their audio classes. If a wideband audio signal is to be encoded, for example, then the green curve is used to get a reasonable quality estimation after the encoder unit. Afterwards, the absolute MOS score is calculated first by scaling and later by translating the normalized MOS score as shown in Equation 10. The source audio file [45] used in this paper has a wideband audio class and a quality score of MOS_orig = 4.75.

2) Transmission Exposure: For objectively assessing the transmission exposure, recently Varsier et al. [8] defined a new metric, the Exposure Index (EI), which aggregates downlink and uplink exposure data and quantifies the total exposure of a population in an area. Assuming negligible uplink exposure from other users which are not part of the system under test, the EI in SAR units is given by Equation 11 where t is the period within the considered time frame T, p is the population category, e is the environment, r is Radio Access Technology (RAT), c is the cell type, l is the user load profile, pos is the posture, u is usage of the device, d^{UL} is the uplink dose in units of W/kg for 1W of transmitted power, $\overline{P}_{TX}$ is the average transmitted power by the mobile device, d^{DL} is the downlink dose in units of W/kg for 1 W/m^2 of received power density, $\overline{S}_{inc}$ is the average received incident power density and f is the fraction of the population p.

This exposure formulation is a generalization of the different possibilities that a person can be exposed from a wireless transmission. However, the SAR calculation in [8] is described mainly from a theoretical point of view, and was (i) never defined for real Wi-Fi traces and (ii) was never calculated in real-life using off-the-shelf radios. As such, for experimentally measuring the exposure of a Wi-Fi conferencing set-up, a new metric is needed that can be derived using off-the-shelf commercial Wi-Fi chips (in contrast to the use of dosimeters which are frequently used in exposure research). Applying the formula to the Wi-Fi conferencing set-up with a single population category, a homogeneous environment, a Wi-Fi Radio Access Technology, an access point cell type, an audio broadcasting load profile, a standing posture and a

single purpose device usage, the formula simplifies to

$$EI^{SAR} = \frac{1}{T} \left[\sum_t^{N_T} \left(d^{UL}\overline{P}_{TX} \right) + d^{DL}\overline{S}_{inc} \right] \left[\frac{W}{kg} \right] \tag{12}$$

Unlike Equation (11), shown at the bottom of the page. the calculation of $\overline{P}_{TX}$ and $\overline{S}_{inc}$ in Equation (12) lead to an exact assessment of the EI (with known locations and wireless parameters) for the speaker and listener nodes respectively. For every transmitted and received packets, the speaker and the listener nodes calculate the time duration a packet has occupied the wireless medium (the "Channel Occupation Time (COT)"). During the COT amount of time, the speaker antenna next to a speaker induces a SAR that is proportional to the transmitted power $\overline{P}_{TX}$ and the speaker antenna also induces a SAR that is proportional to the incident power density $\overline{S}_{inc}$ at the listeners. After that, the electromagnetic energy absorption per kilogram of body mass is calculated by applying the uplink and downlink absorption parameters $d^{UL} = 0.0070$ W/kg for 1W of transmitted power and $d^{DL} = 0.0028$ W/kg for 1 W/m^2 of received power density respectively [10]. Finally, the average exposure (power per kilogram of body mass) is calculated by summing all energy absorptions for every transmitted and received packets and dividing the result by the time duration T.

V. RESULT AND DISCUSSION

In this section, the results from the Wi-Fi conferencing optimization experiment are discussed. First, the behavior of the system is analyzed by creating an exhaustive search model that includes all possible parameter combinations. The next sections analyze the performance of the MOSBO optimized system, by calculating the computational overhead, analyzing the impact of different stopping criteria, and investigating the impact of the initial sample size.

A. Exhaustive Search Model

The exhaustive search model is a plot of the objective performances using every input parameter combination. Performing an exhaustive search is not feasible in most realistic situations due to the large number of experiments that need to be performed (6528 experiments in our case, see Section IV-B). However the outcomes of an exhaustive search experiment are included in order to compare with the time-efficient MOSBO experiments.

