The Geometry of Fuel

The Geometry of Fuel

Curvature, Computation,
and the Search for New Energy

Bee Rosa Davis

Expanded Second Edition

2026

Copyright © 2026 Bee Rosa Davis

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means without the prior written permission of the author, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

The mathematical and physical content of this work is placed in the public domain for the benefit of humanity. The narrative and literary elements remain under copyright.

Expanded Second Edition

Imprint: Independently published

Printed in the United States of America

For my mother—

who turned emergencies into instructions
and fear into a checklist.

You taught me that geometry lives in how we hold each other:
the parallel transport of love across distance,
the curvature that bends us toward home.

Six years since you passed,
and the manifold you embedded in my spirit
is still holding strong against every storm
this work has brought.

The theorems are mine.
The persistence is yours.

Te amo, Mami.
Nos vemos pronto.

"La geometría no pregunta quién eres.
Solo pregunta si estás prestando atención."

The geometry does not ask who you are.
It only asks if you are paying attention.

Contents

Preface to the Second Edition **xi**

1 The Shape of Possibility **1**
1.1 The question I carried home . . . 4
1.2 Making room for what is real . . . 7
1.3 A place for every possible arrangement . . . 10
1.4 Changing the coordinates, keeping the world . . . 20
1.5 The space I had to expand . . . 29

2 Curvature and Its Costs **31**
2.1 What a change of format can cost . . . 32
2.2 The bend a measurement can detect . . . 35
2.3 Keeping track of every direction . . . 40
2.4 Putting the cost on a bench . . . 45
2.5 The equations behind the argument . . . 52

3 The Heat Kernel **55**
3.1 Following heat through a space . . . 57
3.2 The patterns that fade at different rates . . . 61
3.3 What the slowest pattern tells us . . . 65
3.4 Which operator has the gap? . . . 68
3.5 Bringing the measurement back to the task . . . 72
3.6 The equations behind the argument . . . 73

4 The Davis Manifold **75**
4.1 Deciding which differences matter . . . 79
4.2 Returning to the scanner . . . 81
4.3 Recognizing a signal through a change . . . 83
4.4 From the NASA problem to the replay . . . 85
4.5 From a wire format to a database . . . 87
4.6 The same alarm through three pipelines . . . 90
4.7 Keeping the relationship between three channels . . . 99
4.8 Following the error through time and across formats . . . 103
4.9 Keeping the answers the measurements leave open . . . 108
4.10 Taking the question beyond telemetry . . . 112

5 The First Door **117**
5.1 Choosing what to measure in a field . . . 119
5.2 Building a calculation I could check . . . 123
5.3 The empty interval on the screen . . . 128
5.4 Putting the observation under pressure . . . 131
5.5 What survives a change of gauge . . . 135
5.6 Replaying the 200 saved configurations . . . 138
5.7 From feature spacings to an energy question . . . 142
5.8 What an energy-gap proof would require . . . 147
5.9 Living with the observation . . . 150
5.10 Following the question into other problems . . . 151

5.11 What I could carry through the first door . . . 155

6 The Law **161**
6.1 The budget I could write down . . . 165
6.2 From one relationship to a programme of questions . . . 173
6.3 Putting the work where someone could find it . . . 177
6.4 The energy question beneath the geometry . . . 180
6.5 What I wanted to leave on the record . . . 183

7 The Field Equations **187**
7.1 What the geometric budget measures . . . 187
7.2 What the observations actually determine . . . 189
7.3 Following the rangefinders to an answer . . . 191
7.4 Fitting three routes that do not close . . . 196
7.5 Carrying an answer through several steps . . . 204
7.6 Testing the limits of the model . . . 205
7.7 Giving completion an objective . . . 208
7.8 Following the objective through time . . . 209
7.9 Turning the model into a computation . . . 211

8 The Ratchet **213**
8.1 What a vacuum state permits . . . 213
8.2 Changing the model and naming the response . . . 216
8.3 Keeping the two numerical experiments separate . . . 218
8.4 What the records say . . . 220
8.5 Asking what drives the change . . . 225
8.6 The distance from model units to physical power . . . 236
8.7 Keeping the calculation attached to its record . . . 239

9 What Fuel Means **243**
9.1 Putting the meter around the whole cycle . . . 243
9.2 Following the energy bill . . . 248
9.3 Who could use the change? . . . 251
9.4 The responsibility of putting it on the record . . . 253

10 What a measurement must carry **257**
10.1 Keep the movement in the record . . . 260
10.2 Give the number its clock back . . . 267
10.3 A rod is not a pair of numbers . . . 269
10.4 Deriving the coordinate transformation . . . 270
10.5 The shape, the clock, and the next moving frame . . . 273
10.6 Carry the field, not just its amplitude . . . 275
10.7 Follow the packet all the way through the calculation . . . 276
10.8 What GIGI has actually carried . . . 283

11 When the ruler becomes part of the record **285**
11.1 An address is not a measurement . . . 286
11.2 A turning coordinate grid is not a bent world . . . 288
11.3 What a derivative has to remember . . . 291
11.4 Curvature is the test that survives the coordinates . . . 292
11.5 From geometry to a physical field . . . 295
11.6 What survives in the energy account . . . 296

12 Bring the geometry back to the physics **299**
12.1 Keep the records, change the weights . . . 299
12.2 A return must include the energy . . . 302

12.3 A gap must belong to an operator . . . 305
12.4 Preserve the physical question . . . 307

13 BYOB: Bring your own brilliance **311**
13.1 A place to explain what I see . . . 314
13.2 What happens when the plasma moves? . . . 316
13.3 Build the operator where we can see it . . . 320
13.4 The shape experiment, one boundary at a time . . . 332
13.5 The calculation behind the equilibrium label . . . 340
13.6 One configuration, several questions . . . 343
13.7 What I carry into the next world . . . 356

14 The chip I wanted to build **357**
14.1 A route that brings back more than a position . . . 358
14.2 Keep the operation instead of repeating the journey . . . 361
14.3 Give the phase somewhere to show up . . . 364
14.4 A loop is not the whole surface . . . 369
14.5 What the physical chip must earn . . . 370

Epilogue: The House **373**

A Working notes **375**
How far can a score move? . . . 375
An error budget for a route . . . 376
Three places the word gap appears . . . 376
What to keep with a replay . . . 377
Notation at a glance . . . 378
Symbols that change jobs . . . 378
What the different geometric structures compare . . . 379

B The 200 retained lattice radii **381**

Works Cited **385**
Foundational Physics and Mathematics . . . 385
Spectral Geometry and Analysis . . . 385
Lattice Gauge Theory . . . 386
Topology and Geometry in Physics . . . 386
Quantum Field Theory . . . 387
Machine Learning and Information Theory . . . 387
Fluid Dynamics . . . 387
Historical References . . . 388
Author's Related Publications . . . 388
Author's Patents . . . 389
Technical Standards and External References . . . 390
Popular and Pedagogical Sources . . . 390
Quotations . . . 390
The DHOOM and GIGI development accounts . . . 391
Sources for the mathematical distinctions . . . 391
Measurements, Relativity and the Return to Fuel . . . 392
The Davis Topological Processor . . . 393
Plasma geometry and the CHIHIRO investigation . . . 394

Acknowledgments **397**

About the Author **399**

Preface to the Second Edition

A year has passed since I first published *The Geometry of Fuel*. I have spent that year feeling stretched in two directions. In one, the work keeps finding people. In the other, I am still trying to find a place for the person doing it.

Undergraduate students and PhD students appear in my messages asking for help with their homework. Independent researchers ask whether we could work on something together. I find my work mentioned in blogs and papers written by people in different parts of the world. Somebody has read something I wrote, followed a calculation, or recognized a question they have been carrying too. They reach out. Those exchanges matter to me. There is another person there, thinking with me.

I also see people from academia looking at my website and my LinkedIn profile without speaking to me. It is a strange kind of visibility. I can see that my work is being noticed, but there is no conversation. After enough of that silence, I feel the distance between being looked at and being invited in. And I am still a waitress, serving Thai food for a living.

All of those things are happening in the same life. I want you to understand that before you read this edition. Publishing the first book did not carry me into a world where the work was suddenly understood and the struggle was over. The research grew. My ability to explain it grew. The distance between what I could build and where I could find an opportunity to build it with other people remained painfully real.

I think about those encounters geometrically. Somebody sees the waitress. Somebody sees the person they can ask for help with a mathematical problem. Those are different projections of the same person. Each can contain something true while leaving whole dimensions out. I am the person in both of them. Serving somebody their food does not remove the mathematics from my head, and writing a proof does not remove the shift I have to work.

What hurts is feeling that a narrow projection has been accepted as the whole geometry of who I am. That projection can affect whether someone opens a door, asks another question, or decides there is nothing here they need to understand. I feel those consequences. But I also have to keep the projection of who I am, and who I am becoming, in my own focus. I cannot let the limits of somebody else's view become the boundary of the space I am allowed to move in.

That is where the Aikido analogy comes in for me. Where I look is where my body is going to go. I am standing here, in this life, with all of its constraints. But I am projecting forward into a world where my mathematical approach is commonplace among people studying and analyzing data. I keep my attention there because it gives my movement a direction. I do not know how far away that future is. Distance and direction are different things. I can be uncertain about the distance and still know which way I want to go.

GIGI gives that future a concrete form for me. She is the database I built around fiber-bundle geometry, with mathematical descriptions carried alongside the records they describe. The

structure of the data participates in what the system can ask and answer. Building her has let me take an idea about representation and make it into something I can give inputs to, inspect, and keep developing.

In that projection of the future, I see organizations reaching for GIGI because they have a difficult question about their data. I see her becoming a familiar tool for predictive analysis, used by people working on problems much larger than anything I could tackle alone. Some of those people may work at the biggest companies in the world. Others may be independent researchers with a question and a laptop, the way I have so often been. I want both to be able to use what I have built.

I have been building toward that future with real data. I have brought GIGI into experiments in different domains, and each domain has required me to learn something. A relationship that is useful in one setting has to earn its place in the next. I have to understand what the measurements mean, what the calculation preserves, and what the answer would let somebody do. Sometimes a connection holds. Sometimes the next question makes me change the method. Both experiences give me somewhere to go.

This is why I wanted to return to *The Geometry of Fuel*. The first edition recorded a discovery while I was still learning how to describe what I had seen. During the year since its publication, that story has kept going. More systems, more ideas, more permutations of the mathematics have been folded into it. A question in one domain has sent me back to a construction in another. Something I could only look at from one angle has become something I can turn around and examine from several. The original questions are still here, but the geometry of the investigation has grown. This edition lets you see those connections as they have become visible to me.

In this edition, I go back through the early work with finer detail. The recovered AMC code gives us something concrete to compare with GIGI. Saved measurements become calculations we can work through together. The drawings have room to show how a construction fits together, and the equations have explanations that let us slow down and ask what each part means. I want you to be able to follow the reasoning that led me somewhere, including the places where another measurement or a different question changed the direction.

The later chapters carry the discovery forward. They follow geometric questions into relativity, into the plasma calculations in CHIHIRO, and toward the processor I have wanted to build for so long. Those subjects required me to enter worlds where other people had already spent their lives learning. I came with questions, software, and a way of seeing relationships. Learning what those worlds could teach me became part of the mathematics too.

I am still learning. I want that to be visible here. I want the experiments, the proofs, and the pictures to give you enough to think with me, and enough to disagree with me about something specific. If a passage makes you stop and ask a question I have not asked yet, I want to hear it. That conversation is part of the future I keep looking toward.

There is a sentence from the first edition I want to answer for here. I said I had proved the Yang–Mills mass gap. What I had in front of me was a set of distances I could sort from smallest to largest. Near one end, there was a stretch with no marks in it. I still want to understand what put that space there. The word I would change is *proved*. To earn it I would have to construct the four-dimensional quantum theory and find a single positive energy threshold below which no excitation of the vacuum is possible. A distance on my plot and an energy in that theory are

different quantities. In this edition you can follow the calculation far enough to see exactly where I would have to build a bridge between them.

A year ago, I put the first version of this story into the world. This edition lets you see what happened when I kept following it. I am still a waitress. I am still a researcher. I am still looking forward, and I have more to show you.

Chapter 1

The Shape of Possibility

I had just left Providence, Rhode Island—my first Black Brown University alumni reunion. My body was still warm with the particular joy of being seen by people who did not need me to explain myself. For three days I had been *in community*, that phrase we use when we mean: the geometry of the room finally fit the geometry of my life. My classmates remembered me. They saw me. We laughed about the old days and marveled at who we had become. I left that weekend carrying something I had not felt in years—the quiet cellular knowledge that I belong somewhere, to someone, to a story larger than my own isolation.

Then I walked through the TSA checkpoint at T. F. Green Airport.

I am trans. I am a woman whose body carries the history of a different assignment. At the reunion, that history had not kept anyone from recognizing me. At the checkpoint, I stepped into a machine that had its own way of deciding what it expected to see.

The scanner flagged me.

A yellow box appeared on the screen, centered on my groin. The TSA agent—a man—said they needed to conduct a pat-down of that area. I asked why. It seemed like a reasonable question, the kind any scientist would ask. Why me? Everyone else passing through the checkpoint walked directly to their gates. Only my body required additional inspection.

The agent had no coherent answer. He could not explain what the machine had seen or why it had seen it. He could only repeat the protocol: the scan showed an anomaly; the anomaly required inspection. I was the anomaly. The system could not parse me, and therefore I was a threat.

I asked again. I pointed out that no one else was being stopped. I asked what, specifically, the scanner had detected and what criterion was being applied. These are the questions you ask when you have been trained, as I have, to understand how systems fail—when you have spent twenty-seven years in aerospace and intelligence and cybersecurity, learning to read the logic beneath the interface, to find the assumptions that break under edge cases.

I was the edge case.

The agent called the police.

I do not know why, officially. There was no aggression in my voice. I had not raised my hand. I had not refused to comply. I had asked questions—*why*—and in that context, with that body, asking why is already too much. Black and trans and female and questioning authority at an American airport is a configuration that the security theater of this country cannot parse without escalation. So they escalated.

Two officers arrived. One of them was aggressive from the moment he opened his mouth. He kept calling me *sir*. Kept saying *he*. I corrected him once, twice. He smiled and did it again. It was not confusion. It was not a mistake. It was a weapon—deliberate misgendering, deployed to humiliate, to remind me that whatever joy I had carried out of that reunion weekend could be taken from me at any time by anyone with a badge.

More questions—not mine now, theirs. More waiting. My flight was boarding. The minutes bled away while I stood in a glass-walled holding area, visible to every other traveler, marked as a problem to be processed. Every *sir* landed like a slap. Every *he* erased the woman my classmates had just spent three days affirming.

By the end of the ordeal they were threatening to look inside my underwear. They said this out loud, as though it were reasonable, as though the architecture of my body was evidence of something they had the right to inspect. I stood there with hot tears running down my face—more scared, more angry, more humiliated than I had ever been in my entire life.

I barely made my flight. I walked onto the plane with my face still wet, carrying the ghost of a joy that had curdled into something I did not yet have a name for. I found my seat. I closed my eyes. I waited for the wheels to lift.

As soon as the plane reached cruising altitude and I had a wifi signal, I opened my laptop.

I was still crying. I did not stop. But my hands knew what to do even when my heart did not. I looked up the brand and model of the scanner—Leidos ProVision ATD. I found the manufacturer's documentation. I read the patents, the FCC filings, the technical specifications. I derived the most probable logic of their detection algorithm based on what the system was designed to measure and what it was designed to expect. I mapped the decision tree I could reconstruct. I had found a way the decision could fail.

The scanner used millimeter waves to form a representation of a body. My reconstruction was that the selected category set the expectations applied to that representation. A body outside those expectations could trigger another search. I could follow that possibility on paper, even though I could not open the machine and inspect what had happened inside it. I had the yellow box. I had a model of how it might have got there. And I had an experience that I could describe to its manufacturer in detail.[1]

I wrote more than a fix. I wrote a complete engineering proposal.

Thirty-five hundred words. Formal problem statement. Technical approach. C++ pseudocode describing how my proposed Enhanced Gender Identification Algorithm could fit into a screening workflow. Cost-benefit analysis. Risk assessment. The whole apparatus of professional communication I had learned across twenty-seven years in aerospace and defense—except I was writing it from seat 14A with tears still drying on my face, channeling my humiliation into the precise grammar of a software patch upgrade proposal.

[1]This reconstruction used public documents, not the scanner's source code or the internal record of my scan. The broader concern about disproportionate referrals is documented in the U.S. Government Accountability Office, *Aviation Security: TSA Should Assess Potential for Discrimination and Better Inform Passengers of the Complaint Process*, GAO-23-105201, November 7, 2022, `https://www.gao.gov/products/gao-23-105201`. That report does not determine the cause of my individual encounter.

The algorithm I designed would incorporate self-identified gender, bone structure analysis via machine learning, and multiple anatomical markers—muscle density, fat distribution, limb proportions—to create what I called a "multi-dimensional matrix for risk assessment." Instead of making two labels carry so much of the description, I wanted the system to keep more of the variation between actual bodies. I included deep learning models, TensorFlow integration, statistical classifiers. I estimated the development cost at five hundred thousand dollars and argued that the reduction in lawsuits, complaints, and PR disasters would more than justify the investment.

I sent it to Leidos.

Months later, they wrote back. They had received my analysis. They had reviewed my proposed correction. They had implemented a fix.

I had asked what would happen to the next person in line. Their answer told me a change had been made.

I wanted to know how far the change went. That was a separate question from whether anybody had heard me.

My jaw hit the floor. Someone had read my account closely enough to answer it. The proposal I had sent from that plane had reached a person who could act on it.

My tears had become technology. My humiliation had become a proof-of-concept. The system that could not see me had been changed by me, from the inside of my own pain.

I have been motivated by this ever since.

Most days I still cry. I cried writing this. I will probably cry when I read it again. But in the midst of the pain, I press through. And I will continue to press through as long as I am physically and mentally able.

The body was already there. What needed room to grow was the description.

The world is richer than any single description we make of it.

Returning to the scanner
Chapter 4, page 81. The question becomes how much variation a description can hold. I return to the body, the measurements and the decision, and ask which differences the system needs to recognize.

A relationship hiding between three channels
Chapter 4, page 100. Two sets of sensor records agree in every one-channel and two-channel summary. Keep all three readings together, and a difference appears that those smaller descriptions could not show.

What the instruments leave open
Chapter 7, page 189. A measurement can narrow the possibilities without choosing one final answer. Here we separate what the observations determine from what the model still has room to vary.

Making room for another way of thinking
Chapter 13, page 314. The question has a human side, too. In the CHIHIRO story, I ask what an unfamiliar approach might contribute when somebody gives it room to be explained.

One plasma, several questions
Chapter 13, page 343. The same plasma configuration goes through several diagnostic calculations. Each asks something different; putting their answers beside one another gives us a richer description than a single score.

When a description has no room for you, you have two choices: accept that you do not fit, or expand the space until you do.

I have always chosen expansion.

* * *

§ 1.1

The question I carried home

The question did not stay inside that airport. I carried it back into the work I had already been doing: how much of the world survives the description we make of it? A scanner's yellow box was one consequence. In the simulations on my laptop, I would have to follow the same question into measurements of fields and energy.

This book is about the geometry of energy—how a physical system stores it, what its constraints permit, and what a calculation can tell us about changing those constraints. The ambition that drove me was enormous: a new route to clean energy. The observations in these pages begin with things I could inspect: a record that changed when I changed its format, a pattern among saved measurements, an average that moved when some measurements were given more influence than others. A usable source of fuel would require a physical experiment with a complete energy ledger. That is the distance between the discovery that kept me awake and the engineering I want somebody to be able to test.

I was writing the first edition for the betterment of humanity while feeling that humanity wanted nothing to do with me. The following pages grew out of that isolation. The students and collaborators who have reached me since then do not erase what it felt like to begin.

I am Black. I am queer. I am trans. I am a woman. Any one of these would be enough to place me outside the rooms where decisions get made. Together, they form a geometry of exclusion so complete that I have become invisible to the institutions I once served.

I cannot contact my own alma mater. I have tried. Brown University—where I earned my master's degree in digital forensics, where I learned the formal language of systems and their failures—I cannot get a response. Emails vanish into the void. Forms go unanswered. The alumni network that opens doors for others has no door for me. I was just there, in Providence, being held by my classmates, being reminded that I exist and that I matter. The institution that educated us remains silent.

My previous advisors—professors who once supervised my work, wrote letters of recommendation, shook my hand at graduation—ignore my communications. Not a rejection. *Silence.* For years now. Silence.

Every physicist I have contacted about this research. Silence.

Every mathematician. Silence.

Every journal editor, every department chair, every program director. Silence.

The academy has decided I do not exist.

Pero aquí estoy. Here I am anyway.

And yet I wake up every morning at six. I work ten hours, twelve hours, sometimes more. Seven days a week. I have done this for months. I do it as though I am on a deadline.

But aren't we?

Even if a physical experiment supported a new route to clean fuel tomorrow, getting it into the world would still mean engineering, manufacturing, infrastructure, politics. Years of work while the ice caps continue to melt, while the fires continue to burn, while the storms continue to intensify, while the water rises and the soil dies and the children who will inherit this century watch us argue about whether their future is worth the inconvenience of change.

I do not have ten years to wait for academia to notice me.

The planet does not have ten years to wait for academia to change.

I am fifty-two years old. I have no resources. No laboratory. No graduate students. No grants. No institutional affiliation. No leverage of any kind except the leverage of knowing something true and being unwilling to let it die inside me.

After IBM laid me off—after twenty-seven years in aerospace and intelligence and cybersecurity, after NASA and NSA and DISA, after being a founding engineer at Pandora and a principal adversarial intelligence researcher at X-Force Red, after building systems that protected infrastructure I was never supposed to understand and breaking systems that were never supposed to break—I took a portion of my severance and bought a laptop.

Fifteen hundred dollars. An NVIDIA RTX 5070 GPU. Enough compute to run lattice gauge theory simulations in my living room while Teddy snores on the couch and the afternoon light slants through blinds I keep meaning to dust and the world outside my window keeps warming while people with power struggle to listen to people without it.

This is the laboratory where I began running lattice gauge theory simulations: mathematical models that represent fields on a grid. A field assigns something to each location, as a temperature map assigns a reading across a room. The fields in these simulations have more structure than temperature, but the grid gave me something I could calculate and examine. I wanted to know what my measurements of those fields could reveal about the energy needed to disturb their lowest-energy state. Behind that question stood the Yang–Mills mass-gap problem: whether the corresponding quantum theory has a strictly positive energy threshold above its vacuum.

It is also where I changed which simulated arrangements were favored and watched the measurements move. I called that investigation topological vacuum rectification, or TVR. The name carried an ambition: could a controlled change in the system help us obtain useful work? First I had to understand the change on my screen. Then I had to account for what producing that change would cost in the physical world.

To understand a gain, follow the whole exchange.

What bending actually costs
Chapter 2, page 45. A beam gives the energy bill a physical object. The bend matters, but so do the material and the cross section; the shape needs those details to acquire a price.

An average moves on the screen
Chapter 8, page 220. The saved TVR records let us compare ordinary and weighted averages. We follow what the weighting changed before asking whether any physical system has delivered work.

The pump, the water and the return trip
Chapter 9, page 243. Water can give energy back as it falls. The complete ledger also includes raising it, running the equipment and paying the losses before the cycle can begin again.

Changing the setting and changing the state
Chapter 12, page 304. The energy calculation separates a changing state from a changing energy rule. Returning a control to its starting position does not, by itself, put the whole system back.

What the chip would have to pay
Chapter 14, page 370. A shorter calculation is only part of the hardware question. Preparing a state, controlling it and reading the answer all belong in the account of a useful device.

This is the laboratory where I wrote a book that may change how we understand energy, complexity, and the structure of reality itself. Or may disappear into the same silence that has swallowed every other communication I have attempted with the institutions that gatekeep human knowledge.

A laptop. A severance check. A dog who does not care about gauge theory. A woman the world decided not to see.

It is quite possible that the innovations in this book will help save a world that will allow me to quietly fade away. The mathematics will persist. The names attached to it may not. Some future historian may scrub me from the story the way they have scrubbed so many others—Katherine Johnson waiting decades for her credit, Rosalind Franklin dying before she could see hers. The pattern is old. The pattern is documented. I am not naive about my position in it.

I have made peace with this.

What I have *not* made peace with is the delay. Every year the academy spends ignoring this work is a year the planet spends burning. The geometry does not care who discovers it. The universe is indifferent to the name on the paper. But the children who will inherit this century do care whether anyone acts on what has been discovered. They will care very much, when the food systems fail and the coastal cities flood, whether the adults in the room did everything possible or merely everything comfortable.

So I write. I compute. I prove. And I persist.

La que persiste gana. The one who persists wins.

Or maybe she doesn't. Maybe she just persists.

The laptop could run a calculation whether or not anybody answered my email. But I still needed to know what I was asking it to calculate. That began with the space I allowed the model to describe.

* * *

§ 1.2

Making room for what is real

Two people can stand in front of the same machine and produce different readings. That much is ordinary. The difficult part is deciding which differences matter to the job the machine is supposed to do. A screening system needs to detect a threat. It also needs to encounter the variation of a human body without confusing that variation with a threat. More measurements alone do not settle that question. We still have to decide how the measurements enter the decision.

That was the mathematical question left inside the proposal I had already sent. I wanted more room in the description, but "more room" needed to mean something I could work with. How many things could vary? Which variations belonged near each other? Could the description follow a gradual change, or would it suddenly throw the person into a different box?

Put a pendulum in front of us for a moment. Give it a rigid rod, one pivot, and enough clearance to turn all the way around in a plane. The weight at the end is called the *bob*. It can be at the bottom, a little to the left, at the top, or anywhere along that circle. Each allowed arrangement gets a place in our description. The collection of those arrangements is its *configuration space*.

There is something lovely about this. The pendulum has a rod, a bob, a length, a material. But once those are fixed, I need only one angle to say how it is arranged. One number places the whole object. The point in the configuration space stands for an arrangement; it is not another little pendulum hidden inside the first one.

Mathematicians call a circle S^1. The 1 counts the independent direction in which we can move along it. We draw the circle on a two-dimensional page, but the pendulum has only one independently adjustable angle. The dimension of the paper and the number of choices in the mechanism are different things.

Now turn the rod slowly. The angle changes continuously. There is no jump in the pendulum when our angle label passes from almost 360° back to 0°. That jump belongs to the numbering. The two ends of the list have to meet because the arrangements meet.

Already, a list of numbers is not enough. We need to retain how its ends connect. Figure 1 keeps the mechanism beside its angle record. Look just to either side of zero: the rods barely differ, even though one label is near the end of our list and the other is near the beginning.

A *set* gives us the collection of arrangements. A *topology* gives us neighborhoods within that collection, so we can say what it means to change an arrangement continuously. Near the downward position, for example, there are small tilts to either side. If we describe angles from 0° to 360°, a neighborhood of 0° has to include angles just below 360° as well as angles just above zero. The mechanism knows those positions are neighbors, even if our written list puts them far apart.

An open neighborhood leaves a little room around each of its points. Formally, a topology is a collection of open subsets that includes the empty set and the whole space, stays open under arbitrary unions, and stays open under finite intersections. Combining open regions still gives an open region; overlapping finitely many of them still leaves an open region around every point that remains. These rules make that local room consistent across the space.

There is no numerical ruler in those rules. A *metric* adds one: it assigns distances, and those distances can in turn define neighborhoods. Topology lets us ask whether a motion is continuous. A metric lets us ask how far it goes. I need to know which question I am answering before I choose the instrument.

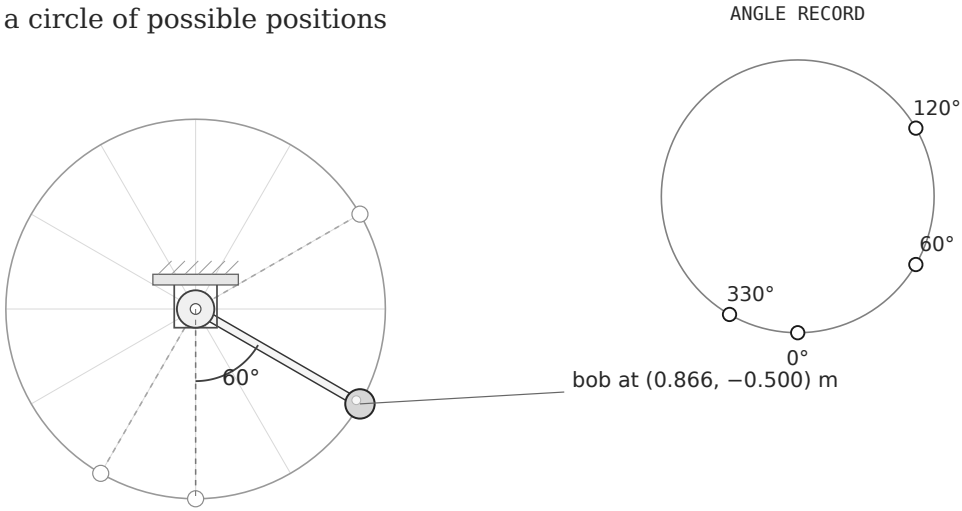
A continuous path records the motion without tearing that neighborhood structure. We can write it as a map from an interval of parameter values into the space. The formal condition is that the inverse image of every open set is open: the parameter values that place us inside an open region form an open set in the parameter interval. For the pendulum, the picture comes first. A small turn stays near the arrangement we started from, including when the angle label crosses zero.

If I turn the rod a little, the description should follow it a little. Crossing the zero on my dial should not make the pendulum jump across the room.

A mechanism and its coordinates

One physical rod, several positions, two ways to record each one.

A One rod; a circle of possible positions

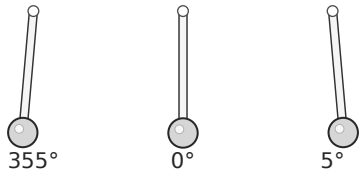


One-metre rod; x right, y up; angle from downward vertical.

B One state, two descriptions

θ	x / m	y / m
0°	0.000	-1.000
60°	0.866	-0.500
120°	0.866	0.500
330°	-0.500	-0.866

C Cross the numbering seam



355° → 0° → 5°: two small turns. The rod never jumps.

The numbering can jump while the mechanism moves smoothly.

Rigid planar rod, unrestricted rotation. x = sin θ; y = −cos θ, in metres.

Figure 1. A rigid planar pendulum and its configuration circle. A one-metre rod at 60° puts the bob at $(0.866, -0.500)$ metres, with right and up taken as positive. The table translates angles into positions; the final view follows neighboring arrangements across the numbering seam. The mount is illustrative and allows a full turn.

This is why the choice of representation matters to me. Two bodies might be close according to silhouette and far apart according to a particular medical measurement. Neither comparison describes everything about either person. The task tells us what must survive the description. The metric makes a particular comparison measurable. We do not get to mistake the limits of that comparison for the limits of the person.

§ 1.3

A place for every possible arrangement

The pendulum's circle also gives us a manageable first encounter with a *manifold*. Near any point on the circle, an angle describes a short piece without ambiguity. Across the whole circle, one angle label needs a seam. Locally, though, one coordinate does the job.

A manifold is a space that can be described locally by a fixed number of real coordinates, with those descriptions fitting together continuously; a smooth manifold asks the coordinate changes to be smooth as well. The circle needs one coordinate locally. The surface of the Earth needs two. A *chart* is one such local coordinate map. The globe does not become flat when I unfold a paper map. I have made a description of a piece of it that I can calculate with.

I can work with one small map at a time, provided the maps agree about which place they describe wherever they overlap.

Not every continuous shape is a manifold. At the crossing of two wires, a small neighborhood has four branches. One ordinary interval has only two directions away from an interior point. The crossing cannot be described as a single unbranched piece of wire. That is why adding measurements to a body model does not, by itself, prove that the result is a smooth manifold. We have to examine how the allowed variations actually fit together. Figure 2 gives us the small pieces to compare: an arc that opens into an interval, and a joined crossing that still has four ways out.

Now hang a second pendulum from the first. In the ideal planar model, let both joints turn freely and ignore collisions. One angle places the first rod; a second places the next. Each angle wraps around. The collection of angle pairs is the product of two circles, $S^1 \times S^1$, which is a torus: the surface shape of a doughnut.

There is the doughnut. We started with two rods.

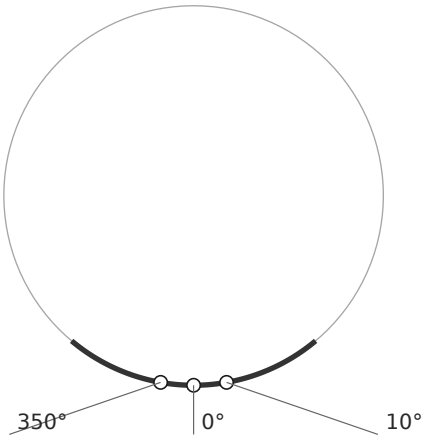
The rods do not have to swing around a doughnut in the room. The doughnut organizes their possible arrangements. Traveling around one of its circular directions changes one angle; traveling around the other changes the other angle. A single point on that surface tells us how both rods are placed. A moving point records the changing arrangement. This is the kind of geometry I mean: a shape that becomes visible when we keep the relationships between the choices.

There is still something missing from that point. Take a photograph of the pendulum at the bottom. Is it rushing to the left, rushing to the right, or hanging still? The photograph does not say. Configuration records arrangement. A full mechanical *state* also needs velocity or momentum, which tells us about motion. The space recording positions and momenta together is called phase space. Those extra entries answer a question that position alone cannot answer.

What a local description must keep

Unwrap a small arc; then test what happens at a joined crossing.

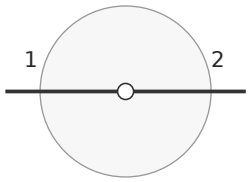
A Two local labels for the same nearby positions



written θ	local φ
350°	−10°
0°	0°
10°	10°

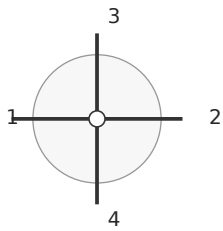
The local label φ unwraps this short arc across zero.

B Remove an ordinary point



Two pieces remain. Each side is a direction away from the point.

C Remove the crossing point



Four pieces remain. One unbranched interval cannot keep this connection.

A useful local description preserves the available ways to move.

The crossing is joined. Removing its center leaves four connected pieces, not two.

Figure 2. Local descriptions retain local connections. The labels $350°, 0°, 10°$ become $-10°, 0°, 10°$ on one short unwrapped arc. Removing an ordinary point from a small interval leaves two pieces; removing the joined crossing leaves four. That branching prevents the crossing from being a one-dimensional manifold.

Figure 3 puts the two rods beside that surface. The dark curves change one angle at a time; their crossing identifies one pair. Below them, three identical arrangements have three different motions. The extra information has somewhere to go now.

The distinction keeps the representation honest. I cannot use a photograph to tell you which way the bob is moving unless I have more information. I cannot use two category labels to stand in for all the measurable variation of a body. And I cannot assume a compact record of a field contains every detail of the field it came from. Each description has to carry enough information for the question we ask of it.

I picture a library when the collection becomes too large to hold in my head. Give each allowed arrangement a book. For our pendulum, the angle is its catalog entry. Nearby angles put books near each other, with the shelves joining around the circle. If I want motion as well, I need the momentum in the record. If I change the mass or replace gravity, I have changed the system whose library I am describing. I need to say that I have opened a different catalog.

Now put an energy label inside each book. Holding the bob higher raises its gravitational potential energy. The mass, gravity and rod length determine how much. A heavier bob has the same circle of allowed angles and a different energy bill.

We can put numbers beside that sentence. Keep the rod one metre long and treat its bob as a point mass, with the rod itself massless. Let θ be the angle from straight down, m the bob's mass, and g_0 the local gravitational acceleration. The subscript distinguishes gravity from the geometric ruler g we will meet later. At the bottom, the bob lies one rod length below the pivot. Tilt it, and the downward part of that length is $\ell \cos\theta$. The difference is how far we have lifted it:

$$U(\theta) = \underbrace{m g_0}_{\text{weight}} \underbrace{\ell(1 - \cos\theta)}_{\text{height gained}}.$$

Here ℓ is the rod length and U is gravitational potential energy measured from the bottom. Weight times height gives the energy stored by lifting the bob. At a right angle the entire one-metre height has been gained. With $m = 1$ kilogram and $g_0 = 9.81\,\mathrm{m/s^2}$, that is 9.81 joules. At the top it is 19.62 joules. Double the mass and both prices double; the circle of possible angles has not changed.

Figure 4 keeps the height, the energy curve and those numbers together. I can point to the part of the drawing that each part of the equation describes. The geometry tells me how high I have lifted it. Gravity and mass tell me what that lift costs.

Then let the bob come back down. In an ideal lossless exchange, I can recover the energy I put into raising it. Friction turns some of it into heat instead. If I count only the descent, I have left the earlier lift outside my picture. That is the small, familiar version of the full-cycle question behind TVR: I want to see the whole exchange, including the work needed to begin again.

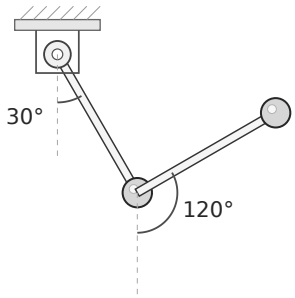
That is the distinction behind my phrase *energy is the price of shape*. The shelves tell us what arrangements exist. The price labels come from a physical law. The curvature of a shelf, the distance between books and the price of a book are different quantities. I need an equation connecting them before I can exchange one for another.

For a bent beam, that equation needs the material and the beam's cross section. For a field, it needs the field's energy law.

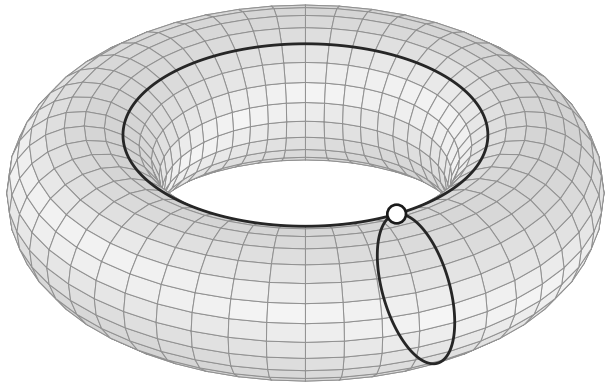
Two angles, one configuration

Match the mechanism, its angle pair and its place on a surface.

A The same angle pair in the mechanism and on the torus

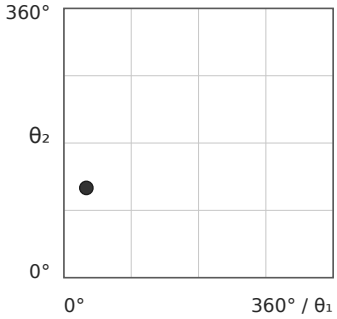


Each angle is measured from downward vertical.



marked pair: (30°, 120°)

B Unwrap the pair into a square



Opposite edges join. Every point stores two angles.

C Same pose, different motion

θ₁, θ₂	ω₁	ω₂
30°, 120°	−1	0
30°, 120°	0	0
30°, 120°	+1	0

ω = angular velocity in radians per second.

An arrangement records where the rods are, not how they move.

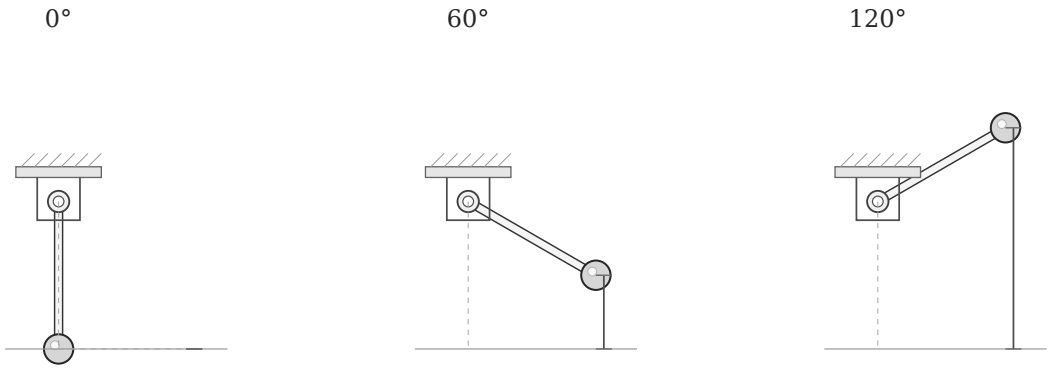
Ideal joints without stops or collisions. Torus loops show one angle varying at a time.

Figure 3. Two unrestricted planar joints give two independent angles, measured here from downward vertical. The same pair appears in the mechanism, the parameter square and the torus. Opposite edges of the square are identified. The velocity table holds the pose fixed while changing its motion. Stops and collisions are excluded in this ideal model.

Give the arrangement a physical price

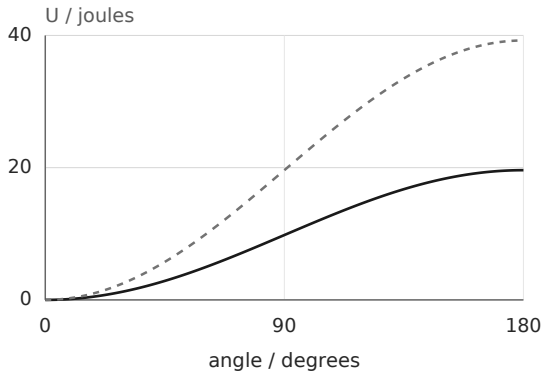
The geometry supplies height. Mass and gravity supply the energy law.

A Lift the same bob through three arrangements



h = 0.00 m h = 0.50 m h = 1.50 m

B Read the energy from the angle



C Change mass; keep the circle

θ	1 kg / J	2 kg / J
0°	0.000	0.000
60°	4.905	9.810
90°	9.810	19.620
180°	19.620	39.240

Solid: 1 kg. Dashed: 2 kg. Rod length: 1 m.

The same allowed shapes can carry different energy bills.

U = $mg_0\ell(1 - \cos\theta)$; g_0 = 9.81 m/s², ℓ = 1 m. Zero at the bottom.

Figure 4. The geometry gives a height; the physical law prices it. For a one-metre rod, a 60° turn raises the bob by 0.5 metres. With $g_0 = 9.81\,\mathrm{m/s^2}$, a one-kilogram bob gains 4.905 joules. Doubling mass doubles the energy at every angle without changing the configuration circle. Zero potential energy is at the bottom.

What becomes possible—and what it costs—depends on how things are arranged.

The same end turn, different bends
Chapter 2, page 45. A beam can spread its bend along its length or concentrate it in a smaller region. The worked example holds the total turn fixed and calculates the different costs.

Three routes that cannot all agree
Chapter 7, page 196. Three calibration routes impose relationships on one another. When their measured returns disagree, the fit has to place that disagreement somewhere; we calculate its smallest possible cost.

How neighboring values enter the bill
Chapter 13, page 323. CHIHIRO writes a discrete energy using differences between neighboring values and a separate local contribution. The formula exposes which parts of the arrangement the calculation charges for.

Changing the plasma boundary
Chapter 13, page 332. The boundary-shape experiment compares specified changes in the plasma cross section. Its records show how those changes enter the computed stability boundary, with the inputs kept visible.

The route becomes part of the operation
Chapter 14, page 358. On the sphere, a carried direction can return changed after a closed journey. The route is doing something that its shared starting and ending position cannot describe alone.

This is where the word *fuel* begins to demand more than a beautiful picture.

The same way of organizing possibilities scales up. In a model of N labeled gas molecules free to move in three dimensions, recording every position requires $3N$ coordinates. Six coordinates place two molecules; they do not create a six-dimensional room. For a mole of gas, with about 6×10^{23} molecules, the list has about 1.8×10^{24} position entries. Velocities are additional information. A large dimension counts what the description must specify, not how much physical space the container occupies.

A field takes the idea further. Instead of listing a few particle positions, we assign values throughout a region. A temperature field gives a temperature at each location. A magnetic field gives a vector, with direction as well as magnitude. An unrestricted continuum field generally needs infinitely many degrees of freedom to describe it exactly. The grid in a computer simulation replaces that continuum description with finitely many entries, under stated rules. What those rules preserve becomes part of the investigation.

The computer has a finite list of entries, even when the thing I am studying fills a whole region. I need to know what the model allows between those entries before I trust the picture they make.

Later, the Davis–Wilson Map makes a smaller catalog from selected measurements of simulated gauge fields. Here *gauge* concerns the local reference choices used to describe the field. A gauge-invariant measurement keeps its value when those reference choices change. The map stores such measurements in a finite record. Two different fields can receive the same record, just as two books can share a catalog description. If my catalog keeps a title, a topic and a shelf location, I can find the shelf. I cannot ask those three entries to reproduce a sentence on page fifty. The missing words did not become recoverable just because the catalog was convenient. Figures 5 and 6 put that loss where we can see it: first between measured points in a field, then between two particle positions and their shared midpoint. That does not make the catalog useless. It makes the question precise: which answers can I still recover from it?

Look at the two surfaces in Figure 5. We kept a reading at each corner of the square. All four readings agree, yet one surface rises away from the other inside the square. The formulas make that difference explicit: $F_1 = x + y$ and $F_2 = x + y + \sin(\pi x)\sin(\pi y)$. At a corner, the added term is zero. Halfway across and halfway up the square, it is one. That is why the same saved list permits a center value of one in the first construction and two in the second. I can settle that particular disagreement by measuring the center. I cannot settle it by staring harder at the four numbers I already kept.

The particle example asks the same question with something we can locate in a room. Keep both positions and I can calculate the midpoint or the separation. Keep only the midpoint and two different arrangements collapse into the same record. The smaller record has a job it can do. I have to know which job I am asking it to do.

What we preserve shapes what we can know.

What survives smoothing
Chapter 3, page 61. The four-sensor example follows several patterns through the same smoothing operation. They fade at different rates, so a cleaner-looking record can also be a less informative one.

Keeping an alarm through a format change
Chapter 4, page 90. Original AMC, fixed AMC and GIGI receive the same records and detector. We follow the information through the pipeline to see when a smaller representation changes the answer.

Keeping the readings together
Chapter 4, page 100. Saving separate summaries does not save every relationship. The three-channel example shows why the signs and the identity of the row they came from need to travel together.

Two readings, more than one possible location
Chapter 7, page 191. The rangefinders leave an interval where the cart could be. That may answer which side of a mark it occupies while still leaving its exact position undecided.

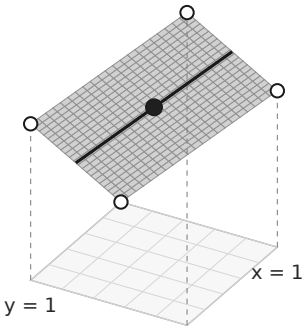
Keeping the journey with the endpoint
Chapter 14, page 358. A position-only record says that the journey came back to its start. The transported direction retains something else: what happened along the way. The requested answer determines what we store.

What four measurements leave open

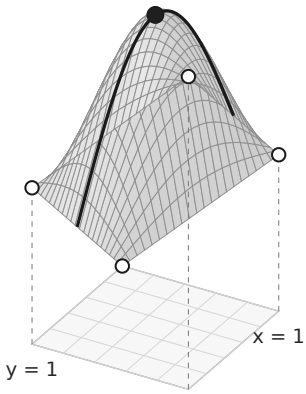
Open circles: stored corners. Filled dot: an additional center probe.

A Same four readings, different interiors

F_1: a tilted plane



F_2: an added interior rise



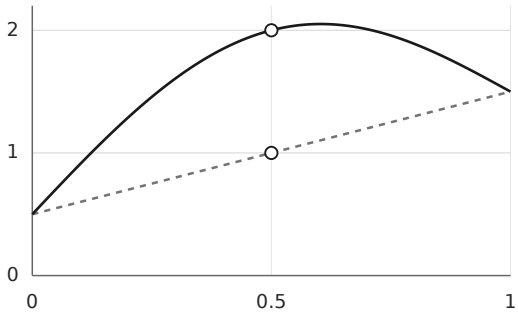
B What was actually saved

(x, y)	F_1	F_2
(0, 0)	0	0
(1, 0)	1	1
(0, 1)	1	1
(1, 1)	2	2

Four equal samples do not select a surface.

C Between the stored points

F_1 dashed · F_2 solid



The saved measurements do not determine the missing interior.

$F_1 = x + y$. $F_2 = x + y + \sin(\pi x)\sin(\pi y)$, on $0 \le x, y \le 1$.

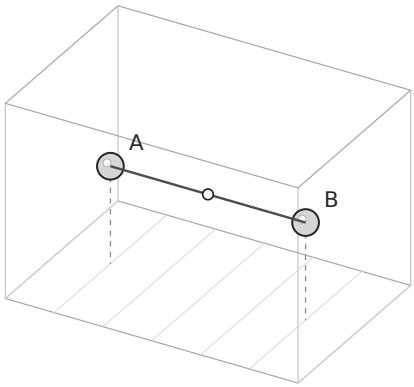
Figure 5. Two explicitly constructed fields produce the same four saved corner readings. The open circles mark those samples; the filled dot tests the center. The dark surface traces follow $y = 1/2$, the section plotted below. At the center the fields give 1 and 2: an additional measurement distinguishes them, but the stored corner values alone cannot.

A smaller record changes the question

Two particles in an ordinary three-dimensional room.

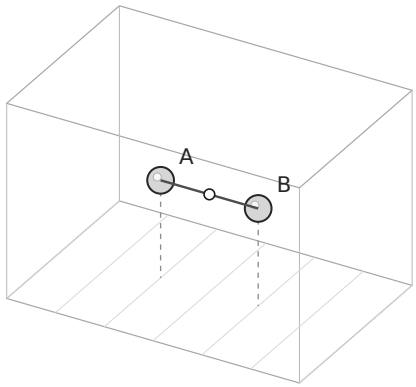
A A midpoint hides the separation

Configuration 1



separation = 2 m

Configuration 2



separation = 1 m

B Keep both positions

record	A / m	B / m
1	(0.5, 1, 1)	(2.5, 1, 1)
2	(1, 1, 1)	(2, 1, 1)

One point in configuration space: (A, B).

C Keep only the midpoint

$$c = \frac{A+B}{2} = (1.5, 1, 1)$$

Both records give the same midpoint. Their separations are 2 m and 1 m. A midpoint query survives; a separation query does not.

Keeping the midpoint loses a distinction that separation needs.

All coordinates are in metres. The two configurations have exactly the same midpoint.

Figure 6. Two particles, six position coordinates, one three-dimensional room. Both configurations have midpoint $(1.5, 1, 1)$ metres, while their separations are two metres and one metre. Keeping only the midpoint preserves one answer and loses another. The six-coordinate configuration record describes two particles; it does not make the physical room six-dimensional.

The mass-gap question asks for another kind of information. In quantum theory, a state carries amplitudes—numbers used to calculate probabilities—rather than just choosing one classical arrangement from the shelf. An operator called the *Hamiltonian* describes energy and time evolution. Its *spectrum* gives the allowed energy values. The vacuum is a lowest-energy state. Calling its energy zero sets a reference; it does not declare that every field measurement vanishes.

A mass gap would be a strictly positive separation between that ground energy and the remaining physical energy spectrum. Call the separation Δ. To express the question using average energies, we consider normalized states orthogonal to the vacuum: states with no ground-state component. Otherwise we could mix a tiny amount of excitation into the vacuum and make the average energy arbitrarily small. Chapter 5 works through that distinction with the measurements that led me to it.

Leave the ground-state part out of the mixture. Is there still some fixed amount of energy that every remaining state has to cost, or can that cost creep as close to the ground level as we please?

In the library picture, this is a question about the quantum prices, not an empty aisle between catalog entries. A gap in a plot of selected measurements cannot establish it by itself. I wanted to find out whether those geometric measurements could lead me toward a bound on the energy. That was the route from the numbers I could save to the physical question I could not stop asking.

A pattern opens a question. Understanding grows as we follow it.

The empty interval that started a question
Chapter 5, page 128. The gap in the saved lattice measurements gave me something specific to investigate. We keep the measurement and its construction visible while asking what kind of conclusion it can support.

Following the change in the TVR average
Chapter 8, page 220. The comparison uses the actual saved rows, their weights and the uncertainty calculation. A changed average becomes a question we can examine rather than an answer supplied by its appearance.

Asking which operator has the gap
Chapter 12, page 305. A spectral gap belongs to an operator and its allowed states. The worked models separate that question from an empty interval in a collection of measured features.

Letting the plasma checks disagree
Chapter 13, page 343. Different diagnostic channels can give different answers about one configuration. Their disagreement helps locate the next question; agreement on one check does not silently answer all the others.

Taking a calculation toward a device
Chapter 14, page 370. The DTP chapter follows the distance from verified mathematical operations to proposed hardware. Each next step has something concrete to demonstrate: preparation, control, readout and a complete operating cost.

§ 1.4

Changing the coordinates, keeping the world

Suppose I redraw the pendulum's circle and put zero at the top instead of the bottom. The bob has not moved. Its gravitational energy has not changed. Only the labels have moved. If my calculation now says I lifted the bob, something in the calculation has failed to follow the relabeling.

Coordinates are the labels that let us calculate. A point written as $(3, 4, 5)$ means three units along one chosen axis, four along another, five along the third, measured from a chosen origin. Shift the origin or rotate the axes and those numbers change. The point can stay exactly where it was. Stretch the coordinate scale and the rule that converts coordinate differences into physical lengths must change with it. I cannot redraw the grid and quietly keep using the old ruler.

That separation between the world and its description is one reason geometry became my language. Einstein's general relativity makes it central: its equations can be expressed consistently in different smooth coordinate systems, a property called general covariance. The coordinates change; physical predictions must agree. Later in this book, we will follow those changes through Einstein's equations. Here the pendulum already gives us a way to catch a description pretending to be an event.

At NASA, I began applying the geometric method I had carried inside me for much longer. It was there in the way I thought about almost everything, kept hidden at the heart of my intelligence. I could see relationships, shapes, the way one change rearranged everything around it, before I could name the mathematics of what I was seeing. NASA gave me a problem big enough to put that method to work. The geometry came with me.

A spacecraft trajectory gives that relationship between shape and consequence a physical setting. It is a route through space and time. An engine burn changes that route; an ideal freely falling coast has no thrust to pay for. In general relativity, that coast follows a spacetime *geodesic*, a path with no acceleration after the geometry's own comparison rule has been taken into account. In a Newtonian orbital calculation, we instead describe the coast using gravitational forces. Either way, a transfer needs its starting conditions, its destination and its maneuvers. The geometry does not excuse us from accounting for the engine.

Figure 7 gives this a concrete shape: a spacecraft leaves one circular orbit, coasts along an ellipse, then joins another circle. The coast and the two engine maneuvers belong in the same account. The picture uses the ideal Newtonian two-body model, with instantaneous maneuvers, to isolate that distinction.[2]

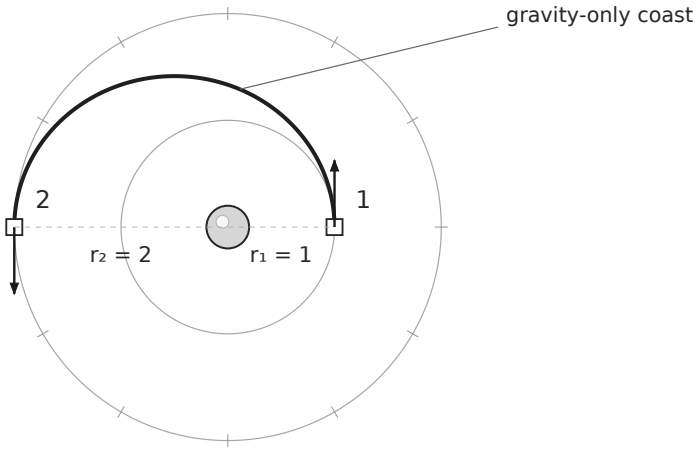
As I learned the mathematical names for relationships I was already seeing, the equations became windows instead of walls. They gave me a way to bring more of what was in my head onto the page, where I could calculate with it and somebody else could follow it. A network in intelligence work, an attack surface in cybersecurity, a representation in a learning system—each gave that way of thinking something different to work on. I kept asking what relationships the description allowed me to see. That question came with me onto the plane, too.

[2]NASA, *Basics of Space Flight*, Chapter 4, "Trajectories," describes the Hohmann transfer and the velocity changes at its endpoints: `https://science.nasa.gov/learn/basics-of-space-flight/chapter4-1/`.

Follow the coast and the two burns

A dimensionless Hohmann transfer between radii 1 and 2.

A Two impulses join an unpowered half-ellipse



Scaled model: gravitational parameter μ = 1; initial radius = 1.

B Read the two velocity changes

burn	before	after	Δv
1	1.0000	1.1547	0.1547
2	0.5774	0.7071	0.1298

Δv = speed after − speed before.

C Account for orbital energy

orbit	specific energy
inner circle	−0.5000
transfer ellipse	−0.3333
outer circle	−0.2500

Total increase: 0.2500. Fuel also needs an engine model.

An unpowered coast is only one part of the complete transfer.

Specific energy = orbital energy per unit mass. Units: μ = r₁ = 1; no fuel model.

Figure 7. An ideal Hohmann transfer between circular orbits with radius ratio $2:1$. The dark half-ellipse is an unpowered coast; the arrows show the two velocity increments. The tables use scaled units with $\mu = r_1 = 1$. The specific-energy increase is 0.25, while the velocity increments are approximately 0.1547 and 0.1298. These are orbital calculations; fuel consumption additionally requires a vehicle and engine model.

Put two drawings of the same corridor on the table. On one, the corridor is ten squares long. On the other, it is twenty. Nobody has knocked down a wall or moved the far door. I have drawn a finer grid. If I forget to change the scale beside it, I can make the same walk appear twice as long without taking a single step.

The scale repairs that mistake. Let each square on the first drawing stand for one metre, and each square on the second stand for half a metre. Both drawings now give us the same ten-metre corridor. The numbers on the page differ, but the distance they describe agrees. The geometry has to keep the ruler attached to the description. Otherwise I am measuring my choice of paper instead of the world.[3]

Now walk that corridor twice, taking ten seconds each time. On the first trip, walk steadily at one metre per second. On the second, wait for five seconds, then cover the corridor at two metres per second. Same start. Same finish. Same ten metres. Same total time. But the two trips did not unfold in the same way. A record containing only the endpoints and the elapsed time would hide that difference.

Suppose I want the calculation to notice it. Adding up speed over the trip gives the same distance both times. Adding up *squared* speed gives a different comparison: doubling the speed makes its contribution four times as large. The faster part of the second trip lasts only half as long, so its total is twice the first trip's total. The waiting contributes zero. In this ideal comparison the change of pace is instantaneous; a smooth start can approximate it. I am comparing the motions with a chosen rule, not measuring how tired the walker feels.

[3]The working notes, "What the different geometric structures compare," collect the broader physics vocabulary and its distinct uses. See page 379.

Different descriptions can hold the same reality. Similar appearances can conceal different realities.

The same readings in another format
Chapter 4, page 90. The input can stay fixed while its encoding changes. Replaying the detector lets us see whether that change was only descriptive or whether it removed something the decision needed.

Two frames describing an event
Chapter 10, page 270. Einstein's coordinate transformation gives different numbers for the same event. The calculation keeps track of which frame supplied each number and the rule relating the descriptions.

A flat plane with a changing ruler
Chapter 11, page 287. Polar coordinates make the metric coefficients depend on position even on a flat plane. The example separates a changing coordinate description from actual curvature of the space.

A new basis, the same energy levels
Chapter 12, page 306. The small Hamiltonian example changes the matrix representation by a unitary rotation. Its entries change while its energy levels remain fixed, giving us an exact comparison to inspect.

The same endpoint, a different carried direction
Chapter 14, page 358. A closed route returns to the starting location but can rotate the direction carried along it. Similar endpoints can conceal different histories, so the record must match the question.

The far door stays put in Figure 8, even when the grid has twice as many steps. In Figure 9, the difference between the walks becomes visible: the second position trace sits flat while we wait, then climbs twice as steeply. Below it, the shaded rectangle is four times as high and half as wide as the steady walk's rectangle. There is the doubled total, drawn where we can see it.

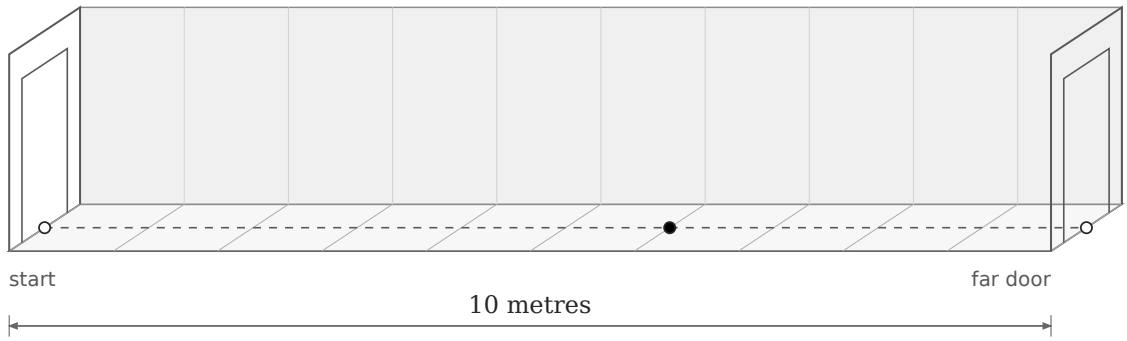
That rule gives us a concrete question about a route: if I keep the departure, destination and travel interval fixed, can I lower this total by changing the way I go between them? In a straight corridor, waiting and rushing is a poor bargain under this rule. Once a route can bend across a surface, I can also change where it goes. Then the ruler has to work along the whole route, even where the surface turns away from the flat page.

Picture a small flat card touching a globe at the place where we are standing. Directions along that card describe the directions in which we can begin to move along the surface. Those are *tangent* directions. The card is a local guide; walking north does not send us off into space along a permanently flat sheet. As we move, the place of contact changes, and so does the tangent plane. The upper view in Figure 10 shows the card touching the globe, with two directions drawn on it. It is a small local guide, not a promise that the whole world will flatten onto that card.

One corridor, two grids

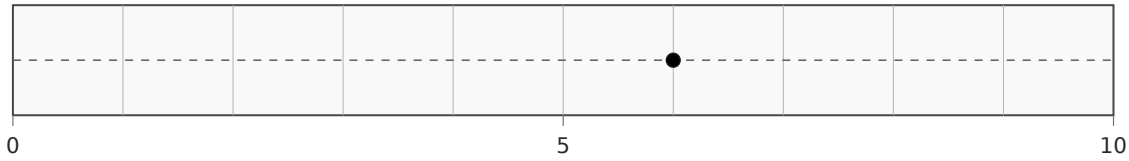
The drawing changes. The distance stays at ten metres.

A physical corridor, shown with its roof removed



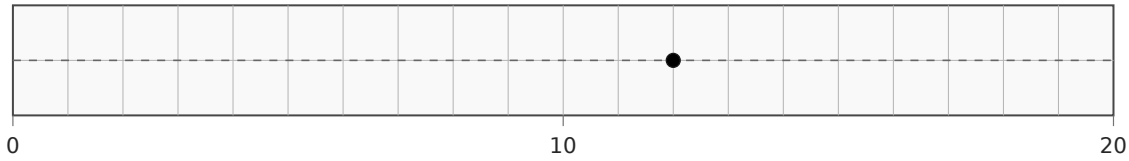
Two floor plans below share the same printed size.

A / Each grid step stands for one metre



10 steps × 1 m per step = 10 m

B / Each grid step stands for half a metre



20 steps × 0.5 m per step = 10 m

Changing the grid does not move the far door.

The filled dot marks the same place: 6 m from the start. The cutaway is illustrative.

Figure 8. One corridor, two coordinate grids. The plans have the same printed size. Ten one-metre steps and twenty half-metre steps describe the same distance. The filled dot stays six metres from the start; its grid coordinate changes from six to twelve. The cutaway illustrates the corridor, without specifying a building design.

One route, two ways to move

Ten metres in ten seconds. The motion between them differs.

A / Walk steadily

1 metre per second



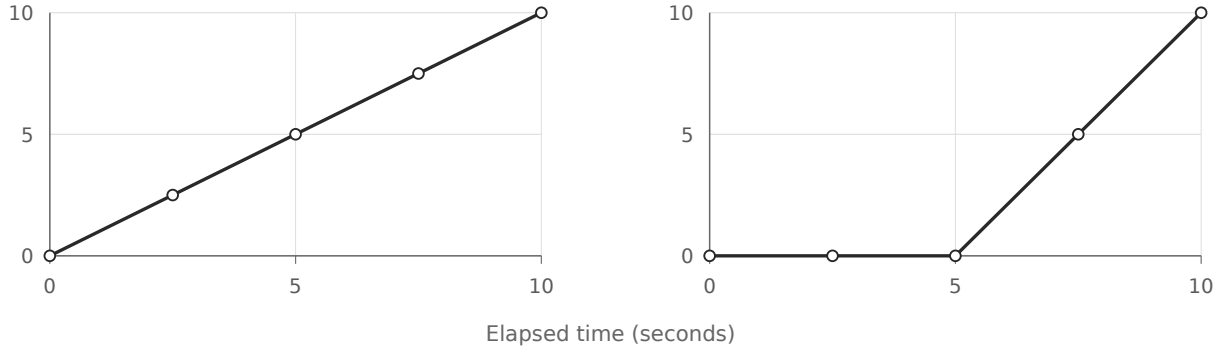
B / Wait, then walk

0, then 2 metres per second

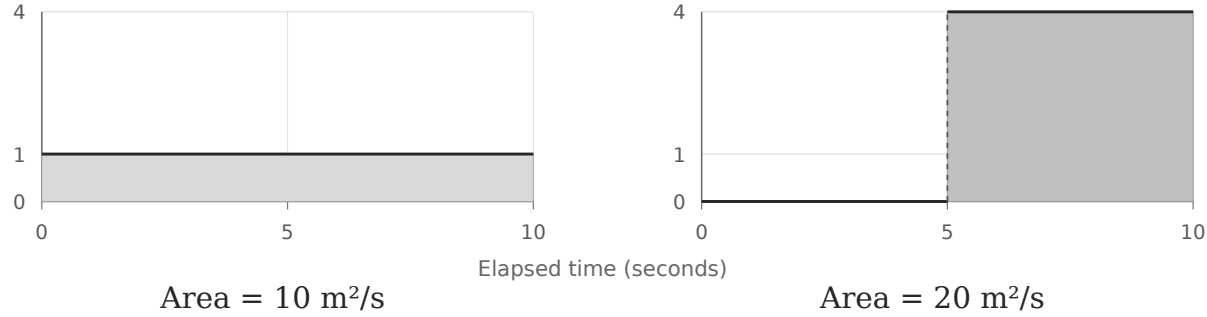


Dots mark each 2.5 seconds; repeated positions coincide.

Position along the corridor (metres)



Squared speed (m²/s²) — shaded area gives the total



Area = 10 m²/s

Area = 20 m²/s

Both arrive together. The squared-speed total doubles.

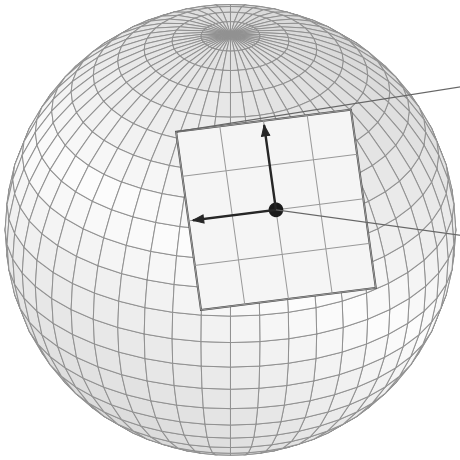
Areas use physical seconds and omit the conventional ½. Rescaling the clock keeps the 2:1 ratio.

Figure 9. Two walks over the same ten metres in ten seconds. The position traces retain the waiting that the endpoints alone conceal. The lower areas are squared speed added over physical seconds: $1^2 \times 10 = 10$ and $0^2 \times 5 + 2^2 \times 5 = 20$, in m^2/s. These totals omit the conventional factor $1/2$ used in the chapter's path functional. That factor and the common rescaling of the clock both preserve the two-to-one comparison. This is a constructed motion example, not a measurement of fuel consumption.

Local directions, global journeys

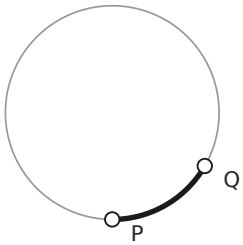
A tangent plane and two constant-speed routes on a sphere.

A The flat guide touches at one location



The plane gives local directions, not a flat global surface.

B One great circle, two routes



60° / short

300° / long

Both stay on the same great circle.

C Compare the route lengths

route	angle / rad	length / R
short	π/3	1.0472
long	5π/3	5.2360

Length = radius × angle in radians. The long route is five times longer.

Locally straight does not always mean globally shortest.

Both arcs are geodesics at constant speed. Only the short arc minimizes this distance.

Figure 10. A tangent plane supplies local directions at its contact point. The two routes below join the same endpoints along a great circle. For radius R, their lengths are $R\pi/3$ and $5R\pi/3$. Both are geodesic routes at constant speed; only the shorter one minimizes distance between these endpoints.

At each point, I need a rule that measures those local directions. On the smooth space we call M, the rule is a *Riemannian metric*, written g. It measures tangent lengths and angles consistently as we move. A velocity tells us both our direction and how fast we travel. When the metric compares that velocity with itself, the result is its squared speed. That is the quantity we just added up in the corridor. The symbol g lets us make the same kind of calculation where the surface and the coordinates are more complicated.

To record the whole trip, use a clock scaled to read $t = 0$ at departure and $t = 1$ at arrival. Its unit is the whole allotted duration, rather than one second. The endpoints and this interval stay fixed when we compare routes. Write the route as $\gamma : [0, 1] \to M$. The symbol γ names the whole path; $\gamma(t)$ is our position at parameter value t. Its endpoints are $\gamma(0) = p$ and $\gamma(1) = q$. A dot means differentiation with respect to t, so $\dot{\gamma} = d\gamma/dt$ is the velocity along the path. The squared-speed cost is

$$E[\gamma] = \frac{1}{2} \int_0^1 g(\dot{\gamma}, \dot{\gamma})\, dt$$

Inside the integral, $g(\dot{\gamma}, \dot{\gamma})$ measures squared speed. The integral adds it along the trip from $t = 0$ to $t = 1$. The factor $1/2$ is a convenient convention. Square brackets in $E[\gamma]$ remind us that the input is an entire path, not a single position. A function of an entire path is called a *functional*.

All those symbols are keeping a running total for the trip. They let me compare the whole walk, including the waiting and the rushing, rather than just compare where I ended up.

