

THE RESEARCH LEDGER

Engineering Intelligent Hybrid VLC/RF Networks

A Graduate Introduction to Modelling,
Controlling, and Governing
Optical–Wireless Systems

NGO TRUNG KIEN

Contents

Preface	vii
Graphical Abstract	xi
Notation and Abbreviations	xxvii
I Foundations of the Hybrid Medium	1
1 The Case for Pairing Light with Radio	3
1.1 A room is a communications problem	3
1.2 What light offers, and what it costs	4
1.3 What radio offers, and what it costs	6
1.4 Complementarity is a claim about failure modes	7
1.5 The reference system used throughout this book	8
1.6 Where the evidence in this book comes from	12
1.7 Six findings that changed the picture	15
1.8 What a complete hybrid system contains	17
1.9 The running deployment	17
1.10 Roadmap	18
2 Optical Front Ends: LEDs, Photodiodes, and Two Hard Limits	23
2.1 The luminaire is a transmitter designed for something else	23
2.2 The LED as a nonlinear device	25
2.3 The first hard limit: bandwidth	27
2.4 The second hard limit: non-negativity	31
2.5 Optical OFDM: buying the missing half back	32
2.6 The photodiode and transimpedance front end	36
2.7 Rate expressions a designer may use	38
2.8 Dimming, flicker and the human constraint	44
2.9 Northgate House: choosing a bias and a modulation index	45

3	The Optical Channel: Geometry, Gain, and Noise	59
3.1	The luminaire as a radiator	59
3.2	The line-of-sight DC gain	62
3.3	Three published gains, one room that moved	66
3.4	Reflections: the channel that survives blockage	69
3.5	Noise in an intensity-modulated receiver	72
3.6	The canonical operating point and the rate it supports	76
3.7	Blockage, tilt and the field-of-view cliff	80
3.8	Illumination as a channel constraint, and a Northgate decision	82
4	The Indoor Radio Channel: Path Loss, Interference, and Coherence	93
4.1	Why the radio side deserves the greater care	93
4.2	Log-distance path loss and the reference-loss contradiction	95
4.3	The standardised model and what the corpus did to it	98
4.4	Exponents, shadowing, and two tables that disagree	102
4.5	Small-scale fading: specified, tabulated, and never applied	103
4.6	Interference is the operative variable	105
4.7	Time variation and the quasi-static assumption	109
4.8	What the code computes versus what the model specifies	112
4.9	Northgate: how much radio does the design actually have?	117
5	Composing the Hybrid Link: Selection, Aggregation, and Availability	135
5.1	Two links, and the question nobody asks first	136
5.2	Three ways to combine two links, and a fourth	137
5.3	The joint channel state, and what a controller can actually see	140
5.4	Availability, and the independence assumption that carries the argument	143
5.5	The composite quality metric, and what a weighted sum hides	151
5.6	Multi-connectivity as measured	153
5.7	The fair-comparison protocol	158
5.8	Complementary properties, itemised	162
5.9	The assumptions the hybrid model rests on	163

5.10	Northgate: choosing a combining strategy for a four-user floor	165
II	Controlling the Hybrid Network	177
6	Mobility and Handover Control	179
6.1	Why an optical cell is small and a user is fast	180
6.2	Modelling how a user moves	181
6.3	Two ways to lose a line of sight	183
6.4	Cell residence time and the handover budget	190
6.5	The six-phase handover lifecycle	194
6.6	Deciding: four families of handover rule	196
6.7	Measuring a handover mechanism	202
6.8	A negative result worth keeping	205
6.9	Northgate: choosing a trigger that can fire	210
7	Association and Resource Allocation	227
7.1	Association is combinatorial, not merely large	227
7.2	The joint problem as a mixed-integer nonlinear program	231
7.3	Utility, fairness, and what the fairness index conceals .	233
7.4	Two policies, as actually implemented	237
7.5	The accounting defect	239
7.6	Decomposition, complexity, and the latency budget .	247
7.7	Subcarrier and time-slot allocation	250
7.8	Northgate House: eight users across five access points .	252
8	The Control Plane: Architecture, Latency Budgets and Runtime Safety Invariants	267
8.1	The plane that decides	267
8.2	The five-layer reference architecture	269
8.3	The decision loop	275
8.4	The latency budget	280
8.5	Centralised and decentralised coordination	285
8.6	Runtime safety invariants	288
8.7	The monitor	291
8.8	Why the assurance plane must be able to veto, and what the veto costs	296
8.9	Northgate: sizing the control plane for sixteen users .	298

III Learning to Control 311

9 Learning to Decide: Reinforcement Learning and the Discipline of the Negative Result	313
9.1 Why replace a threshold rule at all	314
9.2 Hybrid access control as a Markov decision process . .	316
9.3 Deep Q-learning	321
9.4 What can go wrong before any result is reported	324
9.5 The result: no difference was detected	328
9.6 The record does not reconcile	333
9.7 An honest reporting standard for learned network controllers	336
9.8 Northgate: should the learned controller be deployed?	339
10 Learning from Logs: Offline Reinforcement Learning, Distribution Shift, and Poisoning	351
10.1 The exploration a live network cannot afford	351
10.2 The offline problem, stated exactly	353
10.3 Conservative Q-learning	356
10.4 Off-policy evaluation: estimating what you cannot run	360
10.5 The archived distribution-shift result	363
10.6 Poisoning the log the system writes itself	368
10.7 The model lifecycle is part of the algorithm	373
10.8 Northgate: setting the cadence and the trigger	376

