Scalable, resource and locality-aware selection of active scatterers in Geometry-based stochastic channel models

Benjamin Rainer[1], Markus Hofer[1], Stefan Zelenbaba[1], David Löschenbrand[1], Thomas Zemen[1], Xiaochun Ye[2], Peter Priller[3]
[1] Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna, Austria
[2] Institute of Computing Technology, CAS, Bejing, China
[3] AVL List GmbH , Hans-List-Platz 1, 8020 Graz, Austria

***Abstract*—In this paper we adopt and modify a well-known locality-aware hashing scheme to the problem of active stochastic scatterer selection in vehicular non-stationary geometry-based stochastic channel models (GSCM). We show, how under relaxed assumptions on the query set an efficient selection of active stochastic scatterers during simulation is computationally feasible. The proposed approach enables real-time simulation and emulation of large-scale GSCMs by restricting the active stochastic scatterer set to meet given resource constraints. We showcase our approach by introducing a GSCM that is bootstrapped via OpenStreetMap data. The stochastic scatterers are placed automatically along buildings, traffic signs and vegetation. We validate and investigate the impact of the proposed approach on the accuracy of a GSCM by means of second order statistics of the time- and frequency-varying fading process. For validation and performance evaluation we parameterize our GSCM using a vehicular wireless channel measurement campaign conducted in the inner city of Vienna. The impact of selecting only a subset of scatterers is then evaluated using the calibrated GSCM.**

I. Introduction

Testing, verification and validation of vehicle-to-infrastructure and vehicle-to-vehicle (V2X) communication systems is seeing increased importance, in both, science and industry. Due to safety regulations and the lack of reproducibility, testing V2X communication systems cannot be conducted on the road in urban areas. Therefore, a reliable, real-time capable, and as realistic as possible digital twin is needed throughout all the ISO/OSI layers. In this paper we focus on the simulation of wireless communication channels by means of geometry-based stochastic channel models in urban areas. Thus, providing the foundation for a realistic and real-time capable hardware in the loop simulation/emulation of V2X communication systems.

Geometry-based stochastic channel models (GSCM) have been successfully applied to a variety of scenarios ranging from simple to highly mobile scenarios [1]. GSCMs use simplified ray tracing methods (e.g, 2D ray tracing, 2.5D ray tracing) and model the non line-of-sight components, such as those caused by diffuse scattering, penetration/refraction, and diffraction at edges by stochastic scatterers. The parameters of the scatterers' distributions are obtained by measurements. Due to the nature of GSCMs these models can be computed very fast, even satisfying the constraints for real-time channel emulation, trading-off complexity and accuracy, recently shown in [2]. However, real-time capable GSCMs represent small areas with a low number of scatterers leading to low accuracy.

An approach to reduce the computational complexity of diffuse scattering in GSCMs is presented in [3]. The complexity reduction of diffuse scattering has been mainly accomplished by the use of discrete prolate spheroidal sequences (DPSS) and their approximate basis coefficient calculation introduced in [4]. In [5] a GSCM for high speed railway is presented and the notion of active scattering region (ASR) and geometry clusters are introduced. The introduced active scattering regions shall provide means to model the (diffuse) scattering from the environment within certain geometry clusters holding a certain type of scatterers. However, the parameters (e.g., life time/distance, size/shape, energy distribution) of such geometry clusters and ASRs are derived from measurements. In this paper we seek for an scatterer selection approach which is independent of the actual scenario.

The scenarios modeled with GSCMs are currently rather limited in extend (e.g., short highway sections, simplified intersections, etc.). However, for assessing the wireless communication performance in a V2X scenario we need to cover more than a few intersections or road canyons. Ideally, we would like to dynamically load geometry data (e.g., from OpenStreetMap) being able to model the wireless communication performance even for a city. GSCMs covering large geographical areas face two problems. First, the problem of placing the stochastic scatterers (i.e., traffic signs, vehicles, buildings). Second, the computational complexity increases tremendously due to the vast amount of stochastic scatterers in the scenario. In this paper we approach both of these problems. The first problem is solved by automatically placing the stochastic scatterers using map data and the second one by introducing a well-known hashing concept from computer science to GSCMs, allowing to select relevant scatterers with respect to given resource constraints, i.e., maximum number of stochastic scatterers which can be evaluated in real-time.

We introduce a method which uses OpenStreetMap and

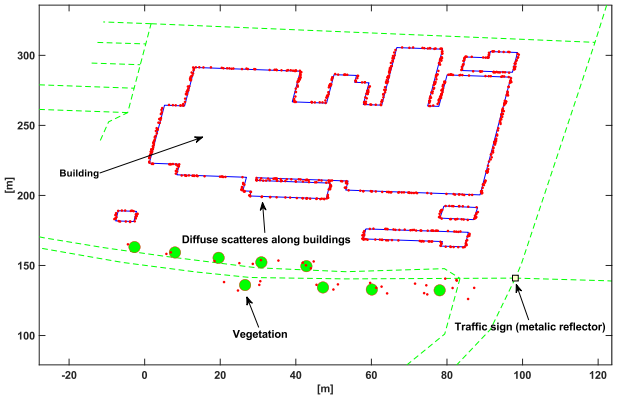


Fig. 1: Example of an automatically initialized geometry in our GSCM.

