A Survey of Efficient Ray-Tracing Techniques for Mobile Radio Propagation Analysis

Tetsuro IMAI\(^{ta}\), Member

SUMMARY  With the advances in computer processing that have yielded an enormous increase in performance, numerical analytical approaches based on electromagnetic theory have recently been applied to mobile radio propagation analysis. One such approach is the ray-tracing method based on geometrical optics and the uniform geometrical theory of diffraction. In this paper, ray-tracing techniques that have been proposed in order to improve computational accuracy and speed are surveyed. First, imaging and ray-launching methods are described and their extended methods are surveyed as novel fundamental ray-tracing techniques. Next, various ray-tracing acceleration techniques are surveyed and categorized into three approaches, i.e., deterministic, heuristic, and brute force. Then, hybrid methods are surveyed such as those employing Physical optics, the Effective Roughness model, and the Finite-Difference Time-Domain method that have been proposed in order to improve analysis accuracy.

key words: ray-tracing, acceleration technique, hybrid model, mobile radio propagation

1. Introduction

In mobile communication environments, radio propagation is very complicated and so many studies have been undertaken to clarify and model it based on measurements. However, it currently takes a great deal of time and effort to do so because there are many available frequencies and the cell configuration is complex. On the other hand, with the great advances in computer processing, numerical analytical approaches based on electromagnetic theory have recently been applied to mobile radio propagation analysis. One such numerical analytical approach is the ray-tracing method based on geometrical optics (GO) and the uniform geometrical theory of diffraction (UTD). In the ray-tracing method, as shown in Fig. 1, the required propagation characteristics for system design are simply calculated by tracing rays from the transmitter (Tx) to the receiver (Rx) considering interactions with structures, i.e., reflection, transmission, and diffraction. Fig. 2 shows an example of traced rays.

In this paper, ray-tracing techniques are surveyed that have been proposed in order to improve computational accuracy and speed. This paper is organized as follows. In Sect. 2, the fundamental ray-tracing techniques are described. Section 3 presents ray-tracing acceleration techniques, and these techniques are categorized into three approaches. In the latest trend, Sect. 4 reports on ways of extending the ray-tracing method that are currently being actively investigated to improve analysis accuracy. Section 5 presents our conclusions.

2. Fundamental Ray-Tracing Techniques

There are two main methods for tracing rays. One is the imaging method (or image method), and the other is the ray-launching method (this is sometimes called the Shooting and bouncing ray (SBR) method or Brute-force ray-tracing method) [1]. The features determining ray-trace accuracy and the computational amount are different. In this section, we first briefly describe these methods, and then focus on extension methods to improve the accuracy and decrease the computational burden.

2.1 Imaging Method

In this method, components of structures such as planes are extracted, and then the shortest path ray from the Tx to the

---

\(^{ta}\)The author is with NTT DOCOMO INC., Yokosuka-shi, 239-8536 Japan.

a) E-mail: imait@nttdocomo.co.jp

DOI: 10.1587/transcom.2016EBI0002
Moreover, when the maximum number of interactions, $n$, is set, the total number of combinations is given by $1 + \sum_{i=1}^{N} M(M - 1)^{i-1}$, as shown in Fig. 4.

The computational amount (or speed) of the ray-tracing method is proportional to the number of ray paths that should be searched. Here, these are candidates of ray paths because the paths to be finally canceled are included. Especially, in the imaging method, the number of candidates is equal to the number of combinations of objects (planes or edges). For example, there are two combinations corresponding to the second-order ray path in Fig. 3 as indicated below.

- $(Tx) \rightarrow$ plane $\#1 \rightarrow$ edge $\#1 \rightarrow (Rx)$
- $(Tx) \rightarrow$ edge $\#1 \rightarrow$ plane $\#1 \rightarrow (Rx)$

When we generalize this, the number of combinations corresponding to the $n$th-order ray paths in the environment where $M$ objects exist is given by $M(M - 1)^{n-1}$, as shown in Fig. 4.

Moreover, when the maximum number of interactions, $N$, is set, the total number of combinations is given by [1], [2]. Therefore, the number of Tx images corresponding to the second-order ray path is only one, because positions of multiple Tx images for above-mentioned two combinations of planes are overlapped at $Tx''$ as shown in Fig. 5. In the special case, it is easy to identify the Tx image positions corresponding to $n$th-order ray paths, so ray paths can be searched without considering combinations of planes. This approach has so far been applied to ray tracing in a rectangular room, a straight tunnel with rectangular cross-section, and a street canyon. However, literature [5] suggests that
this approach can be applied to a more complex environment if the above-mentioned conditions are satisfied and its computational amount becomes a polynomial order.