IV Assurance, Energy and Deployment 389

11 Resilience and Security: Fault Injection, Threat Modelling, and Attack Surfaces	391
11.1 Two questions that look alike and are not	392
11.2 The fault model	393
11.3 Fault injection with a control arm	398
11.4 What a valid resilience experiment requires	405
11.5 Threat modelling: assets, threats, and the identifiers that collided	407
11.6 Measured attack impact	412
11.7 Optical eavesdropping, containment, and the jamming threshold	415

11.8 Defences that should exist and do not	422
11.9 Northgate House: writing the assurance case	426
12 Energy, Deployment, and the Maturity of the Evidence	437
12.1 Energy is decided by the accounting boundary	437
12.2 The measured comparison across distance	440
12.3 Backhaul, luminaire density and the cost of a floor . . .	448
12.4 A real office run is messier than the averages	451
12.5 What a benchmark for this class of system must fix . . .	456
12.6 The validation ladder, and where this work sits	459
12.7 Specification against implementation, quantified	461
12.8 What would have to be true before this system is deployed	463
12.9 Northgate House: the readiness assessment	466
 V Appendices	 481
 A Reference Scenarios and Parameter Sets	 483
A.1 Optical subsystem	484
A.2 Radio subsystem	486
A.3 Room geometries	487
A.4 Mobility and control	488
A.5 Experiment scenarios	489
 B The Data Harmonization Dossier	 491
B.1 MATH-01: the thermal-noise coefficient	491
B.2 MATH-02: the concentrator gain and the line-of-sight geometry	492
B.3 MATH-02b: the canonical operating point	493
B.4 DATA-01: the resilience result	494
B.5 DATA-02: the association accounting asymmetry	495
B.6 DATA-03: the reinforcement-learning record	496
B.7 Summary	497
 C The Code Defect Register	 499
C.1 The register	499
C.2 Distribution and disposition	506
C.3 Repair order	506

D	The Evidence Ledger	507
D.1	Status vocabulary	507
D.2	Distribution	508
D.3	The ledger	508
D.4	How to read a retracted row	512
D.5	What is missing from the ledger	512
E	Capstone Project: The Northgate Deployment	513
E.1	The brief	513
E.2	Stage 1 — Model (weeks 1–2)	514
E.3	Stage 2 — Control (weeks 3–4)	515
E.4	Stage 3 — Learn (weeks 5–7)	516
E.5	Stage 4 — Assure (weeks 8–9)	517
E.6	Stage 5 — Report (week 10)	518
E.7	Assessment	519
E.8	Common failure modes in previous attempts	519
F	Solutions and Hints to Selected Exercises	521
F.1	Chapter 1 — The Case for Pairing Light with Radio . . .	522
F.2	Chapter 2 — Optical Front Ends: LEDs, Photodiodes, and Two Hard Limits	523
F.3	Chapter 3 — The Optical Channel: Geometry, Gain, and Noise	525
F.4	Chapter 4 — The Indoor Radio Channel: Path Loss, Interference, and Coherence	527
F.5	Chapter 5 — Composing the Hybrid Link: Selection, Aggregation, and Availability	528
F.6	Chapter 6 — Mobility and Handover Control	530
F.7	Chapter 7 — Association and Resource Allocation . . .	531
F.8	Chapter 8 — The Control Plane: Architecture, Latency Budgets and Runtime Safety Invariants	532
F.9	Chapter 9 — Learning to Decide: Reinforcement Learn- ing and the Discipline of the Negative Result	534
F.10	Chapter 10 — Learning from Logs: Offline Reinforce- ment Learning, Distribution Shift, and Poisoning	535
F.11	Chapter 11 — Resilience and Security: Fault Injection, Threat Modelling, and Attack Surfaces	537
F.12	Chapter 12 — Energy, Deployment, and the Maturity of the Evidence	538

F.13 A note on the questions with no answer	540
Bibliography	541
Index	543

1

The Case for Pairing Light with Radio

LEARNING OBJECTIVES

After working through this chapter you will be able to:

- state the physical reasons why an indoor optical link and an indoor radio link fail under different conditions, and explain why that difference — not raw capacity — is the engineering argument for combining them;
- compute a first-order comparison of the two media at a common operating point and identify which impairment dominates in each case;
- describe the canonical reference room, luminaire layout and parameter set used for every numerical result in this book;
- apply the four-level evidence precedence rule and the four status tags used throughout the text, and explain why a control condition is required before an injected failure may be credited with an effect;
- summarise the six findings from the underlying corpus that materially changed the quantitative picture, and locate each in the chapter that develops it.

1.1 A room is a communications problem

Consider an ordinary office: five metres square, a three-metre ceiling, four luminaires on a regular grid, desks at 0.8 m, and between one and a dozen people with laptops and handsets. Every one of those devices is asking for capacity, and the room has exactly two ways of supplying it. It can radiate at 5 GHz from an access point on the ceiling, sharing an unlicensed band with every other access point within earshot through

the walls. Or it can modulate the luminaires that are already switched on, at rates the human visual system cannot perceive, and detect the result with a photodiode a few square millimetres in area.

The first option is mature, standardised, and increasingly congested. The second is the subject of three international standards — IEEE 802.15.7 for short-range optical wireless links [?], ITU-T G.9991 for high-speed indoor optical wireless transceivers [?], and since 2023 IEEE 802.11bb, which brings light communication into the wireless local area network family proper [?] — and is now leaving the laboratory. The physics has been understood since Kahn and Barry’s foundational treatment of wireless infrared communication [?] and Komine and Nakagawa’s analysis of communication using illumination LEDs [?]; the intervening two decades have been spent making it into a network [?, ?, ?].