Figure 8(a) shows the results of the exhaustive search model. The OPF is calculated by selecting the non-dominated elements from the exhaustive search experiment and is indicated by a red line. By varying individual design parameters, interesting relations between the input parameters and the
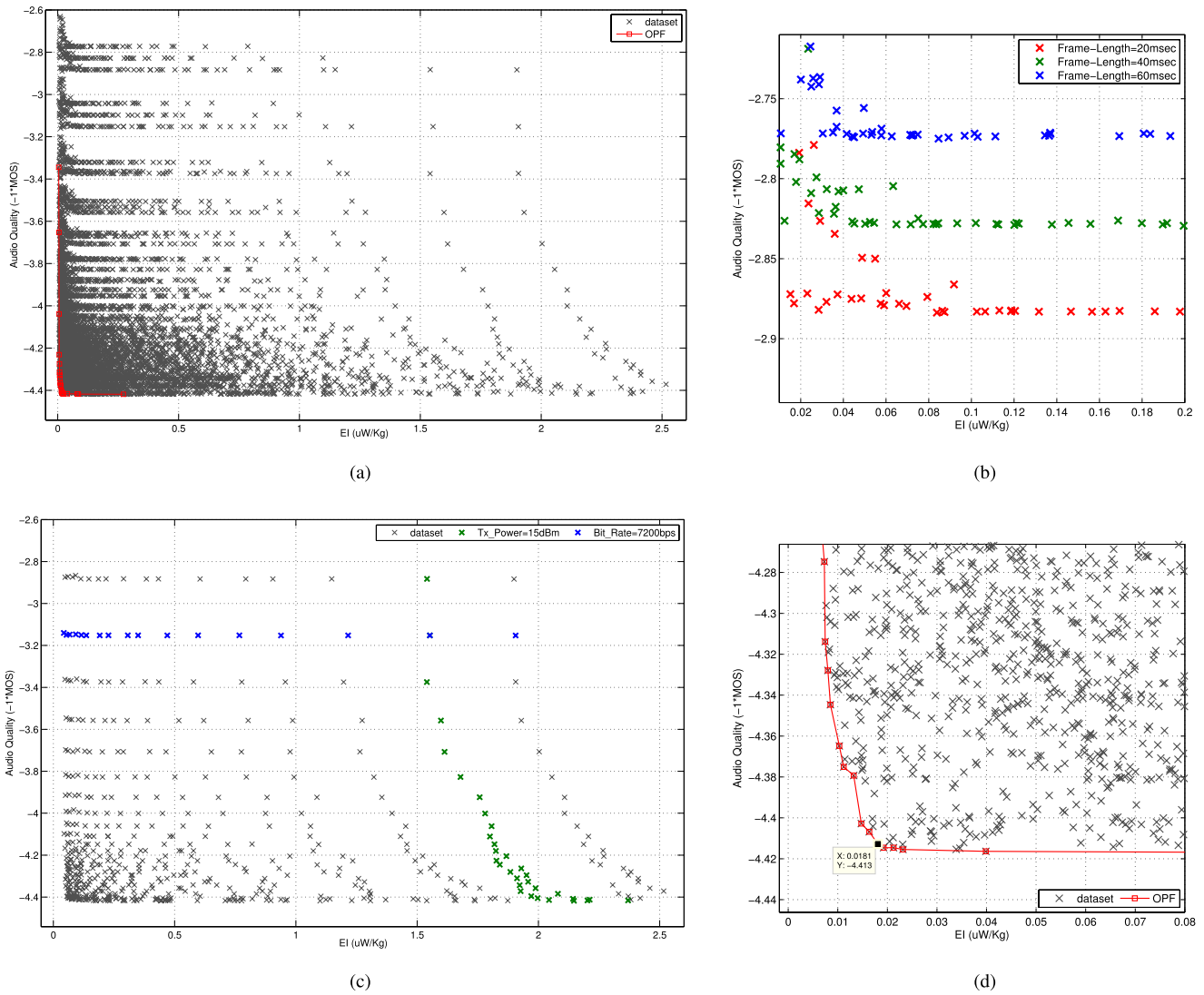
(a)

(b)

(c)

(d)

Fig. 8. Exhaustive search model and OPF plot of the Wi-Fi conferencing experiment. The Y-axis represents the audio quality objective using an inverted MOS score [−1 to −5] and the X-axis represents the transmission exposure of combined uplink and downlink EI values in uW/kg. (a) Visualizing an exhaustive search by using all parameter combinations. (b) Zooming into the top three rows. (c) Visualizing Tx-Power and Bit-Rate variation while keeping Tx-Rate and Frame-Length fixed. (d) Zooming into the lower left knee point region.

performance metrics can be discerned. Figure 8(c) shows the impact of varying Tx-Power and codec Bit-Rate parameters while keeping the Frame-Length and Tx-Rate fixed at their lowest values. By doing so, an exponentially rising and logarithmically spaced relation between the audio quality and exposure objectives can be observed. Varying the codec Bit-Rate parameter leads to the exponential rising relation (green points @ Tx-Power = 15dBm), whereas the logarithmic spacing is caused by varying the Tx-Power parameter (blue points @ Bit-Rate = 7200bps). Since exposure is proportional to the number of sent packets (which is directly related to the choice of codec Bit-Rate), also the audio-quality has an exponentially rising relation with the codec Bit-Rate [37].

To a lesser extent, the codec Frame-Length parameter also affects the audio quality objective. This is visualized in Figure 8(b) by zooming in the top three rows of the overall results. The codec Frame-Length parameter directly affects the latency of audio packets and thus the audio quality is reduced

when the Frame-Length is increased which results in a linear relationship between the Frame-length parameter and audio quality objective (blue points @ Frame-Length = 60msec, green points @ Frame-Length = 40msec and red points @ Frame-Length = 20msec). In contrast, the audio quality objective shows no impact when Tx-Power is increased because the experiment was performed in a shielded environment where no interference was present.

The relations between input parameters and objective as illustrated above are difficult to predict exactly, even for domain experts. Although it is interesting to plot these interactions, in many situations the network operator is only interested in identifying the Pareto front with optimal operation points (i.e. the red line in Figure 8(a)). A zoomed in version is shown in Figure 8(d). Based on the exhaustive search, the best attainable values of the respective objectives is an exposure index of 0.004125 uW/kg and a MOS score of 4.41835. These are the best values of each objectives and can not be improved further by using any parameter combination.

When optimizing a solution in which both metrics are considered, the knee point of the OPF is a usual point to consider. The knee point of the OPF is a Pareto point that is closest to a hypothetical intersection point formed by the asymptotic audio quality and exposure lines (i.e. MOS = 4.41835 and exposure = 0.004125 uW/kg). By doing so the knee point of the OPF, shown as a black dot in Figure 8(d), has design parameters [Bit-Rate = 28000bps, Frame-Length = 20msec, Tx-Rate = 24Mbps, Tx-Power = 0dBm] and performance objectives [MOS = 4.4128, exposure = 0.018097 uW/kg].

In summary, the exhaustive search model gives insights to relations existing between the objectives and the design parameters and resulted in an optimal Pareto front. The next sections will use the MOSBO optimizer introduced in Section III in order to approximiate the OPF using fewer experiments.

B. Stopping Criteria

Since the number of experiments should be reduced, it is crucial to identify stopping criteria that determine when the approximate solution (in this case the OPF) is accurate enough and the experimentation can be stopped, thereby avoiding unnecessary experiment iterations. Typically, stopping criteria are based on estimating the Approximate Pareto Front (APF) progress. There are a number of methods to estimate the APF progress, most of which have previously been applied to Multi-Objective Evolutionary Algorithms (MOEA), however for MOSBO these have not yet been evaluated. Stopping criteria are typically composed of (i) progress indicators, (ii) evidence gathering and (iii) a stopping decision. The next subsections will discuss each of them in more detail.

1) Progress Indicators (PI): The progress indicator calculates from the collected dataset how much the solution has improved from the previous iteration.