2.2 Ray-Launching Method

In this method, multiple rays are launched in all directions, and then the rays that arrive at the Rx are searched by tracing them. Details regarding the procedures in this method are shown in Fig. 6.

First, multiple rays are launched in all directions with an angular interval of $\Delta \Omega$. Each ray is traced until the termination conditions are satisfied such as the number of interactions and the length of the ray path. If the ray intersects a plane, the ray is divided into a reflected ray and a transmitted ray at the reflection point. Here, the intersection point is recognized as a reflection point, $Q_{R}$ (this is the same as the transmission point). If the ray intersects an edge, the intersection point is recognized as a diffraction point, $Q_{D}$, and multiple diffracted rays are conically retransmitted from $Q_{D}$ in directions with the angular interval of $\Delta \phi_{D}$ (here, $D_{in} = D_{out}$). On the Rx side, a reception sphere is set around the Rx in advance, and the rays incident inside the sphere are recognized as received rays [1], [6], [7]. Note that the procedure for judging whether or not the ray intersects an edge is the same; however, a cylinder should be set around the edge instead of a sphere.

The computational amount (or speed) of the ray-launching method is proportional to the number of traced rays. As mentioned above, the number of traced rays increases after reflection and diffraction. So, when parameters;

- the number of launched rays from Tx, $N_{ray}$,
- the number of retransmitted rays from $Q_{D}$, $N_{Dray}$,
- the maximum number of reflections, $N_{R}$,
- the maximum number of diffractions, $N_{D}$,

are set, the number is given by

$$S_{launch} = N_{ray}^{2N_{R}} (N_{Dray})^{N_{D}}.$$  \hspace{1cm} (2)

Here, it is assumed that many structures exist in analysis area, and the length of the ray path is not considered as the termination condition. $N_{Dray}$ is set much larger than 2 in general. So, the maximum number of diffractions makes a large impact on the computational burden in ray-launching method.

(1) Model for ray-launching from Tx

In the ray-launching method, the rays should be uniformly launched, basically. This is easy for 2D ray launching. However, in 3D ray launching, it is difficult to launch an arbitrary number of rays because there are only five kinds of regular polygons: tetrahedron, cube, octahedron, dodecahedron, and icosahedron. Although we could set aside uniform ray launching, in this paper, the proposed method in [6] and [7], in which rays can be launched approximately uniformly, is described. Hereafter, we call this method the geodesic sphere method.

In the geodesic sphere method, a geodesic sphere is set around the Tx, and the rays are launched from geodesic vertices. Details are given hereafter.

First, the basic ray-launch points are determined to be vertices of an icosahedron as shown in Fig. 7(a). At this point, the number of vertices (rays) is 12 and the angular interval between rays, $\Delta \Omega$ is 63 deg. Next, each side is equally divided by $N$, i.e., each face is divided into small regular triangles. Figure 7(b) shows a case where $N = 4$. Here $N$ is referred to as the tessellation frequency. Vertices of the small triangles are also determined as ray-launch points. As a result, the total number of launched rays is $(10N^2 + 2)$, and the angular interval between rays, $\Delta \Omega$ is approximated by

![Fig. 5 Example of special case, where each object is parallel to x or y axis.](image)

![Fig. 6 Ray-launching method. (a) 3D view, (b) Top view.](image)
Fig. 7  Geodesic sphere method. (a) Icosahedron and basic ray-launch points, (b) Division into small regular triangles \((N = 4)\).

Fig. 8  Size of Reception sphere.

\[ \Delta \Omega \approx \frac{1}{N} \sqrt{\frac{4\pi}{5\sqrt{3}}} = \frac{1.205}{N} \text{ [rad]} = \frac{69.0}{N} \text{ [deg]} \quad (3) \]

(2) Model for receiving rays at Rx

As shown in Fig. 8, when the size of the reception sphere is excessively small, it cannot catch any rays. Conversely, when the size is excessively large, a ray is counted more than once. These are causes of analytical error. Therefore, it is important to determine the optimal size.

The distance between neighboring rays, namely the spatial resolution, \(\Delta l\), is given by the product of the path length, \(d\), and the angular interval, \(\Delta \Omega\), i.e., \(\Delta l = d \Delta \Omega\). Therefore, in general, the size of the reception sphere is controlled adaptively, depending on the spatial resolution. In [1] and [7], it is recommended that the radius of the reception sphere be \(d \Delta \Omega / \sqrt{3}\) in 3D ray launching and \(d \Delta \Omega / 2\) in 2D ray launching.