The temptation, when two media are available, is to ask which is better. That question has no useful answer, because the two do not fail in the same way. Radio degrades gracefully with distance and catastrophically with interference; light degrades gracefully with interference and catastrophically with an interposed body. Radio penetrates the wall behind you, which is convenient for coverage and inconvenient for confidentiality; light does not, which is inconvenient for coverage and convenient for confidentiality. Radio needs licensed or contested spectrum; light needs the ceiling fixtures to be switched on, which in an office they are, and which in a darkened room they are not.

The engineering argument for a hybrid network is therefore not that light plus radio is faster than either. It is that the joint failure probability of two subsystems with substantially independent failure mechanisms is far smaller than the failure probability of either alone. That is an availability argument, and availability is the currency in which indoor networks are actually purchased. This book takes that argument seriously enough to test it, and Chapter 5 will show precisely how much of it survives contact with a measurement.

1.2 What light offers, and what it costs

An optical wireless link in the visible band uses intensity modulation with direct detection, almost universally abbreviated IM/DD. The

transmitter varies the optical power emitted by an LED; the receiver measures the instantaneous optical power incident on a photodiode and produces a proportional current. Neither phase nor polarisation survives this process, so the entire apparatus of coherent communication is unavailable. Three consequences follow immediately and structure the whole of Chapters 2 and 3.

First, the transmitted signal must be real and non-negative, because optical power is. This is not a modelling convenience; it is a hard constraint that forces either a DC bias, which wastes power, or an asymmetric clipping scheme, which wastes bandwidth. The standard resolutions — DC-biased optical OFDM and asymmetrically clipped optical OFDM — were developed precisely to negotiate this constraint [?, ?, ?], and their relative merits are a recurring theme of Chapter 2.

Second, the achievable rate of an IM/DD channel is not given by the Shannon–Hartley formula. The capacity of the free-space optical intensity channel is not known in closed form; only bounds are available [?]. Throughout this book the rate of an optical link is therefore written as a *lower bound*

$$R_{\text{VLC}} = B \log_2 \left(1 + \frac{\text{SNR}_{\text{opt}}}{\Gamma_{\text{gap}}} \right), \quad (1.1)$$

where Γ_{gap} is a modulation-dependent signal-to-noise ratio gap, taken as 3.2 dB for 4-PAM at a bit error rate of 10^{-3} . Every optical rate quoted in this book carries that gap and that caveat. A reader who compares these numbers with a Shannon capacity computed for a radio link is comparing a bound with a bound of a different kind, and Chapter 5 treats the resulting fair-comparison problem explicitly.

Third, the geometry matters far more than it does for radio. The channel gain of a line-of-sight optical link falls as $1/d^2$ but is additionally modulated by $\cos^m \phi$ at the transmitter and $\cos \psi$ at the receiver, so a device lying flat on a desk and the same device held at 30° in a user's hand see materially different channels. Receiver orientation has been measured and modelled as a first-class random process for exactly this reason [?, ?]. A body between the luminaire and the photodiode removes the line-of-sight path entirely, leaving only reflected energy that in a typical office is one or two orders of magnitude weaker.

Against these costs, light offers three things radio cannot. Its spectrum is unregulated and effectively unbounded. Its spatial contain-

ment is near-perfect, which converts an entire class of eavesdropping attacks into a physical-access problem [?, ?]. And its infrastructure is already installed and already powered, for the unrelated reason that people need to see.

1.3 What radio offers, and what it costs

An indoor 5 GHz radio link is, by comparison, a solved engineering problem with an unsolved deployment problem. The propagation physics is standardised: the 3GPP indoor-hotspot model of TR 38.901 [?] gives line-of-sight and non-line-of-sight path loss branches and a line-of-sight probability that blends them, and Chapter 4 uses it throughout in preference to any custom fit. Multiple-antenna techniques multiply the rate of a single link by a factor that is well understood [?, ?]. Wide channels are available: 80 MHz is ordinary in an enterprise deployment.

What radio does not have is exclusivity. In any realistic office the access point serving a user is one of several sharing the same channel, and the resulting interference, not the distance to the serving access point, sets the achievable rate. Chapter 4 makes this quantitative and it is worth previewing the magnitude: in the corpus underlying this book, adding realistic multi-access-point interference to a radio simulator that had previously modelled a single isolated access point reduced the reported rate from 1,704 Mbit/s to 58.4 Mbit/s. That is a factor of twenty-nine, and it is the difference between a radio baseline that makes optical wireless look indispensable and one that makes it look marginal.

