



Degree Project in the Field of Technology Engineering Physics and the Main Field of  
Study Applied and Computational Mathematics

Second cycle, 30 credits

# **Optimal Transceiver Placement in Retail Environments**

A Path Tracing Approach

**VICTOR HERNADI  
LEANDRO CAROCCA**



# Optimal Transceiver Placement in Retail Environments

## A Path Tracing Approach

VICTOR HERNADI

LEANDRO CAROCCA

Date: June 28, 2024

Supervisors: Per Enqvist, Serge de Gosson de Varennes

Examiner: Per Enqvist

School of Engineering Sciences

Host company: Pricer AB

Swedish title: Optimal placering av sändtagare i detaljhandelsmiljöer

Swedish subtitle: En strålföljningsansats



## Abstract

This thesis explores the optimization of transceiver placement in retail environments to ensure efficient communication with electronic shelf labels. The study is driven by the challenges of placing transceivers in complex store layouts with varying product densities and configuration. To address these challenges, the thesis introduces a two-component approach comprising simulation and optimization. The simulation leverages path tracing methodology from computer graphics to model infrared light propagation from transceivers within a three-dimensional store representation, accounting for obstacles, reflection and the emissive properties of transceivers. The optimization component aims to minimize the number of transceivers needed for reliable communication.

Our findings demonstrate that this approach is not only feasible but also highly flexible, as it easily adapts to a wide range of geometrical settings. We identified the need for rigorous convergence criteria and improvements in light sampling strategies to enhance accuracy. Sensitivity analysis highlights the critical role of high-quality simulations in reducing computational efforts and improving optimization outcomes. We investigate both cooperative (where transceivers work together) and non-cooperative (independent transceivers) communication models, ultimately recommending the cooperative model despite its slightly higher computational cost.

Future research should focus on modeling and representing the costs associated with communication, such as energy consumption, and further evaluating communication models in more complex store geometries, especially dynamic ones. This flexible approach can serve as a robust method for similar optimization problems, particularly where ease of modeling, flexibility, rapid iteration, and computational efficiency are important.

## Keywords

Retail technology, Wireless communication, Transceiver placement, Coverage problem



## Sammanfattning

Detta examensarbete behandlar optimering av placeringen av sändtagare i butiksmiljöer för att säkerställa tillförlitlig kommunikation med elektroniska hylletiketter. Studien motiveras av de utmaningar som finns vid placering av sändtagare i komplexa butiksmiljöer med varierande produktensitet och utformning. För att möta dessa utmaningar introducerar vi en tvåkomponentsansats bestående av simulering och optimering. Simuleringen använder sig av strålföljning, en teknik från världen av datorgrafik, för att modellera infrarött ljus från sändtagare inom en tredimensionell butiksrepresentation, med hänsyn till hinder, reflektion och sändtagarnas utsändningsegenskaper. Optimeringskomponenten syftar till att minimera antalet sändtagare som behövs för tillförlitlig kommunikation.

Våra resultat visar att denna metod inte bara är genomförbar utan också mycket flexibel, eftersom det går lätt anpassa sig till en mängd olika geometriska villkor. Vi identifierade behovet av noggranna konvergenskriterier och bättre strategier för stokastiskt urval av ljus. Känslighetsanalys av optimeringen visar att högkvalitativa simuleringar är av stor betydelse för att kunna minska beräkningskostnaden och förbättra optimeringskvaliteten. Vi undersöker både samverkande (där sändtagare samarbetar med varandra) och icke-samverkande (självständiga sändtagare) kommunikationsmodeller och rekommenderar slutligen den samverkande modellen trots dess något högre beräkningskostnad.

Framtida forskning bör fokusera på att modellera och representera de kostnader som är associerade med kommunikation, såsom energiförbrukning, och ytterligare utvärdera kommunikationsmodeller i en större mängd olika butikslayouter, i synnerhet dynamiska miljöer. Denna flexibla metod kan fungera som en robust metod för liknande optimeringsproblem, särskilt där det är viktigt med enkel modellering, flexibilitet, snabb iteration och beräkningsmässig effektivitet.

### Nyckelord

Detaljhandelsteknik, Trådlös kommunikation, Placering av sändtagare, Täckningsproblem



## Acknowledgments

We would like to express our sincere gratitude to Prof. Per Enqvist, our supervisor, for his guidance, encouragement, and support throughout the course of this research. Additionally, we are deeply thankful to Pricer AB, especially to Dr. Serge de Gosson de Varennes for his willingness to provide us with this project. His unwavering support and knowledge have significantly contributed to the success of this project.

Stockholm, June 2024

Victor Hernadi

Leandro Carocca



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem . . . . .	2
1.2	Purpose . . . . .	2
1.3	Methodology and Delimitations . . . . .	3
1.4	Structure of the thesis . . . . .	4
<b>2</b>	<b>Light Simulation</b>	<b>5</b>
2.1	Requirements . . . . .	5
2.2	Strategy . . . . .	6
2.3	Radiometry . . . . .	7
	2.3.1 The Fundamental Quantities of Radiometry . . . . .	8
2.4	The Light Transport Equation . . . . .	11
2.5	Measuring Radiance . . . . .	14
2.6	Path Integral Formulation . . . . .	15
2.7	Path Sampling . . . . .	19
	2.7.1 Random Walk . . . . .	20
	2.7.2 Next Event Estimation . . . . .	22
	2.7.3 Multiple Importance Sampling . . . . .	23
2.8	BRDF . . . . .	24
	2.8.1 Properties . . . . .	25
	2.8.2 Ideal Diffuse Surfaces . . . . .	25
	2.8.3 Microfacet Theory . . . . .	26
	2.8.4 Sampling . . . . .	30
2.9	Transceiver Model . . . . .	32
2.10	Test Retail Environment . . . . .	33
2.11	Results and Discussion . . . . .	35
	2.11.1 Measured Radiance . . . . .	35
	2.11.2 Number of Samples . . . . .	38

<b>3</b>	<b>Optimization</b>	<b>41</b>
3.1	Requirements . . . . .	41
3.2	Communication Models . . . . .	42
3.2.1	Non-Cooperative Communication . . . . .	43
3.2.2	Cooperative Communication . . . . .	44
3.3	Solvers and Heuristics . . . . .	47
3.4	Test Strategy . . . . .	48
3.5	Results and Discussion . . . . .	48
3.5.1	Non-Cooperative Communication . . . . .	48
3.5.2	Cooperative Communication . . . . .	51
3.5.3	Comparing Non-Cooperative and Cooperative Com- munication . . . . .	53
3.5.4	Sensitivity . . . . .	54
3.5.5	Symmetry . . . . .	57
<b>4</b>	<b>Conclusions and Future Work</b>	<b>59</b>
4.1	Conclusions . . . . .	59
4.2	Future Work . . . . .	60
	<b>References</b>	<b>61</b>

# List of Figures

2.1	Radiant flux and irradiance. . . . .	9
2.2	Radiance. . . . .	10
2.3	Invariance property of radiance . . . . .	10
2.4	Illustration of the reflectance equation. . . . .	12
2.5	Three-point-form of the LTE. . . . .	16
2.6	Factors contributing to a measurement due to a path of length three. . . . .	17
2.7	3D view of the test environment. . . . .	34
2.8	Top-down view of the environment where dots (●) are ESLs and crosses (×) are transceivers. . . . .	34
2.9	Measurements due to transceiver at bottom left. . . . .	36
2.10	Measurements due to transceiver at bottom right. . . . .	37
2.11	Measurements due to transceiver at the center. . . . .	37
2.12	Measurements due to transceiver at top-middle. . . . .	38
2.13	Radiance measurements depending on the number of samples. . . . .	39
3.1	Demonstration of excessive communication. . . . .	42
3.2	Optimal transceiver placement using 100 000 samples. . . . .	50
3.3	Clustering of transceivers when using the greedy algorithm. . . . .	51
3.4	Optimal transceiver placement using 100 000 samples. . . . .	53
3.5	Excess signal strength above activation threshold. . . . .	54
3.6	Sparsity of the measurement matrix with a threshold cut-off applied. . . . .	55
3.7	Histogram of signal strength samples. . . . .	57



# List of Tables

3.1	Optimal transceiver count using non-cooperative communication. . . . .	49
3.2	Transceiver count using non-cooperative communication obtained by the greedy algorithm. . . . .	51
3.3	Optimal transceiver count using cooperative communication. . . . .	52



# Chapter 1

## Introduction

The retail industry has seen significant technological advancements in recent years, with electronic shelf labels (ESLs) becoming increasingly common. ESLs allow for real-time updates of product prices and information, enhancing the flexibility and efficiency of retail operations. Unlike traditional paper labels, ESLs can be updated wirelessly, ensuring that prices are always accurate and consistent across all points of sale.

Transceivers are crucial components in this communication system. They serve as intermediaries between the central control system and the ESLs, transmitting data to and receiving data from the ESLs to ensure the displayed information is always current. The placement of these transceivers within the retail environment is critical to ensure comprehensive communication coverage.

ESLs can utilize various communication technologies. This thesis focuses on those that communicate using infrared (IR) light, a technology usually chosen because of its low energy consumption and cost-effectiveness. However, IR communication is challenging in several aspects. IR signals do not penetrate walls, so obstacles like shelves, products, and even people can significantly influence communication. The system's communication capabilities can vary throughout the day due to these dynamic obstacles, further complicating optimal transceiver placement.

The complexity of retail store layouts, with varied shelf configurations, aisles, and product densities, poses significant challenges for optimal transceiver placement. These complexities can lead to communication blind spots, where some ESLs do not receive signals effectively, compromising the system's reliability. Addressing these challenges is crucial for maintaining the operational efficiency and reliability of ESL systems.

Currently, Pricer AB, a leading provider of ESL systems, relies on experiential knowledge and in-store testing to determine the placement of transceivers. While somewhat effective, this approach is time-consuming and often results in installing more transceivers than necessary. This not only increases the cost but also the system's complexity. There is a clear need for a more systematic and efficient approach to optimize transceiver deployment.

This thesis aims to develop a robust framework for optimizing the placement of transceivers in retail environments. The framework aims to minimize the number of transceivers required while ensuring comprehensive and reliable communication with all ESLs. By addressing this optimization problem, the thesis seeks to reduce costs, improve system efficiency, and enhance the overall reliability of ESL systems in retail settings.

### **1.1 Problem**

The central problem addressed in this thesis is the optimization of transceiver deployment in retail environments. The goal is to develop a systematic framework that minimizes the number of transceivers while ensuring robust communication with all ESLs. This involves overcoming challenges posed by complex store geometries and varying product densities, which can create communication blind spots. Additionally, this thesis investigates different communication models to understand which approach is best in terms of complexity, accuracy, and computational effort. We explore simpler models where transceivers cannot cooperatively communicate with ESLs (signals do not cumulatively add) as well as models where they can.

### **1.2 Purpose**

The purpose of this thesis is twofold: first, to provide Pricer AB with an optimized deployment strategy that reduces the number of transceivers needed, thereby lowering costs and improving efficiency; second, to contribute to the engineering and academic community by developing and validating a novel approach for optimizing infrared communication networks in complex environments. Additionally, this degree project serves as the culmination of our engineering education, showcasing our ability as new engineers to tackle and solve real-world practical problems. This project stands as a testament to the skills and knowledge we have acquired, demonstrating our readiness to contribute to the field of engineering.

## 1.3 Methodology and Delimitations

To achieve these goals, the research methodology is divided into two main components: simulation and optimization.

**Simulation Component** The objective is to comprehend how a transceiver, placed at a specified location within this virtual space, emits infrared light and how this emission propagates throughout the geometric setting. This is essential to assess the communication capabilities between transceivers and electronic shelf labels. This component uses a path tracing methodology inspired by computer graphics. The benefits of this approach are its simplicity in handling arbitrary and complex geometries which makes it suitable since no store are alike, with different layouts and product densities. Additionally, it is possible to leverage existing tools for geometry production and the capabilities of modern hardware, specifically tailored to work on this type data, allows simulations to be executed efficiently on standard workstations.

**Optimization Component** The second component involves the formulation of an optimization problem aimed at minimizing the number of transceivers required for efficient communication with the electronic shelf labels. The optimization formulation is designed to be agnostic to the intricacies of the simulation process, allowing for the independent development of both components. By separating the problem into these two components – simulation and optimization – not only facilitates their individual progress but also grants the freedom to enhance each component's sophistication and complexity without being constrained by inter-dependencies.

**Delimitations** A key delimitation of this thesis is that while the framework provides a robust methodology for optimizing transceiver placement, it does not develop a cost model that accounts for factors such as energy consumption. These additional costs can be easily incorporated into the proposed framework in future work. Additionally, while field testing is acknowledged as crucial for validating the practical effectiveness of the proposed solutions, it is beyond the scope of this thesis. Furthermore, proprietary details about Pricer's hardware are not disclosed; the proposed framework is general enough to account for any type of emissive transceiver. However, a simple digital retail test environment is devised where transceivers are loosely inspired by their real-life counterparts. Another delimitation is that this study only considers static

geometry, not accounting for dynamic changes such as people moving or products being restocked, which can affect IR communication.

## 1.4 Structure of the thesis

Special care and consideration have gone into making the main chapters of this thesis as standalone as possible, aligning with the approach of keeping the simulation and optimization components independent.

Chapter 2 contains all the information, justification, and theory on how the light is simulated. It details the methodologies used, the principles behind the simulation, and the rationale for the chosen approaches. This chapter serves as a comprehensive guide to the simulation aspect, providing the necessary background and details to understand the light propagation and its impact on communication within the retail environment.

Chapter 3 focuses on the optimization process, only assuming the existence of a signal matrix (or equivalent) that describes signal strength between ESL positions and transceiver positions. This chapter is independent of the simulation details, except for a sensitivity analysis that evaluates how the quality and computational effort of the optimization is affected by the quality of the simulation. It discusses various communication models, evaluates their effectiveness, and provides a detailed account of the optimization methodology and methods used to solve the problem.

By structuring the thesis this way, we hope that readers can benefit from it regardless of which component they are interested in. Each chapter is designed to (hopefully) stand on its own, accounting for respective components of the project. This structure is also meant to allow readers to understand the individual contributions of the simulation and optimization components, while also seeing how these components work together through the sensitivity analysis.

We conclude the thesis with chapter 4, which presents the conclusions and future work. Here we summarize the key takeaways from the thesis, discuss the implications of our findings, and outline potential directions for future research to improve and expand upon this work.

# Chapter 2

## Light Simulation

### 2.1 Requirements

There are several key considerations that guide the development of our framework. The primary goal is to evaluate the communication capabilities between transceivers and electronic shelf labels. In this context, capturing every minute electromagnetic detail of light distribution is not essential. Instead, the ability to handle arbitrary and complex geometry characteristic of retail stores is of key interest. These environments vary significantly in layout, shelf arrangement, and open spaces, necessitating a flexible approach to simulation. Additionally, the simulation must consider the material properties of different geometrical entities within the store and should be designed to minimize the effort required to model these entities and their properties. Leveraging existing tools for geometry production is preferable, and the simulation process should yield results within reasonable time frames without necessitating extraordinary computational resources.

Quantum electrodynamics, while offering an exhaustive description of light-matter interaction, provides a level of detail that exceeds our requirements. The focus on light's macroscopic behavior within the retail environment negates the need for such microscopic modeling, especially considering the significant computational resources it demands. A more conventional alternative, employing classical electromagnetism and solving Maxwell's equations through finite difference schemes, as is common in computational electrodynamics, would result in high accuracy. However, this approach is computationally expensive, particularly for large geometries, and does not align with our goal, where a complete description of the light's electromagnetic field throughout the entire computational domain is

not required. Additionally, this method necessitates powerful tooling for discretization and setting up boundary conditions in complex geometrical configurations.

Therefore, we have been inspired by the methodologies used in computer graphics. This domain concentrates on calculating measurements for virtual sensors – each pixel on a screen. This approach is similar to our objective of evaluating measurements for virtual sensors in retail spaces, where, unlike a concentrated screen area, sensors are distributed throughout the store. For instance, in video games or animated movies, the goal is to simulate light propagation and interaction within a scene, which essentially mirrors our task of determining light distribution in retail stores. Notably, computer graphics typically prioritize perceptual accuracy over absolute physical accuracy. Our application, not being constrained by the need for real-time processing, allows the inclusion of more detailed physical phenomena, which might otherwise be omitted in a video game due to performance requirements. The availability of consumer hardware and GPUs, specifically designed for such operations, ensures that these simulations can be executed efficiently on standard workstations.

Another aspect of these methods is the discrete representation of geometry using simple triangles, coupled with the intuitive description of material properties.\* Powerful, readily available tools can generate such complex geometries, making it easy to swap out different geometrical entities based on consumer data or a predefined collection of retail assets, such as shelves, cans, jars, and other common items. This approach liberates us from the complexities associated with meshing intricate geometries and establishing the necessary boundary conditions for methods that attempt to integrate Maxwell's equations directly.

## 2.2 Strategy

In the realm of computer graphics, the primary objective often involves solving the global illumination problem. This entails determining the incoming light on a sensor, accounting for both direct illumination from light sources and indirect illumination, which results from light reflecting off geometry within the scene. This concept is also central to our study, as indirect illumination will significantly contribute to the light received by electronic shelf labels.

Several approaches exist to address the global illumination problem. One

---

\*Triangles are also the primitive which GPUs are designed to work efficiently with.

method, rooted in the principles of radiative heat transfer, is the radiosity method, introduced to the graphics community in [1]. This method is both conceptually straightforward and mathematically accessible. An advantage of radiosity is its calculation of a view-independent light distribution, allowing for measurements throughout the scene without recalculation. However, the method heavily relies on the discretization of geometry and is particularly challenged by regions of rapidly changing light intensity such as shadows. Additionally, in its original formulation it is limited to diffusive surfaces. Subsequent refinements, as demonstrated by [2] and [3], have generalized the method to include non-diffusive materials, with further extensions for some specular and translucent materials by [4]. Despite these advancements, the increasing complexity and memory requirements of these methods necessitate sophisticated data structures and mesh refinement strategies.

For our purposes, relying on complex mesh refinements is impractical; we strongly prefer a simpler approach to handling geometry. Consequently, we have chosen to adopt the path tracing method, introduced by Kajiya in [5], to solve the global illumination problem. Path tracing naturally simulates a variety of physical phenomena, using Monte-Carlo integration, without the need for explicitly modeling these phenomena. While it is mathematically straightforward and produces unbiased results, it is also computationally demanding. Fortunately, while real-time processing is not a requirement for our framework, the operations necessary for path tracing are still supported by the capabilities of modern hardware, ensuring that simulations can be completed within a reasonable time frame.