A disadvantage to the ray-launching method is that the rays that arrive at the Rx cannot be obtained rigorously due to the discrete launching rays. Regardless of how narrow the angular interval of the launched rays is, analytical error that is caused by spatial resolution exists. However, this means that rough analysis results can be obtained in a short time. This is an advantage to the ray-launching method.

2.3 Other Methods

(1) SBR-image method

The SBR-image method was proposed in [8] to improve the computational efficiency of the imaging method using the ray-launching method. The procedure for this method is given hereafter.

1) Perform ray-launching method.
2) Set the candidates for the ray paths based on the results of Step 1.
3) Perform imaging method for the candidates.

In the case of Fig. 9, the ray path \((Tx) \rightarrow \text{plane} \#1 \rightarrow \text{plane} \#2 \rightarrow \text{(Rx)}\) is obtained after performing the ray-launching method, and then rigorous reflection points, \(Q_R^{(1)}\) and \(Q_R^{(2)}\), are determined by performing the imaging method.

In this method, the computational efficiency is improved because the number of combinations of the objects is restricted to the most probable one. If we consider ray launching as the main part, it can be interpreted that the imaging method is used in order to modify rigorously the positions of the interactions as the rays that arrive at the Rx.

(2) Ray-jumping method

As mentioned above, a disadvantage to the ray-launching method is that the rays that arrive at the Rx cannot be obtained rigorously due to the discrete launching rays. Regardless of how narrow the angular interval of the launched rays is, analytical error that is caused by spatial resolution exists. Since the resolution depends on the path length, the error is not uniform in the analysis area. Here, in radio propagation analysis, it is desired that the spatial distribution of the error be uniform, if we accept error. So, the ray-jumping method was proposed [9].

In this method, first the required spatial resolution, \(\Delta l_{req}\), is set. Next, the rays are launched from the Tx similar to that in ray-launching method, and the rays are traced. If the spatial resolution, \(\Delta l\), becomes equal to \(\Delta l_{req}\), the number of rays is increased \(\alpha\) times (here, \(\alpha > 1\)). This procedure is repeated until the termination conditions are satisfied the same as in the ray-launching method. Figure 10 shows the case where \(\alpha = 2\). As a result, the spatial resolution in the analysis (except near the Tx) is expressed by

\[ \Delta l_{req} \leq \Delta l \leq \alpha \Delta l_{req}. \quad (4) \]

So, when the value of \(\alpha\) tends to be one, the spatial resolution in the analysis area becomes uniformly \(\Delta l_{req}\), theoretically.
The computational amount normalized by that of the ray-launching method, \( \rho \), is approximated as

\[
\rho \approx \frac{\alpha}{\alpha + 1}.
\] (5)

Note that the theoretical limitation is \( \rho = 0.5 \), i.e., the computational amount is half that of the ray-launching method.

In [9], the explanation of this method considers only 2D ray tracing. However, extension to 3D ray tracing is easy by using the idea of a geodesic sphere. In the geodesic sphere model mentioned above, the angular interval between rays is given by Eq. (3). So, the spatial resolution is expressed by

\[
\Delta l = d \Delta \Omega = \frac{d}{N} \sqrt{\frac{4\pi}{5\sqrt{3}}}.
\] (6)

Therefore, first we launch rays with \( N = 1 \) from the Tx. If the path length reaches \( \Delta l_{\text{req}} / \Delta \Omega \), we subsequently launch new rays with \( N = 2 \). Thereafter, repeat this process. Note that the tessellation frequency should always be increased by 2 fold.

3. Ray-Tracing Acceleration Techniques

When assuming a realistic environment, any ray-tracing acceleration technique would be helpful in order to perform the analysis within a realistic amount of time. So, many proposals have been reported, and these are categorized into three approaches: deterministic, heuristic, and brute force.

3.1 Deterministic Approach

This is an approach that applies the known acceleration algorithms in order to accelerate ray tracing. By applying this approach, the accuracy of propagation analysis is not degraded. Note that the techniques described in this section have also been used in computer graphics.

1. Visibility graph

In the imaging method, the key point for acceleration is how many combinations can be restricted in advance. On the other hand, in ray-launching method, that is how fast the intersection points of a ray with an object can be searched.
The number of layers is identical to the interaction order. The objects in the 1st layer can be seen from the Tx, the objects in the kth layer can be seen from the related objects in the (k – 1)th layer, and the objects that the ray interacted last can be seen from the Rx. In the imaging method, the combinations shown in Fig. 4 are restricted in this way.

The Binary Space Partitioning (BSP) and Bounding-Volume (BV) algorithms are applied to generate visibility graphs in high speed.