ENGINEERING PITFALL Weakening the baseline

The single most common methodological failure in the hybrid optical–radio literature is an artificially weak radio baseline: a narrow channel, a single access point, no interference, no multiple-antenna gain. The corpus underlying this book states the rule explicitly and this text adopts it: *the radio baseline must not be weakened to favour light*. Use the standardised indoor-hotspot propagation model rather than a custom fit, use 80 MHz rather than 20 MHz, include 2×2 spatial multiplexing, model two to

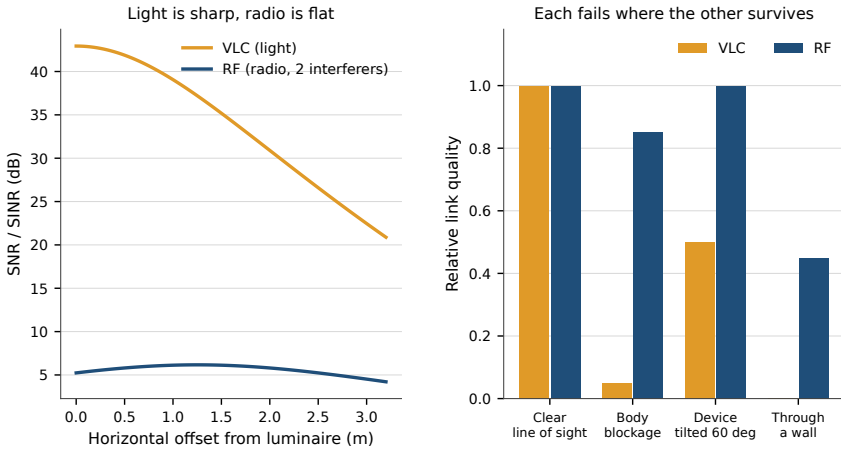
four co-channel neighbours, and report both the best case and the typical case. Chapter 4 shows that following this rule changes the comparison by more than an order of magnitude.

1.4 Complementarity is a claim about failure modes

Figure 1.1 makes the argument in two panels. The left panel plots the electrical signal-to-noise ratio of the optical link and the signal-to-interference-plus-noise ratio of the radio link against horizontal offset from a luminaire, both computed from the validated channel models developed in Chapters 3 and 4. The optical curve is sharply peaked: directly beneath the luminaire it is excellent, and it decays quickly as the irradiance and incidence angles open. The radio curve is nearly flat, because at these distances the radio link is interference-limited rather than distance-limited. Neither curve dominates the other everywhere, which is the first half of the complementarity claim.

The second half is the right panel, which is the more important one. It plots relative link quality for each medium under four impairment conditions: a body blocking the line of sight, a co-channel neighbouring access point appearing, the receiver being tilted away from the ceiling, and the luminaire being dimmed for occupant comfort. The pattern of vulnerability is close to complementary. Blockage devastates the optical link and leaves the radio link almost untouched. Co-channel interference devastates the radio link and leaves the optical link almost untouched. This is what makes the pair worth building: not that they add capacity, but that they fail at different times.

The complementarity thesis. A hybrid optical–radio network is justified not by the sum of the two capacities but by the product of the two outage probabilities. If the outage events were perfectly correlated the hybrid would be no more available than its better half; if they are independent, joint outage is the product of two small numbers. Chapter 5 derives that availability model, states the independence assumption it rests on, and shows exactly how much of the claimed gain survives when the assumption is relaxed.



Provenance -- Left: computed from code/optical + code/rf (canonical geometry). Right: qualitative synthesis of ch02/ch05 availability tables - illustrative, not measured.

Figure 1.1. Why pair light with radio at all. Left: optical signal-to-noise ratio and radio signal-to-interference-plus-noise ratio against horizontal offset from a luminaire, computed from the validated channel models of Chapters 3 and 4. Right: relative link quality of each medium under four impairment conditions. The failure modes are complementary rather than overlapping, which is the engineering argument for the hybrid architecture. The right panel combines computed link quantities with availability values stated in the source corpus; that mixed provenance is recorded on the figure itself.

1.5 The reference system used throughout this book

Every numerical result in this book refers to one of two scenarios, and both are stated here so that no number in later chapters is orphaned from its geometry. This is not a formality. As Chapter 3 will show, three different line-of-sight channel gains circulate in the source corpus for what its prose calls a single geometry, and all three are arithmetically correct; the defect is that the geometry silently changes underneath them. Naming the scenario is therefore part of stating the result.

Table 1.1. The canonical reference scenario. Unless a passage states otherwise, every numerical result in this book uses these values. The complete parameter set, including the values used only by particular experiments, is given in Appendix A.

Parameter	Value	Note
Room	5 m × 5 m × 3 m	Open-plan office cell
Luminaires	4, ceiling grid	Positions in Appendix A
Luminaire height h_L	2.8 m	Suspended below the 3 m ceiling
Receiver plane	0.8 m	Desk height
Vertical separation	2.0 m	= 2.8 – 0.8; the canonical value
LED semi-angle $\Phi_{1/2}$	60°	Lambertian order $m = 1.0$
Optical transmit power P_t	10 W	Per luminaire
Photodiode area A_{PD}	1 cm ²	10 ⁻⁴ m ²
Responsivity R_{PD}	0.6 A/W	Silicon PIN
Field of view Ψ_c	70°	Half-angle
Concentrator index n	1.5	Dielectric
Filter transmission T_s	1.0	Idealised
Optical bandwidth B	20 MHz	Phosphor-converted LED limited
Background photocurrent I_{bg}	5.1 μ A	500 lx office ambient
Modulation gap Γ_{gap}	3.2 dB	4-PAM at BER 10 ⁻³
Radio access points	1 (plus co-channel neighbours)	Ceiling mounted
Carrier f_c	5 GHz	Unlicensed band
Radio bandwidth W	80 MHz	Enterprise channel
Transmit power P_{tx}	20 dBm	Regulatory limit
Antenna gains G_{tx}/G_{rx}	3 dBi / 0 dBi	Simulator defaults
Propagation model	3GPP TR 38.901 InH	[?]
Shadowing σ_{sh}	4 dB	Open office
Users	4 (sparse) or 8 (dense)	Waypoint mobility