In the subsequent sections, we will develop the theory and background necessary to understand path tracing, and explain the specific formulation employed in this thesis.

## 2.3 Radiometry

To describe the electromagnetic radiation in our environment, we employ techniques from radiometry – a macroscopic model of radiative transfer that details how light propagates and interacts with its surroundings. Radiometry adopts a geometrical perspective, utilizing rays to represent light propagation. As a result, optical phenomena arising from the wave-like behavior of light, such as diffraction and interference, are not accounted for. This simplification is justified for our goals, as our focus lies on the interaction of light with objects significantly larger than its wavelength.

In classical geometrical optics, several assumptions are at play. Light

is assumed to be unpolarized, with its spectrum as the sole defining characteristic. Moreover, as previously mentioned, we disregard wave-like behaviors like interference; hence, light rays do not interact with one another. We are essentially treating light as rays confined to linear paths in space.

Another assumption is the instantaneous propagation of light, implying that the light distribution is always in equilibrium. Additionally, we assume linearity, meaning that if we add or remove a light source, we expect the principle of superposition to apply.

These assumptions allow for a straightforward summation of all incoming light onto electronic shelf labels, as light is always in equilibrium and free from interference behavior. Furthermore, the linearity assumption allows us to simulate, for example all, light sources and their interactions independently.

### 2.3.1 The Fundamental Quantities of Radiometry

We proceed by introducing the key quantities in radiometry, assuming for mathematical purposes that energy is continuous. This allows the use of differential calculus in our description of energy transfer.

We start with the concept of radiant flux, also known as radiant power, which is the rate at which energy is transferred per unit of time. It quantifies the flow of energy to or from a surface over a period of time, see fig. 2.1a. Radiant flux,  $\Phi$ , is defined by:

$$\Phi = \frac{dQ}{dt} \quad (2.1)$$

where  $Q$  denotes the radiant energy emitted, reflected, or received, and  $t$  is time.

Next, we consider irradiance, which is the radiant flux received per unit area of a surface. It measures the energy reaching a point  $\mathbf{x}$  on a surface with a normal vector  $\mathbf{n}$ , as shown in fig. 2.1b. Irradiance,  $E$ , is defined as:

$$E(\mathbf{x}) = \frac{d\Phi(\mathbf{x})}{dA(\mathbf{x})}. \quad (2.2)$$

Conversely, radiant exitance refers to the radiant flux leaving a point  $\mathbf{x}$  on a surface.

Finally, we introduce the concept of radiance, a measure of the radiant flux emitted, reflected, or received per unit area perpendicular to a differential direction, per unit solid angle. Intuitively, radiance,  $L$ , quantifies the energy arriving or emanating from a differential area  $dA^\perp$ , perpendicular to a

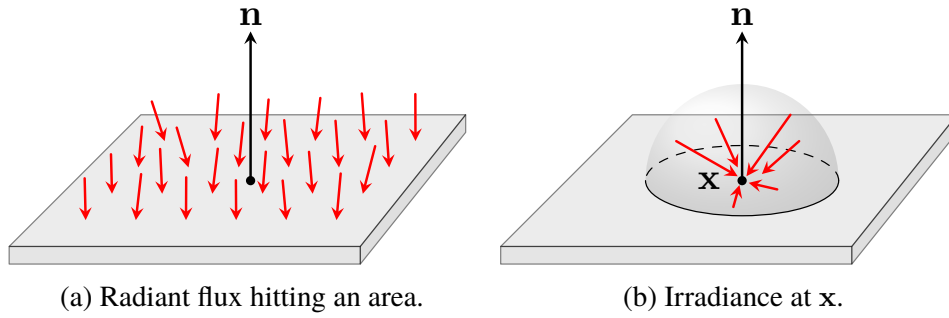


Figure 2.1: Radiant flux and irradiance.

differential solid angle  $d\omega$ , and is defined as:

Finally, we introduce the concept of radiance, a measure of the radiant flux emitted, reflected, or received per unit area perpendicular to a differential direction, per unit solid angle. Intuitively, radiance,  $L$ , quantifies the energy arriving or emanating from a differential area  $dA^\perp$ , perpendicular to a differential solid angle  $d\omega$ , and is defined as:

$$L(\mathbf{x}, \boldsymbol{\omega}) = \frac{d^2\Phi(\mathbf{x}, \boldsymbol{\omega})}{d\omega dA^\perp(\mathbf{x})}. \quad (2.3)$$

Using perpendicular area introduces challenges as radiance varies with the relative orientation of the incident direction and the surface. To consider surfaces at various angles relative to the ray direction, the projected area is used. This is achieved with the relation  $dA^\perp = (\mathbf{n} \cdot d\boldsymbol{\omega}) dA = \cos\theta dA$  – a relationship known as Lambert’s law which is illustrated in fig. 2.2b. With this formulation, we may calculate radiance at the projected area  $dA(\mathbf{x})$ , not only the perpendicular area  $dA^\perp(\mathbf{x})$ , see fig. 2.2a; the radiance expression becomes:

$$L(\mathbf{x}, \boldsymbol{\omega}) = \frac{d^2\Phi(\mathbf{x}, \boldsymbol{\omega})}{(\mathbf{n} \cdot d\boldsymbol{\omega}) d\omega dA(\mathbf{x})} = \frac{d^2\Phi(\mathbf{x}, \boldsymbol{\omega})}{\cos\theta d\omega dA(\mathbf{x})}. \quad (2.4)$$

Radiance is a useful and practical quantity because it remains constant along rays in empty space. To understand this, consider a two surfaces separated by a distance  $r$ , as illustrated in fig. 2.3. Consider the energy leaving  $dA(\mathbf{x})$  towards  $dA(\mathbf{x}')$ , which we denote  $d^3Q(\mathbf{x} \rightarrow \mathbf{x}') = L(\mathbf{x} \rightarrow \mathbf{x}') \cos\theta d\omega dA(\mathbf{x}) dt$  where  $\cos\theta = \mathbf{n} \cdot \boldsymbol{\omega}$ . The differential solid angle is given by  $d\omega = dA^\perp(\mathbf{x}')/r^2 = \cos\theta' dA(\mathbf{x}')/r^2$  where  $\cos\theta' = \mathbf{n}' \cdot \boldsymbol{\omega}'$ , and we may write  $d^3Q(\mathbf{x} \rightarrow \mathbf{x}') = L(\mathbf{x} \rightarrow \mathbf{x}') \cos\theta dA(\mathbf{x}) \cos\theta' dA(\mathbf{x}') dt$ . Similarly, the energy arriving at  $dA(\mathbf{x}')$

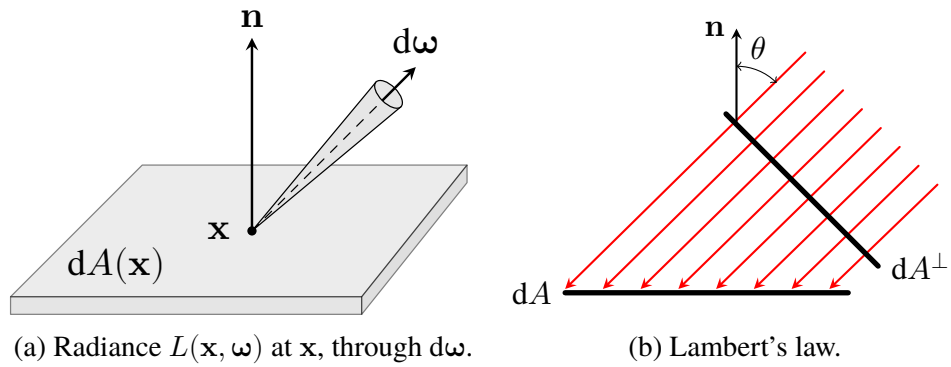


Figure 2.2: Radiance.

due to  $dA(\mathbf{x})$  is  $d^3Q(\mathbf{x}' \leftarrow \mathbf{x}) = L(\mathbf{x}' \leftarrow \mathbf{x}) \cos \theta' d\omega' dA(\mathbf{x}')dt$  and we get, again using  $d\omega' = dA^\perp(\mathbf{x})/r^2 = \cos \theta dA(\mathbf{x})/r^2$ , that  $d^3Q(\mathbf{x}' \leftarrow \mathbf{x}) = L(\mathbf{x}' \leftarrow \mathbf{x}) \cos \theta dA(\mathbf{x}) \cos \theta' dA(\mathbf{x}')dt$ . Since there is no medium, or possibly a lossless medium, the principle of conservation of energy necessitates that no energy is lost, i.e.,  $d^3Q(\mathbf{x} \rightarrow \mathbf{x}') = d^3Q(\mathbf{x}' \leftarrow \mathbf{x})$  which immediately gives  $L(\mathbf{x} \rightarrow \mathbf{x}') = L(\mathbf{x}' \leftarrow \mathbf{x})$ . This invariance property of radiance along a light ray makes it suitable to associate radiance with every light ray in simulations.

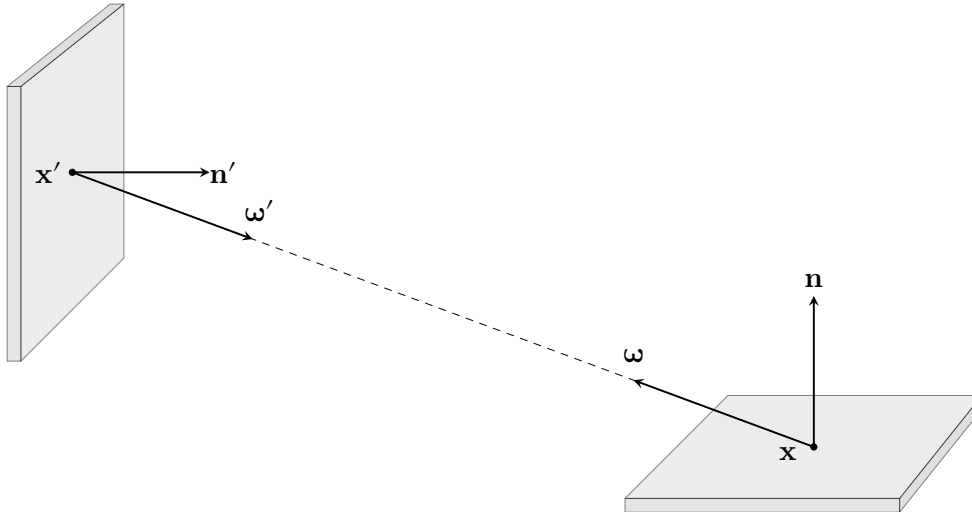


Figure 2.3: Invariance property of radiance

## 2.4 The Light Transport Equation

Light interaction with surfaces involves processes such as reflection, refraction, and absorption, dependent on the material properties. In the context of this thesis, we only consider opaque materials, thus we omit refraction. When light arrives at a point  $\mathbf{x}$  on a surface, traveling along direction  $\omega_i$ , the irradiance at  $\mathbf{x}$  is  $dE(\mathbf{x}, \omega_i) = L_i(\mathbf{x}, \omega_i) \cos \theta_i d\omega_i$  where  $L_i(\mathbf{x}, \omega_i)$  is the incident radiance. Assuming reflection occurs at the same point, the relationship between the incoming radiance and the reflected radiance,  $dL_o(\mathbf{x}, \omega_o)$ , in the direction  $\omega_o$  is described by:

$$f_r(\mathbf{x}, \omega_i, \omega_o) = \frac{dL_o(\mathbf{x}, \omega_o)}{dE(\mathbf{x}, \omega_i)} = \frac{dL_o(\mathbf{x}, \omega_o)}{L_i(\mathbf{x}, \omega_i) \cos \theta_i d\omega_i}. \quad (2.5)$$

This function,  $f_r$ , known as the *bidirectional reflectance distribution function* (BRDF), characterizes the reflective properties of a surface. A comprehensive discussion of the BRDF is provided in section 2.8. To compute the reflected radiance at point  $\mathbf{x}$ , we solve for  $dL_o(\mathbf{x}, \omega_o)$  and integrate over all directions in the unit hemisphere,  $\Omega$ , giving us the so called *reflectance equation*:\*

$$L_o(\mathbf{x}, \omega_o) = \int_{\Omega} f_r(\mathbf{x}, \omega_i, \omega_o) L_i(\mathbf{x}, \omega_i) \cos \theta_i d\omega_i. \quad (2.6)$$

In fig. 2.4, we show an illustration of the reflectance equation. It is important to keep in mind that the direction vectors,  $\omega_i$  and  $\omega_o$ , *always* point *out* from the surface. The reflectance equation expresses that the total reflected radiance is the integration of all incident radiance contributions over the hemisphere, weighted by the BRDF. This formulation presents a localized model of light interactions within a scene. The model is further refined by considering that a point  $\mathbf{x}$  not only reflects light based on the BRDF but may also emit light if it is a light source. The principle of energy conservation necessitates that the total exitant radiance  $L_o(\mathbf{x}, \omega_o)$  equals the sum of the reflected radiance, as dictated by the reflectance equation, and the emitted radiance in the outgoing direction, denoted by  $L_e(\mathbf{x}, \omega_o)$ . This leads to the *light transport equation* (LTE),<sup>†</sup> a fundamental concept in light transport theory introduced by James

\*If light is allowed to refract, the equation is sometimes called the scattering equation, or surface scattering equation, and the integration is carried out over the entire unit sphere instead of the unit hemisphere.

<sup>†</sup>In Kajiyā's paper, the equation is termed the rendering equation; however, given that our focus is not on rendering, this paper adopts the term light transport equation, which is also commonly used.

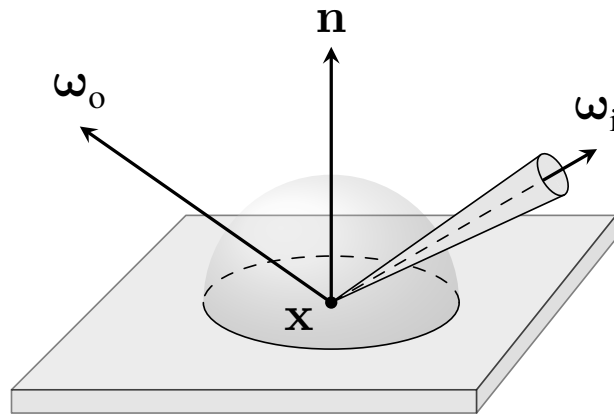


Figure 2.4: Illustration of the reflectance equation.

Kajiya in [5], which is formulated as:

$$L_o(\mathbf{x}, \boldsymbol{\omega}_o) = L_e(\mathbf{x}, \boldsymbol{\omega}_o) + \int_{\Omega} f_r(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\mathbf{x}, \boldsymbol{\omega}_i) \cos \theta_i d\boldsymbol{\omega}_i. \quad (2.7)$$

The LTE embodies an intuitive concept: the light at each point on a surface within an environment is essentially the cumulative effect of light from all other surfaces. This interdependence means that the illumination of each surface is a consequence of the light from every other surface.

However, an inconvenience is the presence of two distinct radiance terms, the exitant radiance and the incident radiance,  $L_o$  and  $L_i$  respectively. To establish an explicit relationship between these terms, the approach from the standard textbook [6] is adopted. Considering that radiance remains constant along light paths in the absence of a medium, the incident radiance at a point can be described using the exitant radiance from another point. This relationship is expressed as  $L_i(\mathbf{x}, \boldsymbol{\omega}) = L_o(r(\mathbf{x}, \boldsymbol{\omega}), -\boldsymbol{\omega})$ , where the *ray-casting function*  $r(\mathbf{x}, \boldsymbol{\omega})$  is introduced. This function emits a ray from point  $\mathbf{x}$  in direction  $\boldsymbol{\omega}$ , and returns the first surface intersection. Should the ray not intersect with any surface, the function outputs a null point  $\Lambda$ , signifying an absence of intersection and resulting in  $L_o(\Lambda, \boldsymbol{\omega}) = 0$ .\* Incorporating this

---

\*The symbol  $\Lambda$ , representing a null intersection in light transport theory can also be found in [7]. In a different realm, Donald Knuth use  $\Lambda$  to denote null links in data structures, symbolizing links leading nowhere.

relationship in eq. (2.7) results in the following formulation of the LTE:

$$L_o(\mathbf{x}, \boldsymbol{\omega}_o) = L_e(\mathbf{x}, \boldsymbol{\omega}_o) + \int_{\Omega} f_r(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_o(r(\mathbf{x}, \boldsymbol{\omega}_i), -\boldsymbol{\omega}_i) \cos \theta_i \, d\boldsymbol{\omega}_i. \quad (2.8)$$

While simpler, since we only deal with one radiance term, the geometric relationship between surfaces are implicitly defined by the ray-casting function. To understand the geometric relationships in the scene better, we can transform the LTE we have studied so far – the hemispherical formulation – to a surface formulation.

In the surface formulation we integrate over all surfaces,  $\mathcal{A}$ , instead of all directions of the hemisphere,  $\Omega$ . Of course, only points on surfaces that are mutually visible should contribute radiance. Fortunately, the ray-casting function allows us to conveniently define a visibility function,  $V$ . For every  $\mathbf{x}, \mathbf{x}' \in \mathcal{A}$ , we define the visibility function as:

$$V(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if } \exists \boldsymbol{\omega} r(\mathbf{x}, \boldsymbol{\omega}) = \mathbf{x}', \\ 0 & \text{otherwise.} \end{cases} \quad (2.9)$$

The visibility function simply specifies the visibility between two arbitrary points  $\mathbf{x}$  and  $\mathbf{x}'$ . Then, we instead write  $L_o(r(\mathbf{x}, \boldsymbol{\omega}_i), -\boldsymbol{\omega}_i) = L_o(\mathbf{x}', -\boldsymbol{\omega}_i) V(\mathbf{x}, \mathbf{x}')$ . Note that this is the exitant radiance *from* a point  $\mathbf{x}'$  in the direction of  $-\boldsymbol{\omega}_i$ , that is, towards  $\mathbf{x}$ . The differential solid angle  $d\boldsymbol{\omega}_i$  subtended by the differential surface  $dA(\mathbf{x}')$  at  $\mathbf{x}'$  is:

$$d\boldsymbol{\omega}_i = \frac{\cos \theta'_o}{\|\mathbf{x}' - \mathbf{x}\|^2} dA(\mathbf{x}') \quad (2.10)$$

where  $\cos \theta'_o$  is the angle between the normal at  $\mathbf{x}'$  and  $-\boldsymbol{\omega}_i$ . Inserting into eq. (2.8), we get:

$$L_o(\mathbf{x}, \boldsymbol{\omega}_o) = L_e(\mathbf{x}, \boldsymbol{\omega}_o) + \int_{\mathcal{A}} f_r(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_o(\mathbf{x}', -\boldsymbol{\omega}_i) V(\mathbf{x}, \mathbf{x}') \frac{\cos \theta_i \cos \theta'_o}{\|\mathbf{x}' - \mathbf{x}\|^2} dA(\mathbf{x}').$$

It is customary to combine the geometric factors into a single term  $G$ , defined as:

$$G(\mathbf{x}, \mathbf{x}') = V(\mathbf{x}, \mathbf{x}') \frac{\cos \theta_i \cos \theta'_o}{\|\mathbf{x}' - \mathbf{x}\|^2} \quad (2.11)$$

which leads to the surface formulation of the LTE:

$$L_o(\mathbf{x}, \boldsymbol{\omega}_o) = L_e(\mathbf{x}, \boldsymbol{\omega}_o) + \int_{\mathcal{A}} f_r(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_o(\mathbf{x}', -\boldsymbol{\omega}_i) G(\mathbf{x}, \mathbf{x}') dA(\mathbf{x}'). \quad (2.12)$$

The recursive nature of the LTE – evident from the radiance term appearing on both sides of the equation – while conceptually simple and elegant, presents significant computational challenges. There are no analytical solutions for this equation except in the most simple scenarios which fall outside our scope of interest. Therefore, the primary focus of the remaining subsections of this section is to develop efficient computational strategies for solving the LTE for environments involving arbitrary geometry.