GPS data to place different types of stochastic scatterers. The proposed GSCM and its scatterer selection scale in accuracy and computational efficiency with respect to the given resource constraints. We validate our approach by modeling a measurement campaign of an urban scenario investigating the impact of selecting a subset of stochastic scatterers on the resulting statistics of the fading process such as RMS delay and Doppler spread. The scientific contributions of the paper are as follows:

- We present a GSCM that can be bootstrapped by OpenStreetMap data and optional GPS trajectories.
- We formulate the search and selection of active (diffuse) scatterers as an approximate nearest neighbor problem and solve it using locality-sensitive hashing (LSH).
- We extend this locality aware selection of scatterers to the case of a constant number of query points (transmitters and receivers) by projection to a higher dimensional vector space.
- We validate the proposed approach by investigating its impact on the statistics of the fading process.

II. Geometry-based stochastic channel model

We briefly review how we proceeded in [6]. We use OpenStreetMap data for automatically instantiating our GSCM which is implemented using MATLAB. Besides buildings, roads, rail tracks, traffic lights/signs and vegetation, we also provide the possibility of importing GPS trajectories for defining the mobility of mobile discrete scatterers (e.g., cars, pedestrians, etc.), receivers and transmitters. If the GPS sampling rate is lower than the assumed stationarity region duration [7] we interpolate the trajectories using splines and use the given stationarity duration as supporting points. This leads to a smoother simulated time- and frequency-variant channel response. Diffuse scatterers are placed alongside building as described in [6]. Fig. 1 depicts an example geometry, extracted from OpenStreetMap data. Metallic objects like traffic signs and lights are represented as static discrete scatterers. Vegetation is modeled by diffuse scatterers such as the scattering from walls (automatically and i.i.d placed).

Our channel model differentiates between the following contributions: LOS between Tx and Rx, reflections from static discrete scatterers (SD), reflections from moving discrete scatterers (MD) and diffuse scattering (D). We assume a wide sense stationary fading process with uncorrelated scattering for a given period of time/spatial region, henceforth referred to as stationarity region. The discrete time-variant frequency response for a single stationarity region is calculated by

$$\begin{aligned} h[x,w] = \eta^{\mathrm{LOS}} e^{-j2\pi w \Delta f \tau^{\mathrm{LOS}}[x]} + \\ \sum_{i=1}^{N_{\mathrm{SD}}} \eta_i^{\mathrm{SD}} e^{-j2\pi (f+w\Delta f)\tau_i^{\mathrm{SD}}[x]} + \\ \sum_{i=1}^{N_{\mathrm{MD}}} \eta_i^{\mathrm{MD}} e^{-j2\pi (f+w\Delta f)\tau_i^{\mathrm{MD}}[x]} + \\ \sum_{i=1}^{N_{\mathrm{D}}} \eta_i^{\mathrm{D}} e^{-j2\pi (f+w\Delta f)\tau_i^{\mathrm{D}}[x]}, \end{aligned} \tag{1}$$

where discrete time $x \in \{0, \ldots, X-1\}$, X is the number of samples within one stationarity region, and discrete frequency $w \in \{-\lfloor \frac{B}{2\Delta f} \rfloor, \ldots, \lfloor \frac{B}{2\Delta f} \rfloor - 1\}$, f denotes the carrier frequency and Δf denotes the subcarrier spacing. The computational complexity of using a sum of exponentials can be significantly reduced by using approximate discrete prolate sequences [4]. But, still the number of scatterers heavily impact the computational complexity. If we can bound the number of scatterers such that we do not loose much accuracy, an optimized implementation of (1) becomes feasible. Even a implementation in a FPGA is possible if we have a fixed bound on the number of scatterers.

III. Locality Aware Active Scatterer Selection

For the rest of this paper let's assume that we are capable of computing (1) and the intersection algorithm in real-time if the number of (diffuse) scatterers does not exceed $M \in \mathbb{N}$. In [5] the concept of active scattering regions has been introduced. However, as discussed before these strongly depend on their parameters extracted from the measurement and, thus, the ASRs have to be recomputed and extracted even if the scenario changes slightly. We pursuit a more general approach. Since we place the (diffuse) scatteres in the geometry, we need an efficient possibility to select (diffuse) scatterers depending on a specific criteria, general enough to be used in every scenario and without the need of parameterization from measurements. A criteria that suits our goal is the path loss of a plane wave component that originates from scattering. We want to select those scatterers for which the path loss between Rx and Tx (or any number of query points) are below a certain threshold (radius of the ball around each scatterer). Figure 2 depicts this idea of using the path loss of a path Tx $\rightarrow$ scatterer $\rightarrow$ Rx. We do only consider the scatterer for further computation if both, the transmitter and the receiver are contained within its r-ball.

More formally:

Problem 1. *Let $P \subset \mathbb{R}^m$ with $|P| = n$, $Q(t) \subset \mathbb{R}^m$ be a set of query points at time $t \in \mathbb{R}^+$ and $d : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^+$ be*

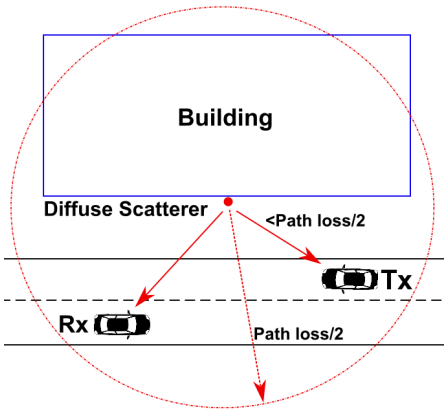


Fig. 2: Example of an r-ball with a diffuse scatterer being its center. The scatterer is only taken into account if both query points (Tx and Rx) are within its r-ball.

a metric. We want to find a set G of at most $M \in \mathbb{N}$ points which fulfill:

$G(t) = \bigcap_{\boldsymbol{q} \in Q} \{\boldsymbol{p} \in P : \boldsymbol{q}(t) \in B(\boldsymbol{p}, r)\}$, *where* $B(\boldsymbol{p}, r) := \{\boldsymbol{x} \in P \cup Q : d(\boldsymbol{p}, \boldsymbol{x}) < r\}$ *and* $r > 0$.