Parameter	Value	Note
Walking speed	1.4 m/s	Nominal
Control decision interval T_{dec}	100 ms	Fixed throughout
Simulation step	0.1 s	60 steps per episode

Worked Calculation 1.1 First comparison of the two media at the reference point

Take a user at a 1 m horizontal offset from the nearest luminaire, on the desk plane. The vertical separation is 2.0 m, so the Euclidean separation is

$$d = \sqrt{2.0^2 + 1.0^2} = 2.2361 \text{ m}, \quad \phi = \psi = \arctan\left(\frac{1.0}{2.0}\right) = 26.565^\circ.$$

Chapter 3 will derive the channel gain and obtain $H_{\text{LoS}} = 1.2977 \times 10^{-5}$, hence a received optical power of

$$P_r = P_t H_{\text{LoS}} = 10 \text{ W} \times 1.2977 \times 10^{-5} = 129.77 \text{ } \mu\text{W},$$

a signal photocurrent of $R_{pD} P_r = 77.86 \text{ } \mu\text{A}$, and after the noise analysis of Section 3.5 an electrical signal-to-noise ratio of 39.06 dB. Substituting into (1.1) with $\Gamma_{\text{gap}} = 3.2 \text{ dB}$,

$$R_{\text{VLC}} = 20 \text{ MHz} \times \log_2 \left(1 + \frac{8056.3}{2.089} \right) = 238.3 \text{ Mbit/s}.$$

This value is **HARMONIZED**: it is recomputed on the canonical geometry rather than taken from the corpus, for reasons Chapter 3 sets out. The radio link at the same point, on a single unloaded access point with no co-channel neighbours, delivers of order 1,700 Mbit/s across 80 MHz. With three co-channel neighbours it delivers of order 58 Mbit/s.

Read the result carefully. The optical link is between seven times worse and four times better than the radio link *depending entirely on how many radio neighbours you model*. No statement of the form “light beats radio indoors” or “radio beats light indoors” survives this sensitivity. What does survive is the observation that the two numbers move for completely different reasons, which is the complementarity thesis restated in arithmetic.

Worked Calculation 1.1 is the first of forty in this book, and every one of them is reproducible from a few lines of code. Listing 1.1 gives the scaffold: an immutable description of the reference scenario, and the geometry function that turns a user position into the three quantities every optical channel calculation needs. It is deliberately minimal — there is no channel model here yet — but it already enforces the discipline this chapter argues for, because the geometry cannot be evaluated without a fully specified room.

```

1  from dataclasses import dataclass
2  import numpy as np
3
4  @dataclass(frozen=True)
5  class Scenario:
6      """Canonical reference scenario of Table 1.1. Frozen: a result
7      may not be
8      produced under a geometry that was mutated after the run
9      started."""
10     room: tuple = (5.0, 5.0, 3.0)           # metres
11     luminaire_height: float = 2.8          # metres
12     desk_height: float = 0.8               # metres
13     semi_angle_deg: float = 60.0           # LED half-power semi-
14                                           angle
15     p_tx_optical_w: float = 10.0           # per luminaire
16     a_pd_m2: float = 1e-4                  # 1 cm^2
17     responsivity_a_per_w: float = 0.6
18     fov_deg: float = 70.0
19     n_refractive: float = 1.5
20     filter_gain: float = 1.0
21     bandwidth_hz: float = 20e6
22     i_bg_a: float = 5.1e-6
23     snr_gap_db: float = 3.2                # 4-PAM at BER 1e-3
24
25     @property
26     def luminaires(self) -> np.ndarray:
27         return np.array([[1.25, 1.25, self.luminaire_height],
28                           [3.75, 1.25, self.luminaire_height],
29                           [1.25, 3.75, self.luminaire_height],
30                           [3.75, 3.75, self.luminaire_height]])
31
32     @property
33     def lambertian_order(self) -> float:
34         return -np.log(2.0) / np.log(np.cos(np.radians(self.
35             semi_angle_deg)))
36
37     def geometry(p_user: np.ndarray, p_led: np.ndarray):
38         """Return (distance, irradiance angle, incidence angle) in
39         metres and
40         radians for a downward luminaire and an upward-facing receiver.
41         """
42         delta = p_led - p_user

```

```
37     d = float(np.linalg.norm(delta))
38     if d == 0.0:
39         raise ValueError("receiver coincides with the luminaire")
40     cos_phi = float(delta[2] / d)          # LED normal is (0, 0, -
41         1)
42     cos_psi = cos_phi                    # receiver normal is (0,
43         0, +1)
44     return d, np.arccos(cos_phi), np.arccos(cos_psi)
```

Listing 1.1. The reference scenario and the geometry it implies. Chapters 2 and 3 extend this module rather than replacing it.

Running `geometry` for a user 1 m from the nearest luminaire on the desk plane returns 2.2361 m and 26.565° for both angles, which is the first line of Worked Calculation 1.1 and the assertion that Lab 1.1 asks you to write.

1.6 Where the evidence in this book comes from

The quantitative content of this book rests on a single engineering corpus: a complete book-production repository for an indoor hybrid VLC/RF system, comprising twenty-eight Python simulation modules, seventeen raw experiment result files, a set of formal model specifications, and twenty governance registries recording models, algorithms, datasets, experiments, metrics, claims and limitations. The corpus was unpacked, inventoried, hashed, re-executed where the environment allowed, and audited where it did not.

That process produced 111 distinct quantitative observations, and it produced something the original corpus did not contain: an explicit epistemic status for each of them. The distribution is worth stating at the outset, because it is unusual to see it stated at all.