## 2.5 Measuring Radiance

The computation of radiance is fundamentally aimed at measuring incident radiance on one or several sensors, in order to determine their responses. This necessitates a short review of sensor response, primarily informed by Nicodemus' work [8]. However, the discussion here employs more contemporary notation from the field of light transport.

Responsivity, the output or response of a device used in radiation measurements, quantifies the output signal, denoted as  $S$ , relative to the incident radiation. The intentionally broad representation of  $S$  accommodates various potential units of measurement, such as voltage or current, depending on the sensor type. Assuming a linear relationship between the sensor response and the incident radiant power, we define *flux responsivity* as:

$$W(\mathbf{x}, \boldsymbol{\omega}) = \frac{dS(\mathbf{x}, \boldsymbol{\omega})}{d\Phi(\mathbf{x}, \boldsymbol{\omega})}. \quad (2.13)$$

This definition takes into account the response variation based on position and direction, thereby characterizing flux responsivity as the sensor response per unit of radiant power received at point  $\mathbf{x}$  from direction  $\boldsymbol{\omega}$ .

The sensor's response is expressed as:

$$dS(\mathbf{x}, \boldsymbol{\omega}) = W(\mathbf{x}, \boldsymbol{\omega}) d\Phi(\mathbf{x}, \boldsymbol{\omega}) = W(\mathbf{x}, \boldsymbol{\omega}) L_i(\mathbf{x}, \boldsymbol{\omega}) \cos \theta d\boldsymbol{\omega} dA(\mathbf{x}).$$

Consequently, the total output signal or the total sensor response is the cumulative effect of radiance received at all sensor points, from every

direction. This total response, denoted  $M$ , is calculated using the *measurement equation*:

$$\begin{aligned} M &= \int_{\mathcal{A}} \int_{\Omega} W(\mathbf{x}, \boldsymbol{\omega}) L_i(\mathbf{x}, \boldsymbol{\omega}) \cos \theta \, d\boldsymbol{\omega} \, dA(\mathbf{x}) \\ &= \int_{\mathcal{A}} \int_{\mathcal{A}} W(\mathbf{x}, \boldsymbol{\omega}) L_o(\mathbf{x}', -\boldsymbol{\omega}) G(\mathbf{x}, \mathbf{x}') \, dA(\mathbf{x}') \, dA(\mathbf{x}) \end{aligned} \quad (2.14)$$

where  $\mathcal{A}$  is the set of all surfaces, and in the last equality we expressed the inner integral as an integral over  $\mathcal{A}$ , as we did for the hemispherical LTE in section 2.4. It is required to define  $W$  for all surfaces and directions to enable integration. In practical terms, this implies that sensors form a surface within the scene, with  $W$  typically being non-zero for only a limited portion of this domain, i.e., on sensors.

## 2.6 Path Integral Formulation

In this section, the goal is to recast the LTE, transforming it from an integral equation into a problem of integration. Our objective is to develop an *explicit* integral formulation that, upon calculation, yields the irradiance at a specific point on a surface within the environment. By rephrasing the problem in terms of integration, we open the door to a vast array of established general-purpose integration methods, with Monte-Carlo integration being of particular relevance.

The approach adopted is primarily based on the principles outlined in Veach's seminal thesis [9]. Readers seeking a more rigorous mathematical exposition, especially regarding the measure theory underpinning the integration process, are encouraged to consult his work.

We begin by writing the surface formulation of the LTE in the so called *three-point form*, illustrated in fig. 2.5. The exitant radiance from a point  $\mathbf{x}$  to a point  $\mathbf{x}'$ , formerly denoted  $L_o(\mathbf{x}, \boldsymbol{\omega})$ , we now denote as  $L_o(\mathbf{x} \rightarrow \mathbf{x}')$  where  $\boldsymbol{\omega}$  is the unit vector in the direction of  $\mathbf{x}' - \mathbf{x}$ . Similarly, the BRDF of the form  $f_r(\mathbf{x}', \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$  we now denote as  $f_r(\mathbf{x} \rightarrow \mathbf{x}' \rightarrow \mathbf{x}'')$  where  $\boldsymbol{\omega}_i$  is the unit vector in direction of  $\mathbf{x}' - \mathbf{x}$  and  $\boldsymbol{\omega}_o$  is the unit vector in direction of  $\mathbf{x}'' - \mathbf{x}'$ . Adopting this notation in the surface formulation of the LTE, eq. (2.12), we

get the three-point form of the LTE:

$$L(\mathbf{x}' \rightarrow \mathbf{x}'') = L_e(\mathbf{x}' \rightarrow \mathbf{x}'') + \int_{\mathcal{A}} f_r(\mathbf{x} \rightarrow \mathbf{x}' \rightarrow \mathbf{x}'') L(\mathbf{x} \rightarrow \mathbf{x}') G(\mathbf{x}, \mathbf{x}') dA(\mathbf{x}) \quad (2.15)$$

where we omit the subscript for exitant radiance since the direction is now obvious.

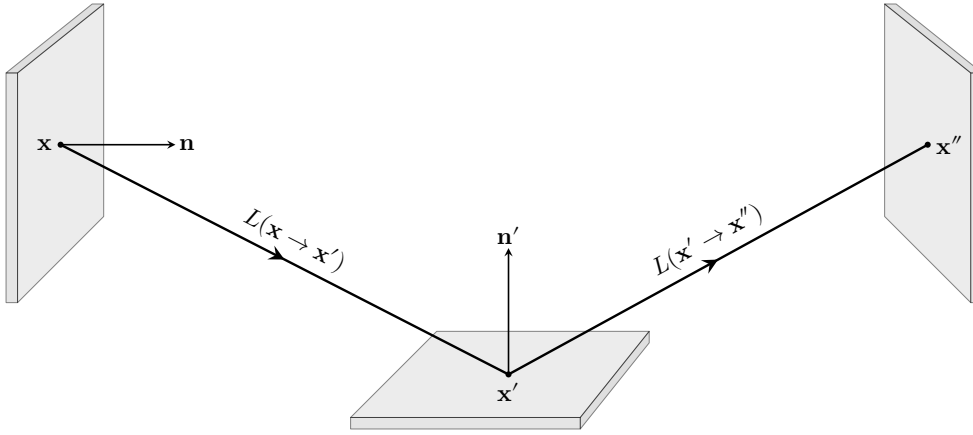


Figure 2.5: Three-point-form of the LTE.

The radiance due to a point  $\mathbf{x}'$  at a sensor, using the current notation in eq. (2.14), is then:

$$M = \int_{\mathcal{A}} \int_{\mathcal{A}} W(\mathbf{x}' \rightarrow \mathbf{x}) L(\mathbf{x}' \rightarrow \mathbf{x}) G(\mathbf{x}, \mathbf{x}') dA(\mathbf{x}') dA(\mathbf{x})$$

and then recursively expanding the radiance term using eq. (2.15) which results in:

$$M = \sum_{k=1}^{\infty} \int_{\mathcal{A}^{k+1}} W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) G(\mathbf{x}_i, \mathbf{x}_{i+1}) \prod_{i=0}^k dA(\mathbf{x}_i).$$

To really get an understanding what this equation represents, we explicitly

write out a few terms:

$$\begin{aligned}
M &= \sum_{k=1}^{\infty} \int_{\mathcal{A}^{k+1}} W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) \\
&\quad \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) G(\mathbf{x}_i, \mathbf{x}_{i+1}) \prod_{i=0}^k dA(\mathbf{x}_i) \\
&= \int_{\mathcal{A}^2} L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) W(\mathbf{x}_0 \rightarrow \mathbf{x}_1) dA(\mathbf{x}_0) dA(\mathbf{x}_1) \\
&+ \int_{\mathcal{A}^3} L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) \\
&\quad \cdot f_r(\mathbf{x}_0 \rightarrow \mathbf{x}_1 \rightarrow \mathbf{x}_2) G(\mathbf{x}_1, \mathbf{x}_2) W(\mathbf{x}_1 \rightarrow \mathbf{x}_2) dA(\mathbf{x}_0) dA(\mathbf{x}_1) dA(\mathbf{x}_2) \\
&+ \dots
\end{aligned}$$

Here, the  $k^{\text{th}}$  term corresponds to the radiance measured at a point on the sensor, resulting from all potential light paths of length  $k$  originating from a light source. For example, the fourth term would correspond to a path of length  $k = 3$ , with its integrand being the product of the factors in fig. 2.6. By integrating over all paths of a specific length  $k$ , we obtain the total measured radiance contributed by all paths of length  $k$ , and this is indeed exactly the  $k^{\text{th}}$  term.

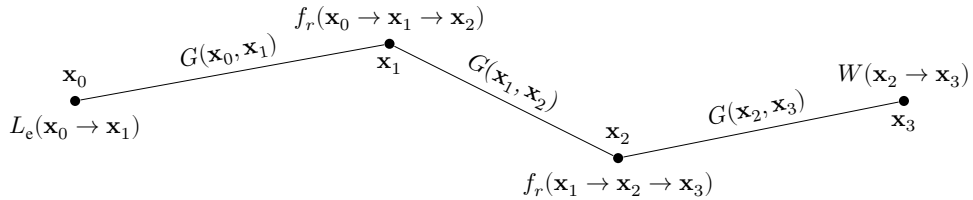


Figure 2.6: Factors contributing to a measurement due to a path of length three.

In [9], Veach formalizes this concept; he defines  $\mathcal{P}_k$  as the set of all paths with a finite length  $k$ . Each path in this set, represented as  $\bar{\mathbf{x}}_k = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k)$ , comprises  $k + 1$  vertices that all lie on some surface in  $\mathcal{A}$ . On this set of paths, a measure  $\mu_k$  is defined as:

$$\mu_k(D) = \int_D \prod_{i=0}^k dA(\mathbf{x}_i)$$

where  $D \in \mathcal{P}_k$  is a set of paths, which indeed means that  $d\mu_k = \prod_{i=0}^k dA(\mathbf{x}_i)$ . Next, Veach defines the path space  $\mathcal{P}$  as  $\mathcal{P} = \bigcup_{k=1}^{\infty} \mathcal{P}_k$  and extends the

measure  $\mu$  to the path space as  $\mu(D) = \sum_{k=1}^{\infty} \mu_k(D \cap \mathcal{P}_k)$ , thus the measure of a set of paths is indeed the sum of the measures of the paths of each length. This allows us to formally write the *path integral formulation* of the LTE as:

$$\begin{aligned} M &= \int_{\mathcal{P}} f(\bar{\mathbf{x}}) d\mu(\bar{\mathbf{x}}) \\ &= \sum_{k=1}^{\infty} \int_{\mathcal{P}_k} f(\bar{\mathbf{x}}_k) d\mu(\bar{\mathbf{x}}_k) = \sum_{k=1}^{\infty} \int_{\mathcal{A}^{k+1}} f(\bar{\mathbf{x}}_k) dA(\mathbf{x}_0) \dots dA(\mathbf{x}_k) \end{aligned} \quad (2.16)$$

where we defined the *measurement contribution function*:

$$\begin{aligned} f(\bar{\mathbf{x}}_k) &= W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) \\ &\quad \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) G(\mathbf{x}_i, \mathbf{x}_{i+1}). \end{aligned} \quad (2.17)$$

This reformulation of the LTE offers several benefits, as previously noted, it allows for the representation of the sensor's entire measurement as a single integral. More significantly, this reformulation shifts the focus to the path of light as the primary entity, rather than traditional radiometric quantities like radiance. Conceptually, envisioning light as traversing paths that emanate from light sources simplifies the understanding. The generation of these paths is highly flexible; for instance, paths can originate from the sensor, or they may involve randomly sampling points on surfaces – it is not even necessary for paths to originate from the light source. This flexibility becomes particularly advantageous when employing Monte-Carlo integration; the Monte-Carlo estimate of eq. (2.16) is simply  $f(\bar{\mathbf{X}})/p(\bar{\mathbf{X}})$  where the paths  $\bar{\mathbf{X}}$  are selected from a probability distribution  $p(\bar{\mathbf{x}})$ . Essentially, the process involves sampling paths according to a probability density, resulting in an unbiased estimate of the integral:

$$\mathbb{E} \left[ \frac{f(\bar{\mathbf{X}})}{p(\bar{\mathbf{X}})} \right] = \int_{\mathcal{P}} \frac{f(\bar{\mathbf{x}})}{p(\bar{\mathbf{x}})} p(\bar{\mathbf{x}}) d\mu(\bar{\mathbf{x}}) = \int_{\mathcal{P}} f(\bar{\mathbf{x}}) d\mu(\bar{\mathbf{x}}) = M.$$

Indeed, the difficulty shifts towards generating paths in an efficient manner. The challenge lies not merely in creating any paths, but paths that are representative of the actual light transport in the scene and are also computationally feasible.

A problem that arises is that we need to estimate an infinite number of terms but we do not have infinite computation. Generally, paths that are

longer carry less energy than shorter paths since energy is absorbed whenever light reflects. Thus, we could for example decide on a maximum allowed path length. However, this would introduce bias since we would consistently underestimate the integral. We effectively want to put more emphasis on paths of shorter length than those of longer lengths, but do so without introducing bias. To do so, we will use *Russian roulette*, a technique to randomly skip evaluating terms that contribute less. Say, at each term in the infinite sum eq. (2.16) we conduct a Bernoulli trial that decides whether or not to continue the evaluation of the sum. Let  $q_i$  denote the probability that trial  $i$  terminates the sum, so the probability of term  $k$  surviving is  $\prod_{i=1}^k 1 - q_i$  then consider the following estimator and its expected value:

$$\mathbb{E} \left[ \prod_{i=1}^K \frac{1}{1 - q_i} \int_{\mathcal{P}_K} f(\bar{\mathbf{x}}_K) d\mu(\bar{\mathbf{x}}_K) \right] = \sum_{k=1}^{\infty} \int_{\mathcal{P}_k} f(\bar{\mathbf{x}}_k) d\mu(\bar{\mathbf{x}}_k) = M.$$

Thus, the application of Russian roulette does not introduce bias. Then, in the path integral framework we can do successive Bernoulli trials until termination to decide the path length of each sample and weight it accordingly. A Russian roulette terminated path of length  $k$  then has the estimator:

$$\frac{f(\bar{\mathbf{x}}_k)}{p(\bar{\mathbf{x}}_k)} \prod_{i=1}^k \frac{1}{1 - q_i}. \quad (2.18)$$

We are effectively creating paths of random lengths, and as long as we weight them correctly with their probability we are not introducing bias. It is important to note that there are no restrictions on the probabilities  $q_i$ , they are arbitrary and can depend on the path in some manner.

## 2.7 Path Sampling

There are many possible ways to generate a path  $\bar{\mathbf{x}}_k$ , the simplest one being sampling random points on surfaces in the scene in a uniform manner. However, it is highly unlikely these points form a polyline without intersecting some geometry, that is each consecutive point is mutually visible. To combat this, we can construct a path incrementally – point to point – thus making sure the path stays connected. This is effectively a random walk along surfaces in the scene.

In this section, we first consider a simple random walk starting from the light source, then we will consider a version of random walks that tries to

terminate at sensors, as well as discuss paths in the opposite direction – from the sensor towards light sources. Then, we discuss a powerful way to combine estimators using multiple importance sampling, as described by Veach in [9].

### 2.7.1 Random Walk

The primary method for constructing random walks is through ray tracing. To determine the next vertex,  $\mathbf{x}_i$ , in the path, we trace a ray from  $\mathbf{x}_{i-1}$  in a randomly sampled direction, until it intersects with a surface. This means that the probability of an interior point being a part of the path is dependent on the sequence of preceding points. In other words, the probability that  $\mathbf{x}_i$  is part of the path is  $p(\mathbf{x}_i | \mathbf{x}_{i-1}, \dots, \mathbf{x}_0)$ . For simplicity, we will refer to these probabilities as  $p(\mathbf{x}_i)$ , bearing in mind that they may actually be conditional probabilities. Additionally, the path is constructed with respect to solid angle – due to sampling directions – but we need the densities to be with respect to area. Say we generated  $\mathbf{x}_i$  as the intersection of a surface and the ray originating from  $\mathbf{x}_{i-1}$  in the randomly sampled direction  $\omega_o^{(i-1)}$  such that we have  $\omega_o^{(i-1)} = \mathbf{x}_i - \mathbf{x}_{i-1}$ , assumed to be of unit length. The solid angle subtended at  $\mathbf{x}_{i-1}$  per unit surface area at  $\mathbf{x}_i$  is given by the transformation factor  $\cos \theta_i^{(i)} / \|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2$  where  $\theta_i^{(i)}$  is the incident angle at  $\mathbf{x}_i$ . Then, the densities are related by:

$$p(\mathbf{x}_i) = p_{i-1}(\omega_o^{(i-1)}) \frac{\cos \theta_i^{(i)}}{\|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2}$$

where  $p_{i-1}(\omega_o^{(i-1)})$  is the density which  $\omega_o^{(i-1)}$  was sampled according to, measured with respect to solid angle. Now, the path probability density function for the path  $\bar{\mathbf{x}}_k$  is simply the joint density given by the product of each vertex density:

$$p(\bar{\mathbf{x}}_k) = \prod_{i=0}^k p(\mathbf{x}_i) = p(\mathbf{x}_0) \prod_{i=1}^k p_{i-1}(\omega_o^{(i-1)}) \frac{\cos \theta_i^{(i)}}{\|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2}$$

where the first vertex is generated according to some a priori distribution.

