Indoor Localization, Activity Recognition and their Applications - Research Article

International Journal of Distributed Sensor Networks

International Journal of Distributed Sensor Networks
2019, Vol. 15(8)

DOI: 10.1177/1550147719862215
journals.sagepub.com/home/dsn
SAGE

Analysis of space labeling through binary fingerprinting

Marouan Mizmizi and Luca Reggiani

Abstract

In the context of fingerprinting applications, this article presents the performance analysis of a type of space labeling based on the binary quantization of the received signal strength indicator. One of the common drawbacks of fingerprinting is the large data size and consequently the large search space and computational load as a result of either vastness of the positioning area or the finer resolution in the fingerprinting grid map. Our approach can be considered, for example, when we use very small, inexpensive beacons, like those based on bluetooth low energy technology, radio frequency identification, or in the future context of the Internet of Things. One of the interesting properties of this deployment is that it can be interpreted as a form of space labeling or encoding since space is divided into cells, and each cell is associated to a binary codeword with the corresponding scalability of the spatial resolution. Here, it developed the performance estimation, exploiting the association of this deployment to an error correcting code. The analysis and numerical and experimental results allow a deeper understanding of the impact of the proposed solution and show that it is robust and computationally efficient with respect to the traditional fingerprinting technique.

Keywords

Wireless sensor networks, positioning, fingerprinting

Date received: 7 August 2018; accepted: 1 June 2019

Handling Editor: Antonio Lazaro

Introduction

Due to recent developments in hardware electronics and communications, wireless personal area networks (WPANs) and wireless sensor networks (WSNs) have found outstanding importance in diverse applications such as industrial, medical, public services, and many other fields. At the same time, Internet of things (IoT) is expected to become the pervasive infrastructure of the information society. In this context, position information is an important enabler of new and future value-added services.

In the last decade, we observed a constant growing effort, both from scientific communities and from the industry in order to design a global solution for the indoor positioning and navigation that achieves high precision, reliability, and cost effectiveness. The availability of such IPS (Indoor Positioning Systems) will enable advances in location-aware applications, pervasive computing, and ambient intelligence. According to Wirola et al.,[1] the high-accuracy indoor positioning will be based on dedicated positioning-specific tags, and this is because of the trade-off between costs and performance. In fact, the wireless local area network (WLAN) technology offers the best solutions in terms of costs and this thanks to the reuse of available infrastructures although it has many limitations in terms of accuracy and reliability. A solution based on bluetooth low energy (BLE) technology offers several advantages. Indeed, it ensures that the mobile terminals require no

Politecnico di Milano, Milano, Italy

Corresponding author:
Marouan Mizmizi, Politecnico di Milano, 20133 Milano, Italy.
Email: marouan.mizmizi@polimi.it

new hardware components and that the radio components are already in mass-production keeping the costs of deployment and new applications low. As a result, it represents a valid candidate for a large-scale solution.

The targets' position can be estimated via geometrical, Bayesian, or pattern-matching approaches.[2–4] The geometrical approaches use range-based techniques such as triangulation and multilateration. The range is estimated from the received signal exploiting different signal metrics, for example, time of arrival (ToA), angle of arrival (AoA), and received signal strength indicator (RSSI). The Bayesian approaches, used mainly for tracking application, compute the probability density function of the target position, providing its mean value as the estimated position and the variance as the uncertainty of the estimate. The pattern-matching approaches or fingerprinting (FP) take advantage of location dependent (LD) features of the signals received by static reference stations, or beacons: these signals, typically RSSI measures, are exploited as unique signatures associated with the target locations. First, a radio map containing stored LD parameters measured over predetermined points (grid points) is built during an off-line or training phase and then the target position is estimated via pattern matching between measured LD parameters and those previously recorded. In He and Chan,[5] it is possible to find some recent trends in two of the major research areas for FP localization: advanced localization techniques and efficient system deployment. Among the advanced approaches, some techniques are based on location-specific characteristics of multipath instead of RSSI as in Wu et al.[6] and Jin et al.,[7] while, in the specific area of IoT applications, the RSSI database is assisted in the online phase by other methods for increasing positioning accuracy, as magnetic or dynamic measures from other sensors or prediction models, as in Lin et al.[8] On the contrary, where complexity and costs are a hard constraint, the FP techniques are used since they can be really easy to implement and are cost-effective, while maintaining a reasonable degree of accuracy. Among the main complexity issues, we can observe that the search space during the online step can be computationally intense, either because the deployment area is wide (e.g. smart city, hospitals, or large factories) or because it is based on devices, such as BLE tags, with strong limitations on power consumption and limited hardware capabilities. In order to overcome this restraint, the works in Arya et al.[9] and Saha and Sadhukhan[10] have proposed clustering and spatial filtering techniques that limit the positioning algorithm to a subset of reference points (RPs) in order to narrow down the search space and focus on the relevant subset of RPs.

In this context, in the past, we have focused on the possibility of reducing further complexity and requirements of FP systems based on simple beacons in limited indoor environments. In Mizmizi and Reggiani,[11] we have investigated the role of quantization in the RSSI information. The obtained results showed that the computational complexity can be reduced, without loss of precision, by adapting the RSSI quantization with respect to the variance of the measurement noise. In Mizmizi and Reggiani,[12] we have focused on the simplest quantization, with two levels, exploring its relations with binary encoding. We have introduced a new scheme of a specific binary representation of the RSSI signatures and the measures (binary fingerprinting or BFP). This novel design is appropriate when the beacons are characterized by very limited size, cost, and computational capability, like in the BLE or in the future technologies for the pervasive IoT. In addition to the pure localization, BFP can be used for creating an ordered partitioning and labeling of the space, available in a simple and inexpensive way to each device in the covered area, with numerous potential applications in the area of context awareness.

In this article, we consider the design presented in Mizmizi and Reggiani,[12] developing the analytical estimation of the space discrimination provided by this novel spatial labeling. Therefore, the main contributions of this article are

- Analytical analysis of the FP technique in two cases: in the former, the number of RSSI quantization levels is assumed infinite, while, in the latter, the quantization levels are reduced to 2 obtaining the BFP. Moreover, the former is used always as a benchmark for the system;
- Interpretation of BFP as a spatial encoding with scalable resolution and logarithmic search of the solution;
- Validation of the numerical results, through experimental tests and channel models derived by experimental measures. The numerical findings, either by means of appropriate analysis, simulations, and experimental measures, show the impact of different parameters and channel conditions on the system performance.