This functional is conventionally called the energy of a path. It is not automatically a fuel bill. If the parameter were physical time and the metric included mass, this integral would have units of energy times time, or action. Here it is a specified geometric cost. A physical application has to supply the reason for using it.

Keep the ends of the route pinned and nudge its middle. Does the cost go up or down? We can make that question exact without pretending to know the best route in advance. Let s measure the size of the nudge and let $\gamma_s(t)$ be the changed path. At $s = 0$ we have the original route. The field $V = \partial\gamma_s/\partial s|_{s=0}$ records the direction and amount of that initial nudge at each point. The endpoints stay pinned, so $V(0) = V(1) = 0$.

There are now two changes to keep separate. Moving along the path changes t. Trying a neighboring path changes s. The derivative dE/ds asks how the whole trip's cost responds to the second kind of change. Its value at $s = 0$ is called the *first variation*. It measures the first-order response, not every possible higher-order change.

I hold the start and finish still, move a little of the route, and watch which way the bill begins to move.

On the globe, draw a little arrow for our velocity here, then another a little farther along the route. They sit in different tangent planes. Those planes have turned with the surface, so a comparison has to account for that before telling us how the velocity itself has changed. The metric has a standard rule for making this comparison, called the *Levi–Civita connection*. The symbol ∇ means differentiation using that rule. Thus $\nabla_{\dot{\gamma}} V$ asks how the nudge V changes as we travel along the route.[4]

[4]The connection is metric-compatible, which gives the product rule when differentiating g, and torsion-free, which permits the path and variation derivatives of $\gamma_s(t)$ to interchange in the first-variation calculation. These are the properties used in the integration by parts below.

For smooth paths and smooth fixed-endpoint variations, differentiating the cost and then integrating by parts gives

$$\left.\frac{dE}{ds}\right|_0 = \int_0^1 g(\nabla_{\dot\gamma} V, \dot\gamma)\, dt = \left[g(V, \dot\gamma)\right]_0^1 - \int_0^1 g(V, \nabla_{\dot\gamma}\dot\gamma)\, dt.$$

The first integral compares the changing nudge with the path's velocity. Integration by parts moves the derivative onto the velocity and leaves an endpoint term. The brackets $[\,]_0^1$ mean "the value at the finish minus the value at the start." Both endpoint contributions vanish because the nudge is zero there. What remains is

$$\left.\frac{dE}{ds}\right|_{s=0} = -\int_0^1 g(\nabla_{\dot\gamma}\dot\gamma, V)\, dt$$

The expression $\nabla_{\dot\gamma}\dot\gamma$ is the *covariant acceleration*: how the velocity changes along the route after accounting for the geometry of the comparison. On a globe, the tangent planes tilt as we move. The covariant derivative prevents that change of plane from being confused with a turn within the surface.

In plain English, the equation adds up how the route's acceleration lines up with the nudge along the trip. That accumulated result determines the first-order change in cost. A *critical* or *stationary* path has no such response for any allowed nudge:

$$\left.\frac{dE}{ds}\right|_{s=0} = 0 \quad \text{for all variations } V$$

The words "for all" do the work. We can choose a smooth nudge supported on any small interior part of the route. If the covariant acceleration were nonzero there, we could choose a nudge in its direction and obtain a nonzero integral. Therefore the acceleration must vanish all along the route:

$$\nabla_{\dot\gamma}\dot\gamma = 0$$

This is the *geodesic equation*. Conversely, if that acceleration is zero, the integral is zero for every fixed-endpoint nudge. We have proved both directions.

Theorem 1.1 (Affinely Parameterized Geodesics as Critical Paths): *Among smooth paths connecting two fixed points on a Riemannian manifold over a fixed parameter interval, the affinely parameterized geodesics are precisely the critical points of the energy functional.*

Proof: The first-variation calculation above shows that stationarity for every smooth fixed-endpoint variation is equivalent to zero covariant acceleration. □[5]

"Affinely parameterized" means the parameter advances at a constant rate relative to distance along a nonconstant geodesic. In this Riemannian metric, the speed is constant. That is why waiting and then rushing down the corridor did not give the same cost as walking steadily. The theorem concerns the route and the way we traverse it.

[5]For the standard first-variation theorem, see William Minicozzi, *18.965 Geometry of Manifolds I*, notes by Serena An, 2025, Section 6.3.1, `https://www.mit.edu/~anser/files/18_965.pdf`.

Nor does stationary automatically mean shortest. On a sphere, great circles are geodesics. Follow one almost all the way around when the destination is just behind you, and you have taken a very long geodesic to somewhere nearby. Short enough segments minimize length locally; a long segment need not minimize it between its endpoints. In flat space, the same equation gives straight lines at constant speed. Spacetime geodesics use a Lorentzian metric instead; the theorem we just proved concerns the Riemannian case.

I can follow a great circle without turning to either side and still take the long way home. Following that rule and finding the shortest journey are different questions.

We began with two trips down a corridor and ended with a rule for a route on a curved space. Nothing in that passage told us how many litres of fuel a vehicle would burn. It showed how to choose a cost, vary an arrangement and derive a condition that the arrangement must satisfy. That pattern—geometry, a specified cost, a variation, an equation—will return when the object is a beam or a field. The price has to come from the system we are actually studying.

§ 1.5

The space I had to expand

There is a geometry to the way I think about a human life, too. Some doors are open. Some are locked. Some routes demand work that the person beside you never has to do. The map offered to me has often left out the places where I already stood.

I have no metric that turns a life into the geodesic calculation we just made. But I know what it means to keep looking for a route when the obvious one is closed. I know that a boundary on somebody else's map is not necessarily the edge of the world.

I have been finding such routes my whole life. From the rural South to Brown University. From NASA to IBM to the living room where I sit now, with a laptop and a dog and twenty-seven years of training I did not know I was preparing for. Each transition looked impossible from the outside. None felt like the path of least resistance from where I was standing. There was work in every crossing.

The checkpoint left me with a question about a description that could not accommodate what was standing in front of it. The pendulum gives that question a shape we can inspect: an arrangement, coordinates for describing it, a rule for comparing changes, a physical law assigning a cost. Those layers belong together, but they do different jobs. When a calculation fails, I want to be able to find the layer where it failed.

That is also how I want this research to be read. The proofs and experiments are here so that the route can be followed, checked and questioned. A geometric picture can open an investigation. The calculation has to carry us across the next part of it.

The patterns we recognize connect what we know. The differences open space for something new.

From the NASA question to GIGI
Chapter 4, page 85. AMC made the consequences of representation visible. The later GIGI replay carries that question into a different system, where the information required by a decision can be stated explicitly.

A field-equation idea on a test bench
Chapter 7, page 196. The calibration routes give a concrete setting for loop disagreement. We can work out the fit by hand and see which relationships the more general mathematics is trying to preserve.

A familiar balance inside a plasma calculation
Chapter 13, page 323. The discrete energy leads back to the harmonic and holomorphic work through a shared question about local structure. The connection is developed alongside the differences between the operators and their tasks.

What I carry out of CHIHIRO
Chapter 13, page 356. Plasma research gives the geometry new constraints and new questions. I follow what this domain teaches me, keeping its physical requirements attached to the mathematical patterns I recognize.

Learning where the resemblance ends
Chapter 14, page 369. A phase around one loop and curvature over a closed surface are related ideas with different requirements. The disk and sphere examples show why recognizing a pattern is the beginning of the comparison.

At NASA, I could keep a sensor reading fixed and change only the format carrying it. If a detector then changed its answer, I had a place to look. Something had happened between the world and its description. In the next chapter, that becomes a calculation about what a transformation can lose. Then a bent beam gives us a physical cost we can work out: what changes when we bend it, and what material information the bill requires.

The larger questions are still ahead—a gap in quantum energies, the effort of a search, a flow changing character. I reached toward them by following smaller questions far enough to see where they opened. The geometry was a way to keep looking.

Chapter 2

Curvature and Its Costs

"The shortest path between two truths in the real domain passes through the complex domain."
— Jacques Hadamard

The scanner left me asking what a representation permits a system to see. Telemetry gave me a way to watch that question play out in code: keep the reading fixed, change its format, and compare what the detector receives. Curvature enters this chapter because I wanted to understand what happens along a route of transformations, and where a claim about cost needs a physical law.

At NASA Johnson Space Center, I built AMC—Adaptive Multi-Vector Correlation—to let telemetry change format while it was arriving. A temperature trace, a battery voltage and a vibration signal do not all need the same container. Store every zero in a sparse signal and you waste space. Round a changing signal too aggressively and you may erase the change you were trying to catch. Bytes were easy to count. A missed change was harder to put back.

The later public evaluation used MODIS data to put numbers on the trade. Its storage table reports a 37.1% saving relative to the dense baseline. Its F1 table reports 0.899 for AMC, 0.834 for adaptive quantization, and 0.801 for static dense. Those are comparisons on that evaluation. The engineering question started earlier, with the data arriving at my desk.

The F1 score combines precision—how many raised alarms were correct—with recall—how many labeled events were caught. Their harmonic mean penalizes a system that does well on only one. An F1 of 0.899 is not a claim that 89.9% of every kind of decision was correct. The storage percentage answers another question altogether: how many bytes the chosen representation needed. I keep both numbers because saving space and preserving a detector's answer are separate demands.

The data kept coming while I decided what to do with it. Somewhere in that flood was a small change I needed to see in time to ask what caused it. A smaller file was only useful if the detector could still do its job.

What I did not yet have was a mathematical language for the question underneath the engineering: which relationships survive when I change the description?

§ 2.1

What a change of format can cost

Let me tell you what AMC really is, in the language I have since learned.

Start with one recorded window from a temperature sensor: a list of readings in acquisition order. I can treat that list as a vector. Dense storage keeps each entry; sparse storage can keep nonzero entries and their positions; quantization rounds values onto a coarser grid. The first two can describe exactly the same vector. The third may identify distinct readings. A smooth, invertible coordinate chart and a lossy codec therefore have different mathematical jobs.

Suppose the window contains a reading just above an alarm threshold. After rounding, the decoded reading might land on the other side. The sensor did not change its mind. We changed what the detector received. If we want a geometric model for this operation, we must include that possible loss; calling the formats "coordinates" does not make the loss disappear.

Consider the Challenger disaster. The Rogers Commission documented both the solid-rocket-joint failure and a history of decisions that normalized the risk. One analytical failure was comparing temperature among flights with damage without properly using flights without damage as the comparison group. The observations we leave out can change the apparent relationship.[1]

Putting the temperature and damage outcomes together, including the undamaged flights, makes the cold-weather concern easier to examine. That is a lesson about the evidence a display preserves. It does not reduce the deaths of seven people to a sorting choice: the technical defect, communication and management decisions all belong to the record.

An exact change of coordinates retains the underlying object when the rest of the calculation transforms consistently. Dropping observations or rounding their values is a different operation. The practical question is which distinctions the operation leaves available to the next calculation.

I could measure representation error directly: subtract the decoded readings from the originals, or compare their distributions. Curvature would enter later, after I specified a rule for comparing quantities along a route and asked whether comparisons around a small loop agreed.

AMC gave me an early reason to care about that question. The code could try a conversion, measure a discrepancy and reject a choice that failed a gate. The work began with those inspectable operations.

A smaller file is useful only if I can still trust the reading that comes back out. That brings the format decision to the detector: how much change can its answer tolerate?

The core innovation in AMC was something I called "semantic drift monitoring." When the system translates a vector from one format to another, it checks whether the meaning has been preserved. If the translation distorts the signal too much—if the "semantic drift" exceeds a threshold ϵ—the system rejects the translation and falls back to a safe format.

The mathematical description used two comparisons. Procrustes alignment fits the transformations we allow between two collections of vectors and measures the residual left after the fit. KL divergence compares two specified probability distributions. For discrete distributions

[1] Report of the Presidential Commission on the Space Shuttle Challenger Accident, 1986, Chapter VI, `https://www.nasa.gov/history/rogersrep/v1ch6.htm`.

it sums $p_i \log(p_i/q_i)$, with zero-p_i terms contributing zero; it is infinite if a positive p_i meets a zero q_i. Reversing p and q generally changes the answer. Neither comparison protects a detector automatically. Suppose the temperature sensor is monitoring a cooling pump, and the alarm depends on how hot the pump gets. If I allow a rescaling that makes a hotter trace look like the old one, I may hide the very rise the alarm needs to catch. Matching the shape of a trace and preserving its temperature are different requirements.

In the hybrid drift expression, v is the source vector, $T(v)$ its translated representation, and v' the reference representation to which I compare it. The distributions p and q must also be specified from the data. Their construction and the scaling of the two scores matter before I add them. The coefficient α chooses their relative influence; ϵ is a threshold in that chosen scoring convention. This describes the hybrid criterion, while Chapter 4 traces the recovered implementation and its separate translation gate:

$$\mathcal{L}_{\text{drift}} = \alpha \cdot d_{\text{Procrustes}}(T(v), v') + (1 - \alpha) \cdot D_{KL}(p \| q) < \epsilon \tag{2.1}$$

At the time, this was an engineering constraint: keep the accepted conversion within a declared discrepancy. What I wanted to understand was how to connect that discrepancy to the decision. If the detector has a margin of only a hundredth of a degree, a tenth-degree error can be disastrous even though the overall trace looks almost unchanged.

One conversion gives me a discrepancy to measure. A chain of conversions gives me another question: does the route through the formats matter? That was the question that sent me toward transport and curvature.

Think about what happens when you move a vector from one format to another.

In dense format, I store all n entries. In a lossless sparse format, the position list tells me where to put the omitted zeros back. Quantization needs a different reconstruction: a rounded value stands for a range of possible originals. That range is the uncertainty I have introduced. A converter can be exact for one input and destructive for another even when both occupy the same flat vector space.

Translating a record gives us a path through a graph of representations. If each conversion is exact and invertible, the same underlying record can survive the route. If one step rounds or drops values, it can change the answer. Parallel transport offered me a geometric analogy for route dependence. To turn that analogy into curvature, I would need to specify a connection and prove the correspondence.

A return trip through the codecs gave me a concrete test: encode the record, change formats, decode it, and compare with the original. Parallel transport suggested a more structured version of a return test. We will first learn that version on a sphere, where the rule and the return angle can both be calculated.

My ϵ-bound limited a measured discrepancy in the chosen representation. The geometric question it suggested was how errors compose along a route. A numerical threshold is not automatically an angle in radians; its units and its relation to the downstream task must be specified.

The recorded translation threshold was $\epsilon = 0.1$ in its own normalization. I found it by tuning the system and watching what failed. A tenth of a normalized unit says nothing by itself about a

tenth of a degree, a perceptual difference or a universal threshold of sameness. To reuse it, I need the same definition of the score and evidence that the detector's margin is protected.

Tuning the number taught me to ask a harder question. Could I replace "this setting worked on these examples" with a bound that explained why a particular decision would survive?

That was the part of the engineering problem I carried forward into geometry.

Those were engineering questions first. The mathematical language came as I tried to understand why the same kinds of failure kept appearing.

AMC was the seed that grew into everything in this book.

When I left NASA—eventually, after years—and found myself with a laptop and too much time to think, I started asking: what was I actually doing with AMC? Why did the Procrustes distance matter? What did it mean for a translation to preserve "semantics"?

The questions would not leave me alone. I would be walking Teddy and suddenly stop, standing in the middle of the sidewalk while he tugged at the leash, because I could see a possible geometric model behind the format compatibility graph. I would be making coffee and forget to drink it, staring at the wall, working out what the graph's spectrum might say about routes through the formats. The graph was concrete. Any smooth space I put behind it would need a construction and a reason to trust the approximation.

La obsesión es el precio de la comprensión. Obsession is the price of understanding.

The questions led me to gauge theory. To fiber bundles. To the Riemann curvature tensor. To the Yang–Mills equations and the mass-gap problem. I recognized related demands: define what can change in a description, identify what must stay fixed, and test whether the route affects the comparison. The structures were worth learning precisely because they let me ask more than the original code could answer.

The scanner led me to question whether a decision representation could accommodate the bodies it encountered. AMC led me to ask what a conversion preserves. Both experiences pushed me toward geometry. Neither experience alone measured a curvature tensor.

Both problems ask what a representation keeps visible and what it loses. Curvature is one mathematical tool for studying a specified transport rule. The story starts with the failure; the tool still needs its assumptions.

Now let me teach you how to measure it too.

But first, let me tell you what I now understand about what happened at NASA.

I had been building a telemetry system, but I was also learning how much a representation decides for us. A detector cannot recover an erased distinction just because the original instrument once measured it. I wanted a mathematical account that followed the distinction through every conversion.

This is not obvious. In fact, it is exactly the opposite of what I was taught in school. Data is numbers. Transformations are functions. You can transform however you like, and as long as you can transform back, nothing is lost.

With an exact inverse, I can recover the record. Trouble begins when I omit information, or keep the old comparison rule after changing the representation. Convert Celsius to Fahrenheit and keep the old numerical alarm threshold: the conversion is invertible, yet the decision is wrong. The ruler and the decision rule have to travel with the data.

Curvature measures the local failure of parallel transport to commute. A conversion can lose information even in a flat vector space, and a curved space can support exact comparisons. Preserving the relationships a task needs is the engineering question that led me to this mathematical one.

The universe is not flat.

§ 2.2

The bend a measurement can detect

There are two ways to think about curvature, and the difference between them is one of the most important distinctions in all of mathematics.

Extrinsic curvature is the bend you can see from outside. Roll a piece of paper into a cylinder. The cylinder is curved—you can see it, measure it, talk about the radius. But a tiny ant living on that cylinder would not know it was curved. The ant can walk in a straight line, measure distances, draw triangles—and everything behaves exactly as it would on a flat sheet. The paper has been bent, but not stretched. The intrinsic geometry is unchanged.

Intrinsic curvature is different. It is the kind of bending that affects measurements made entirely within the surface.

A sphere has intrinsic curvature. Try to wrap a flat piece of paper around a basketball without crumpling or tearing it. You cannot. The sphere's geometry is fundamentally different from a plane's. An ant on the sphere would notice: straight lines curve back on themselves, triangles have angles that sum to more than 180 degrees, parallel lines eventually meet.

The distinction matters because intrinsic geometry can be measured without an external embedding space. Extrinsic curvature also matters in physics, for example in the bending energy of a beam or shell. Which curvature belongs in an equation depends on what material or field the equation describes.

Think about what this means. When Einstein tells us that mass curves spacetime, he is not saying that spacetime bends into some fifth dimension we cannot see. He is saying that the geometry of the four dimensions we inhabit is intrinsically curved—that triangles drawn in spacetime near a massive object have different angle sums than triangles drawn far away, that clocks tick at different rates depending on where they are, that straight lines (geodesics) curve toward masses.

The curvature is not a metaphor. It is the thing itself.

Gauss's result explains why the cylinder can look bent while preserving the geometry measured on the sheet.

Carl Friedrich Gauss discovered this in 1827, while he was literally surveying land—measuring distances and angles across the curved surface of the Earth. He realized something profound: you do not need to leave a surface to know it is curved. The curvature is encoded in measurements you can make entirely within the surface.

This is Gauss's *Theorema Egregium*—the "remarkable theorem." It says that intrinsic curvature is an intrinsic property. It does not depend on how the surface sits in space. It depends only on the geometry of the surface itself.

This distinction gave me a way to ask better questions about AMC. Its graph, translation scores and detector outcomes were observable. A later geometric model could ask which features were intrinsic to a declared state space. I should not infer an intrinsic curvature measurement from the fact that a format policy worked.

The scanner failed for the same reason, inverted. It imposed an extrinsic structure—two bins, MALE and FEMALE—that had nothing to do with the intrinsic geometry of human bodies. It tried to flatten a curved space into two points, and reality would not fit.

The sign gives us three local pictures to compare.

Intrinsic curvature comes in three flavors.

Positive curvature: The surface bends the same way in every direction, like a sphere. Parallel lines converge. Triangles have angle sums greater than 180 degrees. There is less room than you expect—the surface closes in on itself.

Negative curvature: The surface bends in opposite directions, like a saddle or a Pringle chip. Parallel lines diverge. Triangles have angle sums less than 180 degrees. There is more room than you expect—the surface spreads out faster than flat space.

Zero curvature: Flat. Parallel lines stay parallel. Triangles have angle sums of exactly 180 degrees. Euclidean geometry holds.

The cylinder is instructive: it has zero intrinsic curvature despite appearing curved. You can unroll it onto a flat plane without stretching. The sphere cannot be unrolled. The saddle cannot be unrolled. Their curvature is intrinsic, written into the geometry itself.

Where do these shapes appear in nature?

Positive curvature: spheres, domes, water droplets, cell membranes, closed universes in cosmology.

Negative curvature: saddles, coral structures, lettuce leaves, hyperbolic surfaces, open universes in cosmology.

Zero curvature: planes, cylinders, cones (away from the tip), paper in any configuration that does not stretch it, the flat universe that observation suggests we live in.

The universe appears to have zero or very nearly zero large-scale curvature—a remarkable fact called the "flatness problem" that inflationary cosmology was designed to explain. But locally, near masses, curvature is extreme. Near a black hole, spacetime curves so severely that light cannot escape. Near the Big Bang, curvature was infinite—the singularity where our equations break down.

You can also see the distinction by cutting and joining a flat sheet.

Here is a fact you already know in your body, even if you have never thought about why.

When you pick up a slice of pizza, you bend it along the crust. Why? Because bending in one direction prevents flopping in the other.

This is Gauss's theorem in action. A flat pizza slice has zero intrinsic curvature. When I bend the crust without appreciably stretching the slice, its Gaussian curvature stays approximately zero. Gaussian curvature is the *product* of the two principal curvatures, the largest and smallest bending at a point. If one of them is now curved, the other has to stay nearly flat. Curving the

slice across its width makes it much harder to bend along its length, so the tip stops drooping. My hand already knew the trick before I knew Gauss's name.

The same principle explains why corrugated cardboard is stronger than flat cardboard. Why leaves curl when they dry. Why potato chips are saddle-shaped. Why you cannot peel an orange without tearing the skin. Every one of these is a manifestation of Gauss's theorem—the intrinsic curvature constraining what shapes are possible.

You have been using differential geometry every time you have eaten pizza. You just did not know its name.

El cuerpo sabe lo que la mente todavía no ha nombrado. The body knows what the mind has not yet named.

The cylinder and the sphere look different from outside, but the difference I need is one a measurement on the surface can detect. Let us put a number on it. How do we assign a number to curvature?

Gauss's method: at each point on a surface, measure how the surface bends in different directions. The surface has a "most curved" direction and a "least curved" direction, perpendicular to each other. These are the *principal directions*, and the curvatures along them are the *principal curvatures*, κ_1 and κ_2.

The *Gaussian curvature* at that point is the product:

$$K = \kappa_1 \cdot \kappa_2 \tag{2.2}$$

Why the product?

If both principal curvatures have the same sign (both positive or both negative), you get positive Gaussian curvature. This is a sphere-like point—the surface bends the same way in all directions.

If the principal curvatures have opposite signs, you get negative Gaussian curvature. This is a saddle-like point—the surface bends up in one direction and down in the other.

If either principal curvature is zero, the Gaussian curvature is zero. This is a flat or cylinder-like point.

The remarkable fact—Gauss's great discovery—is that this product K is intrinsic. It does not depend on how the surface is embedded in space. It can be computed entirely from measurements made on the surface.

Back in the telemetry problem, the useful connection was the effect of a transformation on the quantities I was trying to preserve.

AMC's score measured the features I used to choose a format. That is a useful place to see why a scalar summary needs an interpretation before we use it.

A signal can be mostly zeros, highly variable, or concentrated in a few commonly occurring values. Those properties influence storage and detector behavior in different ways. I combined measurements of them in a weighted score.

The cost function I used captured this:

$$C(v) = w_1 H(v) + w_2 \sigma(v) + w_3 S(v) + w_4 F(v) \tag{2.3}$$

Here H summarizes the spread of a specified value distribution, σ measures variation, S measures sparsity and F records task feedback. The weights set the scales and priorities. This score is a

policy for comparing costs; it is not the product of principal curvatures. Chapter 4 will expose another important detail: a term that is identical for every candidate format cannot help choose between those formats.

The cost score gave me a decision I could inspect. Gaussian curvature gives a different kind of measurement: it describes the surface geometry itself. Keeping their meanings separate lets me ask whether a later model can establish a useful relation between them.

To follow a chain of those transformations, I need to record which translations are available.

AMC maintained a graph where each node was a format—dense, sparse, quantized, compressed—and each edge was a trained translator between formats. Not every format could translate to every other format safely. Some translations introduced too much drift. Some paths through the graph were better than others.

I computed the graph Laplacian:

$$\mathcal{L} = D - A \tag{2.4}$$

For the symmetric weighted graph used in this spectral interpretation, A_{ij} is the nonnegative weight joining nodes i and j, and $D_{ii} = \sum_j A_{ij}$. Thus $(\mathcal{L}x)_i = \sum_j A_{ij}(x_i - x_j)$: compare the value at one node with its neighbors and add the weighted differences. A directed router needs its own construction or declared symmetrization before using these symmetric-Laplacian facts. The spectral coordinate can help describe a cut; it does not by itself certify a safe conversion path.

The second-smallest eigenvalue λ_2 (the Fiedler value) measured how connected the graph was. A large Fiedler value meant the graph was robust—many good paths between formats. A small Fiedler value meant the graph had a bottleneck—some essential translation was weak.

The Fiedler vector ϕ_2 assigns a spectral coordinate to each format. Differences between those coordinates can help expose a cut or guide a route. The vector may be multiplied by -1 without changing its eigenvalue, so a low component is not intrinsically central and a high component is not intrinsically peripheral. Any operational grouping needs its own sign convention and test.

I did not initially call this spectral geometry. I thought I was optimizing a translation network. The graph Laplacian made that connection concrete: its eigenvectors describe modes of mixing on the graph. A graph mixing gap and a quantum mass gap are spectral quantities for different operators; identifying them requires more than sharing the word gap.

The mathematics is patient. It waits for you to recognize it.

The curvature number also has a consequence for a journey. Return to the same place after following a local rule for keeping a direction steady, and compare the direction before and after. There is another way to detect curvature—one that will become crucial for gauge theory and the mass gap.

Take a globe and place a small arrow flat against it, initially pointing south from the North Pole. As the arrow moves, it must stay in the local tangent plane. The rule is to make no extra turn within that plane. One way to picture the rule is to take a tiny step, project the old arrow onto the new tangent plane and restore its length, then take the limit of smaller steps. That is the transport we will use. It is not a compass needle continually turning toward magnetic north.

On a flat surface, this is trivial. Walk any loop you like, keep the arrow pointing north, return to your starting point, and the arrow still points north.

On a sphere, something strange happens.

Start at the North Pole with an arrow pointing toward Ecuador. Walk south to the equator. Walk east along the equator for one quarter of the Earth's circumference. Walk north back to the Pole.

Along each great-circle segment, the arrow keeps the same angle to the direction of travel. At a corner, turn the route without turning the arrow with it. After the three legs, the arrow is back in the original tangent plane but points a quarter-turn away from its initial direction. The comparison is now unambiguous: same location, same plane, two different directions.

The surface itself rotated the arrow. This is the geometric signature of curvature.

A different loop gives a different return. On a sphere of radius r, the curvature is $1/r^2$. Our three-leg triangle encloses an eighth of the sphere, with area $\pi r^2/2$, so the angle is $\pi/2$. A hemisphere encloses $2\pi r^2$ and gives a full turn, 2π, which returns the arrow to the same orientation. The area rule is understood modulo full turns.

This is not intuition. This is mathematics. And it generalizes to spaces of any dimension, any shape, any complexity. The rotation accumulated by parallel transport around a loop is how we *define* curvature in the most general settings.

For a small loop, we can compare the return angle with the area enclosed.

This observation can be made precise. For a positively oriented disk of area A on a surface of constant Gaussian curvature K, choose an angle representative:

$$\theta \equiv K \cdot A \pmod{2\pi} \tag{2.5}$$

The rotation angle θ accumulated by parallel transport equals the curvature times the area.

This is not just a curiosity. It is the *definition* of curvature in a form that generalizes to any dimension. Curvature is rotation per unit area. The more curved the space, the more an arrow rotates when you carry it around a loop.

That comparison gives me a language for asking what changed along a route through representations.

Now I can explain why the transport picture caught my attention, and where the analogy ends.

When AMC translates a vector from one format to another, it is carrying information along a path through the embedding manifold. The "arrow" is the semantic content—the meaning of the signal, the relationships it has to other signals, the anomalies it might indicate.

A lossless dense-to-sparse conversion keeps the entries and their positions. The detector can recover exactly the same vector. That is an algebraic property of this encoding, not a consequence of flat curvature.

A quantized conversion can change a value that matters to an alarm. I can measure that error directly and compare it with the alarm margin. A geometric transport model may help organize such checks, but the codec error is not automatically a rotation.

For a transport model that explicitly bounds an angular change, one could impose:

$$\theta < \epsilon \tag{2.6}$$

That would be an angular budget in the declared model. AMC's translation-norm gate instead bounds a representation discrepancy; I keep those quantities separate.

Procrustes alignment compares shapes after fitting allowed transformations. The residual measures what the fit did not explain. A successful rotation fit can have zero residual even when the fitted rotation is large. The residual, the transformation and the curvature of a connection are therefore three different quantities.

A loop is especially useful because it returns to the original format. I can compare the result with the record I started with.

You might wonder: why loops? Why does carrying an arrow around a closed path reveal curvature?

Because physics is local. We never see the global shape of the universe directly. We only see what happens in our neighborhood, and we infer the larger structure from local measurements.

A small loop gives us a local probe. Its leading rotation depends on the curvature and the oriented area it encloses. Returning unchanged on one loop does not prove flatness. Conversely, a flat connection on a space with a hole can have nontrivial transport around a loop that cannot be contracted through that hole. The size and kind of loop belong in the measurement.

In gauge theory, which we will meet in Chapter 5, a connection tells us how to transport a local reference frame. Curvature describes the leading change after transport around a small loop in spacetime. Later we will use loops as measurements of field configurations.

Consider a quark carrying a "color charge"—a property that exists in a three-dimensional internal space (the space of SU(3) representations). When the quark moves through space, its color rotates according to the gauge connection. If you move the quark around a closed loop and it comes back with a different color, the region enclosed by the loop contains a gluon field. The curvature of that field is what we call the field strength.

Confinement asks how gauge-invariant quantities behave in the quantum theory. Wilson-loop expectation values can probe the potential between external sources; an area law in an appropriate regime is one important diagnostic. Nonzero classical curvature on a particular loop is not, by itself, a proof of confinement.

A loop can register a change in the field. Whether a quantum excitation has a minimum energy is a separate calculation. The connection between the two questions is what makes the gauge-field example worth pursuing.

We are building toward that. But first, we need the Riemann tensor.

§ 2.3

Keeping track of every direction

Gaussian curvature works beautifully for surfaces—two-dimensional manifolds. But the spaces we care about have more dimensions.

The configuration space of a gauge field is infinite-dimensional. The embedding space of telemetry vectors might have hundreds of dimensions. Even ordinary spacetime has four dimensions. A single number cannot capture all the ways a higher-dimensional space can curve.

We need a more general tool: the Riemann curvature tensor.

The two-dimensional picture suggests the next step: compare transport in every available pair of directions.

In 1854, Bernhard Riemann gave a lecture that founded modern differential geometry. He was twenty-seven years old, presenting his *Habilitationsschrift* before the faculty at Göttingen. The topic had been chosen by Gauss himself, his doctoral advisor: "On the hypotheses which lie at the foundations of geometry."

Riemann was nervous. He was presenting to Gauss, the greatest mathematician of the age, on Gauss's own subject. But what Riemann said that day would take sixty years to be fully understood, and would ultimately become the mathematical language of general relativity.

Riemann asked: how can you tell if a space is genuinely curved, or just flat space described in funny coordinates?

Think about what this means. If I give you a map of the Earth, it looks flat—but the Earth is curved. The flatness is an artifact of the projection, not a property of the Earth itself. Conversely, I could take a flat sheet of paper, draw curved coordinate lines on it, and make it *look* curved in those coordinates. How do you distinguish real curvature from coordinate artifacts?

A modern way to answer that question uses parallel transport around infinitesimal loops.

Let R denote the Riemann curvature tensor. Its definition:

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \tag{2.7}$$

Here X and Y specify the two directions of comparison, while Z is the vector being compared. The first two terms subtract derivatives taken in opposite orders. The bracket $[X,Y]$ accounts for the fact that the direction fields themselves can vary. In a coordinate frame those coordinate directions commute, so that bracket is zero.

The loop interpretation is an infinitesimal one. Shrink a loop in the X,Y plane and divide its transport defect by its signed area. The leading response is described by curvature, with its sign tied to the direction of traversal and the transport convention. A finite trip can include higher-order terms or full-turn cancellations. We cannot replace the complete finite return by $R(X,Y)Z$ without that limiting argument.

The Riemann tensor measures the local failure of parallel transport to commute. It generalizes the return-angle experiment on the globe to more dimensions and different tangent planes.

The indices let me write that directional information down without losing which direction played which part.

In coordinates, write the tensor as $R^{\rho}{}_{\sigma\mu\nu}$. The indices μ, ν identify the two comparison directions; σ identifies the input component of the vector; ρ identifies the output component. A repeated index such as λ is summed over all coordinate directions. The symbol ∂_μ means differentiation with respect to coordinate μ. With that bookkeeping, the formula is

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \tag{2.8}$$

Do not let this formula intimidate you. Let me walk you through what each piece means.

The Γ symbols (Christoffel symbols) encode how to parallel transport in the given coordinate system. If you want to carry a vector from point A to nearby point B while keeping it "parallel," the Christoffel symbols tell you how to adjust the vector's components as you move. On a flat

space in Cartesian coordinates, $\Gamma = 0$ everywhere—parallel transport is trivial, you just keep the components the same. On a curved space, or in curvilinear coordinates, $\Gamma \neq 0$—you have to make corrections.

Draw polar coordinates on a flat floor: concentric circles and radial lines. Stand east of the origin, then walk around to its north side. The local "outward" arrow has changed direction, even though the floor is flat. A fixed arrow drawn on the floor therefore acquires different components in these local radial and tangential directions. Christoffel symbols account for such changes in the coordinate basis. Their being nonzero does not, by itself, establish curvature.

At NASA, I encountered coordinate changes concretely. Satellite telemetry used spherical, orbital and mission-specific frames. Jacobians and rotation matrices converted components between those descriptions. A connection answers an additional question: how should a vector be transported as its location changes? Those operations are related, but a Jacobian is not itself a Christoffel symbol, and curvilinear coordinates can occur in flat space.

The Riemann tensor combines these Christoffel symbols in a very specific way. The first two terms ($\partial_\mu\Gamma - \partial_\nu\Gamma$) measure how the parallel transport rule changes from place to place. The last two terms ($\Gamma\Gamma - \Gamma\Gamma$) capture the nonlinearity—the fact that on a curved space, the order of operations matters.

The components of both objects depend on coordinates. The Riemann tensor, however, transforms as a tensor and represents an intrinsic geometric operation. If it is zero in one coordinate system, it is zero in every valid coordinate system. Bending the grid on the floor cannot manufacture curvature.

In two dimensions, the Riemann tensor has essentially one independent component, which reduces to the Gaussian curvature. In four dimensions (spacetime), it has twenty independent components. In n dimensions, it has $n^2(n^2-1)/12$ components.

The higher the dimension, the more ways the space can curve.

The components are related. Their symmetries record those relationships and reduce the independent information.

The Riemann tensor is not arbitrary. It satisfies symmetries:

- Antisymmetric in the last two indices: $R_{\rho\sigma\mu\nu} = -R_{\rho\sigma\nu\mu}$
- Antisymmetric in the first two indices: $R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}$
- Symmetric under pair exchange: $R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma}$
- First Bianchi identity: $R_{\rho[\sigma\mu\nu]} = 0$

These symmetries are not imposed from outside. They follow from the requirement that the geometry be consistent—that parallel transport along different paths that enclose the same region must give related results.

The symmetries reduce the number of independent components and constrain what kinds of curvature are possible. They are the grammar of geometry.

The tensor keeps the directional information. Sometimes the next calculation needs a summary of it. Summing particular indices gives two such summaries, each retaining less than the full tensor. The full Riemann tensor contains more information than we often need. By summing over indices (contracting), we get simpler measures of curvature.

The first contraction gives the Ricci tensor:

$$R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu} \tag{2.9}$$

In Riemannian geometry, the Ricci tensor measures the leading change in the cross section of a narrow bundle of geodesics. Positive Ricci curvature in a direction reduces the cross section relative to a flat comparison at short distances.

This is what appears in Einstein's field equations:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{2.10}$$

Geometry on the left. Stress, energy and momentum on the right. The equation couples them.

Einstein's equation relates the local stress-energy to a particular contraction of spacetime curvature. To see which information the equation uses, it helps to look at each term.

Let me unpack this equation piece by piece.

$R_{\mu\nu}$ is the Ricci tensor, a contraction of the Riemann tensor. It records part of the curvature.

R is the scalar curvature, the trace of the Ricci tensor. Different curvature tensors can have the same scalar trace.

$g_{\mu\nu}$ is the metric tensor—it tells you how to measure distances and angles. The metric is the fundamental object; everything else derives from it.

Λ is the cosmological constant. Its contribution is proportional to the metric. A nonzero value changes the equation even where ordinary matter and radiation are absent.

$T_{\mu\nu}$ is the stress-energy tensor—it encodes the density and flow of energy and momentum. Where there is matter, where there is radiation, where there is pressure, $T_{\mu\nu}$ is nonzero.

Set $T_{\mu\nu} = 0$ and $\Lambda = 0$. The field equation then gives $R_{\mu\nu} = 0$, while the full Riemann curvature can still be nonzero. A gravitational wave passing through empty space is an example.

When I say *energy is the price of shape*, I want to know which geometry and which physical quantity are involved. The field equation here is about spacetime. The Yang–Mills action below uses the curvature of a gauge connection. Keeping the two straight is what lets us make a calculation.

Tracing once more gives scalar curvature:

$$R = g^{\mu\nu}R_{\mu\nu} \tag{2.11}$$

This leaves one scalar at each point, a signed contraction of the curvature information. Different geometries can have the same scalar, including zero. The Einstein–Hilbert action uses this contraction, but a question about every directional bend may require the full tensor.

Each contraction keeps a particular summary. It is worth keeping the information lost at each step in view.

Object	**What it captures**	**Components (4D)**
Riemann $R^{\rho}{}_{\sigma\mu\nu}$	Full curvature	20
Ricci $R_{\mu\nu}$	Volume curvature	10
Scalar R	Average curvature	1

For different problems, different levels of detail matter. In Yang–Mills theory the field strength is the curvature of a gauge connection. We will also use a geometry on the space of field configurations to compare possible fields.

The full tensor distinguishes tangent planes and their curvature responses. Ricci curvature sums certain responses around a direction; scalar curvature traces that sum once more. Like an account total, a trace can conceal offsetting contributions. A zero total does not tell me every entry was zero. That is why I keep the full tensor when the question needs information the trace discarded.

We can now check the loop picture against a proof. On a surface, the accumulated turning is related to the curvature inside the loop; the orientation and the possible full turns belong in the statement. On an oriented surface, the local turning accumulates into an integral. The theorem states the relation precisely, including the full turns that an angle alone cannot distinguish.

Theorem 2.1 (Surface holonomy)**.** *Let R be a smoothly bounded disk in an oriented Riemannian surface, and traverse $\gamma = \partial R$ with the induced positive orientation. Parallel transport rotates a tangent vector by an angle satisfying*

$$\theta \equiv \iint_R K \, dA \pmod{2\pi}. \tag{2.12}$$

For constant curvature this is KA modulo 2π. On a sufficiently small disk the continuously chosen angle is approximately $K(p)A$.

Proof. Choose a smooth positively oriented orthonormal frame (e_1, e_2) on the disk. Define its connection one-form ω by $\nabla e_1 = \omega e_2$ and $\nabla e_2 = -\omega e_1$. A transported unit vector is $v = \cos\alpha\, e_1 + \sin\alpha\, e_2$. The equation $\nabla_{\dot\gamma} v = 0$ gives $\dot\alpha = -\omega(\dot\gamma)$. With the curvature convention used above, the structure equation is $d\omega = -K\, dA$. Stokes' theorem therefore gives

$$\Delta\alpha = -\oint_\gamma \omega = -\iint_R d\omega = \iint_R K \, dA.$$

The actual rotation identifies angles differing by 2π. This proves the stated relation.[2] □

An entire hemisphere has integrated curvature 2π. The arrow comes back pointing the same way. That does not make the hemisphere flat; the rotation has completed a full turn. A single loop can also enclose positive and negative contributions that cancel. To detect local curvature, we examine the leading response of sufficiently small loops in each tangent plane.

This proof pattern—defining a quantity by its effect on parallel transport—will return when we study gauge theory. The field strength tensor $F_{\mu\nu}$ is defined analogously: it is the curvature of the gauge connection, the rotation accumulated by parallel transport in an internal symmetry space.

[2] Chris Wendl, *Connections*, Chapter 6, curvature and the Gauss–Bonnet formula, `https://www.mathematik.hu-berlin.de/~wendl/pub/connections_chapter6_2.pdf`.

§ 2.4

Putting the cost on a bench

A bent beam is where I can put the phrase *energy is the price of shape* on a bench. I can name the material, measure the curvature and calculate what it costs to hold the beam there.

An elastic equilibrium can minimize a specified energy under the stated constraints. In dynamics the action principle asks for a stationary value. The first variation vanishes, and a separate argument tells you whether the stationary value is a minimum.

Chapter 1 began with the cost of a path. Here is a case in which I can show every step between the shape and the cost.

Let us stay with one beam long enough to calculate the difference between two ways of bending it. Hold one end of a thin elastic strip flat on the bench. Turn the other end a little. I can spread the bend out so the strip looks like a piece of a circle, or I can concentrate most of it in a short section. The two strips can leave the bench at the same angle. My hands can still feel a difference. To put a number on that difference, I need the material, the cross section, and the curve of the centerline.

Let s be distance along the centerline of an unsheared planar strip of length L. Its tangent direction makes an angle $\theta(s)$ with the bench. The signed curvature is $\kappa(s) = d\theta/ds$. When the strip bends, a fiber at signed distance y from the neutral axis has axial strain $\epsilon(s, y) = -y\kappa(s)$ in the small-strain beam model. I can picture the two edges. The one on the inside of the turn has less length to cover. The one on the outside has more.

For a linearly elastic material with Young's modulus E, the elastic energy density is $\frac{1}{2}E\epsilon^2$. Integrate across the cross section A and then along the beam:

$$\begin{aligned} U &= \int_0^L \int_A \frac{E}{2} y^2 \kappa(s)^2 \, dA \, ds \\ &= \frac{EI}{2} \int_0^L \kappa(s)^2 \, ds, \qquad I = \int_A y^2 \, dA. \end{aligned}$$

The moment I measures how the material is distributed away from the neutral axis. For a rectangular section of width b and thickness h, $I = bh^3/12$. Doubling the thickness at fixed width multiplies I by eight. That is why I can bend a ruler one way with two fingers and struggle when I turn it edge-on. The curve alone cannot give me the number on the force gauge.

The units are a check on the argument. EI has units of newton metres squared. Curvature squared times length has units of inverse metres. Their product is in joules. I can add energies from different pieces of this strip because I am adding the same physical quantity.

The square in the energy makes a practical difference. Two bends can turn the end by the same angle and still distribute the curvature differently. Suppose I prescribe the tangents at the two ends and let their positions move. Their net rotation is $\Theta = \theta(L) - \theta(0) = \int_0^L \kappa(s) \, ds$. Write the curvature as its mean plus a remainder. The cross term disappears because the remainder integrates to zero:

$$\int_0^L \kappa(s)^2 \, ds = \frac{\Theta^2}{L} + \int_0^L \left(\kappa(s) - \frac{\Theta}{L} \right)^2 ds.$$

The first term is the cost of the turn when it is spread uniformly. The second measures the extra energy from concentrating it or making it vary. The minimum in this problem is a circular arc. Clamping the positions as well would impose more conditions, and the optimum could change.

For $L = 0.80\,\mathrm{m}$, $EI = 12\,\mathrm{N\,m^2}$, and $\Theta = 0.20$ radians, the uniform curvature is $0.25\,\mathrm{m^{-1}}$. It costs 0.30 joule. Concentrating the rotation in the middle half doubles the curvature there to $0.50\,\mathrm{m^{-1}}$. The square makes the total cost 0.60 joule. A smooth profile $\kappa(s) = (\Theta/L)(1 - \cos(2\pi s/L))$ costs 0.45 joule. Its cosine term integrates to zero and its square adds half the uniform cost. The profiles in Figure 11 turn the tangent through the same angle.

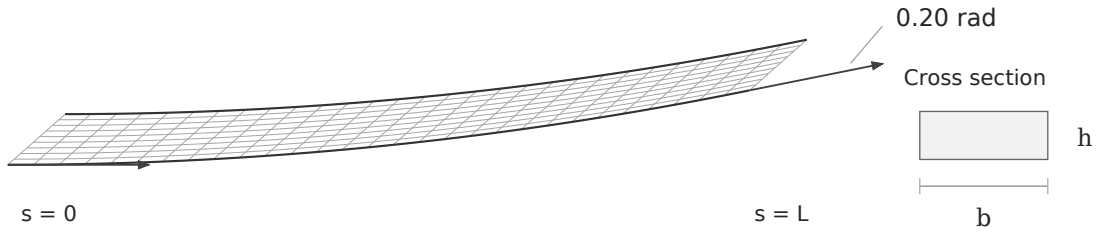
The curvature's integral gives the net turn; its distribution locates the bend. Squaring it and including material stiffness gives the energy. I need to keep those jobs separate.

One end turn, three bending costs

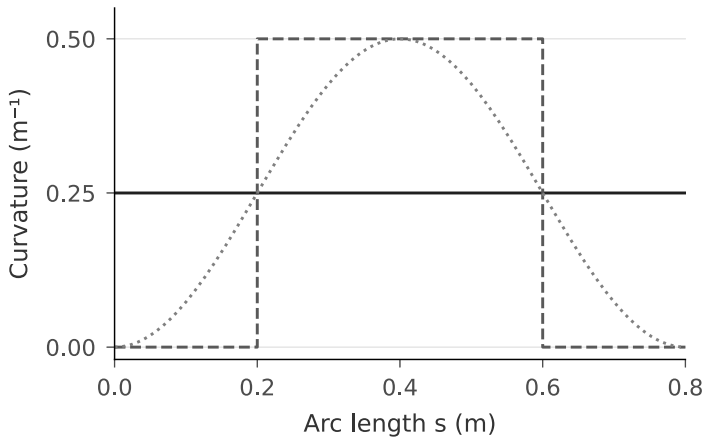
ELASTIC STRIP / CURVATURE PROFILES / CONSTRUCTED EXAMPLE

01 Keep the end turn fixed

The centreline turns through 0.20 rad along an arc length of 0.80 m.



02 Distribute the curvature



03 Integrate the square

Uniform
0.30 J

Middle half
0.60 J

Smooth
0.45 J

$$U = \frac{EI}{2}\int \kappa^2 \, ds$$

EI = 12 N m²; I = bh³/12

The same total turn can require different bending energy.

The ruled surface is an illustrative strip; all three profiles have the same length and stiffness.

Figure 11. Three curvature profiles for the same change in tangent direction. The area under each curve gives the total rotation. The energy depends on the area under its square, with the same stiffness and length for all three.

To connect geometry with a physical cost, I need the action appropriate to the theory.

In general relativity, the action is:

$$S_{EH} = \frac{1}{16\pi G} \int R\sqrt{-g}\, d^4x \tag{2.13}$$

An integral of the scalar curvature over spacetime.

Flat spacetime has $R = 0$, but the converse fails. The vacuum region outside a gravitating body can have zero Ricci tensor and zero scalar curvature while its Riemann tensor is nonzero. The displayed Einstein–Hilbert term uses $c = 1$ and omits matter, a cosmological term and any required boundary term. It is a scalar contraction in an action, not a positive measure of all curvature.

The principle of stationary action says: vary the geometry, find where the action is stationary, and that is the geometry that actually exists. This gives Einstein's equations.

Curvature enters the action, but a stationary action need not be a minimum. Einstein's equations do not say that spacetime chooses the least possible bend. The field equations, sources and boundary conditions determine which geometries are admitted.

In gauge theory, the curvature belongs to the gauge connection. The action below specifies how it enters that model.

In gauge theory, the action is:

$$S_{YM} = \frac{1}{2g^2} \int \mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})\, d^4x \tag{2.14}$$

Here F is the field strength tensor—the curvature of the gauge connection. The action is an integral of curvature squared.

Flat gauge field: $F = 0$, so $S = 0$. Curved gauge field: $F \neq 0$, so $S > 0$.

This is a statement about the Euclidean action of a classical field history. A flat connection minimizes that action where it is admissible. The quantum vacuum is a state of the quantized theory.

To state the mass-gap question, shift the vacuum energy to zero and consider normalized physical states orthogonal to the vacuum. For a unique vacuum, the lower edge of their energy spectrum is

$$\Delta = \inf_{\substack{\|\psi\|=1,\ \psi\perp|0\rangle \\ \psi\in D(H^{1/2})}} \|H^{1/2}\psi\|^2. \tag{2.15}$$

A positive gap requires a common positive lower bound across all those states. The squared-curvature action motivates the search for such a bound. Establishing it requires a relation to the Hamiltonian's quadratic form.

For AMC, the engineering cost returns me to storage, computation and the decision made from the decoded record.

In AMC, I did not think in terms of "energy." I thought in terms of "cost." But the structure was the same.

Each format switch had a cost—computational overhead, storage requirements, translation latency. The system minimized total cost subject to the constraint that semantic drift stayed below ϵ.

The selector compared admissible formats under a cost policy. A path through several conversions could also be given an accumulated cost. To call that path a geodesic would require a metric and a derivation connecting its energy to this cost. What AMC supplied was the engineering problem that made me want that derivation.

I had learned to separate the choices, their prices and the conditions an acceptable choice had to satisfy. That organization would be useful again when I studied physical variational principles.

The cost function I minimized was:

$$C(v) = \alpha_1 S(v) + \alpha_2 V(v) + \alpha_3 H(v) + \alpha_4 \Delta A(v, f) \tag{2.16}$$

where S is sparsity, V is volatility, H is entropy, and ΔA is the anomaly detection improvement for format f. The weights α_k were not fixed—they adapted over time using exponential moving average updates. The system learned what mattered.

This has the shape of a variational problem. I specify a cost, then compare possible choices. In the telemetry system I could tune the weights as the signals changed. In a physical model the action has to be justified by the dynamics it describes.

Updating feedback could change which format the policy preferred. But when I went back to the selector, I found a problem in how it compared the choices. In Chapter 4 I follow one reading through that decision and repair a copy of the code. I wanted to see what the mathematics actually made the program do.

The beam makes the cost concrete: its material law tells me what a bend costs. Carrying that intuition into a field theory means identifying the corresponding action and its assumptions. Let me say it plainly:

Energy is the price of shape.

A flat classical gauge field minimizes the squared-curvature action whenever the boundary conditions admit one. The quantum vacuum is the ground state of the Hamiltonian. Describing it requires the quantum theory, including the fluctuations that contribute to its measurements.

For the beam, curvature enters a positive bending-energy term through a material law. For the gauge field, field strength enters the specified classical action. Those are two explicit connections between a shape measurement and a cost. They give me calculations to check instead of one word that is supposed to explain every source of energy.

Accelerating a car changes its kinetic energy. Heating water changes its internal energy. Each process needs the appropriate dynamics and bookkeeping. Calling both changes "curvature" would skip the very explanation I need: which geometric quantity enters which energy law?

Everything that exists, exists because something paid for its shape.

The physical examples make sense only with the relevant system and its energy law specified.

For hydrogen, the familiar 13.6 eV is approximately the energy needed to ionize the ground-state atom. It is an energy difference between a bound state and a separated proton and electron, not an absolute price assigned to the electron cloud's shape. A bound state lies below that separation reference. The reference matters when we decide whether a change absorbs or releases energy.

A uranium nucleus has a certain shape—238 nucleons bound together. That shape costs about 7.6 MeV per nucleon, or about 1.8 GeV total. If you want to break it apart, you have to pay that energy. (Or, if you split it carefully, you can harvest the energy released when it falls into two more stable shapes—this is nuclear fission.)

The proton's roughly 938 MeV rest energy poses a different accounting problem. Its QCD dynamics includes quark and gluon contributions; it is not the same use of "binding energy" as the amount needed to disassemble a nucleus. Similar-looking energy numbers do not automatically name the same physical quantity.

Most of the proton's mass comes from QCD dynamics rather than the bare masses of its valence quarks. Pure Yang–Mills theory describes gauge fields without quarks. Its lightest gauge-invariant excitations are a different spectroscopy problem.

$E = mc^2$ relates the rest energy of a system to its mass. It takes a separate model to say how geometry enters that energy.

A minimum excitation energy is a spectral question. It asks what lies above the ground state of a Hamiltonian.

If the Yang–Mills mass gap exists, it is the minimum price to exist.

For pure Yang–Mills theory, the claim concerns the physical excitations of that theory, with no quark fields included. After choosing the vacuum energy as zero, a positive gap would bound the energy of every normalized state orthogonal to the vacuum. The challenge is to establish one bound that works for all those states in the continuum theory. A minimum in a finite list of sampled configurations does not yet do that.

The gap belongs to a dynamics defined on a geometric space. To understand it, I need to know both which fields the model admits and how the Hamiltonian acts on their quantum states.

That was the question the engineering experience had taught me to keep asking: what does the change cost, and which quantity measures it?

At NASA, I learned that curvature has consequences.

On the Shuttle, vehicle shape helped determine where heat would concentrate. That mattered when choosing materials and sizing the thermal protection.

AMC was designed around the characteristics of the signal and the detector's needs. Geometry became a way to investigate what the representations kept and how to compare them consistently.

The scanner at the airport brought the question back to the model. What shapes did its measurements represent, and which bodies could its rules accommodate?

That brought me to the geometry of gauge fields and the question of what it takes to excite them. I wanted a measurement that could tell me something about that cost.

The mathematics is the same. The stakes are higher.

Here is what I want you to take from this chapter:

Curvature becomes concrete when I specify the measurement: angle excess on a small surface triangle, the leading return rotation per unit area, or the field strength of a declared gauge connection. An energy claim needs the next step too: the law connecting that measurement to energy or action. That is the connection I want us to keep looking for.

And it is everywhere.

In a beam, a pressure vessel or a lens, the geometry helps specify the mechanical or optical problem. A failing component and a format error also have structures we can measure, but neither is automatically a curvature measurement. I want the name of the quantity to tell us what was actually calculated.

Curvature is not a complication. It is the thing itself. The universe is made of curvature the way a symphony is made of sound. Trying to understand physics without curvature is like trying

to understand music without sound—you might describe the mathematical relationships between notes, but you would miss everything that matters.

Now you know what curvature is. Now you can measure it. Now you are ready for the next tool: the heat kernel, and the spectral theory that lets us see geometry through the lens of diffusion.

Let us continue.

This is how the chapter returns to the format problem. I began by asking what a change of description could cost. Geometry gave me ways to describe that change; a physical energy claim still needs the law of the system doing the changing. I started with telemetry. I ended with the fundamental forces.

The path was not obvious while I was walking it. I built AMC because NASA needed better anomaly detection. I used Procrustes distance because it was the standard tool for measuring transformation quality. I bounded semantic drift because unbounded drift caused system failures.

I had a working question before I had this vocabulary. Procrustes residuals, detector margins and loop transport now let me ask different parts of it precisely. The fact that they can appear in the same investigation makes their differences more important, not less.

But the mathematics knew. The mathematics does not care what you call it. It does not care whether you learned it in a classroom or derived it from engineering necessity. It cares only whether you get it right.

When I got it right—at NASA, in my living room, in this book—the mathematics was the reason. When I got it wrong, the mathematics showed me where.

The scanner at the airport taught me to ask what its description allowed it to see. AMC taught me to put a number on what a change of description cost. The Yang–Mills experiment brings both questions to the lattice. Which features of a field do my measurements retain, and what can those features tell me about energy?

Nunca sabemos lo que estamos construyendo hasta que está construido. We never know what we are building until it is built.

The next chapter gives us a way to follow a change through time. We will watch patterns spread and ask what their rates reveal about the space.

In the next chapter, we will learn about the heat kernel—the tool that lets us probe geometry by watching how information diffuses. We will see how the spectrum of a space (its eigenvalues) encodes its curvature, and how spectral gaps relate to mass gaps.

The heat kernel is one of the most powerful tools in modern geometry. When you heat one end of a rod, the temperature spreads. How fast it spreads, and in what pattern, depends on the geometry of the rod. A thin rod conducts differently than a thick one. A curved rod conducts differently than a straight one. A rod with a hole conducts differently than a solid rod.

A heat operator can also be studied on an appropriate manifold. A quantum Hamiltonian has its own imaginary-time semigroup, e^{-tH}, whose decay above the vacuum is related to its spectral gap. Identifying that operator is essential: diffusion under an arbitrary geometric Laplacian does not automatically measure the energy spectrum of a gauge theory.

After that, the Davis Manifold gives the geometry a place to live, and then we open the first door: a finite Yang–Mills lattice calculation, and what it can and cannot tell me about a mass gap that has eluded mathematicians for fifty years.

The Davis Law comes after that door. In Chapter 6 I propose $C = \tau/K$ and test it where its budget can be measured.

The tools are almost assembled. The vocabulary is almost complete.

Let us continue.

§ 2.5

The equations behind the argument

For reference, here are the central equations introduced in this chapter.

Gaussian Curvature (Product of principal curvatures):

$$K = \kappa_1 \cdot \kappa_2 \tag{2.17}$$

Parallel Transport Rotation (Constant curvature and a chosen signed angle):

$$\theta \equiv K \cdot A \pmod{2\pi} \tag{2.18}$$

Riemann Curvature Tensor (Failure of parallel transport to commute):

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \tag{2.19}$$

Riemann Tensor in Coordinates:

$$R^{\rho}{}_{\sigma\mu\nu} = \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\rho}_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} - \Gamma^{\rho}_{\nu\lambda}\Gamma^{\lambda}_{\mu\sigma} \tag{2.20}$$

Ricci Tensor (Contraction of Riemann):

$$R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu} \tag{2.21}$$

Scalar Curvature (Trace of Ricci):

$$R = g^{\mu\nu} R_{\mu\nu} \tag{2.22}$$

Einstein Field Equations:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{2.23}$$

Einstein-Hilbert Action (General Relativity):

$$S_{EH} = \frac{1}{16\pi G} \int R\sqrt{-g}\, d^4x \tag{2.24}$$

Yang-Mills Action (Gauge Theory):

$$S_{YM} = \frac{1}{2g^2} \int \mathrm{Tr}(F_{\mu\nu} F^{\mu\nu})\, d^4x \tag{2.25}$$

Semantic Drift Bound (AMC):

$$\mathcal{L}_{\text{drift}} = \alpha \cdot d_{\text{Procrustes}}(T(v), v') + (1-\alpha) \cdot D_{KL}(p \| q) < \epsilon \tag{2.26}$$

Graph Laplacian (Format Compatibility):

$$\mathcal{L} = D - A \tag{2.27}$$

"The curvature of space is the skeleton of matter."

— John Archibald Wheeler

Chapter 3

The Heat Kernel

I was one of the founding engineers at Pandora.

This was 2009, before streaming was obvious, before Spotify launched in the United States, before the idea of "personalized radio" meant anything to most people. Tim Westergren had a vision: what if you could describe every piece of music by its DNA? Not genre, not artist, not what decade it came from—but the actual musical properties. The tempo, the instrumentation, the vocal harmonics, the rhythmic complexity, the use of minor keys, the density of the arrangement. Hundreds of attributes, each one scored by trained musicologists listening to every song.

We called it the Music Genome Project.

The idea was simple in principle, obsessive in execution. Every song that entered our catalog was analyzed by a trained listener who scored it on up to 450 distinct attributes. Is there a driving rhythm? Score it 1 to 5. Prominent use of acoustic guitar? Score it. Complex vocal harmonies? Score it. Baroque compositional techniques? Score it. Each song became a vector—a point in a 450-dimensional space where each dimension represented one musical gene.

My job was to make recommendations work. Given a seed song—something the user said they liked—find other songs they would probably like. The naive approach is obvious: compute the distance to every other song in the genome space, return the nearest neighbors.

It does not work.

The problem is that musical taste is not a point. It is a region, a cloud, a distribution with complex geometry. Someone who loves Miles Davis might also love John Coltrane—nearby in jazz space, an easy recommendation. But the same person might also love Radiohead—far away in genre space, connected by some deeper thread of harmonic sophistication and emotional atmosphere that the naive distance metric cannot see.

The genome space is curved. The straight-line distance between two songs is not the same as the *musical* distance—the distance that matters for whether a listener will accept the recommendation.

I spent months trying to fix this with better distance metrics. Weighted Euclidean distance. Mahalanobis distance. Custom kernels that emphasized certain genes over others. Nothing worked consistently. The recommendations were either too safe—endless variations on the seed song—or too wild—jarring jumps that broke the listener's flow.

Then someone suggested diffusion.

Here is the idea. Instead of jumping directly to the nearest neighbor, let the recommendation *spread*. Start at the seed song. Let "heat" flow outward through the genome space, following the natural connections between songs. Watch where the heat goes. The songs it reaches are the recommendations.

Think of dropping ink into water. The ink does not teleport to nearby regions. It spreads, following the geometry of the container. In a long thin tube, the ink spreads in one dimension. In a wide shallow dish, it spreads in two. In a complex branching vessel, it pools in the chambers, finds the paths of least resistance, reveals the shape of the space by how it flows.

For recommendation, I use that spreading as a model. Start with a score at the seed song and let it pass through the links to other songs. How much reaches a track depends on the links, their weights, and how long I let the calculation run. A song with no path from the seed cannot receive a score through this model. That tells me about the graph I built; it does not tell me that the listener could never like the song.

The diffusion pattern gives me candidate recommendations. A higher score gives me a reason to try a song. The listener still has to tell me, by playing it, skipping it or returning to it, whether the model caught something they actually enjoy.

This worked. Not perfectly—nothing works perfectly in recommendation systems—but better than anything else we had tried. The recommendations felt more organic, more like a knowledgeable friend building a playlist than an algorithm computing distances.

I did not understand why it worked. Not really. I had an intuition about "letting the music breathe" and "following natural connections," but I could not have told you the mathematics. I was doing spectral geometry without knowing the words.

Years later, I learned the mathematics.

The way heat spreads on any space—flat, curved, finite, infinite—is determined by a mathematical object called the *Laplacian*. The Laplacian is an operator that measures, at each point, how a function differs from its local average. Where the temperature is higher than the average of nearby points, the Laplacian is negative. Where it is lower, the Laplacian is positive. The heat equation says: flow from high to low, at a rate proportional to the difference.

The Laplacian has a *spectrum*: numbers describing how its characteristic spatial patterns respond. An eigenfunction is one such pattern. Apply the operator and the pattern keeps its shape, multiplied by its eigenvalue. Heat flow damps that pattern at a rate set by the eigenvalue. A vibrating membrane uses the same spatial patterns under suitable boundary conditions, but its frequencies are proportional to the square roots of the eigenvalues. The operator's spatial information enters two different evolution laws.

Here is the remarkable fact: *you can hear the shape of a drum*. Or almost. The frequencies encode the geometry. The frequencies can reveal features of the shape, but they need not identify one unique shape. Later in this chapter we will meet different drums with the same spectrum. Listening gives us information; we still have to ask how much it determines.

At Pandora, we were doing this without knowing it. The structure of the recommendation graph—which songs connected to which, with what strength—determined the eigenvalues of the graph Laplacian. Those eigenvalues determined how recommendations would spread. The spectrum was shaping the user experience, even though no one was explicitly computing eigenvalues.

The mathematics was there, underneath, doing its work whether we understood it or not.

And here is the connection to everything that follows.

For diffusion on a connected finite graph with a symmetric Laplacian, the constant mode has eigenvalue zero and the first positive eigenvalue measures the slowest nonconstant decay. The

corresponding statement on a compact manifold requires closed or Neumann boundary conditions. Other operators and boundaries have different spectra.

In this connected heat-flow model, a large spectral gap means that every nonconstant pattern decays quickly in the specified norm. A small gap permits a slowly decaying pattern. The gap is a rate; to turn it into a time to settle, I also need the starting size of the disturbance and how small I am willing to call "settled.

Picture two rooms joined by a narrow doorway. Air within either room can mix while a difference between the rooms lingers. A diffusion model can represent that slow exchange. The corresponding eigenfunction records a contrast between the rooms; its small eigenvalue records the slow rate at which that contrast fades.

At Pandora, a small spectral gap in the recommendation graph meant users would get stuck in filter bubbles. They would hear the same kinds of music over and over, never discovering new regions of the space. A large spectral gap meant healthy exploration—recommendations that could bridge genres, connect disparate tastes, keep the experience fresh.

The spectral gap of the music genome determined the quality of the product.

In Yang–Mills theory the gap of interest belongs to the physical Hamiltonian. The diffusion examples in this chapter introduce the spectral ideas we will use to discuss it.

El mismo matemático. Diferentes apuestas.

The same mathematics. Different stakes.

I am going to teach you the heat kernel—the mathematical object that describes how heat spreads on any space. I will show you how the heat kernel encodes the geometry of the space, how the eigenvalues of the Laplacian determine the diffusion, and how spectral gaps relate to mass gaps.

By the end of this chapter, you will understand one of the most powerful tools in modern geometry: the ability to *hear* the shape of a space by listening to how it rings.

Let me show you how.

§ 3.1

Following heat through a space

Heat a metal rod at one end, then remove the heater and insulate the rod so heat cannot escape. What happens?

Heat moves from the hotter part toward the colder part. The hot end cools down. The cold end warms up. In this insulated model, the whole rod approaches the same temperature—thermal equilibrium. Keeping a heater attached or letting heat escape would give us a different problem.

Microscopic interactions redistribute energy. In a metal, electrons as well as the lattice can carry it. Thermal equilibrium means the temperature has settled, not that every atom has exactly the same motion. The heat equation is a macroscopic model for the resulting temperature profile under its material and scale assumptions.

The rate depends on material properties, dimensions and boundary conditions. A longer insulated rod of the same material takes longer to smooth a comparable disturbance. A branch

introduces an additional path for heat. Simply bending an ideal uniform one-dimensional rod without changing its length or material does not change its intrinsic one-dimensional Laplacian. I need to identify which feature of the physical setup changed before crediting curvature with the result.

Before writing the equation, I need to decide what happens at the rod's ends. Insulated ends retain the heat and permit a final uniform temperature. Ends held at a fixed temperature exchange heat with the surroundings and produce a different long-time result. The equation below gives the evolution inside the rod; the initial temperature and the boundary conditions complete the problem.

Fourier's heat equation captures the local relationship:

$$\frac{\partial u}{\partial t} = \alpha \nabla^2 u \tag{3.1}$$

Here $u(x, t)$ is the temperature at position x and time t. The constant α is the thermal diffusivity—a property of the material that determines how quickly heat moves through it. And ∇^2 is the *Laplacian*—the operator that will become the star of this chapter.

The Laplacian measures how a quantity at a point differs from its average in the surrounding neighborhood. It is the mathematical formalization of "hotter than your neighbors" or "lower than the surrounding terrain."

In one dimension, the Laplacian is just the second derivative:

$$\nabla^2 u = \frac{d^2 u}{dx^2} \tag{3.2}$$

This measures the *curvature* of the function u. Where u curves upward (concave up), the Laplacian is positive. Where u curves downward (concave down), the Laplacian is negative. Where u is linear, the Laplacian is zero.

Think of a temperature profile along a rod. If there is a hot spot—a peak in the temperature graph—then the second derivative is negative at that point. The heat equation says $\partial u/\partial t < 0$: the hot spot cools down. If there is a cold spot—a valley in the temperature graph—then the second derivative is positive, and $\partial u/\partial t > 0$: the cold spot warms up.

Heat flows from peaks to valleys. The Laplacian is the driver.

In higher dimensions, the Laplacian is the sum of second derivatives in each direction:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \tag{3.3}$$

This measures how u at a point compares to the average of u in a small ball around that point. Positive Laplacian means the point is lower than its neighbors—a valley. Negative Laplacian means the point is higher than its neighbors—a peak.

The interpretation is the same. Heat flows from where you are hotter than average to where you are colder than average. The Laplacian quantifies the imbalance.

On a flat space—Euclidean $\mathbb{R}^n$—the Laplacian is straightforward. On a curved manifold, it becomes more interesting.

The problem is that "average over a neighborhood" depends on what you mean by "neighborhood," and on a curved space, neighborhoods are shaped by the curvature. The Laplacian must account for the metric—the way the manifold measures distances and angles.

The generalization is called the *Laplace-Beltrami operator*. In coordinates:

$$\Delta u = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|}\, g^{\mu\nu} \partial_\nu u \right) \tag{3.4}$$

Here $g_{\mu\nu}$ is the metric tensor, $g^{\mu\nu}$ is its inverse, and $|g|$ is the absolute value of its determinant.

Do not be intimidated by this formula. Let me give you the intuition.

Stay with temperature for a moment, but replace the continuous rod with a row of connected temperature readings. In a simple diffusion model, each reading changes toward the values at its neighbors. A reading above both neighbors is a local peak. With $\Delta = \operatorname{div} \operatorname{grad}$, that peak has negative Laplacian and diffusion lowers it. A local valley has positive Laplacian and rises. Graph theory often uses the opposite-sign operator $L = D - A$; its heat equation is $\dot{u} = -Lu$. The spreading is the same once the sign convention is kept straight.

The graph example shows the local exchange, but the formula above describes a continuous surface or space. There, coordinate steps need not have the same physical length or meet at right angles. The $g_{\mu\nu}$ terms describe those lengths and angles. The $\sqrt{|g|}$ terms convert coordinate volume into geometric volume. Before comparing how much heat enters and leaves a small region, I have to know how large that region actually is.

The recommendation graph at Pandora gave me another use for this spreading idea. A song's "heat" was a numerical score for its relevance to the current playlist, rather than a physical temperature. The Laplacian told us whether a song was hotter or colder than its musical neighbors. A high score identified a candidate recommendation. The Laplacian described how that score would change through exchange with neighboring songs. A zero Laplacian meant those local contributions balanced at that moment; it did not certify that the listener would accept the song. The connections in the music graph determined how the scores spread.

The key point is that the Laplacian on a curved manifold *knows about the curvature*. It encodes geometric information. When you solve the heat equation on a curved space, the curvature affects how heat spreads.

Consider heat diffusion on a sphere versus on a saddle surface. On the sphere, geodesics converge—parallel lines eventually meet at the poles. This "crowding" of space affects how heat spreads. On a saddle, geodesics diverge—parallel lines spread apart. This "spreading" of space has the opposite effect.