Studying the Monte-Carlo estimator  $f(\bar{\mathbf{x}}_k)/p(\bar{\mathbf{x}}_k)$  shows:

$$\begin{aligned} & \frac{W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) G(\mathbf{x}_0, \mathbf{x}_1) \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) G(\mathbf{x}_i, \mathbf{x}_{i+1})}{p(\mathbf{x}_0) \prod_{i=1}^k p_{i-1}(\boldsymbol{\omega}_0^{(i-1)}) \frac{\cos \theta_i^{(i)}}{\|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2}} \\ &= \frac{W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) \prod_{i=1}^k G(\mathbf{x}_{i-1}, \mathbf{x}_i)}{p(\mathbf{x}_0) \prod_{i=1}^k p_{i-1}(\boldsymbol{\omega}_0^{(i-1)}) \prod_{i=1}^k \frac{\cos \theta_i^{(i)}}{\|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2}} \end{aligned}$$

then, using eq. (2.11), we have that:

$$\frac{G(\mathbf{x}_{i-1}, \mathbf{x}_i)}{\frac{\cos \theta_i^{(i)}}{\|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2}} = V(\mathbf{x}_{i-1}, \mathbf{x}_i) \cos \theta_0^{(i-1)}$$

where we also used that  $\|\mathbf{x}_i - \mathbf{x}_{i-1}\|^2 = \|\mathbf{x}_{i-1} - \mathbf{x}_i\|^2$  which substituted into the estimator above yields:

$$\frac{W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) \prod_{i=1}^{k-1} f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1})}{p(\mathbf{x}_0) \prod_{i=1}^k p_{i-1}(\boldsymbol{\omega}_0^{(i-1)})} \prod_{i=1}^k V(\mathbf{x}_{i-1}, \mathbf{x}_i) \cos \theta_0^{(i-1)}$$

but by construction, using ray tracing, we have that  $\prod_{i=1}^k V(\mathbf{x}_{i-1}, \mathbf{x}_i) = 1$  so the Monte-Carlo estimator is:

$$\frac{f(\bar{\mathbf{x}}_k)}{p(\bar{\mathbf{x}}_k)} = W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) \frac{L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1)}{p(\mathbf{x}_0) p_0(\boldsymbol{\omega}_0)} \prod_{i=1}^{k-1} \frac{f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) \cos \theta_i}{p_i(\boldsymbol{\omega}_i)}$$

where we no longer explicitly denote the outgoing direction. If we apply Russian roulette, i.e., the path is terminated with probability  $q_i$  at each bounce, then a path of length  $k$  has the following estimator:

$$\begin{aligned} \frac{f(\bar{\mathbf{x}}_k)}{p(\bar{\mathbf{x}}_k)} &= W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) \frac{L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1)}{p(\mathbf{x}_0) p_0(\boldsymbol{\omega}_0)} \\ &\quad \prod_{i=1}^{k-1} \frac{1}{1 - q_i} \frac{f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) \cos \theta_i}{p_i(\boldsymbol{\omega}_i)}. \end{aligned}$$

As mentioned in a previous section, the probabilities  $q_i$  may depend on the path: we can simply let the  $q_i$  be inversely proportional to how much radiance the path carries at that bounce. Thus, there is a higher probability of terminating paths with estimated low contribution.

The natural question that arises at this point is the choice of density functions,  $p_i(\omega_i)$ , to sample directions from. While one could sample the directions over the hemisphere in a uniform manner, it is essential to acknowledge that, in reality, the behavior of light reflection is contingent upon the material properties of the reflective surface. Thus, sampling the BRDF to generate the directions should yield paths that are more physically accurate. In fact, this is importance sampling since we are prioritizing directions where more light is reflected and thus we produce paths that are more representative of the environment, effectively reducing the variance of the estimator. The practical details of sampling from a BRDF is discussed in section 2.8.4.

## 2.7.2 Next Event Estimation

A different technique is to make random walks but try to connect the path to a sensor chosen at random to not introduce bias. In fact, the sensor can be sampled at every step of the path. That is, we make a random walk starting from the light source, and at every step we sample a random sensor and connect the path to that sensor if visibility allows; then, we reuse the previous vertices of that path, excluding the vertex at the sensor, and continue the random walk until Russian roulette terminates the walk. The  $i^{\text{th}}$  step of a path of length  $k$  then has the estimator:

$$\frac{f(\bar{\mathbf{x}}_i)}{p(\bar{\mathbf{x}}_i)} = \frac{L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1) W(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i) f_r(\mathbf{x}_{i-2} \rightarrow \mathbf{x}_{i-1} \rightarrow \mathbf{x}_i) G(\mathbf{x}_{i-1}, \mathbf{x}_i)}{p(\mathbf{x}_0) p_0(\omega_0) p_s(\mathbf{x}_i) \prod_{j=1}^{i-2} \frac{f_r(\mathbf{x}_{j-1} \rightarrow \mathbf{x}_j \rightarrow \mathbf{x}_{j+1}) \cos \theta_j}{p_j(\omega_j)}}$$

where we omit the Russian roulette weights for brevity. Here,  $p_s(\mathbf{x}_i)$  is some probability distribution over the sensors of the scene. This technique is commonly referred to as *next event estimation*. In fact, one usually reverses the direction of the path: starting the path at the sensor and then sampling light sources at every step. This is conceptually easier since the sensor for which we are conducting the measurement for is fixed before the path begins. For a

particular sensor, the estimator for the  $i^{\text{th}}$  step of a path of length  $k$  is:

$$\frac{f(\bar{\mathbf{x}}_i)}{p(\bar{\mathbf{x}}_i)} = \frac{W(\mathbf{x}_0 \rightarrow \mathbf{x}_1) L_e(\mathbf{x}_i \rightarrow \mathbf{x}_{i-1}) f_r(\mathbf{x}_i \rightarrow \mathbf{x}_{i-1} \rightarrow \mathbf{x}_{i-2}) G(\mathbf{x}_i, \mathbf{x}_{i-1})}{p(\mathbf{x}_0) p_0(\boldsymbol{\omega}_0) p_e(\mathbf{x}_i)} \prod_{j=1}^{i-2} \frac{f_r(\mathbf{x}_{j+1} \rightarrow \mathbf{x}_j \rightarrow \mathbf{x}_{j-1}) \cos \theta_j}{p_j(\boldsymbol{\omega}_j)}$$

where  $p_e$  is some probability over the light sources.

### 2.7.3 Multiple Importance Sampling

There are both drawbacks and benefits of each sampling technique; sampling the light sources can be good when the light sources are small and the material is diffuse, but if light sources are large and the material is less diffuse it might be better to sample the BRDF. Neither technique is necessarily best for an arbitrary scene. We would like to get the best of both worlds by combining these techniques that preserves the strength of each one. This is what multiple importance sampling (MIS) achieves.

The idea behind MIS is to sample according to both\* techniques and then weight the samples to create an estimator that have lower variance for a wide class of integrands, what Veach calls a robust estimator.

The idea is as follows, say we have two sampling distributions  $p_a$  and  $p_b$  and a sample from each,  $\bar{\mathbf{X}}$  and  $\bar{\mathbf{Y}}$  respectively. Then, the estimator is:

$$w_a(\bar{\mathbf{X}}) \frac{f(\bar{\mathbf{X}})}{p_a(\bar{\mathbf{X}})} + w_b(\bar{\mathbf{Y}}) \frac{f(\bar{\mathbf{Y}})}{p_b(\bar{\mathbf{Y}})}$$

where  $w_a$  and  $w_b$  are the weight functions. A study of the expectation:

$$\begin{aligned} \mathbb{E} \left[ w_a(\bar{\mathbf{X}}) \frac{f(\bar{\mathbf{X}})}{p_a(\bar{\mathbf{X}})} + w_b(\bar{\mathbf{Y}}) \frac{f(\bar{\mathbf{Y}})}{p_b(\bar{\mathbf{Y}})} \right] &= \int_{\mathcal{P}} w_a(\bar{\mathbf{x}}) \frac{f(\bar{\mathbf{x}})}{p_a(\bar{\mathbf{x}})} p_a(\bar{\mathbf{x}}) d\mu(\bar{\mathbf{x}}) \\ &+ \int_{\mathcal{P}} w_b(\bar{\mathbf{y}}) \frac{f(\bar{\mathbf{y}})}{p_b(\bar{\mathbf{y}})} p_b(\bar{\mathbf{y}}) d\mu(\bar{\mathbf{y}}) \\ &= \int_{\mathcal{P}} (w_a(\bar{\mathbf{x}}) + w_b(\bar{\mathbf{x}})) f(\bar{\mathbf{x}}) d\mu(\bar{\mathbf{x}}) \end{aligned}$$

shows that we need  $w_a(\bar{\mathbf{x}}) + w_b(\bar{\mathbf{x}}) = 1$  to have an unbiased estimator. In [9], Veach argues more formally about the requirements of these weight functions,

---

\*Multiple importance sampling is not limited to only two techniques, but can be used for an arbitrary amount of techniques.

as well as providing several ways to calculate the weights. He proves that no other strategy is much better than the so called balance heuristic:

$$w_i(x) = \frac{p_i(x)}{p_a(x) + p_b(x)} \quad (2.19)$$

which give us the following MIS estimator:

$$\frac{f(\bar{\mathbf{X}})}{p_a(\bar{\mathbf{X}}) + p_b(\bar{\mathbf{Y}})} + \frac{f(\bar{\mathbf{Y}})}{p_a(\bar{\mathbf{X}}) + p_b(\bar{\mathbf{Y}})}.$$

Please note that we only present a special case of the theory Veach develops in his thesis, and the reader is encouraged to consult his thesis for a more complete and general description.

If we now apply this technique, the  $i^{\text{th}}$  step of a path has the estimator:

$$\frac{f(\bar{\mathbf{x}}_k)}{p(\bar{\mathbf{x}}_k)} = W(\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k) \frac{L_e(\mathbf{x}_0 \rightarrow \mathbf{x}_1)}{p(\mathbf{x}_0)p_0(\boldsymbol{\omega}_0)} \prod_{i=1}^{k-1} \frac{1}{1 - q_i} \frac{f_r(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) \cos \theta_i}{p_i(\boldsymbol{\omega}_i)}.$$

where  $p_l$  and  $p_b$  are the probability distribution functions used for each sampling technique, the first one for the lights and the second one for sampling according to the BRDF. We also defined the path throughput up to vertex  $i$  as:

$$\beta_i = \prod_{j=1}^i \frac{f_r(\mathbf{x}_{j+1} \rightarrow \mathbf{x}_j \rightarrow \mathbf{x}_{j-1}) \cos \theta_j}{p_j(\boldsymbol{\omega}_j)}. \quad (2.20)$$

In effect, this method is estimating the integral in two different ways – light sampling and BRDF sampling – and combining the result. Technically, this could be done one after the other, but it is more common to take both samples and combine them at each step of the path.

## 2.8 BRDF

The study of electromagnetic wave scattering has been studied for a long time for radar technology, but these models for reflection are usually studied assuming light behaves like a wave. The original work from Beckmann and Spizzichino [10] serves as an introduction to such models as well as an introduction to microfacet theory and their distributions. Torrance and Sparrow [11] later formulated these scattering principles in the framework of geometrical optics, presenting the general form of the bidirectional reflection

distribution function (BDRF) adopted in this paper.

We begin by discussing some simple properties the BRDF should have to accurately model physically plausible reflections, and then a review of the microfacet theory, then present some BRDF models from the literature, and finally we show how to sample from them. The reader should note that we are strictly talking about *reflection* functions, thus these properties are not necessarily true for the more general scattering functions that are found in the literature, commonly called BSDF.

### 2.8.1 Properties

The BRDF should model light interactions with surfaces according to physical laws, thus we require the following three properties [6] to hold for any physically plausible BRDF:

- *Positivity*: No surface can have negative reflectance, mathematically  $f_r(\mathbf{x}, \omega_i, \omega_o) \geq 0$ .
- *Helmholtz reciprocity*: The surface reflection should be invariant of the direction of light, mathematically  $f_r(\mathbf{x}, \omega_i, \omega_o) = f_r(\mathbf{x}, \omega_o, \omega_i)$ .
- *Energy conservation*: Light can only be absorbed or reflected, i.e., the exitant light can never exceed the incident light, mathematically  $\int_{\Omega} f_r(\mathbf{x}, \omega_i, \omega_o) \cos \theta_o d\omega_o \leq 1, \forall \omega_i$ .

In general a material's reflective properties are anisotropic, meaning that if the surface is rotated around the surface normal, the reflective properties may change. This is true for some materials such as brushed aluminium, cloth and hair [12]. However, it is common to assume, with good approximation, that many surfaces' reflective properties are isotropic, that is the reflection is only dependent on the incident and exitant zenith angle. Additionally, it is common to assume that surfaces are homogeneous, i.e., the reflective properties of the surface does not vary across it, i.e.,  $f(\mathbf{x}, \omega_i, \omega_o) = f(\omega_i, \omega_o)$ .

### 2.8.2 Ideal Diffuse Surfaces

Certain materials, such as paper, are characterized by their diffuse nature, which implies they reflect light evenly across the reflective hemisphere. This means that the material's diffuse reflection appear equally bright from any viewing angle. These are commonly referred to as *Lambertian* surfaces. The

BRDF for Lambertian surfaces are constant and is defined as follows:

$$f_r(\mathbf{x}, \omega_i, \omega_o) = \frac{\rho}{\pi} \quad (2.21)$$

where  $\rho \in [0, 1]$  represents the *reflectance* or *albedo* of the surface [6]. However, no material is purely diffuse or purely specular; most exhibit a combination of both, resulting in what are known as glossy materials. To model the specular characteristics of these materials, we turn to microfacet theory.

### 2.8.3 Microfacet Theory

The common way to model geometry in a scene is to discretize surfaces into triangles. However, if surface features are small, many triangles are needed to capture the geometric details. It is computationally infeasible to discretize microscopic features of a material's roughness. Instead, we can model these microscopic features with a statistical model. The idea of such a statistical model is to model the surface as a collection of reflective *microfacets*. These microfacets, are described by their normal vector, and they are given by a distribution function with respect to the surface's ideal normal  $\mathbf{n}$ . A surface is then smooth if there is a small variation in the microfacet normals relative to the ideal normal, and vice versa.

In the microfacet model by Torrance and Sparrow [11], each microfacet is perfectly specular, thus only those facets with a normal:

$$\omega_h = \frac{\omega_i + \omega_o}{\|\omega_i + \omega_o\|} \quad (2.22)$$

known as *half-angle vector*, contribute to specular reflection due to incoming light in direction  $\omega_i$  towards outgoing direction  $\omega_o$ . The incident flux  $d^2\Phi(\omega_h)$  on microfacets with their normal contained in  $\omega_h$  is given by the definition of radiance eq. (2.4):

$$d^2\Phi(\omega_h) = L_i(\omega_i)(\omega_h \cdot \omega_i) d\omega_i dA(\omega_h).$$

The number of facets per unit surface area with normals contained by solid angle  $\omega_h$  is  $D(\omega_h) d\omega_h$  where  $D(\omega_h)$  is the facet normal distribution. Then, to get the number of facets we simply multiply by the differential surface area, i.e.,  $D(\omega_h) d\omega_h dA$  is the number of facets in  $dA$  with normals contained in

$\omega_h$ , that is  $dA(\omega_h) = D(\omega_h) d\omega_h dA$  which inserted above gives:

$$d^2\Phi(\omega_h) = L_i(\omega_i)D(\omega_h)(\omega_h \cdot \omega_i) d\omega_i d\omega_h dA.$$

The fraction of incident flux to exitant flux is governed by Fresnel's equations and thus we say:

$$d^2\Phi(\omega_o) = F(\omega_i, \omega_h) d^2\Phi(\omega_h)$$

where  $F$  is the Fresnel term. In general, the Fresnel term would depend on the index of refraction, but for brevity we omit it. Additionally, we add a geometrical attenuation term, which models microfacets that might occlude each other, as:

$$\begin{aligned} d^2\Phi(\omega_o) &= F(\omega_i, \omega_h)G(\omega_i, \omega_o) d^2\Phi(\omega_h) \\ \implies L(\omega_o) &= \frac{F(\omega_i, \omega_h)G(\omega_i, \omega_o)L_i(\omega_i)D(\omega_h)(\omega_h \cdot \omega_i) d\omega_i d\omega_h}{\cos \theta_o d\omega_o} \end{aligned}$$

and then the definition for the BRDF eq. (2.5) gives:

$$f_r(\omega_i, \omega_o) = \frac{dL(\omega_o)}{dE(\omega_i)} = \frac{F(\omega_i, \omega_h)G(\omega_i, \omega_o)D(\omega_h)}{\cos \theta_i \cos \theta_o} (\omega_h \cdot \omega_i) \frac{d\omega_h}{d\omega_o}.$$

It can be shown that the Jacobian term is:\*

$$\frac{d\omega_h}{d\omega_o} = \frac{1}{4(\omega_h \cdot \omega_i)} = \frac{1}{4(\omega_h \cdot \omega_o)}$$

which inserted into the previous BRDF equation yields the BRDF model of Torrance and Sparrow:

$$\begin{aligned} f_r(\omega_i, \omega_o) &= \frac{F(\omega_i, \omega_h)G(\omega_i, \omega_o)D(\omega_h)}{4 \cos \theta_i \cos \theta_o} \\ &= \frac{F(\omega_i, \omega_h)G(\omega_i, \omega_o)D(\omega_h)}{4(\mathbf{n} \cdot \omega_i)(\mathbf{n} \cdot \omega_o)}. \end{aligned} \tag{2.23}$$

Note that we omitted any dependence on position in this derivation, but this BRDF may of course vary across the surface.

Depending on the material, different choices of  $F$ ,  $G$  and  $D$  may be suitable. There is rich literature concerning the choice of these, but we will present some common ones that model a wide range of materials with good

---

\*In general the Jacobian term requires absolute value, but since we do not consider refraction it is omitted.

approximation. The reader should note that each surface in the scene may of course have have different parameters, and if necessary even different BRDF model for more exotic materials.