The remainder of this work is organized as follows: the "System model" section describes the network scenario and the "Review of FP techniques" section reviews the FP techniques. In the "BFP design" section, we introduce the BFP approach and, in the "Analysis of FP performance" section, the analysis of both BFP and FP techniques. Finally, in the "BFP performance"

section, the numerical results are reported and discussed.

System model

The considered network model is the same as used in Mizmizi and Reggiani,[11,12] that is, an asynchronous sensor network containing a number of target devices over a limited squared area on a single floor, with two-dimensional coordinates (x, y).

In the area, there is a set of N_B fixed nodes called beacons (BS) with known positions $\mathbf{P}_i^{BS} = \{x_i, y_i\}, \{i = 1, 2, \ldots, N_B\}$. A rectangular grid is defined over the two-dimensional area, and any estimate of a target location is limited to the points on this grid. The grid of points has a resolution of Δ meters. Assuming that the grid spacing results in K_x points along the x coordinate and K_y along the y coordinate, we have a total number of positions in the area

$$K_T = K_x \times K_y \tag{1}$$

Any position can be represented by a triplet with label (x, y, z), where x and y represents the $2D$ coordinates on the floor plane while z represents the height of the antenna at that particular grid position. The coordinate $z = 0$ is assumed for all the points unless otherwise mentioned, and hence, the coordinates $(x, y)_i$ denote the location of the ith FP signature.

The experimental measures are taken in a room at Dipartimento di Elettronica, Informazione and Bioingegneria (DEIB) of Politecnico di Milano, whose map is sketched in Figure 1.

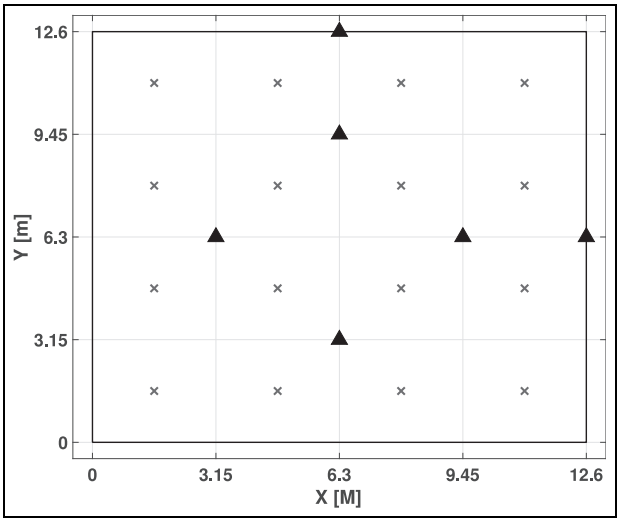

Figure 1. Map of the environment used for the experimental measurements. The triangles represents the BSs and the crosses are the FP signature locations.

Signal model and measures

The most common model adopted for the real RSSI, recorded and stored during the off-line phase, responds simply to the received power under log-normal shadowing and receiver isotropic antenna gain 0 dB, that is

$$RSSI(d) = EIRP - L_0 - 10 \cdot \alpha \cdot log_{10}\left(\frac{d}{d_0}\right) - L_{SH} \tag{2}$$

where the Effective Isotropic Radiated Power $EIRP$ incorporates the transmitter power and the antenna gain, L_0 is the average propagation loss at the reference distance d_0 (usually 1 m), α is the path loss exponent (PLE), d is the distance between transmitter and receiver, and $L_{SH} \sim \mathcal{N}(0, \sigma_{SH}^2)$ is the log-normal fluctuation due mainly to obstacles in the environment.[13]

According to the relative position between the point on the grid and the beacon, the values of L_0 and α depend also on the line of sight (LoS) or non-LoS (NLoS) conditions. In fact, each RSSI measured by the target during the online phase is affected by a measure error W with variance σ_W^2. This additional error is due to

- Channel random fluctuations (or multipath fading), widely studied in outdoor and indoor environments[14]
- System impairments,[15] such as differences in target device types, user orientation, environmental changes, mobile devices in different places, or heights (pockets, bags, hands, etc.)

Therefore, the RSSI measures are modeled with an additional random log-normal component, uncorrelated to the channel shadowing component in equation (2)

$$RSSI_{MEAS}(d) = RSSI(d) + W \tag{3}$$

where $W \sim \mathcal{N}(0, \sigma_W^2)$. The experimental measures in the area in Figure 1 have confirmed the validity of the model in equation (3) and they have returned the distributions of the RSSI levels from each beacon in the covered space; an example is reported in Figure 2. In order to extend the possibility of analysis and simulation to other environments and to different sizes, we have also used the experimental data for deriving the parameters of the model in equation (2), separately for the signal coming from each beacon, by means of a linear regression. The numerical values, averaged among all the beacons for the sake of brevity, are reported in Table 1.

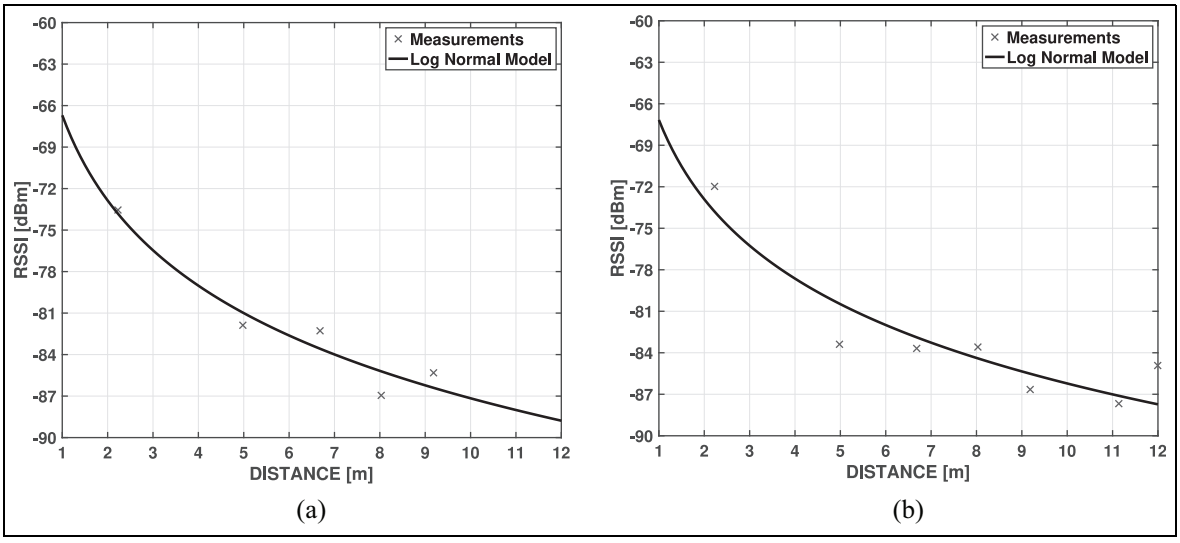

Figure 2. RSSI levels from beacons 3 and 5 in the environment used for the experimental measures: (a) beacon 3 and (b) beacon 5.