(2) BSP algorithm

The BSP algorithm handles the analytical space (or analysis area) in the front and back of the planes. Here, we assume the environment shown in Fig. 12(a). Here, the arrow direction represents the normal vector of the plane. The BSP algorithm divides the analytical space into subspaces in the front and back of the planes, in sequential order. Figure 12(b) shows a definition of the dividing order, and is referred to as a BSP tree. In this example, the space is divided into two subspaces on the basis of plane #1. Next, the subspace in front of plane #1 is divided into two subspaces on the basis of plane #7, and the subspace in back of plane #1 is divided to two subspaces on the basis of plane #4. This procedure is repeated until all planes are added to the BSP tree. Although there are no rules for the dividing order, it is said that it is better that the sizes of the ultimately obtained subspaces be equal. By referring to the BSP tree, the visible objects can be found quickly. So, the visibility graph can be generated at high speed. More detailed information regarding the way of applying the BPS algorithm is given in [11]–[13].

(3) BV algorithm

The BV algorithm accelerates the searching of visible objects by handling multiple neighboring objects together [10], [14]–[16]. In this algorithm, as shown in Fig. 13(a), the analysis area is divided into multiple blocks (size: $\Delta L \times \Delta L$), where the height of each block, $\Delta H_b$, is defined by the height of the highest structure in the block. Note that the block is referred to as a bounding-volume.

For example, as shown in Fig. 13(b), the visible objects are searched using the procedures below (in an outdoor case).

1) Identify the highest $B_j$ in the visible range of bounding-volumes from the observer, taking ground level into consideration.
2) Select the highest building, $m$, in $B_j$, and draw additional lines of sight in both the horizontal and vertical planes.
3) Delete BVs that are completely hidden from the observer’s view by selected building $m$, from the candidates of subsequent searches by referring to the additional lines.
4) Re-build the list of buildings in other $B_k$ that are partially hidden, and delete the hidden buildings from the list of $B_k$ and modify the height $\Delta H_b$ of $B_k$ if necessary.
5) Save building $m$ in a database as a visible structure, and modify height $\Delta H_b$ of $B_j$ using the remaining struct-

![Fig. 12 BPS algorithm. (a) Analysis model, (b) BPS tree.](image_url)
By repeating Steps 1 to 5 until there are no more building candidates in the visible range, all the buildings visible from the observer's point of view can be eventually obtained. This algorithm allows deleting multiple hidden structures (or buildings) at once, which speeds up the processing. In this paper, we assume that the sizes of all BVs are the same; however, an approach in which the size is defined depending on the spatial density of the structures was proposed in [10]. The effect of the BV algorithm is theoretically and experimentally shown in [15] and [16]. Reference [16] in particular has reported that the running time can be reduced by half by optimizing the size.

(4) Space division approach

The reason why searching for the intersection points of a ray with an object (or searching for an object in visible range) takes a great deal of time is that the searching is performed in round-robin manner. This is done because mutual positional relationships between rays and objects are not clear. One solution is the space division approach. In a 2D approach, the analytical space is divided into grids (or cells), where the information of structures is given in the horizontal plane. In a 3D approach, the analytical space is divided into voxels (which is a coined word from a volume element), where the information of structures is given. In the following, the 2D approach is explained.

Two methods have been proposed using the 2D approach. One is the rectangular grid method. In [17], it was reported that the computational time with the rectangular grid approach is on average 86% less than that using the above-mentioned visibility graph. The other method is the triangular grid method, which is more effective than the rectangular grid [18], [19].

In the triangular grid method, the analytical space is divided into triangular grids as shown in Fig. 14(a). In the figure, the solid lines are identical to the edges of the structures, and the dashed lines are dummy lines for analysis. Here, the total number of the triangular grids, \( N_{\text{triangle}} \), and the number of edges of the triangular grids, \( N_{\text{edge}} \), are given by

\[
N_{\text{triangle}} = 2 (N_v - 1) - N_b
\]

\[
N_{\text{edge}} = 3 (N_v - 1) - N_b
\]

respectively, where \( N_b \) is the number of vertexes of a polygon surrounding the analysis space (the rectangle in Fig. 14(a)) and \( N_v \) is the total number of vertexes of the triangular grids (here, \( N_b \) is included) [18]. In Fig. 14(a), \( N_b \) and \( N_v \) are 4 and 26, respectively. So, \( N_{\text{triangle}} \) and \( N_{\text{edge}} \) are 46 and 71, respectively. In this way, the number of triangular grids (and edges) that is needed to divide the space is uniquely determined, and this is an advantage to the triangular grid method. In addition to this, a fast algorithm for ray tracing exists. The procedures are given hereafter (see Fig. 14(b)).