A commonly used model due to Cook and Torrance in [13], for usage in Computer Graphics, uses the microfacet distribution which Beckmann and Spizzichino presented in [10], usually referred to as the Beckmann. His distribution, where  $\cos \theta_h = \mathbf{n} \cdot \boldsymbol{\omega}_h$ , is as follows:

$$D_a(\boldsymbol{\omega}_h) = \frac{e^{-\tan^2 \theta_h (\cos^2 \phi_h / \alpha_x^2 + \sin^2 \phi_h / \alpha_y^2)}}{\pi \alpha_x \alpha_y \cos^4 \theta_h} \quad (2.24)$$

$$D_i(\boldsymbol{\omega}_h) = \frac{e^{-\tan^2 \theta_h / \alpha^2}}{\pi \alpha^2 \cos^4 \theta_h} \quad (2.25)$$

where eq. (2.24) is the anisotropic version, and with  $\alpha = \alpha_x = \alpha_y$  the isotropic version is eq. (2.25).<sup>\*</sup> Here,  $\sigma$  is the root mean square slope of the microfacets such that  $\alpha = \sqrt{2}\sigma$ . The geometrical attenuation factor,  $G$ , depends on the microfacet distribution and geometry of microspheres to ensure conservation of energy. Cook and Torrance used a one-dimensional microfacet geometry model where the facets are parallel grooves, which guarantees energy conservation for an arbitrary microfacet distribution. Their proposed term is:

$$G(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \boldsymbol{\omega}_h) = \min \left\{ 1, \frac{2(\mathbf{n} \cdot \boldsymbol{\omega}_h)(\mathbf{n} \cdot \boldsymbol{\omega}_i)}{\boldsymbol{\omega}_i \cdot \boldsymbol{\omega}_h}, \frac{2(\mathbf{n} \cdot \boldsymbol{\omega}_h)(\mathbf{n} \cdot \boldsymbol{\omega}_o)}{\boldsymbol{\omega}_i \cdot \boldsymbol{\omega}_h} \right\}. \quad (2.26)$$

However, in the more recent work [15] from Walter et al., they recommend against this attenuation term since it contains, inter alia, features not observed in real surfaces. They instead derive approximate attenuation terms for several distributions, based on Smith's work, that are on the form:

$$G(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \boldsymbol{\omega}_h) \approx G_1(\boldsymbol{\omega}_i, \boldsymbol{\omega}_h)G_1(\boldsymbol{\omega}_o, \boldsymbol{\omega}_h). \quad (2.27)$$

Here, the function  $G_1$  depends on the microfacet distribution. In their work, they derive the following attenuation factor for the isotropic Beckmann

---

<sup>\*</sup>In Cook and Torrance's work, they only considered the isotropic distribution, but for completeness we present the anisotropic version found in [14].

distribution:

$$G_1(\boldsymbol{\omega}, \boldsymbol{\omega}_h) = \frac{2}{1 + \operatorname{erf}(a) + \frac{1}{a\sqrt{\pi}}e^{-a^2}} \quad (2.28)$$

$$\approx \begin{cases} \frac{3.535a+2.181a^2}{1+2.276a+2.577a^2} & \text{if } a < 1.6, \\ 1 & \text{otherwise} \end{cases}$$

where  $a = (\alpha \tan(\mathbf{n} \cdot \boldsymbol{\omega}))^{-1}$ . According to Walter et al., the approximation carries a relative error less than 0.35% and is much cheaper to evaluate than the exponential and, especially, the error function.

Another modern microfacet distribution, popularized by Disney and Epic Game's Unreal Engine, is the so called GGX distribution:

$$D_a(\boldsymbol{\omega}_h) = \frac{1}{\pi \alpha_x \alpha_y \cos^4 \theta_h \left(1 + \tan^2 \theta_h \left(\frac{\cos^2 \phi_h}{\alpha_x^2} + \frac{\sin^2 \phi_h}{\alpha_y^2}\right)\right)^2} \quad (2.29)$$

$$D_i(\boldsymbol{\omega}_h) = \frac{\alpha^2}{\pi \cos^4 \theta_h (\alpha^2 + \tan^2 \theta_h)^2} \quad (2.30)$$

where eq. (2.29) is the anisotropic version given by [7], which simplifies to the isotropic version, eq. (2.30), when  $\alpha = \alpha_x = \alpha_y$ . Walter et al. presents the following attenuation term for the isotropic version:

$$G_1(\boldsymbol{\omega}, \boldsymbol{\omega}_h) = \frac{2}{1 + \sqrt{1 + \alpha^2 \tan^2(\arccos(\mathbf{n} \cdot \boldsymbol{\omega}))}} \quad (2.31)$$

$$= \frac{2}{1 + \sqrt{1 + \alpha^2 ((\mathbf{n} \cdot \boldsymbol{\omega})^{-2} - 1)}}.$$

The Fresnel term is independent of the  $D$  and  $G$  term, but care must still be taken since the Fresnel effects are strongly dependent on the conductive properties of the material. In general, the reflectance is described by different terms depending on the polarization of the light. However, assuming light is unpolarized, the Fresnel term is given by:

$$F = \frac{1}{2} (R_s + R_p) \quad (2.32)$$

where  $R_s$  and  $R_p$  are the reflectance of  $s$  and  $p$  polarized light, respectively.

They are defined as:

$$R_s = \left| \frac{Z_i \cos \theta_i - Z_t \cos \theta_t}{Z_i \cos \theta_i + Z_t \cos \theta_t} \right|^2 \approx \left| \frac{n_i \cos \theta_i - n_t \sqrt{1 - \left(\frac{n_i}{n_t} \sin \theta_i\right)^2}}{n_i \cos \theta_i + n_t \sqrt{1 - \left(\frac{n_i}{n_t} \sin \theta_i\right)^2}} \right|^2,$$

$$R_p = \left| \frac{Z_i \cos \theta_t - Z_t \cos \theta_i}{Z_i \cos \theta_t + Z_t \cos \theta_i} \right|^2 \approx \left| \frac{n_i \sqrt{1 - \left(\frac{n_i}{n_t} \sin \theta_i\right)^2} - n_t \cos \theta_i}{n_i \sqrt{1 - \left(\frac{n_i}{n_t} \sin \theta_i\right)^2} + n_t \cos \theta_i} \right|^2.$$

Here,  $Z_i$  and  $Z_t$  represent the wave impedances of the incoming and transmitted (refracted) light waves, respectively, while  $\theta_t$  denotes the angle of refraction. It is often reasonable to ignore the magnetic properties of a medium since, for most materials, the magnetic permeability,  $\mu$ , is close to the permeability constant,  $\mu_0$  [16]. Then, the wave impedance can be effectively characterized by its refractive index, expressed as  $Z = Z_0/n$ , with  $Z_0$  being the impedance of free space and  $n$  the index of refraction. By applying this approximation and utilizing Snell's law to eliminate the angle of refraction, we get the second form of the equations above.

## 2.8.4 Sampling

When we are sampling a direction from a BRDF that is based on a microfacet model, we are actually interested in sampling a microfacet normal and then calculate a direction on the hemisphere using the law of the reflection. To this end, we need a density function that is normalized over the unit hemisphere of the idealized surface where the normal is given. Consider a small patch of area,  $A$  around the surface normal  $\mathbf{n}$ , and the area of microfacets contained within  $A$  with normals within  $d\omega_h$  is  $dA_h = D(\omega_h) d\omega_h A$ . The signed projected area of microfacets, i.e.,  $(\mathbf{n} \cdot \omega_h) dA_h$ , contained within  $A$  onto  $A$  should indeed equal the area of  $A$ , otherwise there would be either overlapping microfacets or gaps between microfacets. Mathematically, we require:

$$A = \int_{A_h} (\mathbf{n} \cdot \omega_h) dA_h = \int_{\Omega} (\mathbf{n} \cdot \omega_h) D(\omega_h) d\omega_h A$$

$$\implies 1 = \int_{\Omega} (\mathbf{n} \cdot \omega_h) D(\omega_h) d\omega_h = \int_{\Omega} D(\omega_h) \cos \theta_h \sin \theta_h d\theta d\phi$$

which indeed is the normalization criteria. It is a straight forward calculation to show that both Beckmann and GGX satisfy the normalization criteria. From this, it is clear that our probability density function, in spherical coordinates, is:

$$p(\theta_h, \phi_h) = D(\theta_h, \phi_h) \cos \theta_h \sin \theta_h.$$

Now we derive a sampling procedure for the Beckmann and GGX distribution; we will only consider isotropic distributions, and we refer the reader to Burley's paper [17] for a discussion of the anisotropic distributions. We proceed using standard inverse transform sampling for multivariate densities.\*

We begin with Beckmann's distribution, eq. (2.25), which gives us the following density function:

$$p(\theta_h, \phi_h) = \frac{\sin \theta_h}{\pi \alpha^2 \cos^3 \theta_h} e^{-\tan^2 \theta_h / \alpha^2}$$

with the marginal density and cumulative distribution:

$$\begin{aligned} p_\theta(\theta_h) &= \int_0^{2\pi} p(\theta_h, \phi_h) d\phi_h = \frac{2 \sin \theta_h}{\alpha^2 \cos^3 \theta_h} e^{-\tan^2 \theta_h / \alpha^2} \\ \Rightarrow P_\theta(\theta_h) &= \int_0^{\theta_h} p_\theta(\theta'_h) d\theta'_h = 1 - e^{-\tan^2 \theta_h / \alpha^2} \end{aligned}$$

and then the conditional density and cumulative distribution for  $\phi_h$  is:

$$\begin{aligned} p_\phi(\phi_h | \theta_h) &= \frac{p(\theta_h, \phi_h)}{p_\theta(\theta_h)} = \frac{1}{2\pi} \\ \Rightarrow P_\phi(\phi_h | \theta_h) &= \int_0^{\phi_h} p(\phi'_h | \theta_h) d\phi'_h = \frac{\phi_h}{2\pi}. \end{aligned}$$

Now, consider two uniform random variables  $\xi_1, \xi_2 \in [0, 1)$ , then by inverse transform sampling, we have that  $P_\theta^{-1}(\xi_1)$  has the same distribution as  $\theta_h$ , and  $P_\phi^{-1}(\xi_2)$  has the same distribution as  $\phi_h | \theta_h$ . Of course, in this case  $\phi_h$  and  $\theta_h$  are independent and we drop the conditionality. Then, simply inverting the

---

\*Even though both density functions we consider are separable we use the general approach of considering marginal and conditional densities to show that it is not necessarily more difficult for other density functions.

cumulative distribution functions gives us the sampling formulas:

$$\theta_h = \arctan \sqrt{-\alpha^2 \ln(1 - \xi_1)}, \quad (2.33)$$

$$\phi_h = 2\pi\xi_2. \quad (2.34)$$

Similarly, for the GGX distribution, eq. (2.30), we get the density function:

$$p(\theta_h, \phi_h) = \frac{\alpha^2 \sin \theta_h}{\pi \cos^3 \theta_h (\alpha^2 + \tan^2 \theta_h)^2}$$

which has the following marginal and cumulative distribution

$$p_\theta(\theta_h) = \int_0^{2\pi} p(\theta_h, \phi_h) d\phi_h = \frac{2\alpha^2 \sin \theta_h}{\cos^3 \theta_h (\alpha^2 + \tan^2 \theta_h)^2}$$

$$\Rightarrow P_\theta(\theta_h) = \int_0^{\theta_h} p_\theta(\theta'_h) d\theta'_h = \frac{\tan^2 \theta_h}{\tan^2 \theta_h + \alpha^2}$$

and for  $\phi_h$  we get the exact same as for Beckmann. Then, by the principle of inversion sampling, the sampling formulas are:

$$\theta_h = \arctan \left( \alpha \sqrt{\frac{\xi_1}{1 - \xi_1}} \right), \quad (2.35)$$

$$\phi_h = 2\pi\xi_2. \quad (2.36)$$

## 2.9 Transceiver Model

In the context of this paper, the light sources are transceivers. A transceiver has some type of emission profile that describes how its emitted light is distributed. Assuming the transceivers are point sources, their emission profiles can be completely described by their *emission function*  $\eta(\theta, \phi)$ . This function describes the radiant intensity (Watt per steradian) emitted by the transceiver in direction  $(\theta, \phi)$ . Of course, one could model a transceiver as having an isotropic emission function, but no such assumption is necessary in the framework of this paper. The only requirement is that  $\eta$  is non-negative and continuous on its domain:  $\theta \in [0, \pi/2]$  and  $\phi \in [0, 2\pi)$ . An example of such a function could be:

$$\eta(\theta, \phi) = \hat{I} \sin 2\theta$$

which is rotationally symmetric around the zenith, with  $\hat{I}$  being the max radiant intensity. This would correspond to a transceiver which sends out most of its light at a 45% degree angle with respect to the ceiling. Furthermore, the directions can be importance sampled by sampling from the density function:

$$p_{\eta}(\theta, \phi) = \begin{cases} \frac{3}{2} \frac{\sin 2\theta}{2\pi} & \text{if } \theta \in [0, \pi/2], \\ 0 & \text{otherwise,} \end{cases}$$

with the corresponding sampling formulas:

$$\begin{aligned} \theta &= \arcsin \sqrt[3]{\xi_1} \\ \phi &= 2\pi\xi_2 \end{aligned}$$

derived using the same principles as in section 2.8.4. Of course, for the best result, the emission function should be constructed from experimental data of real transceivers.

## 2.10 Test Retail Environment

In constructing our experimental setup, we have created a basic retail environment that includes shelving units and a section resembling a freezer. The shelves are fitted with electronic shelf labels (ESLs), which are intentionally larger than their real-life counterparts. This is done to reduce computational costs, as it allows us to consider only a few hundred ESLs instead of several thousand. This is justified by the assumption that ESLs positioned closely together will receive similar illumination levels.

Additionally, all surfaces in this environment exhibit Lambertian reflectance, mimicking a paper-like material with an albedo  $\rho = 0.7$ . This approach is a significant simplification, but it has been deemed necessary due to the considerable effort involved in accurately modeling a typical grocery store and the extensive data required to determine appropriate material properties for each surface. It is important to view this test environment as an initial step in assessing our framework. We emphasize that the methodologies outlined in this paper are applicable to a wide range of geometries and material.

The shelving units, each extending to a height of two meters, are arrayed with three rows of ESLs. Above, potential transceivers are distributed across a grid on the ceiling, which is situated at a height of four meters. A three-dimensional representation of this test environment is depicted in fig. 2.7, while a top-down perspective is provided in fig. 2.8.

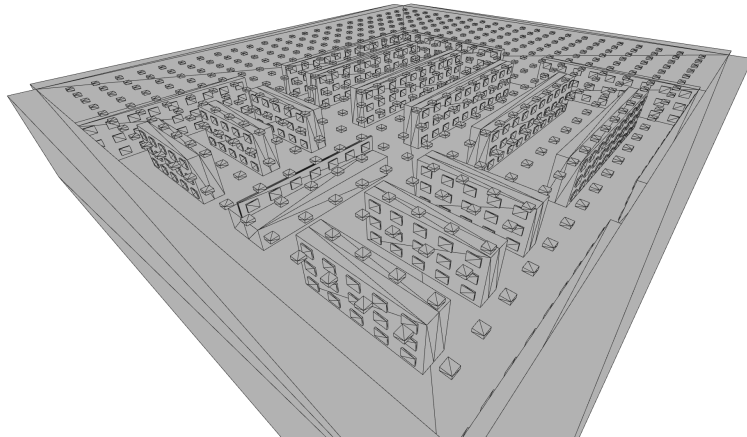


Figure 2.7: 3D view of the test environment.

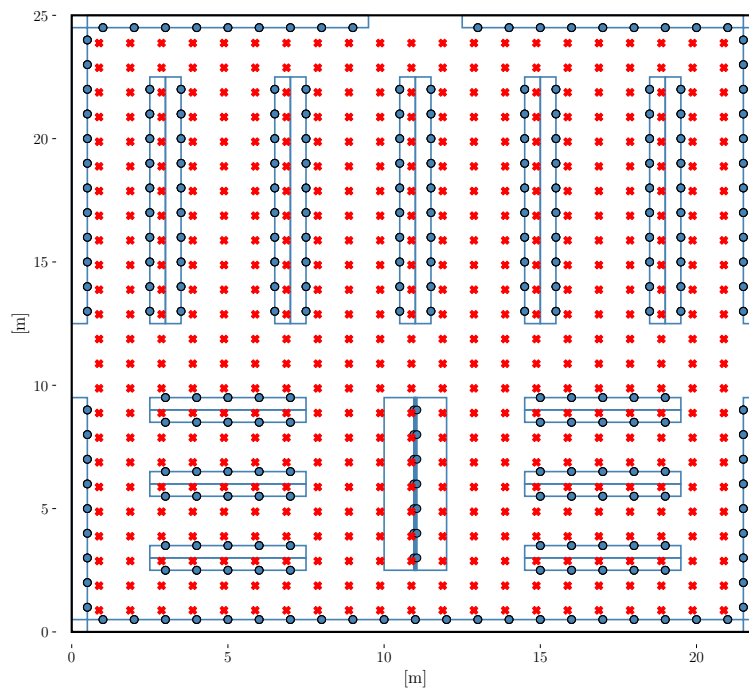


Figure 2.8: Top-down view of the environment where dots (●) are ESLs and crosses (×) are transceivers.

Furthermore, the signal strength of the transceivers has been deliberately reduced to yield a more nuanced light distribution, facilitating a more straightforward comparison between different transceivers. This adjustment is also aimed at simplifying the subsequent evaluation of the optimization

framework, ensuring that the effects of varying transceiver strengths on the distribution are clearly discernible and can be easily assessed.

## 2.11 Results and Discussion

In this section, we present and discuss the simulation results obtained from the retail environment described in section 2.10. The electronic shelf labels (ESLs) within this environment are categorized into three levels, based on their vertical positioning. ESLs positioned on the lowest shelf section constitute level one, those on the middle shelf section and the freezer section are categorized as level two, and ESLs on the top shelf section are categorized as level three.

### 2.11.1 Measured Radiance

Our presentation of simulation results is limited to a selected subset of transceivers. These selections are intended to assess the simulation's validity. Additionally, we refrain from reporting specific numerical values for a few reasons. The transceiver model used in this simulation is only loosely based on real-world counterparts, and the ESL responsivity values are proprietary to Pricer and thus not disclosed in detail; instead we use arbitrary values that bears some resemblance to reality. Consequently, our analysis is qualitative, focusing on relative measurements rather than specific values.

Accordingly, we have depicted the results on a logarithmic relative scale. It is important for readers to interpret the figures with this in mind: ESLs represented by larger and darker dots indicate significantly higher measured radiance. However, we also report a filtered version using an *arbitrary* threshold value to allow for a more intuitive analysis of which ESLs the transceiver might be able to activate.

As previously mentioned, our first sanity check is to verify the expected symmetry, judging by figs. 2.9 to 2.12 we achieve the anticipated symmetry. While the symmetry is not perfectly mirrored, a consequence of relying on statistical methods, this still serves as a strong indication that the simulation is yielding reasonable results. Additionally, it is evident in all figures that ESLs located nearer to the transceiver exhibit significantly stronger measured radiance. Of course, there are some outliers; for instance, in fig. 2.12, there are ESLs, situated far away in the bottom, that show a high radiance value. However, they seem to be within line of sight of the transceivers and should thus not constitute any concern.