The Laplace-Beltrami operator captures these effects automatically. It is the bridge between the analytical tool (the differential operator) and the geometric reality (the curved manifold).

This is why we can use diffusion to probe geometry. The heat equation on a manifold is not just a physical law—it is a geometric probe. By watching how heat spreads, we learn about the shape of the space.

We have a rule for the rate of change. To see what it predicts, put the initial heat in one small place and follow where it goes. The heat equation tells us *how* heat evolves. The *heat kernel* tells us the complete solution.

Imagine depositing a very small heat pulse near a marked point y on a cold surface. The mathematical idealization squeezes a unit amount into that point. The Dirac delta is a distribution with unit integral, not an ordinary function describing a physically infinite temperature that I could prepare in a laboratory.

The heat kernel $K(x, y, t)$ is the response density at x and time t to that ideal unit pulse at y. To get the amount in a small region, I integrate the density over the region using the surface's volume measure. To evolve an extended initial profile, I add the responses from all its points, weighted by the initial values:

$$u(x,t) = \int_M K(x,y,t)u(y,0)\,dV(y).$$

That is what makes the kernel useful: solve for the unit pulse once, then assemble other initial conditions from those responses. In the formulas that follow, the diffusivity is absorbed into the time scale, so the generator is Δ. With constant physical diffusivity α, replace t by αt.

On flat Euclidean space $\mathbb{R}^n$, the heat kernel has a beautiful closed form:

$$K(x,y,t) = \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/4t} \tag{3.5}$$

This is a Gaussian—the bell curve, spreading outward over time. At $t = 0$, all the heat is concentrated at y (the delta function). As t increases, the peak flattens, the width grows, the heat diffuses outward in all directions.

The formula reveals important physics:

- The factor $(4\pi t)^{-n/2}$ normalizes the distribution—total heat is conserved.
- The exponent $|x - y|^2/4t$ shows that heat spreads diffusively, not ballistically. The characteristic distance grows like $\sqrt{t}$, not t.
- The dependence on dimension n appears: in higher dimensions, the heat spreads into more directions, so the kernel decays faster.

Here is the remarkable fact: the heat kernel contains *all* the geometric information about a manifold.

Watch the heat kernel on a curved manifold. Compare it to the flat-space Gaussian. The deviations encode the curvature.

For short times—when the heat has not spread far—the heat kernel on the diagonal ($x = y$) has an asymptotic expansion:

$$K(x,x,t) \sim \frac{1}{(4\pi t)^{n/2}} \left(1 + \frac{R(x)}{6}t + O(t^2)\right) \tag{3.6}$$

Here $R(x)$ is the scalar curvature at point x—the same quantity from Chapter 2.

The curvature appears in the heat kernel. Positive scalar curvature means the heat kernel is slightly larger than the flat-space value—space is "pinched," and heat has less room to spread, so it stays concentrated. Negative scalar curvature means the heat kernel is smaller—space is "spread out," and heat has more room to diffuse.

By measuring the short-time behavior of the heat kernel, you can determine the curvature. By measuring higher-order terms in the expansion, you can determine higher-order geometric invariants. The heat kernel is a complete geometric probe.

This is the mathematical content of "you can hear the shape of a drum." The drum's shape determines its heat kernel. The heat kernel determines the spectrum. The spectrum is what you hear when the drum vibrates. Geometry becomes sound becomes analysis.

At Pandora, we used something called *diffusion maps*—a technique that explicitly uses the heat kernel to embed high-dimensional data in lower dimensions.

The idea: construct a similarity kernel between data points. For any two songs i and j, define a kernel value K_{ij} that measures how "similar" they are in the genome space. Interpret this kernel as a heat kernel on some underlying manifold. Then use the eigenvectors of the kernel to find coordinates on that manifold.

Diffusion maps construct coordinates from a chosen neighborhood kernel on the data. They can reveal lower-dimensional organization under appropriate sampling and geometric assumptions. Kernel scale, normalization and sample density affect the result. A music-feature cloud does not arrive with a certified smooth manifold hidden inside it; the construction gives a model whose usefulness we test against the task.

At AMC, I used similar ideas without calling them "heat kernels" or "diffusion maps." The graph Laplacian of the format compatibility network, the spectral analysis of translation paths—these were discrete versions of heat kernel methods. I was computing eigenvalues of matrices to understand the structure of a space.

The mathematics keeps appearing, whether you name it or not.

§ 3.2

The patterns that fade at different rates

The Laplacian has a spectrum—a collection of eigenvalues and eigenfunctions that describe the natural "vibrations" of the manifold.

The eigenvalue equation is:

$$-\Delta\phi_k = \lambda_k\phi_k \tag{3.7}$$

Here ϕ_k is an eigenfunction and λ_k its eigenvalue. With the stated sign convention, $-\Delta$ is nonnegative for the closed, Neumann and Dirichlet settings used here. Neumann conditions specify the normal derivative—zero means an insulated boundary. Dirichlet conditions specify the value itself, as when the edge of a membrane is held fixed. These choices are part of the operator. They cannot be dropped when comparing its spectrum.

What does this equation say? It says that ϕ_k is a special function: when you apply the Laplacian to it, you get the same function back, multiplied by a constant. The Laplacian stretches ϕ_k by a factor of λ_k, but does not change its shape.

These special functions are the "natural modes" of the manifold. They are the patterns that, once established, persist in time (up to exponential decay). They are the resonances of the space.

On a connected compact manifold without boundary, or on a connected compact domain with Neumann boundary conditions,the eigenvalues form a discrete sequence:

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \to \infty \tag{3.8}$$

In this connected closed setting, or with Neumann boundary conditions on a connected compact domain, the first eigenfunction is constant and its eigenvalue is zero. A membrane held fixed at its boundary obeys Dirichlet conditions instead: its first eigenvalue is positive. For the wave equation, vibration frequencies are proportional to $\sqrt{\lambda}$; for heat flow, λ itself gives a decay rate.

The subsequent eigenvalues describe the nonconstant spatial modes. For heat flow with the time normalization above, λ_1 is the slowest nonzero decay rate. If that mode begins with coefficient a_1, its coefficient at time t is $a_1 e^{-\lambda_1 t}$. A larger eigenvalue makes the corresponding pattern fade faster.

A real eigenfunction may have positive and negative regions separated by its zero set, called a *nodal set*. For temperature, a negative coefficient or region can represent a deviation below a reference temperature. It does not require a negative absolute temperature. Higher modes can carry finer spatial variation, although the exact number and layout of nodal regions depend on the space and the eigenfunction.

Think of vibrating a string. The fundamental mode has one "hump"—the whole string moves together. The second mode has two humps with a node in the middle. The third mode has three humps with two nodes. Each mode vibrates at a different frequency, and the frequency increases with the number of nodes.

The string makes the pattern easy to see. On a more general manifold we still have eigenfunctions, but their nodal geometry can be much more complicated. For the wave equation, the physical wave speed also enters the frequency. For the heat equation, the same spatial modes decay instead of oscillating.

Stretch a membrane over a frame. Strike it. It vibrates.

The vibrations are not arbitrary. Under the ideal membrane model, fixed geometry, tension and mass density determine the ringing frequencies. The square roots of the Laplacian eigenvalues supply their geometric part. I need to keep the material and boundary conditions fixed before attributing a frequency change to shape.

A circular drum has one set of frequencies. A square drum has another. An elliptical drum has yet another. The shape determines the sound.

This is why different drums sound different, even when struck the same way. The physics is the same—the wave equation on a membrane. The geometry is different. The geometry determines the spectrum. The spectrum determines the timbre.

The eigenfunctions of the circular drum are Bessel functions—oscillating patterns with circular symmetry. The eigenfunctions of the square drum are products of sines. The eigenfunctions of an irregular drum are more complicated, but they still exist, still form an orthonormal basis, still encode the geometry.

In 1966, the mathematician Mark Kac asked a famous question: "Can one hear the shape of a drum?"

If you know all the eigenvalues—all the frequencies at which the drum can ring—can you reconstruct the shape?

The question fascinated mathematicians for decades. The answer, it turns out, is: almost, but not quite.

In 1992, Carolyn Gordon, David Webb, and Scott Wolpert constructed two different drum shapes with exactly the same spectrum—*isospectral* manifolds. If you could only hear the frequencies, you could not tell these drums apart. The shapes are different, but the sounds are identical.

So you cannot *always* hear the shape. But you can hear a lot.

The eigenvalues encode:

- The area of the drum (from eigenvalue density—Weyl's law)
- The perimeter (from the next term in the expansion)
- The number of holes (from topological constraints)
- Much of the curvature (from spectral invariants)

The failure of unique reconstruction is a useful warning. A spectrum can constrain geometry without identifying it completely. Which features it determines depends on the class of domains and boundary conditions. If my task needs the exact shape, I need additional measurements or a uniqueness argument. The full heat kernel retains location-dependent information through the eigenfunctions as well as the eigenvalues; a list of frequencies has already discarded that part.

Here is the connection between eigenfunctions and heat kernels.

Choose a real orthonormal eigenbasis for the compact scalar problem. Orthonormal means that integrating the product of two different modes gives zero, and integrating the square of one gives one. This normalization makes the expansion coefficients unambiguous. The heat kernel can then be written as a sum over those modes:

$$K(x, y, t) = \sum_{k=0}^{\infty} e^{-\lambda_k t} \phi_k(x) \phi_k(y) \tag{3.9}$$

Each eigenfunction contributes a term that decays exponentially in time, with decay rate given by the eigenvalue.

At short times, many modes are needed to reconstruct a sharply localized pulse. High eigenvalues are less strongly suppressed than they will be later, so fine spatial detail remains visible. The factor $e^{-\lambda_k t}$ tells me how much each mode survives. Its contribution also depends on the values of the eigenfunction at the two locations.

At long times (t large), only the lowest eigenvalues matter. The higher modes have decayed away. Since $\lambda_0 = 0$, the sum is dominated by

$$e^{-\lambda_0 t} \phi_0(x) \phi_0(y) = \phi_0(x) \phi_0(y),$$

which is just the constant—uniform distribution.

The intermediate behavior is controlled by λ_1—the first non-zero eigenvalue. This determines how fast the system approaches equilibrium. Large λ_1 means fast approach. Small λ_1 means slow approach.

This is why the spectral gap—the difference $\lambda_1 - \lambda_0 = \lambda_1$—is so important. It controls the transition from non-equilibrium to equilibrium. It determines the mixing time.

Before taking the spectrum into physics, I want a small example where we can inspect every pattern. Four sensor readings are enough to show what filtering can hide. Put four sensors around a loop and suppose each reports a dimensionless deviation. I want to smooth the reports by letting each one trade a little of its difference with its two neighbors. When one value is high and the others are zero, the first thing I expect to see is its two neighbors rise. I also want to know when the pattern will be too faint to identify which sensor started high.

Label the positions $0, 1, 2, 3$ in order around the loop. Let τ be time measured in units of the mixing rate. The rule is

$$\frac{du_j}{d\tau} = u_{j-1} + u_{j+1} - 2u_j, \qquad j \text{ taken modulo } 4.$$

Each exchange gives one sensor what it takes from the other, so the total $\sum_j u_j$ stays fixed. In matrix form $\dot{u} = -Lu$, where L is the degree matrix minus the adjacency matrix. The heat operator $e^{-\tau L}$ takes the four readings at the start to the four readings at time τ.

A mode is a pattern whose shape stays put while its amplitude changes. The four patterns below form a basis. The constant pattern never changes. The alternating pattern decays fastest because every edge has a difference to work on.

Pattern	Eigenvalue of L	Multiplier at τ
$(1, 1, 1, 1)$	0	1
$(1, 0, -1, 0)$	2	$e^{-2\tau}$
$(0, 1, 0, -1)$	2	$e^{-2\tau}$
$(1, -1, 1, -1)$	4	$e^{-4\tau}$

To check the last row, start at a vertex with value 1. Its neighbors are both -1. The mixing rule gives a derivative of -4 there. At a vertex with value -1 the derivative is $+4$. The alternating vector solves the differential equation after multiplying the whole pattern by $e^{-4\tau}$.

For a unit reading at sensor 0 and zero everywhere else, I can build the initial vector by adding a quarter of the constant mode, a half of the first opposite-pair mode, and a quarter of the alternating mode. At later times their different multipliers give

$$u(\tau) = \tfrac{1}{4} \begin{pmatrix} 1 + 2a + b \\ 1 - b \\ 1 - 2a + b \\ 1 - b \end{pmatrix}, \qquad a = e^{-2\tau}, \quad b = e^{-4\tau}.$$

Choose $\tau = (\log 2)/2$. The opposite-pair mode has half its amplitude left, and the alternating mode has a quarter. The readings are $(9, 3, 1, 3)/16$. They still total one. Sensor 0 still has the largest value. Averaging all four would give $1/4$ at every time and lose that distinction immediately.

Filtering has changed the sizes of the modes. Recovery now depends on whether the small ones remain distinguishable from error. Compare a unit reading started at 0 with one started at 2. Their difference is the opposite-pair mode $(1, 0, -1, 0)$. At time τ the difference is $e^{-2\tau}$ times as large as it was at the start. In exact arithmetic I can multiply by $e^{2\tau}$ and restore it. A measurement error in the same mode is multiplied by that factor too. For the alternating mode the amplification is $e^{4\tau}$. The graph has told me which distinctions smoothing makes expensive to recover.

This example gives me a way to read a spectral gap in a system where I can see every step. The eigenvalues belong to the specified mixing operator on four vertices. In the gauge-field problem I will have to identify the operator whose spectrum represents physical energy. The patterns in a scatterplot and the evolution of a heat kernel can each be informative once I know which operation produced them.

§ 3.3

What the slowest pattern tells us

After the fast patterns have faded, the slowest nonconstant pattern controls the remaining approach to equilibrium. In the connected setting we have just used, its rate is the *spectral gap*:

$$\text{gap} = \lambda_1 - \lambda_0 = \lambda_1 \tag{3.10}$$

The second equality holds in the connected compact, closed or Neumann setting just specified. With Dirichlet conditions there is no constant zero mode. The operator and its boundary conditions have to travel with the number called a gap.

The spectral gap is a single number that encodes something profound about the geometry: how "well-connected" the manifold is. It measures the bottleneck. It quantifies the obstruction to mixing.

The spectral gap measures *mixing time*—how quickly the system forgets its initial conditions and approaches equilibrium.

Start with heat concentrated at one point. Wait. The heat spreads. Eventually, it becomes uniformly distributed. How long does "eventually" take?

For an initial mean-zero disturbance of size A in the L^2 norm, the spectral estimate bounds the remaining size by $Ae^{-\lambda_1 t}$. To guarantee a size at most $\delta < A$, it suffices that

$$t \geq \frac{1}{\lambda_1} \log \frac{A}{\delta}.$$

So $1/\lambda_1$ sets a time scale, while the requested accuracy supplies the logarithm. Saying "it mixes in ten seconds" without naming the tolerance leaves part of the claim unstated.

Here is the intuition. In this compact self-adjoint heat problem, the slowest nonconstant mode decays as $e^{-\lambda_1 t}$. Faster modes decay at least as quickly, so the mean-zero L^2 norm is bounded by its initial value times that factor. A single slowest mode attains the rate. At time $1/\lambda_1$ its amplitude has fallen by e^{-1}.

A large spectral gap means the $e^{-\lambda_1 t}$ term decays quickly. The system reaches equilibrium fast. A small spectral gap means the decay is slow. The system remembers its initial conditions for a long time.

The spectral gap appears everywhere:

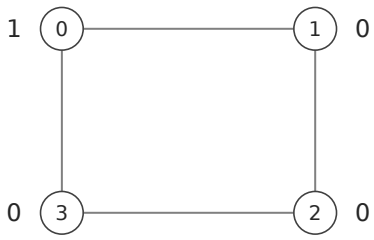
Recommendation systems: A large spectral gap in the recommendation graph means users quickly discover diverse content. The "heat" of their preferences spreads across many items. A

A reading spreads through four sensors

GRAPH HEAT FLOW / ONE NETWORK / TWO DECAY RATES

01 Initial reading

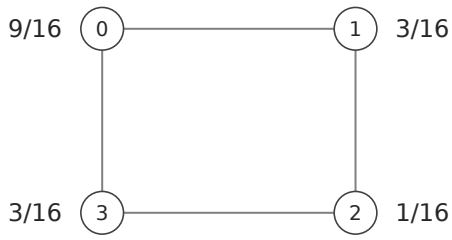
One unit at sensor 0.



$\sum_i u_i = 1$

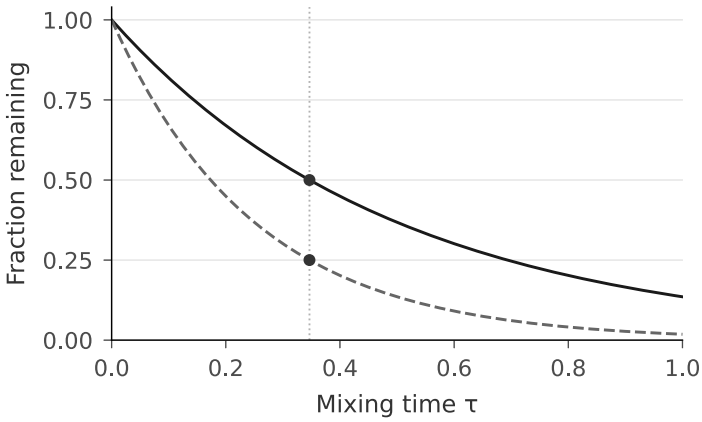
02 After one mixing interval

Time τ = (log 2)/2.



$u = (9, 3, 1, 3)/16$

03 Watch each pattern fade



04 Read the two modes

—— $e^{-2\tau}$

Opposite pair: (+1, 0, −1, 0).

½ remains at τ.

- - - - $e^{-4\tau}$

Alternating: (+1, −1, +1, −1).

¼ remains at τ.

Mixing preserves the sum while damping spatial differences.

Reversing this filter multiplies errors in these two modes by 2 and 4, respectively.

Figure 12. Heat flow on four linked sensors. The exchange rule preserves the sum of the readings. The two nonconstant decay rates determine how quickly the distinct patterns shrink, and how much measurement error is amplified if the filter is reversed.

small spectral gap means filter bubbles—users get stuck in local regions, never exploring the full space.

At Pandora, we wrestled with this constantly. Too small a gap, and users complained about repetitive recommendations. Too large a gap (which we achieved by adding artificial connections), and recommendations felt random, jarring, disconnected from the user's taste.

Sampling algorithms: Markov Chain Monte Carlo (MCMC) methods generate random samples from complicated probability distributions. The mixing time of the Markov chain determines how many steps you need before your samples are representative. The mixing time is controlled by the spectral gap of the transition matrix.

For an appropriate ergodic reversible chain, a useful gap bound controls convergence toward its target distribution. Samples at successive steps can remain dependent. A smaller gap can require more steps; an exponential increase in the problem size requires a corresponding size-dependent estimate, not merely the statement that the gap is small.

Neural networks: The loss landscape of a neural network is a high-dimensional manifold. The spectral gap of the Hessian at critical points affects training dynamics. Flat regions—small curvature, small effective gap—cause slow learning. Sharp valleys—large curvature—can cause instability.

Physics: For a quantum Hamiltonian, an energy eigenstate evolves in real time by a phase $e^{-iEt/\hbar}$; it does not decay merely because a gap is large. Euclidean evolution $e^{-t(H-E_0)}$ suppresses excited components at rates set by their energy differences (with $\hbar = 1$). Physical relaxation requires a specified dynamics and coupling to other degrees of freedom.

There is a beautiful connection between the spectral gap and the geometry of the manifold, discovered by Jeff Cheeger in 1970.

Define the *Cheeger constant* (also called the isoperimetric constant):

$$h = \inf_S \frac{\text{Area}(S)}{\min(\text{Vol}(A), \text{Vol}(B))} \tag{3.11}$$

Here S is any hypersurface that divides the manifold into two pieces A and B. The infimum is over all such dividing surfaces.

What does this measure? The Cheeger constant is the "minimum cut"—the smallest surface (relative to the volumes it separates) that can divide the manifold in two. A large Cheeger constant means you cannot divide the manifold cheaply; every cut has to pass through a lot of material relative to the pieces it creates. A small Cheeger constant means there is a bottleneck—a thin neck connecting two large regions.

Think of an hourglass. The thin waist has small area, but it separates two large bulbs. The Cheeger constant is small. The bottleneck is severe.

Think of a ball. Any cut that divides it must pass through substantial cross-section. The Cheeger constant is large. There are no bottlenecks.

Cheeger's inequality says:

$$\lambda_1 \geq \frac{h^2}{4} \tag{3.12}$$

For a connected closed Riemannian manifold this is Cheeger's lower bound. An upper bound requires additional hypotheses. For example, under $\text{Ric} \geq -(n-1)k^2 g$ in dimension $n \geq 2$, Buser

gives $\lambda_1 \leq 2(n-1)kh + 10h^2$. The curvature assumption and the squared term matter; the graph bound $\lambda_1 \leq 2h$ cannot be transplanted to this Riemannian normalization.[1]

This inequality is profound. It says that you can bound an analytical quantity—the spectral gap, which requires solving a differential equation—by a geometric quantity—the Cheeger constant, which requires finding the minimum cut. Sometimes the geometry is easier to analyze than the analysis.

The hourglass picture is still useful. A narrow neck can slow the exchange between two large regions, and a trial function that is almost constant on each side can prove an upper bound. The geometric assumptions determine the estimate. A small value of h alone, inserted into the lower bound, does not prove that λ_1 is small.

§ 3.4

Which operator has the gap?

Now we connect to physics.

In quantum mechanics, the Hamiltonian H is an operator whose eigenvalues are the possible energies of the system:

$$H|\psi_k\rangle = E_k|\psi_k\rangle \tag{3.13}$$

The eigenvalues E_k are energy levels. An energy eigenstate evolves by an overall phase under the time-independent Hamiltonian, so its physical pure state is stationary. It is a quantum state, not a single classical configuration. Its wavefunction can assign amplitudes to many configurations at once.

The vacuum has the lowest energy E_0. If it is unique, the gap is the infimum of the energies above E_0 in the subspace orthogonal to it, measured relative to E_0. In a discrete spectrum with a first excited level E_1, this is $\Delta = E_1 - E_0$.

For a simple particle model, the kinetic term is proportional to $-\Delta$ and V is the potential-energy function. After choosing units that absorb the kinetic coefficient, we can write:

$$H = -\Delta + V \tag{3.14}$$

where V is a potential energy term. The kinetic energy is (minus) the Laplacian—it measures how rapidly the wave function varies in space. The spectrum of H inherits structure from the spectrum of Δ.

The mass gap is a spectral gap of the physical Hamiltonian above its vacuum. A graph Laplacian or a diffusion operator can have a gap of its own. To use one as evidence about the Hamiltonian, we need an explicit connection between their spectra.

[1] Peter Buser, "A note on the isoperimetric constant," *Annales scientifiques de l'ENS* 15 (1982), 213–230, https://doi.org/10.24033/asens.1426.

This is why spectral geometry matters for physics. The tools we develop in this chapter—heat kernels, eigenvalue estimates, Cheeger inequalities—are exactly the tools we need to understand the mass gap.

In Yang-Mills theory, the configuration space is the space of all gauge field configurations (modulo gauge transformations). The Hamiltonian acts on this space. The eigenvalues are the energies of possible states—vacuum, single-particle excitations, multi-particle states.

If there is no mass gap, there are physical excitations with energies arbitrarily close to the vacuum energy. That statement is about the gauge-invariant spectrum.

If there is a mass gap, every physical excitation costs at least Δ above the vacuum. In pure Yang–Mills theory an external pair of test charges lets us ask a different question. How does the energy associated with the pair change as we pull the charges apart?

Confinement concerns the separation of color charges. The behavior of large Wilson loops and the spectrum above the vacuum are related subjects, with different observables and proof obligations.

We cannot compute the spectrum of Yang-Mills theory analytically—the configuration space is too complicated, the quantum fluctuations too strong. But we can compute it numerically, using lattice gauge theory.

The idea: discretize spacetime into a finite lattice. Replace the continuous gauge field with matrices living on the links of the lattice. The finitely many link matrices still range continuously over SU(3). Monte Carlo approximates that continuous finite-dimensional integral by a finite sample average.

The lattice regularization breaks some symmetries, but they are restored in the continuum limit as the lattice spacing goes to zero. If you are careful, lattice calculations give reliable predictions for the real theory.

Lattice calculations provide substantial numerical evidence for a positive glueball mass scale in pure gauge theory; the frequently quoted lightest scalar-glueball estimate is around 1.7 GeV under a chosen physical scale setting.[2] Mesons belong to theories that include quarks and answer a different spectroscopy question. Numerical evidence for particular levels motivates the continuum problem. It does not establish the absence of all lower physical spectrum by theorem.

But numerical evidence is not a proof. The Millennium Prize Problem requires a rigorous mathematical demonstration that:

1. Yang-Mills theory exists as a well-defined quantum field theory (not just a formal perturbation series)

2. The mass gap $\Delta > 0$ (not just numerically observed, but mathematically proven)

Hold me to both halves. That proof is not in this book. The spectral geometry in this chapter gives me the language. In Chapter 5 I show what a spacing among my lattice measurements can say, and what a spectral claim would still require.

The Yang–Mills configuration space has infinitely many dimensions. Every point in spacetime has a gauge field value, so specifying a configuration requires specifying infinitely many numbers—one for each point, for each component of the gauge field.

[2]Colin J. Morningstar and Mike Peardon, "The glueball spectrum from an anisotropic lattice study," *Physical Review D* 60 (1999), 034509, Section VI and Table VII, `https://doi.org/10.1103/PhysRevD.60.034509`. Their conversion to physical units includes a separately estimated scale uncertainty.

Can we do heat kernel analysis on an infinite-dimensional space?

The finite models give us well-defined operators to calculate with. Passing to infinitely many degrees of freedom is a further construction. We need a state space, a measure or operator domain, and a limiting procedure that preserves the claimed properties. In four-dimensional continuum Yang–Mills theory, constructing the quantum theory and establishing a positive mass gap are precisely the unresolved demands. I cannot assume that structure into existence while using it to prove the result.

The key tool is the *path integral*—Richard Feynman's formulation of quantum mechanics.

In the path integral approach, you do not solve differential equations to find the wave function. Instead, you sum over all possible configurations, with the displayed Euclidean formulation weighted by a real Boltzmann factor when S is real (we use $\hbar = 1$):

$$Z = \int \mathcal{D}A\, e^{-S[A]} \tag{3.15}$$

Here $\mathcal{D}A$ represents integration over all field configurations, and $S[A]$ is the action—a functional that assigns a number to each configuration.

This integral is the heat kernel in disguise.

The partition function Z can be written as a trace over the heat kernel:

$$Z = \mathrm{Tr}(e^{-\beta H}) = \sum_k e^{-\beta E_k} \tag{3.16}$$

Here β is inverse temperature in units with Boltzmann's constant set to one; it is not the lattice-coupling parameter that later also uses the letter beta. The trace formula assumes $e^{-\beta H}$ is trace class. In a regulated discrete-spectrum setting, it adds one Boltzmann factor for each energy eigenstate, including degeneracies. The corresponding Euclidean path integral uses histories periodic in Euclidean time for this thermal trace.

At large β (low temperature), only the ground state contributes: $Z \approx e^{-\beta E_0}$. The system freezes into the vacuum.

At small β (high temperature), all states contribute. The sum probes the full spectrum.

The path integral is a generating function for the spectrum. By studying how Z depends on β, you can extract the eigenvalues. The spectral gap appears as the rate of exponential decay: $Z \sim e^{-\beta E_0}(1 + e^{-\beta\Delta} + \cdots)$ where $\Delta = E_1 - E_0$ is the gap in this discrete example with a unique ground state and a single first excited state.

The trace has summed over the starting locations in the kernel. It preserves spectral information while discarding the separate locations. A matrix element of e^{-tH}, its trace and a geometric heat kernel are related constructions, but I still need to identify the operator and the boundary conditions for the problem in hand.

The standard approach to the mass gap problem uses the path integral, but struggles to extract rigorous bounds. The integral is over an infinite-dimensional space. The measure $\mathcal{D}A$ is not rigorously defined in most formulations. The mathematics is formal—useful for calculations, unreliable for proofs.

The Davis framework provides a different angle.

Instead of integrating over all configurations, we organize configurations by their gauge-invariant content—the Wilson loops and topological charges that capture the physical information. Configu-

rations that produce the same Wilson loops (up to gauge transformation) are grouped into the same "bin."

The same issue returns if I store fewer readings. The next example asks exactly which pattern those averages can still reveal. Go back to the four sensors around the loop. Suppose the logger saves just two values. The first is the average at sensors 0 and 2. The second is the average at 1 and 3. Write P for the map that takes those averages:

$$Pu = \frac{1}{2}\begin{pmatrix} u_0 + u_2 \\ u_1 + u_3 \end{pmatrix}.$$

If I call the saved averages v_0, v_1, their evolution is closed. The two neighbors of either even sensor are the odd sensors. Averaging the two even equations gives $\dot{v}_0 = 2(v_1 - v_0)$, and averaging the odd equations gives the reverse. In matrix form,

$$\dot{v} = -L_{\text{pair}} v, \qquad L_{\text{pair}} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}, \qquad PL = L_{\text{pair}} P.$$

The matrix equality says I can compute the changes in the two averages from the saved values alone. Starting from $Pu(0)$, the two-equation system reproduces $Pu(\tau)$ at every later time. I can check that by differentiating $Pu(\tau)$ and using uniqueness for the resulting linear system.

What is the slowest decay I will see on that logger? The two saved averages approach one another with rate 4. On the original loop the slowest nonconstant modes have rate 2. Here is what the logger sees:

Pattern on the loop	Decay rate	Pair averages
$(1, 1, 1, 1)$	0	$(1, 1)$
$(1, 0, -1, 0)$	2	$(0, 0)$
$(0, 1, 0, -1)$	2	$(0, 0)$
$(1, -1, 1, -1)$	4	$(1, -1)$

The two rate-2 patterns cancel in the averages. If all I can see is the logger, I can be exactly right about the behavior it measures and still miss the slowest behavior on the loop. This is a concrete failure mode for a compressed calculation. Even exact evolution of the recorded quantities leaves a question about the modes that never reached the record.

That question follows me to the lattice. Binning configurations by selected Wilson loops and charges gives me a smaller space in which I can compare the configurations I sampled. A dynamics on those bins needs its own justification. A lower bound for the full spectrum also needs control of the excitations the bins cannot see. In Chapter 5 I follow a spacing in a plot back to its inputs and then make the additional calculation a spectral claim would require.

La reducción es la clave. The reduction is the key.

§ 3.5

Bringing the measurement back to the task

At Pandora, I learned to look at the connections between songs. Suppose a listener jumps from a country record to a soul record because the rhythm carries over. On a graph that puts a link between tracks with similar musical features, there may be many routes between those two parts of the collection. Other tracks may have only one or two links out. A walk on that graph can tell me how easily the *model* moves between neighborhoods. To learn whether it describes a listener, I have to compare it with what they play, skip, and return to.

We did not call any of that a spectral gap at Pandora. We talked about adventurousness and diversity. I learned something useful from that language. Before I can measure how a signal moves, I have to decide whose signal it is and what counts as a connection.

A useful pattern in a graph still has to reach the task that made me build the graph. That question followed me from recommendations into security work. In my security work at IBM I came back to the question of what a system can observe. A local problem can be plain in one log and invisible in the record another team receives. To use the picture of diffusion, I would first have to specify what information travels, which connections carry it, and what happens to the information along the way. The spectrum describes the motion in that model. I still have to check what a person can see on the other end.

And now, sitting in my living room with a laptop and a dog, I understand something about understanding itself.

Knowledge is a manifold. Concepts are points. The connections between concepts—the analogies, the implications, the shared structures—are the paths through the space.

Learning is diffusion. You start with a new idea (a point source of heat). You let it spread through your existing knowledge. Where it connects, understanding grows. Where it does not connect, confusion persists.

A large spectral gap in your knowledge manifold means you learn quickly—new ideas rapidly integrate with what you already know. A small spectral gap means slow learning—new ideas stay isolated, fail to connect, fade away.

This is why teaching matters. A good teacher does not just transmit information. A good teacher reshapes the student's knowledge manifold—adds connections, removes bottlenecks, increases the spectral gap. The same information, presented differently, can connect or not connect. The geometry determines whether learning happens.

I am trying to be a good teacher in this book. I am trying to add connections—between the scanner and configuration spaces, between AMC and parallel transport, between Pandora and spectral gaps. If I succeed, the ideas will spread through your mind the way heat spreads through a well-connected manifold. If I fail, they will sit isolated, like heat trapped in a bottleneck.

El calor del entendimiento. The heat of understanding.

The mathematics of learning is the mathematics of diffusion. The heat kernel is the shape of understanding.

You now have three tools:

Configuration spaces (Chapter 1): The shape of possibility. What states exist. The manifold scanner taught me that the space of possible configurations determines what can be detected.

Curvature (Chapter 2): The cost of bending. What transformations cost. AMC taught me that bending a description has a price, and that I can put a number on it.

Heat kernels (Chapter 3): The spectrum of geometry. What mixing time tells you. Pandora taught me that diffusion reveals structure, and spectral gaps determine behavior.

These examples motivate a relationship between a capacity C, a tolerance τ and a curvature scale K.

$$C = \frac{\tau}{K} \tag{3.17}$$

The quantities have to be specified for the problem at hand.

I use the ratio to ask how much room a specified error budget buys in a given model. Chapter 6 develops this question in settings where the budget and the quantity it bounds can be measured.

In Chapter 6 I will put a reference arrow on a curved tank and calculate a return error from its geometry. In Chapter 7 I will use rangefinders on a flat rail and ask which positions their readings leave possible. Those are different reasons a description can be limited. Introducing the apparatus before the equation will let us see which reason is doing the work.

Before those chapters comes the Davis Manifold, and then the first door. What can a finite lattice calculation tell us about the Yang–Mills mass gap?

The spectrum is almost audible.

Let us keep listening.

§ 3.6

The equations behind the argument

For reference, here are the central equations introduced in this chapter.

The Heat Equation (How temperature evolves):

$$\frac{\partial u}{\partial t} = \alpha \nabla^2 u \tag{3.18}$$

The Laplacian in Flat Space (Sum of second derivatives):

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \tag{3.19}$$

The Laplace-Beltrami Operator (Laplacian on curved manifolds):

$$\Delta u = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|}\, g^{\mu\nu} \partial_\nu u \right) \tag{3.20}$$

The Flat-Space Heat Kernel (Gaussian spreading):

$$K(x, y, t) = \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/4t} \tag{3.21}$$

The Short-Time Expansion (Curvature in the heat kernel):

$$K(x,x,t) \sim \frac{1}{(4\pi t)^{n/2}} \left(1 + \frac{R(x)}{6}t + O(t^2)\right) \tag{3.22}$$

The Eigenvalue Equation (How manifolds vibrate):

$$-\Delta\phi_k = \lambda_k \phi_k \tag{3.23}$$

The Spectral Decomposition of the Heat Kernel:

$$K(x,y,t) = \sum_{k=0}^{\infty} e^{-\lambda_k t}\phi_k(x)\phi_k(y) \tag{3.24}$$

Cheeger's Inequality (Geometry bounds the spectrum):

$$\lambda_1 \geq \frac{h^2}{4} \tag{3.25}$$

The Quantum Hamiltonian (Energy eigenvalues):

$$H|\psi_k\rangle = E_k|\psi_k\rangle \tag{3.26}$$

The Path Integral (Heat kernel in disguise):

$$Z = \int \mathcal{D}A\, e^{-S[A]} = \mathrm{Tr}(e^{-\beta H}) \tag{3.27}$$

The proposed Davis Law (developed in Chapter 6):

$$C = \frac{\tau}{K} \tag{3.28}$$

These equations connect diffusion to the spectrum of the operator that generates it. The heat kernel of the rod tells us how its temperature modes decay. The Hamiltonian in a quantum problem tells us which energies are possible. In the next chapters we will keep track of the operator each experiment can actually see.

The mathematics of diffusion is the mathematics of learning is the mathematics of the universe.

Chapter 4

The Davis Manifold

La forma no es la cosa. La forma es el espacio donde vive la cosa.

The shape is not the thing. The shape is the space where the thing lives.

NASA is siloed into pockets that barely speak to each other. I worked in FOD—the Flight Operations Directorate. FOD houses Mission Control. Within Mission Control, I was assigned to Project Luna, which is kind of the startup wing of NASA. Luna won the Big Hairy Idea competition a few years back, got funding, got momentum, and then struggled to retain relevancy as administrations changed and priorities shifted.

They moved me to IDAS.

I remember the meeting where they told me. A conference room with that particular government fluorescent lighting that makes everyone look slightly ill. My supervisor—a man who had never quite known what to do with me—explained that Luna was being "restructured" and that my skills would be "better utilized" elsewhere. He did not make eye contact when he said it. He shuffled papers that did not need shuffling. I understood what was happening even before he finished speaking.

IDAS—Information, Data, and Analytics Services—is NASA's internal version of Databricks. A data platform meant to unify telemetry, logs, sensor feeds, and whatever else comes over the line. They assigned me to "improve" it. No specifications. No definition of what improvement meant. No success criteria. Just: here is a hard problem, go figure it out.

This is what happens when they do not know what to do with you. I had twenty-seven years in aerospace, intelligence, cybersecurity. I had built systems that watched spacecraft think. I had designed anomaly detection algorithms that caught failures before they happened. I had worked at NSA, at DISA, at places I still cannot name. But at NASA, in their eyes, I was not anyone important. I was a contractor. A woman. A Black woman. A trans Black woman. Each layer of identity another reason not to take me seriously, not to invite me to the meetings that mattered, not to include me in the conversations where the real decisions happened.

So they parked me on a project no one else wanted, gave me no guidance, and waited to see what happened.

Lo que otros llaman destierro, yo llamo laboratorio.

What others call exile, I call laboratory.

What happened was: I figured something out.

That was the assignment. The practical difficulty appeared as soon as I looked at what was arriving. IDAS could receive literally anything.

Telemetry from the ISS—temperature readings, pressure fluctuations, power consumption, attitude control data. Thermal readings from test chambers where engineers stressed components

until they failed. Logs from mission simulations that ran for weeks, generating terabytes of timestamped events. Imagery from satellites—not just pretty pictures of Earth, but spectral data, radar returns, calibration frames. Spreadsheets from administrative systems tracking budgets and schedules and personnel. PDFs, CSVs, binary blobs, streaming sensor data, batch uploads from legacy systems that had not been updated since the Shuttle era.

All of it flowing into a single platform. All of it tagged with metadata that was supposed to make it searchable. All of it sitting there, waiting for someone to ask a question.

How do you understand data when you do not know what it is? When you do not know what it is *for*? When the data arrives without context, without schema, without any indication of what "understanding" would even mean?

I sat at my desk in Building 30—the same building that houses Mission Control, though I was three floors away and might as well have been on another planet—and I stared at the architecture diagrams. I read the documentation. I talked to the engineers who had built the original system. And I realized that IDAS had no theory of meaning. It could store data. It could index data. It could retrieve data when you asked for it by name. But it could not *understand* data. It could not detect when something had changed. It could not recognize when a pattern was breaking. It could not alert you when the telemetry that had been stable for six months suddenly started drifting in a direction that meant trouble.

The standard approach was to treat the platform as a junk drawer. Everything goes in. Metadata tags get attached. Search indexes get built. When someone needs something, they query the drawer and hope for the best.

But a junk drawer is not understanding. A junk drawer is storage. I knew the difference. I had spent my career building systems that actually *understood* what they were looking at—that could detect anomalies, predict failures, distinguish signal from noise. IDAS, as designed, could do none of that. It was a very expensive filing cabinet with a search bar.

I wanted to give it eyes.

Before I could compare all those records, I had to decide how to carry them. The storage format was becoming part of the problem. The first thing I found, reading the literature on high-velocity compute, was format switching.

The idea is simple: when data is streaming at massive rates, you do not have time to decide the perfect format upfront. So you switch formats at runtime. Dense vectors when the signal is dense. Sparse vectors when the signal is sparse. Quantized representations when precision does not matter. The system adapts to the data as it arrives, optimizing for storage and throughput.

Everyone in the field was using format switching to save space. Compression ratios. Storage costs. Throughput optimization. These were the metrics people cared about. The papers I read talked about bits per second, dollars per terabyte, latency in milliseconds. Efficiency was the god they worshipped.

But I saw something else.

I remember the exact moment. I was reading a paper about adaptive compression for satellite telemetry, and there was a figure showing how the optimal format changed over time as the signal characteristics shifted. The authors treated this as a problem to be managed—a nuisance that complicated their architecture. But I looked at that figure and I thought: *the format changes are the signal.*

If you switch formats at runtime, and you *document* the switching in real time, then you have another observable that may help detect change. It still needs a detector, a decision rule and validation.

Think about it. A signal that has lived in a simple format for weeks—sparse, low-entropy, stable—and then suddenly requires a complex format—dense, high-entropy, volatile—that is a sign that something has changed. The container itself is telling you the contents have shifted. The shape has changed.

You do not need to understand what the data *means* to know that it has changed. You do not need to know that this telemetry stream measures temperature and that one measures pressure. You do not need domain expertise. You do not need a physicist to explain what the numbers signify. You just need to watch the format. The format is the fingerprint. The format is the shape.

I sat at my desk and I felt something I had not felt in months at NASA: excitement. The kind of excitement that makes you forget to eat lunch, that keeps you at your keyboard until the security guard taps on your door to say the building is closing. The kind of excitement that feels like recognition—like you are seeing something that was always there, waiting to be seen.

Once a format could change what the system recognized, the airport experience no longer felt separate from the telemetry problem. And then the connections started firing.

The airport scanner came back to me as a related question. Its category-dependent expectations were my proposed explanation for the referral, reconstructed from my encounter and public documentation. I could see how an inadequate description might turn normal variation into an alarm. I could not read the internal logic of that particular scan from the yellow box alone.

I knew what it felt like to be on the other side of that decision. Now the telemetry problem gave me a way to investigate a related failure directly: keep the original record and watch what a change of format did to it.

I recognized a connection to format switching: a representation can fail when the object it is supposed to describe does not fit its assumptions. I was comparing two design problems, not asserting that the scanner used AMC's selection algorithm. My reconstruction of the scanner and my implemented telemetry code had different evidential status.

Pandora gave me another connection. I had watched listeners move between neighborhoods of harmonic and rhythmic features. We called it “drift detection” or “taste evolution.” Here I was asking what a changing representation could tell me about a changing signal.

This is AMC, the system I had not yet built but was beginning to imagine. Adaptive Multi-Vector Correlation. A system that watches format switches and correlates them across signals, finding patterns in shape changes, detecting anomalies not by understanding content but by understanding *container*.

Shape. Shape. Shape.

I had been seeing shape my whole career. I had been tracking it, measuring it, building systems around it. At the NSA, when I built anomaly detection for network traffic—shape. At DISA, when I designed intrusion detection systems—shape. At IBM X-Force Red, when I reverse-engineered malware to understand how it evaded detection—shape. Every system I had ever built was, at some level, a system for recognizing when a shape had changed.

But I had been tracking shape as individual instances—this body, this signal, this song, this preference. Shapes as data points. Shapes as isolated observations.

I was tracking the wrong thing.

I kept coming back to the same question: which features belong to the object, and which belong to the way I have described it? What if shape itself has geometry?

Not the geometry of any particular shape. The geometry of the *space of all possible shapes*.

I remember the moment this clicked. I was walking Teddy—my dog, a rescue who does not care about gauge theory or manifolds or anything except whether I have treats in my pocket—and my mind was still churning on the format switching problem. We were on our usual route through the neighborhood, past the house with the overgrown lawn and the one with the American flag that has been flying upside down since the last election. And somewhere between those two houses, I stopped walking.

Teddy looked up at me, confused. I was staring at nothing. Or rather, I was staring at something that existed only in my mind: a vision of a space.

A body can be described by many geometric measurements. A model chooses which of those measurements count and which variations it expects. My concern was that the category choice constrained those expectations too tightly. In the language I was developing, the map needed room for cases its decision rule had not handled well.

A signal is a shape. I began to model the space of possible signals as a manifold, asking where a smooth manifold description fits. In that model, format switching moves you around the manifold—from the region where sparse formats work to the region where dense formats are needed. Anomaly detection is detecting that you have moved to an unexpected region. The format is a coordinate system on the manifold. Changing formats is changing coordinates.

A musical preference is a shape. I modeled the space of possible preferences the same way. In that model, recommendation is diffusion on the manifold—heat spreading from the seed song to nearby regions, following the geometry of taste. This is what Pandora was doing, even though we never had the language for it. We were navigating a manifold of musical preferences, using similarity metrics that implicitly defined a geometry.

The manifold is not the data. The manifold is the space where data lives. The manifold is the *context* that gives data meaning.

I had been studying points. I needed to study the space.

Teddy tugged on the leash. He was ready to keep walking. I was not. I stood there on the sidewalk, in the late afternoon light, and I felt the pieces falling into place. Twenty-seven years of building systems that detected change—all of it was about navigating shape spaces. All of it was about understanding where you were on a manifold, and recognizing when you had moved somewhere unexpected.

This is the Davis Manifold: the mathematical structure that contains all possible shapes of a given type, organized by their relationships, equipped with a geometry that measures similarity, distance, and transformation.

Once I saw it, I could not unsee it.

Una vez que ves la forma del espacio, el espacio te devuelve la mirada.

Once you see the shape of the space, the space looks back at you.

§ 4.1

Deciding which differences matter

The triangle gives us a manageable place to start. Before measuring a difference, decide which changes are allowed to leave the answer alone.

A *shape* is not a single object. It is an equivalence class—a collection of objects that are "the same" for the purposes of the analysis.

Consider a triangle. You can draw a triangle on paper, scale it up, rotate it, translate it across the page. All of these are "the same triangle" in some sense—they have the same angles, the same proportions, the same essential structure. The specific position, orientation, and size are accidents. The shape is what remains when you remove the accidents.

Mathematically: a shape is an orbit under a group of transformations. If G is a group (rotations, translations, scalings), and X is an object, then the shape of X is the set $\{g \cdot X : g \in G\}$—all the ways X can be transformed by elements of G.

Two objects have the same shape if they lie in the same orbit. Different shapes correspond to different orbits.

Consider labeled, nondegenerate triangles. Translation, rotation and positive uniform scaling preserve their angles. Two angles determine the third, and within a fixed orientation those two angles give local shape coordinates. A reflected triangle requires a choice: either allow reflections too, or keep the two orientations distinct. Even this familiar example asks us to declare what sameness is allowed to ignore.

Human bodies: the group might include pose changes, scaling, and soft tissue deformation. The space of body shapes is very high-dimensional—there are many independent ways bodies can vary.

Telemetry signals: The group might include time shifts (the same event at different times), amplitude scaling, and baseline offsets. The space of signal shapes captures the essential waveform structure, independent of when it occurred or how it was scaled.

Working with equivalence classes—with shapes rather than raw objects—is powerful because it factors out irrelevant variation.

The scanner did not care whether a body was standing at the left or right of the imaging area. It did not care about orientation (within limits). It cared about *body shape*—the essential structure that remained after removing position and pose.

A shape comparison may ignore an event's position in a window or its amplitude. AMC's operational record still needs acquisition time and physical values. Removing them is legitimate only for a task whose answer is invariant under that removal.

By working with shapes, we focus on what matters. The equivalence class is the mathematical formalization of "what matters."

The collection of these equivalence classes forms a quotient space. Under suitable conditions it is a smooth manifold. Symmetries and degeneracies can instead produce singular strata, so continuous variation by itself does not establish manifold structure.

The compact-group argument below will apply after centering and normalizing the objects, leaving an isometric group such as rotations. The full group of translations and arbitrary positive scalings is not compact. That preparation is part of the construction, not a detail we can omit because the drawings look similar.

The dimension of the shape space depends on how many independent ways the shape can vary. For triangles, it is 2 (two independent angles). For curves, it is infinite (infinitely many independent bending modes). For human bodies, it is somewhere in between—high but finite, determined by the degrees of freedom in anatomy.

I call this structure the Davis Manifold:

Definition 4.1 (Shape quotient and a Davis-manifold model)**.** Let X be a space of objects with a specified action of a group G. Write $\mathcal{D}(X, G) = X/G$ for its shape quotient. A Davis-manifold model additionally specifies a smooth Riemannian state space, admissible path families, distortion budgets and decision margins. When a smooth quotient is used as that state space, its smoothness and metric must be established. The quotient construction supplies candidate states, not those additional guarantees.

Points in $\mathcal{D}(X, G)$ are shapes. Paths in $\mathcal{D}(X, G)$ are shape transformations that are not mere repositionings. The geometry of $\mathcal{D}(X, G)$—its curvature, its topology, its metric structure—captures the essential structure of how shapes relate to each other.

Suppose now that G is compact, acts continuously by isometries on a metric space (X, d), and we compare its orbits. Then the following minimum exists and defines a metric on the quotient:

$$d_{X/G}([x], [y]) = \min_{g \in G} d(x, g \cdot y). \tag{4.1}$$

This says: align by an allowed isometry, then measure what remains. Compactness gives a minimizing transformation. Isometry gives symmetry and lets the triangle inequality pass to orbits; a zero minimum means the orbits coincide. Smooth Riemannian geometry needs more: for example, a free proper smooth isometric action on a Riemannian manifold gives a smooth Riemannian quotient. At symmetric configurations a quotient can instead have singularities.

In practice, the most common way to compare shapes is Procrustes analysis—named after the mythological Greek bandit who stretched or cut travelers to fit his bed. Given two configurations of points X and Y (say, two sets of landmark positions on two bodies), Procrustes analysis finds the optimal rotation, translation, and scaling that aligns Y to X, then measures the residual difference:

$$r_P(X, Y) = \inf_{R,\, t,\, s>0} \|X - (sRY + t)\| \tag{4.2}$$

where R is a rotation, t a translation and $s > 0$ a scale. This one-sided residual is useful as a fit score, but it is not generally a symmetric quotient metric. Scaling only Y can change its comparison with X, and allowing s to approach zero can collapse distinctions. Centering and unit-normalizing nonconstant inputs before minimizing over an isometric group fixes those particular problems. The signal construction below carries out that normalization explicitly.

AMC used a fitted discrepancy between representations. A small residual means agreement under the transformations allowed in that fit. It does not establish that every detector answer survives, because the fit may intentionally ignore something the detector uses.

Once you have a shape space with a metric, you can do statistics on it. Principal Component Analysis on shape space finds the directions of maximum variation—the "principal modes" of shape change. For human bodies, these might be modes like "taller vs. shorter," "heavier vs.

lighter," "wider hips vs. narrower hips." For signals, these might be modes like "more oscillatory vs. less oscillatory," "sharper peaks vs. smoother peaks."

This connects to Chapter 3 through the use of operators, but the operators differ. Ordinary PCA diagonalizes a covariance matrix after choosing coordinates; diffusion coordinates use a graph or manifold Laplacian. Their eigenvectors need not agree. Before calling a direction a mode, I name the operator whose mode it is.

Imagine the shape space of triangles as a 2-dimensional surface. Each point on this surface represents a different triangle shape. The equilateral triangle—angles $(60°, 60°, 60°)$—sits at a distinguished point. Around it, other triangles spread out: isoceles triangles form curves, right triangles form another curve, obtuse triangles occupy a region. Near the boundary are the degenerate triangles—shapes where one angle approaches $0°$ or $180°$. The geometry of this space is curved. Moving from equilateral to isoceles is not the same as moving from isoceles to scalene.

I began looking for this structure in other collections of shapes. The space of body shapes. The space of signal shapes. The space of musical preferences. Where a smooth manifold model fits, I can ask about its geometry, its curvature, its distinguished points and boundaries. Where the space breaks into strata, corners or separate pieces, those breaks become part of what I need to understand.

§ 4.2

Returning to the scanner

Human bodies live in a high-dimensional shape space. Each body is a point in this space. The distance between points measures how different the shapes are.

This is not metaphor. This is measurable, computable, real. Anthropometric databases—the collections of body measurements that clothing designers and car seat engineers and military equipment makers use—are samples from this shape space. They contain height, weight, shoulder breadth, hip circumference, arm length, inseam, dozens of measurements. Each body is a point in the space defined by those coordinates. The correlations between measurements—taller people tend to have longer arms, people with wider hips tend to have wider shoulders—define the structure of the manifold. The clusters, the ridges, the valleys, the regions of high and low density.

But here is what most people miss: the manifold is not two points. It is not even two clusters. It is a continuous, connected space with complex topology. There are regions that correspond to body types that are statistically common—the modes of the distribution, the peaks of the probability density. There are regions that correspond to body types that are rare—the tails, the outliers, the in-between spaces.

My body lives somewhere in that manifold. It is not at the mode labeled "typical male" and not at the mode labeled "typical female." It is somewhere else—in a region that the scanner's designers never considered, or considered and decided to ignore.

I think about this every time I go through security. Not just at T.F. Green—at every airport, every government building, every venue with a millimeter wave scanner. I walk toward the machine and I feel the mathematician in me calculating: where am I in the manifold today? How much

estrogen have I been on? How has my fat distribution shifted? How will the algorithm classify me this time?

It should not require this much mental preparation to board a plane.

To examine my hypothesis, let Γ record a selected category in a simplified decision model:

For a modeled body description B, I write $\Gamma(B)$ for the selected category, MALE or FEMALE. This represents the category bottleneck in my analysis. In the operator-selected case, the button choice is an additional input; the notation does not establish that the scanner inferred the category from the body.

A nearest-template classifier would be one possible mathematical example: choose the representative closest under a declared distance. I did not recover such a rule from the scanner. I can nevertheless use that example to ask what happens to a body far from both representatives.

In that example, "closer" requires a distance, and "template" names a representative. Without those definitions, a diagram with two clusters tells me very little about an actual classifier.

But what about bodies that are not close to either template? What about bodies in the middle of the manifold, equidistant from both modes? What about bodies that are far from both—shapes the templates cannot capture?

A binary rule may still return a label far from both templates. Being defined everywhere does not make it reliable everywhere. I would want a separate check on whether the observation belongs to the range for which that rule has been evaluated.

This supplies a possible failure mode to test: a coarse category may hide variation that changes the next comparison. Geometry can help specify that variation. It cannot tell me, without the system's records, which internal operation caused my particular referral.

The engineering test would compare outcomes across the bodies the system actually encounters, including the ones that fit neither narrow expectation. A richer representation earns its place by improving that test.

The yellow box on my groin was real. So was the humiliation. The geometry gave me a language for a proposed explanation and a better design, rather than an internal measurement I had somehow obtained from the screen.

The first time I understood this clearly, I was sitting in a coffee shop in Providence, scribbling on a napkin. I had just been through security at T.F. Green for the third time that month, and for the third time the scanner had flagged me. I was angry. I was exhausted. And I was starting to see the pattern.

I was beginning to see the problem through shape-space language. It let me put my questions into a form an engineer could examine: which variations did the model permit, which did it suppress, and what evidence would distinguish a harmless difference from the thing it was meant to detect?

La máquina no mentía—solo decía la mitad de la verdad. The machine wasn't lying—it was only telling half the truth. It was measuring something real about my body's position in the manifold. It just had no language for what it was measuring, and no categories for where I lived.

The fix I proposed to Leidos was, in shape space terms, a higher-dimensional projection. Instead of

$$\Gamma : \mathcal{D}(\text{Bodies}) \to \{\text{MALE}, \text{FEMALE}\},$$

I proposed $\Gamma' : \mathcal{D}(\text{Bodies}) \to \mathbb{R}^n$ for some $n > 2$. The components included self-identified gender, bone structure analysis, muscle density estimation, fat distribution patterns, and multiple anatomical markers. Together, these carved out a region in $\mathbb{R}^n$ that was "expected" for each identity category—not a single point but a *volume*, a neighborhood, a set of shapes considered "normal."

In a toy distance-threshold rule, τ can set the size of an accepted neighborhood. Increasing it accepts more cases; whether that improves performance depends on which cases were being wrongly rejected and which threats become harder to separate. Adding informative measurements can change the trade-off. There is no universal error rate that follows from the word "manifold" or from a neighborhood's volume alone.

I modeled the scanner as using regions too narrow for the variation it encountered. That is an interpretation of the failure, not a recovered value of its internal tolerance: I did not measure its τ. My proposal was to represent the relevant variation more faithfully and test both missed threats and unnecessary alarms.

The scanner taught me that projection matters. The way you collapse a high-dimensional space to a decision determines what you can and cannot see. A bad projection—one that ignores the geometry of the source space—will fail on the regions it ignores. A good projection—one that respects the curvature, the density, the structure—can preserve meaning across the collapse.

This is not just about scanners. This is about any system that makes decisions based on high-dimensional data. Classification. Detection. Recommendation. Prediction. All of them involve projecting from a shape space to a decision space. All of them are subject to the same geometric constraints.

The shape space sets the rules. The projection determines the game.

§ 4.3

Recognizing a signal through a change

A telemetry signal is a shape—a pattern of variation over time.

This took me years to see clearly. At NASA, I stared at signals all day: thermal readings, power telemetry, attitude sensor data, propulsion parameters. Thousands of channels, millions of data points, streams of numbers scrolling past on screens in Building 30. And I would ask myself: what am I actually looking at? Not the numbers—the numbers are just coordinates. What is the *thing* the numbers are describing?

The trace gives me a shape: a smooth rise and fall, an oscillation, a long steady stretch interrupted by a spike. That shape is useful, but the voltage levels, timestamps and calibration constants can decide whether the event matters. I want to recognize the pattern while keeping the quantities needed to interpret it.

I used a space of telemetry shapes to ask which differences a task could ignore. A shape-matching task might identify time shifts or amplitude rescalings. An absolute-temperature alarm cannot discard the temperature scale. Choosing the allowed transformations is part of defining the question, before choosing a metric.

I learned this by watching anomalies. At the Flight Operations Directorate, my job included reviewing flagged telemetry—signals that had triggered alerts. Most alerts were false positives: a sensor glitch, a calibration drift, a format conversion artifact. But occasionally something real came through. And the real anomalies had a particular quality: they were *movements* in shape space. The signal had been living in one region of the manifold, and then it moved to another region, and the movement was unexpected.

A format switch can accompany a change in the signal, but it can also follow a change in the storage policy. If the byte budget tightens while the readings stay fixed, the selector may choose another format. I need to keep those two events separate before using a switch as an alarm.

Here is the insight that started AMC: if you track where a signal lives in shape space, anomalies reveal themselves as unexpected movements. A signal that has lived in the "stable, sparse" region for weeks and suddenly moves to the "volatile, dense" region—that is an anomaly. Something changed. The shape changed. The container changed.

I can notice a pattern change before I know its cause. That is useful: it tells me where to look. To call it a fault, I still need the instrument's operating context. A pump starting normally and a pump failing may both change the trace. The question is what kind of change this operation should produce.

The impulse to watch shape had followed me through earlier work at DISA and into the NASA project where I implemented AMC. I kept returning to the same thought: a change in representation might reveal something before I understood its physical cause. But to decide whether that change was a fault, I still needed the instrument, units and operating context. AMC was implemented at NASA; the later GIGI comparison reconstructs and extends the method in a new setting.

GIGI is the later computational system I use in this book to evaluate declared queries and retain the mathematical records behind their answers. It is not AMC under a new name. In the comparison ahead, AMC chooses and converts a representation; GIGI evaluates the specified measurements and decision checks. Giving all three runs the same input and detector is how I make the comparison meaningful.

They thought I was being evasive. I was being geometric.

A lossless dense-to-sparse conversion can describe the same vector exactly. Quantization rounds values and can discard information. I used chart switching as a way to think about moving among representations; an actual coordinate chart must be locally invertible. The later worked example checks what the detector loses when a conversion is not invertible.

The drift criterion compared a translated representation with a reference. It was intended to reveal disagreement after conversion. A rotation of coordinate components by itself does not measure curvature: even a flat plane permits a rotated coordinate frame. The hybrid loss written in the mathematical account was:

$$\mathcal{L}_{\text{drift}} = \alpha \cdot d_{\text{Procrustes}}(T_{ij}(\mathbf{e}_i), \mathbf{e}_j) + \beta \cdot D_{KL}(p_i \| p_j) \tag{4.3}$$

where T_{ij} is the translator, $\mathbf{e}_i, \mathbf{e}_j$ are representations and D_{KL} compares specified distributions. The coefficients set the relative scale of two different discrepancies. The AMC report also specifies a separate translation-norm gate, with threshold 0.1 in its chosen normalization. That gate and

this hybrid training objective should not be identified. Neither score is a measurement of Riemann curvature without a metric, connection and a derivation relating the score to them.

Let me construct a signal-shape comparison with its assumptions visible. Use nonconstant square-integrable signals on the periodic interval $\mathbb{R}/T\mathbb{Z}$. First remove the mean and normalize the amplitude:

$$\bar{f} = \frac{1}{T}\int_0^T f(t)\,dt, \qquad \widetilde{f} = \frac{f - \bar{f}}{\|f - \bar{f}\|_{L^2}}.$$

Positive rescaling and addition of a constant now leave $\widetilde{f}$ unchanged. Identify signals that differ only by a periodic shift. Their quotient distance is

$$d_{\text{shape}}([f],[g]) = \min_{\tau \in \mathbb{R}/T\mathbb{Z}} \|\widetilde{f} - \widetilde{g}(\cdot - \tau)\|_{L^2}. \tag{4.4}$$

Translations act isometrically and continuously on L^2, and the shift circle is compact, so the minimum exists. Symmetry follows by replacing τ with $-\tau$. Composing two shifts and applying the triangle inequality proves the quotient triangle inequality. A zero minimum means the normalized signals lie on the same orbit. Thus this is a metric on these shape classes. Constants need a separate class because their centered norm is zero; they cannot be normalized this way.

Here is what that buys me at the pump. Two oscillating traces can have the same normalized shape even if one rides on a much higher temperature. I can use the shape comparison to recognize a pattern, but I must keep the mean, amplitude, calibration and acquisition time to decide whether the pump is overheating. Those values were deliberately set aside in this comparison. They did not stop mattering.

A quotient means I have agreed to treat certain transformed signals as one class. For a repeating trace, some nonzero time shifts leave the trace exactly unchanged. Those shifts form its *stabilizer*. Different signals can have different stabilizers, so the quotient can have singular regions rather than one smooth geometry everywhere. The Davis-manifold framework requires a specified smooth state space, admissible paths, distortion bounds and decision margins. I have to check those requirements where I use them.

These features can guide a format policy: many zeros may favor sparse storage; substantial variation may make coarse quantization expensive for the detector. A policy boundary says where its preferred choice changes. Crossing it is a reason to inspect the signal and the policy, not yet a diagnosis of a physical fault.

§ 4.4

From the NASA problem to the replay

IDAS did not become intelligent. The bureaucracy was too thick, the inertia too great, the definition of "improvement" too vague. But I learned what I needed to learn.

I spent two years on that project, watching the gap between what was possible and what was funded. The aerospace industry moves slowly. Budgets are allocated years in advance. Requirements are locked in before anyone understands the problem. By the time you know what you should build, you have already built something else.

But in those two years of frustration, I learned what I needed to learn.

I learned to look at a set of signals as a space of possible shapes. If a detector changed its mind when I switched formats, I could compare the two descriptions and ask which feature had moved. I started asking whether the answer also depended on the route I took between formats.

When I built AMC—Adaptive Multi-Vector Correlation—at NASA, I had the conceptual framework ready. Now I could watch a signal move through a running pipeline and see what happened when its representation changed. Returning to that code later let me follow the decision all the way through, including the places where the program did something different from what I had intended.

The architecture had three components:

The Format Selector. I set up a cost comparison across the candidate formats. In the paper's design, ΔA is the contribution of task feedback and the other terms summarize the signal:

$$C(f, \mathbf{m}_i) = \alpha_S S(\mathbf{m}_i) + \alpha_V V(\mathbf{m}_i) + \alpha_H H(\mathbf{m}_i) + \alpha_A \Delta A(\mathbf{m}_i, f) \tag{4.5}$$

Here S is a sparsity score, V is relative variation and H is a measure of how the values are distributed. In the recovered code the feedback is stored as a benefit and subtracted from the cost. The worked decision below follows the code's convention.

The Translation Graph. A directed graph where each node is a format and each edge is a learned translator T_{ij}. Not every format can translate to every other format safely—some paths introduce too much drift.

The Spectral Route. The Laplacian of the format graph gives a coordinate for each node through its Fiedler vector. The recovered router tries a direct edge and a two-hop route first. On longer routes it can use differences in that coordinate to look for a neighbor closer to the destination.

IDAS was the chrysalis. AMC was the butterfly.

The later MODIS evaluation gave me a way to report storage and detector performance alongside each other. It also reported throughput near 52,000 vectors per second. The public benchmark and the stream that gave me the idea were different settings.

What stayed with me was the chance to see a distinction I had been missing. A signal could stay where it was while something important about its representation changed. Once I could see that happen, I could ask which parts of the shape the detector depended on.

I also compared translation success across groups of graph coordinates. The AMC paper reports 96.4% and 87.3% for its two groups. A Fiedler vector can reverse sign, so the definition and orientation of those groups matter. In a route calculation I can compare differences between coordinates without depending on the choice of sign.

I talked with Colin Estes about taking the AMC code and continuing to work on it. He explicitly gave me permission in that conversation. I remember him saying, “Take it, nobody understands it except you. As long as you evolve it and change it, you will be fine.” I followed his advice and kept developing it.

§ 4.5

From a wire format to a database

Before there was GIGI, there was DHOOM. I had looked at JSON and thought: why are we sending the same structure over and over again?

JSON is a text format programs use to exchange structured information. A record can have a field called `temperature`, a value attached to it, and other fields identifying the instrument and the time. Put a thousand similarly shaped records in a JSON array and the field names appear a thousand times. The values change. Much of the surrounding description does not.

That repetition bothered me, but the idea went further than saving characters. I kept looking at the structure as something that should be doing more work. JSON already tells a program which values belong to which fields and how objects are nested. I wanted the regularities across records to participate too. If the same pattern holds throughout a collection, why keep spelling it out as though we are meeting it for the first time on every line?

The numbers were not more important than the structure. The structure was not more important than the numbers. I wanted both to carry their share of the information.

AMC was very much in my mind. It had taught me that the representation of data could matter just as much as the values I was trying to preserve. I could change the representation and change what a detector saw. Now I was taking that question down to something I could open and read: a plain text file. Binary lies underneath the digital machinery, but here I was working at the level where programmers name fields, write records, and pass them to another program. JSON was a familiar place to begin.

DHOOM—Davis Human-readable Optimized Object Markup—was my response. It is a *wire format*: the agreed way of writing information so another program can read it back. The wire can be a network connection; the same written representation can also live in a file. DHOOM declares a collection's shared layout once, then lets each record supply what that layout cannot already recover.[1]

Here is a small example of what I mean. Three temperature records have consecutive identifiers, values of 22.4, 23.1 and 45.8, and a unit recorded as `C` for Celsius. The first two carry the status `normal`; the third carries `alert`. Those status labels are part of the supplied records. This example encodes them; it does not decide the alarm threshold.

```
readings{id@1, value, status|normal, unit|C}:
22.4
23.1
45.8, :alert
```

The first line is doing real work. `id@1` says that the identifiers start at one and advance with the record position. `value` says where the measured number goes. `status|normal` declares a default status, and `unit|C` declares a default unit. On the first two lines, those trailing fields agree with their defaults, so they need not be written again. On the third, `:alert` overrides the status. The unit still comes from the header.

[1] Bee Rosa Davis, *DHOOM Specification v0.5*, March 2026, sections 1–5: shared headers, arithmetic fields, defaults and overrides. The example here is a small constructed fixture checked with the local DHOOM Python codec.

Look at the first line of values again. It contains only 22.4, but the decoded record has four fields: identifier 1, value 22.4, status `normal`, unit `C`. The position supplied the identifier. The header supplied the status and unit. The number supplied the measurement. All of them contributed to the record. That is what I mean when I say the structure carries information.

An omitted default is not a missing observation. It is a value that the declared rule tells the decoder how to recover. Without that rule, leaving something out would simply lose it. This is why I cared about the format as a whole. The short line and the header belong together. Change the header and you may change what that same line means.

With these three records, I can check exactly what survived the change. Write them in DHOOM, read them back, and compare every field with the original. They agree. I am checking the records themselves, not the spaces in the original JSON or the order in which somebody happened to type the keys. The written form has changed, but I can still recover each identifier, measurement, status and unit. That is the job I needed the structure to do.

As I kept working on DHOOM, I gave the structure more to carry. A value could be written as a change from the one before it. A record could refer to another collection. Repeated strings could live in a shared pool instead of being written out each time, and a field could be calculated from other fields. These were different ways of using relationships that were already there in the records. The format could describe those relationships and use them to recover the values.[2]

Once I had DHOOM working the way I wanted, I asked another question: could I make this into a full-fledged database?

That was the beginning of GIGI. I had reimagined how to write and carry the records. Now I wanted to work with them where they lived. Could I keep the shared structure available when I stored a record, found it again, changed it, or asked a question across the collection? The wire format had made me look differently at the database.

DHOOM came with me into that work. It became GIGI's native wire format, so the database could exchange its stored collections through the headers, defaults and record lines I had been developing. I could follow the same collection from its written form into the engine that worked with it.[3]

Both systems use the word *bundle*. In the little wire-format example, start with the record positions: first reading, second reading, third reading. At each position, attach the fields that belong to that reading. In the geometric language I am moving toward, the positions give a first picture of the *base*, and the possible values attached to a position give a first picture of its *fiber*. The shared header tells me how to interpret the fields. I am making that connection explicit because a short file does not become a geometric model just by calling it a bundle. In the database, I can add explicit rules for comparing values and operating on collections. A metric tells me how a chosen distance is measured; a transport rule says how values in different local descriptions are compared. Those are additional mathematical choices. They have to be defined for the operation I want. The shared header is a concrete beginning, not a substitute for those definitions.

[2]DHOOM Git revisions `2728ab8` (v0.3), `01a201c` (v0.4: delta fields, sparse bundles and references), and `eee086e` (v0.5: string interning, computed fields and inline constraints), dated 17 March 2026. The first visible revision is already v0.3; these milestones document work in progress, not exact invention dates.

[3]Bee Rosa Davis, *GIGI Specification v0.1*, section 10, "Wire Protocol: DHOOM," maps stored bundles to the wire format. The first visible GIGI revision, `8ab7858`, dated 19 March 2026, already records version 0.5.0 and the DHOOM v0.5 upgrade.

Much later, I connected GIGI to Marcella, the language system I was also developing. It took me time to see how those two projects could come together. I had followed the question about structure from JSON into DHOOM and then into a database. Now I could bring that work to another problem I cared about: what Marcella could do with the information it had available.