Problem 1 states that given a set $P \subset \mathbb{R}^m$ of scattering points we want to obtain those scattering points for which the points in Q are included in the same r-balls. $G(t)$ denotes the set of scatterers for which the transmitter and receivers are included in their respective r-ball at time instant t. Please note, P has no dependence on time and is of constant size. Thus, we would like to have a data structure which allows a pre-computation which must not be efficient but in turn yields **a computational fast lookup during simulation**. Here we want to emphasize that the lookup shall only return M possible scatterers, as M depicts our resource constraint such that the channel impulse response can be evaluated in real-time.

We observe that the dimension of our data is $m = 2$ or $m = 3$ and $d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x} - \boldsymbol{y}||_p$. The data structure which would fit such a problem is the minimal spanning tree (MST) but since our query points' position changes over time, and so do the distances between the query points and the scatterers, which would require the MST to be rebuilt for every lookup. These thoughts instantly lead to hash functions. More specifically, in our case we need hash functions which are locality sensitive/aware with respect to a metric according to Def. III.1 presented below [8].

A. LSH for active scatterer selection

Before we dive into the details, let us review the definition of locality sensitive hash functions [8]:

Definition III.1. *A family $\mathcal{H} = h : S \to U$ of hash functions is called (r, cr, p_1, p_2)-sensitive for a metric $d : S \times S \to \mathbb{R}^+$ if $\forall v, q \in S$:*

i) if $q \in B(v, r) \implies \mathbb{P}(h(v) = h(q)) \geq p_1 \ \forall h \in \mathcal{H}$,
ii) if $q \notin B(v, cr) \implies \mathbb{P}(h(v) = h(q)) \leq p_2 \ \forall h \in \mathcal{H}$.

A locality-sensitive hash (LSH) family is only useful if the probabilities satisfy $p_1 > p_2$ and the radius satisfies $r < cr$ which we establish by setting $c := 1 + \varepsilon$, $\varepsilon > 0$. In our case we want to exploit the property of not being one-to-one such that the hash function(s) shall collide (project onto the same value) if points are *near* otherwise they should not collide. The interpretation of Def. III.1 is as follows. If $\boldsymbol{v}$ is in the r-ball of $\boldsymbol{u}$ their hashes shall collide with probability of at least p_1. Conversely, if $\boldsymbol{v} \notin B(\boldsymbol{u}, cr)$ we expect a collision probability of smaller than p_2.

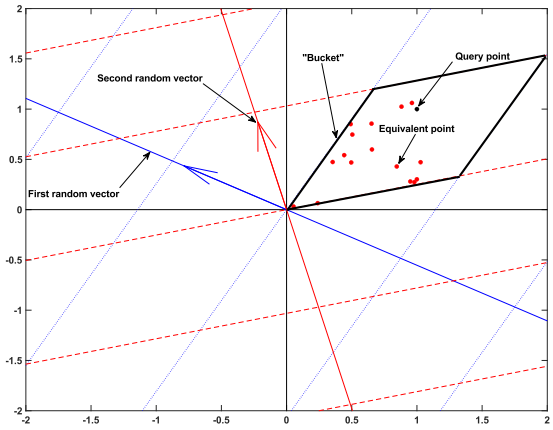


Fig. 3: LSH example with query point $(1, 1)^T$.

In order to apply the LSH scheme presented in [8] we construct k hash functions $h : \mathbb{R}^m \to \mathbb{Z}$, $h(\boldsymbol{v}) := \left\lfloor \frac{<\boldsymbol{a}, \boldsymbol{v}> + b}{r} \right\rfloor$ using the random projection principle, where $\boldsymbol{a}$ is drawn from m independent distributions which have to be $p - stable$ and b is drawn from $U(0, r)$. For example we construct two hash functions $h(\boldsymbol{v})$, where we set $r = 1$, $\boldsymbol{a} \sim \mathcal{N}(0, I_{2\times 2})^2$ for illustration purposes and for the inner product we use the inner product induced by the euclidean norm. Then $< \boldsymbol{a}, \boldsymbol{v} >$ is the subspace basis coefficient of the orthogonal projection of $\boldsymbol{v}$ on the subspace $\text{lin}\{\boldsymbol{a}\}$ and $\lfloor < \boldsymbol{a}, \boldsymbol{v} > + b \rfloor$ identifies into which bucket the point $\boldsymbol{v}$ falls. Figure 3 depicts the buckets spanned by the two hash functions and the result of querying an indexed hash table using a query point. The lookup/index table is created by hashing 50 points with standard normal distributed components. We define the relation $\boldsymbol{u} \sim \boldsymbol{v} \iff |\{i \in \mathbb{N} : h_i(\boldsymbol{u}) = h_i(\boldsymbol{v}), 1 \leq i \leq k\}| \geq R$, which let's us express the similarity of two points based on the number of colliding hash functions $R \leq k$. Please note that $\boldsymbol{u} \sim \boldsymbol{v}$ is only an equivalence relation if $R = k$ because for $R < k$ the transitivity property does not hold ($\boldsymbol{u} \sim \boldsymbol{v}$ and $\boldsymbol{v} \sim \boldsymbol{y}$ it must also hold that $\boldsymbol{u} \sim \boldsymbol{y}$).