The variation in measurements across different levels is also noteworthy. In general, there is no considerable difference across levels, but in certain instances, for example in fig. 2.10 there is some variation. This is to be expected to some degree, given that the transceiver's emission function, as described in section 2.9, is designed to emit most of its radiation at an angle to the ceiling, rather than uniformly.

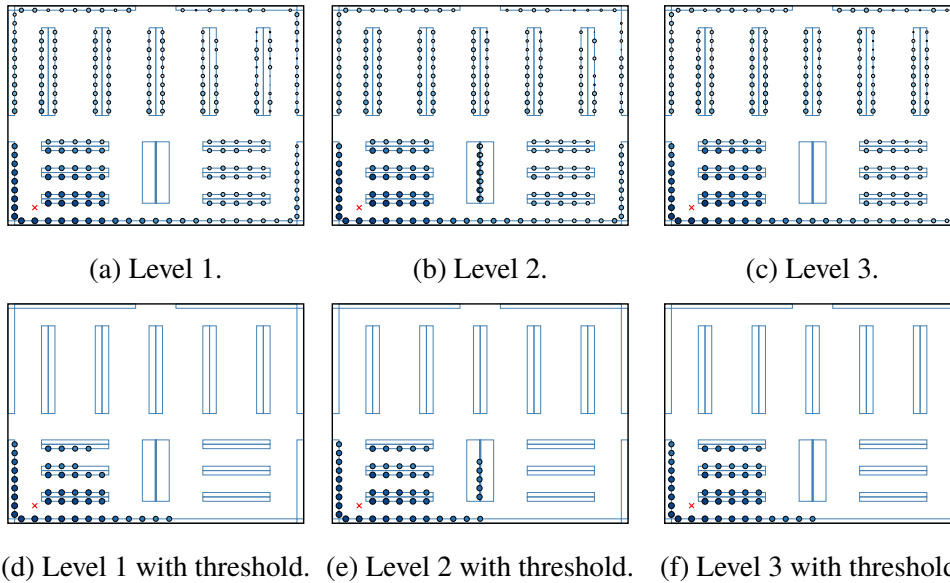


Figure 2.9: Measurements due to transceiver at bottom left.

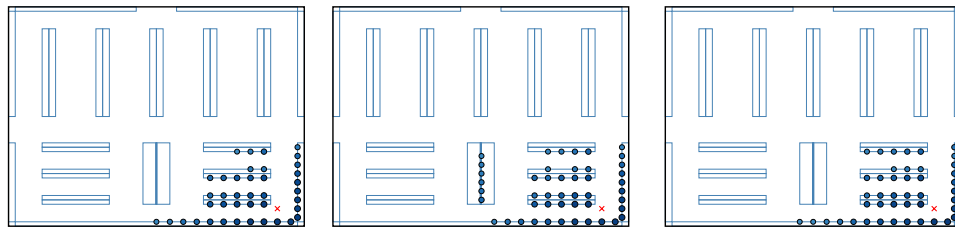
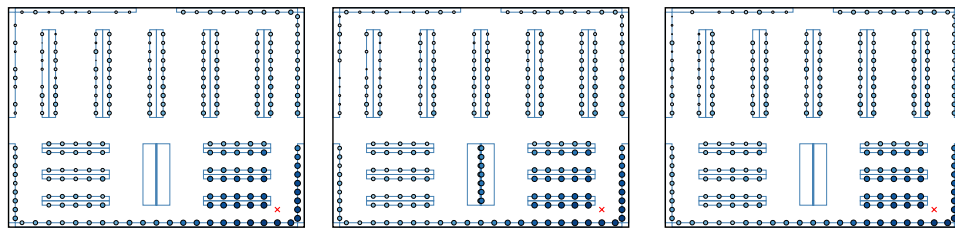


Figure 2.10: Measurements due to transceiver at bottom right.

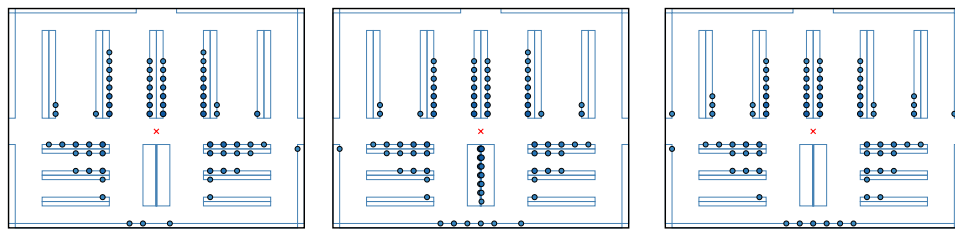
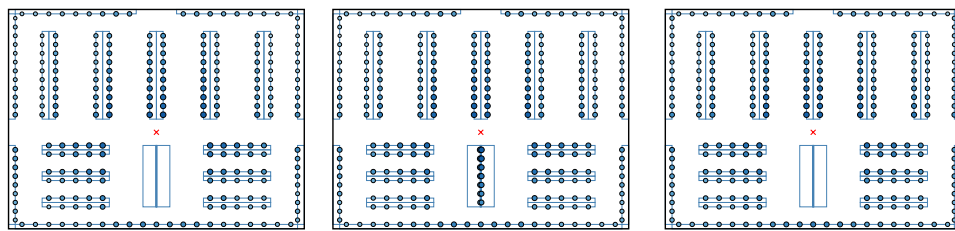


Figure 2.11: Measurements due to transceiver at the center.

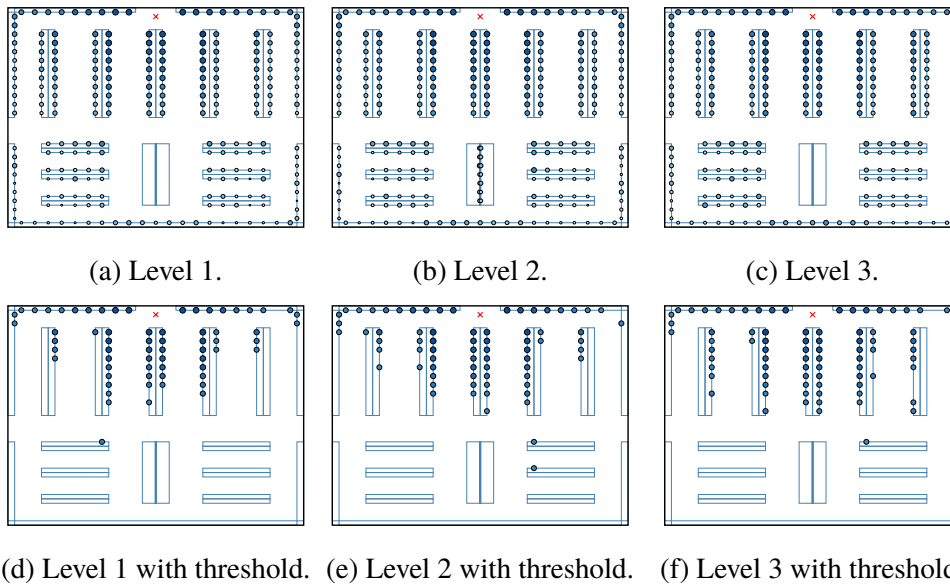


Figure 2.12: Measurements due to transceiver at top-middle.

### 2.11.2 Number of Samples

The sample count significantly influences the outcomes, as depicted in fig. 2.13. This figure demonstrates the distribution of signals from the central transceiver to all ESLs at level two, varying by the number of samples used. The impact of sample size is particularly pronounced when implementing a consistent cut-off threshold across all measurements. Notable discrepancies are seen when comparing results from 5 000 samples, as seen in fig. 2.13d, with those from 25 000 samples, as seen in fig. 2.13e. The frequency of statistical outliers reduces significantly as the sample count increases. However, outliers persist, especially near the bottom edge, but these are entirely absent when the sample count reaches 100 000, as shown in fig. 2.13f. This phenomenon likely stems from the relatively non-obstructed view between these ESLs and the transceiver, leading to increased signal strength accumulation during the initial ray bounces. Nevertheless, with a sufficient number of samples, as we see in fig. 2.13f, these ESL's signal strength eventually falls off.

The observed variance largely results from the uniform random light sampling technique employed. Although uniform random sampling is not inherently bad, the specific context of this study – featuring hundreds of transceiver light sources – exposes its limitations. The large amount of transceivers and the high probability of occlusion, particularly due to tall

shelving, significantly increase the probability that a randomly selected light fails the visibility test, rendering many samples ineffective.

In general, it is quite challenging to determine the convergence of the simulation. In fields like computer graphics, noise in the imagery can visually indicate a lack of convergence. However, in this scenario, such visual indicators are not apparent. One potential approach involves monitoring the variation in signal measurements between each ESL and transceiver pairing and terminating the simulation when the variation becomes sufficiently small. This method is complicated by the fact that the analysis does not only focus on the total received signal but also on the contribution from each light source. Monitoring the contribution changes from each light source poses no challenge, but establishing a good termination criterion for each light source simultaneously is more difficult.

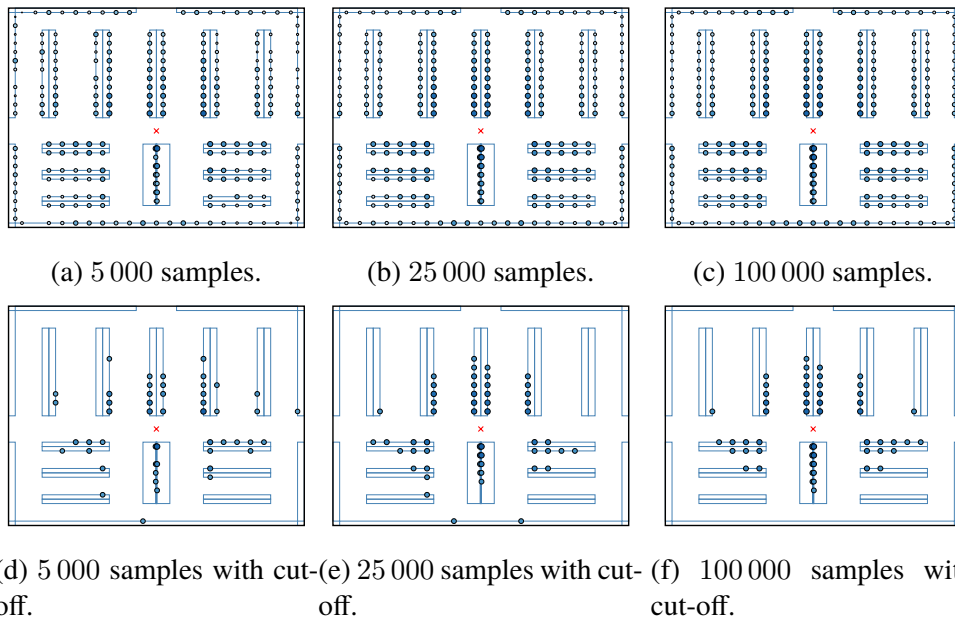


Figure 2.13: Radiance measurements depending on the number of samples.



# Chapter 3

## Optimization

### 3.1 Requirements

This study aims to optimize the number of transceivers necessary to meet the communication demands of a retail environment utilizing electronic shelf labels (ESLs) and transceivers. A key consideration is the mitigation of excessive communication, a phenomenon observed when ESLs experience varying frequencies of communication. This discrepancy leads to inefficiencies, particularly for less active ESLs, which spend energy processing irrelevant messages. For example, as illustrated in fig. 3.1, using a single transceiver for all ESLs results in unnecessary energy consumption: even if only the ESL on the left shelf requires updates, those on the right shelf participate in the communication process (see fig. 3.1a). Conversely, using two transceivers, as shown in fig. 3.1b, enables targeted communication, potentially reducing energy waste by coupling ESLs with specific transceivers. Although the benefit of an additional transceiver may not be immediately apparent in such simple configurations, in more complex scenarios, the reduction in excessive communication can lead to considerable energy savings.

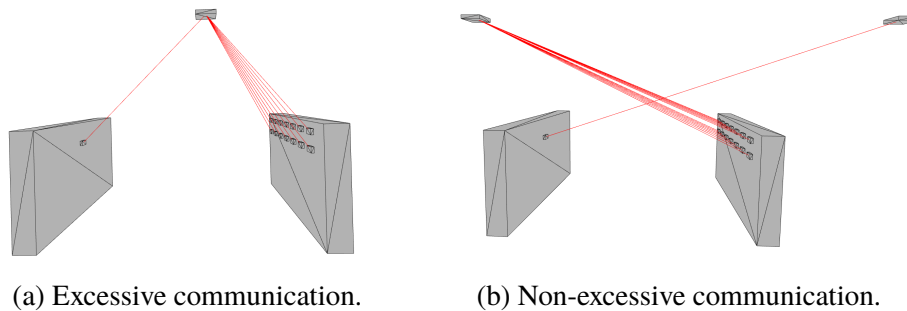


Figure 3.1: Demonstration of excessive communication.

Accordingly, the framework should empower users to incorporate these considerations into their cost analysis, allowing them to weigh the expense of additional transceivers against the savings from reduced excessive communication. There should be no limiting assumptions imposed on these costs, affording users the flexibility to adapt their assessments based on factors such as transceiver models, ESL dimensions and model, and so on.

Moreover, the optimization framework is designed to require minimal specifications regarding input data. It only needs a selection of potential transceivers, a grouping of ESLs, and a matrix (or an equivalent description) that captures the signal strength between each transceiver and ESL. Importantly, the source of the signal measurement matrix is not restricted to the simulation method presented in this thesis; it may come from any source.

We now formally state the optimization problem this framework intends to solve. Consider a set of potential transceivers, denoted  $\mathcal{T}$  and a set of ESLs, denoted  $\mathcal{E}$ . We introduce binary decision variables  $T_j \in \{0, 1\}$  for each  $j \in \mathcal{T}$ , where  $T_j = 1$  if transceiver  $j$  is selected, otherwise  $T_j = 0$ . Additionally, we are given a set of measurements,  $M_{ij}$ , representing the signal strength that ESL  $i$ , where  $i \in \mathcal{E}$ , would receive from transceiver  $j$ . Each ESL has a predefined activation threshold  $\tau_i$  and a safety factor  $\gamma_i$ . The primary goal is to identify the smallest subset of transceivers from  $\mathcal{T}$  such that all ESLs meet their communication requirements.

## 3.2 Communication Models

We will explore multiple approaches to modeling these communication requirements, starting with the scenario where signal cooperation among transceivers is not considered, followed by the scenario that permits cooperative communication.

### 3.2.1 Non-Cooperative Communication

In the non-cooperative model, we assume that the signals from different transceivers do not cumulatively contribute to the activation of an ESL. Under this premise, we define binary variables  $\ell_{ij}$ , as follows:

$$\ell_{ij} = \begin{cases} 1, & \text{if } M_{ij} \geq \gamma_i \tau_i, \\ 0, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{E}, \forall j \in \mathcal{T}. \quad (3.1)$$

This matrix, referred to as the *reach matrix*, indicates whether a transceiver  $j$  can reach ESL  $i$  with its signal. A key requirement is that each ESL can communicate with at least one transceiver. The activation of a single transceiver can potentially satisfy the communication requirements of several ESLs. This negates the need for defining specific ESL-transceiver pairings. In fact, upon its activation, it necessarily communicates with all reachable ESLs. The following constraint ensures that each ESL is able to communicate with at least one transceiver:

$$\sum_{j \in \mathcal{T}} \ell_{ij} T_j \geq 1, \quad \forall i \in \mathcal{E}. \quad (3.2)$$

When formulating the objective function, our aim is twofold: to minimize the total number of transceivers and to reduce redundant communications. The latter is achieved by minimizing the total number of connections between ESLs and transceivers, and thus a reduction in redundant coverage. However, these objectives may sometimes conflict. To effectively quantify the trade-offs involved in activating additional transceivers and establishing more communication links, we introduce costs for active transceivers and for communication links. The cost associated with activating a transceiver, denoted  $c_j^{\mathcal{T}}$  for each  $j \in \mathcal{T}$ , might encompass installation, maintenance, or other related expenses. Similarly, we assign a cost,  $c_i^{\mathcal{E}}$  for each  $i \in \mathcal{E}$ , to the communication links. This cost could reflect the energy requirements for communication with an ESL. The objective function, therefore, evolves into minimizing the following expression:

$$\min \left\{ \sum_{j \in \mathcal{T}} c_j^{\mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} c_i^{\mathcal{E}} \ell_{ij} T_j \right\}. \quad (3.3)$$

For completeness, we write the entire formulation in a compact matrix format; say there are  $m$  ESLs, linearly indexed from 1 to  $m$ , and similarly

for  $n$  transceivers; then let  $A$  be the matrix such that  $A_{ij} = \ell_{ij}$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ , and for the costs  $\mathbf{c}^{\mathcal{T}} = (c_1^{\mathcal{T}}, \dots, c_n^{\mathcal{T}})^{\top}$  and  $\mathbf{c}^{\mathcal{E}} = (c_1^{\mathcal{E}}, \dots, c_m^{\mathcal{E}})^{\top}$ , we can write  $\mathbf{c} = \mathbf{c}^{\mathcal{T}} + A^{\top} \mathbf{c}^{\mathcal{E}}$ . Then, the problem may be stated as

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{Z}_2^n} \quad & \mathbf{c}^{\top} \mathbf{x} \\ \text{s.t.} \quad & A \mathbf{x} \geq \mathbf{1}_{m \times 1} \end{aligned} \quad (3.4)$$

where the solution vector is  $\mathbf{x} = (T_1, T_2, \dots, T_n)^{\top}$ .

### 3.2.2 Cooperative Communication

In contrast to the non-cooperative model of communication, the cooperative communication model allows multiple transceivers to work together, cumulatively contributing to the activation of an ESL. This approach not only mirrors a more realistic scenario but also allows for more flexibility in optimization.