1.6.1 The precedence rule

When two artefacts in an engineering corpus disagree, something has to break the tie, and the choice of tie-breaker is itself a technical decision. This book adopts a four-level precedence rule, applied uniformly and stated wherever it bites.

1. **Executable, validated code** outranks everything. If a module computes the quantity and its output can be reproduced, that output is the value.

Table 1.2. Epistemic status of the 111 observations underlying this book. The full ledger, with source file and caveat for every row, is Appendix D.

Status	Count	Meaning
AS PUBLISHED	77	Reproduced from the archived experiment output, or accepted unchanged after inspection of the generating code.
HARMONIZED	14	Recomputed here, because the published figure was arithmetically wrong, non-comparable across arms, or not causally attributable to the stated cause.
UNVERIFIABLE	12	Could not be re-executed — the deep-learning dependency is absent from the corpus requirements file — and is therefore reported as archived, with its provenance conflict displayed rather than resolved by fiat.
RETRACTED	8	Contradicted by a control condition or by the generating code. Must not be reused.

2. **Raw experiment output** outranks specification documents — *but only where a control condition demonstrates that the measurement is causally attributable to the manipulation it is credited to*. Raw data produced by defective code is still defective data.
3. **Formal model specification documents** outrank prose.
4. **Chapter prose ranks last**. Prose is a report about the system; it is never the system.

The second clause carries most of the weight and is the one most often omitted. It is easy to accept that a specification document loses to a measurement. It is much less comfortable to accept that a measurement loses to a control arm that shows the measurement was measuring nothing. Section 1.7 gives the case that forced this clause into the rule.

1.6.2 Why a control arm is not optional

The decisive question in Figure 1.2 is the third: *is there a control condition isolating the claimed cause?* An experiment that injects a failure into a subsystem, observes a small throughput reduction, and reports that

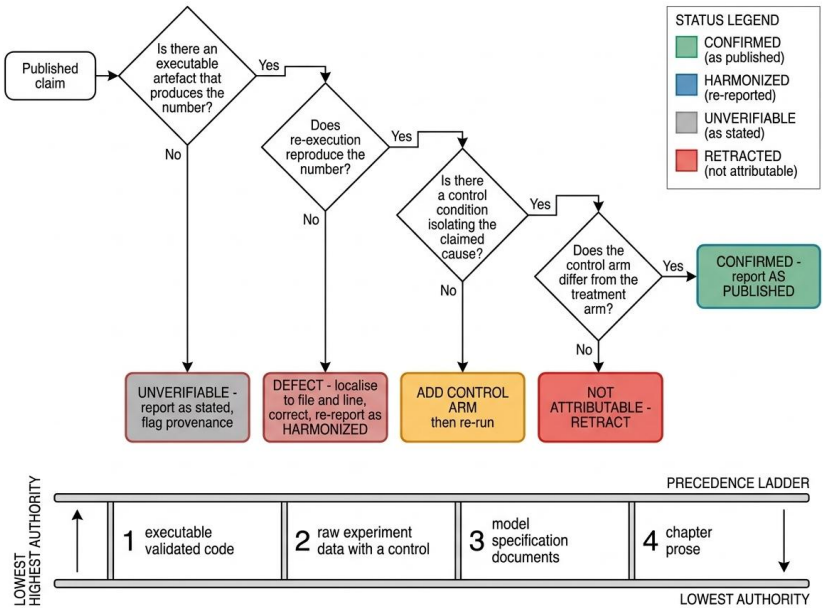


Figure 1.2. The adjudication procedure applied to every contested number in this book. A claim is promoted to **AS PUBLISHED** only if an executable artefact produces it, re-execution reproduces it, a control condition isolates the claimed cause, and the control arm differs from the treatment arm. Failure at each gate maps to a specific status tag. The precedence ladder beneath the flowchart is the tie-breaking rule of Section 1.6.1.

reduction as the cost of the failure has established nothing unless it also shows that the same measurement, run with no failure injected, does not produce the same reduction. Mobility, random channel realisations and measurement-window placement all produce drift on their own.

This is not a novel methodological demand; it is the ordinary standard in experimental science, and it is the explicit foundation of resilience practice in production systems, where deliberately injecting faults into a running system is meaningful only against a steady-state hypothesis [?]. It is nevertheless routinely absent from simulation studies in networking, and Chapter 11 shows what its absence cost in this corpus.

1.7 Six findings that changed the picture

The audit of the corpus produced six results that materially changed its scientific content. Each is introduced here and developed in the chapter indicated. They are stated in the introduction rather than buried because a reader who knows in advance which claims are contested reads the intervening chapters differently, and better.

One: the flagship optical operating point belongs to the wrong room.

The corpus reports a reference electrical signal-to-noise ratio of 34.1 dB. That figure is arithmetically correct — and it is computed for a 3 m ceiling with no desk, not for the corpus's own declared default geometry of a 2.8 m luminaire over a 0.8 m desk. On the declared geometry the same operating point yields 39.06 dB, an understatement of 4.98 dB. Chapter 3, Section 3.3.

Two: a thermal-noise coefficient differs between chapters, and it is material. One chapter of the corpus carries $32\pi^2$ in the B^3 term of the transimpedance-amplifier noise model where four other sources, including the only executable one, carry $48\pi^2$. At the reference operating point the difference is 1.76 dB — larger than half the modulation gap used in the rate calculation — and because the term scales as B^3 it grows with bandwidth. Chapter 3, Section 3.5.