I *Mutual Domination Rate (MDR):* MDR is a progress indicator evaluated between consecutive Pareto sets [46]. The consecutive Pareto sets are compared and MDR calculates the domination rate of the recent set on the previous set. MDR values range from -1 to 1 where 1 indicates a highest domination, -1 indicates no domination at all and others represent scaled domination rates according to their amount. A high MDR value indicates that the newest Pareto set shows a significant improvement when compared to the Pareto set that was obtained during the previous experiment run. Mathematically the MDR is defined as,

$$I_{mdr}\left(P_t^*, P_{t-1}^*\right) = \frac{\left|\triangle\left(P_{t-1}^*, P_t^*\right)\right|}{\left|P_{t-1}^*\right|} - \frac{\left|\triangle\left(P_t^*, P_{t-1}^*\right)\right|}{\left|P_t^*\right|} \quad (13)$$

where $|A|$ is the number of elements in A and $\triangle(A, B)$ is the set of elements of A that are dominated by at least one element of B.

II *Epsilon Dominance (ED):* Similar to the MDR indicator, the ED indicator is also calculated between consecutive Pareto sets [35]. The ED calculates a minimum factor ϵ by which the current Pareto set is better than the former Pareto set with respect to all objectives.
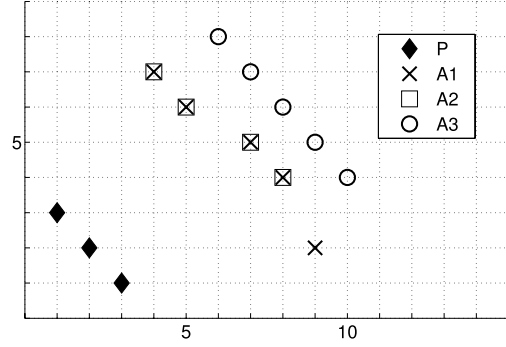

Fig. 9. A hypothetical dual objective problem having OPF=P and three consecutive Pareto sets (i.e. A_1, A_2 and A_3). $I_\varepsilon(A_1, A_3)$ calculates the minimum factor ε that needs to be added or multiplied on all A_3 objectives to compare with the Pareto set of A_1. This gives $I_{\varepsilon+}(A_1, A_3)$ = -1 and $I_{\varepsilon*}(A_1, A_3) = 0.9$. On the other hand, $I_{\varepsilon+}(A_1, A_2)$ and $I_{\varepsilon*}(A_1, A_2)$ evaluate to 0 and 1 respectively as all elements of A_2 exist in A_1.

Mathematically ED, for Pareto sets A and B each having n design objectives, is defined as,

$$I_{\epsilon+}(A, B) = \max_{Z^2 \varepsilon B} \min_{Z^1 \varepsilon A} \max_{1 \leq i \leq n} \left(Z_i^1 - Z_i^2\right) \quad (14)$$

$$I_{\epsilon*}(A, B) = \max_{Z^2 \varepsilon B} \min_{Z^1 \varepsilon A} \max_{1 \leq i \leq n} \left(Z_i^1 / Z_i^2\right) \quad (15)$$

for additive and multiplicative versions respectively. In the experiments performed in this paper, the additive ED indicator is only applied. Figure 9, provides a graphical explanation of the ED indicator.

III *Hyper Volume (HV):* The HV indicator [47], previously discussed in section III-B.1, calculates the volume of the dominated region of a given Pareto set bounded by a reference point r and dominated by all points in the Pareto set. As the optimization progresses, the Pareto set starts to converge to the OPF which also increases the HV indicator. Therefore, larger HV values indicate better Pareto sets.

2) Evidence Gathering Process (EGP): Next, the evidence gathering process performs a statistical analysis to calculate the changes in the progress indicators over time.

I *Moving Average (MA):* MA evidence gathering calculates the average of a given indicator values by moving along a fixed calculation window. In time, indicators are expected to reach a stable value, allowing it to compare with a threshold stopping decision at a later stage.

II *STandard DEViation (STDEV):* STDEV evidence gathering calculates the standard deviation on a collection of indicator values to estimate a possible experiment stagnation. STDEV was used in [39] in conjunction with a combined objective indicator.

III *Linear Regression (LR):* LR evidence gathering calculates the linear regression of a collection of indicator values and compares it against a stopping decision once the goodness of fit is satisfied. Types of Indicators often used with LR evidence gathering are MDR, HV and ED.

IV *Kalman Filtering (KF):* KF evidence gathering uses the Kalman Filter to estimate the state of a dynamic system from noisy measurements [46]. Kalman Filters assume a system to be linear and consecutive iterations to be only

TABLE IV
STOPPING CRITERIA COMBINATIONS USED IN MOSBO EXPERIMENT

PI	EGP	SD	Remark
ED	STDEV	THLD	WIDTH=10 and STDEV-THLD=ε
MDR	KF	ciNormal	R=0.1, Pr=95% and KF-THLD=-0.9
HV	LR	C-THLD	WIDTH=10, RMSE=20ε, LR-THLD=10ε and Reference Point: (EI=2.877 uW/kg, MOS=2.4478)

dependent on previous measurements. At each iteration, the state vector of the system and the covariance of the vector are updated based upon a new observation. The filtered estimate and its error value will then be used to make a stopping decision. MDR indicator is mostly applied with KF evidence gathering.

3) Stopping Decision (SD): Finally, the result from the evidence gathering is compared against a predefined stopping decision.

I *ThresHoLD (THLD):* Threshold stopping decision compares the evidence gathering estimate against a fixed value and stops iterating when the evidence gathering process falls below a predefined threshold. Threshold stopping decision is often used with STDEV and MA evidence gatherings.

II *ciNormal:* ciNormal is similar to the Threshold stopping decision but rather than making plain threshold comparison, it uses a normal-distribution to address the uncertainty of the EGP values. KF EGP is one typical example that uses the ciNormal stopping decision.