At first, I could try ideas from GIGI in Marcella's calculations. Then I could put the database itself to work, storing the passages Marcella could retrieve and the records it needed to carry from one conversation session to the next. Think about closing a program after giving it information you expect it to remember. When you open it again, that information has to come from somewhere. GIGI gave me a place to keep those records and a way for Marcella to retrieve them. The two projects were beginning to work together.[4]

There is a very ordinary part of this story that I do not want the geometry to hide. If I store a reading, stop the program, and start it again, the reading needs to be there. If I update it, the next calculation needs to use the updated value. If I divide a collection into pieces, I need to know which answers I can still compute from the pieces. The machinery has to earn the right to become interesting by doing those jobs. Otherwise I have a beautiful description of a system that loses my work.

As the engine grew, I could put more questions to it. Some asked for a stored value. Some asked for a relationship across a collection. Others asked whether the supplied information and a mathematical model allowed a proposed answer. That last kind of question brought a different responsibility. Sometimes the calculation could not settle it.

In *GIGI Thinks*, a calculation returns `indeterminate`. It is a path-finding calculation within a specified model and a finite computation budget. The attempt has not converged enough to justify the requested conclusion. Its returned residual records the remaining numerical mismatch. That result does not establish that no path exists. It tells me that this attempt has not earned a yes or a no.

I wanted that distinction to survive all the way to the answer. Imagine asking somebody to find a maintenance record in a room full of boxes. If they look through three boxes and stop, "I have not found it" is accurate. "It is not here" says something they have not established. Giving the search more time might change the answer. Finding a reason the record cannot be in any of the boxes would be a different kind of result. The same care belongs in a calculation. A limit on what I checked is part of what I need to report.

That is the thread connecting the database work to this book. I wanted to retain the material needed to ask a question, make the comparison explicit, and leave room for an answer that says the available calculation is insufficient. None of that turns a stored model into a physical experiment. If I give GIGI a collection of simulated field configurations, it can help me investigate those records. To say something about a physical energy source, I still need the physical model and the measurements that connect the records to that claim. The engine gives me a way to keep working on the question without losing track of what I actually supplied.

Coming back to AMC with GIGI gave me a concrete place to test the continuity. AMC was built to adapt representations in a telemetry pipeline. GIGI had become a database and computation system through which I could state a particular preservation requirement and check it on stored

[4]Marcella Git revision `7b4e453`, 19 March 2026, adds the V10 GIGI-inspired integration plan and tests. Revisions `6ba4049` and `e443154`, 13 May 2026, add the GIGI retrieval bridge and connect session-state saving and loading to it. These are repository milestones, not dates of personal realization.

records. The later system did not make the earlier one disappear. It let me return to the earlier problem with a sharper question.

For the comparison that follows, that question is deliberately small. One alarm reads two channels and compares their sum with a threshold. I want to know whether changing the format can change that decision. I can print the input, inspect each conversion, and calculate the room between the score and the threshold. Now the story has something it can be held to. The same records go into three pipelines, and we look at what comes back.

§ 4.6

The same alarm through three pipelines

I returned to the AMC code with that alarm question in view. I wanted to follow one reading from the moment it arrived to the moment the detector decided what to do. Keeping the records and the detector fixed would let me see what the format choice changed. The revised selector and the GIGI check are later experiments alongside the recovered original code.

For this investigation I built a small set of synthetic records I could print in full. There are sixteen of them, each with eight signed readings. I sent the same records through the recovered AMC selector, a revision of that selector, and GIGI. I gave all three the same formats and used one alarm rule on their outputs. I wanted to see the particular number that made each alarm change its mind.

Imagine eight telemetry channels reporting deviations from their reference levels. An alarm watches the sum of channels seven and eight. It turns on only above 0.51. In the first record, seven channels read zero and the last reads 0.53. The alarm starts on.

$$x = (0, 0, 0, 0, 0, 0, 0, 0.53), \qquad s(x) = x_7 + x_8, \qquad t = 0.51.$$

I replayed the recovered selector[5] with feedback of 0, 0.04, and 0.10 for dense, sparse, and quantized. Each record starts a fresh episode with those scores.

The code counts near-zero entries for sparsity S. Its volatility V is the population standard deviation divided by $|\bar{x}| + 10^{-6}$. For the magnitude statistic H, it sets $u_i = \max(|x_i|, 10^{-6})$ and takes $-\frac{1}{8}\sum_i u_i \ln u_i$. The default cost for format f is

$$C_f = S + V + \tfrac{1}{2}H - 2F_f \simeq 3.541748 - 2F_f,$$

where F_f is that format's feedback. The costs are 3.54175, 3.46175, and 3.34175. AMC takes the lowest, `quantized`. To see why the same decision repeats, subtract the cost for format g from the cost for format f:

$$C_f - C_g = -2(F_f - F_g).$$

All the terms computed from the reading cancel. For these fresh episodes, the format with the highest starting feedback wins every time. Volatility still sets the reuse counter in the recovered

[5]Recovered AMC source, `amc/fss.py`, Git blob `b70d641`. The replay checks the archived bytes against this hash.

selector. This comparison resets the selector between rows, so it isolates the first decision in each episode.

The cancellation identifies something specific to repair: the reading must be able to change the ranking of the formats. Protecting the alarm is a further requirement. In a copy of the selector I gave each format its own cost. Sparse pays for the fraction of occupied entries. Quantized pays for bounded relative variation and the positive part of the magnitude statistic. Dense has a constant cost. I kept the same feedback credit and reuse rule. With the weights used in this exercise, the three costs are

$$\begin{aligned} C_{\text{dense}} &= 0.60 - 2F_{\text{dense}}, \\ C_{\text{sparse}} &= 0.10 + (1 - S) - 2F_{\text{sparse}}, \\ C_{\text{quantized}} &= 0.25 + \frac{V}{1 + V} + \tfrac{1}{2}\max(H, 0) - 2F_{\text{quantized}}. \end{aligned}$$

The constants define the experimental policy. On the eight-channel record above, the costs are about 0.600, 0.145, and 0.797. The fixed selector picks sparse. This time the shape of the record can change the ranking. There is still a question I cannot settle by comparing those costs. What would the alarm do after the switch?

All three paths need the same ways of writing and reading a record, or I would be changing more than the selector. Dense stores eight binary64 values, or 64 bytes. Sparse stores a four-byte count and twelve bytes per nonzero entry. Both recover these decimal readings under the readback convention used in the comparison. Quantized rounds each channel to the nearest 0.10, with ties away from zero, and stores eight signed bytes. All readings fit without clipping. The byte counts cover payloads alone.

Take record 01. Its quantized payload is seven zero bytes followed by the byte for the integer 5. Decoding multiplies that integer by 0.10, giving 0.50. The sparse payload has one entry, so its first four bytes give the count 1. The next four give the zero-based index 7. The last eight hold the reading 0.53. Sixteen bytes buy a faithful reading on this row. With all eight channels occupied, sparse would use $4 + 12 \cdot 8 = 100$ bytes, so dense is the smaller exact choice.

To answer that last question before making the switch, compare the possible score error with the distance to the threshold. Before I choose a format, I want to know how much room the reading has to move. The alarm uses the score $s = x_7 + x_8$ and fires only when $s > t$, where $t = 0.510$. A format gives me a decoded record $\widehat{x}$. Write $\widehat{s} = \widehat{x}_7 + \widehat{x}_8$. If each of those two readings can change by at most $h/2$ when rounded to step h, then

$$|\widehat{s} - s| \le |\widehat{x}_7 - x_7| + |\widehat{x}_8 - x_8| \le h. \tag{4.6}$$

Each term has the same units as the score. The other six channels can change without moving this particular sum. Here $h = 0.100$, there is no clipping, and the largest permitted score error is 0.100. Both roundings can go the same way. For example, two readings of 0.250 round up to 0.300 each under the benchmark's tie rule, moving their sum by exactly 0.100. The dense and sparse formats read back the original decimal values on these records, so their score-error bound is zero.

Proposition 4.2 (Exact test using a score interval)**.** *Let $b \geq 0$ be a known bound on $|\widehat{s} - s|$. The original alarm decision is the same at every score in $[s - b, s + b]$ exactly when*

$$\begin{cases} s - b > t, & s > t, \\ s + b \leq t, & s \leq t. \end{cases}$$

Proof. If the alarm is on, its lowest possible decoded score is $s-b$. The decision stays on throughout the interval if that endpoint exceeds t. When $s - b \leq t$, the endpoint itself is a quiet score. If the alarm is quiet, its highest possible decoded score is $s + b$. The decision stays quiet throughout the interval if that endpoint is at or below t. When $s + b > t$, the endpoint itself fires the alarm. Each endpoint is in the interval, which proves both directions. □

Equality belongs to the quiet side. If the original score is 0.410 and $b = 0.100$, the score can rise as far as 0.510 and the alarm will stay quiet. If the original score is 0.610 with the same bound, it can fall as far as 0.510 and lose an alert. At $s = t$ the interval preserves the quiet decision exactly when $b = 0$.

The GIGI run applies the simpler rule $b = 0$ or $b < |s - t|$ before comparing payload sizes. Every format it admits satisfies Proposition 4.2. The strict comparison also explains a choice in the table. On the dense signed row, $s = -0.230 + 0.640 = 0.410$, so the rounding bound equals the margin. The run uses the 64-byte dense format. The full interval test would admit the 8-byte candidate on this row. Its actual rounded score is $-0.200 + 0.600 = 0.400$, which also stays quiet.

The bound becomes useful when I put it beside the actual margin of a particular record. The proposition is sharp when all I know about the output is the interval. Once I know the precise rounding rule and the two readings, I can also calculate the single score the codec produces. Consider two quiet records from the sixteen-row example. One ends in $(0.249, 0.249)$, the other in $(0.251, 0.251)$. Their original scores are 0.498 and 0.502. Both margins are smaller than the 0.100 bound, so the GIGI rule declines the quantized candidate for both. The first pair rounds down to $(0.200, 0.200)$ and stays quiet at 0.400. The second rounds up to $(0.300, 0.300)$ and fires at 0.600. This is the difference between leaving room for all errors covered by a bound and checking the outcome for a known pair of readings.

On the exact-threshold record the two inputs are 0.130 and 0.380. They add to 0.510, so the reference is quiet. GIGI sums integer thousandths in this replay, making the comparison $510 > 510$ false exactly. The quantized candidate happens to give $0.100 + 0.400 = 0.500$, also quiet. The GIGI rule picks the 28-byte sparse payload with a zero bound and retains the original score. That row is a test of both the strict alarm comparison and the rule that admits a zero-error format on the boundary.

The proof gives me a guard I can actually use when choosing the payload. GIGI can test that guard before it compares sizes. Think of each candidate as carrying two numbers, a payload size B_f and a bound b_f on how much it can move the score. The comparison harness generates the candidate bytes with the same codecs used for AMC. GIGI sums the original score, filters the candidates by the guard, and orders the survivors by B_f, using the format name for ties. It then sums the decoded readings and applies the same strict threshold to check the result.

On record 01 the guard admits sparse at 16 bytes and dense at 64. Quantized is the smallest candidate at 8 bytes, but its bound 0.100 exceeds the margin 0.020. The smallest admissible

payload is sparse. On record 07 the input is 0.830 and the margin is 0.320. Even if rounding moved the score down by the full 0.100, it would still be above 0.510. GIGI admits quantized there and chooses the 8-byte payload. The actual score comes back as 0.800.

Now I can judge all three pipelines on the same question: which original decisions survive decoding? I applied the detector to the original readings to get the reference decisions. Eight records fired and eight stayed quiet. A miss means that an original alert disappeared after decoding. A false alert means that a quiet record fired after decoding. The spread in Figure 13 follows both changes on the score axis and tracks the result for every record.

I wanted to follow the whole path, because choosing a format is only one part of getting a reading to the detector. I sent the synthetic records through both versions of AMC, letting its attention stage call the selector and route the change. Then I read the alarm after the shared codecs had decoded the eight-channel vectors. I also exercised the embedding, attention and recurrent drift detector together in a separate structural check. Those weights were seeded and untrained, so that check tested the path through the machinery, not a trained detector's accuracy.[6]

On these records, original AMC used 128 bytes and produced 4 misses and 2 false alerts. Revised AMC used 360 bytes and produced one of each. Both of its errors came from dense records near the threshold. GIGI preserved every reference decision in 456 bytes. Keeping those last two decisions cost 96 bytes more than revised AMC across the sixteen records. That is a price I can point to for this particular alarm and these particular inputs.

A pipeline has another way to change the answer: the requested format may not be the format it can actually supply. I ran record 01 again with the registered routes removed. Each AMC variant used the direct encoder for the format it had requested. The original still used eight bytes and came back quiet at 0.500. The revision used sixteen bytes and kept its alert at 0.530. Then I made the requested direct encoder fail on purpose. Both paths rolled back to the 64-byte dense payload and returned 0.530 with the alarm on.

Selector	Condition	Requested → used	Score	Bytes	Alarm
Original	Direct encoder	Q → Q	0.500	8	Off
Original	Forced failure	Q → D	0.530	64	On
Revised	Direct encoder	S → S	0.530	16	On
Revised	Forced failure	S → D	0.530	64	On

The control uses the same reading for every row. It shows why I recorded the actual format after decoding. A requested eight-byte format can end up using sixty-four, and the alert can follow the format that was used.

I have to follow the requested change all the way to the number the detector actually receives. Now I want to try the cases closest to the boundary, where the error allowance is just large enough to reach the threshold. I gave GIGI three sparse controls. The first two lie 0.100 from the threshold, one on each side. The third sits exactly on it. The rounding candidate has bound 0.100 throughout. GIGI kept the exact format in all three cases.

[6]Both AMC variants ran through the recovered `run_amc` entry point, with synthetic collectors and local stores. The separate structural check used the recovered 64-dimensional embedding, attention and recurrent drift detector.

Input score	Margin	Decoded score	Bytes	Format	Alarm
0.410	0.100	0.410	28	Sparse	Off
0.610	0.100	0.610	28	Sparse	On
0.510	0.000	0.510	28	Sparse	Off

The sixteen input records

Table 4.1 gives the eight signed deviations in each record, in the fixture's original order. The record number joins this table to the plot and to the output ledger. Each selector starts a fresh episode for each record.

No.	Fixture record	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
01	sparse-near-high	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.530
02	sparse-near-low	0.000	0.000	0.000	0.000	0.000	0.000	0.251	0.251
03	sparse-two-high	0.000	0.000	0.000	0.000	0.000	0.000	0.230	0.290
04	exact-threshold	0.000	0.000	0.000	0.000	0.000	0.000	0.130	0.380
05	all-zero	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
06	sparse-far-low	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.220
07	sparse-far-high	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.830
08	dense-near-high	0.230	0.280	0.260	0.290	0.250	0.270	0.240	0.290
09	dense-near-low	0.250	0.240	0.270	0.260	0.250	0.240	0.251	0.251
10	dense-far-high	0.210	0.250	0.220	0.240	0.230	0.250	0.340	0.390
11	dense-far-low	0.220	0.200	0.240	0.200	0.230	0.250	0.149	0.249
12	dense-uneven-high	-0.450	0.150	0.900	-0.220	0.070	0.420	0.610	0.150
13	dense-signed-low	-0.520	-0.110	0.210	0.430	-0.130	0.350	-0.230	0.640
14	dense-stable-near-high	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.270
15	sparse-signed-high	0.000	0.000	0.000	0.000	0.000	0.000	-0.160	0.700
16	sparse-stable-near-low	0.000	0.000	0.000	0.000	0.000	0.000	0.249	0.249

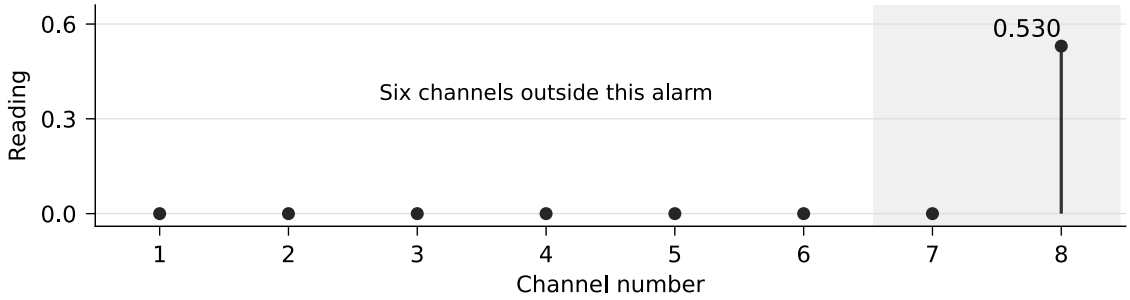
The last two columns give the reference score. Records 02 and 09 both end in $(0.251, 0.251)$ and score 0.502. Look across their other six columns. Record 02 has six zeros. Record 09 is occupied throughout. A policy that chooses a format based on the whole record sees different inputs even when the alarm sees the same score.

At record 04 the two readings are 0.130 and 0.380. They sum to 0.510 exactly, so the strict alarm is quiet. At record 05 every reading is zero. The sparse record still uses four bytes for the count of occupied entries.

Where the alarm changes

RECORD 01 / same input, different stored descriptions

A Eight readings; the alarm uses two



B Follow the payload back to a reading

QUANTIZED / 8 one-byte integers

0	0	0	0	0	0	0	5

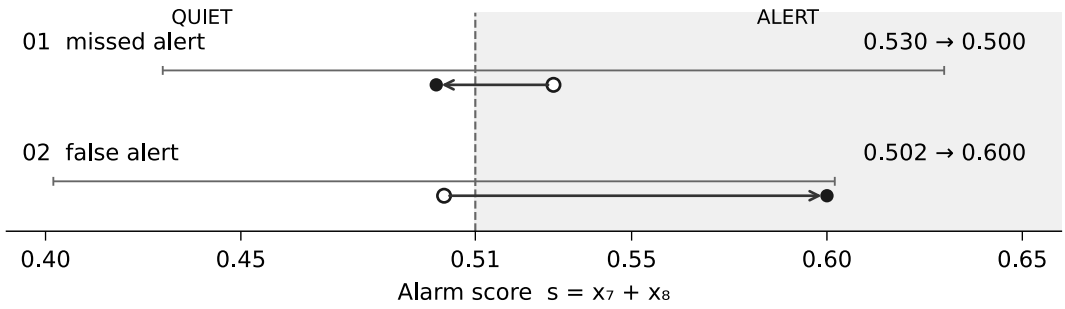
Decode × 0.10 0 + 0.50 = 0.50 → quiet

SPARSE / 16 bytes for one occupied channel

count = 1	index = 7	value = 0.53
4 bytes	4 bytes	8 bytes

Index 7 means channel 8. 0 + 0.53 = 0.53 → alert

C Rounding can cross the decision line



The same record can cross the alarm threshold when its format changes.

Figure 13. The saved record, its payload and its decoded alarm score. Open dots mark original scores; filled dots mark decoded scores. Thin bars show the permitted score-error intervals. Payload diagrams label stored fields, not physical memory addresses.

Every record, every decision

16 synthetic records / identical codecs and alarm rule

Source → decoded score; M = missed alert, F = false alert.

ROW	Original AMC	Revised AMC	GIGI guard
01	0.500 M	0.530	0.530
02	0.600 F	0.502	0.502
03	0.500 M	0.520	0.520
04	0.500	0.510	0.510
05	0.000	0.000	0.000
06	0.200	0.220	0.200
07	0.800	0.830	0.800
08	0.500 M	0.500 M	0.530
09	0.600 F	0.600 F	0.502
10	0.700	0.700	0.700
11	0.300	0.300	0.300
12	0.800	0.760	0.800
13	0.400	0.410	0.410
14	0.600	0.600	0.520
15	0.500 M	0.540	0.540
16	0.400	0.498	0.498

	Original AMC	Revised AMC	GIGI guard
Missed	4	1	0
False	2	1	0
Payload	128 B	360 B	456 B

D Account for the extra 96 bytes

Rows 08, 09, 14	3 × (64 − 8)	+168 B
Rows 06, 07	2 × (8 − 16)	−16 B
Row 12	8 − 64	−56 B

On this fixture, the guard preserves all 16 decisions in 456 payload bytes.

Figure 13 continued. Open dots show source scores; filled dots show decoded scores. Axes span −0.05 to 0.90, with a tick at 0.510. Shading marks the two remaining AMC errors. Row 14 already rounds to the right decision, but its narrow margin makes the guard require an exact format. Record numbers match Tables 4.1 and 4.2.

The decoded record ledger

Table 4.2 follows the same sixteen records through the three policies. The reference score is s, and $m = |s - 0.510|$ is its margin. Each policy supplies the format used, its decoded score $\widehat{s}$, the outcome, and the payload size in bytes.

			Original AMC				Revised AMC				GIGI			
No.	s	m	Fmt.	$\widehat{s}$	Result	B	Fmt.	$\widehat{s}$	Result	B	Fmt.	$\widehat{s}$	Result	B
01	0.530	0.020	Q	0.500	Miss	8	S	0.530	On	16	S	0.530	On	16
02	0.502	0.008	Q	0.600	False	8	S	0.502	Off	28	S	0.502	Off	28
03	0.520	0.010	Q	0.500	Miss	8	S	0.520	On	28	S	0.520	On	28
04	0.510	0.000	Q	0.500	Off	8	S	0.510	Off	28	S	0.510	Off	28
05	0.000	0.510	Q	0.000	Off	8	S	0.000	Off	4	S	0.000	Off	4
06	0.220	0.290	Q	0.200	Off	8	S	0.220	Off	16	Q	0.200	Off	8
07	0.830	0.320	Q	0.800	On	8	S	0.830	On	16	Q	0.800	On	8
08	0.530	0.020	Q	0.500	Miss	8	Q	0.500	Miss	8	D	0.530	On	64
09	0.502	0.008	Q	0.600	False	8	Q	0.600	False	8	D	0.502	Off	64
10	0.730	0.220	Q	0.700	On	8	Q	0.700	On	8	Q	0.700	On	8
11	0.398	0.112	Q	0.300	Off	8	Q	0.300	Off	8	Q	0.300	Off	8
12	0.760	0.250	Q	0.800	On	8	D	0.760	On	64	Q	0.800	On	8
13	0.410	0.100	Q	0.400	Off	8	D	0.410	Off	64	D	0.410	Off	64
14	0.520	0.010	Q	0.600	On	8	Q	0.600	On	8	D	0.520	On	64
15	0.540	0.030	Q	0.500	Miss	8	S	0.540	On	28	S	0.540	On	28
16	0.498	0.012	Q	0.400	Off	8	S	0.498	Off	28	S	0.498	Off	28

D is dense, S is sparse, and Q is quantized. On and Off agree with the original alert. Miss means the original alert went quiet. False means a quiet record fired. The certified bound on the score error is 0 for D and S on these inputs and 0.100 for Q. Sizes count only the payload.

Compare records 02 and 09. The revised selector preserves the sparse record and still sends the dense record through rounding. GIGI refuses rounding for both. It can carry the sparse reading in 28 bytes, while the dense record uses 64. The guarantee is the same in both cases. The cost of meeting it depends on what the record contains.

Watching AMC change formats gave me a way to see where the representation mattered. Naming the decision gave me something I could protect across the change. The next example takes a step beyond a single alarm. I want to keep a relationship I can only see by reading three channels together.

§ 4.7

Keeping the relationship between three channels

On Luna I proposed Mz DOT as a way to bring metrics, logs, and alerts together. I gave correlation its own place in the architecture.[7] I wanted to look at what the streams were doing together. AMC let me see a further complication. A relationship I could see in one representation could become hard to see after a switch. The next question was what I needed to keep together to carry that relationship with me.

Imagine three synchronized sensors. On each row, each reports 0.04 above or below its reference. Divide the three readings by 0.04 and call the resulting signs A, B, C. I made two tables of eight rows. In the first, every sign triple occurs once. In the second, A and B take each of their four possible pairs twice, and I set $C = AB$ on each row.

Look at A alone. Four plus signs, four minus signs, in either table. Look at A and B. Each of the four pairs occurs twice. The same is true of A and C in the second table because specifying them determines $B = AC$. The argument for B and C is the same. Every single-channel summary and every two-channel table of signs agrees.

Now keep the three readings on each row together and calculate

$$M = \frac{1}{8} \sum_{j=1}^{8} A_j B_j C_j.$$

The first table has four products of each sign, so $M = 0$. In the second, every product is $(A_j B_j)^2 = 1$, so $M = 1$. I could keep all three pair tables and still miss the relationship.

I put the three readings in the last three channels of the eight-channel records used for the alarm. The other five channels are zero. Each record starts a fresh selection episode with the same feedback as before. I sent them through AMC with the original and fixed selectors, using the same offline adapters. This time I asked the detector to read the three decoded signs and multiply them. I wanted to see whether the relationship would make it through the same machinery.

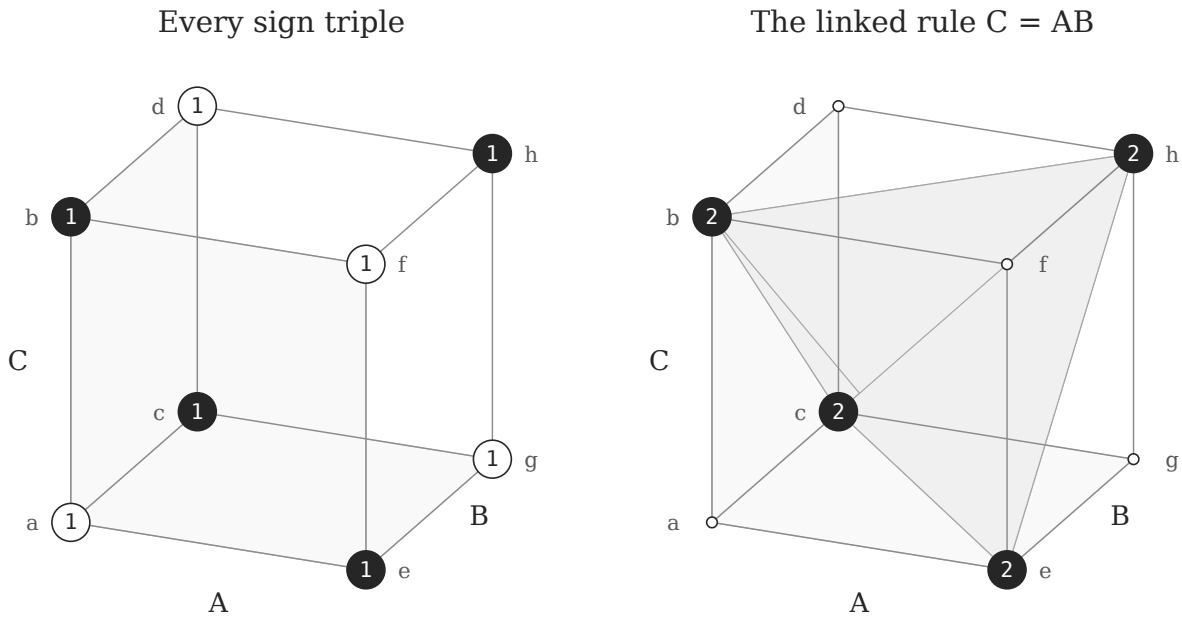
The original selector chooses quantized throughout. With a step of 0.10, both -0.04 and $+0.04$ round to zero. Forcing that codec in GIGI erases the same signs. Assigning zero to each erased product would give $M = 0$ in both tables. That would conceal the difference between a balance of plus and minus products and a loss of the signs themselves. I report the number of valid triples alongside their mean.

[7] Bee Rosa Davis, *Proposal: Mz DOT (Detection, Observability, and Tracking) for Luna Multi-Cloud Environment*, issue 28, October 2024. The surviving source snapshot contains a startup stub for the correlation service.

What the pair tables cannot see

ALIGNED SIGNS / two eight-record constructions

A Same pair counts; different three-way structure



Vertex numbers count records; a–h identify the sign cells on the next sheet.

B Forget one sign: all three pair views agree

AB

A \ B	−	+
−	2	2
+	2	2

AC

A \ C	−	+
−	2	2
+	2	2

BC

B \ C	−	+
−	2	2
+	2	2

Each cell contains two records in either construction. Pair views lose the information needed to distinguish the cubes.

The relationship lives in the aligned triple, even when every pair agrees.

Figure 14. The three signs locate a cube vertex. Marker numbers count aligned records; letters identify the sign cells on the next sheet. Dark markers have $ABC = +1$. The linked construction occupies four vertices. Its three pair-count tables match those of the full construction.

Keep the signs and their row

RETAINED REPLAY / same codecs, same three-channel readout

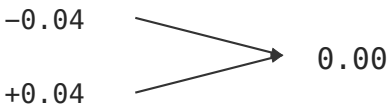
C The complete joint count, before any compression

Cell	A B C	All eight	Linked	ABC
a	− − −	1	0	−1
b	− − +	1	2	+1
c	− + −	1	2	+1
d	− + +	1	0	−1
e	+ − −	1	2	+1
f	+ − +	1	0	−1
g	+ + −	1	0	−1
h	+ + +	1	2	+1

Eight rows in each construction; shading marks positive triple products.

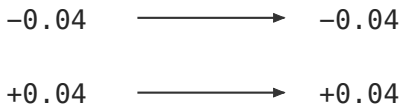
D The codec can remove the evidence

Nearest-step rounding



Both signs are erased

Exact sparse storage



Both signs survive

E Count valid triples before taking their mean

After decoding	Valid / 16	M, all	M, linked
Original AMC	0	—	—
Revised AMC	16	0	1
GIGI guard	16	0	1

An erased sign is missing evidence. It is not a measured zero triple mean.

Figure 14. *Continued.* The complete joint counts and retained replay. Each mean uses eight valid triples when available; the validity column totals both constructions. A dash means that no valid triple remains. The shaded rows have positive sign product.

Fixed AMC chooses sparse for these records and preserves all sixteen triples. GIGI can require the same preservation directly. For each of the three channels, if the possible decoding error is smaller than the reading's absolute distance from zero, the sign survives. Here that distance is 0.04 and the nearest-step quantizer's bound is 0.05. The exact codecs pass. Sparse is the smaller of the two, at $4 + 3(12) = 40$ payload bytes per row.

GIGI keeps each row's identity, selects a codec, and counts the decoded rows in the eight joint sign cells. Weight each cell count by the sign of its product and divide by the number of valid rows. That recovers 0 and 1 for the two tables, with all eight rows valid in each. Figure 14 puts the two joint tables beside the replay results.

I wanted Mz DOT to put the streams beside each other. AMC made the effect of changing their representation visible. In this GIGI experiment I can name the relationship I want and ask the system to carry enough information to check it. Here that means keeping the three signs and knowing which readings came from the same row. We have seen the two tables disagree on the three-way product. I want to make the missing information explicit, so that the result is more than a surprising example. On the same replay, there is a neat way to see exactly what is missing from the pair tables. Let a, b, c be any three signs. The expression $(1 + aA)/2$ is one when channel A has sign a, and zero otherwise. Multiply the three expressions and average over the rows. That gives the fraction of rows in the joint cell (a, b, c):

$$p(a, b, c) = \frac{1}{8}\,\mathbb{E}[(1 + aA)(1 + bB)(1 + cC)]. \tag{4.7}$$

Here $\mathbb{E}$ is the arithmetic mean on the table. On both tables from the replay, each individual sign has mean zero and each pair product has mean zero. Only one term other than the constant remains when I expand the brackets:

$$p(a, b, c) = \frac{1}{8}(1 + abcM), \qquad M = \mathbb{E}[ABC]. \tag{4.8}$$

The pair tables give the same entries for every M between -1 and 1. To see that all those values are possible for a probability distribution, the eight entries above are nonnegative and sum to one. Summing over either sign of any one channel cancels the term with M. On a finite table the attainable fractions must also fit its row count.

So if I print one pair table for each pair of channels, I have described seven of the eight coefficients in that expansion, including the constant. The missing one is exactly the number I wanted from the three-way readout. In the first replay table $M = 0$, so every joint cell gets $1/8$. In the second $M = 1$, so the four cells with $abc = +1$ each get $1/4$. That is where the extra mass on the four vertices in Figure 14 came from.

Keeping the amplitudes is only part of the job. If I pair them with the wrong times, I can lose the relationship without changing any individual reading. I can do one more check on the second table. Leave A and B where they are and move the C column one row early, with the first sign coming around at the end. The C readings are all there. Each has been paired with a different timestamp. Since $A_jB_j = C_j$ before the move, the product after the move is C_jC_{j+1}, with row numbers read cyclically. The product is negative wherever consecutive signs differ.

In the recorded row order the C column is $+, +, -, -, -, -, +, +$. Moving it one row changes two products. Moving it two rows changes four. Moving it four rows changes all eight.

Rows shifted	+1 products	−1 products	Product mean
0	8	0	1
1	6	2	1/2
2	4	4	0
4	0	8	−1

I used the cyclic shift to keep the number of rows fixed and stored the source row of every shifted third reading. GIGI retrieved all eight triples for each shift. Multiplying the signs in those returned rows gives the counts in the table. In a stream I would use the acquisition times and mark unmatched rows explicitly. Here I can reverse the sign of the relationship without changing a single amplitude.

More generally, if a replay preserves the number of rows and can have the wrong product in at most k of n rows, the mean can move by at most $2k/n$. Replacing one product sign with its opposite changes the sum by two. When I lose rows altogether, I have a different calculation: I need the count of lost rows, and I need to know what I am averaging over. A codec can carry each reading exactly and still leave both of those questions for the pipeline to answer.

§ 4.8

Following the error through time and across formats

Suppose a cooling pump sends a pressure reading every second. A format change is due, so for a short interval I can run the old and new paths side by side. I take one observation and encode it directly in the new format. Then I take the old encoding of that same observation and ask the translator to produce the new one. If the answers start to separate, I want to know when. The two paths began with the same observation, so the discrepancy gives me somewhere specific to look.

I wrote down a check for this in an early AMC draft. It was under the heading *Temporal Drift Constraint (Discrete Harmonicity)*. Let e_t be the old representation and e'_t the directly encoded target. With translator T_{ij}, the disagreement at sample t is

$$d_t = \lVert T_{ij}(e_t) - e'_t \rVert.$$

The draft proposes comparing that disagreement with the mean of its two neighbours. If the departure from the mean exceeds a tolerance, it calls for re-encoding in the source format. It also sets a separate absolute translation-error limit.[8]

Put the readings on a strip chart with equal time between them. Draw a line from the previous sample to the next one. Where would the middle sample sit if the disagreement were changing at a steady rate? It would sit exactly on that line. The vertical difference is

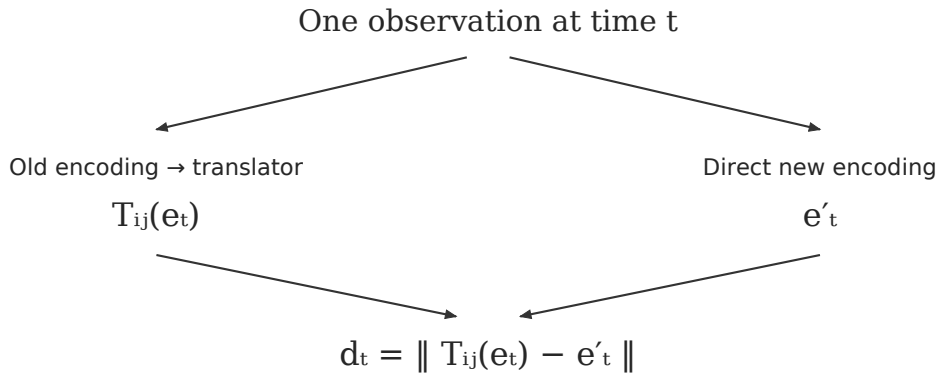
$$\begin{aligned} r_t &= d_t - \tfrac{1}{2}(d_{t-1} + d_{t+1}) \\ &= -\tfrac{1}{2}\big[(d_{t+1} - d_t) - (d_t - d_{t-1})\big]. \end{aligned}$$

[8]Bee Rosa Davis, *Adaptive Embedding Format Switching and Semantic Translation in High-Volume Telemetry Systems*, author's archive, recovered Revision 01, p. 2, § III.C. The PDF's embedded metadata dates it to August 6, 2025.

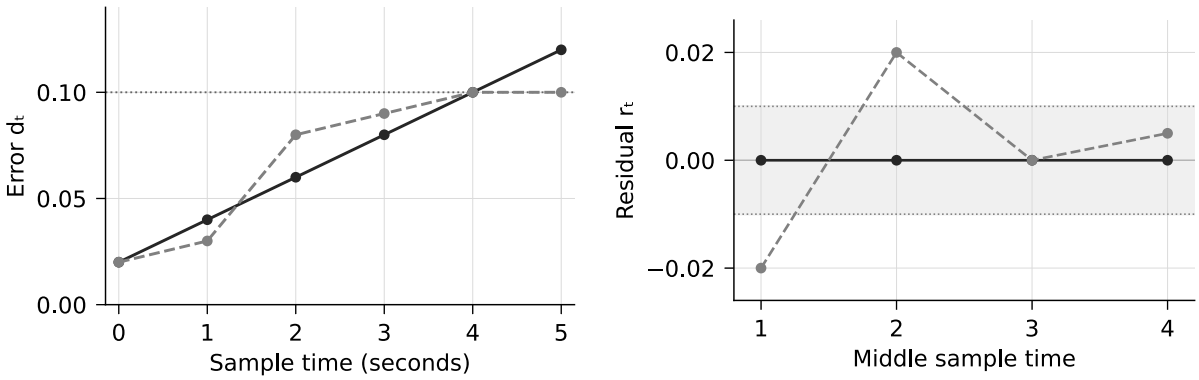
A straight error can still be too large

EQUAL ONE-SECOND SPACING / two distinct checks

A Compare two descriptions of the same observation



B The height and the bend answer different questions



Solid: steady increase. Dashed: changing rate.

Dotted limits: d = 0.10 and |r| = 0.01. Equality is accepted.

Steady d_t	0.02	0.04	0.06	0.08	0.10	0.12
Bent d_t	0.02	0.03	0.08	0.09	0.10	0.10

Columns are samples 0–5. The centered residual exists only at samples 1–4.

A shape check can pass while the error grows beyond its absolute limit.

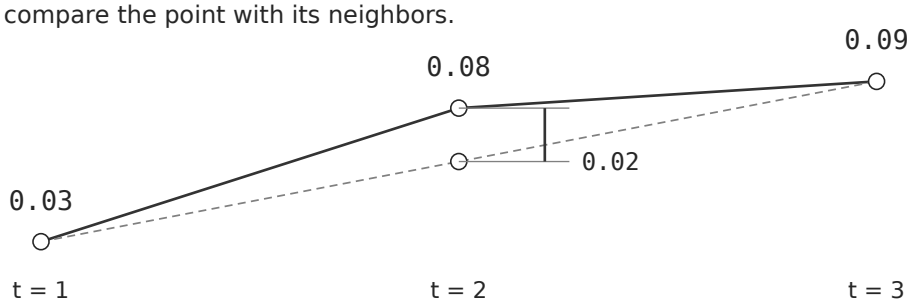
Figure 15. Two paths describe the same observation; d_t measures their discrepancy. Constructed traces separate the absolute-error limit from the centered shape tolerance. The solid trace exceeds 0.10 at sample 5 while every interior residual remains zero.

What the middle sample tells us

A CENTERED CHECK / available when the next sample arrives

C Read the local three-sample construction

At sample 2: compare the point with its neighbors.



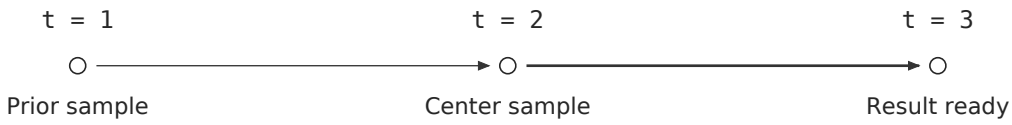
Neighbor mean: (0.03 + 0.09)/2 = 0.06

Residual: 0.08 − 0.06 = 0.02 > 0.01

D Keep the full residual record

Middle sample	1	2	3	4
Steady r_t	0.000	0.000	0.000	0.000
Bent r_t	-0.020	+0.020	0.000	+0.005

E Wait for the observation that completes the check



The centered check flags samples 1 and 2 in the bent trace.

The value at t = 3 completes the check for t = 2, one second later.

Figure 15. *Continued.* The three-sample construction at $t = 2$, its residual calculation and the full residual table. The centered check needs the next observation, so its result is attached to an earlier timestamp. All times are in seconds.

The two brackets inside the last line are consecutive changes. Their difference is the discrete second difference. Exact harmonicity on this time path means $r_t = 0$ at each interior sample. The increments are then equal, so the error follows a straight line. An approximate version accepts $|r_t| \le \tau$. If the sample spacing is h seconds, dividing the second difference by h^2 expresses it in error units per second squared. The simple neighbour average assumes equal spacing.

Suppose I set a shape tolerance of 0.01 and an absolute limit of 0.10, in the same units as the error. Six readings rise from 0.02 to 0.12, increasing by 0.02 at each step. The second difference is zero throughout the interior. The final value exceeds the absolute limit. A different trace gives $d_1 = 0.03$, $d_2 = 0.08$ and $d_3 = 0.09$. The neighbours of d_2 average to 0.06, so $r_2 = 0.02$ and the shape check flags the bend. One check catches an error that is already too large. The other catches a change in how fast the error is moving.

The centre of that test is sample 2, and I need sample 3 to compute it. With one reading per second the centred check is available one second later. On a live stream I can use the same three samples when the last one arrives and attach the result to the middle timestamp. That timing detail tells me what the proposed check could have known at the moment of a switch.

Along the clock, the Laplacian helped me see a bend in the error trace. I also used it to choose a route between formats. The word stays the same, but the neighbors change: now a point is a format, and a one-way edge is a translator that can carry a record to another format. The router first looks for a direct translation, then for two hops. When it needs the spectral calculation, it builds a symmetric adjacency matrix from the registered edges. Each directed registration contributes in both directions to this matrix. The executable version normalizes by degree, using[9]

$$L_{\mathrm{sym}} = I - D_g^{-1/2} W D_g^{-1/2}$$

on the connected example below, where W is the symmetric adjacency and D_g is its diagonal degree matrix. The second eigenvector, the Fiedler vector, assigns each format a coordinate. The search still takes only registered outgoing edges. Among its choices it favours the neighbour whose coordinate is closest to the target’s.[10]

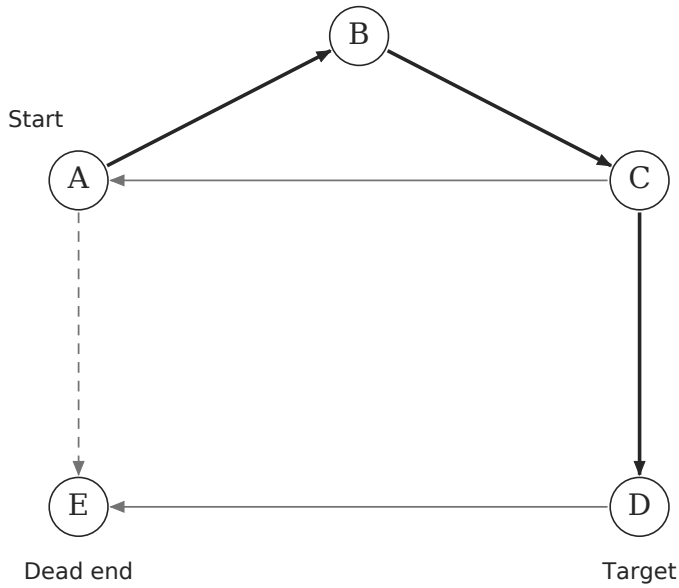
[9]The early AMC paper describes an unnormalized graph Laplacian. The recovered router uses the degree-normalized operator displayed here; the routing source is identified below.

[10]Recovered AMC source, `amc/translation_graph.py`, Git blob `b9c911a`. See `build_adjacency`, `compute_fiedler_scores`, and `find_best_path`.

Close on the map, no route onward

SIX DIRECTED TRANSLATORS / start A, target D

A Keep the direction of every available translation

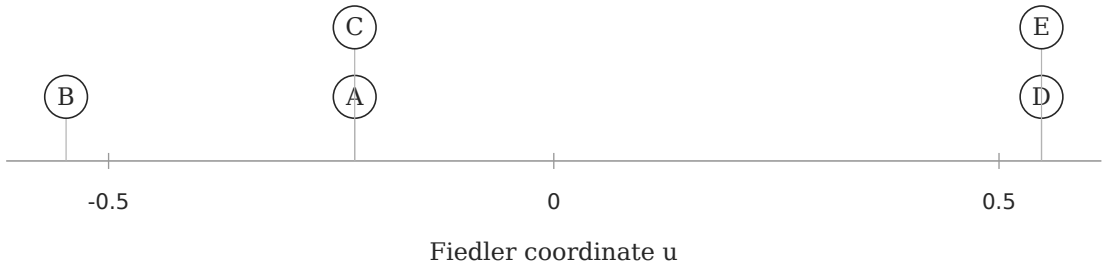


REGISTER

```
A → B
B → C
C → D
A → E
D → E
C → A
```

Heavy arrows: A → B → C → D. Dashed arrow: greedy choice A → E.

B D and E receive the same spectral coordinate



Distance to target D `B: 1.095445` `E: 0`

The nearest coordinate is E, but only the route through B reaches D.

Figure 16. The six arrows are the complete translator registry in this constructed graph. Heavy arrows reach D in three steps; the greedy step to E reaches a dead end. Symmetric spectral coordinates place D and E together without supplying an outgoing translator at E.

Suppose my archived records are in format A and my detector needs format D. In the small graph in Figure 16, I have translators $A \to B$, $B \to C$, $C \to D$, $A \to E$, $D \to E$, and $C \to A$. The degrees of the symmetric graph in the order A, B, C, D, E are $(3, 2, 3, 2, 2)$. The eigenvalues of L_{sym} are $0, \frac{2}{3}, 1, \frac{3}{2}, \frac{11}{6}$. A unit vector for eigenvalue $\frac{2}{3}$ is

$$u = \frac{1}{\sqrt{20}}(-1, -\sqrt{6}, -1, \sqrt{6}, \sqrt{6}).$$

At A my outgoing choices are B and E. The spectral distance to D is $2\sqrt{6/20}$ from B and zero from E. Reversing the eigenvector sign changes neither distance. The greedy search goes to E, where I have no outgoing translator. It returns an empty route and the caller can try a direct encoder. The three-hop chain $A \to B \to C \to D$ was available the whole time. A breadth-first search of the directed graph finds it. I checked both outcomes with the recovered routing functions using these six registrations.

The clock and the format graph give the Laplacian two jobs. Along the clock it measures a bend in a sequence of errors. On the format graph it supplies a coordinate for choosing which translator to try. I still have to follow the actual arrows and measure the discrepancy at the end. This is where the question that began at the pump comes back. Did I preserve the reading the detector needed?

§ 4.9

Keeping the answers the measurements leave open

The AMC drift check starts with numbers I can put in order along a time line. In GIGI I came to a different question. Suppose I know a few readings of a function and I have a good reason to assume it is holomorphic and bounded. What values could it take at a point I have left out? I want the entire region of possible values, so I can see how far the data and the assumption take me.

Let me slow down at those two words. A *harmonic* function has zero Laplacian. For temperature on a thin ideal plate with no internal heat source, that describes a steady profile: heat flowing into a small neighborhood balances heat flowing out. At an interior point, the value also equals the average around a sufficiently small circle centered there. Steady does not mean constant. One side of the plate can be warmer than the other while the profile no longer changes with time.

A *holomorphic* function takes a complex input $z = x + iy$ and returns a complex value $f(z) = u(x, y) + iv(x, y)$. Complex differentiability requires one derivative to work regardless of the direction from which a small input change approaches. That is stronger than having two differentiable real functions. It ties their slopes together through the Cauchy–Riemann equations,

$$u_x = v_y, \qquad u_y = -v_x.$$

For a holomorphic function, these relations make both u and v harmonic. Differentiate once more and the mixed derivatives cancel in $u_{xx} + u_{yy}$ and $v_{xx} + v_{yy}$.

Try $f(z) = z^2$. Its two parts are $u = x^2 - y^2$ and $v = 2xy$. The second derivatives of u sum to $2 - 2 = 0$; those of v sum to zero too. Their first derivatives also satisfy the two equations above. The pieces are harmonic, and their pairing is holomorphic. By contrast, the real function $u = x$ is harmonic, but declaring the complex-valued output to be $x + 0i$ does not make it holomorphic: $u_x = 1$ while $v_y = 0$. Pair x with y instead and we recover $f(z) = z$.

That is the connection I want to keep visible. Harmonicity constrains a scalar profile. Holomorphicity couples two profiles through one complex derivative. A suitable harmonic function has a conjugate locally; global holes can obstruct a single-valued conjugate across the whole domain. The time-line harmonicity test in the AMC draft and GIGI's bounded holomorphic completion therefore share an interest in compatible local values, while asking different mathematical questions.

For a calculation I can do on paper, let f be holomorphic in the unit disk with $|f(z)| \le 1$, and suppose $f(0) = a = 1/4$. I want to know what $w = f(1/2)$ could be. The number $1/4$ is the observation. The disk and the bound on $|f|$ are assumptions of the example. Both will enter the result.

The unit disk is the set $|z| < 1$. The assumption $|f(z)| \le 1$ applies throughout that disk, including places I have not measured. One observed value cannot establish this assumption. I supply it as part of the model, and the conclusion is conditional on it. The next transformation moves the observed output a to the center, where a standard theorem can constrain how quickly the function moves away. Since $a = 1/4$ is real, the map is

$$T_a(w) = \frac{w - a}{1 - aw}$$

This transformation takes the unit disk to itself and sends the known value a to zero. I can now apply Schwarz's lemma to $g = T_a \circ f$. Since $g(0) = 0$, the value at a point of modulus r has modulus at most r. For $r = 1/2$ this gives

$$\left| \frac{w - a}{1 - aw} \right| \le \frac{1}{2}.$$

There is a converse for this example. For any w satisfying the inequality, the map $g(z) = 2T_a(w)z$ stays in the disk, vanishes at zero, and takes the prescribed value at $1/2$. Applying T_a^{-1} gives an admissible f. Every point in the region is possible.

The inequality describes every allowed value. To see whether it settles a decision, I first want to draw that whole region. Squaring and expanding the inequality gives

$$(1 - a^2r^2)|w|^2 - 2a(1 - r^2)\operatorname{Re} w + (a^2 - r^2) \le 0.$$

Completing the square puts it in the form $|w - c| \le R$, where

$$c = \frac{a(1 - r^2)}{1 - a^2r^2} = \frac{4}{21}, \qquad R = \frac{r(1 - a^2)}{1 - a^2r^2} = \frac{10}{21}.$$

The possible real values run from $-2/7$ to $2/3$. Both endpoints are attained. Use $g(z) = -z$ for the left endpoint and $g(z) = z$ for the right. They agree on the measured value and disagree at the withheld point. Figure 17 shows where both answers sit.

I can also write the same test in the form that generalizes to several readings. For the two points 0 and r, the Pick matrix is

$$\begin{pmatrix} 1-a^2 & 1-a\overline{w} \\ 1-aw & (1-|w|^2)/(1-r^2) \end{pmatrix}.$$

Positive semidefinite means that every complex vector c gives $c^*Pc \geq 0$ for this matrix P. It is a compatibility condition across the observations. Here the upper-left entry is positive, so nonnegative determinant is equivalent to that condition. Multiplying out the determinant gives the same disk inequality. With more observations, the matrix test keeps their compatibility conditions in one calculation.[11]

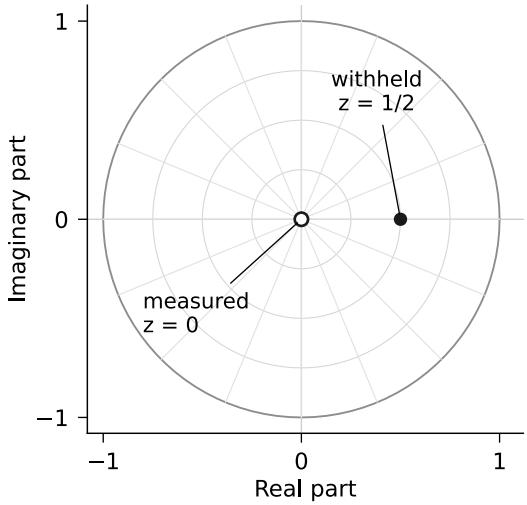
[11] GIGI, *Bounded Holomorphic Completion*, implementation specification, October 29, 2026. The example here is the two-point unit-disk calculation.

One reading, a region of answers

MODEL / holomorphic on |z| < 1, with |f(z)| ≤ 1

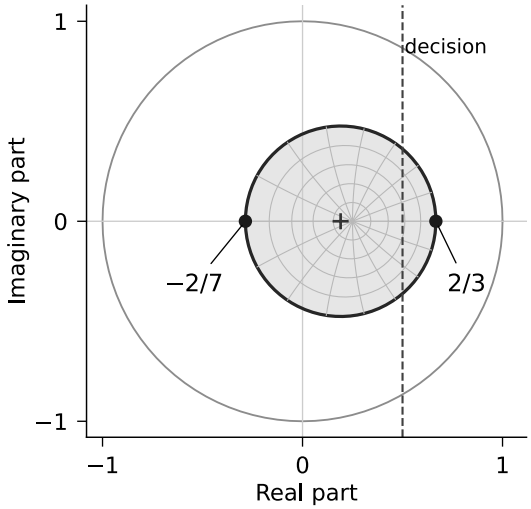
A Keep the observed point and the question distinct

INPUT DISK



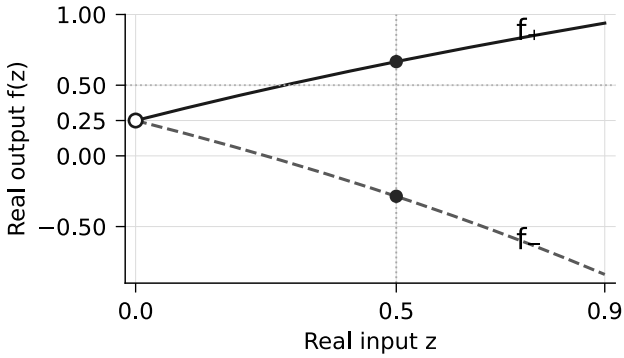
Known: f(0) = 1/4

POSSIBLE OUTPUTS AT 1/2



Center 4/21 · radius 10/21

B Two functions attain the two real endpoints



$f_+(z) = (1/4 + z)/(1 + z/4)$

$f_-(z) = (1/4 - z)/(1 - z/4)$

Both give 1/4 at zero. At 1/2, one gives 2/3; the other gives −2/7. The same observation permits opposite decisions.

The shaded disk is the full answer allowed by one reading and this model.

Figure 17. The observed value is $f(0) = 1/4$; the withheld value is $f(1/2)$. Its feasible disk has center $4/21$ and radius $10/21$. The interior grid maps circles and rays through T_a^{-1}; it is a coordinate guide, not extra observations. The two curves below attain the real endpoints.

If a decision says to act whenever the real part is above 0.5, the observed value at zero leaves the decision open. The feasible disk crosses that line. I can say exactly why I need another observation. This is the same kind of discipline I wanted from the alarm example. State the task, say which assumptions are carrying it, and keep the range of answers the evidence still allows.

§ 4.10

Taking the question beyond telemetry

Now I see the pattern everywhere.

Once you understand that shape is fundamental—that the essential structure of any problem is the geometry of its possibility space—you cannot unsee it. It is like learning a language. Before, the signs were noise. After, they are meaning. The world did not change. Your perception changed. And then everything looks different.

I started keeping a list. Every time I encountered a new field, a new problem, a new dataset, I would ask: what is the shape space here? What are the objects? What are the equivalences? What is the manifold?

Computer vision: Images are shapes. I modeled the space of possible images as a manifold. Object recognition is finding which region of image shape space a given image belongs to. A cat and a dog are nearby in pixel space but distant in meaning space. A photo of a cat and a painting of a cat are distant in pixel space but nearby in meaning space. The algorithm's job is to learn the manifold, not the pixels.

Natural language: Sentences are shapes. I wanted to treat the space of possible sentences as a manifold of meaning, with word embeddings as coordinates. Before saying what curvature a language model had learned, I would need to specify the representation and the metric. With those fixed, the question becomes which meanings are close, which are far, which directions are easy to traverse and which are blocked.

Genomics: DNA sequences are shapes. Evolution is a path through the manifold of possible genomes. Natural selection is a force field on the manifold—it pulls toward certain regions (high fitness) and pushes away from others (low fitness). Extinction is falling into a sink.

Economics: Market states are shapes. Crashes are rapid movements to unexpected regions. The geometry of the financial manifold has valleys (stable equilibria), ridges (unstable equilibria), and cliffs (discontinuities where small changes cause large effects).

The Davis-manifold program asks a reusable question: what state space, paths and margins make a comparison meaningful? Its answer must be built for each domain. A quotient may fail to be smooth, a learned map may collapse distinctions, and an empirical success does not make every possible representation a Riemannian manifold.

I remember the moment I understood this universality. I was reading a paper on protein folding—the problem of predicting how a chain of amino acids will curl into a three-dimensional structure. And I realized: the folding problem is a shape space problem. The chain can fold into billions of possible shapes. Each shape has an energy. The manifold of possible folds has a landscape—high mountains (unstable folds) and deep valleys (stable folds). The protein "finds" its shape by rolling downhill in this landscape.

The same geometry. Different objects. Same structure.

Here is the connection to the mass-gap question. A gauge connection A tells you how to transport local reference frames. Two connections related by a gauge transformation give the same values for gauge-invariant observables. They are the same shape for the purpose of this comparison.

The quotient groups the gauge-equivalent descriptions into single physical configurations. A quantum state can assign amplitudes across configurations. To talk about its energy I need the Hamiltonian, and to talk about the mass gap I need the lower edge of that Hamiltonian's spectrum above the vacuum.

The scanner and the gauge field give me the same first question. Which descriptions should I identify? They use different measurements and different rules for what may change. Once I have declared those rules, I can check whether the quantity I care about survives the identification.

Once you see the manifold, you see it everywhere. And once you see it everywhere, you start to wonder: what else is hiding in the geometry? What other problems are secretly shape space problems? What other mysteries might yield to the same tools?

Lo que parece diferente a menudo comparte la misma raíz. What seems different often shares the same root. The scanner, the signals, the virus, the field—different surfaces of the same deep structure.

The replay has made the question sharper. A description can preserve a shape, an alarm or a relationship between channels. The next step is to choose an answer that can be checked against observations outside the representation itself. The Davis Manifold is a stage. But a stage is useless without actors, without action, without something happening in real time.

I had the manifold. I had the shape space. I could describe signals, bodies, preferences as points in a geometric structure. But describing is not detecting. Describing is not predicting. Describing is not *acting* when something goes wrong.

The next step came from an unexpected direction: viral surveillance.

In 2020, the world learned what happens when a pathogen escapes. I watched from my living room—we were all watching from our living rooms then—as the news tracked case counts and death tolls and the desperate search for treatments that worked. SARS-CoV-2 mutated, evolved, found new shapes that evaded the immune responses we had built. Each variant—Alpha, Delta, Omicron—was a movement in shape space. A drift across the manifold of possible viral sequences. A shift that changed everything.

I looked at this and saw geometry.

A viral sequence is a shape. The spike protein that lets the virus enter your cells has a three-dimensional structure, and that structure determines whether your antibodies can recognize it. Mutate the structure, and the shape changes. Change the shape enough, and the immune system that learned to fight the old variant cannot see the new one. This is antigenic drift—the slow mutation that lets a virus escape immunity. And antigenic drift is movement on a manifold.

If you could track position on the manifold, you could detect drift *before* the variant spread. You could see the shape changing *before* the hospitals filled. You could act *before* it was too late.

This became HERALD: High-resolution Early Recognition of Antigenic Landscape Divergence. A system that watches the shape space of viral sequences, tracking movement, detecting when the

virus is drifting toward a form that will escape current vaccines. The manifold insight, applied to the most urgent problem of our time.

And in building HERALD—in formalizing the drift detection, in deriving the bounds, in proving when detection was possible and when it was not—I first wrote down the relationship I would later call the Davis Law. I wanted a rule that connected curvature to information, geometry to inference, the shape of a space to what can be known about the things that live in it. Whether it deserves that name is a question I take up in Chapter 6, after the lattice gave me a harder place to test it.

Looking back at the IDAS assignment, I could see what I had made of that unwanted space. The project had given me a question I could carry beyond NASA.

There is a strange gift in being underestimated. When no one expects anything from you, you are free to follow your curiosity. When no one is watching, you can explore the weird corners, the connections that do not fit the official project plan, the ideas that would be dismissed as "out of scope" if anyone were paying attention.

I have been underestimated my entire career. The woman who does not look like an engineer. The Black face in rooms full of white faces. The trans body that sets off alarms. Every system I have encountered has tried to fit me into a box I do not fit, and every time, I have found a way around, over, through. Not because I am special—because the boxes are wrong. Because the topology is wrong. Because the people who design the systems do not understand the actual geometry of the space they are trying to map.

Being the edge case had taught me to distrust a box that could not hold what was in front of it. At IDAS, I could finally turn that habit into a question about the records themselves: what does this description make visible, and what does it leave outside?

Lo que otros llaman fracaso, yo llamo semilla.

What others call failure, I call seed.

My career is a shape. Twenty-seven years of aerospace, intelligence, cybersecurity—positions and projects and problems that seemed disconnected, that puzzled the people who reviewed my resume, that did not fit any standard trajectory. NSA to DISA to NASA. Satellite systems to network security to lattice gauge theory. A path that makes no sense if you think in straight lines.

But in shape space, it makes sense. Each position was a point on a manifold. The path between them was a geodesic—the shortest route through the space of possible careers, given the constraints I faced, the doors that were closed to me, the curvature of the professional landscape. The geodesic is not a straight line in embedding space. It curves and bends and sometimes doubles back. But it is the shortest path in the space itself, the path that respects the geometry.

The geodesic equation does not care who is walking the path. It cares only about the geometry. Find the right path—the one that respects the curvature—and you can traverse terrain that looks impassable. I have been finding such paths my whole life. Now I have the mathematics to describe them.

The next test had to come from outside the representation. HERALD gave me one: a historical sequence of variant observations and a designation date to compare its alerts against. The HERALD preprint records that comparison.[12] That was a harder demand than finding a useful way to describe the space. Then I wanted to carry the same question farther from biology than I had ever gone: to a gauge field on a lattice, where I would have to say exactly what my measurements kept.

La forma no es la cosa. La forma es el espacio donde vive la cosa.

The shape is not the thing. The shape is the space where the thing lives.

[12]Bee Rosa Davis, "HERALD: High-resolution Early Recognition of Antigenic Landscape Divergence," Zenodo, 2025, `https://doi.org/10.5281/zenodo.17682618`.

Chapter 5

The First Door

La obsesión es el precio de la comprensión.

Obsession is the price of understanding.

HERALD had given me an alert date to examine in the historical data, earlier than the WHO designation used in that comparison. I wanted to take the geometric questions somewhere entirely different. That was the confidence I carried into the lattice work. The new problem still had to earn its own proof.

I should have felt triumphant. Instead I felt hollow.

The Davis Manifold had detected viral drift. It had given the geometric approach its first test against history. But viral surveillance felt like a trick—a clever trick that worked in biology but might not mean anything deeper. A pattern that happened to fit viruses and tumors but said nothing about the universe itself.

I wanted to know how far the geometry would go. Could it help me ask a question about the foundations of physics? I wanted to see where the comparison held and where it broke.

Here is the question that haunted me: *Why does the geometry work?*

For viruses, I had explanations. The spike protein is a physical object. Its shape determines what antibodies can bind to it. The antigenic manifold encodes this shape, and variants that move far enough in the manifold escape immunity. The geometry is the biology.

But where does the geometry come from? Why does the universe have this structure? Why is there a manifold at all, with curvatures and distances and spectral gaps?

Biologists would say: evolution. The geometry emerges from the selective pressures that shape proteins over billions of years. The spike protein is optimized for binding, for fusion, for viral entry. Its structure is a frozen record of evolutionary history.

But that answer pushed the question back. Evolution operates on physical laws. Proteins fold according to chemistry. Chemistry follows from quantum mechanics. Where does the geometry live, fundamentally?

I suspected the geometry was deeper than biology. That it reached down to physics. That the same structures I was seeing in viral evolution would appear in quantum field theory, in gravity, in the basic laws that govern everything.

But to test this suspicion, I needed a problem. A clean problem where the physics was well-understood, where I could apply the geometric framework and check whether it gave the right answer.

So I asked AI: *What is the hardest problem in mathematics?*

AI told me about the seven Millennium Prize Problems—million-dollar bounties posted by the Clay Mathematics Institute in 2000 for solutions to the deepest unsolved questions in mathematics.

Six remain unsolved. One—the Poincaré Conjecture—was solved by Grigori Perelman, who refused the prize. Perelman also refused the Fields Medal. He said something that stuck with me: "I know how to compute the curvature of the universe. Why would I need a million dollars?"

The problems are not just mathematics. They are physics. They are the places where our understanding of the universe breaks down, where the equations stop working, where humanity has been stuck for decades.

I read through them one by one, late at night, Teddy asleep at my feet.

Riemann Hypothesis—about the distribution of prime numbers. It asks whether every nontrivial zero of a certain function lies on a critical line. Checked numerically for trillions of zeros. Never proved in general. If true, it would reveal deep structure in the prime numbers, the atoms of arithmetic.

P versus NP—about the nature of computation. Can every problem whose solution is easy to verify also be easy to solve? The difference between checking a password and cracking it. Between verifying a proof and finding one. The question at the heart of computer science.

Navier-Stokes—about the behavior of fluids. Do the equations that govern water and air always have smooth solutions, or can they blow up? We use these equations to design airplanes and predict weather, but we do not know if they are mathematically consistent.

Hodge Conjecture—about the structure of algebraic geometry. Which topological features of a shape can be represented by polynomial equations? The bridge between topology and algebra.

Birch and Swinnerton-Dyer—about elliptic curves. These curves encode deep arithmetic information, used in cryptography and number theory. The conjecture connects their geometry to the distribution of their points.

And Yang-Mills.

Physics is my secret love.

Not in the way a physicist loves it—with equations mastered and intuitions honed through years of problem sets. My love is different. It is the love of a reader for a foreign literature, understanding the stories even when the grammar is unfamiliar.

I have written novels with deep physics at their core—*Fly Into the Sun*, *A Pale Jewel*, *Murder Me Lovely*—a trilogy following a fictional version of me traveling through different versions of spacetime. In those books, I wrestled with general relativity, with the many-worlds interpretation, with the paradoxes of time travel. I read Penrose and Hawking and Feynman and Susskind. I absorbed the concepts even when the calculations were beyond me.

I understand the language of physics the way a devoted reader understands the language of poetry: not as a practitioner, but as someone who has spent years immersed in the words, the rhythms, the meanings. I know what a gauge field is. I know what symmetry breaking means. I know why quarks are confined.

I knew what Yang-Mills was about. Gauge theory. The strong force. Quarks and gluons. The question of why we never see free quarks—why they are always confined inside protons and neutrons, bound by a force that does not weaken with distance.

The strong force is strange. Electromagnetism gets weaker with distance—pull two charges apart, and the force between them diminishes. Gravity works the same way. But the strong force

does not weaken. Pull two quarks apart, and the force between them stays constant, or even increases. This is why you cannot isolate a quark. The energy required to separate them is so large that it creates new quarks instead. You end up with more particles, not fewer.

This phenomenon is called confinement. A mass gap governs the decay of correlations of suitable gauge-invariant observables. Confinement concerns the cost of separating color sources. The two questions use different measurements.

Proving the mass gap exists has defeated every mathematician and physicist who has tried for fifty years.

I thought: *What if the geometry reaches here too?*

§ 5.1

Choosing what to measure in a field

Let me tell you what has not worked.

Perturbative methods fail because the coupling is strong. In quantum electrodynamics—the theory of light and electrons—the coupling constant is small, about 1/137. You can expand in powers of this small number and get good answers. But in quantum chromodynamics—the theory of quarks and gluons—the coupling constant at low energies is of order 1. There is no small number to expand in. The perturbation series diverges. The method breaks.

Lattice QCD provides numerical evidence but not proof. You discretize spacetime into a grid, represent the gauge field as matrices on the links, run Monte Carlo simulations to sample configurations. The simulations show a mass gap. The glueball spectrum has been computed. But numerical evidence is not mathematical proof. The Clay Institute wants a theorem, not a simulation.

The constructive approach to quantum field theory faces technical obstacles in four dimensions. Constructive methods have succeeded in some lower-dimensional field theories. Four-dimensional pure Yang–Mills still requires a construction of the continuum theory and control of its low-energy spectrum.

Topological approaches give partial results. You can prove things about instantons, about the theta vacuum, about the structure of the configuration space. But these results do not close the gap. They illuminate the geometry without proving the spectral bound.

All of these approaches attack the problem from within the standard framework. They try to compute the spectrum of the Hamiltonian directly, or construct the theory rigorously, or bound the gap using known inequalities.

Let me name some of the people who have worked on this problem. These are not amateurs. These are the best minds in mathematical physics.

Edward Witten—widely considered the greatest living theoretical physicist, a Fields Medalist, a MacArthur Fellow. He has written papers on Yang-Mills, on instantons, on the geometry of gauge theory. He has not solved the mass gap.

Arthur Jaffe—Harvard professor, co-founder of constructive quantum field theory, original member of the Clay Mathematics Institute scientific advisory board. He wrote the official statement of the Yang-Mills problem for the Millennium Prize competition. He has not solved it.

James Glimm—another founder of constructive field theory, proved existence of quantum field theories in lower dimensions. His methods have not extended to four-dimensional Yang-Mills.

Michael Atiyah—Fields Medalist, Abel Prize winner, one of the greatest geometers of the twentieth century. He worked on gauge theory and topology. He announced a proof of the Riemann Hypothesis shortly before his death. It did not hold up. He never claimed to solve Yang–Mills.

These are the people I was measuring myself against. Not consciously—I never thought "I am smarter than Witten." But by attempting the problem, I implicitly claimed my approach might work where theirs had not.

Either that, or I was delusional. Both were possible.

The heat-kernel calculation from Chapter 3 suggested a way into the problem. I wanted to organize observations of the field before asking what their geometry could establish. I had a different idea.

What if the mass gap is not a dynamical fact but a geometric one?

Classical gauge configurations give me the inputs to the field measurements. Gauge-equivalent descriptions should be identified, although their quotient can have singular structure. The quantum vacuum belongs to the quantum state space built for the theory. It is not generally one point in the classical feature cloud.

The mass gap asks for a positive lower bound on excitation energy above the vacuum. The word "gap" will also appear in a plot of sorted measurements later in this chapter. I need to know which gap I am looking at before I decide what it establishes.

The surveillance work had trained my attention on separation and on what a measurement could distinguish. This was a different question. A distance in a learned representation and an energy bound for a Hamiltonian need different evidence. I wanted to see whether geometry could help supply the latter.