Table 1. Values of the channel model derived from the experimental measures.

Parameter	Value
Average loss at the reference distance L_0 (dB)	−76.0
Pathloss exponent α	2.04
Reference distance d_0 (m)	1.7
RSSI standard deviation σ_{SH} (dB)	0
RSSI online standard deviation (error) σ_W (dB)	4.02

RSSI: received signal strength indicator.

Review of FP techniques

Conventional localization algorithms using signal information like ToA, RSSI, or AoA face a serious performance degradation in indoor environments affected by phenomena like harsh multipath and NLoS. Taking advantage of LD features of the radio signal, there can exist a radio map containing LD parameters measured in predetermined points called grids so that target position can be estimated using pattern matching algorithms. In IPS based on radio frequency (RF) technologies such as WLAN or BLE, FP methods are among the most used,[3] thanks to their simplicity and reliability. There is a variety of measurements that can be used. The most common is the RSSI, but also signal to noise ratio (SNR), link quality information (LQI), channel impulse response, and others can be exploited.

FP is implemented in two basic steps: in the first step, which is called off-line or training phase, LD parameters of the received signal are measured in a grid-based map over the surveyed area; these are stored, and they form the so called radio-map. In the second phase, also called online phase, the target position is estimated by pattern matching between ongoing measurement of LD parameters and the stored radio-map.

The construction of the radio-map begins by dividing the area of interest into cells with the help of a floor plan. The RSSI values of the radio signals transmitted by beacons are collected by a test target inside the cells (or calibration points $\mathbf{P}_k = \{x, y\}_k$) for a certain period of time and stored into a database. The kth element ($k = 1, \ldots, K_T$) in the radio-map has the form

$$\mathcal{M}_k = \left\{\mathcal{R}_k = \{r_k(1), \ldots, r_k(N_B)\}\ ,\ (x,y)_k\right\} \quad (4)$$

where $\mathcal{R}_k$ is the fingerprint vector of measured RSSI from the beacons, and $(x,y)_k$ is the location of the kth fingerprint. The database term $\mathcal{M}_k$ can contain further information, such as orientation or others indicators. The radio-map can be modified or pre-processed before applying it in the location estimation phase. The reason can be the reduction of the memory requirements and/or of the computational cost of location estimation. In addition, different location estimation methods use different characteristics of the fingerprint histogram, such as the mean and the variance.

During the *online* phase, the target collects a vector of measurements (here RSSI) from the beacons

$$\tilde{\mathcal{R}} = \{\tilde{r}_1, \ldots, \tilde{r}_{N_B}\} \quad (5)$$

In order to estimate the position of the target user $\mathbf{p}^{MS}$, two main approaches are used:

- *Deterministic*: the position of the target user $\mathbf{p}^{MS}$ is not considered as a random vector.[16] The main objective is to estimate $\hat{\mathbf{p}}^{MS}$ at each time

step. Usually, the estimate is a convex combination of the calibration points $\mathbf{P}_k$, that is

$$\hat{\mathbf{p}}^{MS} = \sum_{k=1}^{K_T} \frac{w_k}{\sum_{j=1}^{K_T} w_j} \mathbf{P}_k \tag{6}$$

where w_k are the weights applied to the kth calibration point, that can be inversely proportional to the RSSI norms, or

$$w_k = \frac{1}{\| \mathcal{R}_k - \tilde{\mathcal{R}} \|} \tag{7}$$

where $\| . \|$ is a norm, for example, the Euclidean.

The estimation technique described in equation (6) is known as "Weighted K-Nearest Neighbor" (WKNN), and it is one of the most used by the FP algorithms. When all the calibration points use the same weight, it is called "K-NN," and when $K = 1$, it is denoted simply as "NN." In general, K-NN and WKNN can perform better than the NN method, particularly with parameter values $K = 3$ and $K = 4$.[16] However, if the density of the radio map is high, NN method can perform as well as more complicated methods.

- *Probabilistic*: the position of the target user $\mathbf{p}^{MS}$ is considered as a random vector. The idea behind the probabilistic approach is to compute the conditional (a posteriori) probability density function of the state from the measurements. The a posteriori pdf contains all the necessary information for computing an arbitrary estimate of the state and an estimate of the error. Two common estimators can be used: the maximum a posteriori (MAP) and the minimum mean square error (MMSE). The first computes the maximum of the a posteriori pdf and the second computes its mean.

FP-based methods produce accurate estimation of position in indoor environments;[17] they are easy to implement and the cost of the system is low since there is no need of further hardware, being RSSI measurements available in each radio technology. However, they have two main drawbacks: first, the off-line phase is laborious and time consuming, and changes in the environment can also compromise the overall system. Second, the vastness of the radio-map can make computationally heavy the on-line estimation, especially for IoT devices.

BFP design

In Mizmizi and Reggiani,[12] we have presented the BFP design, where the system is seen from a different point of view, related to a *binary code interpretation* of the FP scheme. The $\log_2(K_T)$ bits that enumerate the K_T grid positions and the corresponding FP signatures are now transformed, or *encoded*, in the $N_B \cdot \log_2(L)$ bits of each signature, where L is the number of levels used for each RSSI measure. The resulting code rate is defined as

$$R = \frac{N_B \cdot \log_2(L)}{\log_2(K_T)} \tag{8}$$

where, in the binary case ($L = 2$), we have

$$R = \frac{N_B}{\log_2(K_T)} \tag{9}$$

In order to design the BFP scheme, the main steps are

- Given the area of interest $\mathcal{A}$, which is an arbitrary polygon, we define the smallest square $\mathcal{S}$ with side length s that includes $\mathcal{A}$.
- Define the cell size Δ, so that the number of cells $\hat{K}$ obtained is a power of 2

$$\hat{K} = \frac{s}{\Delta} \tag{10}$$

- Define a beacons placement, according to an iterative procedure, similar to the Gray code encoding process,[18] which allows the elimination of the ambiguities among the binary representations of the cells and the achievement of a minimum Hamming distance, important for an effective distinction of the cells in the space.[12] For a generic number of beacons N_B, the Hamming distance respects the bound

$$D_H \leqslant \left\lfloor \frac{N_B}{N_{B,Min}} \right\rfloor = \left\lfloor \frac{N_B}{2 + \sum_{i=1}^{\log_4(K_T)-1} 4^i} \right\rfloor \tag{11}$$

The numbers of beacons for the minimum $D_H = 1$ and different number of cells are reported in Table 2.