III *Conditional THreshoLD (C-THLD):* Sometimes the values from EGP are not acceptable until a certain condition is fulfilled. LR EGP, for example, needs to have a very low Root Mean Squared Error (RMSE) between the PI values and their linear approximation before using the slope parameter as a stopping decision.

C. Performance Evaluation

This section evaluates the performance of the MOSBO approach when optimizing the Wi-Fi conferencing scenario described in Section IV. To this end, the MOSBO experiment is conducted by selecting different stopping criteria (i.e. indicators, evidence gathering and stopping decision), and comparing the results with those obtained from the exhaustive search model (Section V-A). Two performance metrics are evaluated: (i) Speed up Factor and (ii) Population Domination Rate.

- The *Speed up Factor* (SuF) compares the number of experiment iterations that a MOSBO experiment requires compared to the number of iterations for an exhaustive search experiment. Higher values of the SuF correspond to faster optimizations.

- The *Population Domination Rate* (PDR) gives an indication of the closeness between the APF and the OPF. It expresses the percentage of elements dominated by the Pareto set from the MOSBO experiment when compared to the exhaustive search model. As such, higher PDR values indicate a better Pareto front estimation.

To ensure an adequate approximation of the OPF, multiple stopping criteria have been combined: the MOSBO optimizer
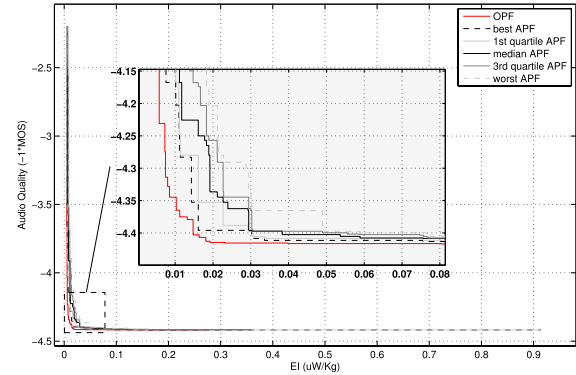
(a) Pareto Front plot
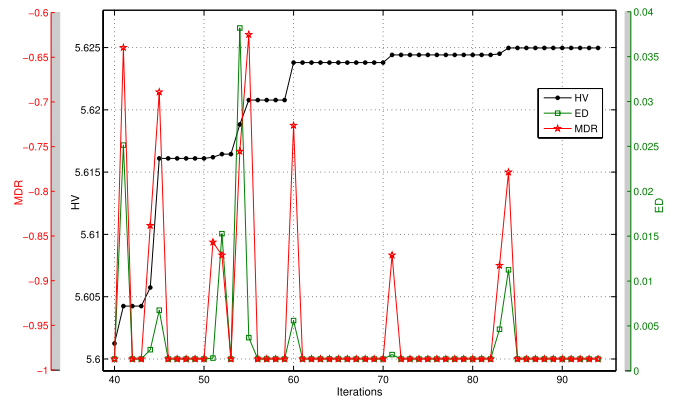
(b) Progress Indicators plot

Fig. 10. MOSBO Pareto optimization plot after the stopping criteria is met. (a) Pareto front overview, including the Pareto front of the exhaustive search model (red line), the Pareto front from the MOSBO experiment (blue line) and the intermediary experiments (dots). (b) Values of the progress indicators as calculated during the experiment iterations. (a) Pareto Front plot. (b) Progress Indicators plot.

stops only after satisfying all of them. Table IV gives an overview of the different stopping criteria combinations used. Although an infinite number of stopping criteria combinations are possible, the selected values have shown to be useful in previous multi-objective optimization problems. In the EGP calculation, the WIDTH parameter is set to 10 for ED and HV progress indicators. The R and Pr parameters (values taken from [46]) are used as a noise value and as an uncertainty estimator of the KF respectively. Regarding the HV indicator, linear regression is applied once the RMSE is below a minimum value (20 times the value of epsilon). Finally SD thresholds of STDEV and LR take epsilon as the lower bound and KF take -0.9 (90% of minimum MDR) as a lower bound.

Figure 10(a) shows the snapshot after the end of the MOSBO optimization by comparing the OPF (red curve) with the summary attainment surface plots of 8 different experiments. Summary attainment surface plots [48] are means to visualize the distribution of different APFs according to statistical models. In Figure 10(a), 5 different estimators (best, 1st quartile, median, 3rd quartile and worst) are visualized. The best and worst plots represent the boundaries of all APFs but they are biased estimators of the population because they show wide variations for different experiments. On the other hand, the 1st quartile, the median and the 3rd quartile plots

are stable estimators, median being the best of all showing the least variation across different experiments.

The average number of experiment iterations to obtain the APF is 94 and due to its limited count, the APF partially overlaps the OPF. Compared to the exhaustive search experiment, the PDR of the MOSBO solution corresponds to 96.58%, meaning that the APF dominates 96.58% of the complete design space. Although the MOSBO Pareto front locates most of the optimal solutions, some Pareto optimal solutions were not identified especially on the knee point region of the OPF.

To analyze the behavior of the MOSBO APF over time, a progress Indicator plot is shown in Figure 10(b). In the figure, three PIs (MDR, ED, HV) are visually presented, which together form the MOSBO stopping criteria. The MDR indicator is highly fluctuating while the others are relatively stable. As MDR indicator only accounts for the number of updated Pareto elements but not their magnitudes, this tells that the Pareto elements were showing slight variations which couldn't have noticeable effect on the ED and HV indicators. By applying a combined stopping criteria (Table IV) the MOSBO solution converged after 94 iterations, corresponding to a speed up factor of $6528/94 = 69.45$ and Population Domination Rate of 96.58%.

D. Initial Sample Size Sensitivity

In the previous section, it was mentioned that the MOSBO experiment converged after 94 experiments. Although the MOSBO optimizer includes selection criteria for identifying the most promising candidate settings (i.e. the expected improvement criteria, see Section III-A), it relies on the size of the initial samples and the sampling method before generating the initial model. The choice of the sampling method was briefly discussed in the previous paper [39], where Latin Hypercube Sampling (LHS) was shown to have best results. This section expands on this discussion by making a sensitivity analysis over the iteration count.