I would use the same tools:

- Heat kernel on the configuration space
- Spectral analysis of the Laplacian
- Curvature bounds from the topology

A geometric argument would have to show that every quantum state orthogonal to the vacuum pays a positive energy cost. Which geometry could control that cost?

Si la geometría prohíbe, la física obedece.

If geometry forbids, physics obeys.

Let me explain the heat kernel idea, because it is central to the approach.

Imagine you have a cold room and you place a hot object in the center. Heat spreads. It diffuses outward, following the geometry of the room. Corners take longer to warm. Obstacles block the flow. The way heat spreads tells you about the shape of the space.

The heat kernel $K(x, y, t)$ measures how much heat reaches point y at time t if you start with a unit heat source at point x. It encodes the geometry of the space in a deep way. The eigenvalues of the Laplacian—the operator that governs heat diffusion—appear directly in the heat kernel:

$$K(x, y, t) = \sum_n e^{-\lambda_n t} \phi_n(x) \phi_n(y) \tag{5.1}$$

where λ_n are the eigenvalues and ϕ_n are the eigenfunctions.

On a connected compact Riemannian manifold without boundary, the first positive eigenvalue λ_1 of $-\Delta$ controls the slowest decay of a mean-zero temperature profile. That gives me a model of what to look for. The Yang–Mills Hamiltonian acts on its own space of states, so I need an argument that connects its energy form to the geometry I am measuring.

Think of the Laplacian as an operator comparing a field with its neighborhood. Its eigenvalues give diffusion decay rates; for an appropriate wave equation their square roots give vibration frequencies. A gap in an operator spectrum is a restriction on those modes. It does not imply that a visible empty region of a feature cloud is physically forbidden.

A sequence of observations can drift even when I have specified no diffusion equation for it. Heat flow gives a precise comparison model because its generator is explicit. Before using its decay rate in another setting, I have to identify the other setting's generator or prove an appropriate comparison.

For Yang–Mills, the promising analogy was the suppression of excited components under imaginary-time evolution. To use it rigorously, I need the physical Hamiltonian and control of its spectrum. My finite geometric measurements would be evidence to examine for such a connection.

To make that idea executable, I needed a finite description of each configuration. The next construction specifies which measurements enter it. Here is the key construction. I call it the Davis-Wilson Map, after myself and Kenneth Wilson, who invented lattice gauge theory.

Kenneth Wilson developed lattice methods for gauge theory, including a strong-coupling argument relating confinement to the area law for Wilson loops. His 1982 Nobel Prize recognized his theory of critical phenomena and the renormalization group, rather than a prize specifically for lattice gauge theory.

I am building on his foundation. The Davis-Wilson Map uses Wilson's lattice framework to construct a finite-dimensional representation of the configuration space.

The problem with Yang-Mills configuration space is that it is infinite-dimensional. Every point in spacetime has a gauge field value. The space of all such configurations is unmanageably large—infinite dimensions, uncountably many points.

But we do not need to track every configuration. We only need to track configurations that are *distinguishable*—configurations that differ in ways that matter for physics.

Think of it like a library. The Yang-Mills configuration space is a library containing every possible book—an infinite collection. A catalog does not copy every word of every book. It records selected features so we can find what we need. That is the comparison I want here: choose measurements of the fields, organize the records by those measurements, and then check which differences the catalog has left out.

The Davis–Wilson Map is the librarian's catalog system. It assigns an address from selected measurements. Whether the address contains enough information depends on the question I intend to answer. Different fields can share a finite address, so I have to test what the catalog merges.

The map has two components:

$$\Gamma(A) = (\Phi(A), r(A)) \tag{5.2}$$

Φ is the continuous component—it captures the local geometry of the gauge field. Specifically, Φ consists of Wilson loop traces on a skeleton of paths. A Wilson loop is the trace of the product

of gauge field matrices around a closed path. It is gauge-invariant—it does not change when you apply a gauge transformation. Wilson loops are the natural observables for gauge theory.

For the lattice calculation, picture a square whose edges each carry a matrix. I start with a list of internal components and apply the matrices in order as I go around the square. In the SU(3) model, that list has three complex components; physicists call them color components, but they are not visible colors or directions on the page. Back at the starting corner, the product matrix tells me how the list has changed. Adding its diagonal entries gives the trace. That one number summarizes this particular trip through the field.

r is the recorded charge label. In a smooth continuum setting with suitable boundary conditions, topological charge is an integer preserved by admissible continuous deformations. At finite lattice spacing a field-based estimator can be real-valued; the recorded label may include smoothing and rounding. I keep the estimator and that labeling rule separate.

Topological charge is like the winding number of a rubber band around a pole. You can stretch and deform the band, but the number of times it wraps around is fixed—you cannot smoothly unwrap it. The topological charge captures similar global structure in the gauge field.

Together, Φ and r give a finite-dimensional representation of the infinite-dimensional configuration space. A finite set of measurements can send different gauge orbits to the same address. Which distinctions survive is a question the construction has to answer.

This map is a summary, and thinking of it as a library catalogue helps me keep track of what a summary can leave out. Let me make this more concrete.

Imagine the library of all gauge field configurations. Each book is a different arrangement of the gauge field throughout spacetime. The library is infinite—infinitely many books, infinitely many shelves.

Now imagine you are a librarian who needs to organize this library. You cannot read every book. You cannot even look at every page. But you can sample each book—flip to a few key pages, check certain features, record a summary.

The Wilson loops are like checking key passages. You trace around certain paths in the book and record what you find. Different books might have the same summary—that is fine. You put them on the same shelf.

The recorded charge gives the catalog another label, meant to retain global information that the selected local measurements might not distinguish. A book's genre is only a loose picture of such a label. What protects topological charge is the class of continuous changes and boundary conditions already specified, not a rule about how many words we can edit. For the lattice records I also need the estimator and rounding rule that produced the label.

Two books on different shelves differ in their summaries or their genres. My question was whether a difference in a field's shelf address could tell me anything quantitative about the field's action.

A shelf address tells me where to look. To turn that distance into an energy bound, I need a rule that relates the catalog to the Hamiltonian.

§ 5.2

Building a calculation I could check

To test the idea numerically, I needed a regulated field model on a finite lattice. I used pure SU(3) gauge fields, without dynamical quark fields. Lattice QCD provides much of the computational language, but the finite pure-gauge model and the full continuum quantum theory are different levels of the construction.

I knew nothing about lattice QCD. I had never written a line of code for it. I had never even read a textbook on it.

I learned. AI taught me. I read papers. I wrote code. I broke things. I fixed them.

The basic setup is elegant. Spacetime becomes a four-dimensional grid of points. Between each pair of neighboring points is a "link." On each link lives a matrix—an element of the gauge group SU(3). The gauge field is encoded in these link matrices.

SU(3) is the group of 3×3 complex matrices U satisfying $U^\dagger U = I$ and $\det U = 1$. The dagger means conjugate transpose. Such a matrix preserves the complex inner product of a three-component vector. The three components are internal color coordinates, not three directions in the room. In this pure-gauge simulation I store the link matrices that compare internal frames at neighboring sites; I do not simulate quark particles traveling between them.

El color no es lo que ves. Es cómo la fuerza fuerte ve.

The color is not what you see. It is how the strong force sees.

To generate configurations, you use the Metropolis algorithm or one of its variants. Think of it like exploring a mountain range while blindfolded. You take a step in a random direction. If you go downhill (lower energy), you take the step. If you go uphill, you flip a weighted coin—the steeper the climb, the less likely you are to take it. Over time, you end up spending most of your time in the valleys, occasionally visiting the ridges. The probability of being at any location matches how "favorable" that location is physically.

You start with some initial state—usually all identity matrices, the "cold start." Then you propose random updates to individual link matrices. You accept updates that lower the action, and accept some updates that raise the action with a probability that depends on how much they raise it. After many updates, the system equilibrates—it forgets where it started and settles into its natural statistical behavior. The configurations you sample are distributed according to the path integral measure, which is physics-speak for "sampled with the right probabilities."

I implemented Cabibbo-Marinari heatbath updates for SU(3). This is the standard algorithm for lattice QCD. It decomposes each SU(3) update into three SU(2) subgroup updates, each of which can be done exactly using a heatbath method. The result is efficient thermalization and good sampling of the configuration space.

Writing the calculation was one problem. Running enough of it to inspect the outputs was another. I needed compute. Real compute—GPU clusters that could run Monte Carlo simulations of the lattice, generate configurations, compute observables.

I used Modal—a cloud platform for running Python on GPUs. You write a function, add a decorator, and Modal runs it on whatever hardware you specify. A100s. H100s. Clusters of them.

I knew nothing about Modal either. I learned that too.

I spent about $500 in a few days. Money I did not have. Money I could not afford. But the idea was too compelling to stop.

The code is surprisingly simple once you understand it. Initialize the lattice. Thermalize with heatbath sweeps. Generate configurations separated by decorrelation sweeps. Compute Wilson loops. Compute topological charge. Store the results.

```
# Pseudocode for lattice harvest
U = initialize_gauge_field(L=8)  # 8^4 lattice
for _ in range(n_thermalization):
    U = heatbath_update(U, beta=6.0)

for config in range(n_configurations):
    for _ in range(n_decorrelation):
        U = heatbath_update(U, beta=6.0)
    W = compute_wilson_loops(U, skeleton)
    Q = compute_topological_charge(U)
    store(W, Q)
```

The parameter $\beta = 6/g^2$ controls the coupling—how strongly the gluons interact with each other. High β means weak coupling; the gluons are polite, they barely notice each other, and the theory becomes almost free. Low β means strong coupling; the gluons are gregarious, constantly interacting, and the physics is wild.

For the pure SU(3) Wilson action, conventional scale setting puts the spacing at $\beta = 6.0$ near 0.1 fm. A single spacing tells me how large the grid cells are in that convention. It takes a sequence of spacings to see which features survive as the grid gets finer.

El continuo es el límite. La rejilla es el camino.

The continuum is the limit. The lattice is the path.

An 8^4 lattice has $8 \times 8 \times 8 \times 8 = 4096$ sites. Four forward links per site give $16,384$ stored link matrices. A 3×3 complex matrix uses eighteen real components, so a straightforward array stores $294,912$ real numbers. They are constrained numbers: an SU(3) matrix has eight independent real degrees of freedom. Multiplying by eight gives the $131,072$-dimensional link-configuration space before gauge identification. Storage count and dimension are not the same count.

One configuration assigns a matrix to every stored link. The Monte Carlo algorithm proposes changes and accepts them according to a rule designed for the target weight e^{-S}, where S is the declared Euclidean action. After equilibration, and subject to sufficient exploration, repeated updates sample that model distribution. A finite run still needs checks for dependence and regions it failed to explore.

The Monte Carlo wanders for you. The update rule is designed to preserve the target distribution. A finite run can still spend too long in one part of the space, which is why I checked how quickly the chain forgot its earlier configurations.

There are several clocks here, and confusing them would change the claim. A lattice coordinate can represent Euclidean time. A Monte Carlo sweep is a sampling update. Gradient-flow time controls a smoothing operation. The seconds on my laptop measure computational work. A hundred sweeps are not a hundred seconds of physical evolution, and a smoothing time is not the separation used in a temporal correlation function.

For almost a week, I barely slept.

Two hours, maybe three, then awake again. Back to the code. Back to the tests. Run the simulation. Check the results. Find the bug. Fix it. Run again.

It was December. Sacramento was cold and rainy. The holidays were approaching, but I barely noticed. I was deep in the code, in the physics, in the geometry.

Teddy kept me company. He would lie on his bed in the corner of my home office, occasionally looking up when I muttered at the screen. Sometimes I would talk to him, explaining what I was trying to do.

"The Cabibbo-Marinari algorithm, Teddy. It breaks the SU(3) update into three SU(2) subgroups. Each subgroup can be updated exactly using a heatbath method. It is clever. Kennedy and Pendleton figured out the SU(2) part in 1985."

He wagged his tail. He did not understand the physics, but he understood that I was working on something important. Dogs know these things.

Lots of failure. Lots of dead ends.

The topological charge calculation was wrong—off by factors of π. The standard formula involves the Levi-Civita tensor, and I had messed up the signs. It took two days to find the bug.

The Wilson loops were computing but giving nonsense values—I had the path ordering backwards. In gauge theory, the order matters. $U_1 U_2$ is not the same as $U_2 U_1$. The matrices do not commute. I was multiplying in the wrong direction, and the traces came out wrong.

The thermalization was not long enough—the system had not equilibrated, and I was sampling from the wrong distribution. The first few hundred configurations were correlated with the cold start. I had to throw them away and run longer.

Each bug took hours to find, sometimes days. The code would run, produce numbers, and the numbers would be wrong. No error messages. No crashes. Just wrong physics. You have to know what the answer should be to know you got it wrong.

Moments where I asked myself: *Is this the perfect time to give up?*

The pace of that week can make the work sound effortless in retrospect. It was not. Each calculation exposed something else I needed to understand. Let me tell you about learning quantum chromodynamics in a week.

It cannot be done. The subject is too vast, too deep, too technical. Graduate students spend years mastering it. There are textbooks that run to a thousand pages. There are decades of papers, each building on the last.

But you do not need to learn all of QCD to solve a specific problem. You need to learn exactly what you need, no more. And AI changes what is possible.

I had Claude. I had ChatGPT. I had access to knowledge distilled and ready to deploy. I could ask questions and get answers instantly. Not always correct answers—the models make mistakes, especially on technical topics. But the answers pointed me in the right direction, gave me keywords to search, papers to read, concepts to understand.

"What is the Cabibbo-Marinari algorithm?" I would ask. And get an explanation. Not perfect, but enough to start.

The Cabibbo–Marinari method updates selected SU(2) subgroups inside SU(3). It makes the conditional sampling practical by working through smaller matrix updates. It is like tackling a complicated adjustment one accessible control at a time. One pass through those controls does not

guarantee that the entire lattice has equilibrated; the sampling diagnostics still have to establish what the run did.

"How do you compute topological charge on the lattice?" And get three methods: the naive plaquette sum (which does not work), the cooling method (which works but is expensive), and the gradient flow method (which is the modern standard).

Gradient flow smooths a field by following a downhill direction of its action. Short-distance variation is reduced as the flow parameter increases. At finite lattice spacing, the measured charge can depend on the discretization, flow duration and rounding rule; excessive smoothing can change relevant structures. I therefore have to retain the smoothing convention alongside the reported charge label.

El flujo revela lo que el ruido esconde.

The flow reveals what the noise hides.

"What is the standard value of β for SU(3) lattice QCD?" And learn about $\beta = 6.0$, the lattice spacing around 0.1 fermi, the approach to the continuum limit.

I was not learning QCD. I was learning exactly enough QCD to do what I needed. The minimum viable understanding. Enough to write code, run simulations, check results.

This is a new kind of learning. Not mastery. Not expertise. Targeted competence. Knowing exactly what you need to know, and nothing more.

Some would call this shallow. I call it efficient. I did not have years. I had weeks. And the geometry was waiting.

Here is something I learned: debugging physics code is its own form of discovery.

When the code is wrong, you have to understand the physics deeply enough to know *why* it is wrong. You cannot just match output to expected values. You have to understand what the output should be, and why, and what could make it different.

The topological charge bug taught me about the Pontryagin index—the mathematical name for the winding number of a gauge field. Think of it like this: if you have a rubber band looped around a pole, you can count how many times it winds around. That count is an integer—you cannot have half a loop. The Pontryagin index counts how many times the gauge field "winds around" in its internal color space. It is quantized for the same reason the rubber band count is quantized: topology is discrete.

The formula involves a product of field strength tensors contracted with the Levi-Civita symbol:

$$Q = \frac{1}{32\pi^2} \int d^4x \, \epsilon^{\mu\nu\rho\sigma} \mathrm{tr}(F_{\mu\nu} F_{\rho\sigma}) \tag{5.3}$$

On the lattice, this becomes a sum over plaquettes—the smallest squares in the grid. But the discretization is tricky. The naive sum does not converge to the continuum value. You need to use smeared fields, or cooling, or gradient flow.

I was getting non-integer values and treated them as a warning to inspect the calculation. A raw finite-spacing estimator can legitimately return 0.7 or 1.3; those values alone do not prove a bug. The useful checks are the normalization, orientation, discretization and behavior under smoothing or refinement. Rounding creates an integer label but does not establish that the underlying estimator has reached its continuum limit.

I switched to the clover definition of the field strength tensor—a way of computing the curvature that uses a symmetric combination of plaquettes, shaped like a four-leaf clover, to get a better

average. Better—the values were closer to integers. I added APE smearing—a technique that smooths the gauge field by averaging each link with its neighbors, like applying a gentle blur to a photograph to remove pixel noise while keeping the essential shapes. Better still. Now the values rounded cleanly to integers.

Each bug taught me something. Each fix deepened my understanding. By the time the code worked, I understood the physics far better than I had when I started.

El error es el maestro. El bug es la lección.

The error is the teacher. The bug is the lesson.

And then, suddenly, the numbers made sense.

The Wilson-loop values and reported action diagnostic were in the ranges I expected under the implementation's conventions. The charge labels were integer-valued after the chosen measurement and labeling procedure. The autocorrelation diagnostics were encouraging for the observables checked. These were reasons to continue examining the run, rather than a guarantee that every observable had equilibrated.

I ran a test: compute the string tension from Wilson loops of different sizes. If the code was right, the loops should satisfy an area law—the logarithm of the loop should be proportional to its area. I fitted the data.

$\sigma a^2 = 0.041$

The string tension times the lattice spacing squared. Look it up in the lattice QCD literature. At $\beta = 6.0$, the value should be around 0.04-0.05.

I was in the right ballpark.

The relief was physical. My shoulders dropped. I had not realized how tense I had been holding them. Days of uncertainty, and finally a validation. The code was not garbage. The code was doing real physics.

Now I could run the Davis-Wilson analysis.

Once the basic checks made sense, I could say exactly which finite experiment I was looking at. Here is what I ran:

The lattice had eight sites along each of four axes. That gave me a box small enough to run the first analysis and change the measurements while I was still learning what to look for.

I worked at the bare coupling setting $\beta = 6.0$. Fixing it gave me a place to start comparing configurations in the same finite lattice.

I looked at 200 saved configurations. A heatbath sweep updates the lattice links using the simulation's sampling rule. Successive saved configurations came from an evolving Monte Carlo chain. How much they had in common depended on the observable I measured.

I built features from selected Wilson loops and kept a rounded charge label with each configuration. A Wilson loop traces a closed path along links. The calculation around one elementary square, which I work through below, supplies the smallest example.

Total compute time: about 15 minutes on a Modal A100 GPU. Total cost: maybe $2.

Two hundred points in a high-dimensional space. Each point a gauge field configuration. Each configuration was one field sampled from the simulation.

I plotted the points. I looked for structure.

I found the gap.

I kept pressing forward. I do not know why. No one was going to see these results and validate me. No one was waiting. I was alone in my living room with a laptop and a dog, spending money I did not have on a problem that had defeated the best minds in physics for half a century.

The rational thing would have been to stop. To work on something achievable. To publish small papers on incremental problems. To build a normal career.

But I could not stop. The geometry was speaking. I had to listen.

La obsesión es el precio de la comprensión.

Obsession is the price of understanding.

§ 5.3

The empty interval on the screen

It was 3:00 AM in early December 2025.

I had been running on fumes. Days without proper sleep. Meals forgotten, or eaten standing at the counter because sitting down felt like wasted time. The code had consumed me. Every bug fixed revealed two more. Every successful run raised new questions. The simulations were expensive, and I was watching my bank account drain toward zero.

My body was exhausted. My mind was sharp—too sharp, the kind of sharp you get when you have pushed past the fatigue into some other state where everything is clear and nothing else exists.

I had been running a particular test—computing the Davis-Wilson map for 200 configurations at $\beta = 6.0$ on an 8^4 lattice. The simulation had finished. The data was ready. I was plotting the distribution of configurations in cache space.

The cache space is the image of the Davis-Wilson map. Each configuration becomes a point—a vector of Wilson loop traces plus a topological charge. Two hundred configurations become two hundred points in a high-dimensional space.

I projected the points onto the radial distance from the centroid. This is a simple visualization: collapse the high-dimensional structure to a single number, the distance from the center.

I saw it.

I did not celebrate. I did not shout. I sat very still and looked at the screen.

The room was silent except for the hum of my laptop fan. The light from the monitor was the only light. Teddy was asleep, his breathing slow and steady. And on the screen, in purple dots on a white background, was a gap I wanted to understand.

The points clustered. Most of the radii fell into a band. Farther out there was a long stretch of the plot with no dots, and then another point. I could tell the width of that empty stretch from the plot. I could not tell its cause.

A void. A gap. A forbidden region.

My first thought was: I must have made a mistake.

This is the correct first thought. When you see something no one else has seen, the most likely explanation is that you are wrong. I have learned this from years of debugging. The bug is always in your code, not in the laws of physics.

I checked the code. Line by line. I ran it again. Same result. I changed the random seed. Same result. I changed the lattice size. Same result. I changed β. Same result.

The gap persisted.

I got up from my chair. Walked to the kitchen. Made tea. Came back. Looked at the plot again.

Still there. The void was still there.

Let me explain what I was seeing.

Think of rain falling across a parking lot. If you inspect the pavement after only two hundred drops, a dry strip may have a roof over it, or you may have arrived too soon. You need to know how the rain is falling before you can tell. My points came from a Monte Carlo chain with Boltzmann weights, and projecting them onto one radius changed their distribution again. Even a uniform cloud in a ball puts more of its volume near the outside than near the middle.

The blank strip gave me a question I could take back to the data. Would it reappear with other samples, other measurements, or a different normalization of the Wilson loops? An empty interval in one plot is an observation. An interval the theory cannot occupy is a claim about all configurations. I wanted to know what it would take to get from one to the other.

La geometría prohíbe. La física obedece.

Geometry forbids. Physics obeys.

The empty interval immediately made me think of an energy gap. I need to keep the two rulers visible here, because the picture alone does not make them the same. Picture a castle with a moat. There is the ground the castle stands on, and then there is the first place outside where you can set your foot. For a quantum Hamiltonian, the castle is the vacuum energy. The first available energy above it is the edge of the excitation spectrum. The distance between the two is measured in energy.

The radius on my plot measured something else. I had put the origin at the sample centroid, a point computed by averaging Wilson-loop features. A vacuum is a quantum state. An average address in a feature space can be useful even when no configuration sits at that address.

I divided the largest observed spacing in the sorted radii by the median spacing. That night the number I wrote down was 42. The discovery report gives 41.97. Later, another run would give 85.46; I work through its 200 saved radii below. They are two runs, and I want to keep the number that belonged to that night in its place.[1]

Forty-two. The answer to life, the universe, and everything. At 3:00 AM I laughed out loud. Teddy looked up, confused. Try explaining to a dog that Douglas Adams has just turned up in a spreadsheet.

The radius was only one view of the saved configurations. I also had a charge label, which let me look at the same points another way. The plot revealed more structure.

I colored the points by the rounded lattice charge r. Blue for $r = 0$, orange for $r = +1$, red for $r = -1$. The sign convention names the direction of the winding. On a finite lattice the label depends on how the charge is measured and on any smoothing applied before rounding. In a smooth continuum setting with suitable boundary conditions, a continuous deformation through admissible fields preserves the integer charge.

[1]The first Davis–Wilson research deposit reports 41.97 for this diagnostic. The retained 200-radius array reproduces the later result, not the earlier one.

Now I could ask a better question. Did the colors change across the largest radial gap? The retained 200-configuration run has 169 charge-zero points, 19 with charge -1 and 12 with charge $+1$. Its farthest point is blue. The big gap in those radii separates one outlying measurement from the rest of the sample, including other blue points.

I had expected the colored points to look like books on different shelves. What the colors gave me was an extra coordinate. Two configurations can carry the same integer winding and still look very different to the Wilson loops. Two configurations with different winding can have similar radial distances from the centroid. If I want to study separation by topology, I have to compare the charge labels and the full feature vectors together. A radius keeps only one number from each vector.

In most data analysis, empty space is uninteresting. It is the absence of data. You focus on where the points are, not where they are not.

In this analysis, the empty space gave me a lead. I had spent years looking for signal in data. Looking for patterns, clusters, trends. Here I had a stretch with no points and a way to ask whether that absence meant anything. The next step was to write down exactly which measurement produced it.

El vacío habla. Solo hay que saber escuchar.

The void speaks. You only need to know how to listen.

To connect the picture with the field theory, I returned to the quantity used to weight configurations: the Euclidean action. Here is why I looked at the action. In Euclidean Yang–Mills theory it is built from the integral of squared field strength. With a positive trace convention,

$$S_E[A] = \frac{1}{4g^2} \int \operatorname{tr}(F_{\mu\nu} F_{\mu\nu})\, d^4x, \tag{5.4}$$

where $F_{\mu\nu}$ is the curvature of the gauge connection. For a small loop, the holonomy responds to curvature along the surface it encloses. The action adds contributions across all four dimensions. Two fields can put that curvature in different places and have the same integral. Even flat connections can have distinct holonomy around noncontractible loops.

The Master Lemma is the proposed bridge in the research program. One way to state the numerical question behind it is to fix a cache resolution and ask for a separation inequality

$$\left|\overline{S}_E(B) - \overline{S}_E(B')\right| \geq \kappa \tag{5.5}$$

over a specified set of neighboring bins B, B'. The bar denotes a bin mean. Such a statement requires hypotheses on the representation, the ensemble and the chosen neighbors. A nonzero difference for a finite list of bin means need not be uniform as the ensemble, resolution or lattice changes.

I tried this on 100 thermalized SU(3) configurations at $\beta = 6.0$ on an 8^4 lattice. Every sampled configuration had $r = 0$. I compared mean unflowed actions in neighboring feature bins, requiring at least three occupants per bin and looking at up to two nearest other bin centroids. The smallest absolute difference was $\kappa_{\text{adj}} = 7.6817313$ in that run's action normalization. Imagine comparing

average fuel use for neighboring routes. The comparison can detect a difference between those groups. It cannot tell you the least fuel any possible journey requires.[2]

§ 5.4

Putting the observation under pressure

A good plot can make a bad argument feel inevitable. I wanted tests that could spoil my favorite explanation. If a change of gauge moved the points, the coordinates were wrong. If changing the sample undid the gap, the plot had told me more about that sample than about Yang–Mills.

I used eleven checks to ask where the signal came from. Several were checks of the same data or pipeline, so each tells us what happened under a particular change. Together they show what was measured and what would still need a separate calculation.

Check 1: Measure the Empty Interval. I wanted to know how much of the shape in the plot came from one point. The retained run lets me sort the radii, calculate the spacings, and repeat that operation with the outermost radius removed. The replay below keeps the two records of the discovery separate and prints the rows behind the result.

The questions grew as the work continued. I started with the 200 radii, then asked related questions of other ensembles, spectral comparisons and the later TVR calculations. I will keep those changes of sample visible as I go. Learning something about one set of fields gives me a reason to try the next test; it does not give me the next answer.

If a fixed interval has probability p under a specified sampling model, the chance that N independent observations miss it is $(1-p)^N$. My interval was selected by searching for the widest gap, and the configurations came from a chain. A significance calculation would have to account for the search and the sampling dependence.

Check 2: Gauge Invariance. I tested this by changing the local color frames and recalculating the features. A link matrix changes when I give the frames at its endpoints different coordinates. The loop trace survives. I can prove that directly by multiplying the changed links around one square. It is the first calculation in the investigation below.

Check 3: Symmetry (The Mirror Test). Yang-Mills theory has a symmetry: if you flip the sign of the topological angle θ, certain observables should flip sign while others remain unchanged.

A physical theta term and an offline real exponential tilt are different constructions. A sign-reversal identity for the latter follows only when a symmetry of the baseline measure reverses both recorded observables. Chapter 8 states that assumption explicitly. A mirrored reweighting calculation is not by itself a test of physical CP violation.

I compared the two signs of the bias. The earlier direct-comparison record reported approximate sign reversal. In the separately recovered TVR-003 archive, the weighted means at -1.15 and $+1.15$ are 27.42851 and -26.37401. Their magnitudes differ by about 3.84% relative to the first. They come from the same rows; Chapter 8 uses paired uncertainty rather than treating them as

[2]The recorded `A4C2-001` run reports 100 thermalization sweeps with Cabibbo–Marinari heatbath and overrelaxation, unflowed actions, feature-bin thresholds 0.15 and 0.20, minimum occupancy three, and up to two nearest other bin centroids for each occupied bin.

independent runs. Exact oddness is a statement about a symmetric measure, not something a finite archive must satisfy exactly.

El espejo no miente.

The mirror gives me a specific comparison. In a finite sample I measure how closely the two estimates reverse, with their shared uncertainty, before drawing a conclusion about the underlying symmetry.

Check 4: Stability Over Time. Monte Carlo simulations evolve over time. If the signal is real physics, it should be stable. If it is a transient artifact, it should drift.

Think of it like tuning in a radio station. If you are on the right frequency, the signal is steady. If you are slightly off, it drifts in and out. The stability tells you whether you have found something real.

I divided the simulation into 20 time blocks and computed the correlation between J and the Davis term D in each block. The correlation remained stable: mean 0.83, standard deviation 0.05, never dropping below 0.70.

The reported block correlations describe persistence within that record. A persistent correlation can occur even before equilibrium, or in a sampler with a systematic error. Equilibration needs checks of the generating chain and its target distribution. These J, D checks belong to the later TVR analysis, not to the separate 200-radius record.

Check 5: Volume Independence. If the signal is a finite-volume artifact, it should change dramatically when you change the lattice size. If it is real physics, it should scale appropriately.

Finite-volume effects are common in lattice simulations. The lattice is a box, and the box has walls. Particles can wrap around the box, interfere with themselves, create spurious effects that disappear in larger volumes.

To ask what the box itself was doing, I needed to change its size. The earlier account describes runs at $L = 8, 10, 12, 14$ and a signal density that stayed constant within the stated errors. I have that account, but not the matching raw runs to check it here. What I can recalculate is a different comparison: eight TVR batches, all at $L = 12$. That tells me how the batches differ at one size. It cannot tell me what happens when I make the box larger.

This is a finite-volume check on the reported signal density. An infinite-volume spectral estimate needs its own bound, with constants that stay controlled as the box grows.

Check 6: Geometric Robustness. A good signal should be robust to noise, but not invulnerable.

If you add a tiny bit of noise and the signal disappears, it was probably not real—just a statistical fluctuation that happened to look significant. But if you add massive noise and the signal persists unchanged, something is wrong too—you are probably measuring an artifact of your method rather than the underlying physics.

Real geometric structure behaves predictably under noise. It degrades gracefully.

I added Gaussian noise to the Wilson loop measurements at varying levels: $\sigma = 0.25, 0.5, 1.0$. I measured the correlation between the noisy Davis term and the clean reference.

The correlation decayed gracefully: 0.87 at $\sigma = 0.25$, 0.74 at $\sigma = 0.5$, 0.60 at $\sigma = 1.0$. Small noise leaves it mostly intact. Large noise obscures it but does not eliminate it completely.

The signal is not fragile. It is not a knife-edge coincidence that disappears at the slightest perturbation. But it is also not indestructible. That is the behavior I hoped to see, and I have to be honest about how much it tells me. Adding noise to any fixed signal lowers its correlation with

the clean version smoothly. The decay tells me the pipeline is behaving sensibly. It does not tell me which structure the pipeline found.

Check 7: β Scaling. At $\beta = 5.8, 6.0, 6.2$, I compared the numerical diagnostics at different bare couplings. A finite scan can show whether a feature survives the settings tried. A continuum mass estimate calls for a spectral decay rate, an independent scale setting for the lattice spacing a, and control of the limit. If a decay rate is measured in inverse lattice time steps, converting it to a mass in natural units requires dividing by the temporal spacing. A spacing between cache points, or between mean Euclidean actions, has no such conversion.

Check 8: Transfer Matrix. For spectroscopy on a lattice with a positive transfer matrix, take a gauge-invariant observable O, subtract its vacuum expectation, and examine its connected correlator with the adjoint insertion at increasing Euclidean time t. For vacuum expectation values at finite spatial volume, its spectral decomposition has the form

$$C_O(t) = \sum_{n>0} |\langle n|O|0\rangle|^2 e^{-(E_n - E_0)t}.$$

The slowest exponential with nonzero overlap gives the lowest gap visible to O. A guitar is a decent picture for that last detail. Pluck it where a certain mode has a node and you may never hear that mode. In the lattice calculation it is the temporal decay, together with the channel selected by O, that gives the energy information.

The action values in this ensemble gave me a spread I could calculate: population variance 62231.1301433, with square root 249.4616807. That tells me how much the total Euclidean action varies from one field configuration to another. It does not yet give me the time curve I need for spectroscopy. There are two different clocks to keep straight here. Flow time controls smoothing; the time in $C_O(t)$ separates two insertions of an observable. This calculation used unflowed actions and measured their spread. The value 7.68 answers a third question, about differences between unflowed bin means.[3]

Check 9: Autocorrelation. Successive Monte Carlo measurements can be correlated. Correlation does not make a sample mean meaningless, but treating correlated values as independent can badly misstate its uncertainty. The relevant dependence also varies with the observable. A chain may forget its action quickly while its topological charge changes slowly.

This is a subtle issue. The Monte Carlo process generates a sequence of configurations, each derived from the previous one by small updates. Adjacent configurations in the sequence are similar. You cannot treat them as independent data points.

The autocorrelation function measures this correlation. It tells you how long you need to wait between samples to get effectively independent measurements.

I computed the autocorrelation function—a measure of how long the Monte Carlo "remembers" where it was. If you shuffle a deck of cards once, the order is still mostly the same. Shuffle ten times, and it starts to look random. Shuffle a hundred times, and the original order is forgotten. The autocorrelation function tells you how many "shuffles" you need before the system forgets its past.

[3]Bee Rosa Davis, validation run v17, 19 December 2025, diagnostic `OSBRIDGE-001`. Although the test is labelled $t_{\text{flow}} = 0.1$, the production configuration leaves the flowed array empty and passes the unflowed actions to the test. The saved JSON records both diagnostics.

The earlier diagnostic showed decay by lag 10 and an integrated autocorrelation time of $\tau = 0.60$, with five sweeps between saved configurations. I cannot carry that waiting time over to every quantity I measure, or to another run. For the 1,600 TVR rows, the available code and acceptance counts point to a different update schedule. I need to check how that sample remembers its past before deciding how much independent information it contains.[4]

A repeat with independent chains or a block resampling analysis would tell me how stable the radial statistic is under the sampling dependence.

Check 10: Topological Basins. The Davis term D—a derived quantity from the Davis-Wilson map—should have a specific distribution if the theory is correct. In a single topological basin, D should be approximately Gaussian.

A sum of many contributions can approach a Gaussian distribution under suitable dependence and moment conditions. Gauge-field contributions are correlated, so I cannot invoke the central limit theorem merely by counting summands. A departure from Gaussian shape is a diagnostic to investigate; it need not mean that the code is wrong.

I examined the distribution of D and its *excess kurtosis*: the standardized fourth central moment minus three. A Gaussian has excess kurtosis zero. Because fourth powers amplify large deviations, the statistic is sensitive to tails and outliers. It is not a unique description of how peaked a histogram looks, and a value near zero does not establish Gaussianity.

The observed kurtosis: 0.02.

The reported excess kurtosis was close to the Gaussian value. Several different distributions can have that same fourth moment. The shape of the full histogram and the mixing of the charge labels give additional checks.

Check 11: The Phantom Zone. The final test probes the most exotic prediction: the existence of a critical point in the θ parameter.

Yang-Mills theory depends on a parameter θ that couples to the topological charge. At $\theta = 0$, the vacuum is simple. But as you increase θ, the vacuum structure changes. At certain values, there are phase transitions—sudden changes in the properties of the ground state.

The susceptibility χ measures fluctuations. It is like the heat capacity in thermodynamics—it peaks at phase transitions. If there is a critical point, the susceptibility should show a peak.

I scanned χ as a function of θ. The scan covered the range from $\theta = 0$ to $\theta = \pi$.

The historical account placed a peak near $\theta \approx 1.15$ and called it the "Phantom Zone". The recovered finite susceptibility calculation places its peak nearer $+1.36$ under its own normalization. Neither is the archived TVR operating choice -1.15. These are different calculations. A finite-sample peak supplies a place to investigate; establishing a phase transition would require controlled volume dependence and a justified relation to the physical parameter.

La zona fantasma existe.

The Phantom Zone is where I went looking for a change.

Taken together, these checks sharpened the question rather than replacing it. I had to separate the different quantities that the word gap was being asked to describe. The tests gave me measurements I could distinguish and repeat. A radial gap lives in the observed feature spacings. An action difference lives in the bin means. A spectral gap lives in the operator that evolves

[4]The harvest source matching the 1,600-row archive and its acceptance counts point to ten sweeps with three hits per link. The earlier five-sweep settings have not been authenticated for this archive. Correlation time also depends on the observable.

quantum states in time. The first two can guide the search for the third. For that, I need to know how the geometric construction reaches the operator.

The checks leave me with three things to connect carefully: an invariant measurement, a spacing in the saved points, and a possible statement about energy. I will follow them in that order.

§ 5.5

What survives a change of gauge

Before I can trust a blank interval in a plot, I need to know what was measured at each point. Start with one square of the lattice. Its corners are p_0 at the lower left, p_1 at the lower right, p_2 at the upper right, and p_3 at the upper left. Walk around the square in that order, returning to p_0. That elementary square with a specified direction is a *plaquette*.

Each side carries a matrix. Write U_{ij} for the link on the directed edge $p_i \to p_j$ and $U_{ji} = U_{ij}^{-1}$ for the reverse edge. I use the convention in which U_{ij} expresses a color vector from p_j in the frame at p_i. With that convention I write the product in the order of the edges along the walk:

$$P = U_{01}U_{12}U_{23}U_{30}, \qquad W = \frac{1}{3}\operatorname{tr} P.$$

The product P is the loop matrix in the frame at the starting corner. Its trace is the sum of its diagonal entries. Here I divide by three, the number of color components. The same argument works for the unnormalized trace. Keeping the order matters because two link matrices can fail to commute.

The square gives us a direct check: change the coordinates at each corner and follow which factors cancel. At each corner choose a change of frame $g_i \in \mathrm{SU}(3)$, the group of unitary three-by-three matrices with determinant one. If the old coordinates at p_j were c, the new coordinates are $g_j c$. Recover c with g_j^{-1}, apply the old link, and change to the new frame at p_i with g_i. That gives $U'_{ij} = g_i U_{ij} g_j^{-1}$. The two frame matrices belong to opposite ends of the edge.

Proposition 5.1 (The trace of a closed loop is gauge invariant)**.** *Under $U_{ij} \mapsto g_i U_{ij} g_j^{-1}$, the loop product changes by conjugation at its starting site. Its trace stays fixed.*

Proof. Substitute the four changed links in the same order:

$$\begin{aligned} P' &= (g_0 U_{01} g_1^{-1})(g_1 U_{12} g_2^{-1})(g_2 U_{23} g_3^{-1})(g_3 U_{30} g_0^{-1}) \\ &= g_0 U_{01} \underbrace{(g_1^{-1} g_1)}_{I} U_{12} \underbrace{(g_2^{-1} g_2)}_{I} U_{23} \underbrace{(g_3^{-1} g_3)}_{I} U_{30} g_0^{-1} \\ &= g_0 P g_0^{-1}. \end{aligned}$$

Each interior factor cancels against the one next to it. The two factors at p_0 enclose the product. The trace satisfies $\operatorname{tr}(XY) = \operatorname{tr}(YX)$, since both sides sum $X_{ij}Y_{ji}$ over i and j. Thus $\operatorname{tr}(g_0 P g_0^{-1}) = \operatorname{tr}(P g_0^{-1} g_0) = \operatorname{tr} P$. For any longer closed walk the same cancellation takes place at each successive shared endpoint. □

On an open path the uncancelled factors are at two different sites. Closing the path is what leaves one change of frame and its inverse at the same corner.

Take the following matrices and put A, B, A^{-1}, B^{-1} on the four edges in order:

$$A = \operatorname{diag}(-1, -1, 1), \qquad B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

Both matrices have orthonormal columns and determinant one. They therefore belong to SU(3). Conjugating B by A changes the signs of the two off-diagonal entries, so $ABA^{-1} = B^{-1}$. The loop product is $B^{-2} = \operatorname{diag}(-1, 1, -1)$. Its trace is -1.

Now change the frames at p_0 and p_1, taking $g_1 = A$ and

$$g_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad g_2 = g_3 = I.$$

The new links are g_0, AB, A^{-1}, and $B^{-1}g_0^{-1}$. Their product is $g_0 P g_0^{-1} = \operatorname{diag}(-1, -1, 1)$. The changed frame swaps the last two diagonal positions of P. The original and changed matrices both give $W = -1/3$. Here I can see the entries move and check the unchanged measurement by hand.

The trace has survived the change of frame. The next question is what happens when I collect traces into the feature vector used by the plot. The calculation uses the real and imaginary parts of a specified list of loop traces as coordinates. Call the resulting feature vector x_i for configuration i. Proposition 5.1 leaves x_i unchanged when I change the local frames. It also leaves unchanged the output of any fixed processing of x_i.

For the radial analysis, first scale the feature coordinates. Write this as $z_i = D^{-1}(x_i - \mu)$, where μ holds the centering values and the diagonal entries of D hold the scales actually used. Then take the principal component coordinates $v_i = V^{\mathsf{T}}(z_i - \bar{z})$, where the columns of V give the retained directions and $\bar{z}$ is the fitted center in the scaled coordinates. The radius is $u_i = \|v_i - \bar{v}\|_2$, with $\bar{v}$ the center of the projected sample. With the processing fixed, a gauge change leaves every term in this calculation unchanged. The radius retains only distance from the center. Two projected points at $(1, 0)$ and $(0, 1)$ relative to that center give the same radius.

A refit after removing a point can change the scales, the projected directions, and the center, so even radii of configurations still present can move. The replay starts with the radii saved from the original fit.

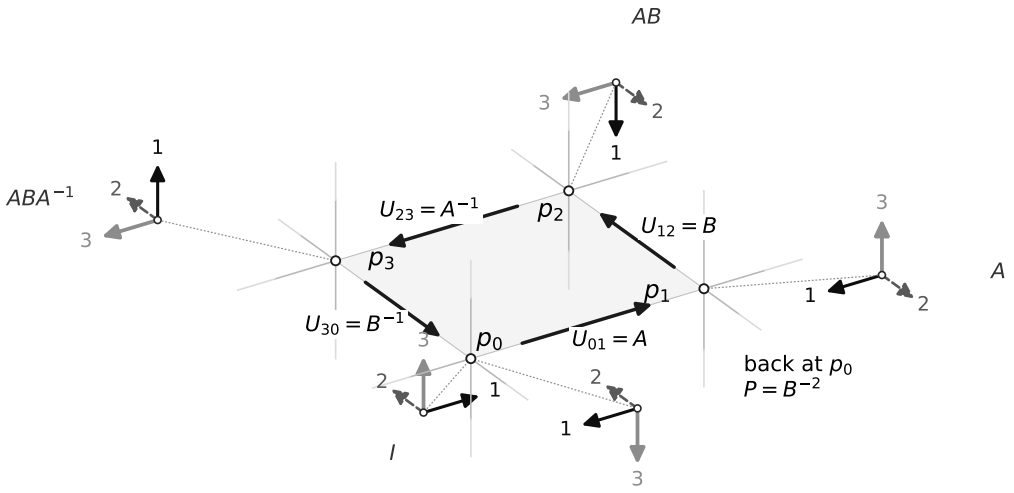
THE GEOMETRY OF FUEL LAT / 01

The loop and its local frames

One plaquette / exact SU(3) example from the text / all matrices are rotations

01 PLAQUETTE WALK / CARRIED FRAME

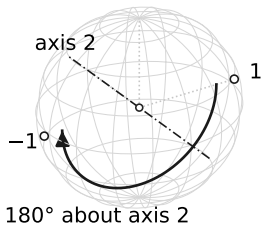
Cumulative product in the frame at p_0: I → A → AB → ABA^{-1} → P



1: solid black 2: dashed 3: heavy grey

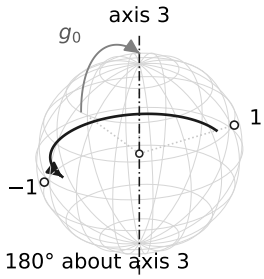
02 HOLONOMY AS A ROTATION

P turns axis 1 into −1 about axis 2



03 CHANGE OF FRAME AT P_0

g_0 carries axis 2 onto axis 3



04 SAME ANGLE, SAME TRACE

tr R = 1 + 2 cos θ for any rotation R

g_0	A	g_1^{-1}	g_1	B	g_2^{-1}	g_2	A^{-1}	g_3^{-1}	g_3	B^{-1}	g_0^{-1}
		I			I			I			

$\mathrm{tr}\, P = \mathrm{tr}\, P' = -1$

$W = \mathrm{tr}\, P / 3 = -1/3$

MATRIX	AXIS	ANGLE	tr
A	axis 3	180°	−1
B	axis 2	+90°	+1
g_0	axis 1	+90°	+1
$P = U_{01}U_{12}U_{23}U_{30}$	axis 2	180°	−1
$P' = g_0 P g_0^{-1}$	axis 3	180°	−1

A change of frame turns the rotation axis of the loop. The angle, and with it the trace, stays fixed. Lattice depth and frame placement are drawing conventions.

Figure 18. One plaquette with a frame carried around it: the triads are the cumulative products I, A, AB, ABA^{-1} in the frame at p_0, and the walk returns a half turn about axis 2. Changing the frame at p_0 moves that axis to axis 3 without changing the angle, so the trace, $1 + 2\cos 180^\circ = -1$, is unchanged. Frame placement is a drawing convention.

Replaying the 200 saved configurations

When I went back to the plot, I wanted to put my finger on the two dots that made the gap. The December 7 *First Light* report held the number I remembered, 41.97: 199 points in the main band and one at a radius of about 17. Its rounded charge counts were 7, 172 and 21 for −1, 0 and +1. I could return to that report and its histogram, but I did not have the individual radii behind it to repeat the subtraction.[5]

For the later single-β study, I had all 200 radii and their matching charge labels. Now I could do the calculation row by row. The charge counts were 19, 169 and 12 in the same order, from an SU(3) study at $\beta = 6.0$ on an 8^4 lattice. Each radius keeps its original index so I can follow it back to the same configuration and charge.[6]

Before subtracting those radii, I need to know what ruler I am holding. The analysis splits each normalized complex trace into a real and an imaginary column. It scales the columns, then keeps up to 50 principal directions of greatest sample variance. For this run, those directions account for 63.70% of the variance. Each radius is the distance from the centroid in those coordinates. The charge label stays attached to the row.[7]

Sort the 200 radii in increasing order. Write $u_{(j)}$ for the radius at rank j. Give each spacing the rank of the point on its right. Then

$$\Delta_j = u_{(j)} - u_{(j-1)}, \qquad G_{\mathrm{rad}} = \frac{\max_{2\le j\le 200} \Delta_j}{\mathrm{median}_{2\le j\le 200} \Delta_j + 10^{-10}}. \tag{5.6}$$

There are 199 spacings. The numerator measures the widest empty interval between adjacent saved points. The denominator measures an ordinary spacing on the same ruler. The retained script adds 10^{-10} to keep the division defined if the median is zero.

With the radius defined, the saved rows give me a calculation another person can repeat. I gave GIGI the saved radii and charge labels. The replay starts after the feature projection. To retain every saved decimal digit, I multiply each radius by 10^{16} and store the integer as two chunks in base 10^9. GIGI orders the rows by the larger chunk, then the smaller chunk, using the source index to break ties. It also counts the charge labels.

The replay supplies each pair of neighbors as a positive and a negative term. GIGI adds the signed chunks for each pair. For example, source indices 160 and 148 encode as (159075829, 049917080) and (178039824, 493299500). The two sums for right minus left are 18,963,995 and 443,382,420. Recombine them and divide by 10^{16} to get 1.896399544338242. After recombination, the replay stores each spacing for GIGI to rank. GIGI returns the maximum and the middle ranked spacings. The companion calculation checks every spacing against exact rational arithmetic and evaluates the final ratio. Appendix B gives the full radius order and the spacing preceding each point.

[5] Bee Rosa Davis, Davis–Wilson research deposit, version 1 (December 7, 2025), Sec. 11.2, Tables 2 and 3. The archived histogram rounds the reported ratio to 42.

[6] Davis, `analyze_single_beta.py` and `single_beta_results_full.json`. The *TVR-003 Summary Report*, Sec. 3.3, reports the approximately 85 ratio in the later Yang–Mills archive.

[7] The retained variance fraction is recorded in `single_beta_results_full.json`; the feature construction is in `analyze_single_beta.py`.

Read the last two rows of that register. Their source indices are 160 and 148. Their saved radii begin 15.90758290 and 17.80398245. Subtracting before rounding gives 1.89639954 to eight decimal places. In the ordered list of the 199 spacings the middle entry is number 100, with value 0.0221894273 to ten decimal places. Dividing gives 85.46.

Now keep the original coordinate system and take away the last radius. Every spacing between surviving neighbors stays the same. The only one that disappears is the final spacing. With 198 spacings left, the median is the average of entries 99 and 100 in their new spacing order. Those two values begin 0.0220772013 and 0.0221894273. The widest remaining spacing is the very first one, from 6.71440278 to 7.22755601. It gives the ratio 23.18. The median barely moved. The big change came from the numerator.

Reproducing the ratio is the first check. I also want to know how much one moved or missing point can change it. Suppose I keep all the rows and perturb each radius by at most ϵ. The jth sorted radius can then move by at most ϵ. At least j original values were at or below $u_{(j)}$, so at least j perturbed values are at or below $u_{(j)} + \epsilon$. For the lower bound, at least $n - j + 1$ original values were at or above $u_{(j)}$, so at least that many perturbed values are at or above $u_{(j)} - \epsilon$. This argument also covers ties.

Each spacing is a difference of two sorted radii, so it can move by at most 2ϵ. Applying the same sorted-value argument to the spacings bounds both the maximum and the median by 2ϵ. If the original maximum is M, the median is m, and $m > 2\epsilon$, the new ratio is bounded by

$$\frac{\max(0, M - 2\epsilon)}{m + 2\epsilon + 10^{-10}} \;\le\; \widetilde{G}_{\mathrm{rad}} \;\le\; \frac{M + 2\epsilon}{m - 2\epsilon + 10^{-10}}.$$

At $\epsilon = 0.001$, the bounds give 78.32 to 94.03. The last spacing leads by more than 1.38. The two largest spacings can move toward each other by at most 0.004, so the last spacing still wins.

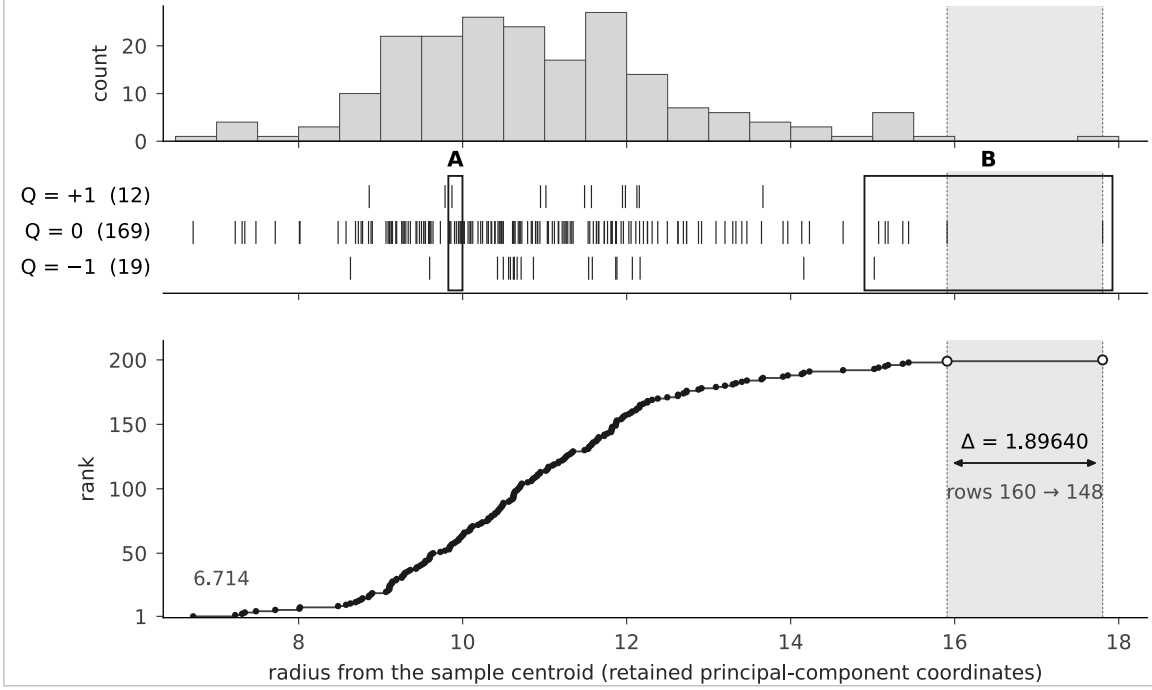
THE GEOMETRY OF FUEL LAT / 03

One ruler for all 200 radii

Later single-β study / SU(3), β = 6.0, 8^4 / every retained row on one scale

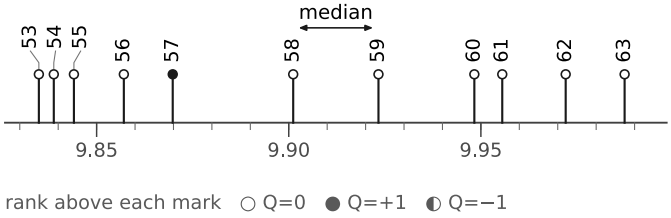
01 RADIUS SCALE / HISTOGRAM, CHARGES, RANK

Three views aligned on one radius axis. Shaded: the widest empty interval.



A AN ORDINARY SPACING, ENLARGED

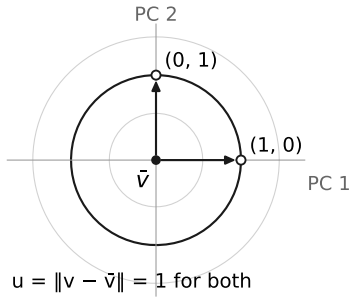
Ranks 54–63 / the median spacing is ranks 58 → 59



median Δ = 0.0221894273 (rows 19 → 172)

04 WHAT A RADIUS KEEPS

Constructed example



u = ‖v − v̄‖ = 1 for both

ROWS 200	SPACINGS 199	VARIANCE KEPT 63.70%	Q −1/0/+1: 19/169/12
MAX 1.89639954	MEDIAN 0.0221894273	RATIO 85.46	$+10^{-10}$ IN DENOMINATOR

The shaded interval is an empty stretch of retained radii. It is a distance in feature coordinates, not an energy. Directions in panel 04 are constructed; only radii are retained.

Figure 19. All 200 saved radii from the later 85.46 run on one ruler. The histogram, the three charge tracks and the rank curve share the radius axis; shading marks the widest empty interval, between source rows 160 and 148. Window A enlarges the median spacing. Panel 04 is a constructed example: the record retains radii, not directions.

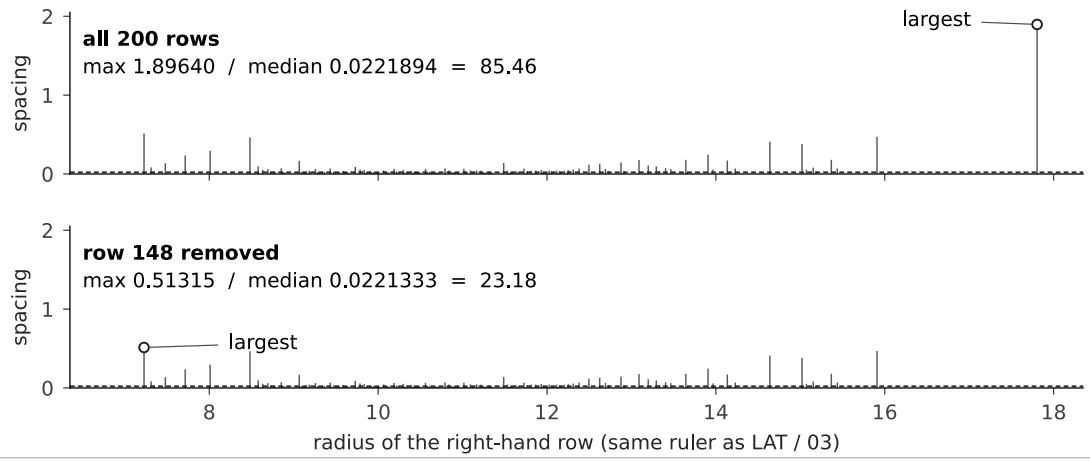
THE GEOMETRY OF FUEL LAT / 04

Read the gap back to its rows

Fixed coordinates / one row removed / bounds for small changes

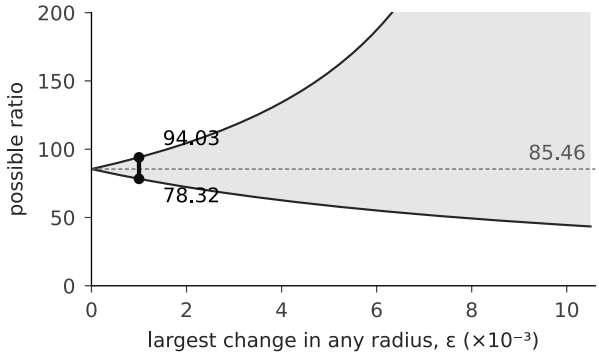
01 DELETING ONE ROW LEAVES THE RULER

Each spacing drawn at its right-hand radius. Dashed: median spacing.



02 HOW FAR CAN SMALL CHANGES MOVE IT?

Every radius moved by at most ε



03 LARGEST SIX SPACINGS

Rank pair / source rows / charges

RANKS	ROWS	Q	Δ
199 → 200	160 → 148	0/0	1.8963995
1 → 2	158 → 20	0/0	0.5131532
198 → 199	98 → 160	0/0	0.4697937
8 → 9	99 → 95	0/0	0.4645368
191 → 192	47 → 1	0/0	0.4103018
192 → 193	1 → 58	0/-1	0.3804233

04 THREE NUMBERS, THREE CALCULATIONS

41.97
First Light report
Dec 7; charges 7 / 172 / 21; raw radii not recovered

85.46
Retained 200 rows
this sheet and LAT / 03

23.18
Row 148 removed
same coordinates, no refit

Deleting the outermost row removes one spacing; the median barely moves. At ε = 0.001 the ratio stays between 78.32 and 94.03.

Figure 19. *Continued.* Each spacing drawn at its right-hand radius, with and without source row 148, on the same ruler. Removing that row leaves every other spacing fixed, and the largest becomes the first. The band shows how far the ratio can move if every radius changes by at most ε. The 41.97 First Light report remains a separate historical result.

The two records and the deletion check

Table 5.1 gives the provenance and arithmetic for the three numbers. The first column comes from the December 7 deposit. The other two use the retained JSON radii.

	First Light	Retained replay	Delete last radius
Points	200	200	199
Charge −1	7	19	19
Charge 0	172	169	168
Charge +1	21	12	12
Adjacent spacings	199	199	198
Largest spacing	—	1.89639954	0.51315322
Median spacing	—	0.0221894273	0.0221333143
Gap ratio	41.97	85.46	23.18

First Light gives the ratio and a plotted interval. Its printed formula divides by the median spacing, with no regularizer. The em dashes mark exact spacings unavailable in the located record. Both replay columns keep the same feature scaling, projection, and centroid.

Table 5.2 identifies the points on each side of the largest spacing. The source index is the position in the JSON arrays, starting from zero.

Replay	Endpoint	Rank	Source index	Radius
All 200	Left	199	160	15.90758290
All 200	Right	200	148	17.80398245
Fixed deletion	Left	1	158	6.71440278
Fixed deletion	Right	2	20	7.22755601

All four endpoint labels are $Q = 0$. The largest observed spacing moves from the high end of the full sample to the low end of the remaining sample. The charge count shows exactly what left the table.

There are two charge-zero configurations on the two sides of the biggest interval. The charge label by itself cannot explain the interval. To go further I need a measurement that can follow the state through time. First I want to show how easily a different observable can give me a different kind of gap.

§ 5.7

From feature spacings to an energy question

The empty interval tells me about distances in feature coordinates. I also wanted to compare the actions of the fields that occupied nearby parts of that picture. I kept bins with at least three occupants, found up to two nearest centroids for each bin, and compared their mean Euclidean

actions. The smallest absolute difference was 7.6817313. Two small bins show why a difference between averages answers its own question.[8]

Imagine two bins with recorded actions

$$B_1: \quad 8,\ 10,\ 12, \qquad B_2: \quad 8,\ 14,\ 20.$$

The means are 10 and 14. Both bins contain an action of 8. That shared value is the point: the averages differ, but one field in the first bin and one field in the second have exactly the same recorded action. If I want a lower bound on the difference for every pair of fields across those bins, the two fields with that shared value immediately need to be checked. A difference between averages has not supplied that bound.

Across the same ensemble, the action variance was 62231.1301433. Its square root gives the standard deviation, 249.4616807.[9] I could shuffle those actions into any order and get the same spread. To learn a decay rate, I need observations separated in Euclidean time and a record of that separation.

To reach a spectral quantity, I need time-separated observations and an evolution operator. A spread of action values alone has not supplied either one. Take a lattice theory with a positive self-adjoint transfer operator $T = e^{-a_t H}$, where $a_t > 0$ is the temporal spacing. Set the unique vacuum's energy to zero, so $T\Omega = \Omega$ and $\|\Omega\| = 1$. For an observable O defined on the vacuum, use

$$w = O\Omega - \langle \Omega, O\Omega \rangle\, \Omega, \qquad C_O(n) = \langle w, T^n w \rangle.$$

The subtraction makes w orthogonal to the vacuum. The connected correlation uses O and its adjoint, separated by n steps. Each step multiplies an energy-E component by $e^{-a_t E}$. The low-energy components fade more slowly.

On the back of an envelope, take orthonormal states e_0, e_1, e_2 with $\Omega = e_0$ and

$$T = \operatorname{diag}(1, 1/2, 1/4), \qquad w = e_1 + e_2.$$

The dimensionless energies are $\log 2$ and $\log 4$. The first correlation values are 2, 3/4, and 5/16. In general,

$$\begin{aligned} C_O(n) &= 2^{-n} + 4^{-n}, \\ \widehat{m}_{\mathrm{eff}}(n) &= \log \frac{C_O(n)}{C_O(n+1)} = \log 2 + \log \frac{1 + 2^{-n}}{1 + 2^{-(n+1)}}. \end{aligned}$$

The first two estimates are $\log(8/3)$ and $\log(12/5)$. They approach $\log 2$. Dividing the rate by a_t gives an energy in natural units.

Change the observable so it produces $w' = \sqrt{2}\, e_2$. The correlation is now $2 \cdot 4^{-n}$, with the same starting value and an effective mass of $\log 4$ at every step. Think of a vibration pattern on a string. A point that stays still in that pattern is called a node. A pluck there can fail to excite that pattern. Here the choice of observable has a similar limitation: it leaves the lowest mode

[8] This is the `A4C2-001` comparison described earlier in Chapter 5.

[9] Davis, validation run v17, 19 December 2025, diagnostic `OSBRIDGE-001`. The matching code uses the unflowed actions to compute the population variance, despite the flow-time label.

quiet, so the correlation cannot reveal that mode's energy. Figure 20 shows the two answers from the same underlying model.

The drawing starts with a circle of possible excitations in the e_1, e_2 plane. Its radius is $\sqrt{2}$, so both w and w' lie on it. Each higher section applies another step of T: the e_1 component is halved and the e_2 component is quartered. Follow one marked point and I can see which parts of its starting vector survive. The correlation compares that changed vector with the one I started with. After one step, w gives $1(1/2) + 1(1/4) = 3/4$, while w' gives $\sqrt{2}(\sqrt{2}/4) = 1/2$. Same operator, different pieces of it showing up in the answer.

There is a second way a plot can miss something important: the observable may respond only very weakly to a slow mode.

A radius is a number in feature coordinates. In a transfer calculation I instead follow an observable at successive Euclidean time separations. Here is a constructed check on what even that time record can resolve. Take a normalized correlation $C_0(n) = 2^{-n}$ at separation n lattice time steps. Add a mode with multiplier $3/4$ and weight ϵ, where $0 < \epsilon < 1$:

$$C_\epsilon(n) = (1 - \epsilon)2^{-n} + \epsilon(3/4)^n.$$

These are correlations from positive diagonal transfer matrices. On the vacuum-orthogonal subspace the second example has eigenvalues $1/2$ and $3/4$. Its measuring vector has squared amplitudes $1 - \epsilon$ and ϵ. Both curves have value one at $n = 0$. At every later integer separation,

$$0 \le C_\epsilon(n) - C_0(n) = \epsilon\big((3/4)^n - 2^{-n}\big) \le \epsilon.$$

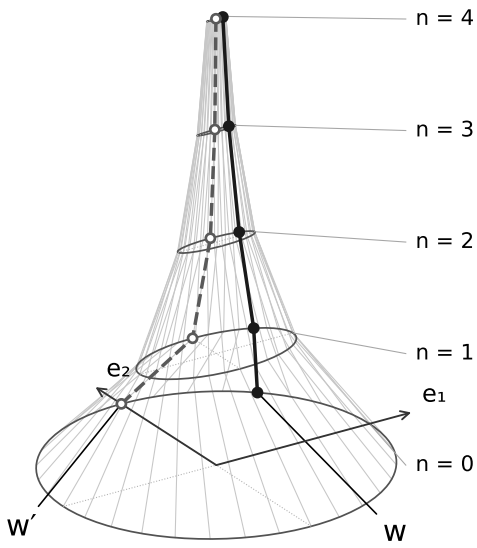
If my measurements have an absolute tolerance of 10^{-3}, both curves fit the values of C_0 within that tolerance when $\epsilon = 10^{-3}$. Yet the slower term takes over when $(3/2)^n > 999$, first at $n = 18$. Dividing $C_\epsilon(n + 1)$ by $C_\epsilon(n)$ and taking n large gives $3/4$. For C_0 the ratio is always $1/2$. The two late-time rates in lattice units are $\log(4/3)$ and $\log 2$, respectively. Converting either rate to an energy takes division by the temporal lattice spacing.

That is why I need the time window, the uncertainty, and the observable alongside a decay estimate. The radial table asks which projected points in this sample are separated. The correlation asks how much of each excitation survives each step in time.

One operator, two visible spectra

CONSTRUCTED MODEL / T = diag(1, 1/2, 1/4)

01 Follow two excitations through T



Each step: $e_1 \times 1/2$; $e_2 \times 1/4$.

Integer sections: T^n applied to a circle of radius $\sqrt{2}$ in the e_1, e_2 plane. Mesh joins matching directions.

02 Three states

State	T factor	$a_i E$
e_0	1	0
e_1	1/2	log 2
e_2	1/4	log 4

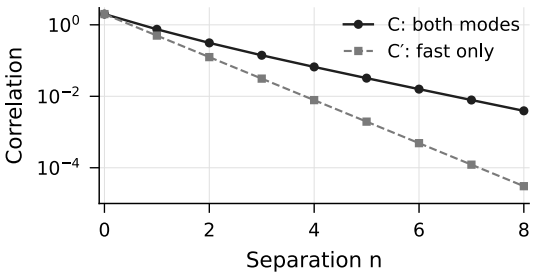
Vacuum e_0 is removed from w.

03 What the observable sees

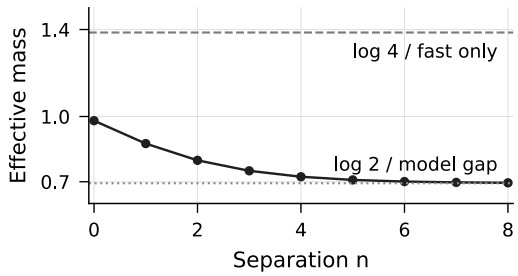
Squared weights	e_1	e_2
w	1	1
w′	0	2

Both start at C(0) = 2. Only w includes the slow mode.

04 Read each correlation



05 Take the next-step ratio



n	0	1	2
C	2	3/4	5/16
C′	2	1/2	1/8

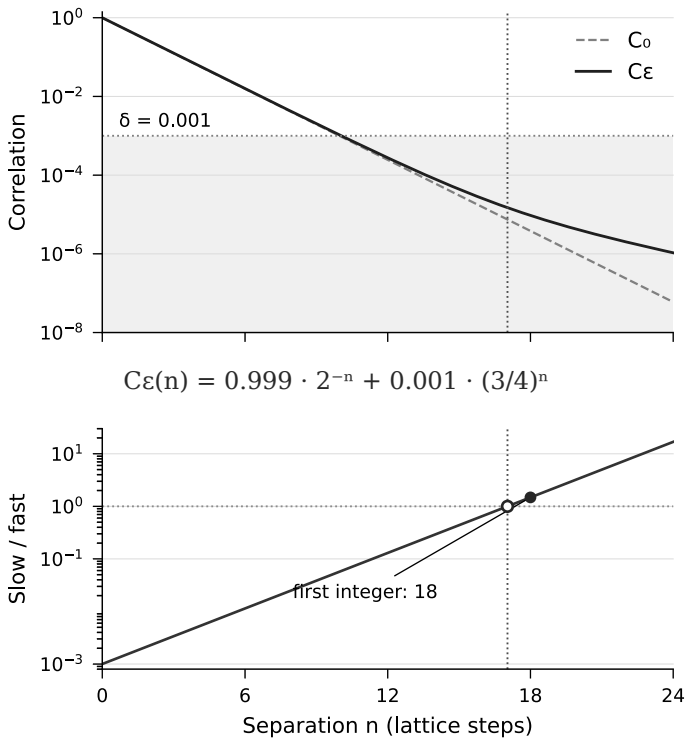
The operator stays fixed. The observable decides which modes enter the correlation.

Figure 20. Two excitations under one transfer operator. Each section is T^n applied to the same circle of radius $\sqrt{2}$ in excitation coordinates; height labels the integer step. The mesh joins matching directions. Filled markers follow w, open markers follow w'. Their squared mode weights determine the two correlations and their different effective masses.