Let $p = 2$, then $||x||_2$ also defines an inner product (this is not the case for, i.e., $p = 1$), we use the fact that the normal distribution is 2-stable and, thus, the distance between $< \boldsymbol{a}, \boldsymbol{v} >$ and $< \boldsymbol{a}, \boldsymbol{q} >$ is distributed according to $\mathcal{N}(0, \frac{||\boldsymbol{v}-\boldsymbol{q}||_2^2}{r^2})$. Since we select k hash functions uniformly, identically and independently (i.i.d.), we observe from Def. III.1 that all k hash functions have to collide which happens with probability greater or equal to p_1^k to fulfill the first property in Def. III.1.

Let P_{D} be the maximum path loss threshold for propagation

paths in dB. We select the radius of the buckets as

$$r := \frac{1}{2} 10^{-\frac{P_\mathrm{D}}{10 p_l}},$$

where p_l denotes the path loss exponent for our propagation paths. This selection ensures that for any two points within $B(\boldsymbol{p}, r)$ the path loss is not lower than $2P_\mathrm{D}$, since for $\boldsymbol{v}, \boldsymbol{u} \in B(\boldsymbol{p}, r) : ||\boldsymbol{u} - \boldsymbol{v}||_p \leq ||\boldsymbol{u} - \boldsymbol{p}||_p + ||\boldsymbol{v} - \boldsymbol{p}||_p < 2r$. Please note, that our goal is not to find the nearest neighbor of certain query points, but to find all *ball centers* for which the given query points are contained within their respective r-balls. Thus, we proceed as described in [8], [9] by selecting L tables of such hash function sets with N hash functions and we define $g_i(\boldsymbol{v}) := (h_{i,1}(\boldsymbol{v}), \ldots, h_{i,N}(\boldsymbol{v}))$, $1 \leq i \leq L$. For a query point $\boldsymbol{q}$ we lookup $g_1(\boldsymbol{q}), \ldots, g_L(\boldsymbol{q})$ and stop searching after we have encountered $\leq 3L$ points. If an encountered point $v_j \in B(\boldsymbol{q}, cr)$ then we return the point. We set $L = \frac{M}{3}$ and $k = 10$ such that with constant probability it holds that if $u \in B(\boldsymbol{q}, r)$ then there $\exists i \in \{1, \ldots, L\} : g_i(\boldsymbol{q}) = g_i(\boldsymbol{u})$ and $\sum_{j=1}^{L} |(P \setminus B(\boldsymbol{q}, cr)) \cap g_j^{-1}(g_j(\boldsymbol{q}))| < 3L$ (guarantees correctness). We outline the initialization phase for $p = 2$ of the scatterer selection in Algorithm 1.

Algorithm 1 Scatterer selection initialization.

Require: $L > 0, P_D > 0, p_l > 0, sourceDim > 0, P$ ▷ P is our set of scatterers

$r = \frac{1}{2} 10^{-P_L / p_l}$
for $i = 1 \to L$ **do** ▷ L hash tables
  for $j = 1 \to N$ **do** ▷ N hash functions
    for $t = 1 \to$ sourceDim **do**
      $a_{i,j,t} \sim \mathcal{N}(0, 1)$
    end for
    $b_{i,j} \sim U(0, 2r)$
  end for
  for $j = 1 \to |P|$ **do**
    $\mathrm{LUT}_{i,j} =$
    $(\lfloor \frac{<\boldsymbol{a}_{i,1}, \boldsymbol{P}_j> + b_{i,1}}{r} \rfloor, \ldots, \lfloor \frac{<\boldsymbol{a}_{i,N}, \boldsymbol{P}_j> + b_{i,N}}{r} \rfloor)^T$
  end for
end for
return LUT, $\boldsymbol{a}$, $\boldsymbol{b}$, r

Let f be the density of our absolute standard normal distributed random variable (folded normal distribution) and with transforming $x = cy$ we have

$$p(c) = \mathbb{P}(h_{\boldsymbol{a},b}(\boldsymbol{v}) = h_{\boldsymbol{a},b}(\boldsymbol{q})) = \int_0^r \frac{1}{c} f\left(\frac{x}{c}\right)\left(1 - \frac{x}{r}\right) dx.$$

By our selection, r is fixed and with $c = \frac{||\boldsymbol{v} - \boldsymbol{q}||_2}{r}$ it is easy to see that our selection fulfills Def. III.1 (i) and (ii) for $p_1 = p(1)$ and $p_2 = p(1 + \varepsilon)$ (with increasing distance of $\boldsymbol{q}$ and $\boldsymbol{v}$ the collision probability monotonically decreases). We further omit the correctness check ($\boldsymbol{v} \in B(q, cr)$) after obtaining a list of possible scattering points because we check whether the selected scattering points are within the LOS of the receiver and transmitter, respectively and whether the inclusion condition (cf. Prob. 1) is fulfilled such that we obtain our set G. The time complexity of a single lookup is given as $O(n^\rho \log_{1/p_2} n)$, where $\rho = \frac{\ln p_1}{\ln p_2}$. We summarize the procedure of obtaining G during simulation in Algorithm 2. The logical(.) operator is defined by logical$(x) = 0$ if $|x| > 0$ otherwise logical$(x) = 1$ and we apply the operator to a vector component-wise.