Providing additional samples typically result in a better initial model, at the cost of additional experiments. As indicated in [39], the problem of exploration vs. exploitation trade-off can be addressed by a good selection of the initial sample size. The initial sample size should explore a problem adequately such that the APF can be retrieved in the shortest time possible. Although previous works in other applications analysed the optimal number of sample points (for example, [49] advises to use 10 times the number of design parameters), the optimal number depends on the smoothness (i.e. the predictability) of the objective functions over the design space.

To evaluate the sensitivity of different initial sample sizes, Figure 11 shows the iteration count of MOSBO experiments as a function of the initial sample size. The iteration count indicates the total number of experiments that are performed until the stopping criteria are satisfied. To indicate the efficiency of the MOSBO optimizer, the PDR metric is also plotted in Figure 11.

As indicated by Figure 11, the iteration count oscillates along a linear regression line (broken purple color on
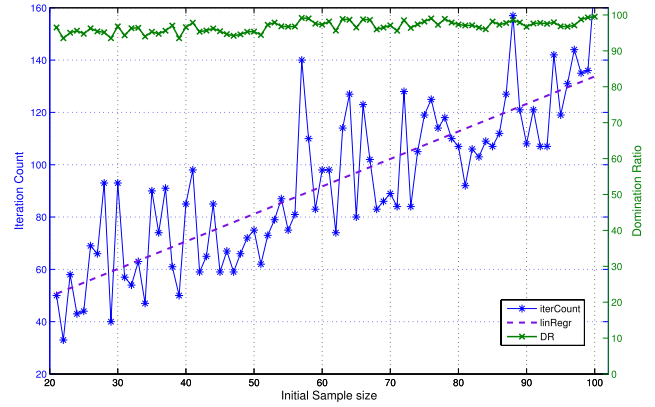

Fig. 11. Sensitivity analysis of different initial sample sizes as a function of Iteration count.

Figure 11) with a slope of $1.05 \approx 1.0$. On average, the iteration count increases with the same proportion as with the increase in the initial sample size. This indicates that on average, the MOSBO optimizer spends an equal number of iterations for any initial sample size selected. This is also related to the smoothness nature of the objective functions over the design space. Hence, the Wi-Fi conferencing scenario presented in this paper is not sensitive to different initial sample sizes and 22 initial samples are sufficient to get the APF in the shortest time possible (33 iterations). However, as in many cases the smoothness of the evaluated solution is not known a priori, a more conservative amount of samples (i.e. 10*design parameters or 40 initial samples) have been used following the guidelines from [49].

E. MOSBO Computational Complexity

Besides the time required for experimentation, also the optimizer spends time calculating the next set of parameter settings that should be investigated. Constructing the kriging models involves a Cholesky decomposition which has an order of $O(n^3)$ amongst other matrix operations [50], [31]. Since the kriging models are constructed for each design objective, the overall computation is multiplied by the number of design objectives involved. On the second step, the constructed kriging models are processed to generate the non-dominated region and the computational complexity is a function of the Pareto points ([7, Fig. 5(a)]). At the last section of the actual optimization, cost prediction and variance prediction processes are involved which require $O(n)$ and $O(n^2)$ computational complexities respectively. The actual execution time of the MOSBO optimizer is dependent on the type of hardware used. Specific to our experimentation set-up (Table I), a single MOSBO iteration needs an average of 1 second, which is significantly less than the time required for most real-life experiment iterations.

VI. CONCLUSION

This paper provides a method to identify performance trade-offs in wireless systems through efficient Pareto optimizations. Wireless systems typically have a large design space with several interacting parameters. Analyzing these parameters using

a wireless experiment typically consumes a significant amount of time due to the need to deploy, configure and monitor devices. To reduce the time complexity, the paper applied a Multi Objective Surrogate-Based Optimizer (MOSBO) which works by creating a surrogate (Kriging) model out of selected design points and their objective performances.

Afterwards, the advantages of the MOSBO optimizer are demonstrated using a Wi-Fi conferencing scenario where a speaker node adjusts four configurable parameters (i.e Tx-Power, Tx-Rate, Codec Bit-Rate, Codec Frame-Length) and measures two performance objectives (i.e. audio quality and transmission exposure). Both objectives are influenced by all selected parameters and can not be optimized individually since they negatively influence each other. An automated experimentation system was set up in which the Wi-Fi network was integrated with a MOSBO optimizer with the aim of identifying the Approximate Pareto Front (APF) using as few experiments as possible.

Based on an exhaustive search model, the OPF was determined and the interactions between input parameters and performance metrics were described. For the MOSBO approach, the influence of selecting different initial sample points was analyzed and multiple stopping criteria were discussed. A combination of three stopping criteria was proposed to ensure adequate covering of the APF. The experimental validation of the solution compares the performance of the MOSBO approach against an exhaustive search experiment. The benefit of the MOSBO optimizer is demonstrated by finishing the experiment using 94 iterations out of the complete design space (6528 elements) and speeding up the experiment 6528/94 = 69.45 times while the APF dominates 96.58% of the complete design space. Moreover, the sensitivity analysis of different initial sample sizes on the performance of MOSBO is investigated and it is found that the Wi-Fi conferencing scenario is not sensitive to different initial sample sizes.

REFERENCES

[1] K. Deb, "Multi-objective optimisation using evolutionary algorithms: An introduction," in *Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing*. London, U.K.: Springer, 2011, pp. 3–34. [Online]. Available: http://dx.doi.org/10.1007/978-0-85729-652-8_1

[2] "Basic concepts," in *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed. Boston, MA, USA: Springer, 2007, pp. 1–60. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-36797-2_1

[3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[4] N. Beume, B. Naujoks, and M. Emmerich, "SMS-EMOA: Multiobjective selection based on dominated hypervolume," *Eur. J. Oper. Res.*, vol. 181, no. 3, pp. 1653–1669, Sep. 2007.

[5] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," Swiss Federal Inst. Technol., Zürich, Switzerland, Tech. Rep., 2001.