- Quantization of each RSSI measures in 2 levels according to the threshold $RSSI_{REF}$, which, in general, depends on the beacons. Therefore, the RSSI coming from the bth beacon is quantized into

$$BFP(b) = \begin{cases} 1 & \text{if } RSSI_{measured} \geqslant RSSI_{REF,b} \\ 0 & \text{if } RSSI_{measured} < RSSI_{REF,b} \end{cases} \tag{12}$$

The threshold $RSSI_{REF,b}$ can be obtained through measurements at the borders that delimit the regions with label 1 and the regions with label 0 for each beacon (in practice, the final threshold is obtained by

Table 2. Number of beacons in order to guarantee $D_H = 1$ in the binary design for different cells grids.

Cells configuration	K_T	$N_{B,min}$ ($D_H = 1$)	W_1
2×2	4	2	4
4×4	16	6	24
8×8	64	22	104
16×16	256	86	424

The code weight W_1 is the number of signatures couples at $D_H = 1$.

averaging the values of the field measured on a set of points along the border). Of course, in the simulations, it can be computed using the model defined in equation (2).

- Discard all the cells that do not belong to the effective area of interest $\mathcal{A}$.

This design is referred to a static situation of the RSSI and of the beacons placement; in presence of a dynamic condition, caused by fluctuations or permanent changes of the channel propagation, the algorithm update is limited to the change of the N_B quantization thresholds $RSSI_{REF,b}$, one for each beacon. This is different from the type of update necessary in the conventional FP, where the K_T (generally more than N_B) single fingerprints of the database are updated periodically. Furthermore, it is possible to update the $RSSI_{REF,b}$ values by means of adaptive algorithms, which change the thresholds according to the errors in the cells selection recognized during the online phase.

Another aspect to remark is that, due to the irregularity of the real RSSI field (e.g. in presence of shadowing) and to the non-exact coincidence between the ideal designed borders (e.g. the lines delimiting the square cells considered here) and the RSSI field lines corresponding to the thresholds $RSSI_{REF,b}$, the real regions corresponding to the binary labels will not correspond exactly to the ideal cells. In our approach, this issue is faced by updating the cells to the new shapes after the computation of the thresholds $RSSI_{REF,b}$. Also, it is possible that more binary labels than those used in the ideal design could appear because of the irregular lines of the RSSI values. In this case, two options are possible: (1) to keep the new number of modified cells as the new label assignment in the region or (2) merge the new cells in order to obtain the same number K_T of cells in the original design.

Analysis of FP performance

The design of an FP-based IPS is a complex task since it is not limited to some general guidelines, but most of the time, it requires a tuning phase after that the system has been deployed, for example, it may be necessary to add or move some beacons, optimize the radio map, etc. The scope of the FP analysis is to help the designer, and to save time and costs. This is possible by predicting analytically the expected performance of the systems. In this section, we propose an analytical solution for both cases, $L = 2$ and $L = \infty$ methods; the proposed solutions are then validated through numerical simulations and experimental results as shown in the "BFP performance" section.

The main metrics we are considering in this analysis are the label selection error (LSE) and the mean square error (MSE). Therefore, given a FP-based localization system, with K_T labels (or cells), the total MSE is given by

$$MSE_{TOT} = \frac{1}{K_T} \sum_{k=1}^{K_T} MSE_k \tag{13}$$

where MSE_k is related to the error contributions coming from devices in the kth label (or cell)

$$MSE_k = \sum_{j=1, j \neq k}^{K_T} P_E(j|k) \cdot d_{j,k}^2 \tag{14}$$

where $d_{j,k}^2$ is the squared distance between the points corresponding to the jth and the kth labels and $P_E(j|k)$ is the conditioned LSE of j given k, or the probability to estimate the cell j when you are located in the cell k. The analytical model is based on the following assumptions:

- The NN estimation method is considered for intercepting a performance upper bound;
- The performance is computed limiting the target positions to the grid points at the centers of the cells. However, the analysis can be extended to all the points in each cell, also considering the issues related to the RSSI irregularities commented at the end of "BFP design" section, whose impact is higher for the points closer to the cells borders.

The performance estimation can be applied also to experimental RSSI measures, including the impact of correlated shadowing. After the off-line FP phase, there are K_T vectors of length N_B with the form

$$\mathcal{M} = \{\mathcal{R}_k, (x,y)_k\} \quad \text{for } k = 1, \ldots, K_T \tag{15}$$

The elements $\mathcal{R}_k(b)$ ($b = 1, \cdots, N_B$) of the vector $\mathcal{R}_k$ are assumed as the true mean of the RSSI from each beacon. Usually, this is achieved by collecting a large number of samples of the RSSI for each orientation of

the target. The elements of the RSSI measurements at an arbitrary grid point k are denoted as

$$\tilde{\mathcal{R}}_{\mathbf{k}} = \{\tilde{r}_{1,k}, \ldots, \tilde{r}_{N_B,k}\} \tag{16}$$

These measures can be obtained directly by experimental measures or they can be modeled according to equation (3). For the analysis, each component in equation (16) is assumed to be a random variable with the following assumptions:

- The random variables $\tilde{r}_{i,k}$ (in dBm) for all i are mutually independent;
- The random variables $\tilde{r}_{i,k}$ (in dBm) are normally distributed;
- The (sample) standard deviation of all the random variables $\tilde{r}_{i,k}$ is assumed to be identical and denoted by σ_W (in dBm).

For $L = +\infty$, the signal distance between the sample vector and the FP vectors is used to determine which of the points on the grid corresponds to the position of the target. The NN technique selects the (x, y) coordinates corresponding to the FP vector with the smallest signal distance from the sample vector as the estimated location.