To formalize this, we introduce a cumulative signal strength variable  $S_i$  for each ESL  $i$ , defined as:

$$S_i = \sum_{j \in \mathcal{T}} M_{ij} T_j, \quad \forall i \in \mathcal{E}. \quad (3.5)$$

As with the non-cooperative model, we impose the constraint that the cumulative signal strength at each ESL must surpass both its activation threshold and safety factor:

$$S_i \geq \gamma_i \tau_i, \quad \forall i \in \mathcal{E}. \quad (3.6)$$

Similar to the non-cooperative model, costs for activating each transceiver, denoted  $c_j^{\mathcal{T}}$ , are incorporated. However, the cooperative model allows for a finer granularity of control. Here, we can penalize not just activation but also excessive signal strength,  $\Delta$ , at each ESL:

$$\Delta_i = S_i - \gamma_i \tau_i, \quad \forall i \in \mathcal{E} \quad (3.7)$$

which is guaranteed to be non-negative due to the previous constraint. This strategy aims to distribute signal strength across all ESLs while minimizing surplus, thereby potentially reducing over-coverage. The objective function, therefore, becomes:

$$\min \left\{ \sum_{j \in \mathcal{T}} c_j^{\mathcal{T}} T_j + \sum_{i \in \mathcal{E}} c_i^{\mathcal{E}} \Delta_i \right\}. \quad (3.8)$$

Of course, this description was for clarity, but the problem can be stated much more simply without introducing extra variables. However, extra variables like the ones we have introduced may still be convenient since solvers usually respond with their values and one can easily evaluate the solution. For mathematical purity, we state the simplest form; the only necessary constraints are:

$$\sum_{j \in \mathcal{T}} M_{ij} T_j \geq \gamma_i \tau_i, \quad \forall i \in \mathcal{E} \quad (3.9)$$

with the the following objective function:

$$\min \left\{ \sum_{j \in \mathcal{T}} c_j^{\mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} c_i^{\mathcal{E}} M_{ij} T_j \right\} \quad (3.10)$$

and we get the exact same form of the formulation as in the non-cooperative model as one might expect. One follows the exact same steps to transform it to matrix form, and is thus omitted here.

In the current formulation, we penalize signal strength to reduce a surplus in signal to mitigate over-coverage. Despite this, permitting a certain degree of surplus signal to reach these unintended ESLs can be advantageous. This approach is justified as the received signal may not exceed the ESL's activation threshold, and therefore, does not pose a significant issue. Instead, we can explicitly try to minimize the number of transceivers that try to establish any form of communication with the the ESLs.

To do so, we will need to define additional variables to capture the individual pairings between transceivers and ESLs. Let  $t_{ij} \in \{0, 1\}$  such that  $t_{ij} = 1$  if ESL  $i \in \mathcal{E}$  communicates with transceiver  $j \in \mathcal{T}$ , otherwise  $t_{ij} = 0$ . Of course, as soon as a transceiver is selected, all pairings related to that transceiver are indeed also activated since an ESL cannot selectively receive signal or not. Thus, we require:

$$t_{ij} = T_j \quad \forall i \in \mathcal{E}, \forall j \in \mathcal{T} \quad \text{and} \quad \sum_{j \in \mathcal{T}} M_{ij} t_{ij} \geq \gamma_i \tau_i \quad \forall i \in \mathcal{E} \quad (3.11)$$

and the number of transceivers ESL  $i$  communicates with is simply:

$$q_i = \sum_{j \in \mathcal{J}} t_{ij}. \quad (3.12)$$

With this approach we can now explicitly minimize the number of transceivers each ESL, surely, communicates with by minimizing:

$$\sum_{i \in \mathcal{E}} c_i^{\mathcal{E}} q_i = \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{J}} c_i^{\mathcal{E}} t_{ij}. \quad (3.13)$$

Since this formulation significantly alters the matrix formulation we, for completeness and implementation purposes, state it. We begin by constructing a vector with all variables using the following linear indexing scheme  $\mathbf{x} = (t_{11}, t_{12}, \dots, t_{1n}, t_{21}, t_{22}, \dots, t_{2n}, \dots, t_{mn}, T_1, T_2, \dots, T_n)^\top$ . Then, we may express the constraints  $\sum_{j \in \mathcal{J}} M_{ij} t_{ij} \geq \gamma_i \tau_i$ , for all  $i \in \mathcal{E}$ , as:

$$(((I_m \ \mathbf{0}_{m,1}) \otimes \mathbf{1}_{1,n}) \odot (\mathbf{1}_{1,m+1} \otimes M)) \mathbf{x} \geq \boldsymbol{\gamma} \odot \boldsymbol{\tau}$$

where  $M = M_{ij}$ ,  $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)^\top$  and  $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)^\top$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . Also, here  $\otimes$  and  $\odot$  denotes the Kronecker and Hadamard product, respectively. Similarly, for the other constraints,  $t_{ij} = T_j$  for all  $i \in \mathcal{E}$  and  $j \in \mathcal{J}$ , we have:

$$((-I_m \ \mathbf{1}_{m,1}) \otimes I_n) \mathbf{x} = \mathbf{0}_{mn,1}.$$

Then, as before, we let  $\mathbf{c}^{\mathcal{J}} = (c_1^{\mathcal{J}}, \dots, c_n^{\mathcal{J}})^\top$  and  $\mathbf{c}^{\mathcal{E}} = (c_1^{\mathcal{E}}, \dots, c_m^{\mathcal{E}})^\top$  as well as introducing a new signal cost  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_m)^\top$  for flexibility; then we have:

$$\begin{aligned} A &= ((I_m \ \mathbf{0}_{m,1}) \otimes \mathbf{1}_{1,n}) \odot (\mathbf{1}_{1,m+1} \otimes M) \\ A_{\text{eq}} &= (-I_m \ \mathbf{1}_{m,1}) \otimes I_n \\ \mathbf{c} &= \begin{pmatrix} \mathbf{c}^{\mathcal{E}} \otimes \mathbf{1}_{n,1} \\ \mathbf{c}^{\mathcal{J}} + M^\top \boldsymbol{\delta} \end{pmatrix} \end{aligned}$$

and the problem is conveniently written as:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{Z}_2^n} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & A \mathbf{x} \geq \boldsymbol{\gamma} \odot \boldsymbol{\tau} \\ & A_{\text{eq}} \mathbf{x} = \mathbf{0}_{mn,1} \end{aligned} \quad (3.14)$$

### 3.3 Solvers and Heuristics

The optimization formulations considered thus far bear a strong resemblance to a classic cover problem, specifically a set cover problem. In particular, the non-cooperative formulation, as represented in eq. (3.4), aligns precisely with a weighted set cover problem. It may be tempting to explore the *exact* cover problem, which seeks a cover where each element is included in precisely one covering set. Such an approach would ideally minimize the issue of over-coverage. However, this represents an excessively rigid constraint and is unlikely to consistently yield feasible solutions, at least in the context of our problem.

Both the set cover (non-cooperative) problem and the cooperative communication problem can be solved using any standard integer linear programming (ILP) solver. Unfortunately, these problems fall into the category of NP-complete problems, making them challenging to solve efficiently. Consequently, the use of heuristics or approximation methods might be necessary, either to identify a feasible solution or to obtain a satisfactory solution.

There exist several algorithms capable of finding solutions to the set cover problem in polynomial time – of course not necessarily optimal solutions. One such algorithm, due to [18], is a greedy algorithm. In the context of our study, this algorithm selects the transceiver that offers the lowest cost relative to the number of ESLs it covers, continuing until all ESLs are accounted for. Despite its simplicity, this method achieves an approximation ratio  $H(m)$ , where  $H(m)$  denotes the  $m^{\text{th}}$  harmonic number and  $m$  the total elements to be covered, i.e., the ESLs. In fact, one cannot find a better approximation ratio unless NP permits superpolynomial time algorithms [19].

We give a brief description of the greedy algorithm. Let  $\mathcal{E}$  be set of all ESLs and  $S_j$  the set of all ESLs transceiver  $j$  can cover (the row indices of non-zero entries in column  $j$  of matrix  $L$ ), and  $c(S_j)$  the cost, or weight, of selecting set  $S_j$ . A formal description is given in algorithm 1. This greedy algorithm, of course, is only applicable to the non-cooperative model, where we indeed have the set cover problem. In cases involving cooperative communication, the problem becomes more complex and we must instead rely on general ILP solvers. For those ILP problems that are too difficult – where even the state-of-the-art ILP solvers struggle – one must turn to approximation or heuristics. However, investigating the wide array of possible heuristics, or meta-heuristics like [20], or approximation algorithms that one may apply to this problem is outside the scope of this paper. Instead, we are more interested

---

**Algorithm 1** Greedy algorithm for weighted set cover

---

```

 $\mathcal{S} \leftarrow \{S_1, S_2, \dots, S_n\}$ 
 $T \leftarrow \emptyset$ 
while  $\mathcal{E} \neq \emptyset$  do
    Choose  $S \in \mathcal{S}$  that minimizes  $c(S)/|S \cap \mathcal{E}|$ 
     $\mathcal{E} \leftarrow \mathcal{E} \setminus S$ 
     $T \leftarrow T \cup S$ 
end while

```

---

in the appropriate way to model the problem at hand, and thus we will rely on general ILP solvers and only evaluate the simple greedy algorithm described in algorithm 1.

### 3.4 Test Strategy

We will evaluate the various formulations discussed using the retail environment outlined in section 2.10. Additionally, we will test the optimization with light simulations conducted with varying sample sizes to determine if the quality of these simulations significantly influences the optimization results.

Furthermore, this study does not investigate the costs associated with transceivers, communication, or signals. Instead, we operate under the presumption that fewer transceivers are always preferable. Therefore, we initially identify the smallest number of necessary transceivers for each setup. Subsequently, we incorporate communication costs in a way that ensures the optimal solution does not exceed the number of transceivers identified when solely minimizing their quantity.

## 3.5 Results and Discussion

### 3.5.1 Non-Cooperative Communication

**Optimal Solutions** In table 3.1, we present the necessary number of transceivers based on the objective function. Additionally, we show the percentage of ESLs communicating redundantly with transceivers. It is important to note that the redundancy here is calculated under the assumption that transceivers *cannot* cumulatively contribute to the activation of the ESLs. Therefore, it requires some care to evaluate them when compared

to how redundant the communication is when cooperative communication is employed. Such a comparison is provided in section 3.5.3. The placement of transceivers for both formulations with 100 000 samples is illustrated in fig. 3.2.

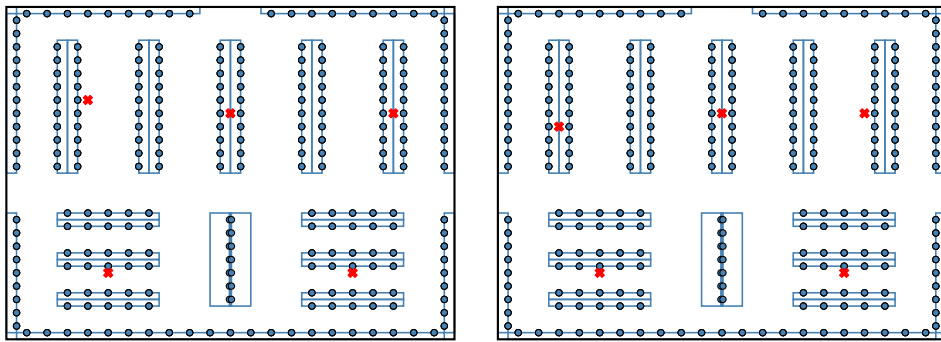
It is apparent that the number of samples substantially influences the required number of transceivers. Specifically, an increase in samples corresponds to a decrease in the number of transceivers needed. This difference is discussed in detail in section 3.5.4, thus the discussion here primarily focuses on results obtained using 100 000 samples, unless otherwise noted.

Furthermore, integrating the additional link-term into the objective function effectively minimizes redundant connections between ESLs and transceivers. The influence of this modification on transceiver positioning is relatively minor, as illustrated in fig. 3.2b, where the locations of the transceivers have shifted marginally from those in fig. 3.2a. This modest adjustment in placement aligns with expectations, considering that minimizing the number of transceivers tends to naturally distribute them more uniformly across the retail environment, thereby already, somewhat, reducing redundant communication.

Additionally, the extra computational effort required to resolve cases with the incorporated link-term is minimal, especially when the sample size is sufficiently large (refer to section 3.5.4 for more details). This observation suggests that incorporating the minimization of additional communication typically does not impose a significant computational burden, making this formulation a preferable choice.

Table 3.1: Optimal transceiver count using non-cooperative communication.

Objective	Transceivers		Transceivers & Links	
	$\sum_{j \in \mathcal{T}} T_j$		$\sum_{j \in \mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} \ell_{ij} T_j$	
Samples	Transceivers	Redundant	Transceivers	Redundant
5 000	8	54%	8	41%
25 000	6	21%	6	17%
50 000	6	27%	6	9%
100 000	5	2%	5	1%



(a) Minimizing only transceivers. (b) Minimizing transceivers and links.

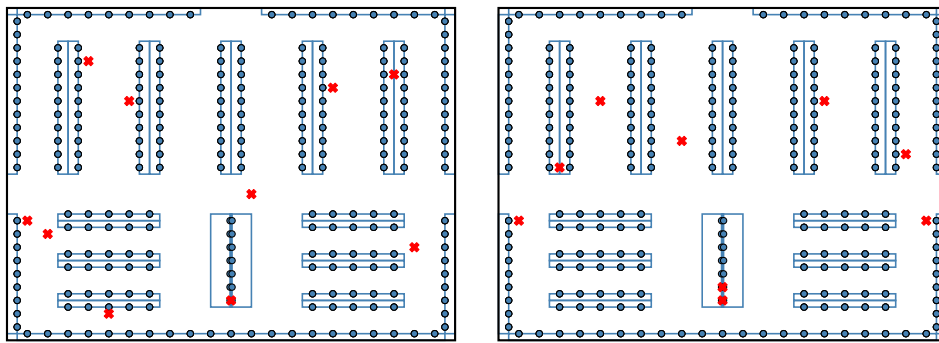
Figure 3.2: Optimal transceiver placement using 100 000 samples.

**Heuristic Solutions** The performance of the greedy algorithm is generally quite poor, with a slight improvement observed when increasing the number of simulation samples, as seen in table 3.2. The primary issue with the greedy approach is its tendency to add additional transceiver placements in the latter stages of selection. This is due to the initial greedy choices leaving small gaps in coverage. Evidence of this issue is the frequent clustering of transceivers in the solutions generated by the algorithm, as depicted in fig. 3.3.

Moreover, the algorithm can be counter-intuitive to use: incorporating an additional link-term in the objective function can result in *increased* redundant communication. It seems that making locally optimal greedy choices is ill-suited for scenarios like ours, where the transceivers may be geometrically close to each other. One potential workaround could be to limit the number of potential transceivers, thereby reducing the likelihood of clustering due to increased separation between the transceivers. Yet, reducing the number of potential transceivers makes the use of general ILP solvers more appealing, as they tend to already perform well with accurate simulations. Consequently, we find little to no advantage in employing the greedy algorithm since the marginal benefits, when compared to ILP solvers, are minimal.

Table 3.2: Transceiver count using non-cooperative communication obtained by the greedy algorithm.

Objective	Transceivers		Transceivers & Links	
	$\sum_{j \in \mathcal{T}} T_j$		$\sum_{j \in \mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} \ell_{ij} T_j$	
Samples	Transceivers	Redundant	Transceivers	Redundant
5 000	10	77%	10	71%
25 000	9	72%	9	72%
50 000	9	70%	8	58%
100 000	9	59%	9	60%



(a) Clustering near the middle-left. (b) Clustering near the bottom-middle.

Figure 3.3: Clustering of transceivers when using the greedy algorithm.

### 3.5.2 Cooperative Communication

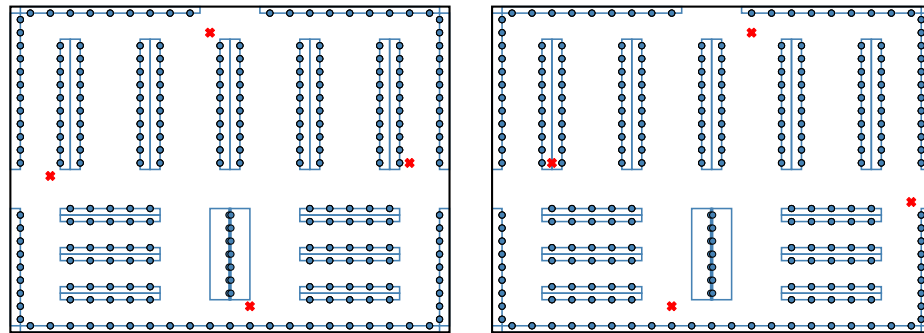
In table 3.3, we present the optimal transceiver count necessary based on the formulation of the objective function, and the number of samples in the light simulation. It is not immediately obvious how the redundancy should be evaluated when employing cooperative communication, and care should be taken when comparing to non-cooperative communication. For a comparison of the approaches, please refer to section 3.5.3. The transceiver placement for all considered objective functions and different number of simulation samples is shown in fig. 3.4. The optimal solution for minimizing both transceiver count and signal aligns precisely with the optimal solution for minimizing transceiver count and links. Consequently, they correspond to the same placement, given by fig. 3.4b.

The number of samples significantly affects the required transceiver count: as the number of samples increases, fewer transceivers are needed. If one were to perform a similar redundancy calculation as done in table 3.1, for instance, defining redundancy for each ESL to be whether the removal of any transceiver would hinder communication with that ESL, then almost 100% of the ESLs would, individually, allow the removal of a transceiver. This high level of redundancy is due to the compact environment and the dense arrangement of the ESLs, where nearly every potential transceiver contributes to the measurement of each ESL. Consequently, reducing redundancy would involve spacing the transceivers as far apart as possible, ideally situating them in each corner and adding an extra one in the center. However, this configuration would call for an extra transceiver (the center one), going against the working assumption that keeping the transceiver count as low as possible is always beneficial.

Differences between the objective functions are marginal in this context; only slight variations in transceiver placement arise from different objective formulations. This minimal variation is due to the uniform costs of ESLs. However, in a more complex setting that incorporates different ESL models with varying weights, the outcomes could substantially differ. In such scenarios, the more detailed formulation that considers each link might be advantageous, otherwise, the model focusing solely on the count of transceivers might be more appropriate since it is simpler and computationally cheaper.

Table 3.3: Optimal transceiver count using cooperative communication.

Objective	$\sum_{j \in \mathcal{T}} T_j$	$\sum_{j \in \mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} M_{ij} T_j$	$\sum_{j \in \mathcal{T}} T_j + \sum_{i \in \mathcal{E}} \sum_{j \in \mathcal{T}} t_{ij}$
Samples	Transceivers	Transceivers	Transceivers
5 000	6	6	6
25 000	5	5	5
50 000	4	4	4
100 000	4	4	4



(a) Minimizing only the number of transceivers. (b) Minimizing transceivers & signal, and transceivers & links.

Figure 3.4: Optimal transceiver placement using 100 000 samples.

### 3.5.3 Comparing Non-Cooperative and Cooperative Communication

Here, the goal is to gauge the efficacy of the non-cooperative solution as an alternative to the cooperative solution, particularly its viability given its lower computational cost (for further details, refer to section 3.5.4).