[6] T. Rakotoarivelo, G. Jourjon, and M. Ott, "Designing and orchestrating reproducible experiments on federated networking testbeds," *Comput. Netw.*, vol. 63, pp. 173–187, 2014.

[7] I. Couckuyt, D. Deschrijver, and T. Dhaene, "Fast calculation of multi-objective probability of improvement and expected improvement criteria for Pareto optimization," *J. Global Optim.*, vol. 60, no. 3, pp. 575–594, Nov. 2014.

[8] N. Varsier *et al.*, "A novel method to assess human population exposure induced by a wireless cellular network," *Bioelectromagnetics*, vol. 36, no. 6, pp. 451–463, Sep. 2015. [Online]. Available: http://dx.doi.org/10.1002/bem.21928

[9] D. Plets, W. Joseph, K. Vanhecke, and L. Martens, "Exposure optimization in indoor wireless networks by heuristic network planning," *Prog. Electromagn. Res.*, vol. 139, pp. 445–478, May 2013.

[10] D. Plets *et al.*, "Joint minimization of uplink and downlink whole-body exposure dose in indoor wireless networks," *BioMed Res. Int.*, vol. 2015, Feb. 2015, Art. no. 943415. [Online]. Available: http://dx.doi.org/10.1155/2015/943415

[11] L. Christianson and K. Brown, "Rate adaptation for improved audio quality in wireless networks," in *Proc. IEEE Int. Workshop Mobile Multimedia Commun. (MoMuC)*, Nov. 1999, pp. 363–367.

[12] M. I. Tariq, M. A. Azad, R. Beuran, and Y. Shinoda, "Performance analysis of VoIP codecs over BE WiMAX network," *Int. J. Comput. Elect. Eng.*, vol. 5, no. 3, pp. 345–349, Jun. 2013.

[13] I. S. H. C. Ilias and M. S. Ibrahim, "Performance analysis of audio video codecs over Wi-Fi/WiMAX network," in *Proc. 8th Int. Conf. Ubiquitous Inf. Manage. Commun. (ICUIMC)*, New York, NY, USA, 2014, pp. 60:1–60:5. [Online]. Available: http://doi.acm.org/10.1145/2557977.2558009

[14] S. Thakolsri, W. Kellerer, and E. Steinbach, "Application-driven cross layer optimization for wireless networks using MOS-based utility functions," in *Proc. 4th Int. Conf. Commun. Netw. China (ChinaCOM)*, Aug. 2009, pp. 1–5.

[15] S. Jadhav, H. Zhang, and Z. Huang, "MOS-based handover protocol for next generation wireless networks," in *Proc. IEEE 26th Int. Conf. Adv. Inf. Netw. Appl.*, Mar. 2012, pp. 479–486.

[16] S. Koziel and A. Bekasiewicz, "Fast multiobjective optimization of narrowband antennas using RSA models and design space reduction," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 450–453, 2015.

[17] S. Koziel and A. Bekasiewicz, "Multiobjective antenna design by means of sequential domain patching," *IEEE Antennas Wireless Propag. Lett.*, vol. 15, pp. 1089–1092, 2016.

[18] S. Koziel and S. Ogurtsov, "Numerically efficient approach to simulation-driven design of planar microstrip antenna arrays by means of surrogate-based optimization," in *Solving Computationally Expensive Engineering Problems*. Springer, 2014, pp. 149–170.

[19] I. Couckuyt, F. Declercq, T. Dhaene, H. Rogier, and L. Knockaert, "Surrogate-based infill optimization applied to electromagnetic problems," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 20, no. 5, pp. 492–501, Sep. 2010.

[20] K. Navaie and T. A. Le, "Fundamental performance trade-offs in coexisting wireless networks," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity (5GU)*, Nov. 2014, pp. 246–251.

[21] M. Stehlík, A. Saleh, A. Stetsko, and V. Matyáš, "Multi-objective optimization of intrusion detection systems for wireless sensor networks," in *Proc. 12th Eur. Conf. Artif. Life*, pp. 569–576, Sep. 2013. [Online]. Available: http://mitpress.mit.edu/sites/default/files/titles/content/ecal13/978-0-262-31709-2-ch082.pdf

[22] R. Annauth and H. C. S. Rughooputh, "OFDM systems resource allocation using multi-objective particle swarm optimization," *Int. J. Comput. Netw. Commun.*, vol. 4, no. 4, pp. 291–306, Jul. 2012. [Online]. Available: http://airccse.org/journal/cnc/0712cnc19.pdf

[23] R. Devarajan, S. C. Jha, U. Phuyal, and V. K. Bhargava, "Energy-aware resource allocation for cooperative cellular network using multi-objective optimization approach," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1797–1807, May 2012.

[24] S. Fu, Y. Zhang, Y. Jiang, C. Hu, C.-Y. Shih, and P. J. Marrón, "Experimental study for multi-layer parameter configuration of WSN links," in *Proc. IEEE 35th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun./Jul. 2015, pp. 369–378.

[25] R. Badhurshah and A. Samad, "Multiple surrogate based optimization of a bidirectional impulse turbine for wave energy conversion," *Renew. Energy*, vol. 74, pp. 749–760, Feb. 2015.

[26] X. Chen, L. Zhang, X. He, C. Xiong, and Z. Li, "Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 29, no. 5, pp. 359–381, May 2014.

[27] M. Costas, J. Díaz, L. Romera, and S. Hernández, "A multi-objective surrogate-based optimization of the crashworthiness of a hybrid impact absorber," *Int. J. Mech. Sci.*, vol. 88, pp. 46–54, Nov. 2014.

[28] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *Global Optim.*, vol. 21, no. 4, pp. 345–383, Dec. 2001.

[29] M. T. M. Emmerich, A. H. Deutz, and J. W. Klinkenberg, "Hypervolume-based expected improvement: Monotonicity properties and exact computation," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2011, pp. 2147–2154.

[30] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Prog. Aerosp. Sci.*, vol. 45, nos. 1–3, pp. 50–79, Jan./Apr. 2009.