Algorithm 2 Scatterer selection lookup.

Require: $LUT, Q, M > 0, N > 0, R > 0, L > 0, \boldsymbol{a}, \boldsymbol{b}, P$ ▷ R number of hash functions that have to match

$\forall 1 \leq k \leq |Q| : U_k = \emptyset$
for $i = 1 \to L$ **do**
  for $j = 1 \to |Q|$ **do**
    $\boldsymbol{q}_{i,j} = (\lfloor \frac{<\boldsymbol{a}_{i,1}, \boldsymbol{Q}_j> + b_{i,1}}{r} \rfloor, \ldots, \lfloor \frac{<\boldsymbol{a}_{i,N}, \boldsymbol{Q}_j> + b_{i,N}}{r} \rfloor)^T$
  end for
end for
for $i = 1 \to L$ **do**
  for $j = 1 \to |P|$ **do**
    for $k = 1 \to |Q|$ **do**
      $s = \sum_{m=1}^{|N|} (1 - \mathrm{logical}(\boldsymbol{q}_{i,k} - \boldsymbol{LUT}_{i,j}))_m$
      if $s \geq R$ **then**
        $U_k = U_k \cup \{j\}$
      end if
      if $|U_k| \geq M$ **then**
        break
      end if
    end for
  end for
end for
return $G = \bigcap_{k=1}^{|Q|} U_k$

B. Locality aware selection of scatterers with a constant number of query points

In this section we assume a fixed number of query points (transmitters and receivers) in our scenarios, which translates into $|Q(t)| = k$ being of constant size for any $t \in \mathbb{R}$, we simply write $|Q| = k$. This assumption is valid for most simulation and measurement campaigns.

For the rest of this section let $\frac{1}{p} + \frac{1}{q} = 1$, $1 \leq p, q \leq \infty$. We define the following linear mapping:

Definition III.2. *Let $k \in \mathbb{N}$ define $l : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^{mk}$ by*

$$l(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \sum_{i=1}^{k} \sum_{j=1}^{m} v_{i,j} \boldsymbol{e}_{(i-1)k+j} = vec(\boldsymbol{v}_1 | \ldots | \boldsymbol{v}_k), \tag{2}$$

where $v_{i,j}$ denotes the j-th component of the i-th vector and $\boldsymbol{e}_i \in \mathbb{R}^{mk}$, $(\boldsymbol{e}_i)_j = \delta_{i,j}$, $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. In short terms this defines the vectorisation of a $\mathbb{R}^m \times \mathbb{R}^k$ matrix which is linear and an isomorphism ($\mathbb{R}^m \times \mathbb{R}^k \cong \mathbb{R}^{mk}$).

We further define $\boldsymbol{q}' := l(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k)$, $\boldsymbol{q}_i \in Q$, creating a single vector using all of the available query points. Revisiting Prob. 1 we may rewrite it to

$$G = \{\boldsymbol{p} \in P : \forall \boldsymbol{q} \in Q : ||\boldsymbol{p} - \boldsymbol{q}||_p < r\}. \tag{3}$$

As we have shown before, G can be directly obtained. However, by the cost of doing the lookup described before for all query points and by computing the intersection on the resulting sets. Our goal is now to show that when the amount of query points is constant during a simulation we can reduce the number of queries to a single query and we omit the intersection of sets which yields G. But, we also increase the dimension of the underlying vector space which has an impact on the volume of r-balls in such spaces.

Algorithm 3 Scatterer selection lookup using l or l_p.

Require: $LUT', \boldsymbol{q}', M > 0, R > 0, N > 0, L > 0, \boldsymbol{a}, \boldsymbol{b}, P'$
$G' = \emptyset$
for $i = 1 \rightarrow L$ **do**
  $\boldsymbol{q}_i = (\lfloor \frac{<\boldsymbol{a}_{i,1}, \boldsymbol{q}'> + b_{i,1}}{r} \rfloor, \ldots, \lfloor \frac{<\boldsymbol{a}_{i,N}, \boldsymbol{q}'> + b_{i,N}}{r} \rfloor)^T$
end for
for $i = 1 \rightarrow L$ **do**
  for $j = 1 \rightarrow |P'|$ **do**
    $s = \sum_{m=1}^{|N|} (1 - \text{logical}(\boldsymbol{q}_i - \boldsymbol{LUT}_{i,j}))_m$
    if $s \geq R$ **then**
      $G' = G' \cup \{j\}$
    end if
    if $|G'| \geq M$ **then**
      break
    end if
  end for
end for
return G'

Proposition III.1. *Let* $\boldsymbol{q}' = l(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k)$, $r > 0$, $1 \leq p \leq \infty$ *and* $||l(\boldsymbol{p}, \ldots, \boldsymbol{p}) - \boldsymbol{q}'||_p < r$ *then* $\boldsymbol{p} \in G$.