For $L = 2$, the Hamming distance is evaluated between competing labels (after the RSSI quantization) and the NN technique selects the (x, y) coordinates corresponding to the BFP vector with the smallest Hamming distance from the received binary vector as the estimated location.

Case $L = +\infty$

The idea behind our approach comes from the signal theory and the corresponding geometrical interpretation of the FP technique. For the sake of simplicity, let us assume to have a system with two beacons: the radio map could be seen as a set of points scattered in the signal space as in Figure 3. Given the online FP, the estimated position is the FP with the minimum signal distance. From a geometric point of view, this means that the online FP belongs to the Voronoi region of the estimated FP. Therefore, the LSE probability of j given k in equation (14) can be computed as

$$P_E(j|k) = \int\int_{V_j} f_k\left(\tilde{\mathcal{R}}_{\mathbf{k}}\right) dV_j \tag{17}$$

where $f_k(\tilde{\mathcal{R}}_{\mathbf{k}})$ is the probability density function of the kth FP, and it is given by a multivariate normal distribution, thanks to equation (3), and V_j is the Voronoi region related to the jth FP

$$V_j = \left\{\tilde{\mathcal{R}} : \| \mathcal{R}_j - \tilde{\mathcal{R}}\|_2 < \| \mathcal{R}_k - \tilde{\mathcal{R}}\|_2\right\} \ \forall k \tag{18}$$

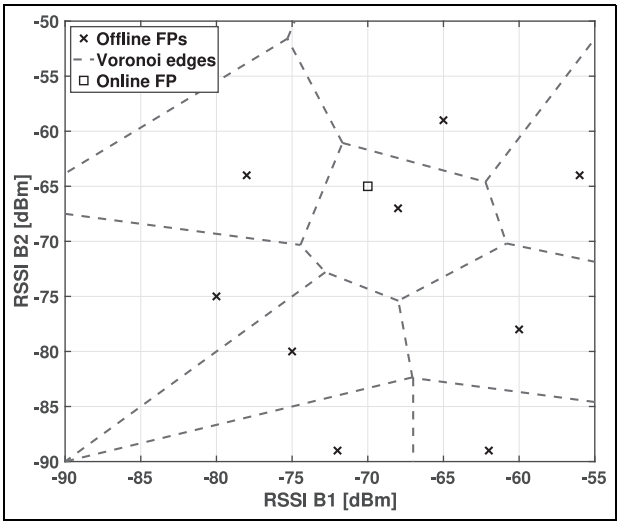

Figure 3. Example of 2D radio map with the corresponding Voronoi's regions.

The integral in equation (17) can be very complex, also with numerical approximated methods, especially when the number of beacons grows. In Swangmuang and Krishnamurthy,[19] they propose to use the concept of proximity graph in order to compute a lower bound solution to the probability of selecting correctly a FP and therefore to optimize the radio-map. However, to give a more general analysis, we are interested in predicting the MSE or the LSE. Therefore, our solution approximates the Voronoi's regions with hypercubes in the multidimensional space, making their calculation possible in all the cases, since the multi-dimensional integral reduces to the product of mono-dimensional integrals. The proposed approach can be summarized as follows:

- Compute the minimum signal distance

$$D_{min} = min\{D_{j,k}\} = min\{\| \mathcal{R}_j - \mathcal{R}_k\|_2\} \ \forall j, k \tag{19}$$

- For each jth FP, define a hypercube with side equal to the minimum signal distance in equation (19) and centered at the jth FP;
- Compute the LSE probabilities in equation (17) integrating the multivariate normal in the regular hypercube;
- Compute the final MSE with equations (14) and (13).

Case $L = 2$

In this case, which can be extended to the non-binary case with $L>2$, we exploit the error correcting code

interpretation. The error probability is estimated by the following steps:

- For each FP vector $\mathcal{R}_k$ in $\mathcal{M}$, compute its binary version *BFP*, according to equation (12)
- The Hamming distance $D_{H,j,k}$ between the kth and the jth FP vector is computed as

$$D_{H,j,k} = \sum_{b=1}^{N_B} \left| BFP_j(b) - BFP_k(b) \right| \tag{20}$$

- The probability $P_E(j|k)$ in equation (14) is computed by considering the $N_{j,k}$ error events $\mathcal{E}_{j,k}^{(i)}$ $(i = 1, \cdots, N_{j,k})$ that cause the mis-detection in the cell j given the considered cell k. If, for example, the Hamming distance $D_{H,j,k} = 1$, $N_{j,k} = 1$ and the error event $\mathcal{E}_{j,k}^{(1)} = \{E_{j,k}^{(1)}(1), \cdots, E_{j,k}^{(1)}(N_B)\}$ will be constituted by the components $E_{j,k}^{(1)}(b) = BFP_j(b) - BFP_k(b)$, all zeros except for one. If we denote with $E_0^{(i)}$ the set of bits in $\mathcal{E}_{j,k}^{(i)}$ equal to 0 (correct bits) and with $E_1^{(i)}$ those equal to ± 1 (incorrect bits), the error probability will be approximated by

$$P_E(j|k) \approx \sum_i P\left(\mathcal{E}_{j,k}^{(i)}\right) = \sum_i \left(\prod_{b \in E_1^{(i)}} p\left(E_{j,k}^{(i)}(b)\right) \prod_{b \in E_0^{(i)}} \left(1 - p\left(E_{j,k}^{(i)}(b)\right)\right) \right) \tag{21}$$

where $p(E_{j,k}^{(i)}(b))$ is the probability of an incorrect beacon detection (i.e. the bit assignment based on equation (12) turns out to be incorrect) and it is affected by the assumptions made on RSSI, which is computed according to equation (3), that is

$$p\left(E_{j,k}^{(i)}(b)\right) = Q\left(\frac{|\mathcal{R}_k(b) - RSSI_{REF,b}|}{2\sigma_W}\right) \tag{22}$$

- When all the set of LSE probabilities $P_E(j|k)$ are computed, equations (14) and (13) are applied for obtaining the final MSE.

To extend the proposed method for considering not only the center of each cell but any random position, equation (22) has to be averaged over the entire cell changing the signal distance at the numerator; it is clear that the points closer to the cell borders will show the highest error contributions. In order to facilitate the integral, it is possible to use the distributions of the signal derived in Appendix 1, which can be considered approximately Gaussian when the cell size is small.