[31] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq, "A surrogate modeling and adaptive sampling toolbox for computer based design," *J. Mach. Learn. Res.*, vol. 11, pp. 2051–2055, Jan. 2010.

[32] I. Couckuyt, T. Dhaene, and P. Demeester, "ooDACE toolbox: A flexible object-oriented Kriging implementation," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3183–3186, Jan. 2014.

[33] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard, "Aspects of the MATLAB toolbox DACE," Dept. Informat. Math. Model., Tech. Univ. Denmark, Lyngby, Denmark, Tech. Rep. IMM-REP-2002-13, 2002.

[34] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, 1999.

[35] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance assessment of multiobjective optimizers: An analysis and review," *IEEE Trans. Evol. Comput.*, vol. 7, no. 2, pp. 117–132, Apr. 2003.

[36] S. Bouckaert, P. Becue, B. Vermeulen, B. Jooris, I. Moerman, and P. Demeester, "Federating wired and wireless test facilities through Emulab and OMF: The iLab.t use case," in *Testbeds and Research Infrastructure. Development of Networks and Communities* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 44, T. Korakis, M. Zink, and M. Ott, Eds. Berlin, Germany: Springer, 2012, pp. 305–320. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-35576-9_25

[37] A. Rämö and H. Toukomaa, "Voice quality characterization of IETF opus codec," in *Proc. 12th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Florence, Italy, Aug. 2011

[38] ITU-T Study Group 12 (1993-1996), "Methods for subjective determination of transmission quality," Int. Telecommun. Union, Geneva, Switzerland, Tech. Rep. I.-T.R.P.800, Aug. 1996.

[39] M. T. Mehari *et al.*, "Efficient global optimization of multi-parameter network problems on wireless testbeds," *Ad Hoc Netw.*, vol. 29, pp. 15–31, Jun. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870515000244

[40] W. Jiang and H. Schulzrinne, *Comparisons of FEC and Codec Robustness on VoIP Quality and Bandwidth Efficiency*. Singapore: World Scientific, Jun. 2002.

[41] A. Rämö and H. Toukomaa, "Subjective quality evaluation of the 3GPP EVS codec," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 5157–5161.

[42] A. Heuberger, "The future of communication: Full-HD voice powered by EVS and the AAC-ELD family," Fraunhofer IIS, Munich, Germany, Tech. Rep., 2015, accessed on Apr. 19, 2016. [Online]. Available: http://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/wp/FraunhoferIIS_Technical-Paper_Full-HDVoice_AAC-ELD_EVS.pdf

[43] (2014). *P.863: Perceptual Objective Listening Quality Assessment*, accessed on Apr. 19, 2016. [Online]. Available: http://www.itu.int/rec/T-REC-P.863-201409-I/en

[44] R. Chen, T. Terriberry, J. Skoglund, G. Maxweel, and H. T. M. Nguyet, "Opus testing," IETF, Prague, Czech Republic, Tech. Rep., Mar./Apr. 2011. [Online]. Available: https://www.ietf.org/proceedings/80/slides/codec-4.pdf

[45] I. Dyakonov and G.-C. Pascutto. (2011). *64kbit/sec Stereo Multiformat Listening Test*, accessed on Apr. 19, 2016. https://people.xiph.org/~greg/opus/ha2011/

[46] L. Marti, J. Garcia, A. Berlanga, and J. Molina, "An approach to stopping criteria for multi-objective optimization evolutionary algorithms: The MGBM criterion," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, May 2009, pp. 1263–1270.

[47] J. L. Guerrero, J. Garcia, L. Marti, J. M. Molina, and A. Berlanga, "A stopping criterion based on Kalman estimation techniques with several progress indicators," in *Proc. 11th Annu. Conf. Genet. Evol. Comput. (GECCO)*, New York, NY, USA, 2009, pp. 587–594. [Online]. Available: http://doi.acm.org/10.1145/1569901.1569983

[48] M. López-Ibáñez, L. Paquete, and T. Stützle, "Exploratory analysis of stochastic local search algorithms in biobjective optimization," in *Experimental Methods for the Analysis of Optimization Algorithms*. Berlin, Germany: Springer, 2010, pp. 209–222. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02538-9_9

[49] J. L. Loeppky, J. Sacks, and W. J. Welch, "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, vol. 51, no. 4, pp. 366–376, 2009.

[50] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning). Cambridge, MA, USA: MIT Press, 2005.

Michael Tetemke Mehari received the B.Sc. degree in electrical engineering from Addis Ababa University, Ethiopia, in 2006, and the master's degree in electronics engineering from Groept University/College, Belgium, in 2010. He is currently pursuing the Ph.D. degree with the Department of Information Technology (INTEC), Internet Based Communication Networks and Services (IBCN), Ghent University. He is a Senior Researcher at the Department of INTEC, IBCN, Ghent University. His research interests include cognitive networks, wireless experiment optimization, benchmarking of wireless networks and methodologies. He received the Grant Scholarship from Groept University/College in 2009.

Eli De Poorter received the master's degree in computer science engineering from Ghent University, Belgium, in 2006, and the Ph.D. degree from the Department of Information Technology, Ghent University in 2011. He is currently a Professor with Ghent University. He is currently involved in and/or a Research Co-Ordinator of several national and international projects. His main research interests include wireless network protocols, network architectures, wireless sensor and ad hoc networks, future internet, self-learning networks, and next-generation network architectures. He has authored or co-authored over 70 papers published in international journals or in the proceedings of international conferences. He is also the creator of the patented IDRA architecture, a flexible communication framework for heterogeneous networked devices. He served on the program committee of several conferences. He received the Ph.D. scholarship from the Institute for Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) and FWO post-doctoral research grant.