However, this states that $G' \subset \mathbb{R}^{mk}$, $G' := \{\boldsymbol{p} \in P : ||l(\boldsymbol{p}, \ldots, \boldsymbol{p}) - \boldsymbol{q}'||_p < r\}$ fulfills $G' \subset G$ but $\forall \boldsymbol{p} \in G$ we have $||l(\boldsymbol{p}, \ldots, \boldsymbol{p}) - \boldsymbol{q}'||_p = (\sum_{\boldsymbol{q} \in Q} ||\boldsymbol{p} - \boldsymbol{q}||_p^p)^{\frac{1}{p}} < k^{\frac{1}{p}} r$ and, thus, we do not have equivalence of the result sets for $1 \leq p < \infty$ and $k > 1$. The root cause for this is the increased dimension of the underlying vector space and the corresponding norm. The LSH scheme demands a p-stable distribution which are known to exist *only* for $p \in (0, 2]$ [10]. If we want to be less restrictive and if it suffices that we provably obtain $G \subset G'$ we may use the following isomorphic linear mapping.

Definition III.3. *Let* $k \in \mathbb{N}$, $1 \leq p \leq \infty$ *then define* $l : \mathbb{R}^m \times \mathbb{R}^k \rightarrow \mathbb{R}^{mk}$ *by*

$$l_p(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = k^{-\frac{1}{p}} \sum_{i=1}^{k} \sum_{j=1}^{m} v_{i,j} \boldsymbol{e}_{(i-1)k+j} = \quad (4)$$

$$k^{-\frac{1}{p}} vec(\boldsymbol{v}_1 | \ldots | \boldsymbol{v}_k),$$

which is linear and an isomorphism.

We directly see that for any $\boldsymbol{u} \in P$: $||l_p(\boldsymbol{u}, \ldots, \boldsymbol{u})||_p = \frac{1}{k^{\frac{1}{p}}} (\sum_{i=1}^{k} ||\boldsymbol{u}||_p^p)^{\frac{1}{p}} = ||\boldsymbol{u}||_p$. However, Proposition III.1 is not valid for l_p. But, we obtain the following result:

Proposition III.2. *Let* $r > 0$, $1 \leq p < \infty$ *and* $\boldsymbol{u} \in G$ *then for* $\boldsymbol{q}' \in \mathbb{R}^{mk}$, $||l_p(\boldsymbol{u}, \ldots, \boldsymbol{u}) - \boldsymbol{q}'||_p < r$.

Using the transformed sets we can approximate the original result set G, by reducing the number of queries to a single query and, therefore removing the necessity for intersecting the individual results sets. However, we increase the memory consumption by a factor of k.

The initialization of our lookup table in Algorithm 1 changes only slightly, depending on whether we use l or l_p to modify our point set into $P'(f) := \{\boldsymbol{v} \in \mathbb{R}^{mk} : \boldsymbol{v} = f(\boldsymbol{u}), \boldsymbol{u} \in P\}$. We apply the same transformation to our query set (which is assumed to be constant). These changes are reflected in Algorithm 3.

The proposed algorithm can be implemented in a fast manner and does only involve simple arithmetic operations, considering the case that the used norm is induced by a scalar product. If the assumption of the constant query set is violated the lookup table must be build every time the query set changes and this would nullify the improvements.

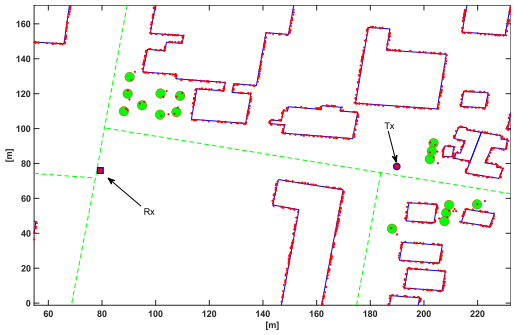
IV. Validation and Performance Evaluation

We validate our GSCM with active scatterer selection applying it to reduce the amount diffuse scatterers used for the calculation of the time- and frequency-selective CIR. We model a vehicular wireless channel measurement campaign conducted in the inner city of Vienna ($48°16'04.9''$N $16°25'56.6''$E). The measurement scenario is presented in Fig. 4a. The reconstructed scenario using our geometry import and the automatic placement of scatterers is depicted in Fig. 4b.

The measurement is conducted using the AIT multi-node channel sounder [11] with a bandwidth of $B = 150$ MHz at a center frequency of 3.5 GHz. The sub-carrier spacing is $\Delta f = 250$ kHz and the snapshot interval is 500 μs. The antennas are mounted on the transmitting node on a Toyota Prius at a height of approximately 1.4 m. For the receiving node we mount the antenna on a Volkswagen T5 at a height of 2 m. The speed of the cars varies between 0 km/h and 40 km/h, where we try to maintain a velocity of 40 km/h. We use $\lambda/2$ dipole antennas for the measurement and simulation. The trajectories for the transmitter and receiver are collected during measurement, having a GPS sampling rate of 1Hz. Thus, we assume a constant velocity between two sampling points. For the simulation scenario we do not model the parked cars along the streets and we do not model other mobile vehicles in the scenario. We conduct for each parameter setting 50 simulation runs.

We estimate the local scattering function (LSF) from the gathered discrete time- and frequency-varying channel transfer function (measurement and simulation) $h[x, w]$ according to [12] using a multi-taper estimate with I orthogonal time-domain and J orthogonal frequency-domain tapers. We collect S samples in time and $Q = 601$ samples in frequency per snapshot, respectively. The stationarity region is indexed in time by $k_\text{t} \in \{1, \ldots, \lfloor S/X \rfloor\}$ and in frequency by $k_\text{f} \in \{1, \ldots, \lfloor Q/N \rfloor\}$ resulting in a region size of $X \times N$. For

(a) Scenario of the measurement campaign (©Google).