1) Find grid \( T_0 \) in which the Tx exists.

2) Calculate vectors \( \vec{u}_i = \vec{r} \times \overrightarrow{TxV_i} \) \((i = 0, 1, 2)\), where \( \vec{r} \) is a direction vector of the ray and \( \overrightarrow{TxV_i} \) is a vector from the Tx to vertex \( V_i \).

Here, we focus on \((\vec{u}_i)_z\) (which is the z component of \(\vec{u}_i\)). If \((\vec{u}_i)_z > 0\), the ray will pass through the right side of vertex \( V_i \). If \((\vec{u}_i)_z < 0\), the ray will pass through the left side of vertex \( V_i \). If \((\vec{u}_i)_z = 0\), vertex \( V_i \) will be hit. So, in Fig. 14(b), the following must be performed.

3) Confirm that the ray hits edge \( e_1 \), and record it with the information of the next grid, \( T_1 \).

4) If \( e_1 \) is not an edge of a structure, calculate vector \( \vec{u}_3 = \vec{r} \times \overrightarrow{TxV_3} \) where \( V_3 \) is a vertex opposite to edge \( e_1 \) in \( T_1 \).

5) Confirm that the ray hits edge \( e_3 \), because of \((\vec{u}_3)_z < 0\), and record it with the information of the next grid, \( T_2 \).

6) If \( e_3 \) is not an edge of a structure, calculate vector \( \vec{u}_4 = \vec{r} \times \overrightarrow{TxV_4} \), where \( V_4 \) is a vertex opposite to edge \( e_3 \) in \( T_2 \).

7) Confirm that the ray hits edge \( e_6 \), because of \((\vec{u}_4)_z < 0\), and record it with the information of the next grid, \( T_3 \).

8) If edge \( e_6 \) is an edge of a structure, calculate the intersection point of \( e_6 \) with a ray, and record it as a reflection point.

In this algorithm, after Step 4, the number of calculations is only one and that calculation is an outer product. So, the ray can be traced quickly. Note that this algorithm can be extended to a 3D case. In [19], this triangular grid method...
is applied with the Vertical Plane Launch (VPL) method, which is described below.

3.2 Heuristic Approach

This is an approach that changes the computational logic based on the heuristic model in order to accelerate the ray tracing. In this approach, the range of ray tracing is basically restricted such that only the rays that contribute to radio propagation remain. In general, a significant increase in speed can be expected, but the accuracy of the propagation analysis is guaranteed only through comparison to measurement results.

Methods (1) through (3) below were proposed to accelerate 3D ray tracing in urban areas, and these cannot be used in combination. Method (4) is based on the metaheuristic approach. This can be applied to an arbitrary propagation environment and can be used in combination with other methods if the basic ray-tracing algorithm is the same.

1) VPL method

The VPL method in [20] basically performs the ray-launching method in the horizontal plane, as shown in Fig. 15. The procedures are given hereafter.

1) Launch the 2D rays from the Tx in the horizontal plane.
2) Trace the rays to the Rx considering the interactions of buildings.
3) Extend the rays to 3D rays based on the Tx and Rx antenna heights, where the transmissions are changed to diffractions on rooftops of buildings.

The order of the computational amount for this method is almost the same as that for the 2D ray-launching method. Therefore, the computational amount can be evaluated based on Eq. (2). The number of launched rays in the 2D ray-launching, \( N_{2D-ray} \), is much less than that in the 3D ray-launching, when \( \Delta V \) is same. Moreover, in the case of the 2D ray-launching, the number of processing of diffracted ray retransmission from an edge is only once, even if plural rays are incident on the edge. Taking these into consideration, the total number of traced rays in the 2D ray-launching, \( S_{2D-launch} \), is expressed as

\[
S_{2D-launch} \leq N_{2D-ray}^{2N} \left( N_{Dray} \right)^{N_D} < S_{3D-launch}. \tag{9}
\]

The accuracy of this method is reported in detail in [21]. Note that this idea can be applied to indoor propagation analysis. The method for this was proposed in [15], and is referred to as the Hybrid Ray-trace (HY-RAYT) method.

2) Sighted Objects-based Ray-Tracing (SORT) method

The SORT method in [14] basically performs the imaging method for buildings that can be seen from the Tx or Rx. The procedures are given hereafter.

1) Search the buildings that can be seen from the Tx or Rx.
2) Trace the rays using the imaging method for the buildings, as shown in Fig. 16(a).
3) Extract the buildings that rays pass through.
4) Re-trace the rays such that the transmissions are changed to diffractions on rooftops of buildings, as shown in Fig. 16(b).