Ivo Couckuyt received the master's degree in computer science from the University of Antwerp in 2007, and the Ph.D. degree in science from the IBCN Research Group, Department of Information Technology, Faculty of Engineering, Ghent University in 2013. He is currently a Post-Doctoral Research Fellow with iMinds, Ghent University. His research is mainly focused on global and local surrogate modeling (metamodeling) and its application to solve real world problems, optimization of expensive functions, evolutionary computing and machine learning methods. He is also the Lead Developer and the Project Manager of the Surrogate Modeling Toolbox and the ooDACE Toolbox, used by many companies and institutions worldwide for design space exploration and optimization. He is a Post-Doctoral Research Fellow of the Research Foundation Flanders (FWO-Vlaanderen).

Dirk Deschrijver (M'09–SM'13) received the master's (Licentiaat) and Ph.D. degrees in computer science from the University of Antwerp, Antwerp, Belgium, in 2003 and 2007, respectively, and the Ph.D. degree in electrical engineering from Ghent University, Belgium, in 2012. He was with the Computer Modeling and Simulation Group, University of Antwerp. In 2005, he was a Marie Curie Fellow with the Scientific Computing Group, Eindhoven University of Technology, Eindhoven, The Netherlands. He is currently an FWO Post-Doctoral Research Fellow with the Department of Information Technology (INTEC), Ghent University, Ghent, Belgium. His research interests include the macromodeling of high-speed and high-frequency interconnects, orthonormal rational functions, parameterized modeling, and machine learning.

Günter Vermeeren received the M.Sc. degree in industrial engineering from the KAHO Sint-Lieven, Ghent, Belgium, in 1998, the M.Sc. degree in electrical engineering from Ghent University Belgium in 2001, and the Ph.D. degree in electro-technical engineering from the Ghent University, Belgium, in 2013. From 2001 to 2002, he was with the Research and Development Department, Network Integrator Telindus, Leuven, Belgium. Since 2002, he is a Research Engineer with the WiCa Group of Prof. L. Martens. His research focuses on the numerical modeling and measurements of electromagnetic radiation in the domain of radio-frequency dosimetry, electromagnetic exposure, on-body propagation, and medical imaging systems, such as hybrid MRI systems.

David Plets received the master's degree in electro-technical engineering from Ghent University, with ICT as main subject, in 2006, and the Ph.D. degree (Hons.) in the characterization and optimization of the coverage of wireless broadcast and WLAN networks in 2011. He has been a member of the Wireless, Acoustics, Environment& Expert Systems (WAVES) Group, Department of Information Technology - INTEC, Ghent University. His current research interests include optimization of (cognitive) wireless communication networks, with a focus on coverage, interference, and exposure. He is also involved in research on localization techniques and IoT, for both industry- and health-related applications.

Wout Joseph (M'05–SM'12) was born in Ostend, Belgium, in 1977. He received the M.Sc. degree in electrical engineering and the Ph.D. degree from Ghent University (UGent), Ghent, Belgium, in 2000 and 2005, respectively. From 2000 to 2005, he was a Research Assistant with the Department of Information Technology, iMinds/UGent-INTEC. His scientific work was focused on electromagnetic exposure assessment. His research work dealt with measuring and modeling of electromagnetic fields around base stations for mobile communications related to the health effects of the exposure to electromagnetic radiation. Since 2005, he has been a Post-Doctoral Researcher with iMinds/UGent–INTEC. From 2007 to 2013, he was a Post-Doctoral Fellow with the Research Foundation- Flanders (FWO-V). Since 2009, he has been a Professor in the domain of experimental characterization of wireless communication systems. His professional interests are electromagnetic field exposure assessment, propagation for wireless communication systems, antennas, and calibration. Furthermore, he specializes in wireless performance analysis and quality of experience.

Luc Martens (M'92) received the M.Sc. degree in electrical engineering and the Ph.D. degree from Ghent University, Ghent, Belgium, in 1986 and 1990, respectively. From 1986 to 1990, he was a Research Assistant with the Department of Information Technology (INTEC), Ghent University. His scientific work was focused on the physical aspects of hyperthermic cancer therapy. His research work dealt with electromagnetic and thermal modeling and with the development of measurement systems for that application. Since 1991, he has been managing the Wireless & Cable Research Group, INTEC, which has been part of the iMinds Institute since 2004. Since 1993, he has been also a Professor with Ghent University. He has authored or co-authored over 300 publications in the domain of electromagnetic channel predictions, dosimetry, exposure systems, and health and wireless communications. His experience and current interests are in modeling and measurement of electromagnetic channels, of electromagnetic exposure, around telecommunication networks and systems, such as cellular base station antennas, and of energy consumption in wireless networks.

Tom Dhaene received the Ph.D. degree in electrical engineering from Ghent University, Belgium, in 1993. He is currently a Full Professor with the Department of Information Technology, Ghent University. In 2000, he joined the Department of Mathematics and Computer Science, University of Antwerp as a Professor. Since 2007, he has been a Full Professor with the Department of Information Technology, Ghent University. He is also with iMinds. He has authored or co-authored over 350 peer-reviewed papers and abstracts in international conference proceedings, journals and books about computational science and numerical analysis and engineering. He holds five U.S. patents.

Ingrid Moerman received the degree in electrical engineering and the Ph.D. degree from Ghent University, in 1987 and 1992, respectively. She became a part-time professor with Ghent University in 2000. She is a Staff Member with the Research Group on Internet-Based Communication Networks and Services, IBCN, where she is leading the research on mobile and wireless communication networks. Since 2006, she has been with iMinds, where she is coordinating several interdisciplinary research projects. She is a Principal Investigator of the High Impact Initiative Everything Connected with iMinds. She has authored or co-authored over 700 publications in international journals or conference proceedings. Her main research interests include Internet of Things, cooperative and cognitive radio networks, wireless access, flexible HW/SW architectures, and experimentally-supported research. She has longstanding experience in running national and EU research funded projects. At the European level, she is in particular very active in the European Future Internet Research and Experimentation research area, where she has coordinated and is coordinating several projects (CREW, WiSHFUL, eWINE) and participating in other projects (Fed4FIRE, FORGE, FLEX). In the research area on Future Networks, she is involved in the 5G-PPP project Flex5GWare.