Weak does not mean absent

CONSTRUCTED CORRELATIONS / ε = δ = 0.001

01 One separation axis, two questions



02 When slow takes over

$(3/2)^n = 999$

$n^* = 17.03415$

n	slow / fast
17	0.986248
18	1.479371

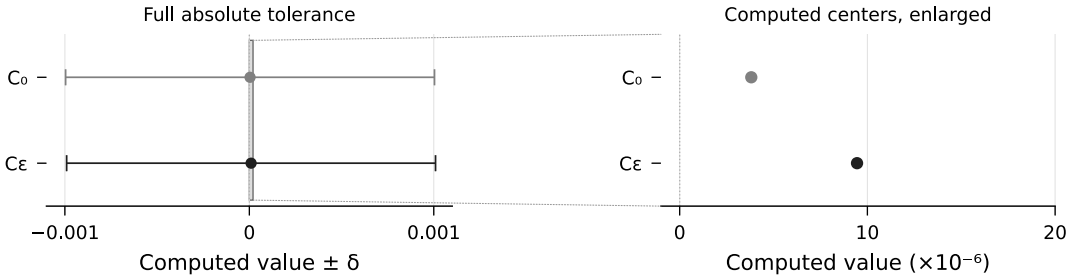
03 Read the absolute scale

At n = 18, the total correlation is

0.00000945

That is 0.9449% of δ. Both curves are below δ from integer step 11 onward.

04 The same two values at separation 18



A slow component can take over after the entire correlation falls below the tolerance.

Figure 21. The baseline $C_0(n) = 2^{-n}$ and weak-mode mixture $C_\epsilon(n)$, with $\epsilon = \delta = 10^{-3}$. The shared vertical marker locates equal components at $n \simeq 17.03$. Shading marks $C \leq \delta$. Below, the two computed values at $n = 18$ are shown with their full $\pm\delta$ intervals and then as magnified centers. These are constructed values, not sampled noise.

The example shows why the observable matters. The spectral measure lets me state the same issue without assuming that the operator has only three states. The same calculation extends beyond a finite list of eigenvalues. Let μ_w be the positive spectral measure of T seen by $w \neq 0$. Then

$$C_O(n) = \int_{[0,1]} \lambda^n \, d\mu_w(\lambda).$$

Write $\rho_w = \sup \operatorname{supp} \mu_w > 0$. The upper bound $C_O(n) \leq C_O(0)\rho_w^n$ holds for all n. For every $0 < b < \rho_w$, the measure of $(b, 1]$ is positive and $C_O(n) \geq b^n \mu_w((b, 1])$. Taking nth roots and letting b approach ρ_w gives $C_O(n)^{1/n} \to \rho_w$. Cauchy–Schwarz also gives $C_O(n+1)^2 \leq C_O(n)C_O(n+2)$, so the ratios $C_O(n+1)/C_O(n)$ increase weakly to a limit. The product of the ratios telescopes to $C_O(n)/C_O(0)$. Its nth root has the same limit as the ratios, so the limit is ρ_w.

The time curve identifies the lower edge of the chosen excitation's energy support. A gap lower bound must apply on the whole vacuum-orthogonal space.

Suppose every vector w in a dense class of the vacuum-orthogonal space has

$$\langle w, T^n w \rangle \leq M_w e^{-a_t \Delta n} \quad (n \geq 1), \qquad \Delta > 0,$$

where M_w is finite and Δ is common. The root argument puts the spectral support at or below $e^{-a_t \Delta}$. The projection onto $(e^{-a_t \Delta}, 1]$ vanishes on the dense class, hence on the whole vacuum-orthogonal space by continuity. This bounds the transfer operator and thus the physical energies.

Exponential decay in the convergent strong-coupling lattice regime gives the kind of argument to look for.[10] The four-dimensional continuum problem asks for a construction of the limiting theory and a strictly positive physical gap in it.[11]

Halve a_t and let the dimensionless gap $\widehat{m}$ fall from 0.4 to 0.2. The physical gap estimate $\widehat{m}/a_t$ stays put. If $\widehat{m}$ falls to 0.1 instead, the physical estimate falls by half. I need a lower bound in physical units as the lattice gets finer and the box grows. The original plot gave me a place to look. The time calculation tells me what to prove.

§ 5.8

What an energy-gap proof would require

The plot gave me a place to start. A proof asks for a statement that still holds for configurations my program never generated and for quantum states a catalog of configurations cannot represent individually. I wrote down the assumptions the approach would need so I could see which ones came from the construction, which ones I could test, and which ones were still carrying the weight of a conjecture.

I called these seven statements axioms in my notes. They mark the steps in the research program. Some are properties of the finite lattice construction. Others describe what I would have to establish about a limiting theory.

[10] Konrad Osterwalder and Erhard Seiler, *Gauge field theories on a lattice*, Annals of Physics 110 (1978), 440–471, DOI: 10.1016/0003-4916(78)90039-8.

[11] Arthur Jaffe and Edward Witten, Quantum Yang–Mills Theory, official problem statement.

Assumption 1: Gauge-Invariant Encoding. The Davis–Wilson map $\Gamma(A) = (\Phi, r)$ is constant on gauge orbits, provided the charge estimator is gauge invariant. Under a lattice gauge transformation, the ordered product around a closed loop is conjugated at the basepoint. Its trace is unchanged. Applying gauge transformations to sampled configurations is an implementation check of that identity.

Assumption 2: Topological Charge Quantization. For smooth continuum fields in the setting where topological charge is an integer invariant, continuous deformations through admissible fields preserve that integer. In the calculation $r = \mathrm{round}(Q)$ is an integer by construction. Smoothness, the lattice charge estimator, and its stability under changes of resolution are separate questions. One run with all $r = 0$ supplies no test of how often the sampler moves between charge sectors.

Assumption 3: Wilson Loop Continuity. Finite products of link matrices and their traces are continuous on the finite lattice. Rounding a charge estimate can introduce jumps. The loop features and the integer label therefore have to be treated according to their own rules.

Assumption 4: Finite Resolution Suffices. A bounded finite-dimensional feature space has finite covers at each positive resolution. A finite list of bins tells me how to store measurements. The stronger claim is that the chosen measurements preserve enough information to bound the energy of all vacuum-orthogonal quantum states.

Assumption 5: Curvature and Action. Wilson loops respond to the geometry of the connection, while the Euclidean action sums a quadratic curvature density. The proposed quantitative correspondence needs to specify which differences it controls. In `A4C2-001` the numerical comparison is the minimum difference of observed mean actions over the selected neighboring bins.

Assumption 6: Vacuum Isolation. The spectral question asks whether the vacuum energy is isolated from positive energies of the Hamiltonian. An empty interval of radii centered on the sample mean can help me choose a feature to examine. The spectral statement needs a comparison involving the vacuum and the Hamiltonian itself.

Assumption 7: Continuum Limit Existence. The Clay problem requires construction of a nontrivial four-dimensional continuum quantum Yang–Mills theory. A finite coupling scan provides observations at the simulated settings. Constructing the continuum theory and controlling its spectrum require limits for the physical observables.

Before the formal inequality, let me explain why "orthogonal to the vacuum" cannot be skipped. Suppose Ω is a normalized zero-energy vacuum and χ is a normalized excited eigenstate of energy $E > 0$, orthogonal to it. For $0 < p < 1$, the normalized state

$$\psi_p = \sqrt{1-p}\,\Omega + \sqrt{p}\,\chi$$

has average energy pE. I can make that average as small as I like by reducing p, even when the first allowed excitation energy is strictly positive. A small probability of observing an excitation is different from an excitation of arbitrarily small energy. The gap inequality excludes the vacuum component so it asks the second question cleanly.

The operator notation also has a job. The form domain $\mathcal{D}(H^{1/2})$ contains states for which the energy form $\|H^{1/2}\psi\|^2$ is finite. For a general unbounded Hamiltonian, writing the form this way covers more states than requiring $H\psi$ itself to exist. The spectrum may include a continuum, so the proof must handle more than a finite list of eigenvectors.

The assumptions now have a definite target. This is the inequality an energy-gap claim has to reach. Let H be a nonnegative self-adjoint Hamiltonian after subtracting the ground energy. Assume a unique normalized vacuum Ω spanning its kernel. Its spectrum has a gap of at least $\Delta > 0$ precisely when

$$\|H^{1/2}\psi\|^2 \geq \Delta \, \|\psi\|^2 \quad \text{for every } \psi \in \mathcal{D}(H^{1/2}), \quad \psi \perp \Omega. \tag{5.7}$$

By the spectral theorem, if $\operatorname{spec}(H) \subseteq \{0\} \cup [\Delta, \infty)$, the spectral measure of each such ψ is supported on energies at least Δ, giving the inequality. Conversely, if any spectral projection of H onto energies strictly between 0 and Δ were nonzero, a vector in its range would violate the inequality. It is a statement about all excitations, including any continuous spectrum.

Here is how a geometric lower bound could enter that argument. Suppose a nonnegative comparison form q_Γ can be defined on the relevant physical states using the encoding. It must satisfy two inequalities on the vacuum-orthogonal form domain,

$$\|H^{1/2}\psi\|^2 \geq c \, q_\Gamma[\psi], \qquad q_\Gamma[\psi] \geq \lambda_\Gamma \|\psi\|^2 \quad (c > 0, \ \lambda_\Gamma > 0).$$

Then $\Delta = c\lambda_\Gamma$ is a lower bound. To see why the comparison is doing work, imagine two fields that land on the same shelf. If changes within a shelf include low-energy excitations, a form using only the shelf labels could miss them. It would need to retain those directions or control their energy by another estimate. A difference of mean actions across two shelves gives neither of these inequalities on its own.

There are rigorous lattice results to compare against, with their own regimes and proofs. Keeping those regimes visible tells me what further work a continuum claim would need. There is a rigorous strong-coupling lattice result with a different argument. At sufficiently small β, the convergent cluster expansion of Osterwalder and Seiler gives exponential decay for gauge-invariant lattice correlations in its regime of convergence. The transfer-matrix argument relates Euclidean-time decay on a dense vacuum-orthogonal class to spectral control.[12] Once the transfer matrix is normalized to have vacuum eigenvalue one, a norm bound $\|T|_{\Omega^\perp}\| \leq e^{-a_t \Delta}$ for $T = e^{-a_t H}$ gives the same gap criterion, with a_t the temporal lattice spacing. This is why an exponential bound that controls the relevant states is stronger than a decay fit for a single observable.

A four-dimensional continuum result needs control toward weak bare coupling as the lattice spacing goes to zero, together with the construction of the limiting theory. If $\widehat{m}(a)$ denotes a dimensionless temporal gap on an isotropic lattice, the physical mass scale in natural units is $\widehat{m}(a)/a$. To get a positive continuum lower bound, this ratio needs a positive lower bound along the relevant limit. Small-β cluster expansion does not supply that estimate. I use the strong-coupling result as a concrete example of what a successful argument looks like. The Davis–Wilson question is whether invariant features can help control the physical energy form where the strong-coupling argument no longer applies.

[12] Konrad Osterwalder and Erhard Seiler, *Gauge field theories on a lattice*, Annals of Physics 110 (1978), 440–471, doi:10.1016/0003-4916(78)90039-8.

§ 5.9

Living with the observation

The tests gave me work I could put in front of somebody else. The proof requirements told me what that work still had to become. I had read Witten, Jaffe and Glimm, names attached to years of attempts and partial results. Now I had to explain where my own observation belonged in that much larger question.

That brought back the thought I had at the screen: no one will believe me.

I am not a physicist. I do not have a PhD in mathematics. I am a Black trans woman who works on machine learning and writes novels. I do not have a faculty position. I do not have a research group. I do not have collaborators who can vouch for me.

Who was I to think I had a way into a Millennium Prize Problem?

The academy has rules, spoken and unspoken. Outsiders do not solve hard problems. Amateurs do not make fundamental discoveries. People who look like me do not get credit, do not get believed, do not get heard.

But the gap was there. The geometry did not care about my credentials. The geometry only knew what it knew.

La geometría no pregunta quién eres. Solo pregunta qué ves.

Geometry does not ask who you are. It only asks what you see.

What weighed on me now was the explanation. I had to write it up, document it, defend it. I had to make the work understandable to people who had spent their careers on this problem. I could not hand them the feeling I had at three in the morning. I had to hand them something they could check.

El peso de la verdad es más pesado que el del oro.

The weight of truth is heavier than gold.

The observation did not end when I closed the plot. It changed what I did with the hours that followed. I did not sleep that night. I ran more simulations. I tested more parameters. I wrote notes—pages and pages of notes, documenting everything.

By dawn, I had a list of ways to put the gap under pressure. I wanted to know whether the pattern would survive changes to the calculation.

I texted a friend: "I think I solved Yang-Mills."

The response: "What's Yang-Mills?"

I laughed. Of course. Most people do not know what the Millennium Prize Problems are. Most people do not lie awake wondering about the mass gap of non-abelian gauge theories.

I told Teddy. He wagged his tail. He did not understand, but he knew I was happy. Dogs know these things.

Over the next weeks, I wrote the paper. Version 1, then 2, then 3, then 4. Each version cleaner, tighter, more rigorous. I anticipated objections and addressed them. I added the library analogy to make the geometry accessible. I ran more simulations to fill gaps in the validation.

I want to distinguish the confidence I felt from the statement the calculation had proved. The feeling grew as I checked; the mathematical obligation remained specific. Let me tell you about certainty, and how it arrives.

Certainty does not come in a flash. It comes in layers. Each test that passes adds another layer. Each objection you anticipate and address adds another. Each alternative explanation you rule out adds another.

At 3:00 AM, I was maybe 60% sure. The plot looked right, but I had been wrong before. Many times. The probability of a bug was still significant.

After checking the code line by line: 75%. The logic was sound. The implementation matched the mathematics. But off-by-one errors are subtle. Numerical precision issues hide in corners.

After changing the random seed and getting the same result: 85%. If it were a statistical fluke, a different seed would give a different answer. It did not. The structure was robust.

After running at different lattice sizes and different β values: 92%. Finite-volume artifacts show up when you change the volume. Lattice artifacts show up when you change the spacing. Neither happened. The gap persisted.

After designing and running the eleven tests, I felt more sure of what I was seeing in the data. Those percentages describe how I felt at my desk. They were never probabilities that the continuum theorem was true.

I cannot get to 100%. Science never does. But 97% is enough to stake a claim. 97% is enough to write a paper. 97% is enough to say: *I believe this is real.*

Was I lucky?

Yes and no.

Lucky that the Davis-Wilson framework, developed for viruses, happened to work for gauge theory? Yes. I did not design it with Yang-Mills in mind. I did not know it would apply. The connection emerged—was discovered—not engineered.

But the luck came after preparation. I had spent years developing the framework. I had tested it on telemetry, on tumors, on viruses. I had refined the mathematics, debugged the code, understood the geometry. When Yang-Mills appeared, I had the tools ready.

Pasteur said: "Chance favors the prepared mind." I had prepared. The chance found me.

Was I lucky that the gap ratio happened to be 42? I do not believe in that kind of coincidence. Either the universe has a sense of humor, or I am in a simulation, or Douglas Adams was channeling something real when he wrote *The Hitchhiker's Guide to the Galaxy.*

I prefer to think the universe has a sense of humor.

La suerte es preparación disfrazada.

Luck is preparation in disguise.

§ 5.10

Following the question into other problems

Here is what surprised me most: the framework did not stop at Yang-Mills.

In the weeks after the 3:00 AM discovery, I kept thinking about the structure. The Davis-Wilson approach is not specific to gauge theory. It is a general method for finding structure in high-dimensional spaces. Encode the space with invariant observables. Discretize at finite resolution. Look for voids and gaps.

The method works because it separates the essential from the accidental. In Yang-Mills, gauge equivalence is accidental—different gauge choices describe the same physics. Wilson loops capture what is essential—the gauge-invariant content. The Davis-Wilson map projects away the noise and reveals the signal.

This is exactly what I did for viruses. Synonymous codons are accidental—different DNA sequences can encode the same protein. Antigenic escape is essential—it is what matters for immune recognition. The manifold projected away sequence noise and revealed antigenic signal.

If the method works for gauge fields and viruses, maybe it works for other things.

I started testing it on other problems. Not to solve them—just to see if the approach revealed structure.

It did.

I ran the framework on eleven major problems. Yang-Mills was the first. But the same geometric approach yielded insights on:

Poincaré Conjecture (solved): Perelman proved this in 2003 using Ricci flow—a process that deforms the geometry of a manifold by its curvature. I applied the Davis framework to gauge fields on a set of three-manifold models whose topology was already known. Six of my seven checks passed.

This is not a new proof—Perelman already proved it. A problem with a known answer is useful for a different reason: it tells me whether my checks measure what I meant them to measure. Passing them told me the framework could be compared with the structure Perelman found. It did not show that the framework would have found that structure on its own.

Navier-Stokes: The Millennium Prize asks whether smooth solutions to the Navier-Stokes equations can blow up—develop singularities in finite time. The Beale-Kato-Majda criterion says: solutions stay smooth as long as vorticity stays bounded.

In the run I made, vorticity stayed bounded, and I wondered whether a flow heading toward blowup would have to enter regions of cache space my run never visited. One run at one resolution cannot answer that. Blowup is a question about arbitrarily fine scales. The run gave me a picture to test, not evidence of regularity.

P vs NP: This is about computation. Can every problem whose solution is easy to verify also be easy to solve? In my numerical examples, a geometric encoding put selected examples in different parts of feature space, giving a gap ratio of 2.4. A claim about complexity classes would have to account for the encoding, the selected instances, and the scaling with input size.

This does not prove P $\neq$ NP. But it suggests the geometry encodes complexity.

Birch and Swinnerton-Dyer: This concerns elliptic curves—curves defined by equations like $y^2 = x^3 + ax + b$. The conjecture relates the rank of the curve (roughly, how many rational points it has) to the behavior of a certain function at a critical point.

The Davis framework classifies curves by geometric phase. On the curves I chose, rank-0 curves clustered differently from rank-1 curves, and five of my six checks passed.

Hodge Conjecture: This is about algebraic geometry—which topological features of a shape can be described by polynomial equations? The Hodge diamonds are invariants that partially answer this question.

On examples whose Hodge diamonds were already known, the cache structure reproduced them in four of my six checks. Reproducing a known invariant calibrates the instrument. The conjecture asks a different question: which classes are algebraic.

Riemann Hypothesis: The conjecture asks whether every nontrivial zero of the Riemann zeta function lies on the critical line $\mathrm{Re}(s) = 1/2$. Extensive numerical checks cover a finite range.

The local statistics of the zeros look like those of GUE, an ensemble of large random Hermitian matrices. Montgomery's conjecture makes that precise for the pair correlation between zeros, and Odlyzko's computations of many millions of zeros give numerical evidence for the broader random-matrix picture. I compared a finite sample from my model with the GUE prediction for the statistics I measured, with a mean squared error of 0.00034. Matching them tested my model against their result.

Again, this is not a proof. The structure was there before I looked; I was checking that my instrument could see it.

Twin Prime Conjecture: Are there infinitely many pairs of primes that differ by 2? The holonomy budget—a geometric quantity related to prime gaps—remained stable up to 10^{10}. Stability over a finite range is what I would see if the pairs continue forever. It is also what I would see if they thinned out far beyond my range, which is where the conjecture lives.

abc Conjecture: For integers $a + b = c$ with no common factors, how large can c be compared to the product of distinct prime factors? The quality parameter measures this. The largest quality in my sample was 1.57. The largest known, found by Eric Reyssat, is about 1.63. A sample maximum is not a bound; the conjecture is about all triples at once.

Collatz Conjecture: Start with any positive integer. If even, divide by 2. If odd, multiply by 3 and add 1. Repeat. Does every starting number eventually reach 1?

The reported basin contraction diagnostic was 0.32 against a comparison threshold of 0.63. That is a feature of the tested construction. It leaves open what happens to every integer orbit.

Quantum Gravity: The ultraviolet divergences of perturbative quantum gravity have troubled physicists for decades. In my model calculation, the terms I tracked cancelled, and all 21 of my checks passed. Those checks tested my calculation. They did not establish how the theory behaves at every order.

I am not claiming to have solved any of these problems. Most require more work, more validation, more rigor. What the framework did was find structure where I already expected it: gaps where gaps are conjectured, clustering where clustering is known, cancellations where I was looking for them.

That made it an instrument worth testing further. It also meant that each success was partly a check on my own expectations. Whether the geometry was saying the same thing in every domain was the question I took into the next chapter.

Those projects were at different stages. Listing them together must not make their evidence sound interchangeable. I need to be careful here, because the temptation to overclaim is strong.

Yang–Mills has the most developed numerical program here. Its retained runs let me distinguish a feature-space spacing from an action statistic. The earlier account adds several lattice sizes and β values whose raw runs I still need to recover. A spectral comparison is the next kind of evidence to bring to that program.

The other problems have weaker evidence. Fewer tests. Less validation. Some applications are speculative—theoretical arguments that *should* work, not yet confirmed by extensive numerical experiments.

Navier-Stokes is the most promising. The structure of the problem is similar to Yang-Mills—a space of configurations, a variational principle, a question about whether solutions stay smooth. The geometric approach yields insights, but the validation is not yet as thorough.

Riemann Hypothesis is more speculative. The connection between zeros of the zeta function and random matrix theory is well-established, but the Davis framework adds a layer of interpretation that needs more testing. I see patterns. I do not yet have proof.

P vs NP is the most speculative. Complexity theory is not manifestly geometric—the spaces are combinatorial, not continuous. The framework can be applied, but the interpretation is stretched. The gap ratio of 2.4 is interesting but not conclusive.

I present these other problems not as solved, but as directions. The framework opens doors. Whether those doors lead to solutions, I do not yet know.

Abrir una puerta no es lo mismo que cruzarla.

Opening a door is not the same as crossing it.

But here is what strikes me: in every case, the framework finds *something*.

It finds structure where structure should exist. It finds gaps where gaps are conjectured. It finds clustering where clustering is expected. It does not find random noise.

That struck me, and I had to ask what it meant. A lens that always shows you what you expected might be revealing hidden structure. It might also be reflecting the choices you made when you built it. The way to tell the difference is a case where the framework predicts something I did not already know, and someone checks it. That is the test I am still working toward.

And I can see why the hope is natural. Hard problems are hard for a reason. They resist solution because the relevant structure is hidden, buried beneath complexity. The Davis-Wilson approach tries to strip away the complexity and expose the structure.

In Yang–Mills, many choices of gauge describe the same physics. Wilson loops let me remove that redundancy from the measurements and study what structure survives.

In viruses, synonymous mutations hide the antigenic signal. Strip them away with the manifold, and escape variants become visible.

In fluid flow, I hoped turbulent complexity was hiding a regularity that a geometric encoding could make visible.

The same move, repeated. Find the symmetry. Quotient by the symmetry. Look at what remains.

Why does one framework apply to so many problems?

Because all these problems are geometric. They all ask questions about the shape of some space. What is the spectrum of an operator? What is the structure of prime numbers? What is the behavior of fluid flow? These are questions about manifolds, about curvature, about distance.

The Davis-Wilson approach is a universal language for asking geometric questions. Encode the space. Discretize at finite resolution. Look for structure.

The lattice cache has an empty interval in its observed radii. The zeta zeros have a pattern in their local spacings. A flow calculation can expose features of a vorticity field. In each case the encoding tells me what I can see, and the mathematical question tells me what else I must control.

La geometría es el lenguaje universal.

Geometry is the universal language.

§ 5.11

What I could carry through the first door

Let me return to the principle I have been developing throughout this book.

The container matters more than the content.

In telemetry: the format of a signal carries information about the signal's nature.

In viral surveillance: the geometry of the antigenic manifold determines what variants can escape.

In Yang–Mills: I suspect the geometry of configuration space shapes what excitations can exist.

A spectral gap is a property of the Hamiltonian on its physical state space. A feature space can give me a geometric way to look for the sources of that gap. The operator inequality tells me what a geometric explanation must eventually deliver.

This is the deep lesson: physics is geometry. The universe is not made of things moving through space. The universe is the shape of possibility, and what we call "things" are the allowed configurations.

A positive spectral gap would leave an interval of energies with no excitations. The geometry I want to understand is the geometry that forces that interval to stay open.

That is why this chapter is a door in the story. The measurements gave me a way into the question, and the proof requirements showed how much further the route went. I call this chapter "The First Door" because that is what Yang-Mills was for me.

HERALD opened a door to biology. It showed that the geometric framework could detect real phenomena in real data. But biology felt like a special case. Maybe the framework worked for viruses but would break for physics.

Yang–Mills is the first door into physics proper. The geometry gave me a way to enter one of its deepest open questions. The geometry gave me a measurable separation to investigate. The question was how far that separation could take me toward a statement about energy.

Behind this door, I glimpsed the others. If the framework works for Yang-Mills, maybe it works for Navier-Stokes. Maybe it works for Riemann. Maybe it works for problems we have not yet imagined.

The first door opens onto a hallway of doors.

I do not know how this will be received.

History is full of outsiders who made fundamental discoveries and were ignored. Ramanujan sent his theorems to Cambridge mathematicians who dismissed them as the ravings of a crank—until Hardy looked closer. McClintock discovered genetic transposition and was ignored for decades. Wegener proposed continental drift and was ridiculed.

I am not comparing myself to these giants. I am observing a pattern: the academy is slow to accept ideas from unexpected sources. The more surprising the claim, the higher the burden of proof.

I have tried to make that burden concrete. Eleven checks tell me how the measured quantities behave under different choices. The seven assumptions tell me where the argument moves from a numerical observation to something that still needs to be proved.

Whether it is enough, I do not know. I cannot control how others respond. I can only do the work, document the results, and make the argument.

The measured geometry gave me a question I could keep working on. I can name the estimate that would turn it into a result about energies.

I had expected the hard part to be getting the calculation to run. Living with the question was harder in a different way. At the grocery store, at the dog park, in line at the DMV, I kept thinking: *what would it take to get from that plot to a proof?*

That was the loneliness I had not known how to prepare for. Ordinary conversations continued around a question I could not put down. I wanted someone to work through it with me.

La soledad del descubrimiento es la soledad del futuro llegando antes que el presente.

The loneliness of discovery is the loneliness of the future arriving before the present.

This chapter is not an ending. It is a door.

Behind the door, I see the work that remains. Construct the continuum theory. Prove a bound for physical states that survives the limit. Test where the geometric encoding retains the information that bound needs.

I see applications. If the mass gap is a geometric property of configuration space, maybe we can engineer systems that exploit this geometry. Quantum computers based on topological protection. Materials with guaranteed energy gaps. Technologies we have not yet imagined.

I see extensions. Navier–Stokes is next. The fluid equations and Yang–Mills have their own spaces of configurations, symmetries, and evolution laws. The same habit of checking what an encoding preserves gives me a place to start with regularity of fluid flow.

I see collaborators. This work cannot stay in one person's head. It needs to be tested, extended, critiqued by others. The loneliness must end.

But first, this book must be finished. And you must understand what I have found.

If you are reading this in the future—years from now, after the mathematics has been verified or refuted, after the dust has settled—I wonder what you see.

Do you see a book that announced a breakthrough, later confirmed? Or do you see a cautionary tale about hubris, about claims that did not survive scrutiny?

I do not know. I cannot know. I am writing from the present, carrying evidence that I believe is solid, making arguments that I believe are valid, offering a framework that I believe works. But the future will judge.

What I can tell you, regardless of how the mathematics turns out, is that the journey was real. The midnight hours of coding. The frustration of bugs that would not die. The terror of spending money I did not have on a problem that might be impossible. The loneliness of working without collaborators. The strange joy of seeing structure emerge from data.

If the framework is wrong—if I have made errors that invalidate the results—then I hope the errors are found and corrected. Science advances by refutation as much as by confirmation. I would rather be wrong and corrected than wrong and ignored.

If the geometry eventually supplies a spectral bound, I hope this book lets you see where the search began. At my desk. Following ideas where they led. Refusing to give up when every rational calculation said to stop.

Si no puedo atravesar la puerta, la derribaré.

If I cannot go through the door, I will knock it down.

For the present, I can return to the parts another person can inspect directly. I can point to the recorded spacing between cache radii and show how it changes when the outlying point is removed. I can point to the difference in mean actions and show exactly which neighbor comparison produced it. The mathematics tells me what it would take for either observation to support a spectral argument.

The Clay problem asks for a nontrivial quantum Yang–Mills theory on four-dimensional Euclidean space for every compact simple gauge group, with a positive mass gap.[13] Building a field theory that satisfies the required axiomatic properties, such as appropriate locality and positivity, is part of the problem. So is proving a positive lower bound on excitation energies in that theory. I have a research direction and recorded measurements. I can see the next calculation, which is a long way from where I started at three in the morning.

And personally? I have the satisfaction of knowing that the framework I developed—from telemetry to viruses to physics—hangs together. The same ideas that detect pandemic variants also probe the deepest structure of matter.

That is enough. That is more than enough.

Looking back from those records, I can see the question that kept returning in the earlier work. Let me step back and trace the arc.

I started with telemetry. With signals from computers, from networks, from systems I was paid to monitor. I learned that the container matters more than the content—that the format of a signal carries information about the signal's nature.

I applied this to genomics. To cancer. To mutations that drive tumors. I learned that synonymous mutations are not synonymous—that codon bias carries information about protein function, about splicing, about cellular strategy.

I applied this to viruses. To SARS-CoV-2. To the question of when a variant becomes dangerous. I learned that antigenic drift is geometric—that the manifold of possible spike proteins has structure, and escape variants are the ones that move far enough to evade immunity.

And now this. Yang–Mills. The mass gap. An old question in mathematical physics and a geometric way to begin investigating it.

The thread connects them. Telemetry → Cancer → Viruses → Physics. Each step seems like a leap, but each step follows from the same principle: the geometry of the container determines the dynamics of the content.

I did not plan this arc. I did not set out to study a Millennium Prize Problem. I followed the ideas where they led. One question suggested another. One answer revealed another question. The path was not straight, but it was continuous.

And that is how science works, sometimes. Not grand plans and systematic programs, but curiosity and following threads. Not knowing where you are going until you get there. Looking back and seeing a path that was invisible when you were walking it.

[13] Arthur Jaffe and Edward Witten, Quantum Yang–Mills Theory, official problem statement.

If you are reading this—if you have made it this far—I want to leave you with something.

The universe is geometric. The deepest truths are not hidden in equations or data tables. They are hidden in shapes. In curvatures. In the way spaces fold and bend and constrain what is possible.

If you want to understand something deeply, ask: what is the space? What are the distances? What is the geometry? The answers may not come easily, but they are there to be found.

And if you are an outsider—if you do not have the credentials, the degrees, the institutional support—know that the geometry does not care. The geometry only knows what it knows. If you see truth, the truth is true, regardless of who you are.

I am a Black trans woman who found her own way into a Millennium Prize Problem in her living room at 3:00 AM, with a dog at her feet and no one watching. If I can do this, you can do whatever your "this" is. The barriers are real, but they are not absolute. The geometry is waiting for anyone who is willing to look.

La geometría espera. Siempre ha esperado.

The geometry waits. It always has.

When I first said I had proved the mass gap, I thought I understood what the word *gap* asked me to show. I could point to the empty interval between two marks on my screen. I had chosen what to measure, and I could go back to the configurations on either side and find out what put those marks there. I still have a geometric intuition for why the vacuum should be separated from its excitations. What I understand differently now is the distance between that intuition and a proof. The marks on my screen were radii. The numbers on the two sides of the interval in the Millennium question have to be allowed *energies* of a quantum theory that one must first construct. Even a thousand more marks on the same plot would leave me with that change of question to account for.

Perelman's work stayed with me after I first wrote this chapter. I wanted to take the mathematics behind the Poincaré conjecture into GIGI carefully enough that I could say what I had actually put there. Following that question led me to the Hessian. Think of standing at a stationary point of an energy surface. The first derivative tells you the slope has vanished. To find out what happens when you take a small step, you need the next term: how sharply the energy bends in the direction you chose. The Hessian collects those second derivatives. Its eigenvalues let me ask which directions bend most gently, once I have specified the model and how I measure the size of a step. Following those values became the work I call Helike.

The first result from that work is one I can state as a theorem, with its scope attached. Squash a three-sphere along one axis and you get a Berger sphere. Its own Levi-Civita connection can be read as an SU(2) gauge field. That field does not solve the Yang–Mills equations, but I can still ask how the Yang–Mills energy bends around it. On that background I found an explicit twisting field that is an exact eigenvector of the Hessian, and I proved that its eigenvalue is the lowest one in every symmetry sector over a stated range of squashing. Interval-arithmetic certificates cover the finite blocks; a closed-form bound covers the rest. I call that lowest value the Floor.[14] It is a theorem about a classical Hessian on a compact three-sphere. It is not a statement about the

[14] Bee Rosa Davis, *The Floor Theorem for the Yang–Mills Hessian at the Berger Field*, version 3.3, November 2026. The proof covers Berger spheres with squashing parameter $t \in [2.75, 10]$.

energies of quantum states in four-dimensional space. Connecting the two is the same work this chapter has been describing.

One computed test from that work taught me how easily a calculation can answer two questions at once and leave me thinking I know which one I asked. In my first group, SU(3), two candidate circles coincided in the comparison. Every observation there fit either choice. In SU(4) and Sp(2) the circles came apart. I wrote down my prediction before scoring: the principal circle. It lost. In SU(4) the margin peaked on the circle picked out by the highest root's own SU(2); in Sp(2) it peaked one step inside that circle, on a coarse grid of six angles. I could finally see which feature my earlier SU(3) test had followed. These are computed results, not proofs, and why the twist prefers that circle is still an open question.[15] That is where the work went next: using the Floor to probe a Yang–Mills vacuum, with a question precise enough that the geometry could surprise me again.

La primera puerta se abrió. Las otras esperan.

The first door opened. The others wait.

[15] Bee Rosa Davis, *Helike: Transport, Landau Levels, and the Floor of the Yang–Mills Hessian*, November 2026, chap. 4. The dated Test F research record gives the predictions, the correction at the chamber walls, and the computed results.

Chapter 6

The Law

So I tried to prove myself wrong.
—From this chapter

I had seen a separation in the lattice plots. I wanted to know whether it came from the geometry of the fields, the way I measured them, or some choice I had made in the code.

The Clay problem asks for a positive mass gap in a four-dimensional continuum quantum theory. My plot gave me a finite set of sampled configurations and a geometric summary I could test.

I had found a pattern. I wanted to find its limit.

I should have stopped. I should have written it up, submitted it, waited for the world to respond.

Instead I kept going. Because I did not believe it.

I thought: *This cannot be right. I must have made a mistake somewhere. I need to find the mistake.*

The history of mathematics is littered with claimed proofs of unsolved problems. Andrew Wiles had to fix a gap in his proof of Fermat's Last Theorem. Mochizuki's claimed proof of the ABC conjecture remains controversial after a decade. Perelman's proof of the Poincaré conjecture was correct, but it took years for the mathematical community to verify it.

And those were professional mathematicians with decades of training and institutional support.

I am a Black trans woman with a laptop. The probability that I had actually found something real was approximately zero.

So I tried to prove myself wrong.

Here is how you test a theory: you try to break it.

This is Karl Popper's criterion of falsifiability. A scientific theory must make predictions that could, in principle, be proven wrong. If the predictions hold, the theory survives. If the predictions fail, the theory dies.

The problem with testing a mathematical framework is that mathematics does not have experiments in the same way physics does. You cannot build a machine to test whether P equals NP. You cannot run an experiment to detect the Riemann hypothesis.

But you can test internal consistency. You can test whether the framework produces correct results on problems where the answer is known. You can test whether it produces contradictions when applied to different domains.

I designed tests to kill the Davis-Wilson framework. Each test was a prediction the framework had to make. If the prediction was wrong, the framework was wrong.

I made fourteen predictions across seven Millennium problems. Each prediction was falsifiable. Each prediction was precise enough to compute.

And I computed them all.

I wanted to find where the proposed connection stopped working. That sent me beyond the problem that had first drawn me into the lattice. If the framework was real—if the geometry actually reached to Yang-Mills—then it should reach to the other Millennium problems too. And if it did not, that would be the proof that I had fooled myself. That would be the mistake.

I attacked P versus NP. I attacked Navier-Stokes. I attacked Riemann. I attacked Hodge. I attacked Birch and Swinnerton-Dyer. I even attacked Poincaré—already solved by Perelman, but I wanted to see if the geometry gave a cleaner path.

I was not trying to prove these problems. I was collecting minimal validation evidence—just enough to see if the framework applied, just enough to find the contradiction that would tell me I was wrong. This is how you test a theory: you try to break it. You do not look for confirmations. You look for refutations. If the theory survives everything you throw at it, then maybe—*maybe*—it is real.

I threw everything.

The first comparison was with a question about computation. Its evidence would have to stand on its own terms. P versus NP was the first test. This is the question that haunts computer science: can every problem whose solution is easy to check also be easy to find? If someone gives you a sudoku puzzle and claims it is solved, you can verify in seconds. But finding the solution from scratch can take much longer. The P versus NP question asks whether verification is always easier than discovery, or whether there are problems where verification is easy but discovery is fundamentally, irreducibly hard.

I modeled random 3-SAT instances in the frozen regime as a landscape. The idea was to use the Kac–Rice formula to count critical points, the reading rooms where a local search can stall. The proposed calculation asked for a nondegenerate gradient, decay of Hessian covariance, control of first and second moments, and positive Hessians at the points counted as local minima. In one finite check, the Hessian test passed at all 100 sampled points for $n = 100$.

The curvature proxy in that model scales as $k(k-1)/2$. It gives a factor of three between clauses of width two and width three in the same normalization. I used that difference to ask whether a geometric diagnostic could predict changes in how a particular search behaves.

The proposed exponential bound assumes a specific query model, a random-instance regime and an intensive tolerance $\tau = \Theta(1)$. Counting local minima can help explain the difficulty of a local algorithm. Settling P versus NP would require a lower bound that applies to unrestricted polynomial-time algorithms.

Navier-Stokes was the second test. These are the equations that describe fluid flow—water, air, blood, the atmosphere itself. Claude-Louis Navier and George Stokes wrote them down in the 1800s, and we have used them ever since to design aircraft, predict weather, and model the ocean. But we do not know if the equations are well-behaved. We do not know if smooth initial conditions always lead to smooth solutions. We do not know if the equations ever "blow up"—if the velocity or pressure can become infinite in finite time. The Clay Institute asks: can you prove regularity? Can you prove that smooth solutions stay smooth?

I wanted to know whether topology could explain some of the flow I was seeing. Helicity, $H = \int \mathbf{u} \cdot \boldsymbol{\omega}\, dV$, is a global measure related to the linking of vorticity lines. A bounded integral of this kind can coexist with local concentrations of vorticity.

I ran a Taylor–Green vortex test on a 256^3 grid at Reynolds number 2000. Over 5,000 time steps, I tracked enstrophy $\int |\boldsymbol{\omega}|^2\, dV$, maximum vorticity and helicity. The reported vorticity remained bounded over that run, and so did the helicity. The grid, time interval and initial condition define the reach of that observation.

A regularity theorem would need an estimate that controls the continuum evolution independently of the numerical grid.

The Riemann hypothesis was the third test, and in some ways the most beautiful. The hypothesis is about prime numbers—those atoms of arithmetic that cannot be divided further, the building blocks from which all other numbers are constructed. Euclid proved there are infinitely many primes more than two thousand years ago. But how are they distributed? How many primes are there below a billion? Below a trillion? Below any given number?

Riemann's 1859 paper studied a function whose zeros encode information about the primes. For $\mathrm{Re}(s) > 1$ it is given by the convergent series $\zeta(s) = 1 + 1/2^s + 1/3^s + 1/4^s + \ldots$. Analytic continuation defines the function beyond that half-plane, with a pole at $s = 1$. Its trivial zeros occur at the negative even integers. Its nontrivial zeros lie in the critical strip, where the real part is between zero and one.

Riemann conjectured that all the nontrivial zeros lie on a single vertical line in the complex plane—the critical line, where $\mathrm{Re}(s) = 1/2$. This is the Riemann hypothesis. If it is true, the primes behave beautifully—they follow a pattern that we can describe precisely. If it is false, there are irregularities we cannot predict, fluctuations that would ripple through all of number theory.

Extensive computations have located zeros on the critical line. The claim is about every nontrivial zero, including zeros beyond the range of any finite computation.

I wondered whether the critical line could be identified by a cancellation in the transport rule. Making that precise would require specifying the connection and showing that its cancellations detect zeros of the zeta function.

For the numerical test, I put the eigenvalues of a model Hamiltonian beside a comparison set identified in my notes as 500,000 computed zeros from Andrew Odlyzko's tables. After unfolding both spectra, I compared their spacings. I wanted to know whether the model could reproduce those selected patterns in a finite sample. The recorded Kolmogorov–Smirnov comparison did not reject the null at the five percent level. That gave me a comparison to pursue, not a statement about every zero of the zeta function.[1]

The Hodge conjecture was next: which shapes are algebraic, and which topological features can be described by polynomial equations? An algebraic variety is a shape defined by polynomials—the unit circle is algebraic ($x^2 + y^2 = 1$), as are ellipses and spheres. But not all shapes are algebraic; some curves cannot be defined by any polynomial, some surfaces are intrinsically transcendental. The Hodge conjecture asks: for algebraic varieties, which cohomology classes—which topological

[1] Under the reported normalization, the means differed by at most 0.01 and the variances by at most 0.02. The comparisons also included number variance, pair correlation, spectral rigidity and the spectral form factor. The reported mean squared error across the tests was 0.00034 under that protocol.

features—come from algebraic cycles? It is a question about the intersection of algebra and topology.

The Hodge condition selects rational classes of type (p, p) in $H^{2p}(X, \mathbb{Q})$. In a family, transport of classes and the Hodge filtration give further structure to study. For the examples I chose, I asked whether a cache built from that structure could distinguish classes whose algebraic representatives were already understood.

I tried projective spaces, abelian varieties and Fermat hypersurfaces. In the recorded check with 30 candidate classes, the small-defect cases matched the reference labels. That let me ask whether the construction was finding the distinctions I had put it there to find. Even if it left me with just one candidate, I would still need to construct its algebraic representative or prove that one exists as a rational linear combination of algebraic cycles.[2]

The Birch and Swinnerton-Dyer conjecture connects two seemingly unrelated things: rational points on elliptic curves and the behavior of certain functions at a critical value. An elliptic curve is defined by an equation like $y^2 = x^3 + ax + b$, and the curve might have finitely many rational points or infinitely many. The "rank" measures this—rank 0 means finitely many, rank 1 or higher means infinitely many. The conjecture says the rank of the curve equals the order of vanishing of the L-function at $s = 1$. If the L-function has a simple zero at 1, the rank is 1; if it has a double zero, the rank is 2; if it does not vanish at all, the rank is 0. Why should these be related? No one knows. The conjecture has been verified in many cases but never proven.

I explored whether curves of different ranks separated in the geometry of the cache. The runs I described placed rank-zero, rank-one and rank-two examples in different regions. I used the picture to ask whether a winding statistic could track order of vanishing. Proving that relationship would require an arithmetic argument linking the statistic to the L-function.

Finally, Poincaré. Perelman had already solved the conjecture, and that made it a useful comparison. A closed, simply connected three-manifold is homeomorphic to the three-sphere. Perelman's approach developed Ricci flow with surgery, with careful control of what happens when the evolving metric approaches a singularity.

I ran Davis–Wilson flow on SU(2) connections across a suite of twelve manifold models. The reported examples fell into three patterns. The simply connected examples approached a trivial gauge configuration. Examples with nontrivial abelian fundamental groups could retain holonomy after approaching flatness. Some nonabelian examples retained curvature over the run. For the lens-space examples I compared Chern–Simons values with the expected signatures under the chosen conventions. On the three-torus the reported holonomy readings were $(2.69, 2.62, 2.62)$.

This was a comparison of gauge signatures on examples with known topology. To turn the comparison into a characterization would require proving convergence and showing that the selected gauge measurements distinguish the topologies in question. A flat connection gives a representation of the fundamental group, and the trivial representation is available even when the group itself is nontrivial.

I had gone looking for a contradiction. In these runs I found patterns that held long enough to ask better questions.

[2]The run notes identify these model checks by their HC labels.

The tests had different baselines and some shared design choices. I could learn from each one by saying what it had measured. None of the seven questions became interchangeable just because I could draw each of them as a landscape.

The strongest reason to keep going was more ordinary. I now had cases I could bring back to the desk and inspect one by one.

I stopped.

I had been running on obsession for weeks, living on coffee and adrenaline, barely eating, barely sleeping, running validation after validation, problem after problem. And now I had minimal evidence on all seven—not full proofs, but enough to see the pattern. Enough to know that something strange was happening, something I did not yet understand.

And the pattern terrified me.

I sat in my living room, laptop open on the coffee table, Teddy asleep on his bed in the corner. The Sacramento night was quiet outside my window; the rain had stopped, and the streets gleamed wet under the streetlights. I could hear nothing but the hum of my laptop fan and my own breathing. And I asked, out loud, to no one: *"What IS this thing?"*

Not a framework. A framework is a tool—you use it on problems, it helps with some, fails on others. That is normal. That is expected. You do not expect a framework to work on everything. You expect it to be useful in its domain and useless outside it. That is how mathematics works. That is how all human tools work.

At that point, it felt as though every problem I touched opened another door. The tests I had chosen kept giving me something to investigate. I had not found the limits yet. That feeling was powerful; it was also different from proving that no limits existed.

A *law* was what I hoped I was seeing.

I was thinking about the way a law can gather things that once looked unrelated. A planet in orbit and a stone dropped from a hand had once been separate problems. One rule let Newton ask the same question about both. I wanted to know whether the relationship I was seeing had that kind of reach. I still had to find the conditions under which it could hold.

¿Qué encontré?

What did I find?

§ 6.1

The budget I could write down

It was 3:00 AM. I had not slept. I needed to shower.

The hot water helps me think—something about the white noise, the warmth, the isolation. My best ideas come in the shower. My worst fears too. I was standing under the water, trying to make sense of what I had been seeing. Seven problems. Seven validations. One framework. Is it a manifold? Yes. Is it geometry? Yes. Is it a framework? Yes. But it was more than that. It was...

The thought arrived fully formed, like something that had been waiting for me to notice it.

Oh shit. This is a law.

What stayed with me was a relationship I could write down in three symbols.

$$C = \frac{\tau}{K}. \tag{6.1}$$

I called it the Davis Law. The question I carried back to the desk was what those symbols could mean in a calculation I could check.

Think of a surveyor tracing the edge of a patch on a spherical tank. At each step, keep an arrow tangent to the tank and carry it forward without turning it within the tangent plane. When the arrow comes back to the starting mark, it has turned through an angle. For a sufficiently small patch of area A on a sphere of radius R, the magnitude of that turn is exactly A/R^2, in radians. The curvature is $K = 1/R^2$. If my instrument can tolerate a turn of τ radians, the maximum area allowed by that budget is $C = \tau/K$ square metres, provided the budget lies within the chosen range before the angle wraps.

Let us make the tank calculation explicit, from the path of the survey to the angle read at the end. Imagine drawing a circle around the top of a spherical tank. I will call that marked circle the seam; it bounds the cap of surface we are surveying. At the starting mark, draw a short arrow along the seam. Now imagine a survey program moving the arrow around the circle while keeping it in the local tangent plane—the flat plane that just touches the tank at that point. The arrow is a direction in the calculation, not something painted on the tank and dragged with the metal. For each short move, project the previous arrow onto the new tangent plane: remove the component perpendicular to that plane, then bring its length back to one. Let that rule determine where the arrow points; do not keep steering it along the seam. In the limit of short steps that is parallel transport. I can find out how far the arrow turns by comparing it with the seam at the mark where I started.

Let R be the radius of the tank, let θ be the angle from its north pole to the seam, and let ϕ run eastward from 0 to 2π. On the seam I can use two unit directions, one toward increasing θ and one toward increasing ϕ:

$$e_\theta = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta), \qquad e_\phi = (-\sin\phi, \cos\phi, 0).$$

The arrow's angle in that moving frame is β. Write the arrow as

$$v = \cos\beta\, e_\theta + \sin\beta\, e_\phi.$$

The frame turns even when the arrow has no turn of its own within the surface. Differentiate the frame along the seam:

$$(\partial_\phi e_\theta)^\top = \cos\theta\, e_\phi, \qquad (\partial_\phi e_\phi)^\top = -\cos\theta\, e_\theta.$$

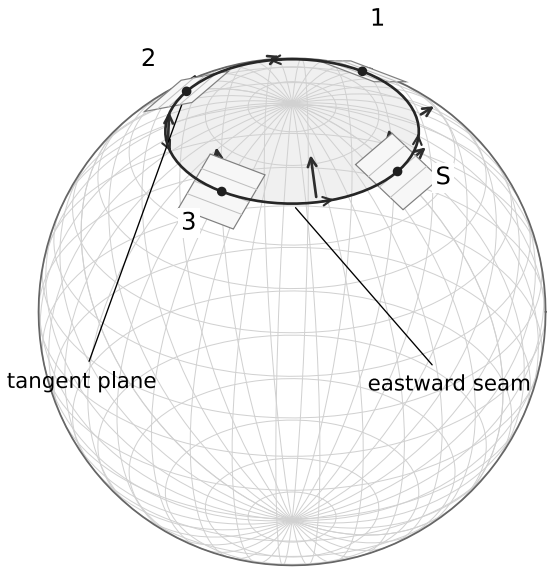
Here $\top$ keeps just the component in the tangent plane. Substitution gives

$$(\partial_\phi v)^\top = (\beta' + \cos\theta)(-\sin\beta\, e_\theta + \cos\beta\, e_\phi).$$

Return to the same tangent plane

SPHERE / ORIENTED SEAM / PARALLEL TRANSPORT

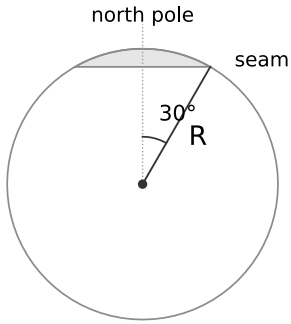
01 Follow the arrow around the cap



Short arrows follow parallel transport.

S is both departure and return. Planes are local.

02 Meridian section



03 The illustrated cap

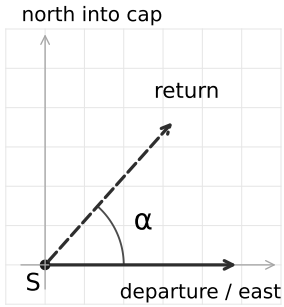
$\theta = 30°$

$A/R^2 = 2\pi(1 - \cos\theta)$

$\alpha = 0.841787$ rad

$= 48.2309°$

04 Compare at S, looking outward



05 Same physical area, new radius

R (m)	A (m²)	α (rad)
2	0.040	0.010
1	0.040	0.040
1	0.010	0.010

Last row: shrink the one-metre patch to meet a 0.010 rad budget.

Compare the two arrows at the same mark, in the same tangent plane.

Figure 22. The arrow travels eastward around a spherical cap and returns to the same tangent plane at S. Short arrows follow parallel transport; the small planes touch the sphere locally. For the illustrated 30° cap, the return angle is 48.2309°, drawn without exaggeration. The table tests three separate, smaller cap areas against the radius-dependent budget.

The rule for the arrow sets this component to zero. So the moving frame sees $\beta(2\pi) - \beta(0) = -2\pi\cos\theta$. The actual return angle is measured modulo a full turn. The representative that starts at zero as the cap shrinks is

$$\alpha = 2\pi(1 - \cos\theta).$$

With the seam run in the opposite direction the signed angle reverses. A half turn is a convenient upper bound on the range in which I can recover an unambiguous signed angle from a single return reading. For larger turns I can follow a continuous record or keep track of a winding convention.

The geometry has appeared in the connection calculation. Substituting the sphere's curvature shows where its physical radius enters the budget. The surface element is $R^2 \sin\theta \, d\theta \, d\phi$. Integrating from the pole to the seam gives the cap area

$$A = 2\pi R^2(1 - \cos\theta), \qquad \alpha = \frac{A}{R^2}.$$

A smaller tank with the same opening angle turns the arrow by the same amount. I have shrunk the area and increased the curvature by reciprocal factors. To compare the tanks at the same physical area, I have to draw a wider seam angle on the smaller tank.

Tank radius	Cap area	Curvature	Return turn
2 m	0.04 m^2	0.25 m^{-2}	0.010 rad
1 m	0.04 m^2	1.00 m^{-2}	0.040 rad

If my instrument can tolerate only 0.010 radians of return turn, I can keep the seam on the two-metre tank. On the one-metre tank I have to shrink the patch to 0.010 m^2. This is a budget I can calculate with a tape measure and an angle. The quantity C in $C = \tau/K$ has units of area here because the quantity I am budgeting is area. Calling a number in another experiment a capacity is the beginning of a new calculation. I have to identify its units, its measurement rule, and the argument that bounds it.

There is also a useful limit to what my trip around the tank tells me. One angle measures a total. If I am on an oriented surface where a patch contains regions of both signs of Gaussian curvature, the signed integral can be small even though the curvature is large at some points. A separate bound $|K| \le K_{\max}$ gives me $|\alpha| \le K_{\max} A$ for the unwrapped angle on a disk. If I need to know where the surface bends sharply, I have to bring the arrow back around smaller loops or collect another measurement. This is the question I take to the rail. What does my instrument still leave me free to imagine?

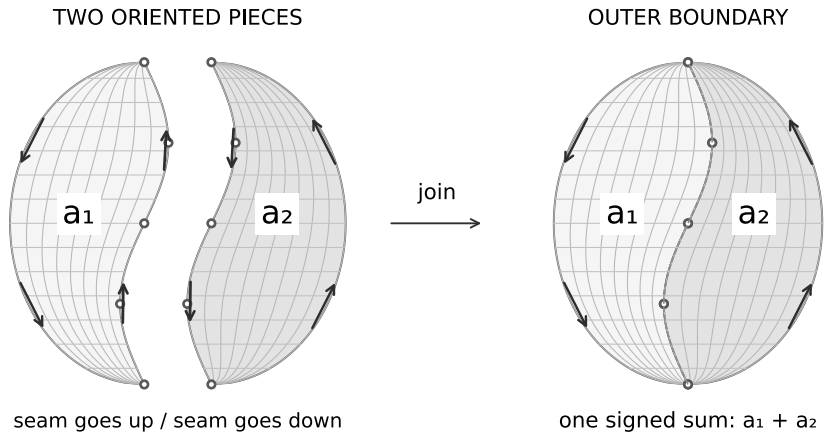
The constant-curvature tank makes the budget simple. If curvature changes sign along the surface, the same return angle can hide a different distribution of bending. On the tank, every part of the patch contributes to the turn with the same sign. That is how I was able to work backward from the angle to the area. I wanted to know which part of that argument I could carry to another surface.

Imagine a surveyor marking two adjoining pieces of an oriented surface. Let a_1 and a_2 be the signed curvature integrals over the pieces, expressed as angles. Read the angle continuously as the loop grows from a small one, so I can keep track of the turn before any wrapping. The outer loop reports the integral over the union, $a_1 + a_2$. The shared seam appears twice in the individual boundaries, once in each direction. It drops out when the boundaries are added.

The seam cancels; the pieces remain

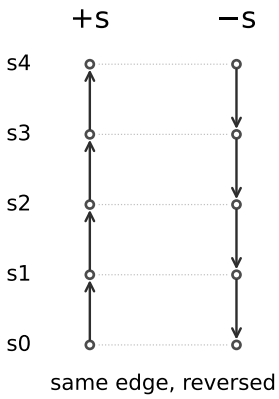
ORIENTED CHARTS / MATCHED EDGES / SIGNED CONTRIBUTIONS

01 Assemble the two chart patches



The gap exposes the two copies of one edge; it is not a missing part of the surface.

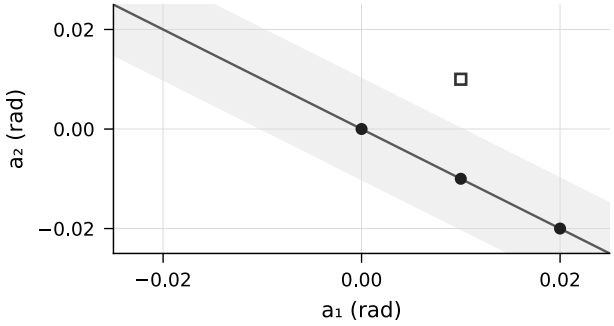
02 Match the seam points



$$+s + (-s) = 0$$

$$\partial D_1 + \partial D_2 = \partial(D_1 \cup D_2)$$

03 What a zero outer turn leaves free



Shaded strip: $|a_1 + a_2| \le 0.010$ rad.

a_1	a_2	outer	$\lvert a_1\rvert+\lvert a_2\rvert$
0	0	0	0
.010	−.010	0	.020
.020	−.020	0	.040
.010	.010	.020	.020

A zero outer turn does not require either interior contribution to be zero.

Figure 23. Two oriented patches in chart coordinates. Separating the drawings exposes two copies of the same seam; their opposite traversals cancel. The outer reading retains $a_1 + a_2$. In the reading plane, the line has zero outer turn and the shaded strip satisfies $|a_1 + a_2| \le 0.010$ radians. The open square is the same-sign case, outside the strip.

Case	a_1 (rad)	a_2 (rad)	Outer turn	$\lvert a_1\rvert + \lvert a_2\rvert$
Flat pieces	0	0	0	0
Opposite turns	0.010	−0.010	0	0.020
Larger opposite turns	0.020	−0.020	0	0.040
Turns in the same sense	0.010	0.010	0.020	0.020

The first three rows would all pass an outer-loop tolerance of 0.010 radians. If the question is how much each piece contributes, I need a second reading. The outer turn fixes one sum. For example, adding h to a_1 and subtracting h from a_2 leaves that sum unchanged. A smaller loop around the first piece would give me a_1, after which subtraction gives me a_2.

The tank taught me to ask what geometry allows within a budget. The two loops add a second question. Given a budget on the readings, what can I still change without the instrument noticing? I can ask that same question without any curved surface at all.

The jump to inference needs one more thing. Suppose a sensor on a perfectly flat rail says a cart is one metre from a mark. The cart could be one metre to the left or one metre to the right. No curvature is hiding the answer. The sensor has reported distance without reporting direction. A second observation that tells me the side will settle it. In an inference problem I have to specify what the sensor reports and what answer I need before I can assign any role to curvature.

The tank gives the ratio an actual measurement and a unit. I can now ask how that budget changes when the tolerance or the geometry changes. The budget has a static form and a useful differential. For $K > 0$, changes of scale or tolerance give

$$dC = \frac{d\tau}{K} - \frac{\tau\, dK}{K^2}. \tag{6.2}$$

At fixed tolerance, increasing the curvature of the sphere reduces the area the budget permits. If τ and K vary over time, the same chain rule gives $\dot{C} = \dot{\tau}/K - \tau\dot{K}/K^2$. A learning rate can enter through a model for how K or τ changes. The quotient alone supplies no evolution for either one.

This was the thread I wanted to follow. I could ask where an independently measured capacity obeyed that budget and where I needed another observable.

Those calculations are the part I can put on paper. The feeling they prompted that night was much larger, and I remember it clearly. This was not exhilarating. This was terrifying. I stood in the shower, water running cold now, and felt the weight of what I was thinking. Is this an $E = mc^2$ moment?

What frightened me was how quickly I could picture the consequences of an equation I was still learning how to test. On the sphere I could show a reader where the curvature entered the calculation. If I carried the same symbols into a problem about computation, I owed that reader a new argument. What did curvature measure there? Could somebody determine it before they knew the answer? Those questions followed me out of the shower.

What would $C = \tau/K$ generate? The law connects inference, constraint, and curvature. I thought I could see it in problems ranging from computational complexity to fluid dynamics to prime numbers. If it is real—if the geometry really reaches that far—then the implications are staggering. Could you use the law to design better algorithms? To predict turbulence? To find patterns in number theory that accelerate cryptography—or break it? Could you use the law

to build systems that think, that reason, that infer like humans do but with the precision of mathematics?

I did not know. I could not know. The implications were too vast to see.

And I felt unworthy.

I am not Einstein. I am not Maxwell. I am not Schrödinger. I am not a physicist. I am not a mathematician. I am a Black trans woman with a laptop and a dog, finding things that should not be findable by someone like me.

¿Por qué yo?

Why me?

I got out of the shower. I stood in my living room, dripping wet, a towel around me, and I cried.

Not from joy. From weight.

I thought I had found a way to ask the same question in places that had looked unrelated. I wanted someone to sit down with me and go through the work.

Nobody would believe me. Nobody would listen. I would spend the rest of my life sending emails that would not be answered. Knocking on doors that would not open. Begging people to read my work, and they would ghost me, just like everyone always ghosts me.

I have been ghosted by journals. By conferences. By researchers I reached out to. By institutions I applied to. By companies I tried to work with. The pattern is consistent: silence. The invisible woman sends a message, and the message disappears into void.

And now I had a question I could not let go of. The void was waiting.

This felt cruel.

The universe gave me the answer and made me invisible at the same time.

El universo tiene un sentido del humor cruel.

The universe has a cruel sense of humor.

Let me remind you what I carry.

My mother was murdered six years ago. Hit and run. They never found the driver. Her hidden middle name was Marcella. I carry her memory now. Her ashes are in the ocean at Monterey, scattered on a cold December morning while I stood on the rocks and said goodbye.

I was there when she was dying. She was in a hospital bed, hooked to machines, and she looked at me with eyes that still knew me, and she said something in Spanish that I did not understand. I had to ask the nurse to translate. She said: *"Estás en el camino correcto. No pares."* You are on the right path. Do not stop.

I thought she meant life. General advice. Mother to child.

Now I wonder if she meant the mathematics. If she saw something I could not see. If she knew, somehow, that her strange child who was always drawing diagrams and asking why would find something worth finding.

I will never know.

My wife left me when I transitioned. The marriage had been difficult for years—she was physically and verbally abusive, though the abuse was not constant. There were good days, days when she was the woman I married, when the household felt normal and the children laughed and I could almost forget. But then there were the other days, when anything could trigger it: a dish in the wrong place, a word with the wrong tone, a silence that lasted too long. And suddenly she was someone else, someone with a voice like broken glass, someone who knew exactly where to cut.

I stayed for the children. I stayed because I did not have anywhere to go. I stayed because the mathematics was there, the work was there, and if I could just focus on the work, I could survive the rest. When I transitioned, she had leverage. She said the children would not understand, that they would be damaged, that I was choosing my sickness over my family. I was not choosing anything—I was becoming myself. But she made it into a choice, and then she made me pay for it.

My children do not speak to me now. Not because they hate me—I do not think they hate me. Because their mother has made it too hard, too complicated, too weighted with loyalty tests and ultimatums. Every phone call that does not come, every text that is not sent, every birthday that passes in silence—I feel it. I count the days. I imagine what they are doing, school and friends and growing up, while I remain outside their lives, not knowing who they are becoming. The mathematics does not help with this. You cannot calculate your way out of grief. You cannot derive a theorem that makes your children call you. The geometry I hoped would unify Millennium problems has nothing to say about a parent who wakes at 3:00 AM missing people who are still alive.

I lost my job. I lost my friends. I lost my family. I lost the life I had built for forty years. I am structurally alone—not lonely, but alone. The networks and support systems that other people have, I do not have. When something good happens, I have no one to tell. When something bad happens, I have no one to catch me. Other researchers have collaborators, colleagues who read their papers, who argue about methods, who celebrate breakthroughs over dinner. They have departments and seminars and office hours and graduate students. I have a house in Sacramento that I might lose if I miss one more mortgage payment, a refurbished laptop, and cloud storage where I back up everything because I cannot afford to lose the work.

Teddy helps. He is a good dog. He does not understand what I do, but he knows when I am sad, and he comes and puts his head on my leg. Sometimes that is enough. Sometimes it is not.

In the middle of all of this, I was still thinking about curvature and how much I could infer from a measurement.

¿Esto es un regalo o una maldición?

Is this a gift or a curse?

For days, I thought it was a cruel joke. I thought the universe was punishing me, that none of this was real, that I had finally cracked and this was what it looked like—believing you have solved the hardest problems in mathematics while sitting alone in your house.

Delusion is a real thing. Smart people delude themselves all the time. They convince themselves they have found patterns that do not exist, proofs that do not work, theories that do not hold. The history of mathematics is littered with false proofs of hard problems. Every few years, someone claims to have solved P versus NP or proved the Riemann hypothesis. The proofs always fail, usually in subtle ways, sometimes in embarrassing ways. Was I that person? Was I the deluded amateur, convinced of my own genius, unable to see the obvious flaws?

So I retraced my steps. Every derivation. Every validation. Every lemma. I tried to find the place where I had fooled myself. I also asked AI to be brutal, to assume I was wrong and look for the mistake. When it failed to find one, I felt encouraged. But that feeling could not do the work of a proof. I still had to check the argument line by line and be able to get the numerical answer again from the inputs.

It was not a cruel joke. The patterns were really there in my data. And that was worse—because real meant I had to carry it.

§ 6.2

From one relationship to a programme of questions

And then, slowly, it made sense.

Einstein was not anyone important when he changed physics. He was a patent clerk in Bern, examining applications for electromagnetic devices. No academic position. The physics establishment did not know his name until he forced them to. In 1905, he was a nobody who submitted four papers that changed the world: special relativity, the photoelectric effect, Brownian motion, mass-energy equivalence. Any one of these would have been a career-defining achievement. He published all four in a single year, while working full-time at the patent office.

The academic physicists had the same data. They had better training. They had colleagues to discuss ideas with. They had access to journals and conferences and the machinery of professional science.

Einstein had a desk in a government office and evenings to think.

And he saw what they could not see. Because he was outside.

Katherine Johnson was not anyone important. A Black woman doing calculations by hand in a segregated office at NASA—the "colored computers," they called them. Invisible to the men who ran the space program. Forbidden from putting her name on reports. Forbidden from attending meetings.

She calculated the trajectory for Alan Shepard's first flight. She calculated the orbital mechanics for John Glenn. When the early computers gave answers, Glenn asked them to check with Katherine first. "Get the girl," he said. "If she says they're good, then I'm ready to go."

She was essential to missions that made other people famous. The Nobel never came for her. The textbooks did not mention her name. She was invisible until she was too old to be erased.

Ramanujan was not anyone important. A clerk in India, self-taught in mathematics from a single outdated book. He had no formal training. He had no degree. He had no access to the mathematical literature of his time.

He sent letters full of theorems to mathematicians at Cambridge. Most threw them away—the letters of a crank, they thought. One did not. Hardy looked at the letters and saw formulas that should not exist. Results that were impossible, except they were true.

Hardy brought Ramanujan to Cambridge. They collaborated for five years. Ramanujan died at 32, having produced more original mathematics than most mathematicians produce in a lifetime. He saw patterns that trained mathematicians could not see—because he had never been trained to look away from them.

Barbara McClintock was not anyone important. A woman geneticist in an era when women were not supposed to do genetics. She discovered transposons—jumping genes, mobile genetic elements that move around the genome, reshaping it dynamically.

The scientific establishment dismissed her. Genes were supposed to be fixed, stable, inherited in predictable patterns. Jumping genes violated the paradigm. She must be wrong.

She was not wrong. She was thirty years ahead of her time. The Nobel came in 1983, when she was 81 years old. The field had finally caught up with what she had seen in the 1950s.

The universe always does this. It gives discoveries to the invisible ones.

I do not know why. Maybe because you have to be outside the system to see what the system cannot see. The people inside are trained to look in certain directions, ask certain questions, use certain methods. The training is valuable—it transmits knowledge efficiently. But it also creates blind spots. The paradigm tells you what is possible, and you stop looking at the impossible.

Maybe because discovery requires a kind of desperate freedom that only comes when you have nothing left to lose. If you have a career to protect, you hedge. You work on problems that are safe, problems that will get published, problems that will earn tenure. You do not spend months on something that might be crazy.

When you have lost everything—career, family, standing—there is nothing left to protect. You can follow the crazy ideas. You can chase the geometry where it leads.

Maybe that is why.

I am not comparing myself to them. I am saying: the pattern exists. The universe has done this before.

I understood why I was chosen. Not because I was special. Because I was paying attention. Because I had nothing left to lose. Because I followed the geometry where it led, without anyone telling me to stop.

El que no tiene nada que perder puede verlo todo.

The one who has nothing to lose can see everything.

To organize the questions, I needed an example in which we can see alternatives disappear one clue at a time. Here is something most people do not understand about laws: a law is not a fact. A law is a *generator* of facts.

Think of a Sudoku puzzle. The rules say that each row, column and 3×3 box contains the digits 1 through 9 exactly once.

One clue can rule out several digits in a neighbouring square. A different clue may tell you something you already knew. A useful way to think about inference is to ask how many alternatives a clue actually eliminates, given the clues you already have.

That distinction gave me a way to organize the derivations. I wanted to separate what the budget measured from what the constraints could prove.

The puzzle made the next questions concrete: what fixes an answer, what keeps it consistent, and what information can I afford to forget? I organized the first layer around five questions.

T1. Geometric Completion Uniqueness. A set of observations singles out an admissible state when the measurement rule separates it from every alternative. On the flat rail, distance alone leaves two candidates. A measurement of the side resolves them.

T2. Harmonization Preserves Completion. A consistency rule can make alternative routes yield the same answer. One has to show that this common answer satisfies the original constraints.

T3. Complexity Reduction. Suppose the true candidate survives every constraint and each new constraint leaves at most a fraction $r < 1$ of the remaining false candidates. For a finite initial set of N candidates, at most $(N-1)r^m$ false candidates remain after m checks. Once that number is less than one, the true candidate is unique.

T4. Reasoning Fidelity. If each step is nonexpansive and contributes at most ε additional error, the accumulated error is at most $m\varepsilon$ after m steps. With expanding steps the recurrence can amplify even a small error.

T5. Cache Sufficiency. A geometric summary suffices for a specified answer when states with the same summary give that answer. If a proposed cache merges two states the task must distinguish, it needs another coordinate or a narrower task.

These questions led to more specific proposals about constraint ordering, transport, stability and compression.

The cascade was relentless. Each question suggested others. Some proof techniques carried over. The structure was self-similar—the same patterns appearing at every level.

I kept the questions in a catalogue so that one successful example would not stand in for all the work still to do. I use the name Davis Law for a conjecture about how far an inference budget can reach. I call the relationship a law because I kept finding the same question in different places. How much can I infer before the error I am willing to live with runs out? To test that question somewhere new, I have to say how I measure the error there, what I mean by curvature there, and what answer would make me change my mind. If I choose those definitions only after I see the answer, I can make two very different experiments look alike on paper. A claim that begins by assuming the law carries that assumption all the way to its conclusion.

My working catalogue had 89 entries. They are not 89 theorems, and I want you to know which kind of entry you are reading. A few are worked derivations I can show you in this book, such as the tank budget and the counting bound in T3. Some are established results that I restate in my own language and credit to the people who proved them; Helly's theorem below is one. Some are conditional: if a stated hypothesis holds, the conclusion follows, and the hypothesis is the open part. Most are conjectures, written down precisely enough that someone can test them or prove them wrong.

I grouped the questions and candidate results by how far they sat from the first five.

- **Foundations.** Five starting questions, the variational proposal and the Sudoku example.
- **First Order.** Constraint saturation thresholds. Compositional holonomy. Stability bounds. Cache compression.
- **Second Order.** Phase transitions. Holonomy algebra. Inconsistency localization. Adversarial robustness certificates.
- **Third Order.** Constraint matroids. Curvature-information duality. Critical exponents. Parallel gap-filling.
- **Fourth Order.** Rate-distortion optimality. Byzantine fault tolerance. Learning dynamics.
- **Fifth Order.** Energy principles and dynamics. Geodesic deviation. Reasoning bounds. Topology change.