(b) Reconstructed scenario from OpenStreetMap.

Fig. 4: Measurement scenario including the trajectories for the transmitter, receiver and the geometry.

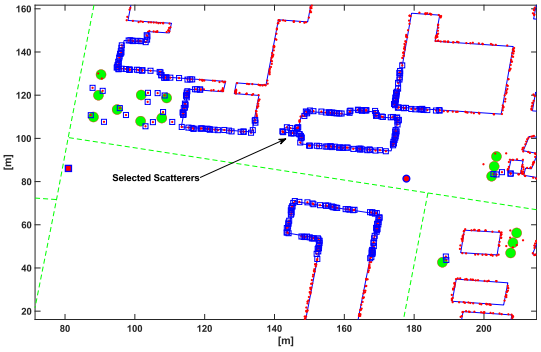


Fig. 5: Example of selected scatterers during simulation for a single time instant. Selected scatterers are highlighted by rectangles.

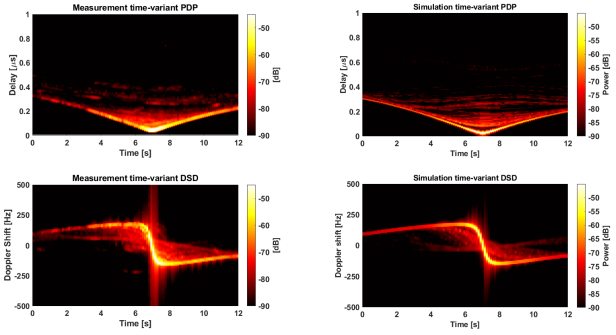


Fig. 6: PDPs and DSDs of the measurement (left) and simulation (right) evaluated by assuming a stationarity duration of $0.12s$.

the measurement and simulation we use a stationarity time $t_s = 0.12$ s [7] and a stationarity bandwidth of $f_{\text{stat}} = 150$ MHz. The time-varying PDP and DSD can be defined as the marginals of the LSF on the delay and Doppler axis, respectively. The PDP and DSD for the measurement and the simulation considering all scatterers in the model is presented in Fig. 6. We further estimate the second central momeents of PDP and DSD by their respective estimators referred to as the RMS delay and Doppler spreads [13].

Fig. 7 depicts the performance of our GSCM without active scatterer selection, taking all 2913 scatterers into account, compared to the conducted measurement by terms of RMS delay spread, RMS Doppler spread and path loss. The auto-

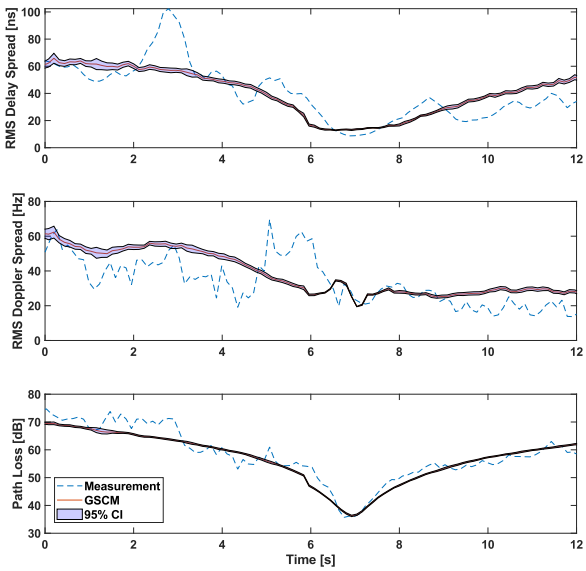


Fig. 7: Comparison of RMS delay spread, RMS Doppler spread and path loss between measurement and simulation.

matically initialized model provides a considerable match with the measurement although missing some dynamic in the RMS delay spread between second 2 and 4 and in RMS Doppler spread between second 5 and 7. In path loss both match quite good. Please not that we compare the average of 50 simulation runs to a single measurement.

We compare the impact of selecting a set of active scatterer by simulating the GSCM using the proposed selection to the case when we consider all scatterers. Therefore, we simulate the GSCM with scatterer selection for a resource constraint $M \in \{300, 500\}$ and a path loss threshold $P_{\text{D}} \in \{35 \text{ dB}, 45 \text{ dB}\}$. Fig. 8 depicts the comparison between the GSCM using all scatterers and the GSCM using only a subset, chosen based on the proposed approach. We see, the higher the number of scatterers selected are and the higher the path loss threshold, the more the RMS delay and Doppler spread approaches the GSCM using all scatterers. However, if we consider a specific resource constraint, we see that the impact on the path loss is negligible when selecting less scatterers.

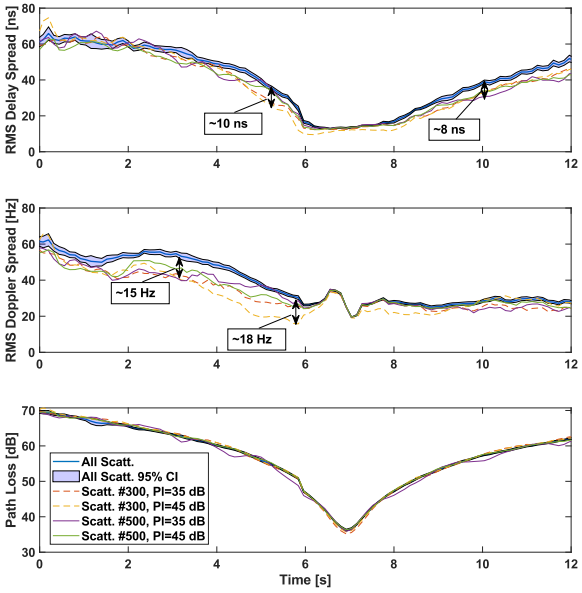


Fig. 8: Comparison of RMS delay spread, RMS Doppler spread and path loss between GSCM simulation taking all scatterer into account and selecting a subset of scatterer based on the proposed approach.