Unsurprisingly, the number of transceivers required decreases when transceivers are allowed to communicate cooperatively. Although the redundancy numbers, as presented in table 3.1 using 100 000 samples, are impressive, we must remember that these numbers are calculated under the assumption of non-cooperative communication. Hence, they should not be directly compared with redundancy values from the cooperative model, which would all be close to 100%. For a more fair evaluation of these two approaches, and to discern any advantages, beyond computational, that the non-cooperative model might offer, we focus on the excess signal, namely, the signal exceeding the activation threshold. We report values solely for scenarios minimizing both transceivers and links within each communication model. Moreover, considering the non-cooperative solution requires five transceivers instead of four, we compare it with both the four-transceiver solution (100 000 samples) and the five-transceiver solution (25 000 samples) from the cooperative model, ensuring a fair assessment. The excess signal data is illustrated in fig. 3.5.

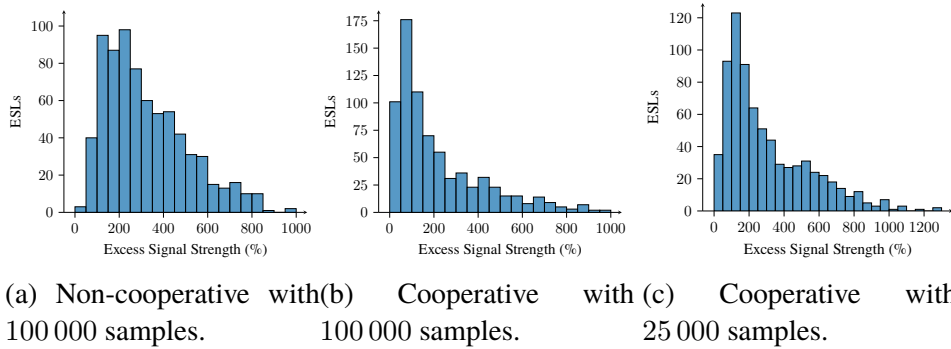


Figure 3.5: Excess signal strength above activation threshold.

When comparing fig. 3.5a and fig. 3.5b, we see that most ESLs in the cooperative solution receive less than 200% excess signal, while a significant portion of ESLs in the non-cooperative solution receive an excess signal ranging between 400 to 500%. Even when the transceiver count is the same – comparing fig. 3.5a and fig. 3.5c – the cooperative solution still outperforms the non-cooperative one. The distribution of signal across all ESLs is significantly better in all scenarios involving cooperative communication.

These insights suggest the non-cooperative model should be a fallback option, primarily when the cooperative model’s computational demands are prohibitive. However, as section 3.5.4 discusses, with proper data preprocessing and high-quality simulations, the cooperative model is typically manageable, making it the preferred choice due to its advantages in transceiver count and signal redundancy, except in the most extreme scenarios.

### 3.5.4 Sensitivity

The initial and most significant observation from the results is the strong influence of the light simulation’s quality, i.e., the number of samples it uses. The optimization’s performance exhibits considerable sensitivity to the quality of the simulation. In instances where the light simulation is not fully converged, resulting in an uneven light distribution, the optimization tends to allocate an excessive number of transceivers to ostensibly fill coverage gaps. However, these perceived gaps are artifacts of an insufficiently converged simulation rather than true gaps in coverage. Our testing indicates that beyond 100 000 samples, there are no notable improvements in the quality of the simulation and optimization, hence results for higher sample counts are not provided.

This sensitivity also significantly impacts the computational load on the

ILP solver. The primary factor is not the number of variables, but the smoothness of the signal distribution within the environment. The presence of statistical outliers and roughness resulting from a poor simulation are the predominant factors increasing computational effort. This is perhaps related to the sparsity of the resulting measurement matrix,  $M_{ij}$ ; the impact of statistical outliers on sparsity becomes quite apparent when visualizing the sparsity pattern of  $M_{ij}$ , particularly after applying a threshold cut-off, as illustrated in fig. 3.6.

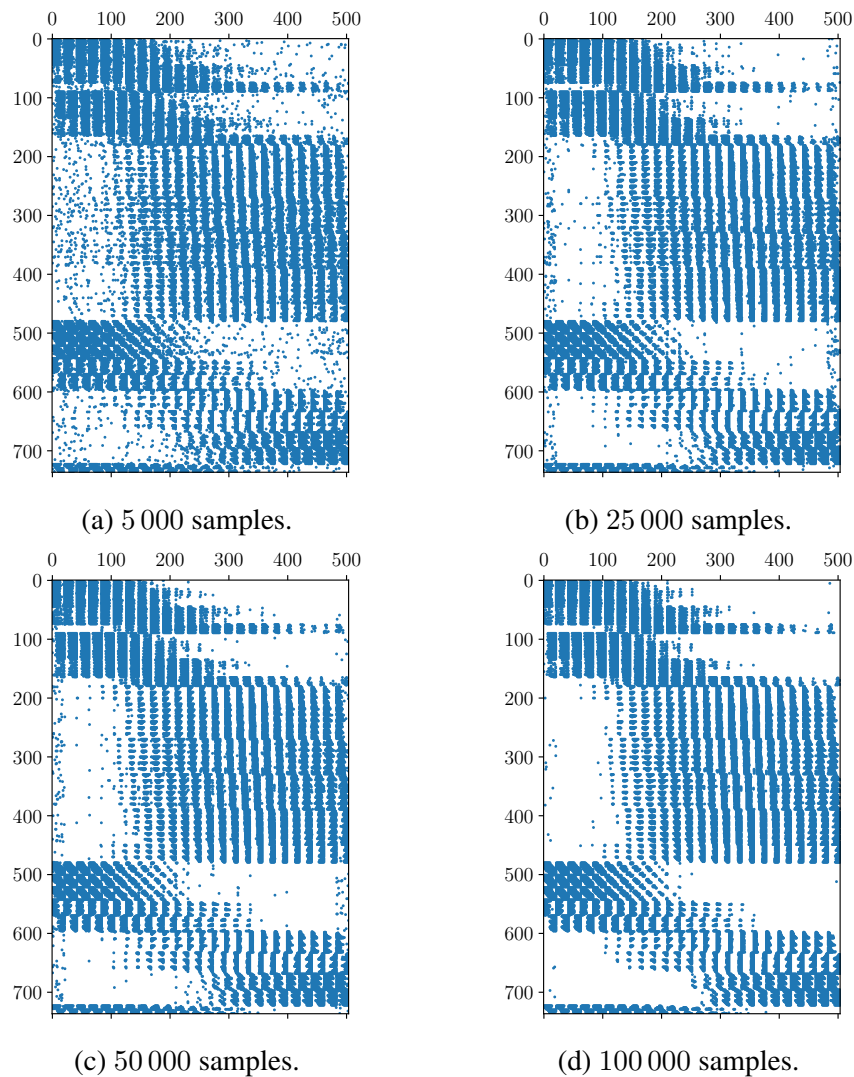


Figure 3.6: Sparsity of the measurement matrix with a threshold cut-off applied.

With a precise and well-converged simulation, the solver completes its task within seconds. In contrast, with a less refined simulation, such as one with only 5 000 samples, the solver’s processing time can extend to one or two days on an average desktop computer. This effect is magnified when considering cooperative communication scenarios, which are more challenging to solve optimally.

While it may not be surprising that a poor simulation heavily skews optimization outcomes, determining the convergence status of a simulation, as discussed in section 2.11.2, can be challenging. Consequently, we emphasize the importance of a high-quality light simulation; considering that even the most accurate light simulations are exponentially faster than the most computationally demanding ILP scenarios, it is advisable to initially pay the cost for a high-quality simulation.

Regardless of simulation quality, in the cooperative model the system tends to not be particularly sparse, especially compared to the typically quite sparse system that arises in the non-cooperative model as we saw in fig. 3.6. Most ESLs receive a small, but non-zero, signal from nearly every transceiver. Although they do not contribute much, they still need to be considered by the solver. However, it is reasonable to assume that such minuscule magnitudes have a negligible impact, particularly because the number of activated transceivers is relatively low compared to the total number of ESLs; the signal strength linked to activated transceivers are significantly higher in magnitude than these noise-like signal measurements. Figure 3.7 displays a typical distribution of signal strengths measured across all transceiver and ESL pairs.

It is evident that the majority of measurements for the ESLs are from transceivers providing low signal contributions. This is not surprising, considering most transceivers and ESLs are both quite far away and occluded from each other. Given this observation, it might be possible to do some initial data preprocessing. In fact, substantial computational gains can be achieved by disregarding all values below a certain threshold; that is, setting  $M_{ij} = 0$  for all  $i \in \mathcal{E}$  and  $j \in \mathcal{T}$  with values below a minimal threshold. By excluding these negligible values, the measurement matrix  $M_{ij}$  transitions from a highly dense matrix, with 0 to 7 percent zeros, to quite a sparse matrix, with 40 to 50 percent zeros, depending on the sample size. This approach significantly reduces the computational load when using the cooperative communication model. In our testing, the solver’s efficiency improved up to 15-fold while still achieving optimal solutions.

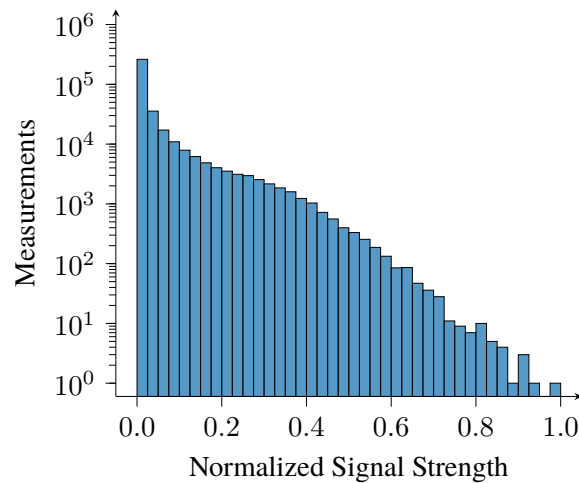


Figure 3.7: Histogram of signal strength samples.

### 3.5.5 Symmetry

Our test retail environment demonstrates strong symmetry; typically, identifying optimal transceiver placement in one half of the store effectively resolves the configuration for the other half. This near symmetric solution is observable in fig. 3.2, albeit not perfectly. The slight deviations from perfect symmetry likely arise from variance in the simulation or the possibility that the observed solution is as optimal as a completely symmetric solution. However, the solver, may not recognize this geometric symmetry, resulting in an unnecessarily large search space.

Additionally, in general, it is quite possible that some transceivers may cover identical sets of ESLs. This scenario introduces a complex form of symmetry for the solver: such interchangeable transceivers can be permuted freely without altering the overall solution, potentially leading to an inefficiently pruned search space by the ILP solver. The full range of strategies that state-of-the-art solvers employ to identify or mitigate such symmetries are typically undisclosed, making it challenging to understand and evaluate the ILP solver's performance.

To mitigate this symmetry challenge, one approach could be to minimize the number of transceivers that are close to each other. While this strategy might risk overlooking more optimal transceiver placements, it is predicated on the assumption that closely positioned transceivers (as in our simulations, approximately one meter apart) generally communicate with the same, or similar, set of ESLs. This assumption minimizes the potential drawbacks of

the approach. Moreover, inaccuracies stemming from imperfect modeling of the store's physical characteristics, such as materials and not accounting for the dynamic nature of grocery items on shelves, are likely more significant. Thus it might be justified considering a less granular grid for potential transceiver placements.

A different approach involves considering orbital branching, a strategy specifically designed to handle symmetry in optimization problems. Orbital branching identifies and groups symmetrical elements, then partitions the search space based on these symmetries, effectively reducing redundant computational effort [21]. Many commercial solvers, including Gurobi and CPLEX, at least implement variations of orbital branching along with additional techniques for symmetry detection and breaking. If one is constrained to working with open-source solvers, it is possible to implement orbital branching through the use of callback functions within the solver, although it may require significant effort.

Nevertheless, one must test on a more extensive set of store layouts, particularly those with varying degrees of symmetry, to validate these claims and evaluate the proposed approaches to mitigate these challenges.

## Chapter 4

# Conclusions and Future Work

### 4.1 Conclusions

This thesis set out to optimize the placement of transceivers in retail environments to enhance the efficiency and reliability of electronic shelf label (ESL) systems. The work was divided into two main components: light simulation and optimization. The light simulation component provided a detailed analysis of infrared (IR) light propagation in a retail store setting, utilizing a path tracing methodology. The optimization component focused on minimizing the number of transceivers required while maintaining robust communication with all ESLs.

Our findings demonstrate that a systematic approach to transceiver placement can significantly reduce the number of transceivers needed, thereby lowering costs and simplifying the deployment process. By using a digital 3D model of a retail store, we were able to simulate various store configurations and product densities. The results showed that both non-cooperative and cooperative communication models have their merits, with cooperative models offering better coverage in complex environments at the cost of increased computational effort.

An important finding is the optimization's sensitivity to signal measurements, which affects both computational efficiency and optimality. High-quality simulations can significantly reduce computational time, highlighting the importance of accuracy in the initial stages. This trade-off is crucial, as poor simulation quality can exponentially increase computational effort.

The investigation reveals that the outcomes of the simulation are reasonable, yet a significant challenge lies in determining when a simulation has converged. This is crucial because statistical outliers can significantly

impact optimization. To address this, we suggest adopting a more rigorous evaluation method by constructing a real-life counterpart to the digital model. This approach not only assesses the model's accuracy but also identifies viable approximations to enhance performance.

The cooperative communication model emerges as the preferable choice due to its realistic and practical solutions, despite being marginally more computationally intensive. Proper preprocessing of signal measurements can enhance its performance to match other models without significant drawbacks.

## 4.2 Future Work

While this thesis has laid a solid foundation for optimizing transceiver placement, several areas remain for future research. One significant area is incorporating dynamic elements into the simulation to account for changes such as the movement of people and restocking of products, which alter the communication landscape.

Additionally, exploring more sophisticated light sampling strategies is necessary. The current method of uniform sampling falls short in environments like retail stores, where occlusions from tall shelves are common. A strategy that samples visible and nearby transceivers, considering the angle to account for the emission profile, could greatly reduce occlusions and improve simulation efficiency.

Future simulations should also incorporate refraction to account for the presence of transparent materials in many stores, such as glass jars, shelves, and doors. This would require adjustments in the Bidirectional Reflectance Distribution Function (BRDF) and surface interaction handling.

Developing a comprehensive cost model is another crucial area for future research. This model should include factors such as energy consumption, transceiver costs, electronic shelf label models, and maintenance costs. Such a model would provide a more holistic optimization strategy and refine the objective function to better differentiate between configurations with equal transceiver counts.

Conducting real-world experiments is essential for validating the simulation and optimization results. Field testing can reveal practical challenges and provide data to refine models further.

In conclusion, this thesis has demonstrated the potential for systematic optimization of transceiver placement in ESL systems. By continuing to refine and expand upon this work, we can develop even more efficient and reliable communication systems for the retail industry.

## References

- [1] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, “Modeling the interaction of light between diffuse surfaces,” *SIGGRAPH Computer Graphics*, vol. 18, no. 3, pp. 213–222, Jan. 1984, ISSN: 0097-8930.
- [2] D. S. Immel, M. F. Cohen, and D. P. Greenberg, “A radiosity method for non-diffuse environments,” *SIGGRAPH Computer Graphics*, vol. 20, no. 4, pp. 133–142, Aug. 1986, ISSN: 0097-8930.
- [3] J. R. Wallace, M. F. Cohen, and D. P. Greenberg, “A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods,” *SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 311–320, Aug. 1987.
- [4] H. Rushmeier and K. Torrance, “Extending the radiosity method to include specularly reflecting and translucent materials,” *ACM Transactions on Graphics*, vol. 9, pp. 1–27, Jan. 1990.
- [5] J. T. Kajiya, “The rendering equation,” *Computer graphics (New York, N.Y.)*, vol. 20, no. 4, pp. 143–150, 1986, ISSN: 0097-8930.
- [6] P. Dutré, K. Bala, and P. Bekaert, *Advanced Global Illumination*, 2nd ed. A K Peters, 2006, ISBN: 9781568813073.
- [7] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 4rd. Cambridge, MA, USA: The MIT Press, 2023, ISBN: 9780262048026.
- [8] F. E. Nicodemus, *Self-Study Manual on Optical Radiation Measurements: Part I—Concepts, Chapters 4 and 5* (NBS TN 910-2). National Institute of Standards and Technology, Gaithersburg, MD, 1978.
- [9] E. Veach, “Robust monte carlo methods for light transport simulation,” Ph.D. dissertation, Stanford, CA, USA, 1997.

- [10] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*. New York, USA: Pergamon Press, 1963.
- [11] K. E. Torrance and E. M. Sparrow, “Theory for off-specular reflection from roughened surfaces\*,” *Journal of the Optical Society of America*, vol. 57, no. 9, pp. 1105–1114, Sep. 1967.
- [12] M. F. Cohen and J. R. Wallace, *Radiosity and Realistic Image Synthesis*. Cambridge, Massachusetts, USA: Academic Press Professional, 1993.
- [13] R. L. Cook and K. E. Torrance, “A reflectance model for computer graphics,” *ACM Transactions on Graphics*, vol. 1, no. 1, pp. 7–24, Jan. 1982, ISSN: 0730-0301.
- [14] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 3rd. Cambridge, MA, USA: Morgan Kaufmann Publishers, 2016, ISBN: 9780128006450.
- [15] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, “Microfacet models for refraction through rough surfaces,” in *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, ser. EGSR’07, Grenoble, France: Eurographics Association, 2007, pp. 195–206, ISBN: 9783905673524.
- [16] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. Cambridge University Press, 2017, ISBN: 9781108420419.
- [17] B. Burley, “Physically based shading at disney,” in *Practical Physically Based Shading in Film and Game Production*, Course Notes, SIGGRAPH, 2012.
- [18] V. Chvatal, “A greedy heuristic for the set-covering problem,” *Mathematics of Operational Research*, vol. 4, no. 3, pp. 233–235, Aug. 1979.
- [19] U. Feige, “A threshold of  $\ln n$  for approximating set cover,” *Journal of the ACM*, vol. 45, no. 4, pp. 634–652, Jul. 1998, ISSN: 0004-5411.
- [20] G. Lan, G. W. DePuy, and G. E. Whitehouse, “An effective and simple heuristic for the set covering problem,” *European Journal of Operational Research*, vol. 176, no. 3, pp. 1387–1403, 2007, ISSN: 0377-2217.
- [21] J. Ostrowski, J. Linderoth, F. Rossi, and S. Smriglio, “Orbital branching,” in *Integer Programming and Combinatorial Optimization*, M. Fischetti and D. P. Williamson, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 104–118, ISBN: 978-3-540-72792-7.



TRITA-SCI-GRU 2024-468  
Stockholm, Sweden 2024

[www.kth.se](http://www.kth.se)