BFP performance

The main advantage from reducing the RSSI to a binary representation is the increment of the computational efficiency. BFP provides two advantages from the computational point of view:

- Decrease of the storage memory (and/or of the time necessary for the transmission of the data necessary for positioning from/to the target devices), thanks to the binary representation of the FP.
- Increase of the algorithmic efficiency, which can be evinced directly from the iterative code construction presented in "BFP design" section: the signature search can proceed in a *logarithmic way*, starting from the most significant bits (the first beacons, with the longest transmission range), halving at each bit the search area, and finishing with the least significant bits, corresponding to the fine division among single cells. Therefore, the logarithmic search is computationally more efficient since the best solutions can be found in $\log_2(K_T)$ steps instead of K_T. In addition, it has an interesting property with respect to the sequential search, from an application point of view: it provides a scalable localization precision, since each algorithm step provides a finer resolution of the area in which the target could be located. The localization process could stop at a number of bits lower than N_B, according to the application or the context.

The validation of the proposed technique is carried out through real experiments and numerical simulations. The case with $L = \infty$ is reported as a benchmark.

Experimental validation

The experimental measurements are collected in a class room of the Politecnico di Milano as described in Figure 1. The beacons used are BLE sensors from Silicon Labs,[20] while the target user is the BLE 112 development board.[20] The beacons transmit continuously BLE packets that contain the MAC address of the transmitter, the receiver board, store for each received packet, the ID of the sender, the RSSI estimated, and the time-stamp. For each of the 16 FP points, we have measured approximately 3000 samples from each beacon (5 min, with a rate of 10 measurements/s).

The results in Figure 4 show the cumulative distribution function of the error from real experiment and numerical simulation, while the analysis for both cases, under the same conditions, shows an average error of

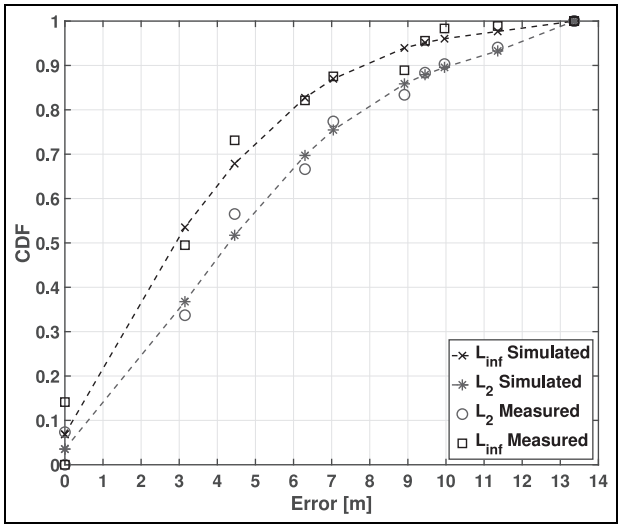

Figure 4. Experimental results obtained in the environment shown in Figure 1.

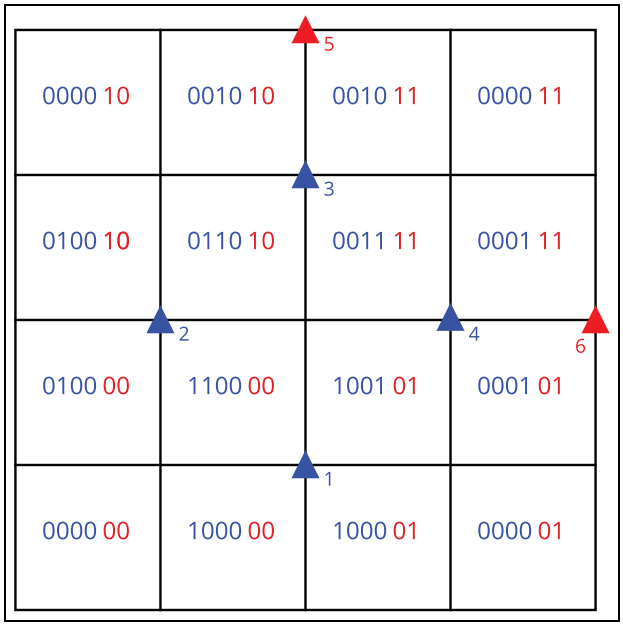

Figure 5. Reference environment for the numerical results from the channel model.

4.6 m for $L = 2$ and 3.1 m for the benchmark, which is in line with the experimental findings.

Numerical validation

In order to extend the investigation of the impact of the main design or channel model parameters on the MSE, we are making the following assumptions for the remaining simulations and analysis, unless differently specified:

- The parameters of the channel model are taken from Table 1. These are derived by analyzing the experimental measurements;
- Elements of the sample vectors $\tilde{\mathcal{R}} = \{\tilde{r}_1, \ldots, \tilde{r}_b, \ldots, \tilde{r}_{N_B}\}$ are assumed independent, with a standard deviation σ_W (dB);
- A square grid $K_T = 4 \times 4$ is considered, with a resolution Δ equal to 3.15 m (as in Figure 1). The reference scenario is a square room with room side $s = 12.6$ m;
- The algorithm NN is used to estimate the target position;
- The binary labels derived for the scenario under consideration are reported in Figure 5;
- Ideal receivers sensitivity ($-\infty$);
- In the simulations, the target locations are extracted randomly among the cells centers and the number of runs is 10^4 for each point in the plot.

Numerical results in Figures 6 and 7 show how the MSE varies as a function of the standard deviation σ_W of the measurement error and of the pathloss exponent, respectively. Each plot compares analytical and

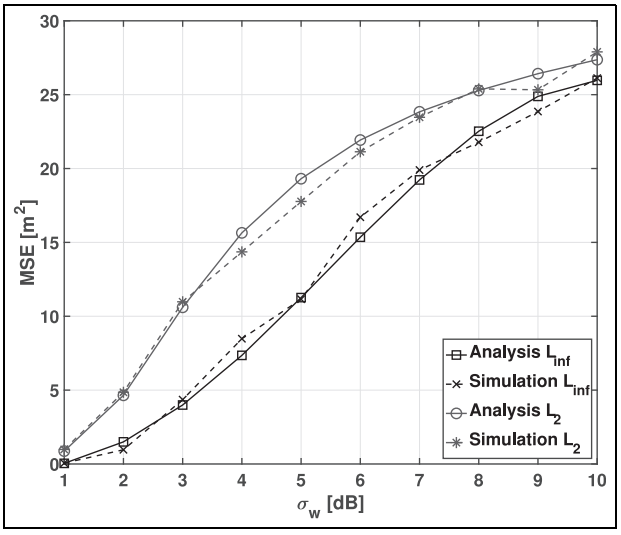

Figure 6. MSE obtained by simulations (dashed lines) and analysis (continuous lines) as a function of the measurement error standard deviation σ_W.