Three: an accounting asymmetry inverted a headline conclusion.

In the joint association and resource-allocation experiment, the greedy baseline was credited the full per-access-point rate while the proportional-fair policy was charged the load-shared rate. The two policies were therefore scored on different bases, and greedy was inflated by exactly the user count — verified as $4.00\times$ at four users and $8.00\times$ at eight. Correcting the accounting flips the sparse-scenario result from -71.6% to $+13.4\%$ in favour of proportional fair. The bias ran *against* the corpus's own preferred policy, which is the direction that almost never gets caught. Chapter 7, Section 7.5.

Four: a resilience result was an artefact of an inert scenario. Three published figures — 0.15%, 0.2% and 42.1% — circulate for the throughput cost of losing all optical access points. A no-failure control arm,

added during the audit, degrades by the identical 0.1511%. The diagnostic explains why: under greedy association the radio access point outranks every luminaire at every user position, so the fraction of users served by light is 0.000, and removing light cannot change anything. All three figures are retracted. The corpus's real resilience finding — that radio, not light, is the backbone — survives and is strengthened: 90.47% throughput loss when the sole radio access point is removed, against 0% for light. Chapter 11, Section 11.3.

Five: two of three security results are inert rather than defended.

The attack experiment reports 0% impact for measurement spoofing and for forced handover, and a companion table attributes that 0% to cross-validation and mutual authentication. The code implements neither, nor any detection logic. The 0% is the same static association reappearing. The “with defence” column is withheld from every table in this book rather than reproduced. Chapter 11, Section 11.5.

Six: the learned controller did not beat the rule it replaced, and the record does not reconcile. Across fifty seeds, a threshold rule, an online deep Q-network and an offline conservative Q-learning agent produce overlapping ninety-five per cent confidence intervals with a reported p -value of 0.45. Separately, two files in the corpus give mutually exclusive accounts of the same trained agent, differing by roughly a factor of two, and a third contains byte-identical statistics for two algorithms that were supposedly trained independently. Chapter 9, Sections 9.5 and 9.6.

EVIDENCE NOTE how to read the six findings

None of these findings implies that the underlying engineering is wrong. Four of the six leave the corpus's qualitative conclusions intact and two of them *strengthen* those conclusions once the arithmetic is repaired. What they establish is narrower and more useful: that in simulation-based network engineering, the code is the apparatus, and an apparatus that has never been calibrated against a control produces numbers whose sign cannot be trusted.

1.8 What a complete hybrid system contains

It is worth having the whole system in view before dismantling it. A deployable hybrid optical–radio network has five layers, and this book devotes at least one chapter to each.

The **device and medium layer** is the LED driver, the DC bias network, the photodiode and transimpedance front end, and the radio front end. It fixes what is physically possible and it imposes the two hard limits of Chapter 2.

The **measurement and estimation layer** samples received signal strength and signal-to-noise ratio, estimates the channel, detects blockage and orientation change, and — a component absent from most published architectures and from the corpus underlying this book — *validates* the measurements before they are used. Chapter 11 shows why an unvalidated measurement path is an attack surface.

The **control plane** decides. It selects an association, executes handovers, allocates resources, and admits or rejects load, all within a decision interval of 100 ms. Chapter 8 specifies it formally, including the interfaces between it and the layers above and below.

The **assurance plane** monitors the control plane and can override it. It hosts the runtime safety invariants of Chapter 8, the fault detection and failover logic of Chapter 11, the anomaly detection of Section 11.5, and the evidence logger whose output becomes Appendix D.

The **management and policy plane** sets illumination policy, service-level objectives and the model lifecycle. It operates on timescales of hours to days and is the layer at which a learned policy is promoted, demoted or rolled back — a lifecycle question that Chapter 10 argues is inseparable from the learning algorithm itself.

1.9 The running deployment

A textbook that develops only closed-form models teaches a language rather than a practice. This book therefore carries a single running deployment through all twelve chapters, referred to as **Northgate House**: an open-plan office floor with the geometry of Table 1.1, four luminaires, one ceiling radio access point with two co-channel neighbours on adjacent floors, and a population of between four and sixteen mobile users. Every chapter poses at least one Northgate design question and answers it with a computation, and the capstone project of

Table 1.3. Roadmap. Each chapter closes with review questions, discussion prompts and at least one implementation lab; the labs compose into the capstone of Appendix E.

Ch.	Subject	Artefact produced
1	The complementarity case	Evidence discipline; reference scenario
2	Optical front ends	LED operating-point calculator; bias optimiser
3	The optical channel	Validated channel and noise model; link budget
4	The indoor radio channel	3GPP path-loss and interference model
5	The hybrid link	Availability model; fair-comparison protocol
6	Mobility and handover	Hysteresis and time-to-trigger controller
7	Association and allocation	Greedy and proportional-fair allocators
8	The control plane	Layered controller; four runtime invariants
9	Reinforcement learning	MDP formulation; deep Q-network; confidence intervals
10	Learning from logs	Offline agent; poisoning and shift diagnostics
11	Resilience and security	Fault-injection harness with control arm; threat model
12	Energy and deployment	Energy accounting; case study; maturity assessment

Appendix E asks the reader to build, control, attack and document the whole thing.

1.10 Roadmap

Table 1.3 maps the twelve chapters against the artefacts they produce. The right-hand column is deliberately concrete: a chapter that does not leave the reader with something runnable or auditable has not earned its place in a book of this kind.