In this method, the target buildings for the imaging method are restricted to the buildings that can be seen from the Tx or Rx. In addition to that, the direction of the propagation is restricted to the forward direction from the Tx to the Rx such as (Tx → buildings in visible range from Tx → buildings in visible range from Rx → Rx). In other words, the partially back propagating rays such as (Tx → the buildings in visible range from Rx → the buildings in visible range from Tx → Rx) are not traced in this method. Note that the electric field
of the ray is weak in general because the path length of the ray is longer than that for the forward propagating ray. Therefore, it is said that this method can reduce the computational amount while maintaining analytical accuracy.

(3) Ray-tracing in vertical and transversal planes

The method in [22] performs ray tracing in the vertical and transversal planes, in which both the Tx and Rx exist as shown in Fig. 17. Hereafter, this method is referred to as VP&TP method in this paper. Although the basic ray-tracing technique is not described in [22], it is considered that the imaging method is applied. Since the combinations of objects are restricted, the computational amount can be reduced. Note that it is possible to apply the ray-launching method in both the VP and TP planes. In [22], the scattering waves are also considered in addition to the rays. The method for this is described in Sect. 4.1(2).

(4) Ray-tracing with genetic algorithm

In the imaging method, the key point for acceleration is how many combinations can be restricted in advance. In order to implement this, the method in [23] and [24] applies the Genetic Algorithm (GA), which represents a meta-heuristic approach. This model is referred to as the GA_RT method.

In the GA, a chromosome is defined as a combination of multiple genes, the corresponding characters of which are different from each other, and an individual is expressed by a chromosome. On the other hand, in the imaging method, as shown in Fig. 4, ray paths are expressed as combinations of objects. This suggests that the GA has a high affinity to the imaging method when each ray path is handled as an “individual.” Figure 18 shows a population model for the GA_RT method. An object, a combination of objects, and a ray path are considered as a gene, chromosome, and individual, respectively. In addition, the “fitness” of each individual is defined as the received power of a traced ray along the corresponding path. The processing procedure in the GA_RT method is described below.

1) Establish the initial path group by randomly extracting $N_c$ paths from the theoretically possible ray paths. Then, go to Step 3.
2) Re-establish the path group by performing “selection,” “crossover,” and “mutation.”
3) Trace the ray and calculate the received power as the fitness for each ray path.
4) Evaluate the fitness for each ray-path.
5) Repeat Steps 2 through 4 until the finish condition is satisfied.

In [25], the effect of this method is reported in an urban environment, where the above-mentioned SORT method is applied together. The computational time is reduced to 20% by using the GA_RT method.

3.3 Brute Force Approach

The processing of ray tracing has good parallelism. In the imaging method, parallel processing can be performed on the “Tx position level,” “Rx position level,” and “combination of objects level.” On the other hand, in the ray-launching method, parallel processing can be performed on the “Tx position level,” “diffraction point level,” and “launching ray level.”

The brute force approach changes the computation architecture based on parallel processing in order to accelerate the ray tracing. This approach does not degrade the accuracy of the propagation analysis. In addition, the total calculation speed is proportional to the number of processing units, basically. However, some idea is needed to address the bottlenecks in practice.

(1) Parallel processing with multiple computers

The system configuration used in [14] and [26] is shown in Fig. 19(a). The system consists of one application server and multiple calculation servers (or distributed servers), and they are connected by a LAN. In the application server, calculation conditions are set and the jobs are distributed to the calculation servers, and then the results from the calculation servers are merged and displayed. This system performs ray
tracing using the SORT method and parallel processing on the Rx position level.

Figure 19(b) shows the evaluation results of the distribution effect. The number of calculation servers is represented on the horizontal axis and the calculation speed is represented on the vertical axis. The calculation speed is normalized by the calculation time for two servers. Term \( N_p \) represents the number of calculation points (or Rx points). When there are fewer calculation points, the distribution effect is small although the number of servers is increased. When there are many calculation points, the calculation speed is enhanced in proportion to the number of servers. This means that the data transmission rate becomes a bottleneck for distribution. So, this should be considered when distributing the servers. Note that [14] and [26] propose an effective distribution method.

(2) Parallel processing with GPUs

Nowadays, General-purpose computing on graphics processing units (GPUs or GPGPUs) has attracted attention in terms of parallel processing. Currently, PCs have multiple CPUs with multi-cores. On the other hand, the GPU board has thousands of GPU cores. So, the number of GPU cores is on a different scale compared to the number of CPU cores, although the clock frequency of GPUs is lower than that of the CPUs. Reference [27] reported on ray tracing with GPUs (especially method of effective memory reference for parallel processing) and its effect.