simulated results with $L = 2$ and $L = \infty$. From this results, we can observe that

- When the propagation becomes more difficult, that is, a higher standard deviation of the measurement error σ_W, the MSE in both cases decreases. However, the gap between the two techniques is lower when σ_W is low or high. This happens since the LSE probabilities become

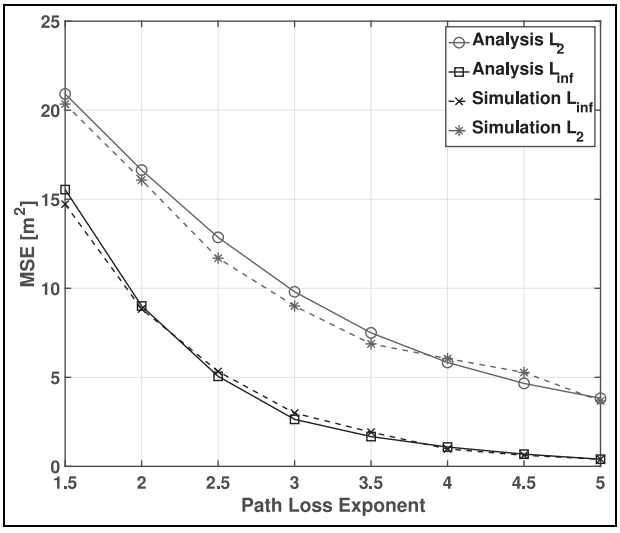

Figure 7. MSE obtained by simulations (dashed lines) and analysis (continuous lines) as a function of the pathloss exponent α.

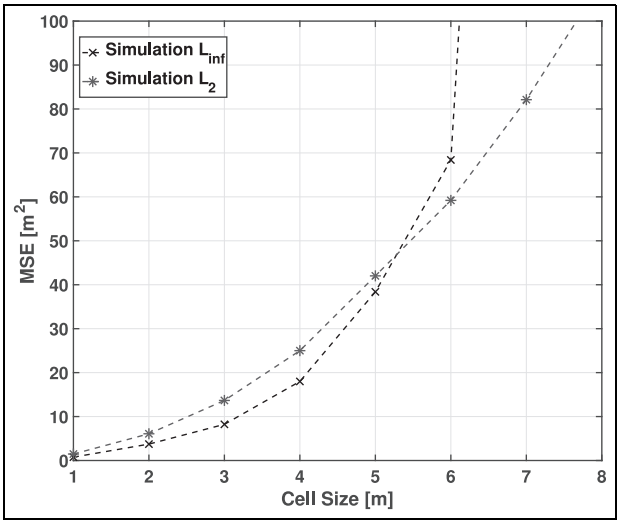

Figure 8. MSE obtained by simulations as a function of the cell size, and for a fixed receiver sensitivity equal to -92 dBm.

comparable and very low or high in these cases, respectively;

- The performance gap around 2 dB between $L = +\infty$ and $L = 2$ can be compensated by a greater number of measures to be averaged in the binary case. In fact, averaging three independent RSSI measures decreases the effective σ_W of about 2 dB. This approach generates also the necessity of a trade-off between positioning quality and latency.
- The Figure 7 shows that, as the propagation becomes more difficult, or the pathloss increases, the MSE falls in both cases. This effect is due to the increase of the average signal distance between all the FPs, which decreases the probability of error.

Finally, Figure 8 shows an interesting property of BFP with respect to $L = +\infty$: when the receivers sensitivity is not ideal but limited, the binary case appears more robust since, when the received signal from a beacon decreases below the minimum detectable level, the corresponding bit is automatically set to zero, which is coherent with the corresponding binary representation of the labels. On the contrary, the $L = +\infty$ case suffers from a reduced precision of the missing RSSI measures, which can be appreciated when the room size increases and some of the farthest beacons signals decrease below the minimum detectable level.

Conclusion

In this article, we have developed the performance analysis in case of binary quantization in the RSSI signatures for FP localization. The study has exploited the design principles of BFP and the related similarities with the binary codes theory. In fact, using a single bit to represent the RSSI, it is possible to make the layout design starting from the concept of Hamming distance between the vectors of the radio map, which is directly related to the localization performance. The analysis and the simulations have revealed the performance compromise between BFP and the ideal case without quantization and the impact of the main channel parameters as well. The BFP looks computationally efficient, often with comparable performance with respect to conventional FP and suited to scenarios in which the computational and storage simplicity are the primary design factors, as for BLE devices, radio frequency identification (RFID) tags, or microsensors. Nevertheless, it is always possible to tune a performance compromise between the number of beacons and the number of measures for each position estimate with a consequent impact on the latency.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Marouan Mizmizi https://orcid.org/0000-0003-4157-2577