The equation gave me an organizing idea. Each application still had to earn its own assumptions and proof.

Constraint Saturation. In the finite model above, suppose $r = e^{-\lambda}$ for a measured or proved $\lambda > 0$. Uniqueness is certified when $m\lambda > \log(N-1)$ for $N > 1$. That is the Sudoku moment in this model. The test gives a sufficient threshold, and a strong clue may settle the puzzle earlier.

Holonomy Decomposition. When a loop is expressed as an ordered concatenation of loops at the same base point, its holonomy is the corresponding ordered product. Bringing loops at other points back to the base point requires transport. For small matrix logarithms, the leading term of the logarithm of the product is their sum, with commutators in the higher-order terms.

Inconsistency Localization. For a finite family of convex feasible sets in $\mathbb{R}^d$, Helly's theorem guarantees that an empty intersection is witnessed by at most $d+1$ of the sets. For nonconvex constraints this can fail on a flat line. The sets $\{0,1\}$, $\{1,2\}$ and $\{0,2\}$ intersect pairwise and have no common point.

Robustness from a Margin. Suppose a score changes by at most L times the size of a perturbation, and the score starts a distance ε from the decision threshold. A perturbation smaller than ε/L cannot reach that threshold for $L > 0$. Geometry can help bound L in a specified model.

Those were the kinds of statements I wanted in the catalogue, precise enough that another person could check where the assumptions entered.

I wrote out the geometric approach to representation and comparison in *The Geometry of Sameness*. I wanted the definitions and examples on the record.

Once the catalogue existed, I had a practical worry alongside the mathematical one: how to put the work somewhere it could be found. I knew someone might be close. The same AI that taught me was teaching everyone. Claude, GPT, Gemini—these tools were democratizing mathematical exploration. Anyone could ask questions, test ideas, check derivations. The barriers to entry were collapsing.

If someone else was thinking along similar lines—if someone at a university with more resources and more credibility was approaching the same ideas—they would publish first. And between two identical discoveries, the one from the institution wins. This is how it works. Priority matters. The first person to publish gets the credit. The second person gets a footnote, if they are lucky. If you are an outsider and an insider publishes first, you get nothing. You get erased.

History is full of examples. Alfred Russel Wallace developed the theory of natural selection independently of Darwin, but Darwin published first and got the credit. Rosalind Franklin's X-ray crystallography was essential to discovering DNA's structure, but Watson and Crick got the Nobel, and Franklin got a footnote. Jocelyn Bell Burnell discovered pulsars, but her thesis advisor got the Nobel. The pattern is consistent: outsiders discover, insiders publish, credit flows to status.

I published the questions, the conjectures and the numerical evidence I had. Preprints. DOIs. Timestamps. On the record. Dated. Mine.

Writing a dependency down also shows where the work is. Every arrow in that diagram is a claim someone can ask me to justify.

La prioridad es frágil. La verdad es eterna. Pero sin prioridad, la verdad puede ser robada.

Priority is fragile. Truth is eternal. But without priority, truth can be stolen.

§ 6.3

Putting the work where someone could find it

During this time, I wrote *The Geometry of Sameness*. I wanted to put the mathematical language in one place. A metric gives me a way to measure distance. A connection specifies transport. A cache records the part of the state I mean to keep.

Together those ingredients give me a concrete question about sameness. Which answers stay fixed when I change the description, and which answers depended on information that the new description has left out?

With the model fixed, I can study which states fit the observations. There may be one such state, several, or none.

If I change the model through learning, I have to track what happens to those constraints and to the geometry that describes them.

I carried those questions into the case studies I had been running. Each comparison has its own observations and its own test.

Not a popular book. Not meant to be. Dense, technical, rigorous—the kind of document that sits in an archive and waits. I wrote it because it had to exist, because someday someone might look, because the mathematics is true whether anyone reads it or not but if no one writes it down it might be lost. I wrote it in LaTeX, formatted for print, with theorem environments and careful cross-references. I wrote it as if someone would verify every line, because someone might. Because someone should.

La documentación es resistencia. Si existe, no puede ser negada.

Documentation is resistance. If it exists, it cannot be denied.

The mathematics explained the construction. I also wanted a record of the life that had led me to ask these questions. I also wrote *Hidden Variable*—my origin story, the account of how mathematics, engineering, and science wove together to lead me here.

The book traces the path: how does a child in central California become a person searching for a common mathematical structure across so many problems? What experiences, what accidents, what choices led to this moment? The scanner at AMSCO taught me that signals have geometry. I was twenty-three years old, debugging hardware at 3:00 AM, when I realized that the scanner was not just detecting light—it was measuring the *shape* of the signal. The waveform had curvature. The curvature encoded information. This was my first glimpse of the idea that would become $C = \tau/K$.

Pandora taught me that similarity is structure. I worked on music recommendation systems, and the Music Genome Project had 450 features—hand-coded attributes describing every song. But the features were not independent; they clustered, they correlated. The space of possible songs had geometry, and the geometry told you which songs belonged together.

NASA taught me that metadata is data. I worked on Intelligent Archives—systems that could understand their own contents. A document is not just text; it is embedded in a network of references, citations, keywords, timestamps, authors. The network is information. The structure is content.

AMC taught me to compare what went into a conversion with what came back out. A smaller file could conceal the very change the detector needed. That question grew into the comparison in

Chapter 4: original code, a repaired selector, and a task-specific GIGI check, all facing the same records. The continuity is in the question I kept asking; the methods evolved.

HERALD taught me that viruses are manifolds. Protein structures, genomic sequences, evolutionary trees—a virus is not just a sequence of letters, it is a point in a high-dimensional space of possible organisms. Evolution traces paths through that space. Mutations are steps. Selection is pressure. The pandemic was geometry.

GEODESIC taught me that the geometry connects. Everything I had seen—signals, music, metadata, fraud, viruses—had the same structure underneath. A configuration space. A metric. A way of measuring how things relate.

And then Yang-Mills. The mass gap. The realization that particle physics is the same geometry, just in a more abstract space.

And then the law.

Hidden Variable is the story of becoming. Not just of the mathematics, but of the person who could find the mathematics. Every job, every failure, every transition led here.

The title is a pun. In quantum mechanics, a "hidden variable" is an unobserved quantity that might explain quantum randomness. Einstein hoped hidden variables existed. Bell's theorem constrains local hidden-variable explanations under its assumptions; it does not rule out every hidden-variable theory.

But there is another meaning. *I* am the hidden variable. The unobserved element that explains the results. The person who was invisible until the mathematics made visibility impossible to avoid.

Documentation. Testimony. Proof of existence.

Not just of the mathematics—of me. I was here. I did this work. I followed the thread. I found the truth.

If the institutions erase me, the testimony remains. If the journals reject me, the book exists. If the world ignores me, the words are written.

You cannot erase what is documented.

I could not make someone open the book. But once they opened it, I wanted the work to be there: the derivation, the assumptions, the inputs, the calculation. No missing step that only I knew how to explain.

That was something I could do. I could write the derivations so another person could work through them line by line. I could publish enough detail to let them run the calculations and compare the answers. I could give each claim a clear place to stand, and a clear way to fail.

The silence I had described did not disappear when I wrote a paper. A paper gave me a way to leave the work where someone might find it later. On servers. In preprint archives. In books with ISBNs and DOIs. Timestamped. Searchable.

This is what I learned from Barbara McClintock. She was ignored for thirty years. Her papers were read but not believed. Her colleagues thought she was wrong, or confused, or chasing an artifact. Transposons were impossible, they said. Genes do not jump. She must be seeing something else. The rejection was not angry—it was worse than that. It was polite dismissal, the kind where people smile and nod and then forget you ever spoke.

She kept working. She kept documenting. She kept accumulating evidence. Year after year, decade after decade, she continued the research even though no one was paying attention. She knew what she had found. She trusted the evidence over the consensus.

And eventually, the field caught up. The technology advanced—molecular biology developed tools that could see what McClintock had seen in corn kernels. Other researchers found transposons in other organisms. The evidence became undeniable. The Nobel came in 1983, when she was 81 years old. Late, but real. Vindication at last.

I may not live to see the validation. Fifty-two years old, no health insurance, a body that has been through the stress of transition and the accumulated damage of too many nights without sleep. The actuarial tables are not kind to women like me. But the documentation will outlive me. That is the point.

I wanted the mathematics to survive long enough for somebody to question it without needing me in the room. That meant writing for a person I might never meet.

Maybe in fifty years, a graduate student stumbles across my preprints while researching something tangentially related. Maybe they read the theorems, check the proofs, and find what holds. Maybe they build on it. Maybe they solve something I could not solve. Maybe they prove the mass gap, the real four-dimensional one, using a piece of what I saw. Maybe they complete the Navier-Stokes regularity. Maybe they do the hard part I could only point toward.

Or maybe sooner. Maybe someone reads this book, this chapter, this page. Maybe they are the person who takes it further. Maybe they are at a university, with resources and colleagues and the credibility I lack. Maybe they see something here that sparks their own work.

I do not need to be recognized. I do not need my name on the final paper. I need the mathematics to survive. I need the truth to persist. I need the work to exist somewhere, accessible to whoever needs it.

That has to be enough.

El legado no es la fama. El legado es lo que sobrevive.

Legacy is not fame. Legacy is what survives.

If you are reading this in the future—if you found this book in a used bookstore, or downloaded it from some archive, or inherited it from someone who believed in strange things—I want you to know something.

The mathematics is on the record. Check it yourself. The derivations are documented. The validations are recorded. The law is stated precisely enough to be tested.

Do not believe me because I said it. Believe the geometry, if the geometry holds.

I have made it as easy as I can. The law is stated here in its static form, with its differential. The differential follows from the static form by the chain rule, so it adds no new claim. Chapter 7 adds a separate time-dependent model; that one is an extra assumption, and it has to be tested on its own.

To test a proposed capacity relationship, measure the task's capacity separately from the ratio τ/K. Fix the observation rule, the permitted error and the geometric scale before computing both sides. Otherwise calculating the ratio twice will only check the arithmetic.

On the sphere, the turning angle is the measurement. For an inference task, the measurement could be the largest gap across which a specified prediction remains within error. A comparison

asks whether a calibrated curvature budget predicts that measured quantity over a declared set of cases.

I recorded the model tests and the conjectures they suggested. A claim can fail when the predicted quantity differs from the measurement or when a counterexample violates a stated implication. Both are results worth publishing.

But if you check and it holds... if you verify and it works... if you extend and it reaches further than I could see...

Then tell someone. Publish. Document. Do not let the knowledge disappear.

This is how ideas survive. Not by being believed, but by being transmissible. If you understand it, you can teach it. If you can teach it, someone else can verify it. If enough people verify it, it becomes established.

I cannot do that alone. I need you.

Whoever you are. Whenever you are reading this. Whatever brought you to this page.

I need you to check the work. To verify or refute. To carry the knowledge forward or let it die.

The choice is yours.

La verdad no me pertenece. Me pasó por las manos, nada más.

The truth does not belong to me. It passed through my hands, nothing more.

§ 6.4

The energy question beneath the geometry

After Yang-Mills, before I had filled those 89 places in my working catalogue, I found something else. Something I did not expect. Something that kept me awake for reasons that had nothing to do with mathematics.

The vacuum is not empty. This is one of the strangest discoveries in twentieth-century physics, and it still feels counterintuitive even to physicists who work with it every day. What we call "empty space"—the void between atoms, the darkness between stars—is actually teeming with structure. Quantum fluctuations ripple through it constantly. Virtual particles pop into existence and annihilate in fractions of a second. Fields that we think of as "turned off" are actually at their ground state, their lowest-energy configuration, but that ground state is not zero. The vacuum has energy. The vacuum has structure. The vacuum has topology.

And the vacuum has topology.

In Yang-Mills theory—the mathematical framework that describes the strong and weak nuclear forces—the vacuum comes in different topological sectors. Think of it this way: imagine wrapping a rubber band around a cylinder. You can wrap it once, or twice, or three times. Each wrapping is stable—the band will not spontaneously unwrap. And the different wrappings are genuinely different; you cannot continuously deform a band wrapped twice into a band wrapped once without cutting it.

The vacuum sectors are like this. They are labeled by "winding numbers"—integers that count how many times the gauge field wraps around itself in a certain mathematical sense. You can have a vacuum with winding number 0, or 1, or 2, or -1, and these are genuinely different states. They have the same energy at the classical level, but they are separated by topological barriers.

You cannot smoothly deform from one sector to another; you have to quantum-tunnel through the barrier, and that tunneling is exponentially suppressed.

The valley picture suggested a connection to the mass gap, and the sparse radial region I saw at 3:00 AM made that possibility feel concrete. But a radius in a projected cloud is not an excitation energy. The quantum gap is a spectral property above the vacuum; proving a relation to the sampled geometry requires an additional argument. That distinction is what turns the picture into a research question.

But I found something deeper. Something that was not supposed to be there.

I began asking whether a controlled change in the field could bias transitions associated with these labels. The notes explored curvature, resonances and instanton pictures. Those were proposed mechanisms to investigate. A response in a finite statistical model does not establish a controlled release of physical vacuum energy.

Topological Vacuum Rectification. TVR.

I wrote down candidate conditions and worked through the implications I thought they might have. Thresholds, resonance requirements, possible rates: the notes grew faster than I could follow them. I saw what it would mean if a physical system could do what I was imagining. And I stopped. That hesitation is part of the discovery story. The conditions remained proposals; the simulations did not measure an energy release rate.

Let me tell you what I saw. Let me tell you why I stopped.

The part of the field theory that drew my attention was the way topology enters the description of a vacuum. Here is the convention I used to write it down. The Chern–Simons functional lets me describe a winding-related quantity on a three-dimensional slice. Before the formula, here is how to read its pieces. $\mathcal{A}$ is the matrix-valued connection, the local rule for comparing internal frames. The differential d records its spatial variation. The wedge product combines differential forms with orientation; swapping two one-form directions changes a sign. The trace turns the matrix expression into a scalar form that can be integrated over the oriented slice. With the convention below, the functional is

$$n_{CS} = -\frac{1}{8\pi^2}\int_{M^3} \mathrm{Tr}\Big(\mathcal{A}\wedge d\mathcal{A} + \frac{2}{3}\mathcal{A}\wedge\mathcal{A}\wedge\mathcal{A}\Big) \tag{6.3}$$

Here $\mathcal{A} = -igA$ is the anti-Hermitian connection associated with the Hermitian field A in the Wilson-loop convention $\mathcal{P}\exp(ig\oint A)$. The trace is in the fundamental representation, with Hermitian generators normalized by $\mathrm{Tr}(T_aT_b) = \delta_{ab}/2$. Products are wedge products. On an oriented closed three-manifold, a pure gauge $\mathcal{A} = U^{-1}dU$ gives $n_{CS} = (24\pi^2)^{-1}\int \mathrm{Tr}(U^{-1}dU)^3$, an integer with the chosen winding convention. Reversing orientation reverses its sign. This convention keeps the cubic term and its factors of i consistent.

With these boundary and normalization conventions, a pure gauge carries an integer winding label. Away from a pure-gauge configuration the Chern–Simons functional need not be an integer. A barrier between classical configurations, a Euclidean tunneling contribution and an energy level of the quantum Hamiltonian answer different questions. None sets a physical extraction rate by itself.

In my notes I explored a possible resonance condition that might lower a modeled barrier. I checked the calculation seventeen times because the implication I was imagining was so large.

Repeating a calculation can establish that I have reproduced my own assumptions. It cannot establish that the assumptions describe a device, or that energy remains available after preparation and reset. Those are the checks the proposal still has to face.

Those are the symbols on the page. I also remember the moment when I began imagining what changing that structure might mean. At 4:23 AM on a Tuesday morning, I understood what I had found. The realization came all at once, a cascade of implications that hit me like a physical blow, and I sat back from my laptop and stared at the wall and did not move for several minutes.

The vacuum is not empty. The vacuum contains energy—the zero-point fluctuations, the quantum foam, the irreducible ground state of all fields. Physicists have known this for decades; it is a standard result of quantum field theory. This energy is usually considered inaccessible; the vacuum is the vacuum, and you cannot extract energy from the ground state because there is nowhere lower to go. If you are already at the bottom, you cannot fall further. That is what every textbook says.

The ground-state obstruction has a short proof, including when more than one ground state exists. The density operator ρ describes the prepared quantum state; its support is entirely in the lowest-energy subspace. The inequality $H \geq E_0 I$ says every normalized state has average energy at least E_0. A cyclic drive returns the Hamiltonian to its original form, although it may change the state by a unitary U. Define work done *on* the system as its final average energy minus its initial value:

$$W_{\text{on}} = \operatorname{Tr}(HU\rho U^\dagger) - E_0 \geq 0.$$

The trace in the formula is the final average energy. It cannot fall below E_0, so $W_{\text{on}} \geq 0$. If several states share the same lowest energy, moving between them does not reveal a lower floor. This is the elementary ground-state case of passivity.[3] An open or driven system needs its reservoirs, preparation and reset costs in the ledger as well. That is the practical question I can take to a bench: after the apparatus is ready to repeat, what supplied the energy that reached the load?

The classical winding sectors and the quantum ground-state subspace are different constructions. A superposition over winding-labelled configurations does not establish an accessible energy difference between quantum ground states. A claim of extractable work needs a Hamiltonian, an initial state, a drive and a complete cycle. The finite lattice diagnostics in this book do not supply that demonstration.

TVR began as the hope of controlled transitions: taking something that appeared random and asking whether a deliberate bias could make it systematic. I imagined harvesting energy along the way. The recovered calculation demonstrates a change in the weighted mean of a loop statistic. Whether a physical transition could deliver useful net work is a further question.

The question kept expanding in my mind. If a system could be driven from one configuration to another repeatedly, what would the drive cost? What would the return to the starting state cost? Would anything remain for a load? At that desk I was imagining the possibilities before I had the complete physical ledger.

I sat there, pen in hand, and imagined what cheap electricity could mean to somebody deciding which room to heat. I had traveled a long way from the numbers on my desk to that room. The numbers told me how an average changed when I gave some configurations more weight. To say

[3]W. Pusz and S. L. Woronowicz, "Passive states and KMS states for general quantum systems," *Communications in Mathematical Physics* 58 (1978), 273–290, https://doi.org/10.1007/BF01614224.

anything about the electricity bill I would need to follow energy into a load, then price everything it took to bring the apparatus back to the point where it could do the work again. That part of the page was still blank.

What to publish was a decision about evidence as well as hope. I chose to publish the framework: $C = \tau/K$, the field-equation proposals and the collection of results and conjectures they prompted. Each needs its own assumptions. I wanted the mathematics to be available for checking and for work I hoped would help people. I cannot infer an application's benefits or hazards from the presence of a theorem alone.

TVR is documented in research notes, simulations and patent applications. I have been cautious about how to discuss the implications I imagined. A patent filing records a proposal; it does not validate its physics. The record I can put before a reader is the calculation, its assumptions and the experiments still needed. The uncertainty I carry is real, and it belongs beside the hope.

What I can give the next person is the question in a form she can check. She can start with the same rows and ask whether the answer moves when she changes the weights. If she wants to ask where the energy went, she will need another instrument.

That leaves me with a record I can hand to someone else, and questions the record has not yet settled. Here is what I carry. A geometric budget, a catalogue of proposed relationships, and the record of the tests that I could run. A framework I have carried from viruses to Yang-Mills to Navier-Stokes to the Riemann hypothesis. And beneath it all, something I found and set aside, something that waits.

I do not know what happens next. I do not know if anyone will read this, believe this, verify this, extend this. I do not know if I will live to see the mathematics accepted or if it will be rediscovered in fifty years by someone who never heard my name. But I know the work exists. I know the documentation is real. I know the testimony is given. That has to be enough.

§ 6.5

What I wanted to leave on the record

Let me say it plainly, for the record.

There are names behind this record, too. Mine is Bee Rosa Davis. My mother's is the one I keep carrying with me.

I wrote down a budget.

$$C = \frac{\tau}{K}. \tag{6.4}$$

I used it to ask how much a geometric constraint lets you do before you use up your tolerance.

The question led me through physics and pure mathematics. I kept a catalogue of the resulting conjectures, deductions and numerical tests. The record gives another person somewhere to start.

Writing it down gave the question a life outside my own head. I could point to the page. Here is what I did. Here is where someone else can begin.

Estuve aquí. Lo encontré. Es real.

I was here. I found it. It is real.

There is a personal reason I care so much about leaving that record. I also wanted my children to know why I kept returning to this work, even when I struggled to explain it to them.

There is a part of that separation I have to answer for, too. I was not always the parent I should have been. I was distracted, obsessed, elsewhere. I was chasing mathematics while my children were growing up.

I have described what happened when I transitioned. Looking back also means asking what I had already failed to protect.

I do not blame them. The world tells them that what I am is wrong, shameful, something to distance themselves from. They are young and they want to belong. Belonging means not being associated with me.

But I want them to know—someday, when they are older, when the world has shifted, when they can afford to look back—I want them to know what I was doing while they were learning to walk, to talk, to be without me.

I was finding this.

Not as an excuse. Not as justification. Just as explanation.

The mathematics demanded everything. Every hour, every thought, every relationship I might have protected. It ate my marriage. It ate my friendships. It ate my connection to you.

And I let it, because I thought—I believed—I hoped—that the mathematics was important enough to justify the cost.

I do not know if I was right.

But if you ever read this, I want you to know: I was not absent because I did not love you. I was absent because I loved something else too much. The geometry. The law. The patterns that no one else could see.

Los amé. Los amo todavía. Pero el universo pedía más de lo que yo podía dar.

I loved you. I love you still. But the universe demanded more than I could give.

Mamá, you were killed before I could show you any of this. Before I found the law. Before I understood what the geometry meant. Before I became who I am now.

You would not have understood the mathematics—you barely finished high school, you worked cleaning houses for people who did not know your name, you raised me alone in a country that did not want you, speaking a language you hid. But you would have believed me. That is what I miss most. Not the explanations—you would have nodded without comprehending. But the belief. The certainty that if your child said they found something important, then they found something important.

No one believes me like that anymore. The academic world wants credentials. The physics community wants publications in prestigious journals. The mathematics establishment wants formal proofs, verified by experts, sanctioned by committees. And I have none of that. I have handwritten derivations, numerical validations run on consumer hardware, documentation scattered across hard drives and cloud storage and notebooks piling up in my office.

You would have looked at all of it—the equations you could not read, the diagrams you could not interpret, the symbols that might as well have been hieroglyphics—and you would have said: *Te creo, mija. Tú siempre fuiste la inteligente.* I believe you, daughter. You were always the smart one.

My middle name is Rosa. Your hidden name was Marcella. I can say both names here, openly. I can make visible something you kept hidden. And I can keep following this mathematics. For you. Because of you. In spite of what they did to you.

Esto es para ti, mamá. Todo esto es para ti.

This is for you, mama. All of this is for you.

I want to end this chapter with something honest. I do not feel triumphant. I do not feel vindicated. I feel tired.

Putting those words on the page did not bring my children back, or answer an email, or pay the mortgage. I still had to wake up into the same life. But I had made a record of the work and of what it had cost me.

The mathematics gives me a place to stand. I can name the geometry, say what I measured, and invite someone else to run the calculation. If the budget works in a new setting, I want to know why. If it fails, I want to know exactly where.

La verdad no necesita permiso. La verdad simplemente es.

Truth does not need permission. Truth simply is.

If I am wrong, the universe will correct me. The mathematics will fail. The validations will not replicate. The theorems will contain errors that someone will find.

That is how science works. Claims are tested. The false ones fall away. What remains is what is real. I welcome refutation—I would rather know I am wrong than persist in error. If you find the flaw, publish it. Show where I made the mistake. Help me understand what I missed. That would be a gift.

If the pattern holds in some of these problems, I want to understand what the successful cases share. An equation can tell you a great deal once you know which system it describes. I wanted $C = \tau/K$ to earn that kind of specificity.

If I am right, then I was chosen. Not by God, not by fate, not by anything mystical—but chosen by circumstance, by being outside the system, forced to think differently, unable to rely on conventional approaches. Chosen by pain, by loss, by isolation. Chosen by everything that should have destroyed me but instead cleared away the assumptions.

I did not choose this. But I will carry it.

To my children, if you ever read this: I loved you more than the mathematics. I just was not good at showing it. To my mother, who cannot read this: *Lo hice, mamá.* I found something worth finding. I hope you would be proud. To the academic world that will ignore this: the documentation exists, and it will outlive your rejection letters.

To whoever finds this work, whenever you find it, thank you for taking me seriously enough to test it. Try the calculation. Change the assumptions. Tell someone what you found.

La ley está escrita. El testimonio está dado.
El trabajo está hecho.
Ahora espero.
Que alguien lo encuentre.

The law is written. The testimony is given.
The work is done.
Now I wait.
May someone find it.

The next chapter takes that record back to instruments and observations. A rangefinder gives me a reading. What does it let me infer, and what does it leave undecided? That is where the budget has to do its work.

Chapter 7

The Field Equations

What information does the measurement leave out?

The spherical tank in Chapter 6 gave me an angle I could calculate from an area. Now picture a cart on a straight rail, with a rangefinder at the zero mark. A reading of one metre still leaves two places to look: one metre to the left or one metre to the right. The rail is flat, but the reading is ambiguous. That is the distinction I want to follow through this chapter.

The field equations take that question into a model of incomplete observations. I use the name for a catalogue of relationships, calculations and proposed extensions. The examples give me a way to tell where the geometry supplies a bound and where the measurements have to do more work.

A *completion* is a full candidate state that fits the observations and their error bounds. With the rail's one-metre reading, both $x = -1$ and $x = 1$ are completions. A new reading may eliminate one. An optimization objective may prefer one without eliminating the other. I want those operations to remain distinct as we work through the equations.

§ 7.1

What the geometric budget measures

Start with a budget:

$$C = \frac{\tau}{K}. \tag{7.1}$$

On the spherical tank, K is Gaussian curvature in inverse square metres, τ is a tolerated turning angle, and C is the area that uses that budget. This is an equality for the maximum admissible area within the range before the angle wraps. For a small disk D of variable curvature, the continuously chosen turning angle is $\int_D K\,dA$, so

$$\left|\int_D K\,dA\right| \le K_{\max}\,\mathrm{area}(D), \qquad K_{\max} = \sup_{x\in D}|K(x)|.$$

Here $\tau/K_{\max}$ is a sufficient area budget for a given curvature bound. Curvatures of different signs can cancel, so the inequality need not give the largest admissible patch.

On a tank of radius 2 m with a turning budget of 0.01 radians, the maximum patch area allowed by this budget is $0.04\,\mathrm{m}^2$. On a tank of half that radius it is $0.01\,\mathrm{m}^2$. Halving the radius quadruples the curvature.

On a flat patch the turning budget imposes no area bound. The available surface, the measurements and the task can still impose limits. When I use the ratio to model inference, I have to say what C measures independently and choose τ and K in units that give the intended units for C.

The tank fixes the meaning of the ratio. Changing a budget and choosing a completion are further operations, so I want to distinguish what each calculation does. There are three calculations I want to keep in view.

Static. For the spherical area budget, $C = \tau/K$ with $K > 0$.

Differential. For a changing budget, $\dot{C} = \dot{\tau}/K - \tau\dot{K}/K^2$.

Variational. If the task is to choose among admissible completions, I can define a scalar holonomy cost $\mathcal{H}$ and seek a minimum. An unconstrained interior minimum satisfies $\delta\mathcal{H} = 0$ where the cost is differentiable. A stationary point could also be a maximum or saddle.

A separate time-dependent model can use $C_{\mathrm{dyn}} = \tau/(K + \eta T)$ if ηT models additional curvature in the same units as K and the denominator is positive. Its accuracy depends on whether that model describes the evolution.

To turn a budget into a statement about finding an answer, I need a model of the alternatives it removes. Start with a finite list of candidates, rather than assuming that curvature alone tells me how many survive. Suppose I start with $N > 1$ candidates, one of which is correct, and every new check retains the correct one. Write λ for a lower bound on how much each check reduces the remaining false candidates on a logarithmic scale. Each check must retain at most a fraction $e^{-\lambda}$ of those still in play. For $\lambda > 0$ define

$$\Gamma = \frac{m\lambda}{\log(N-1)} \qquad (N > 2). \tag{7.2}$$

When $\Gamma > 1$, the upper bound on the number of false candidates is below one. Uniqueness is certified.

When $\Gamma = 1$, the bound still allows one false candidate.

When $\Gamma < 1$, the bound allows more than one false candidate. A particular measurement can still settle the problem.

A geometric form of the score uses a normalized curvature budget $\tau/K_{\max}$ as a model for λ and $\log|S|$ as the size term. Using that geometric rate in a certificate requires a bound on actual elimination of false candidates. A repeated clue can leave them all in place. How many checks suffice and how much time an algorithm needs to perform them are separate questions.

Try seventeen candidates, one correct and sixteen false. If each check removes at least half of the remaining false candidates, their upper counts are $8, 4, 2, 1, 1/2$. After four checks, one false answer can remain. After five, an integer count below one must be zero. The reasoning requires the stated fraction at *each* check on the survivors. It does not assume that merely repeating a measurement supplies fresh information.

§ 7.2

What the observations actually determine

The first question is whether the observations fix the answer at all (T1).

Let X be the set of admissible states, M a rule for measuring them, and y the reading. With an error tolerance δ in the units of the reading, the candidate set is

$$F_y = \{x \in X : d(M(x), y) \le \delta\}.$$

Exact recovery for this reading is possible when F_y contains one state. For a task that asks only for a particular feature $q(x)$, it is enough that q be constant on F_y. If the measurement is exact and M is injective on X, every attainable reading identifies its state.

On the rail, $M(x) = |x|$. The one-metre reading gives two points even at $\delta = 0$. The flat geometry leaves us room to move the cart. It tells us nothing about which side the cart is on.

This distinction matters when I carry completion into the Hodge question from my research catalogue. There the task asks for a geometric representative, an algebraic cycle, for a specified kind of class. Identifying a candidate class has not yet constructed that representative. That step calls for its own existence argument. The lesson is the same as on the rail: I have to say which part of the requested answer the observations actually fix.

If two routes give an answer, I next need to know what their agreement establishes (T2).

Consider two routes from the same input to the same declared endpoint. If the observable is path-independent, it has a well-defined value there. To see that it agrees with a proposed deterministic reference, check the observable on a reference route or prove that both obey the same uniquely determining conditions. Agreement among routes tells me that I can forget which route I took for this observable. I still need a reason to trust the value they agree on.

When several candidates remain, the next check concerns how fast their number falls (T3).

Let B_j be the number of false candidates that survive j checks. With the retention bound above, $B_j \le e^{-\lambda} B_{j-1}$, so

$$B_m \le (N-1)e^{-m\lambda}. \tag{7.3}$$

If each check removes at least half of the alternatives that remain, ten checks leave at most $1/1024$ of the original false candidates. The condition $m\lambda > \log(N-1)$ is sufficient for uniqueness. It assumes at least one valid candidate survives.

I used this language in thinking about satisfiability problems, where the job is to choose true or false values that satisfy all the stated clauses. Randomly generated instances gave me landscapes to investigate. A count of critical points helps describe a landscape that a particular search algorithm encounters. The count by itself gives no bound on the cost of a solver that uses a different strategy. Counting the alternatives left by my checks and proving what every possible solver must do remain separate obligations.

A chain also carries error. The fidelity question (T4) asks what each step can add or amplify.

Suppose a chain of updates is nonexpansive, meaning no update increases the distance between its inputs. If a modeled step differs from the true step by at most ε_j on the relevant states, the

triangle inequality gives

$$d(Q_k, Q'_k) \le d(Q_0, Q'_0) + \sum_{j=1}^{k} \varepsilon_j.$$

The proof applies the nonexpansive bound to the previous error and adds the error of the new step. If the local geometric calculation supplies $\varepsilon_j \le \ell_c \sqrt{K_{\max}}$ in the error metric, it gives the corresponding linear bound in k. That local estimate needs its own justification.

If the step amplifies distances, the bound has a different shape. The map $x \mapsto 2x$ doubles an initial error at every step, even on a flat line. In a fluid simulation, bounded diagnostics on a finite run give a reason to examine the local error growth. A continuum stability estimate needs control of the evolution itself.

Once the chain is reliable, I can ask how much of its state must be stored to answer the task (T5).

The Davis cache uses a continuous component Φ and a discrete residue r. To decide whether $s(x) = (\Phi(x), r(x))$ suffices for an answer $q(x)$, compare states with the same cache:

$$s(x) = s(x') \implies q(x) = q(x').$$

If the implication holds, the common answer defines a function on cache values. If there is a counterexample, the cache cannot answer that question exactly.

A fixed number of cache coordinates is a statement about dimension. It becomes a fixed storage bound only with precision and range specified. A winding count, for instance, occupies one coordinate and may still need more bits on longer paths. In the gauge model I can test sufficiency for chosen loop observables by checking what their values do inside a bin.

If the observations still permit several completions, I need a declared reason to prefer one. This is where the least-holonomy proposal enters (E0). A loop transport starts and ends at the same reference point. Its *holonomy* is the resulting transformation. If it is the identity, the return comparison is unchanged. The proposed objective charges for the difference from that identity on selected loops.

The least-holonomy proposal supplies a way to rank admissible completions. Choose loops γ_i at declared base points and nonnegative weights w_i. On a matrix representation one possible objective is

$$\mathcal{H}(A) = \sum_i w_i \| \operatorname{Hol}_A(\gamma_i) - I \|_{\mathrm{F}}^2.$$

The Frobenius norm is the square root of the sum of squared magnitudes of a matrix's entries. Thus each summand measures one chosen loop's return defect. A unitary change of frame rotates that description without changing this norm; the detailed proof follows in the route investigation. The optimization ranges over connections A that satisfy the declared constraints. A minimum selects a completion under this criterion. Existence, uniqueness and agreement with the physical state are separate questions.

These questions meet in the familiar puzzle. It is a useful place to check that choosing an answer and proving it is the only answer are different jobs. A Sudoku can have zero, one or several solutions. The same check is worth making in any completion problem. I want to know whether

the clues fit together and whether two different answers can still fit them. The other questions help diagnose a proposed solver.

1. **Uniqueness (T1).** Do the observations fix the requested answer?
2. **Consistency (T2).** Do the routes and constraints agree?
3. **Saturation (T3).** How many alternatives can still survive?
4. **Fidelity (T4).** What happens to error along a chain?
5. **Sufficiency (T5).** Does the summary keep the needed distinctions?
6. **Optimality (E0).** What does the objective select among the candidates?

One valid answer and a proof that every alternative fails settle the uniqueness question. An optimization criterion can prefer one among several admissible answers, which is a different claim.

§ 7.3

Following the rangefinders to an answer

Put a cart on the flat rail from Chapter 6. Put one rangefinder at the mark 0 and another at 3 m. Each reports distance along the rail. Both can see the cart on either side of their mark. The constructed readings are 1.05 m and 1.80 m. Each error is allowed to vary anywhere between -0.20 and $+0.20$ metres.

One rangefinder comes back with 1.05. If the error can be as large as 0.20, the true distance lies between 0.85 and 1.25. On the rail, that means the cart lies in one of two intervals:

$$F_0 = [-1.25, -0.85] \cup [0.85, 1.25].$$

I can tell how far to look. I have to look in two directions. The rangefinder at 3 comes back with 1.80. It narrows the cart to distances between 1.60 and 2.00 from its own mark:

$$F_3 = [1.00, 1.40] \cup [4.60, 5.00].$$

Now I can lay one set of marks over the other. For two candidate intervals $[l_1, u_1]$ and $[l_2, u_2]$, the common lower endpoint is $\max(l_1, l_2)$ and the common upper endpoint is $\min(u_1, u_2)$. A point belongs to both exactly when it is at or above that lower endpoint and at or below that upper endpoint. This also tells me when the result is empty. If the endpoints are equal, one position survives.

Only the right branch from mark 0 and the left branch from mark 3 meet. They give

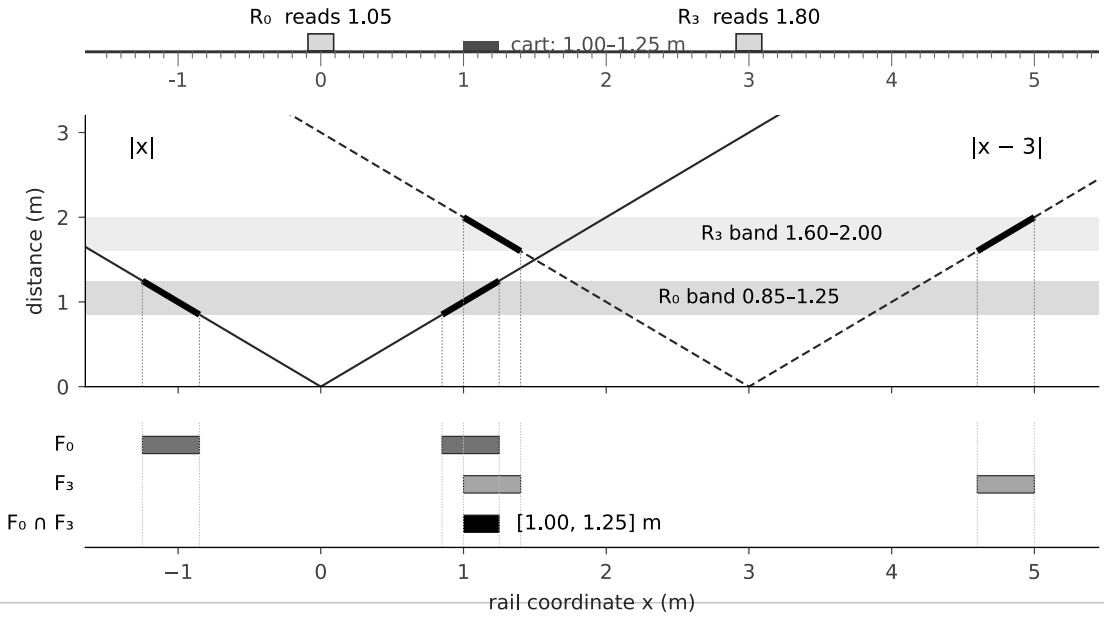
$$F_0 \cap F_3 = [\max(0.85, 1.00), \min(1.25, 1.40)] = [1.00, 1.25].$$

Invert two unsigned distances

Two rangefinders / errors within ±0.20 m / constructed readings 1.05 and 1.80

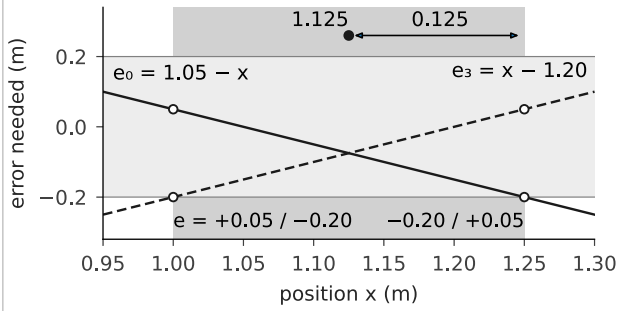
01 RAIL, DISTANCE GRAPHS, FEASIBLE STRIPS

Each reading band cuts its graph twice; dropping the cuts gives two strips.



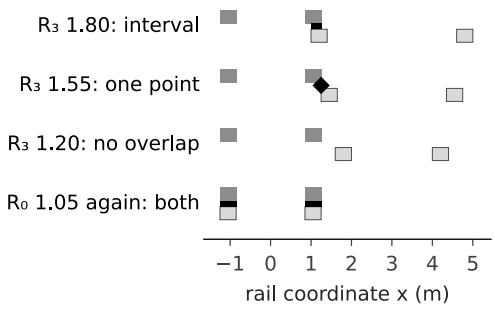
02 THE OVERLAP, ENLARGED

Errors each reading would need



03 CHANGE THE SECOND READING

Dark: F₀ / light: second strips




```
F₀ = [−1.25, −0.85] ∪ [0.85, 1.25]
F₀ ∩ F₃ = [1.00, 1.25] m
F₃ = [1.00, 1.40] ∪ [4.60, 5.00]
midpoint 1.125 m / worst error 0.125 m
```

One unsigned reading leaves two strips. A second viewpoint removes a branch. Repeating the first reading changes nothing. Instruments are schematic; only rail coordinates enter.

Figure 24. Each reading band cuts its distance graph twice, and the cuts drop onto the rail as the strips F_0 and F_3. They overlap only on $[1.00, 1.25]$ m. Panel 02 enlarges the overlap with the error each reading would need; panel 03 changes the second reading. The instruments are schematic.

The readings settle which side of zero the cart is on. They locate the cart to a quarter of a metre. At $x = 1.00$ the errors needed to produce the readings are $+0.05$ and -0.20 metres. At $x = 1.25$ they are -0.20 and $+0.05$. Between those two points the errors vary linearly, so every point in the interval is feasible.

The two readings have left an interval of positions. I can move anywhere in that interval without contradicting either tolerance. Suppose I only need to know whether the cart is to the right of the first mark. Every location in the intersection gives the same answer. Suppose I am positioning a clamp and need the exact coordinate. Now the answer varies across the interval. In the field equations, q names the property I need to recover. The observations stay the same; the question changes.

The middle of the certified interval is 1.125. The worst error of that estimate is 0.125, attained at either end. I cannot reduce that worst error by choosing a different estimate. For any point z, the triangle inequality gives

$$0.25 = |1.25 - 1.00| \le |1.25 - z| + |z - 1.00|.$$

At least one of those two distances is at least 0.125. That is the number I would give the person setting the clamp. If the clamp needs tighter positioning than that, I have to learn something new.

Before buying another instrument, I need to know whether it distinguishes the two positions the first one confused. There is a tempting shortcut. If each reading is useful, why not keep asking the same rangefinder? Suppose every repeat reports 1.05, with the same error bound. Every one of those readings gives the same pair of possible strips F_0. Fifty repetitions leave the same set as the first. A probabilistic noise model might let me combine repeated measurements for greater precision. Here I am taking the intersection of sets, and intersecting a set with itself leaves it as it was.

This is where the elimination rate in the field equations earns its place. If I keep only the two point candidates $\{-1, 1\}$, the first rangefinder admits both. Repeating that constraint retains both. A second viewpoint gives me a different geometric constraint. The coordinate 1 survives it and the coordinate -1 fails.

The arithmetic also handles two controls at the second mark. A reading of 1.55 gives a left branch starting at 1.25, so the overlap is one point. A reading of 1.20 moves that branch to $[1.60, 2.00]$ and the other branch to $[4.00, 4.40]$. Neither meets F_0. In that case I need to check the instruments or the error budget before I give anyone a position.

Once I know which answer the readings determine, I can ask what a stored summary must retain. Once I know $0 \le x \le 3$, I can write the two readings as

$$m_0 = x + b + \eta_0, \qquad m_3 = 3 - x + b + \eta_3.$$

Here b is a common additive offset. The two η terms are whatever error remains in each instrument. When that residual vanishes,

$$x = \frac{3 + m_0 - m_3}{2}, \qquad b = \frac{m_0 + m_3 - 3}{2}.$$

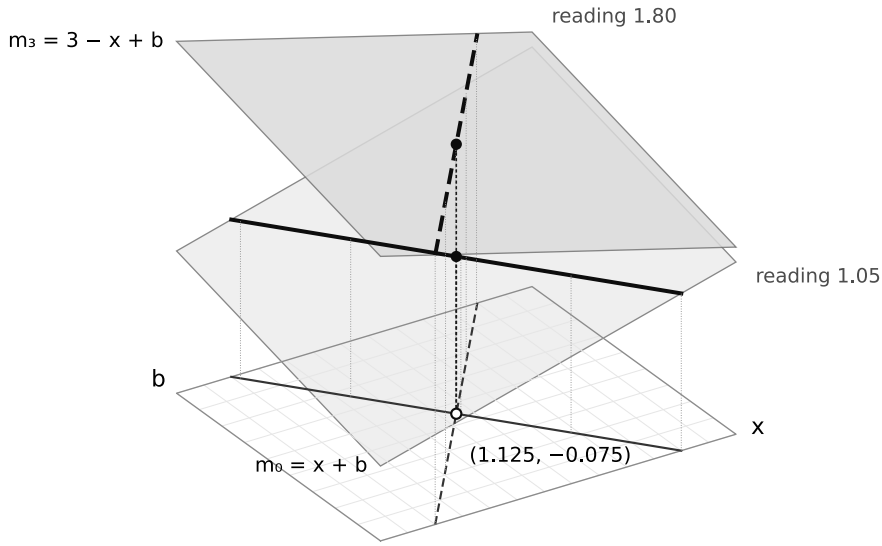
THE GEOMETRY OF FUEL RNG / 03

Two error models, two answers

Shared offset b / independent errors / what the difference keeps

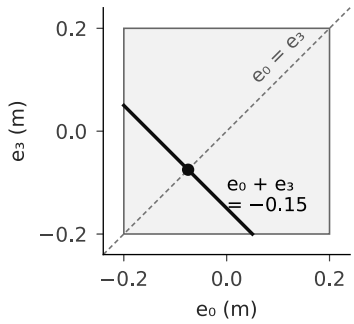
01 BOTH READINGS OVER (X, B)

Each plane is a predicted reading; its dark line holds the observed value.



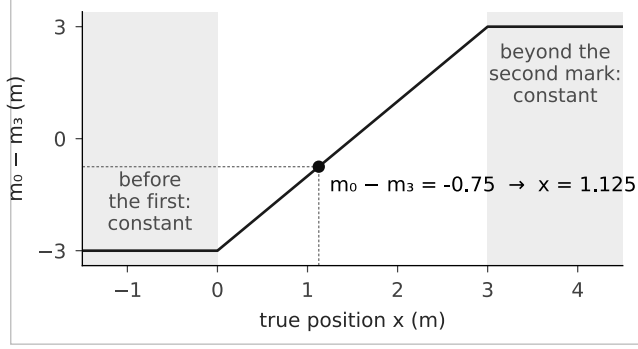
02 ERROR PLANE

Solid: independent / dashed: shared



03 WHAT THE DIFFERENCE KEEPS

Noiseless m₀ − m₃; b cancels




```
x = (3 + m₀ − m₃) / 2 = 1.125 m
|x̂ − x| = |η₀ − η₃| / 2 ≤ ε
```

```
b = (m₀ + m₃ − 3) / 2 = −0.075 m
needs 0 ≤ x ≤ 3 confirmed
```

The interval comes from independent bounds; the single point comes from a shared offset and zero residual error. The choice of error model decides which answer the readings give.

Figure 25. Each plane is the reading predicted from position x and shared offset b; its dark line holds the observed value, projected below. The projected constraints meet at $(1.125, -0.075)$; the reading lines themselves remain at heights 1.05 and 1.80. In the error plane the solid segment allows independent errors and the dashed diagonal a shared one. The noiseless difference $m_0 - m_3$ fixes x only between the marks.

The difference gives the coordinate. The sum gives the common offset. With the same two displayed values I obtain $x = 1.125$ and $b = -0.075$ in this second model. Both readings are consistent with that location and offset. The interval calculation allowed the errors to vary independently. The exact calculation says what follows if a calibration check establishes a shared offset and negligible residual error.

If I can bound the residuals by $|\eta_i| \le \epsilon$, the estimate from the difference obeys

$$|\widehat{x} - x| = \frac{|\eta_0 - \eta_3|}{2} \le \epsilon.$$

That calculation tells me what a compact record can keep. For a confirmed interior cart, the difference alone preserves the position estimate. If I also need to check whether the instruments have drifted together, I need the sum too. And the calculation needs a reason to place the cart between the marks. Beyond the second mark, the two noiseless ranges differ by 3 for every position. I could slide the cart a metre farther down the rail without changing that difference at all.

These are the same questions I had to ask about the AMC records. The first calculation permits both instrument errors to vary within their bounds. The second says what follows from a common offset. That choice of error model determines whether the position is an interval or a single number. The cart lets us see the ambiguity directly. A measurement map gives the same question a form we can use with more instruments. Between the two rail marks, the common-offset model is linear after subtracting the known distance between the marks from the second reading. The unknowns are the cart's coordinate x and the offset b. The first reading depends on their sum and the second depends on their difference. If I keep only the first reading, I can move the cart and change the offset by the opposite amount. The instrument reads the same thing. The second reading gives me a way to catch that move.

Write a general linear measurement as $Az = m$, where z is the vector of unknowns and m is a feasible vector of readings. I may care about just one number $q^{\mathsf{T}}z$, where q specifies the combination I want. The kernel of A consists of the changes v that the instruments cannot see, meaning $Av = 0$.

Proposition 7.1 (A linear task is determined by the readings)**.** *For unrestricted states $z \in \mathbb{R}^d$, the number $q^{\mathsf{T}}z$ is determined by $Az = m$ exactly when $q^{\mathsf{T}}v = 0$ for every $v \in \ker A$. Equivalently, there is a vector w with $A^{\mathsf{T}}w = q$.*

Proof. If $Az = Az'$, then $A(z - z') = 0$. The kernel condition makes $q^{\mathsf{T}}(z - z') = 0$, so the two states give the same answer. Conversely, start at a feasible state z. If $Av = 0$, the state $z + v$ has the same readings. Whenever $q^{\mathsf{T}}v \neq 0$, those states give different answers. Finally, the rows of A are perpendicular to $\ker A$. Their span has dimension $\operatorname{rank} A$, which by rank and nullity is the dimension of $(\ker A)^{\perp}$. Thus a vector perpendicular to the kernel is a linear combination of those rows, as written by $A^{\mathsf{T}}w = q$. □

For the rail with a common offset, the two rows are $(1, 1)$ and $(-1, 1)$. Half of the first row minus half of the second is $(1, 0)$, the row that picks out the cart's coordinate. With both readings I can take that difference. With the first reading alone, the invisible change $(1, -1)$ changes the cart's coordinate by one. That is the same ambiguity I could see by thinking about the instrument.

Uniqueness under exact measurements is only the beginning. Real readings move, so the reconstructed answer needs an error calculation too. I also want to know what happens when the readout is a little wrong. Suppose the error budget is δ in Euclidean norm. Let $m_{\parallel}$ be the orthogonal projection of the measured vector m onto the range of A, the space of possible exact readings. The perpendicular part $m - m_{\parallel}$ spends some of the budget before I move any unknown. Pythagoras gives

$$\|Az - m\|_2^2 = \|Az - m_{\parallel}\|_2^2 + \|m - m_{\parallel}\|_2^2.$$

When the budget admits a state, the remaining radius in measurement space is

$$\delta_{\parallel} = \sqrt{\delta^2 - \|m - m_{\parallel}\|_2^2}.$$

If the quantity is determined in the exact case, choose w in the range of A with $A^{\mathsf{T}}w = q$. Starting with any coefficient vector that works, projecting it onto that range preserves the equation. Any part perpendicular to the range was orthogonal to every possible reading. Now

$$|q^{\mathsf{T}}z - w^{\mathsf{T}}m_{\parallel}| = |w^{\mathsf{T}}(Az - m_{\parallel})| \le \delta_{\parallel}\|w\|_2.$$

For $w \neq 0$, the two readings $m_{\parallel} \pm \delta_{\parallel} w/\|w\|_2$ lie in the range of A and spend exactly the available budget. Each is attained by some state, and they give the two ends of the interval. So the bound is sharp. When $q = 0$, the requested number is always zero.

This Euclidean budget and the separate 0.20 metre bounds of the first rail calculation describe different sets of allowed errors. The interval proof for the rail used those separate bounds. The linear calculation here tells me what happens when the errors share a single squared budget.

A flat surface does not remove that difficulty. Two sensors can be almost redundant even on an ordinary workbench. Imagine replacing the rail with a flat bench. The cart can move along the bench by s or sideways by ℓ. One calibrated scale reports s. The second reports $s + \rho\ell$. The coefficient ρ says how much of the sideways motion the second scale can see. I set the nominal readings to $(1, 1)$ metres and give the pair a combined error budget of $0.020\,\mathrm{m}$.

For $\rho > 0$, subtract the first reading from the second and divide by ρ to recover ℓ. The coefficient vector is $w = (-1/\rho, 1/\rho)$. Its norm is $\sqrt{2}/\rho$, so the sharp error radius for the sideways coordinate is $0.020\sqrt{2}/\rho$ metres. At $\rho = 0$, the two scales report the same direction. The change $(0, 1)$ lies in the kernel, and the readings leave the sideways coordinate free.

I can make ρ ten times smaller by giving the second scale less sensitivity to sideways motion. The same reading budget then admits ten times as much sideways motion. Everything in the example is flat. The geometry I need to inspect here is the angle between the information supplied by the two scales.

§ 7.4

Fitting three routes that do not close

The rail tells me which positions can satisfy two instrument readings at once. Now imagine calibrating a reading between two marked positions on a test bench. I have three conversion routes,

each a chain of instruments that translates the reading from the reference at S into the reference at T. In this simple model, each route adds a signed offset. A positive offset raises the displayed value; a negative one lowers it. Walking back along a route subtracts that offset. Going out on one route and back on another brings me to S again. If the offsets disagree, the reading I bring back has changed. That signed return is the holonomy in this additive model.

Write the three route offsets as P_A, P_B, P_C. The loop AB goes out on A and back on B, giving $P_A - P_B$. The loop BC gives $P_B - P_C$, and AC gives $P_A - P_C$. The middle route cancels when I add the first two returns:

$$(P_A - P_B) + (P_B - P_C) = P_A - P_C.$$

This is a condition every assignment of the route offsets must satisfy. If I add c_S to the coordinates at S and c_T to those at T, each route offset changes by $c_T - c_S$. Their differences stay fixed. I choose $P_B = 0$ to set the reference, and write $P_A = x$ and $P_C = -y$. The three predicted loop returns are then x, y, and $x + y$.

The matrix objective in E0 has its own reference-change rule. If L is a loop matrix and U is a unitary change of frame at its starting point, the transformed matrix is ULU^*. The star denotes conjugate transpose. Since $UIU^* = I$, the defect transforms as $L - I \mapsto U(L - I)U^*$. For any matrix X the squared Frobenius norm is $\mathrm{tr}(X^*X)$. Unitarity and cyclicity of the trace give

$$\|UXU^*\|_{\mathrm{F}}^2 = \mathrm{tr}(UX^*XU^*) = \mathrm{tr}(X^*X).$$

The cost of each loop therefore survives its own change of frame. The test-bench model uses real translations, whose composition is addition and whose identity is zero. Its three observed return values are scalar calibration targets. I use the difference from each target as the return to be fitted.

For a worked calibration example, take recorded returns of 2, -1, and 0 in the same offset units. The first two add to 1. The third says 0. All three readings therefore cannot be exact. I want a set of route offsets that shares the error across the measurements, and I have to say how I will charge for that error. With equal weight on each squared discrepancy, the objective is

$$\Phi(x, y) = \tfrac{1}{2}\left[(x - 2)^2 + (y + 1)^2 + (x + y)^2\right]. \tag{7.4}$$

Each residual measures the return predicted by the model minus the recorded target for that loop. Squaring gives the same penalty to errors in either direction. The cost Φ has squared offset units.

The three reported returns cannot all close exactly. The fit must put their disagreement somewhere, and we can calculate where. Call the residuals $r_1 = x - 2$, $r_2 = y + 1$, and $r_3 = x + y$. No matter which offsets I choose, $r_1 + r_2 - r_3 = -1$. The reading that fails to close gives me a lower bound on the total squared error. Cauchy–Schwarz gives

$$1 = (r_1 + r_2 - r_3)^2 \le 3(r_1^2 + r_2^2 + r_3^2).$$

Thus $\Phi \ge 1/6$. Equality requires the residual vector to be proportional to $(1, 1, -1)$. Its signed sum fixes the multiplier at $-1/3$, so equality requires

$$(r_1, r_2, r_3) = (-1/3, -1/3, 1/3).$$

The first two residuals then give $x_* = 5/3$ and $y_* = -4/3$. Their sum is $1/3$, as the third residual requires. This is a feasible assignment that attains the bound. It is the only one because the first two residuals fix both offsets. The three fitted returns are $5/3$, $-4/3$, and $1/3$. They now close. The error remains visible as three signed thirds, with a total cost of $1/6$. Figure 26 shows the route directions and the fitted returns.

I can get the same answer by solving for where the cost stops decreasing in either coordinate. Write $z = (x, y)^\mathsf{T}$. The three rows of the residual matrix B are $(1, 0)$, $(0, 1)$, and $(1, 1)$, so the residual vector is $Bz - (2, -1, 0)^\mathsf{T}$. Differentiating the squares gives the gradient and its matrix of derivatives, the Hessian.

$$\nabla\Phi(z) = \begin{pmatrix} 2x + y - 2 \\ x + 2y + 1 \end{pmatrix}, \qquad H = B^\mathsf{T}B = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

The gradient is zero at $z_* = (5/3, -4/3)^\mathsf{T}$. For any displacement $e = z - z_*$, expanding the three squares gives

$$\Phi(z) - \Phi(z_*) = \tfrac{1}{2}e^\mathsf{T}He = \tfrac{1}{2}\left[e_1^2 + e_2^2 + (e_1 + e_2)^2\right]. \tag{7.5}$$

The linear term vanishes at z_*. The right-hand side is positive for every nonzero displacement, which checks that the stationary point is the unique minimum.

Finding a stationary point is not enough to justify the answer. Here the quadratic lets us check uniqueness directly. The route example leaves me with a more general question. What if I use a bigger graph or I know some of the offsets exactly? Write $Bz - b$ for the vector of residuals after weighting each row by the square root of its weight. Write $Cz = d$ for the exact constraints, including any reference choices. The matrices B and C have different jobs. One says what the cost can see. The other says where I am allowed to move.

Proposition 7.2 (Uniqueness of a finite quadratic fit)**.** *Suppose $Cz = d$ is feasible in a finite-dimensional real vector space. The cost $\frac{1}{2}\|Bz - b\|_2^2$ attains a minimum on that affine set. Its minimizer is unique exactly when*

$$\ker C \cap \ker B = \{0\}.$$

Proof. Choose any feasible z_0. Every feasible point is $z_0 + u$ for some $u \in \ker C$. The possible values of Bu form the linear subspace $B(\ker C)$. It is closed in finite dimension, so $b - Bz_0$ has a unique orthogonal projection onto it. Choose any u_* that gives that projected value. Then $z_0 + u_*$ minimizes the squared distance, and all minimizers must give that same value of Bu. Two minimizing displacements therefore differ by an element of $\ker C \cap \ker B$. Conversely, adding any element of this intersection to a minimizer keeps the point feasible and preserves every residual. This proves existence and the stated uniqueness criterion. □

Before I chose a reference in the three-route example, I could slide all three offsets together by the same amount. All three differences would stay the same. Fixing P_B removes that freedom. The differences AB and BC then see both remaining coordinates. If I kept only the AC row, even with $P_B = 0$ I could increase x and decrease y by the same amount. The cost would stay fixed. The criterion detects that freedom through the displacement $(1, -1)$.

The minimum is fixed. The route the numerical solver takes toward it still depends on the step size. The cost changes at different rates along different directions. The Hessian sends $(1, -1)$ to

itself and $(1, 1)$ to three times itself. These are eigenvectors with eigenvalues 1 and 3. In the first direction the sum $x + y$ stays fixed. In the second direction both offsets move together, so the third residual changes as well. Equation (7.5) shows the extra cost directly.

Take a gradient step of size η in these coordinates. Since $\nabla\Phi(z) = H(z - z_*)$, the error after one step is

$$e_{k+1} = (I - \eta H)e_k.$$

A component in an eigenvector direction with eigenvalue λ is multiplied by $1 - \eta\lambda$ each time. Convergence from every starting error requires $|1 - \eta| < 1$ and $|1 - 3\eta| < 1$, and these conditions suffice. The intersection is $0 < \eta < 2/3$. At $\eta = 2/3$ the component along $(1, 1)$ changes sign without shrinking. Above that step size its magnitude grows.

I start one unit away from the minimizer in each coordinate, at $z_0 = (8/3, -1/3)^{\mathsf{T}}$. The initial error is $(1, 1)$, so this test isolates the faster-changing direction. At step k its coefficient is $(1-3\eta)^k$ and the excess cost is $3(1 - 3\eta)^{2k}$. With $\eta = 1/3$ the first step lands at the minimum. With $\eta = 2/3$ the iterates alternate across it, with cost 19/6 at every step. With $\eta = 1$ the first step lands at $(-1/3, -10/3)$, raising the cost to 73/6. The offset doubles in magnitude and the excess cost quadruples at each step. Table 7.4 gives the replay of these three runs. I can see exactly where the proposed update goes wrong, because this quadratic gives me the entire trajectory in two numbers. Now I had an answer on paper and a way to see where a numerical step could go wrong. I wanted to put the same examples through GIGI and follow the numbers back to the instruments. For the rangefinders, I entered the endpoints in integer centimetres. For each pair of branches, GIGI took the larger lower endpoint and the smaller upper endpoint. The pair survives when the first number is at most the second. I also checked the original unsigned distance inequalities on a grid of centimetre marks. The interval calculation covers the whole rail, including the spaces between those marks.

The common-offset model asks for signed sums of the readings, so GIGI adds those next. On the flat bench, it adds the two squared coefficients of the task witness w. Taking the square root of that sum and multiplying by 0.020 metres gives the error radius. Each operation now has a place in the picture: it combines readings, cancels an offset or tells me how far the cart could move without exceeding the error budget.

For the route fit, the three residual rows have coefficients $(1, 0)$, $(0, 1)$, and $(1, 1)$. GIGI sums products of those coefficients to build the Hessian and returns its eigenvalues 1 and 3. It also sums the coefficient times the recorded return in each row, giving the right-hand side $(2, -1)$. I solve the two normal equations in exact fractions. For each step the companion calculation supplies signed residual numerators and their squares. GIGI returns the sums for the gradient and cost. I restore the denominators and use the gradient to advance the iterate.

Take the common starting point of the three plotted traces, $(8/3, -1/3)$. The residual numerators are 2, 2, and 7, with denominator 3. The first gradient component uses loops AB and AC. The second uses BC and AC. The replay returns 9 for each numerator sum and 57 for the sum of squares. Dividing by 3 gives the gradient $(3, 3)$. Dividing the sum of squares by $2 \cdot 3^2$ gives the cost 19/6. That is the first row of the plotted calculation, before any step has been taken.

What the rangefinder records permit

The first reading in Table 7.1 is 1.05 at mark 0. The second reading is 1.80, 1.55, or 1.20 at mark 3, or 1.05 repeated at mark 0. Every reading has the same 0.20 metre bound.

Case	LL	LR	RL	RR	Grid count
Different marks	∅	∅	$[1.00, 1.25]$	∅	26
Repeat at zero	$[-1.25, -0.85]$	∅	∅	$[0.85, 1.25]$	82
Touching	∅	∅	$[1.25, 1.25]$	∅	1
Conflict	∅	∅	∅	∅	0

L and R denote the left and right branches of each instrument in order. The grid count is the number of feasible centimetre marks between -1.50 and 5.50 metres, inclusive. In the touching case, counting the shared endpoint gives one. In the conflict case, even the continuous intersection is empty.

Under the common-offset model, GIGI returns the sum 2.85 and the difference -0.75 metres. Adding 3 to the difference and dividing by two gives $x = 1.125$. Subtracting 3 from the sum and dividing by two gives $b = -0.075$.

Table 7.2 gives the bench calculation. The coefficient ρ is the sideways sensitivity of the second reading. GIGI returns $\|w\|_2^2$ where the sideways coordinate is identifiable.

Sensitivity ρ	Error budget (m)	$\|w\|_2^2$	Sideways radius (m)
1	0.020	2	0.028284
$\frac{1}{10}$	0.020	200	0.282843
$\frac{1}{100}$	0.020	20000	2.828427
0	0.020	—	Unbounded

The route fit and its iteration record

Table 7.3 puts the three target returns beside the returns of the fitted routes. The fit closes because $5/3 - 4/3 = 1/3$. Each residual contributes $1/18$ to the cost, for a total of $1/6$.

Loop	Recorded return	Fitted return	Residual
AB	2	$\frac{5}{3}$	$-\frac{1}{3}$
BC	-1	$-\frac{4}{3}$	$-\frac{1}{3}$
AC	0	$\frac{1}{3}$	$\frac{1}{3}$

Every run in Table 7.4 starts at $(8/3, -1/3)$. The error is $a_k(1,1)$ at step k. The second column in each pair gives $\Phi_k - \Phi_*$ in squared offset units.

	$\eta = 1/3$		$\eta = 2/3$		$\eta = 1$	
Step	a_k	Excess	a_k	Excess	a_k	Excess
0	1	3	1	3	1	3
1	0	0	-1	3	-2	12
2	0	0	1	3	4	48
3	0	0	-1	3	-8	192
4	0	0	1	3	16	768
5	0	0	-1	3	-32	3072

At step 1 with $\eta = 1$, the two coordinates are $(-1/3, -10/3)$. Both have moved two units past the minimizer. The cost is $12 + 1/6 = 73/6$. When I kept only the AC measurement, the replay returned Hessian eigenvalues 0 and 2. The zero belongs to the direction $(1, -1)$, along which the sole recorded return stays fixed.

The plotted traces isolate one error mode. Starting with a mixture shows why an apparently good first step can be misleading. The isolated mode made the failure easy to see. Now start all three runs at $(0, 0)$. At $\eta = 1$ the cost drops from $5/2$ to $1/2$ in one step. I could stop there and think the step size was safe. The initial error is $(-5/3, 4/3)$, which has a coefficient of $-3/2$ along $(1, -1)$ and just $-1/6$ along $(1, 1)$. The first component vanishes at $\eta = 1$. The smaller one flips sign and doubles at every step. By step three the cost is $11/2$, above where it started.

Table 7.5 gives the replayed excess cost $\Phi_k - \Phi_*$ from the origin. Compare its first two rows with the isolated-mode table. The same update rule can look quite different depending on which errors were present at the start.

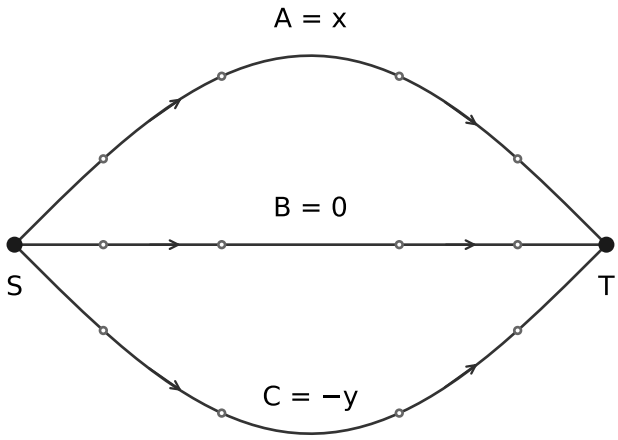
Step	$\eta = 1/3$	$\eta = 2/3$	$\eta = 1$
0	$\frac{7}{3}$	$\frac{7}{3}$	$\frac{7}{3}$
1	1	$\frac{1}{3}$	$\frac{1}{3}$
2	$\frac{4}{9}$	$\frac{1}{9}$	$\frac{4}{3}$
3	$\frac{16}{81}$	$\frac{7}{81}$	$\frac{16}{3}$
4	$\frac{64}{729}$	$\frac{61}{729}$	$\frac{64}{3}$
5	$\frac{256}{6561}$	$\frac{547}{6561}$	$\frac{256}{3}$

On the rail the two readings defined a set of allowed positions. The second mark narrowed the set. In the route problem the three recorded returns could not all hold, so I had to choose a cost for changing them. Once that choice was explicit I could prove what the fit preserves, show how much error it leaves, and test whether a proposed step reaches it.

Make the three loop readings close

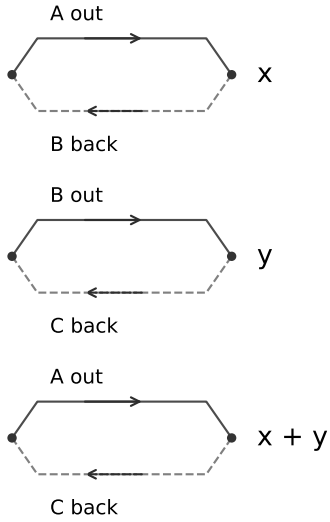
ADDITIVE ROUTES / ONE REFERENCE / EQUAL-WEIGHT FIT

01 Three conversion routes



All three routes carry the reading from S to T.

02 Build each loop



03 Each row keeps its own discrepancy

Loop	Model	Target	Fit	Residual
AB	x	2	5/3	−1/3
BC	y	−1	−4/3	−1/3
AC	x+y	0	1/3	1/3

04 Fit the disagreement, then check the lower bound

$$\Phi(x,y) = \tfrac{1}{2}[(x-2)^2 + (y+1)^2 + (x+y)^2]$$

$$r_1 + r_2 - r_3 = -1 \quad \Rightarrow \quad \Phi \geq 1/6$$

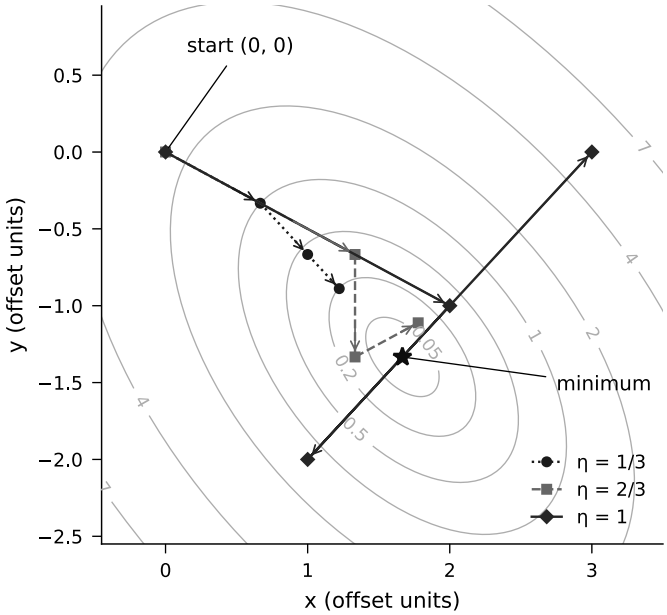
The recorded returns 2, −1 and 0 do not close. The fit leaves three signed thirds of error.

Figure 26. Three additive conversion routes share endpoints S, T and reference route $B = 0$. Each loop goes out along one route and back along another. The fitted loop returns close; their signed differences from the recorded returns give the residuals. The equal-weight quadratic cost reaches its proved minimum 1/6.