In [27], the VPL method is applied to ray-tracing and parallel processing on a 2D ray level is performed with GPUs (NVIDIA Tesla K20c, 0.7 GHz×2496 cores). In the report, it was clarified that the calculation speed of the GPU is 9.82 times higher than that of the CPU (Intel Xeon E5-2687W, 3.1 GHz×8 cores) under the same calculation conditions for ray-tracing. Note that the calculation speed of the CPU with 8 threads is 4.65 times higher than that of the CPU with 1 thread. They concluded that implementation of parallel processing with multiple GPUs is one topic for future work.

4. Extending Ray-Tracing Method

The ray-tracing method has applicable limitations caused by theoretical foundations (GO and UTD). So, to overcome these limitations there have been some interesting proposals.

4.1 Hybrid with Physical Optics

In the ray-tracing method based on GO and UTD, scattering waves cannot be approximated as geometrical rays when the size of the aperture or plane becomes small with respect to the wavelength. Also diffuse-scattering components become large when the roughness of the surface cannot be electrically approximated as a smooth surface. In these cases, the accuracy of the propagation analysis with ray-tracing becomes low. On the other hand, it has been confirmed in detailed measurements in an actual environment that the waves that cannot be recognized as the specular component are observed [28]. Therefore, hybrid methods of Physical Optics (PO) and ray tracing have been proposed.

(1) Application to O-to-I propagation analysis

When waves propagate from outdoors to indoors, the paths passing through windows and doors are dominant. Therefore, it is important to consider apertures such as windows and doors in the analysis of Outdoor-to-indoor (or O-to-I) propagation. PO has been applied to scattering analysis from an aperture [29]. However, when mobile radio propagation is assumed, it is essential to consider multi-paths both outdoors and indoors. Taking this into consideration, a hybrid method of PO and ray tracing was proposed in [30]–[33]. This method is referred to as the RT_PO method in [31]–[33].

In the RT_PO method, as shown in Fig. 20, ray-tracing is performed outdoors (rays are traced from the Tx to the center of the aperture) and indoors (rays are traced from the center of the aperture to the Rx). Here, consequently, the ray paths are expressed as combinations of outdoor paths and
indoor paths. This means that the number of paths is \( M \times N \) (\( M \) and \( N \) are the number of outdoor paths and indoor paths, respectively). The total electric field at the Rx is calculated by summing the electric fields of rays corresponding to the paths considering the scattering effects of the aperture with PO. In this way, by applying PO to scattering at the aperture, the RT-PO method maintains high analytical accuracy. Note that the details on the calculation are described in [31], and a comparison to measurement results is given in [33].

(2) Application to scattering analysis from rough surface

In an outdoor propagation environment, the dominant scatterers are buildings. The shape and size of building are considered in the ray-tracing, but it is assumed that the size of an object is infinite and the surface is smooth when calculating reflection and diffraction coefficients. In actual buildings, the size of its face is limited. In addition, its surface is not smooth because there is unevenness to, for example, windows and veranda, and electric properties (permittivity and conductivity) of components are different from each other. Therefore, in order to consider the scattering from an uneven surface, a PO hybrid was proposed. The waves are scattered from the surface in all directions. So, it takes a great deal of computational time if the paths of all the scattering waves are traced. In order to mitigate the increase in the computational time, two methods were proposed.

One is the method proposed in [34]. In this method, it is assumed that the specular component of scattering waves, i.e., the wave in the specular direction, is dominant. The concrete procedures are given hereafter. First, the ray tracing is performed assuming smooth surfaces, and then the scattering coefficient, which is obtained with PO considering surface unevenness, is applied as an “effective reflection coefficient” instead of the conventional reflection coefficient in the electric field calculation step. As a result, the computational time is almost the same as that for conventional ray-tracing. However, it is expected that the analysis accuracy would become degraded in an environment such that the effect of defuse scattering components is large. Note that [34] reports that the analysis error (RMS error) for the received power becomes approximately 1 dB less than that in conventional ray-tracing as indicated by the results of comparison with measurements.

The second method was proposed in [22], which is mentioned in Sect. 3.2(3). In this method, conventional ray-tracing using the VP&TP method is performed and the electric fields of rays are calculated. Next, the paths of the scattering waves are traced with the number of scattering of one, and the electric fields of the waves are calculated with PO. Finally, the total electric field is obtained by summing the electric fields of the rays and the electric fields of the scattering waves. In this method, the non-specular components of the scattering waves can be considered, but the specular components of the scattering waves are counted as reflected rays on a smooth surface.