This is expected as the first delay taps with deterministic high power components contribute the most to the path loss.

We also obtain a fitting RMS delay spread for all parameters selections, however, this isn't anymore true for the obtained RMS Doppler spread. This is due to the fact, that the lookup table is not sorted and we stop once we found enough points. Thus, it happens that we stop searching for points but we only queried scatterers whose impinging angle do not differ that much. Thus the RMS Doppler spread shows a bigger gap than other statistics. There are time intervals where the RMS Doppler spread of the reduced scatterer set nearly match the baseline. This is the case if all the scatterers with a major contribution to the baseline are within the respective r-ball. For the very same reasons we also see some slight deviations from second 8 to 12, where a lower path loss threshold achieves a better match as more relevant scatterers are selected. Nevertheless, a Kolmogorow-Smirnov test on equivalence of the distribution of stationarity regions among different delay bins (taking into account those which are above the noise threshold), respectively, revealed no statistical difference at a significance level of 5% among the different simulations.

V. Conclusion

In this paper we presented an active scatterer selection approach taking into account resource constraints enabling real-time simulation or emulation of large-scale GSCMs. The selection algorithm is only parameterized via a path loss threshold and a resource constraint. We compare the accuracy of the proposed selection using a GSCM which is parameterized by a measurement for different resource constraints and path loss thresholds. This approach can also be extended for obtaining higher reflection orders (i.e., n-th order) by creating the set of all combinations of scatterer points by stacking each of these to an nm dimensional vector (leaving out self reflections). The results show that our proposed approach does only have a minor impact on the accuracy of the GSCM and allows for fast implementations of the CIR computation as the number of scatterers is bounded from above.

VI. Acknowledgments

This work is funded by the Austrian Research Promotion Agency (FFG) and the Austrian Ministry for Transport, Innovation and Technology (BMK) within the project RELEVANCE (881701) of the funding program transnational projects.

References

[1] J. Karedal, F. Tufvesson, N. Czink, A. Paier, C. Dumard, T. Zemen, C. F. Mecklenbrauker, and A. F. Molisch, "A geometry-based stochastic mimo model for vehicle-to-vehicle communications," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3646–3657, July 2009.

[2] M. Hofer, Z. Xu, D. Vlastaras, B. Schrenk, D. Loschenbrand, F. Tufvesson, and T. Zemen, "Real-time geometry-based wireless channel emulation," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 12 2018.

[3] N. Czink, F. Kaltenberger, Y. Zhou, L. Bernadó, T. Zemen, and X. Yin, "Low-complexity geometry-based modeling of diffuse scattering," in *Proceedings of the Fourth European Conference on Antennas and Propagation*, April 2010, pp. 1–4.

[4] F. Kaltenberger, G. Steinbock, G. Humer, and T. Zemen, "Low-complexity geometry based mimo channel emulation," in *First European Conference on Antennas and Propagation*, Nov 2006, pp. 1–8.

[5] L. Zhou, F. Luan, S. Zhou, A. F. Molisch, and F. Tufvesson, "Geometry-based stochastic channel model for high-speed railway communications," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4353–4366, May 2019.

[6] B. Rainer, M. Hofer, L. Bernadó, D. Löschenbrand, S. Zelenbaba, A. Dakić, and T. Zemen, "Optimized diffuse scattering selection for large area real-time geometry-based stochastic modeling of vehicular communication links," in *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), to be presented*, 2020.

[7] L. Bernadó, T. Zemen, F. Tufvesson, A. F. Molisch, and C. F. Mecklenbräuker, "The (in-) validity of the wssus assumption in vehicular radio channels," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2012, pp. 1757–1762.

[8] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, ser. SCG '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 253–262. [Online]. Available: https://doi.org/10.1145/997817.997857

[9] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, ser. STOC '98. New York, NY, USA: Association for Computing Machinery, 1998, p. 604–613. [Online]. Available: https://doi.org/10.1145/276698.276876

[10] V. Zolotarev, *One-dimensional Stable Distributions*, ser. Translations of mathematical monographs. American Mathematical Society, 1986. [Online]. Available: https://books.google.at/books?id=qfTDAwAAQBAJ

[11] S. Zelenbaba, D. Loeschenbrand, M. Hofer, A. Dakic, B. Rainer, G. Humer, and T. Zemen, "A scalable mobile multi-node channel sounder," *IEEE Wireless Communications and Networking Conference*, pp. 1–5, April 2020.

[12] L. Bernado, T. Zemen, A. Paier, J. Karedal, and B. H. Fleury, "Parametrization of the local scattering function estimator for vehicular-to-vehicular channels," in *2009 IEEE 70th Vehicular Technology Conference Fall*, Sep. 2009, pp. 1–5.

[13] A. F. Molisch, *Wireless communications*. John Wiley & Sons, 2012, vol. 34.