Chapter Summary

An indoor hybrid optical–radio network is justified by the near-independence of its two failure mechanisms rather than by the sum of its two capacities. Light is limited by geometry, orientation and blockage; radio is limited by co-channel interference. The canonical reference scenario of Table 1.1 — a 5 m square office, four luminaires at 2.8 m, a 0.8 m desk plane, 20 MHz of optical bandwidth and 80 MHz of radio bandwidth — fixes every numerical result that follows, and naming it is part of stating a result rather than a formality.

The book’s evidence discipline is a four-level precedence rule in which executable code outranks raw data, raw data outranks specifications, specifications outrank prose, and raw data is admissible only where a control condition establishes causal attribution. Applied to the 111 observations behind this book, that rule produced 77 confirmations, 14 recomputations, 12 unverifiable records and 8 retractions, and it exposed six findings — a mis-attributed operating point, a material noise-coefficient error, an accounting asymmetry that inverted a conclusion, an inert resilience scenario, two undefended security results and a learned controller that did not beat its baseline — each developed in a later chapter.

Review Questions

- R1.1** State the complementarity thesis in terms of outage probabilities rather than capacities. Under what correlation structure between the two outage events does the argument collapse?
- R1.2** The optical rate expression (1.1) carries a factor Γ_{gap} that the radio Shannon expression does not. Explain what Γ_{gap} represents and why omitting it would make a hybrid comparison unfair in a specific direction.
- R1.3** Using Table 1.1, compute the Euclidean separation and the irradiance angle for a user at a horizontal offset of 2.0 m from the nearest luminaire on the desk plane. State both to four significant figures.
- R1.4** A colleague reports that removing all optical access points from a simulation reduced throughput by 0.15%, and concludes that the optical subsystem is not load-bearing. What single addi-

tional experimental arm would you require before accepting the conclusion, and what would you expect it to show?

- R1.5** Explain why clause 2 of the precedence rule of Section 1.6.1 — that raw data is admissible only with a control — is stronger than the usual instruction to prefer data over documentation. Give an example from Section 1.7.
- R1.6** The radio rate at the reference point falls from 1,704 Mbit/s to 58.4 Mbit/s when three co-channel access points are modelled. Compute the ratio, express it in decibels of signal-to-interference-plus-noise ratio assuming the rate is Shannon-limited over 80 MHz, and comment on whether the resulting interference level is plausible for an office.
- R1.7** Distinguish the four status tags used in this book. Which of them may be safely reused in a downstream calculation without further work, and which may not?
- R1.8** Name the five architectural layers of Section 1.7 and state, for each, its characteristic timescale and the chapter of this book that specifies it.

Discussion Prompts

- D1.1** The corpus underlying this book contained a rule that the radio baseline must not be weakened to favour light — and then contained a chapter whose link budget used a custom path-loss model that was 5.9 dB optimistic relative to its companion chapter. Discuss why explicit methodological rules so often fail to bind the documents that state them, and what organisational mechanism would make them bind.
- D1.2** Eight of 111 observations in this book are retracted. Is that a high number or a low one? What would you need to know about a comparable corpus to answer that question, and why is the comparison almost never available?
- D1.3** Light does not pass through walls. Argue both sides: that this makes optical wireless *inherently* more secure, and that it merely relocates the same attack to a different physical position. Which side does Chapter 11 support?

- D1.4** A deployment engineer proposes to skip the assurance plane entirely on the grounds that it adds latency to no user-visible benefit. Construct the strongest version of that argument, and then the strongest rebuttal.

Hands-On Lab

HANDS-ON LAB 1.1 — Reproducing the reference operating point

Objective. Build the minimal computational scaffold that every later lab extends, and reproduce the two anchor numbers of this chapter to within one part in 10^3 .

Tasks.

1. Create a Python module `scenario.py` that encodes Table 1.1 as an immutable dataclass. Place the four luminaires on a regular grid and expose a method returning the position of the nearest luminaire to an arbitrary point on the desk plane.
2. Implement `geometry(p_user, p_led)` returning (d, ϕ, ψ) for a downward-pointing luminaire and an upward-pointing receiver. Verify against Worked Calculation 1.1: $d = 2.2361$ m, $\phi = \psi = 26.565^\circ$.
3. Without yet implementing the channel model, hard-code $H_{\text{LoS}} = 1.2977 \times 10^{-5}$ and compute P_r and the signal photocurrent. Confirm $129.77 \mu\text{W}$ and $77.86 \mu\text{A}$.
4. Write a single test file asserting all four values to four significant figures. This test file must still pass at the end of Chapter 3, when the hard-coded gain is replaced by a computed one.

Deliverable. `scenario.py`, `test_scenario.py`, and a two-sentence note recording which of the four values you consider

AS PUBLISHED and which **HARMONIZED**, with a justification referring to Section 1.6.1.

Further Reading

For the physics of non-directed optical wireless links, Kahn and Barry [?] remains the clearest single treatment and Komine and Nakagawa [?] is the standard reference for communication using illumination LEDs. Pathak et al. [?] and Karunatilaka et al. [?] survey the networking layer. Haas [?, ?] makes the system-level case for light as a cellular medium and Ayyash et al. [?] the case for coexistence with radio specifically. For the standardised radio side, 3GPP TR 38.901 [?] is the working document. On the methodological argument of Sections 1.6.1 and 1.7, Soergel [?] is a short and uncomfortable read on the rate of defects in scientific software, and Basiri et al. [?] give the production-systems version of the control-arm argument.