4.2 Hybrid with Effective Roughness Model

The above-mentioned scattering analysis by PO especially focuses on surface unevenness from relatively large objects such as windows and veranda. However, the actual surfaces of buildings walls are more complex due to small-scale roughness of the surfaces, decorative masonry, pipes, cables, and internal irregularities. Also detection of the sizes and electric properties of these small objects are very difficult. This means applying the PO while considering the contribution of these small objects, which is not realistic. From these viewpoints, the Effective Roughness (ER) model was proposed in [35] and [36].

In the ER model, scattering on a random rough surface is assumed. The parameter values for this model are obtained statistically and experimentally. This means that volume scattering, e.g., scattering from trees, can be considered if the parameter values are obtained. Figure 21 shows the hybrid ER model and ray-tracing method. First, conventional ray tracing is performed and the electric fields of rays are calculated. Next, the paths of the scattering waves are traced with the scattering of one, and the electric fields of the waves are calculated with the ER model. Finally, the total electric field is obtained by summing the electric fields of the rays and the electric fields of the scattering waves. Note that the details of the calculation and a comparison to measurement results are given in [37] and [38].

Investigation of the ER model is now in progress. An improved model has already been proposed in [39]. In addition, in order to perform scattering analysis with high accuracy, the method in which the ER model is applied to point cloud data of structures that are obtained using a 3D laser scanner, has been actively investigated [40], [41].
The Finite-difference time-domain (FDTD) method is a well-known approach for numerically solving Maxwell’s equations [42]. Reference [43] describes a comparison between the FDTD and ray-tracing methods based on the results of [44] as given hereafter.

- Complexity: For the FDTD method, complexity depends mainly on the size of the scenario, whereas for the ray-tracing method it depends mainly on the number of walls.
- Accuracy: The FDTD method is in general more accurate because the number of reflections is not limited unlike in the ray-tracing method.
- 3D extension: The ray-tracing method is in general less computationally demanding than the FDTD method, that is why a 3D version of the model is easier to implement.

Taking these into consideration, [43] proposed a hybrid method of FDTD and ray tracing for O-to-I propagation analysis. In this method, as shown in Fig. 22, 3D ray tracing and 2D-FDTD are applied for outdoor propagation analysis and indoor propagation analysis, respectively. The basic procedures in this method are given hereafter.

1) On the outdoor side, set virtual receiving points on the outside wall of a target building, and then perform 3D ray tracing from the Tx and virtual receiving points.
2) On the indoor side, recognize the virtual receiving points as virtual source points, and perform 2D-FDTD.

Reference [43] reports that the analysis error (RMS error) for the received power in this method is approximately 2.4 dB when compared to the measurement results.

5. Conclusion

Recently, the ray-tracing method has become an important tool for system design and service area design in mobile communications. In this paper, we surveyed ray-tracing techniques from the viewpoints of improving the computation accuracy and speed. The main points are given hereafter.

- Fundamental techniques: the well-known imaging method and ray-launching method were explained, and the methods that extend them were surveyed.
- Ray-tracing acceleration techniques: various methods that have been proposed so far were surveyed categorizing them into three approaches, i.e. deterministic, heuristic, and brute force.
- Extension of the ray-tracing method: Hybrid methods employing, for example, Physical optics, the ER model, and FDTD, that have been proposed in order to improve analysis accuracy were surveyed.

Here, the hybrid methods have been actively investigated, especially as analysis methods for propagation with high frequencies, e.g., mm-waves. We believe that more validation and/or improvement are needed for practical use.

Finally, we excluded the survey of fundamental theory (GO and UTD) in this paper. References [45]–[49] give more information on GO and UTD, and [50]–[54] give more information on slope diffraction and multi-diffraction.

References


Tetsuro Imai was born in Tochigi, Japan, in 1967. He received his B.S. and Ph.D. degrees from Tohoku University, Japan, in 1991 and 2002, respectively. He joined the Radio Communications Systems Laboratories of Nippon Telephone and Telegraph Corporation (NTT), Kanagawa, Japan, in 1991. Since then, he has been engaged in the research and development of radio propagation, antenna systems and system design for mobile communications. He is now Senior Research Engineer of the 5G Laboratory, NTT DOCOMO, INC., Kanagawa, Japan. He received the Young Researcher’s Award in 1998, the Best Paper Awards in 2006 & 2014 and the Communications Society Best Paper Award in 2015 from the IEICE. Dr. Imai is a Member of the IEEE.