References

1. Wirola L, Laine T and Syrjärinne J. Mass-market requirements for indoor positioning and indoor navigation. In: *Proceedings of the 2010 international conference on indoor positioning and indoor navigation (IPIN)*, Zurich, 15–17 September 2010. New York: IEEE.
2. Liu H, Darabi H, Banerjee P, et al. Survey of wireless indoor positioning techniques and systems. *IEEE T Syst Man Cybern: Appl Rev* 2007; 37(6): 1067–1080.
3. Gu Y, Lo A and Niemegeers I. A survey of indoor positioning systems for wireless personal networks. *IEEE Commun Surv Tut* 2009; 11(2): 13–31.
4. Davidson P and Pich é R. A survey of selected indoor positioning methods for smartphones. *IEEE Commun Surv Tut* 2017; 19(2): 1347–1370.
5. He S and Chan SHG. Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Commun Surv Tut* 2016; 18(1): 466–490.
6. Wu Z, Han Y, Chen Y, et al. A time-reversal paradigm for indoor positioning system. *IEEE T Veh Technol* 2015; 64(4): 1331–1339.
7. Jin Y, Soh W and Wong W. Indoor localization with channel impulse response based fingerprint and nonparametric regression. *IEEE T Wirel Commun* 2010; 9(3): 1120–1127.
8. Lin K, Chen M, Deng J, et al. Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings. *IEEE T Automat Sci Eng* 2016; 13(3): 1294–1307.
9. Arya A, Godlewski P and Melle P. A hierarchical clustering technique for radio map compression in location fingerprinting systems. In: *Proceedings of the 2010 IEEE 71st vehicular technology conference*, Taipei, Taiwan, 16–19 May 2010, pp.1–5. New York: IEEE.
10. Saha A and Sadhukhan P. A novel clustering strategy for fingerprinting-based localization system to reduce the searching time. In: *Proceedings of the 2015 IEEE 2nd international conference on recent trends in information systems* (Retis), Kolkata, India, 9–11 July 2015, pp.538–543. New York: IEEE.
11. Mizmizi M and Reggiani L. Design of RSSI based fingerprinting with reduced quantization measures. In: *Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcala de Henares, 4–7 October 2016, pp.1–6. New York: IEEE.
12. Mizmizi M and Reggiani L. Binary fingerprinting-based indoor positioning systems. In: *Proceedings of the 2017 international conference on indoor positioning and indoor navigation (IPIN)*, Sapporo, Japan, 18–21 September 2017, pp.1–6. New York: IEEE.
13. Goldsmith A. *Wireless communication*. 1st ed. Cambridge: Cambridge University Press, 2005.
14. Cotton SL and Scanlon WG. A statistical analysis of indoor multipath fading for a narrowband wireless body area network. In: *Proceedings of the 2006 IEEE 17th international symposium on personal, indoor and mobile radio communications*, Helsinki, Finland, 11–14 September 2006, pp.1–5. New York: IEEE.
15. Kim Y, Shin H, Chon Y, et al. Smartphone-based Wi-Fi tracking system exploiting the RSS peak to overcome the RSS variance problem. *Pervasive Mob Comput* 2013; 9(3): 406–420.
16. Honkavirta V, Perälä T, Ali-Löytty S, et al. A comparative survey of WLAN location fingerprinting methods. *In: Proceedings of the 6th workshop on positioning, navigation and communication*, Hannover, 19 March 2009. New York: IEEE.
17. Seco F, Jimenez A, Prieto C, et al. Survey of mathematical methods for indoor localization. *In: Proceedings of the 6th IEEE international symposium on intelligent signal processing*, Budapest, 26–28 August 2009. New York: IEEE.
18. Fischmann AF. A gray code counter. *IRE Trans Electr Comput* 1957; 6(2): 120–120.
19. Swangmuang N and Krishnamurthy P. An effective location fingerprint model for wireless indoor localization. *Pervasive Mob Comput* 2008; 4(6): 836–850.
20. Silicon Labs. Bluegiga, 2015, https://www.silabs.com/

Appendix 1

When the target location is in any position of the cell, the approximate analysis relies on the derivation of the signal distance probability density function. Let us assume that target location $p = (x_p(l), y_p(l))$, in the cell l, has coordinates x_p and y_p independent and uniformly distributed between a_x, b_x and a_y, b_y, respectively.

$$x_p \sim \mathcal{U}(a_x, b_x) \tag{23}$$

$$y_p \sim \mathcal{U}(a_y, b_y) \tag{24}$$

If we consider N_B beacons, the square of the physical distances between each beacon b and the random point p is

$$\begin{aligned} d_{Fp}(b)^2 &= \left(x_p(l) - x_b\right)^2 + \left(y_p(l) - y_b\right)^2 \\ &= X_{pb}(l)^2 + Y_{pb}(l)^2 \qquad b = 1, \ldots, N_B \end{aligned} \tag{25}$$

where $X_{pb}(l)$ and $Y_{pb}(l)$ are again random variables uniformly distributed. The mean value and the variance of X_{pb}^2 and Y_{pb}^2 are

$$m_{X_{pl}^2}(b) = E\left[X_{pl}(b)^2\right] = \frac{a_{x_b}^2 + b_{x_b}^2 + a_{x_b} b_{x_b}}{3} \tag{26}$$

$$\begin{aligned} \sigma_{X_{pl}^2}^2(b) &= E\left[X_{pl}(b)^4\right] - E\left[X_{pl}(b)^2\right]^2 \\ &= \frac{4a_{x_b}^4 + 4b_{x_b}^4 - 6a_{x_b}^2 b_{x_b}^2 - a_{x_b}^3 b_{x_b} - a_{x_b} b_{x_b}^3}{45} \end{aligned} \tag{27}$$

and similarly for $Y_{pl}^2(b)$. So, being x_p and y_p independent

$$m_{d_{Fp}^2}(b) = m_{X_{pl}^2}(b) + m_{Y_{pl}^2}(b) \tag{28}$$

$$\sigma_{d_{Fp}^2}^2(b) = \sigma_{X_{pl}^2}^2(b) + \sigma_{Y_{pl}^2}^2(b) \tag{29}$$

and the transformation from physical distance to signal follows as

$$\begin{aligned} RSS_{pl}(b) &= -10 \cdot \alpha \cdot \log_{10}\left(d_{Fp}(b)\right) + W \\ &= -\frac{10\alpha}{2ln(10)} \cdot ln\left(d_{Fp}^2(b)\right) + W = PL(b) + W \end{aligned} \tag{30}$$

where $W \sim \mathcal{N}(0, \sigma_W^2)$ is the measurement error and $PL(b)$ is the path loss. Using the approximations

$$E[g(X)] \approx g(m_X) + \frac{g''(m_X)}{2}\sigma_X^2 \tag{31}$$

$$\sigma_{g(X)}^2 \approx \left(g'(m_X)\right)^2 \sigma_X^2 \tag{32}$$

where $g(X) = -\frac{10\alpha}{2ln(10)} ln(d_{Fp}^2(b))$, we obtain the mean value and variance of $PL(b)$ as

$$m_{PL}(b) \simeq -\frac{10\alpha}{2ln(10)} ln\left(m_{d_{Fp}^2}(b)\right) + \frac{10\alpha}{4ln(10)} \frac{\sigma_{d_{Fp}^2}^2(b)}{m_{d_{Fp}^2}^2(b)} \tag{33}$$

$$\sigma_{PL}^2(b) \simeq \left(\frac{10\alpha}{2ln(10)}\right)^2 \frac{\sigma_{d_{Fp}^2}^2(b)}{m_{d_{Fp}^2}^2(b)} \tag{34}$$

We observe that $PL(b)$ results approximately Gaussian in the logarithmic domain if the cell is small enough and its variance turns out to be small with respect to the mean value $m_{PL}(b)$.