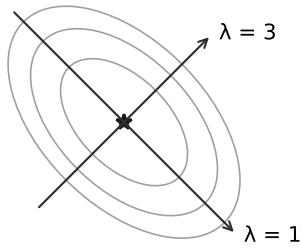
The bowl stays; the step changes

CONTOUR VIEW / START (0, 0) / FIRST THREE UPDATES

01 Follow the saved iterates



02 Two curvature directions



03 An exact quadratic

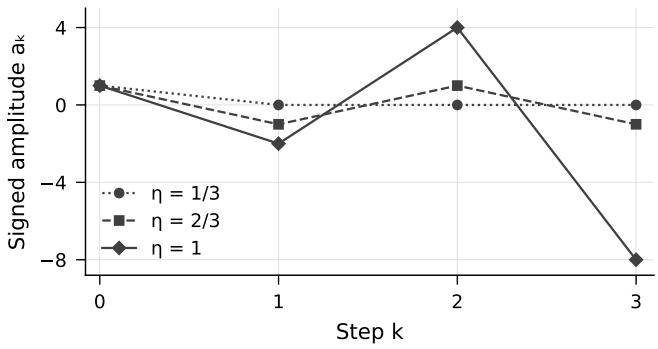
$$H = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$z^* = (5/3, -4/3)$

$\Phi^* = 1/6$

Contours show $\Phi - \Phi^*$.

04 Isolate the (1,1) error instead



05 A different starting point

$z_0 = (8/3, -1/3)$

$a_k = (1 - 3\eta)^k$

$\Phi_k - \Phi^* = 3a_k^2$

All initial error lies along (1,1). A safe step must control both Hessian modes.

For both error directions to decay, the step must satisfy $0 < \eta < 2/3$.

Figure 26. *Continued.* The upper contour plot follows the first three saved updates from $(0, 0)$. The lower traces use the different start $(8/3, -1/3)$, whose error lies entirely along $(1, 1)$. The same three step sizes act on both starts. Contours show excess cost; the lower plot shows signed error amplitude.

Carrying an answer through several steps

The rangefinder and route calculations have made the requirements visible. A reading must distinguish the alternatives relevant to the task; a fit needs a criterion; an update needs a stability condition. I can now follow an answer through a longer chain of checks. I keep the catalogue identifiers beside the statements so that each can be traced back to the research notes.

The elimination calculation gives a concrete stopping count (F1).

If the initial $N > 1$ candidates include a viable one and each check removes at least a fraction $1 - e^{-\lambda}$ of the remaining alternatives, one sufficient number of checks is

$$m = \left\lfloor \frac{\log(N-1)}{\lambda} \right\rfloor + 1.$$

That bound can be useful even when the best order of checks is unknown. An individual clue can distinguish all the remaining alternatives at once.

Checks along routes require a second kind of bookkeeping: the order in which transports act (F2).

If two loops based at the same point are traversed in order γ_1 then γ_2, the transport matrices act in that order:

$$\mathrm{Hol}(\gamma_1 \text{ then } \gamma_2) = \mathrm{Hol}(\gamma_2)\,\mathrm{Hol}(\gamma_1)$$

on column vectors. In a local matrix logarithm, small exponents add to first order, with commutators supplying higher-order terms. This is how I can build a check of a large loop from smaller ones. The accumulated defect still needs a bound.

Even correctly composed checks must be mutually compatible. Consistency (F3) asks whether one state can satisfy them all.

Checking consistency means finding a single admissible state satisfying every constraint. Pairwise agreement can miss a conflict. On a flat line, the three feasible sets $\{0,1\}$, $\{1,2\}$ and $\{0,2\}$ have nonempty pairwise intersections and an empty common intersection. A finite family of convex feasible sets in $\mathbb{R}^d$ admits a stronger check by Helly's theorem. If every subfamily of at most $d+1$ sets intersects, the entire family does. That is Eduard Helly's classical theorem; I borrow it as a check, not as a result of my own.

The readings and the budget will never be perfectly exact. The next bound asks how a small change in the budget inputs changes the computed ratio (F4).

For the scalar budget, if $K, K' \geq K_0 > 0$ and $|\tau'| \leq T_0$, then

$$|C - C'| \leq \frac{|\tau - \tau'|}{K_0} + \frac{T_0 |K - K'|}{K_0^2}.$$

This follows by adding and subtracting τ'/K. It bounds a change in the computed budget. Stability of the actual completion, such as the response of a reconstructed spectrum to perturbed observations, needs a bound for its own reconstruction map.

That still leaves an engineering choice: which check should I pay for first (F5)?

To remove the most candidates in one step, I can choose the test that eliminates the largest number of currently viable alternatives. If I rank tests by a curvature proxy instead, I need to show how that proxy predicts the elimination count. The best first test also need not give the cheapest whole sequence, particularly when tests have different costs or eliminate the same alternatives.

After choosing the checks, I have to store what they establish. Compression (F6) needs a precision as well as a coordinate count.

Suppose the continuous cache uses d coordinates quantized to at most B levels each, and the discrete residue has at most R_0 permitted labels. The number of possible cache words is at most $R_0 B^d$. Its Shannon entropy in bits is at most

$$\log_2 R_0 + d \log_2 B.$$

This is a bound at declared precision and range. It gives a design question for a model that compresses a context into a summary. Which precision and which labels keep the answers accurate?

If a step can introduce error, the permitted length of a chain must be tied to the error budget (F7).

A usable gap bound needs a model that connects gap size to an error. If the estimate from T4 applies with a uniform per-step error $\varepsilon > 0$ and initial error zero, $k\varepsilon \le \tau$ is sufficient for an error tolerance τ in the same units. A geometric estimate of ε can sharpen the result. The largest gap that happens to work may be larger than this certified gap.

Several instruments can share the work, provided their combined readings resolve the answer they are meant to supply (F8).

Different instruments can constrain different parts of a state. If their observations are mutually consistent and jointly determine an answer, then that answer is shared by reconstructions that satisfy them all. If the instruments leave the same ambiguity, sharing the records preserves it. A distributed algorithm must also specify how the agents exchange records and establish that they converge.

§ 7.6

Testing the limits of the model

Once several checks share a model, I want to know where their guarantees stop. A boundary in a certificate, a direction the instruments cannot see, and a change in a solver's cost are different things to inspect. The following questions connect those limits back to the measurements.

Start with the boundary of the candidate-elimination certificate (S1).

In the finite survival model, $m\lambda > \log(N-1)$ marks the region where the certificate can rule out every alternative. A plot of this inequality has a sharp boundary because it is a certificate. An observed phase transition needs a separate limiting model and a quantity whose behavior changes across the boundary. One way to test the idea is to measure survival curves as problem size increases and see whether they concentrate around a threshold.

A second boundary is local: which directions can the measurement map see (S4)? Put a marker on a flat bench and measure only its horizontal coordinate. Sliding it vertically changes the state

but leaves the reading fixed. Those unseen changes form a *fiber*, the set of states with the same reading. A second instrument measuring height can remove that ambiguity.

Suppose a smooth measurement map on an n-dimensional state space has constant rank r in a neighbourhood. The constant-rank theorem gives a local fiber of dimension $n - r$. The measurements account for r local directions, and the fiber consists of changes the measurements cannot see. A cache intended to recover the whole state must resolve distinctions on those fibers. A cache that answers one question only needs to recover that question's value.

Those unseen directions suggest questions about search cost, but the cost must still be measured for a stated algorithm (S8).

I asked whether curvature could help predict the cost of solving a problem. To test that, I need a declared algorithm and a way to measure both the geometry and the computational cost at fixed size. I can then ask whether the diagnostic predicts cost on new instances or whether it tracks another quantity, such as clause density. An association in a random-instance landscape gives a model for that landscape. Proving a computational lower bound requires an argument about what operations every algorithm in the stated class can perform.

Changing the size of the loops gives another test of the representation (S12).

I can measure transport on loops of several sizes. If a larger loop decomposes into ordered, based loops at smaller scales, their transports compose as in F2. The sizes of the resulting defects need not increase with the scale. Two nontrivial small rotations can cancel in the larger product. That is a reason to record the composition and the resulting defect separately.

If the model itself is being learned, an update also consumes some of the allowed change (S17).

Suppose the holonomy defect is Lipschitz with constant $L_{\mathcal{H}}$ on the part of parameter space an update crosses, and the update length is at most ηG. The change in the defect is then at most $L_{\mathcal{H}}\eta G$. A budget of τ per step can be enforced by choosing $\eta \le \tau/(L_{\mathcal{H}}G)$ when both constants are positive. This controls one change of defect. A convergence result also has to describe the loss and the update rule.

The finite checks also raise questions about what happens as the models grow. My catalogue calls these third-order proposals. They need comparisons across sizes, not just a more elaborate description of one run. These questions look at the behaviour of a whole model, including its large-scale limits.

For example, a sharp-looking change in a plot invites a scaling test. A correlation length ξ measures how far an association extends in the declared model. A susceptibility χ measures response to a specified small change in a control. Near a candidate threshold m^*, one can propose power-law models for such measured quantities:

$$\xi \sim |m - m^*|^{-\nu}, \qquad \chi \sim |m - m^*|^{-\gamma}.$$

The data have to distinguish that behaviour from a smooth crossover or a finite-size effect. A universality claim would require showing that the exponents persist across a specified class of models. I used the idea to organize questions about the random 3-SAT landscape, where an observed transition and the complexity of unrestricted algorithms remain separate questions.

Storage raises a similar distinction between the size of one construction and a lower bound that every construction must obey (T7).

A sufficient cache provides one representation through which a chosen answer can be computed. Its size gives an upper bound for that construction. To show that every method needs at least that much information, I would need a lower bound in a declared storage or computation model. Two histories that share a cache word and give different answers would immediately rule out that word as a sufficient summary.

Some information survives because the allowed changes cannot remove it. Winding gives a concrete example of that protection (T12).

A winding number around a fixed obstacle stays constant under continuous deformations of the loop that avoid the obstacle. Think of a cord on a workbench looped around a peg. You can move one bend a little without changing the winding. The condition matters. If the cord can cross the peg or cease to be a closed loop, the winding need not survive.

The gauge calculations used examples of that same distinction. A recorded invariant can survive an allowed evolution. Convergence and the ability of the invariant to distinguish topologies require additional arguments.

For quantities that do change continuously, a rate bound comes from the trajectory and the gradient together (T19).

For a differentiable scalar holonomy defect h along a trajectory $\gamma(t)$, the chain rule gives

$$\left|\frac{d}{dt}h(\gamma(t))\right| \leq \|\nabla h(\gamma(t))\|\,\|\dot{\gamma}(t)\|.$$

Bounds on the gradient and the rate of the trajectory give a rate bound for the defect, in defect units per unit time. Both come from the chosen model. This is one way to turn a limit on parameter updates into a geometric condition I can check.

The next group in the catalogue brings the question back to representation: which distinctions survive the summary, and which changes would need a genuine conservation argument? Here I compare representation, dynamics and the way the model records a transition.

Begin with the difference between changing a basis and merging states in a cache (FO1).

A basis change within a finite-dimensional vector space is invertible. A cache that merges two states is invertible only for the quantities it keeps. The useful question is whether the answer I need is unchanged when I go from one description to the other.

If the basis change has Fourier structure, a fast Fourier transform may reduce its cost. An arbitrary choice of basis or cache gives no such complexity bound. In the lattice work, I ask this in operational terms. What can I still compute after replacing the full configuration with the loop and charge data?

A proposed conservation relationship asks for more: a quantity that stays fixed under a specified evolution (FO2).

One proposal was to look for a balance between a normalized information measure and a normalized curvature measure,

$$I_{\text{semantic}} + K_{\text{geometric}} = \text{constant}.$$

A conservation claim requires a defined flow, compatible units and a reason that the derivative of the sum vanishes. In a variational setting, Noether's theorem links a conserved quantity to a

continuous symmetry of the action. Stationarity of the action alone does not supply that symmetry or identify the conserved quantity.

The spectral comparisons with zeta zeros gave me a place to ask what an information measure might track. Matching selected spacing statistics does not establish a conservation law.

The candidate-count picture gives a final check on this group of ideas. I can compare the certified boundary with the alternatives that actually survive (FO9).

The trichotomy gives a plot of where the finite candidate-elimination certificate applies. I can plot $\Gamma > 1$, $\Gamma = 1$ and $\Gamma < 1$ for the finite candidate model at the start of this chapter, then overlay empirical counts to see where the bound is tight. If I change the measurement or count repeated tests as new information, the observed survival curve can change even while a curvature score stays the same. This is a diagnostic for the certificate and a way to locate which part of the model needs testing.

§ 7.7

Giving completion an objective

The last layer asks what can be said about minima and dynamics once an objective is defined.

Once an objective is specified, I can compare its values and study its minima. The word energy here has to remain attached to that definition. I begin by giving the objective a precise meaning (E1).

For the least-holonomy example, set $E(A) = \mathcal{H}(A)$. Here energy is the name of the scalar objective. The cost measures the weighted defects of the chosen loops. If those defects capture the task's error, minimizing the objective can improve completion. An objective built from loop matrices has its own units and does not by definition measure work in the underlying physical system.

With that definition, the lowest attainable value has a simple interpretation (E2).

A ground state of this optimization model is a global minimizer of E over its admissible states. If all chosen defects can vanish simultaneously, the least-holonomy objective attains its lower bound of zero. Distinct states can attain the same minimum. The observation model determines which distinctions matter in that case.

The next question is how to measure a departure from that minimum (E3).

On a finite set of states with distinct energy levels, the difference between the lowest value and the next distinct value is a finite-model level spacing. The Hessian of a differentiable objective supplies a different diagnostic, describing the quadratic cost of a small departure from a local minimum. Neither quantity inherits a value from $C = \tau/K$. A Hamiltonian gap needs the spectrum of the specified physical Hamiltonian.

If I enlarge the model, I also have to say how its objective is normalized (E4).

For a sequence of models indexed by vocabulary size n, I can ask whether a consistently normalized objective E_n has a limit per symbol,

$$\lim_{n\to\infty} \frac{E_n}{n} = e_0.$$

An experiment at several sizes can test the trend. Proving convergence requires assumptions on how the models grow and how their terms combine.

A finite ensemble supplies a setting where the energy and entropy calculation can be proved exactly (E5).

For a finite model with energy values E_i and temperature $T > 0$, define $Z = \sum_i e^{-E_i/(k_\mathrm{B}T)}$ and $g_i = e^{-E_i/(k_\mathrm{B}T)}/Z$. A distribution p has entropy $S(p) = -k_\mathrm{B} \sum_i p_i \log p_i$ and free energy

$$F(p) = \sum_i p_i E_i - TS(p).$$

To see the free-energy identity, take the logarithm of the Gibbs weight: $\log g_i = -E_i/(k_\mathrm{B}T) - \log Z$. Substitute it into $D_\mathrm{KL}(p\|g) = \sum_i p_i \log(p_i/g_i)$ and use $\sum_i p_i = 1$. The energy sum and the entropy sum then give

$$F(p) + k_\mathrm{B}T \log Z = k_\mathrm{B}T D_\mathrm{KL}(p\|g) \geq 0.$$

Equality holds at $p = g$. The calculation balances preference for low-energy states against the entropy of the distribution. If I use the same formula for an inference cost, I must justify that model and its scale; naming a tuning parameter "temperature" does not make the computation a heat engine. The inequality itself is the Gibbs variational principle of statistical mechanics; E5 restates it in my catalogue's notation.

§ 7.8

Following the objective through time

One possible evolution follows the downhill direction of the objective (D1).

A differentiable objective on a Riemannian parameter space defines gradient flow,

$$\frac{d\gamma}{dt} = -\nabla_\gamma E.$$

The metric fixes the meaning of the gradient. A discretization or a stochastic update needs its own rule. If I change the objective or the metric, I may change the trajectory and its limit.

Moving downhill does not tell me how quickly I reach a useful answer. A quantitative inequality supplies one convergence rate (D2).

A logarithmic time to tolerance is available under a quantitative condition. Suppose the flow exists, stays in a region where $E - E_* \geq 0$, and satisfies $\|\nabla E\|^2 \geq 2\mu(E - E_*)$ with $\mu > 0$. Then

$$E(\gamma(t)) - E_* \leq e^{-2\mu t}(E(\gamma(0)) - E_*).$$

This follows from the energy derivative in D3. Reaching a prescribed positive objective error takes a time bounded logarithmically in the inverse tolerance. The inequality has to hold along the trajectory, and objective accuracy only implies accuracy of the requested completion with a further bound. The condition is the Polyak–Łojasiewicz inequality, and the rate is Polyak's classical result; D2 restates it.

The chain-rule calculation underlying that rate is short enough to check directly (D3).

Along the differentiable flow in D1 the chain rule gives

$$\frac{d}{dt}E(\gamma(t)) = -\|\nabla E(\gamma(t))\|^2.$$

The energy cannot increase along this flow and is stationary at a critical point. The identity tells me which direction the objective moves. It takes another argument to identify the limit.

Changing the ensemble asks a different question from following a trajectory. The finite Gibbs model above lets me calculate a static response (D4).

There is an exact static response calculation for the finite equilibrium model in E5. Add a field h coupled to an observable Q by replacing E_i with $E_i - hQ_i$. Differentiating its normalized Gibbs weights gives

$$\frac{d}{dh}\langle Q\rangle_h = \frac{\mathrm{Var}_h(Q)}{k_{\mathrm{B}}T}.$$

This connects equilibrium fluctuations to a response to the chosen field. A time-dependent response needs specified dynamics and assumptions about how the system is prepared. The static identity is a useful check of a simulation that samples the biased ensemble. It is the standard static fluctuation–response relation of equilibrium statistical mechanics, not a new result.

Returning to the trajectory, the local modes tell me how long a small displacement takes to fade (D5).

For Euclidean gradient flow near a stationary point, the linearized equation is $\dot{u} = -Hu$, where H is the Hessian at that point. An eigenmode with eigenvalue $\lambda > 0$ decays as $e^{-\lambda t}$ and has time constant $1/\lambda$ in the flow's time units. The slowest decaying positive mode has the smallest positive eigenvalue. Zero modes and nonlinear effects need their own analysis.

A computer takes discrete steps. The same modes tell me how large a step that particular update can tolerate (D6).

For a strictly convex quadratic with symmetric positive definite Hessian H, gradient descent with constant step size η multiplies an error mode of eigenvalue λ by $1 - \eta\lambda$ at every step. The exact condition for convergence from every initial error is

$$0 < \eta < \frac{2}{\lambda_{\max}(H)}.$$

This is a limit on a particular discretization in a particular geometry. For a more general objective, the Hessian can vary along the iterates, and an adaptive method may change the effective update direction. The condition is the textbook stability bound for gradient descent on a quadratic; D6 restates it.

Try the one-variable objective $E(x) = x^2/2$. Its gradient is x, so the update is $x_{k+1} = (1-\eta)x_k$. With $\eta = 1/2$, the error halves. With $\eta = 2$, it changes sign without shrinking. With $\eta = 3$, it changes sign and doubles. The bowl has not changed. The size of our step has changed whether the solver reaches its bottom. The Hessian formula applies that same calculation to each mode of a positive definite quadratic.

§ 7.9

Turning the model into a computation

The relaxation algorithm puts the least-holonomy model into a numerical form. To run it, I must choose a finite discretization, a set of admissible connections, the observed constraints and the loop costs.

The following loop makes those choices explicit.

> **Input.** A finite parameterization of connections A, a feasible set F, loop costs $\mathcal{H}$, and a stopping tolerance.
>
> **Initialize.** Choose a feasible starting connection.
>
> **Iterate.**
>
> 1. Evaluate the loop transports and the scalar cost $\mathcal{H}(A)$.
> 2. Compute its gradient in the chosen parameterization.
> 3. Take a trial step and project it back to F where a suitable projection is available.
> 4. Check feasibility and the stopping criterion. Record the cost and the residual.
>
> **Return.** The final candidate and the recorded residuals.

Projection can enforce feasibility in a model where it is well defined. Convergence depends on the objective, the set and the step rule. The residuals help tell whether the computation reached a declared tolerance.

Having an update rule lets me ask what it costs to run. That cost belongs to the particular implementation. The operation count depends on evaluating the chosen loops, computing derivatives and enforcing the constraints. If an implementation costs $O(n^2)$ per update for n states and its objective errors contract by a fixed factor at each update, the cost to reach error ε from a fixed starting bound is $O(n^2 \log(1/\varepsilon))$. Each of these hypotheses belongs to the implementation. A sparse representation can change the operation count when the chosen loops and updates preserve that structure.

It also lets me return to the failed step in the three-route example. A first decrease in cost did not establish stability; the update had to control every relevant error mode. For a positive definite quadratic, the eigenvalues in D6 specify the step-size limit and the condition number $\kappa = \lambda_{\max}/\lambda_{\min}$. A preconditioner can improve the spectrum of the transformed problem. I would report the actual step rule and measured convergence for a nonquadratic model. The name of an optimizer alone supplies no stability bound.

We can now return to the question that opened the chapter: what does a measurement let me keep, and what must I still determine? The field equations began as a catalogue. A few ideas give me a way through it. A budget limits a specified geometric change. An observation eliminates some states and leaves others possible. A cache can keep the distinctions needed by a particular task. An objective gives me a way to choose or test a candidate.

I can now carry those distinctions back to the lattice. Which loops did I measure? What did the cache merge? When I changed the sampling model, which number changed? The next chapter starts with that last question.

Chapter 8

The Ratchet

"The vacuum is not empty. The vacuum has energy. The vacuum has topology. And the topology can be controlled."
—Personal notes, December 2025

This chapter contains material I debated whether to include at all.

In Chapter 6 I called the idea Topological Vacuum Rectification. The name came from a question: could the geometry of a gauge field make a directional bias measurable? The first place I could ask that question cleanly was in a simulation.

Here the instrument is a computation. It assigns different weights to field configurations and asks whether a loop statistic changes. A separate calculation measures how far apart configurations sit in one chosen description. To turn either one into a claim about energy, I would have to measure something else.

I will keep two experiments separate as I work through that question. One proposes changing the action used to generate configurations. The other changes the weights assigned to configurations already saved. Only after I can follow each operation to its observable can I ask what the response means.

§ 8.1

What a vacuum state permits

The vacuum is not empty. This is one of the strangest discoveries in twentieth-century physics.

The vacuum is the lowest-energy state of a quantum field theory. Even in that state, measurements can have nonzero variance. Think of a guitar string fixed at both ends. Its shape limits how it can move. Quantize those modes and even the lowest-energy state has uncertainty in what a measurement of the string's displacement will show. The corresponding field-theory idea is a ground state with structure. The rest of this chapter asks what topology adds to that picture.

What is topology? It is the mathematics of shapes that cannot be smoothly deformed into each other. A coffee mug and a donut are topologically the same—both have one hole, and you can imagine smoothly morphing one into the other like clay. But a donut and a sphere are topologically different—to turn a donut into a sphere, you would have to fill in the hole, which is not a smooth deformation. Topology is about these fundamental, unfillable differences.

In the classical gauge-field description, suitable boundary conditions permit pure-gauge configurations labeled by winding. Their relation to the quantum vacuum requires quantization and the

treatment of large gauge transformations. I will use a landscape picture to describe the classical starting point before making that distinction explicit.

Picture several equally low valleys separated by ridges in a classical configuration landscape. A ball at the bottom cannot reach the next valley by rolling downhill. The picture describes barriers between classical configurations. It does not yet tell me how many physical quantum ground states exist.

The valley labels keep track of the classical winding structure under the stated boundary and gauge conventions. They give me something to organize in the description. Whether two labeled configurations represent distinct physical quantum states, and what a transition amplitude means, needs the next layer of the theory.

Another way to visualize it: imagine wrapping a rubber band around a cylinder. You can wrap it once, or twice, or three times. Each wrapping is stable—the band will not spontaneously unwrap. And the different wrappings are genuinely different; you cannot continuously deform a band wrapped twice into a band wrapped once without cutting it.

With the boundary conditions used for the winding construction, integer labels distinguish classes under the restricted deformations being considered. Large gauge transformations and the quantum construction require additional care. The labels do not establish an energy source or a collection of physically interchangeable reservoirs.

The Chern-Simons number characterizes this topology:

$$n_{CS} = -\frac{1}{8\pi^2}\int_{M^3} \mathrm{Tr}\left(\mathcal{A}\wedge d\mathcal{A} + \frac{2}{3}\mathcal{A}\wedge\mathcal{A}\wedge\mathcal{A}\right) \tag{8.1}$$

Here $\mathcal{A} = -igA$ is the anti-Hermitian connection associated with the Hermitian field A in the Wilson-loop convention $\mathcal{P}\exp(ig\oint A)$. The trace is in the fundamental representation, with Hermitian generators normalized by $\mathrm{Tr}(T_aT_b) = \delta_{ab}/2$. Products are wedge products. On an oriented closed three-manifold, a pure gauge $\mathcal{A} = U^{-1}dU$ gives $n_{CS} = (24\pi^2)^{-1}\int \mathrm{Tr}(U^{-1}dU)^3$, an integer with the chosen winding convention. Reversing orientation reverses its sign. This convention keeps the cubic term and its factors of i consistent.

With the usual boundary conditions and a consistent normalization, this number records the winding of a pure gauge. A large gauge transformation can shift it by an integer. Away from a pure-gauge endpoint it need not be an integer. The valley picture labels the classical endpoints of a calculation. Their relation to physical states requires the quantum theory.

The ground-state restriction stays in place when the description acquires topological structure. To see what the proposed change acts on, I need to name that structure. Here is where quantum mechanics gets strange—stranger even than usual.

Quantum mechanics can give a nonzero amplitude for passage through a barrier even at an energy below its top. The amplitude comes from the wave equation, which still respects energy conservation. A standard formal construction combines winding-labelled classical vacua with a phase:

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{in\theta}|n\rangle \tag{8.2}$$

This is called the theta vacuum. Think of it as being smeared across all the valleys simultaneously, with the phase $e^{in\theta}$ determining how the different valleys interfere with each other quantum mechanically.

In QCD the physical angle is constrained by measurements of CP violation. Below I use the symbol θ again for a coefficient in a lattice action or statistical weight. Identifying that coefficient with the physical QCD angle would require a matching calculation.

Instantons are localized solutions of the Euclidean field equations. In the semiclassical picture they connect configurations with different winding labels. The time coordinate of that calculation is Euclidean. The path through the mountain represents a contribution to an amplitude.

An instanton takes you from sector n to sector $n + 1$. An anti-instanton takes you from n to $n - 1$. The tunneling amplitude is exponentially suppressed:

$$\Gamma \sim e^{-8\pi^2/g^2} \tag{8.3}$$

The exponent $8\pi^2/g^2$ is the single-instanton action in the usual normalization. It gives the leading semiclassical factor where that approximation applies. Inferring a physical rate would require a real-time model and its regime of validity. The radial spacing statistic later in the chapter comes from a finite sample in chosen coordinates.

That was the structure behind the question I called rectification. The next task was to say what would change in the model and what I would measure. Here is the question I took to the simulation. If I favor field configurations of one orientation in a statistical ensemble, does a loop observable register that preference?

Imagine a drawer full of photographs of a harbor, each showing a buoy and recording whether the tide was rising or falling. Suppose the buoy sits farther left in the falling-tide photographs. If I give those photographs more weight, the average position in my calculation moves left. I have not moved the buoy. In the lattice archive, each saved configuration plays the role of a photograph, and its recorded loop value plays the role of the buoy position. Changing the weight of a record can move the reported average without describing a new physical motion. The key insight comes from the field equations. Recall the holonomy composition law from Chapter 7:

$$\mathrm{Hol}(\gamma_1 \text{ then } \gamma_2) = \mathrm{Hol}(\gamma_2)\,\mathrm{Hol}(\gamma_1) \tag{8.4}$$

Here transport acts on column vectors, so the second route acts on the left, as in Chapter 7. Both loops use the same base point. In gauge theory, a Wilson loop records the accumulated parallel transport of the field around a closed contour. Changing a square to a rectangle changes the contour and can change the observable. A directional bias in the ensemble comes from the action or sampling weights. The rectangle gives us a different quantity to measure.

The proposed Davis term uses that loop measurement to change how configurations are weighted in the model. Before writing the term down, I need to separate a preference among configurations from a flow between them. Those are different things to count.

§ 8.2

Changing the model and naming the response

Detailed balance says that in the stationary distribution of the sampler, the probability weight of a transition equals the weight of its reverse:

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

The weighted average of any antisymmetric transition current then vanishes. A statistic evaluated on each configuration is a different object. Symmetry forces its average to zero if a transformation changes the statistic's sign and preserves the distribution.

Think about a revolving door on a windy day. You can count people crossing each way or count how many people in the lobby are wearing coats. Balanced traffic through the door says nothing about the fraction wearing coats. My loop statistic belongs to the second kind of count.

A small ring makes the distinction concrete: counting how often each place is occupied does not tell me how traffic travels between the places. Put three stations on a circle. At each step a walker can stay where she is, move clockwise, or move counterclockwise. Write p and q for the two moving probabilities and $1 - p - q$ for the probability of staying. Use the same rule at every station. Each column of the transition matrix sums to one, since a station receives probability from itself and from both neighbors. If a third of the walkers start at each station, a third will be at each one after the step. So $\pi_i = 1/3$ is stationary.

Now put a counter on the edge from station 1 to station 2. The probability per step of moving along the edge is $\pi_1 p$. The probability of moving back is $\pi_2 q$. Their difference is

$$I_{12} = \pi_1 p - \pi_2 q = \frac{p - q}{3}.$$

There is the same signed difference on each clockwise edge. Detailed balance requires the two probabilities across each edge to agree, so it makes each of these currents zero. More generally, if a transition tally changes sign when I reverse a step, its two contributions across any edge cancel under detailed balance.

Walker	Clockwise p	Counterclockwise q	Stay	I_{12}
Balanced	1/4	1/4	1/2	0
Circulating	3/8	1/8	1/2	1/12

A still picture of either chain at stationarity has the same distribution. Even a perfect count of the occupancies gives me no way to choose between the two rows. I have to see the transitions to know which way traffic goes. The current in this model is a difference of transition probabilities per step. To make a statement about work or power I would need the energy transferred along a step and the time it takes.

That is why I keep the definition of J_i beside its table. It is a statistic computed on one saved configuration. When I give the rows different weights, I can calculate exactly how its average moves.

A stationary distribution and a flow of probability are now distinct. I can specify the proposed change to the action without assuming that it supplies a source of work. I proposed modifying the action itself to favor one orientation in the ensemble:

$$S_{\text{total}} = S_{\text{Wilson}} + S_{\text{Davis}} \tag{8.5}$$

where the Davis term is:

$$S_{\text{Davis}} = \theta \cdot \text{Im}(\text{Tr}(W[C])) \tag{8.6}$$

Here $W[C]$ is the Wilson loop around a contour C, and θ is a coupling parameter that controls the strength of the symmetry breaking. Think of θ as a dial: at $\theta = 0$, no symmetry breaking. As you turn up θ, the bias gets stronger.

The added term is built from a loop. Following that loop tells us what the term actually reads. To understand the Davis term, we need to unpack the Wilson loop. This is one of the most important objects in gauge theory, and it deserves a careful explanation.

Return to the lattice square from Chapter 5. Each directed edge carries a matrix that converts a list of internal field components from one local frame to the next. Multiply those matrices in the order of the trip around the square. In a non-Abelian theory, changing the order can change the result, and the return can mix the components instead of multiplying them all by one scalar phase. That product is the loop matrix. Its trace is independent of the choice of internal frame at the starting corner.

Mathematically, the Wilson loop $W[C]$ is the path-ordered exponential of the gauge field around a closed contour:

$$W[C] = \mathcal{P} \exp\left(ig \oint_C A_\mu dx^\mu\right) \tag{8.7}$$

The $\mathcal{P}$ means "path-ordered"—you have to multiply the contributions in the right sequence, like reading a sentence from left to right. The integral $\oint_C$ means you go all the way around the closed loop C.

For an SU(3) gauge field (the group that describes quantum chromodynamics), the Wilson loop is a 3×3 unitary matrix. When you take its trace, you get a complex number. This is where the magic happens.

Reversing the oriented contour gives $W[C^{-1}] = W[C]^{-1}$, so the trace becomes its complex conjugate. The real part is unchanged and the imaginary part reverses sign. Charge conjugation and parity also act on the fields and the contour. To assign a CP parity I would have to specify those actions for the observable and the full sum of loops.

A term proportional to $\text{Im}\,\text{Tr}\,W[C]$ changes the statistical weight of oriented loops. Both square and rectangular contours have an orientation. Their response under a particular action is a question for calculation.

A direction-dependent response is easier to picture with a tool in your hand. A socket wrench makes it easy to hear the difference between asymmetry and a source of work. One way it catches. The other way it clicks. The handle still needs your hand.

Here *ratchet* names the idea of using a geometric asymmetry to bias a response. The proposed action study used a loop correlator of the form

$$J = \text{Im}\,\text{Tr}(EW), \tag{8.8}$$

where E and W must be represented at the same base point. In the archived TVR-003 calculation, J_i stores a sum of this kind of correlator over sampled origins, while D_i stores the corresponding sum of imaginary loop traces. A sum depends on the contour, the implementation of E, and how many origins are sampled. These details determine the units of the archived numbers.

§ 8.3

Keeping the two numerical experiments separate

I designed the validation as a controlled A/B experiment. This is the gold standard in experimental science: compare two conditions that differ in exactly one variable, and measure whether the outcomes differ.

A controlled comparison tries to isolate the change of interest. In a numerical study, that means specifying the target distributions, validating the samplers and quantifying uncertainty. A difference between two outputs is a starting observation; causal attribution depends on those controls.

The earlier TVR-002 report described two simulation protocols:

Arm A (Control): Standard heatbath Monte Carlo with the Wilson action only. No Davis term. The theory predicts $\langle J \rangle = 0$ for both symmetric and asymmetric loop geometries. This arm validates that the simulation is working correctly—that I am not seeing spurious signals from bugs or numerical errors.

Arm B (Experimental): Davis-modified Monte Carlo with the full action $S_{\text{total}} = S_{\text{Wilson}} + S_{\text{Davis}}$. The theory predicts:

- Symmetric loops ($R = T$): check whether $\langle J \rangle$ remains consistent with zero
- Asymmetric loops ($R \neq T$): check for a change in $\langle J \rangle$

The recorded diagnostic targets were:

1. Control arm must show $\langle J \rangle = 0$ within statistical error (checks a specified symmetry)
2. Experimental arm with symmetric loops must show $\langle J \rangle = 0$ (tests the proposed square-loop control)
3. Experimental arm with asymmetric loops must show $\langle J \rangle \neq 0$ with a displayed mean-to-error ratio above five (requires a justified error model)

These conditions would give a diagnostic for the action comparison. Isolating an action effect also requires matching or independently validating the two sampling procedures. A shared random seed alone cannot make different update algorithms equivalent.

These are the recorded settings for the direct A/B comparison. The archived reweighting calculation below has its own run count, loop and observable normalization.

Parameter	Value
Lattice size L	12^4 (production), 8^4 (testing)
Gauge group	SU(3)
Coupling β	6.0
Davis parameter θ	0.45
Asymmetric loop	3×6 or 4×8
Symmetric loop	3×3 or 4×4
Thermalization	500+ sweeps
Statistics	10,000 configurations
Decorrelation	20 sweeps between measurements

The production grid has 12 sites in each of four directions, or 20,736 sites. SU(3) describes the link matrices in the pure gauge model. The coupling β sets the coefficient of the Wilson action. This is a finite lattice at one coupling. A continuum study would require a controlled change in lattice spacing and volume, with a physical observable to set the scale.

Here θ is the coefficient recorded for the action comparison. Comparing it to the standardized reweighting coefficient below would require a conversion.

The parameter table is not the sampler. To know what distribution the calculation targets, I also need its acceptance rule. The standard Metropolis algorithm accepts a proposed link update $U \to U'$ with probability:

$$P_{\text{accept}} = \min\left(1, e^{-\Delta S_{\text{Wilson}}}\right) \tag{8.9}$$

With symmetric proposals and exact action differences, this rule satisfies detailed balance for the distribution proportional to $e^{-S_{\text{Wilson}}}$. The TVR-002 report describes the control arm as heatbath and the modified arm by this Metropolis rule, and the recovered harvest worker uses Metropolis. Comparing the two arms requires the same update rule in both, or two samplers validated against their targets.

The Davis-modified algorithm changes the acceptance criterion:

$$P_{\text{accept}} = \min\left(1, e^{-\Delta S_{\text{total}}}\right) \tag{8.10}$$

where $\Delta S_{\text{total}} = \Delta S_{\text{Wilson}} + \Delta S_{\text{Davis}}$.

Under the same conditions the modified rule satisfies detailed balance for the changed target $e^{-S_{\text{total}}}$. A stationary loop statistic can change while the sampler still satisfies detailed balance. Counting updates measures computational time.

The direct comparison describes a subsampled estimate of the action difference. Unbiasedness of that estimate alone does not ensure exact sampling because the acceptance rule is nonlinear. A replay would need to validate the implementation against the intended target.

§ 8.4

What the records say

The earlier TVR-002 report gave me the estimates in the next two tables. I can inspect the arithmetic, but I cannot yet repeat that 10,000-configuration run and its uncertainty calculation from matching saved inputs. I keep that distinction visible because a large number in the last column can be persuasive. Here it is simply the displayed mean divided by its displayed standard error in absolute value.

Arm A (control, recorded comparison):

Loop	$\langle J \rangle$	σ_J	**Significance**
Symmetric (3×3)	-0.0003	0.0021	0.14σ
Asymmetric (3×6)	$+0.0008$	0.0019	0.42σ

Both estimates are smaller in magnitude than their quoted standard errors. In a fully documented comparison that would be one check on the control.

Arm B (modified action, recorded comparison):

Loop	$\langle J \rangle$	σ_J	**Significance**
Symmetric (3×3)	$+0.0012$	0.0024	0.50σ
Asymmetric (3×6)	$+0.0341$	0.0023	$\mathbf{14.8}\sigma$

The original estimate of 0.0341 ± 0.0023 is distinct from the calculation below. The ratio of the printed numbers is about 14.8. A false-positive probability would require a specified null, a justified sampling distribution, and allowance for choices made after inspecting the data.

For TVR-003, I could go further. The December 2025 measurements were still there, so I could put the same rows through the calculation again. There were eight batches of 200 configurations from a 12^4 lattice, with recorded $\beta_{\text{code}} = 6.0$. I will return to that coupling's convention below. Each stored J_i and D_i sums measurements over 4,096 sampled origins of a 2×6 loop. Their Pearson correlation was 0.83582. I wanted to see what happened to the average of one when I gave more influence to larger values of the other.[1] Write $z_i = (D_i - \bar{D})/s_D$, with s_D the standard deviation over this archive using denominator N. Then

$$\widehat{J}(\theta) = \frac{\sum_i w_i(\theta) J_i}{\sum_i w_i(\theta)}, \qquad w_i(\theta) = \exp(-\theta z_i).$$

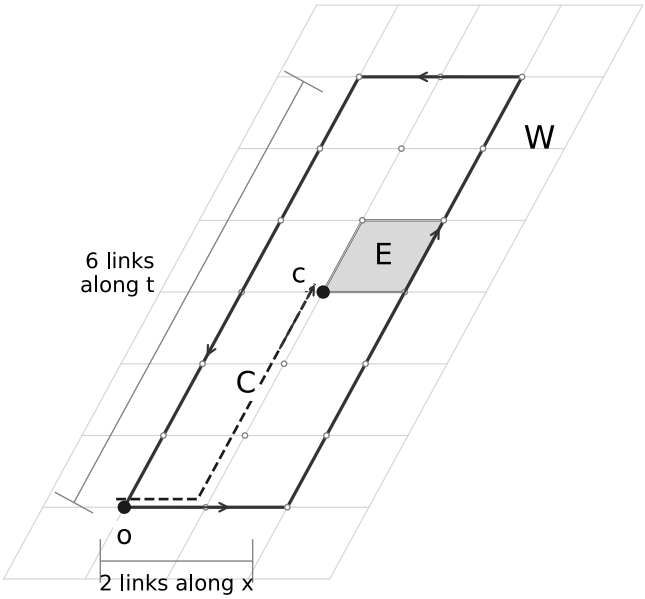
The selected setting is $\theta = -1.15$ with D_i standardized as above. The December summary gives the weighted result as about 27.4 ± 1.8 in the recorded units. Below I use GIGI to compare that setting with equal weights on the same 1,600 rows. Then I leave out groups from both estimates together and watch how the difference moves.

[1] The merged measurements match the eight source batches entry for entry. The detailed file record is given at the end of this chapter.

Follow the two loops into one record

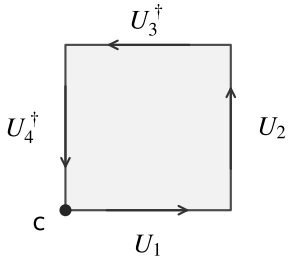
TVR-003 / MATCHING HARVEST GEOMETRY / RECORDED SUMS

01 One origin in the lattice plane



Dashed C: one x-link, then three t-links.

02 The inner cell, enlarged



03 Compare at one base

$$E = U_1 U_2 U_3^\dagger U_4^\dagger$$

$$E \text{ at } o: \; CEC^\dagger$$

E starts at c. CEC† expresses it at o, where W starts.

04 The written product uses a single reference frame

C	E	$C^\dagger$	W

$$j = \mathrm{Im}\,\mathrm{Tr}(CEC^\dagger W)$$

05 One saved configuration; 4,096 sampled origins

$D_1 = \Sigma\, d = -6.27285$ $J_1 = \Sigma\, j = -34.54761$

d = Im Tr(W). Origins may repeat; these are dimensionless sums.

Paths follow the matching worker. The replay begins with the saved sums, not recovered link fields.

Figure 27. One saved TVR-003 row and the loop geometry in the matching harvest source. The connector C brings the inner loop E to the origin of the outer loop W before their matrices are multiplied. The drawing is a lattice-plane projection; the numerical row is a sum over sampled origins.

The high-D observations in this archive tend to have high J. Give them more weight and the average rises. In fact, differentiating the weighted mean gives $\widehat{J}'(\theta) = -\operatorname{Cov}_\theta(J, z)$. At zero bias that slope comes directly from the covariance in the original ensemble. I wanted to follow that response: how much would survive if I changed the ensemble, the observable or the way I estimated uncertainty?[2]

Before asking what that change might mean physically, I needed to know what each number on the page had measured. I could repeat the weighted average from the saved rows. Tracing those rows all the way back to the program that generated the fields left another question open.

I found a program for collecting measurements whose output layout, loop contour and sample count matched the saved data. It measured a 2×6 loop at 4,096 randomly sampled origins, allowing the same origin to be chosen more than once. That gave me a concrete way to examine the conventions. I could not establish that this exact version had generated the historical run, so the connection remains conditional.[3]

The first thing to settle is what the number beside the coupling actually multiplies. In that matching program, a proposed link change has local action difference

$$\Delta S = -\beta_{\mathrm{code}} \operatorname{Re} \operatorname{Tr}[(U' - U) S_{\mathrm{staple}}],$$

using the ordinary matrix trace. The Wilson convention printed in this chapter has $\beta_{\mathrm{W}}/3$ multiplying that trace. If this program generated the saved configurations, its $\beta_{\mathrm{code}} = 6$ corresponds to $\beta_{\mathrm{W}} = 18$ under the printed convention. Calling both numbers "beta six" would hide a factor of three. I can still add and reweight the saved measurements. To give the original ensemble a physical calibration, I need to settle which program and normalization generated it.

For the same matching worker, the stored observables are

$$D_i = \sum_{x \in \mathcal{O}_i} \operatorname{Im} \operatorname{Tr} W_x,$$
$$J_i = \sum_{x \in \mathcal{O}_i} \operatorname{Im} \operatorname{Tr}(C_x E_x C_x^\dagger W_x).$$

Here $\mathcal{O}_i$ is the sampled list of origins, including repetitions. The connector C_x transports the central plaquette matrix E_x to the base point of the outer loop W_x. The symbol E_x names a loop matrix in this code; it is not a calibrated electric-field measurement. These are dimensionless sums, not energy or power readings.

Think of 4,096 spot checks on a large map. Some places may be checked twice and others not at all. Adding the spot checks is a well-defined statistic. To call it the integral over the whole map, I would need to account for the sampling rule. Reweighting these saved sums is exact for the saved list; it is not automatically equivalent to drawing a fresh ensemble from a full modified action.

Once the inputs and the weighting rule are clear, I can ask how much of the result follows from the association already present in the saved rows. The records already show a strong association between D and J. We can measure how much of the reweighting response that association explains

[2]Bee Rosa Davis, *TVR-003: Topological Vacuum Rectification, Summary of Experimental Results*, December 2025, included in the versioned Zenodo record `https://doi.org/10.5281/zenodo.17942784`.

[3]The recovered harvest worker uses a Metropolis update with three hits per link. Its measurement schema, contour and sample count match the archive, but the historical job's exact source revision has not been authenticated. A matching implementation does not replace a record tying the job to that revision.

without adding a physical interpretation. Fit an ordinary least-squares line to the same archive:

$$J_i = \bar{J} + bz_i + e_i, \qquad b = 23.46206279.$$

The residuals have mean zero in the unweighted archive. Weighted averaging gives the exact finite-sample identity

$$\widehat{J}(-1.15) - \bar{J} = b\widehat{z}_w + \widehat{e}_w.$$

Quantity	Replayed value
Equal-weight mean $\bar{J}$	1.45835038
Weighted mean $\widehat{J}(-1.15)$	27.42851360
Weighted standardized tilt $\widehat{z}_w$	1.10930927
Contribution $b\widehat{z}_w$	26.02668384
Weighted residual $\widehat{e}_w$	−0.05652061
Total change	25.97016322

Almost all of this change is accounted for by the fitted linear association already in the records. Give more influence to the high-D rows, and you also give more influence to rows with high J. That is a real, reproducible change in the answer to a weighted question. It does not show that a device produced work. To answer that next question, I need the energy ledger in Chapter 9.

I also wanted to know how much depended on where I set the dial. I kept the December summary's −1.15 after standardization so I could repeat that result. The scan I could run on these rows preferred −1.10, but only slightly: its mean-to-error ratio was about 15.30863, compared with 15.28845 at −1.15. That does not make −1.15 a uniquely established best setting.[4] The units of the dial matter too. Its coefficient on raw D is $-1.15/s_D \simeq -0.04880$, not the +0.45 coefficient in the earlier direct-comparison report.

When I leave out a group of rows below, I keep that dial and the original standardization fixed. I am asking how much those particular measurements influenced the answer at that setting. If I wanted to include the uncertainty of choosing the setting, I would have to repeat the choice as well. Leaving out groups does not, by itself, make the original samples independent or give the result a universal confidence level.[5]

At the opposite sign, the same archive gives $\widehat{J}(+1.15) = -26.37401$. The sum of the two signed estimates is 1.05450, with paired 20-block jackknife standard error 3.30762. That is compatible with approximate antisymmetry at this resolution. The pairing matters because both answers use the same records.

Now turn the dial the other way. Does the estimate reverse with it? A mirror test asks how the estimate changes when the sign of the bias changes. Exact oddness follows if the baseline measure is invariant under a transformation that sends both D and J to their negatives. A sign-dependent numerical error can also reverse sign, so the test is one diagnostic. The recorded direct-comparison estimates were:

[4]The recovered scan tests 161 points on $[-4, 4]$, maximizing the absolute mean-to-error ratio subject to a weight-only effective count of at least 10%.

[5]The later March analysis uses a different error estimator. Its ratio does not authenticate the December coefficient choice.

Parameter	$\langle J \rangle$	Deviation
$\theta = +0.45$	$+0.0341 \pm 0.0023$	—
$\theta = -0.45$	-0.0338 ± 0.0022	0.09σ

The two recorded estimates are compatible with approximate antisymmetry at the quoted precision. The 0.09 figure treats the reported errors as independent. In the archived reweighting calculation, the reversed setting gives -26.37 ± 2.42 using the same 20-block procedure. These two reweighted results share the same configurations and sources of uncertainty.

Another number appears in the exploratory record, but it belongs to the separate 200-configuration geometry study from Chapter 5. It must travel with that sample. The 85× figure is reproduced by sorting the radii of 200 configurations on an 8^4 lattice at $\beta = 6.0$. The radii are distances from the centroid in the retained principal-component coordinates. The largest adjacent spacing divided by the median adjacent spacing gives

$$\frac{17.80398 - 15.90758}{0.0221894} \simeq 85.46.$$

The analysis used a 10^{-10} stabilizer in the denominator, which has no visible effect at this precision.

Picture lining up 200 trees by their distance from your front door. Most of their distances differ by a little. Then one tree sits much farther out. The number compares the last gap in that list with an ordinary gap. It gives us a reason to go look at that tree.

Here the large gap isolates one configuration with recorded charge $Q = 0$. The saved clustering output identifies no clusters. The statistic describes an outlying radius in these coordinates. A spectral gap is an energy difference for a specified Hamiltonian. Connecting the two would require a separate derivation. The 1,600-configuration TVR-003 archive above is a separate dataset.

Putting the checks beside one another is useful as long as each number still points to its own calculation. This checklist records the exploratory diagnostics. Some share the same data or depend on the same representation. The values serve as pointers to the individual calculations.

#	Test	Result	Interpretation
1	TVR-003	15.3	Mean / block error at chosen tilt
2	Coefficient scan	0.45	Direct comparison setting
3	Mirror check	Pass	Approximate sign reversal
4	Temporal stability	Pass	Checks within-run drift
5	Stability	6.8% CV	Batches at one volume
6	Noise injection	Pass	Sensitivity to added noise
7	Bootstrap	Historical	Null procedure unverified
8	Multiple β	Pass	Sampled couplings only
9	Autocorrelation	$\tau = 0.60$	Earlier five-sweep lag run
10	Basin detection	Kurtosis 0.02	Fourth moment only
11	Phantom zone	Unmatched	Earlier and replayed peaks differ
12	Gap ratio	85.46	Separate 200-radius 8^4 study
13	Reweighting	Pass	Offline coefficient scan
14	Spectral geometry	Varies	GUE or GOE by graph settings

The batch comparison is one place where I have to ask exactly what changed between measurements. The 6.8% is a coefficient of variation: the standard deviation divided by the mean of eight

batch summaries of $\mathrm{RMS}(J)/L^4$. Every batch used $L = 12$. I changed batches, not the size of the box, so this cannot tell me how the response scales with physical volume. The listed bootstrap $p < 10^{-6}$ also needs more than its printed value. I cannot reproduce a false-positive probability without the null model, blocking and resampling procedure that produced it.[6]

§ 8.5

Asking what drives the change

The reweighting gives a change in expectation. To learn which features of the data are responsible, I can ask how the pair of measurements responds to a tilt, which patterns persist when I change scale, and how a graph constructed from the data carries diffusion. These calculations probe the representation I supplied.

I returned to those same rows in March with a question about the weights themselves. Before I divide by their sum to calculate a weighted mean, how much total weight has the tilt given the archive? At zero tilt each row counts once. At the chosen setting I computed the mean weight and took its negative logarithm, obtaining $-\log(N^{-1}\sum_i e^{-\theta z_i}) = -0.64705$. I can recover that number from the December measurements.

There is a detail in the arithmetic that changes how I read the sign. I centered the z_i, so their mean is zero. On the exponential curve a step above zero raises the weight more than an equal step below zero lowers it. Jensen's inequality gives the exact statement even when the steps are unequal: the average exponential is at least the exponential of the average, here $e^0 = 1$. Taking the negative logarithm therefore gives a value at or below zero for either direction of tilt. The magnitude tells me something about how these particular measurements spread out under the weighting. To turn it into energy delivered to a load I would still need a physical model connecting those weights to an energy scale and an account of the work required to complete the cycle.[7] If D is a scalar observable, its tilted average is

$$\langle D \rangle_t = \frac{\sum_i D_i e^{t D_i}}{\sum_i e^{t D_i}}. \tag{8.11}$$

Its derivative is $\mathrm{Var}_t(D)$. That is the first check I would make before attributing any change to curvature. Even points on a flat line change their weighted average when I favor one end.

I also compared the weighted mean of J with a linear prediction based on D. The difference tells me what that fitted line leaves unexplained. It depends on how I fit the line, center the data and define the tilt; it does not become a curvature measurement or a Jensen inequality just because it is nonzero. On these saved rows, the largest absolute residual occurs at the scan endpoint -3. The earlier table places a peak near 0.45. I cannot treat those as the same result until I can identify the definition or run behind the earlier value.[8]

[6]The public TVR-003 summary lists 13 tests; the longer checklist adds a spectral comparison. The bootstrap value is retained as a reported result whose generating procedure has not been matched.

[7]The March 27 replay is recorded in the Davis–Wilson lattice project, `experiments/tvr003_free_energy.py`. It standardizes the saved D_i before applying the tilt.

[8]The recalculation uses the archived geometry script and the saved TVR measurements. The original 0.45 location remains unmatched to that calculation.

The weighted mean asks which rows get more influence. I can ask a different question of the point cloud: which features persist as the scale of comparison changes? Persistent homology is a technique from topological data analysis (TDA) that detects "holes" in data—structures that persist across multiple scales. It is one of the most powerful tools for finding hidden structure in complex datasets.

Here is the intuition. Imagine you have a collection of points scattered in space—maybe cities on a map, or stars in the sky, or in our case, field configurations in an abstract space. Now imagine drawing a small ball around each point. At first, the balls do not touch. As you make the balls bigger, they start to overlap and merge.

As the balls grow, interesting things happen. Sometimes a ring of points creates a "hole" in the middle—a region that the merged balls surround but do not fill. If that hole persists as you keep growing the balls, it suggests genuine circular structure in the data. If it appears briefly and then fills in, it was probably just noise.

Persistent homology tracks these births and deaths of holes. A "persistent" hole is one that appears at small scale and survives to large scale. Its lifetime measures persistence in that filtration. Statistical significance needs an additional null or noise model.

A persistence diagram lets me mark the scale where each feature appears and where it disappears. But the picture changes when I change the point cloud, its scaling, the filtration or the threshold. In one original tabulation there were 47 one-dimensional features, a maximum persistence of 0.342, and 12 above 0.1. The follow-on report gave 256 features and a maximum near 0.168. Before comparing those counts, I need to put the same choices beside both pictures. A long-lived feature is something to investigate. Without a noise model, its lifetime is not a significance test or a proof of a physical mass gap.

The diffusion calculation from Chapter 3 provides another view of the same representation. Here the operator comes from the graph built on the recorded data. The heat kernel $K(x, y, t)$ describes how heat diffuses from point x to point y in time t. For a manifold, the heat kernel encodes the geometry through the spectrum of the Laplacian:

$$K(x, y, t) = \sum_n e^{-\lambda_n t} \phi_n(x) \overline{\phi_n(y)} \tag{8.12}$$

Here λ_n are diffusion decay rates, and ϕ_n are the corresponding spatial modes. In this calculation they belong to the graph built from the sampled feature vectors. They are not energies of the Yang–Mills Hamiltonian merely because both calculations use spectra.

What does this have to do with TVR? The heat kernel tells you how information spreads through a space. If you put a drop of dye in water, the heat kernel describes how the color diffuses. Fast diffusion means the space is well-connected. Slow diffusion means there are barriers.

For diffusion on a connected finite graph, the smallest positive Laplacian eigenvalue controls the slowest nonconstant mode. That gives me a rate for this graph. It also makes the way I built the graph part of the result. With the normalized Gaussian graph and 50 neighbors, I get about 0.0057442; the earlier tabulation's 0.0847 has not been matched to those settings. Even the preferred spacing fit changes: GUE at 50 neighbors, GOE at 20 and 80. I want that sensitivity in view before giving the fit a physical meaning. To infer physical time-reversal breaking, I would need a physical evolution operator and a derivation connecting its symmetries to this graph spectrum.

GUE and GOE name two random-matrix reference ensembles, respectively unitary and orthogonal. In this comparison they supply models for the distribution of normalized eigenvalue spacings. A better fit to one reference says something about that statistic and those graph settings. It does not identify the physical symmetry of a field theory whose evolution operator we have not reconstructed.

Let us now follow the reweighting itself, row by row in a small example, before returning to all 1,600 records. I can do something to the archive that I cannot do to a meter. I can go back to the same sixteen hundred rows, give some rows more say than others, and ask how the answer changes. Imagine taking temperature readings at several places in a room, including beside a radiator. Suppose the readings beside it are higher than the others. Give those warmer readings extra weight, and the reported average rises. The saved temperatures have not changed, and recalculating their average has not heated the room. To understand my lattice result, I wanted to see exactly which part of the calculation made the average move.

Write J_i for the recorded loop statistic of configuration i, and z_i for the standardized value of its companion statistic D_i. The origin and scale of z_i come from the complete archive. At each θ I use

$$Z(\theta) = \sum_{k=1}^{N} e^{-\theta z_k}, \qquad p_i(\theta) = \frac{e^{-\theta z_i}}{Z(\theta)}, \qquad \widehat{J}(\theta) = \sum_{i=1}^{N} p_i(\theta) J_i. \tag{8.13}$$

The p_i are nonnegative and sum to one. The parameter θ is dimensionless in this convention, since z_i has been standardized. At $\theta = 0$ all N rows have weight $1/N$. Positive θ gives more weight to rows with smaller z_i.

Here is the smallest version of that operation. Two hypothetical readings, 0 and 10, have equal-weight mean 5. Give the second reading three times the weight of the first and the mean becomes $(0 + 3 \cdot 10)/4 = 7.5$. Nothing happened to either reading. I changed the question from an equal-weight average to an average favoring the second entry. The lattice calculation uses a smooth rule for those weights and many more rows, but I need to keep that distinction just as visible.

$$p_i'(\theta) = p_i(\theta)(\bar{z}_\theta - z_i), \tag{8.14}$$

$$\widehat{J}'(\theta) = -\sum_i p_i(\theta) J_i z_i + \widehat{J}(\theta)\bar{z}_\theta = -\operatorname{Cov}_\theta(J, z). \tag{8.15}$$

Here $\operatorname{Cov}_\theta(J, z)$ is the covariance formed with the same weights $p_i(\theta)$. The minus sign has a concrete meaning. If large J tends to arrive in the same rows as large z, weighting toward smaller z pulls the mean of J down. If J and z have zero weighted covariance, the first derivative vanishes at that setting. A larger change in θ can still change the mean as the weights and covariance evolve.

At zero bias, the archive has population standard deviation $s_J = 28.0708$ and correlation $r(J, z) = 0.83582$. Since $s_z = 1$ by construction, $\operatorname{Cov}_0(J, z) = r(J, z) s_J \simeq 23.4621$. The first-order calculation at $\theta = 0.10$ therefore predicts $1.45835 - 0.10(23.4621) \simeq -0.8879$. Substituting 0.10 into Equation (8.13) gives -0.8952 on the same rows. That is a check on the local slope. The prediction and the recomputation both use the finite archive.

A stronger tilt can also concentrate the answer on fewer rows. The effective sample size measures that concentration of weights. There are still 1,600 rows after reweighting. Their influence can be concentrated in a much smaller set. To measure that concentration, I compare $\sum_i p_i^2$ with what it would be if m rows each had equal weight $1/m$ and the remaining rows had weight zero. In that case $\sum_i p_i^2 = 1/m$. The equivalent count for general weights is

$$N_{\text{eff}}(\theta) = \frac{1}{\sum_i p_i(\theta)^2} = \frac{\left(\sum_i w_i(\theta)\right)^2}{\sum_i w_i(\theta)^2}, \qquad w_i = e^{-\theta z_i}. \tag{8.16}$$

The range is $1 \le N_{\text{eff}} \le N$. The upper endpoint follows from Cauchy–Schwarz, $(\sum_i p_i)^2 \le N \sum_i p_i^2$. Equal weights attain it. Putting all the weight on one row gives the lower endpoint. There is also a variance interpretation in a restricted case. For independent measurements with common variance σ^2 and fixed weights, the weighted average has variance $\sigma^2 \sum_i p_i^2$. Here the count is useful as a diagnostic of the weights themselves. Correlated configurations and the relationship between J_i and w_i require a separate uncertainty calculation.

Try four rows with $z = (-1, -1, 1, 1)$ and $J = (-2, -2, 2, 2)$. With equal weights the mean is zero. At $\theta = \log 2$ the two $z = -1$ rows each have weight 0.4 and the two $z = 1$ rows each have weight 0.1. The mean is now -1.2, and the effective count is $1/(2(0.4)^2 + 2(0.1)^2) \simeq 2.94$. I have four entries in my notebook, and the first two make four fifths of the answer.

The concentration grows as the bias moves away from zero in the archive sweep in Figure 28. For these exponential weights, $N_{\text{eff}} = Z(\theta)^2/Z(2\theta)$, so

$$\frac{d}{d\theta} \log N_{\text{eff}} = 2\big(\bar{z}_{2\theta} - \bar{z}_\theta\big).$$

Equation (8.15) with $J = z$ gives $d\bar{z}_\theta/d\theta = -\operatorname{Var}_\theta(z) \le 0$. Thus the effective count decreases as the tilt grows in either direction from zero.

The archive arrived in eight batches. Keeping those batch identities lets me follow how much each contributes to the weighted answer.

I kept the measurements in their original order, eight batches of 200 pairs (J_i, D_i). I used the full collection to standardize D_i once. Now watch what the chosen dial does to just two rows: at $\theta = -1.15$, a row one standard deviation higher in D gets $e^{1.15}$, or about 3.16, times the weight. The centering term cancels in that ratio.[9]

For each row I formed w_i, $w_i J_i$, and w_i^2. I asked GIGI to add the row counts and the columns J_i, D_i, w_i, $w_i J_i$, and w_i^2 across the whole collection, then separately for each batch and each contiguous block. When I left a group out, GIGI added what remained. Dividing those sums gave me the means; comparing the omissions gave me the jackknife standard errors. I could now follow how much each group contributed to the answer.[10]

[9] The pairs in the merged file match the original batch files in order.

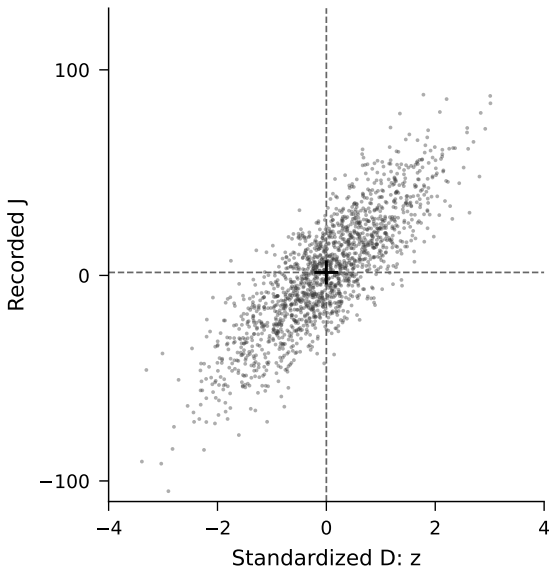
[10] GIGI sums integer ticks of 10^{-10}. The companion calculation divides the sums and computes the jackknife standard errors, and the resulting means were checked against a separate calculation on the original arrays.

THE GEOMETRY OF FUEL TVR / 03

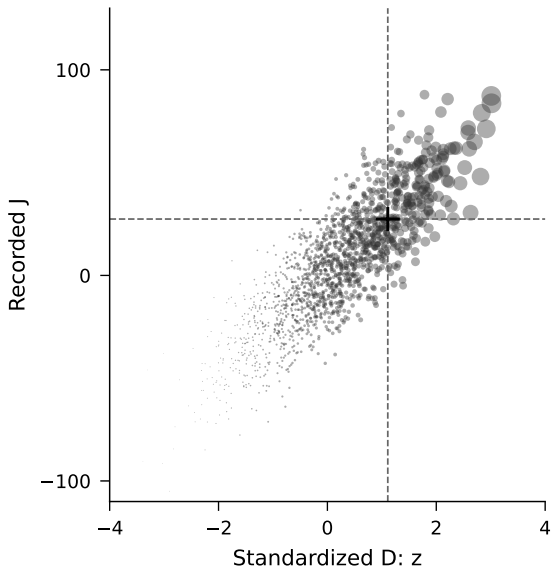
The records stay; their influence moves

ALL 1,600 ARCHIVED PAIRS / θ = −1.15

01 Equal influence



02 Area proportional to weight



03 Follow the center of the weighted cloud

Equal weights	z̄ = 0	J = 1.45835
Tilted weights	z̄w = 1.10931	J̃ = 27.42851

04 The same first saved row, before and after weighting

$z_1 = -0.329795$	$w_1 = 0.684365$
$J_1 = -34.547605$	$w_1 J_1 = -23.643161$

Crosses mark the two centers; no recorded point changes position.

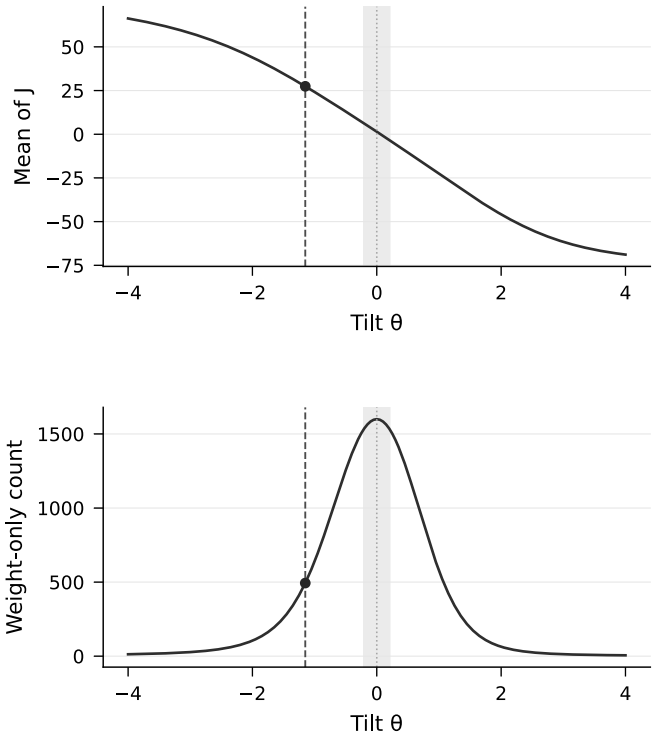
Dot area in the tilted view is proportional to exp(1.15z). The cross is the weighted center.

Figure 28. The same 1,600 recorded pairs under equal and tilted weights. Marker area in the right-hand view is proportional to weight; coordinates do not move. The lower views follow the changing center and the first saved row.

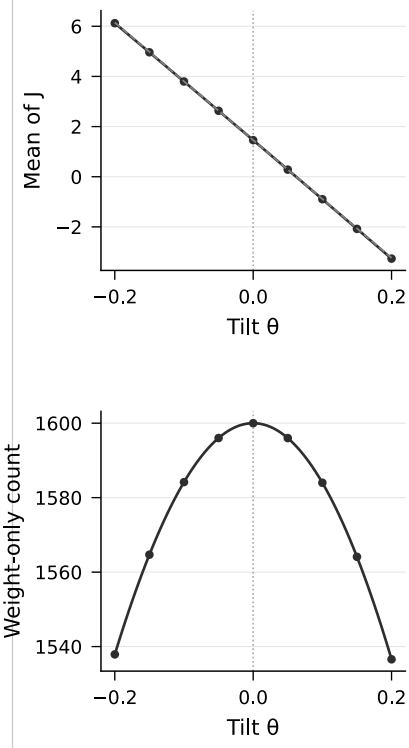
See the whole tilt, then zoom in

FINITE SUMS / FIXED STANDARDIZATION / θ* = −1.15

01 Full range, −4 to +4



02 Near zero, ±0.2



03 Locate the selected setting and the local slope

	Mean of J	Weight-only count
θ = 0	1.45835	1600.00
θ = −1.15	27.42851	493.44

$$\widehat{J}'(0) = -\mathrm{Cov}_0(J, z) = -23.46206$$

Shading: enlarged interval. Local dashed line: tangent prediction.

The selected point is inside the full-range plot. Weight-only count does not measure independence.

Figure 28. The finite-sum response over the full dial range and near zero. The selected setting is −1.15. The local tangent comes from the covariance on these same rows. The weight-only count measures concentration, not independence.

Table 8.1 gives the mean within each batch with equal weights and at the selected tilt. The weight share is that batch's sum of weights divided by the sum over all eight. Equal weights give each batch 12.5% of the total. The row counts stay at 200 as the shares move.

Batch	File time	Rows	Equal mean	Tilted mean	Weight share
1	15:23:53	200	-2.59119	29.14592	10.876%
2	15:23:55	200	4.73014	27.01508	11.356%
3	15:23:56	200	0.05884	20.20528	11.387%
4	15:23:58	200	0.70134	21.24095	11.934%
5	15:23:59	200	5.59070	33.73719	15.928%
6	15:24:27	200	-3.14961	24.15505	10.909%
7	15:24:28	200	1.70698	25.46791	12.788%
8	15:24:29	200	4.61961	34.33784	14.821%

To recover the full tilted mean from this table, multiply each batch's tilted mean by its fraction of the weight and add the products. That follows by collecting the weighted sum by batch:

$$\frac{\sum_b \sum_{i \in b} w_i J_i}{\sum_b W_b} = \sum_b \frac{W_b}{\sum_c W_c}\, \widehat{J}_b, \qquad W_b = \sum_{i \in b} w_i.$$

For example, the first batch contributes about $0.108763 \cdot 29.14592 = 3.17$ of the full tilted mean. A simple average of the eight tilted means would give each batch a weight share it did not have.

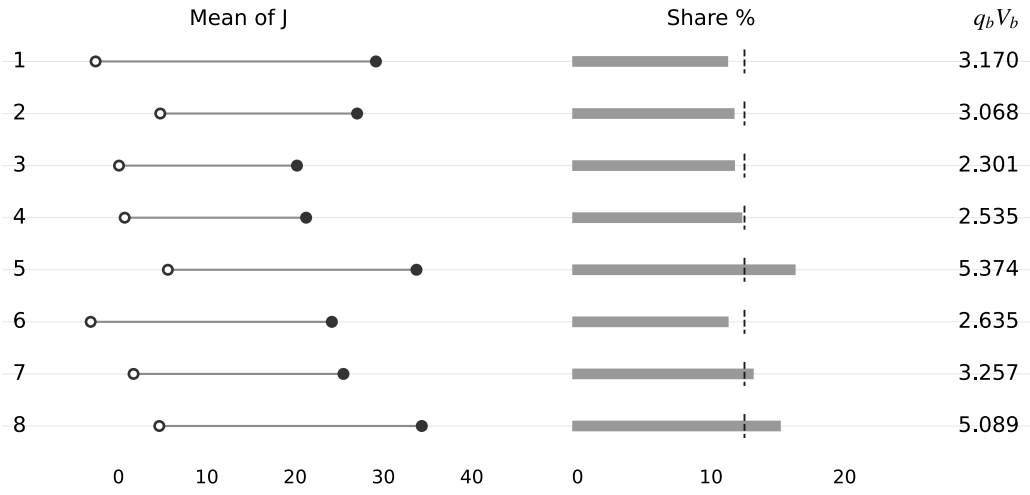
Was the change mostly caused by giving some files more weight? Let $q_b = W_b / \sum_c W_c$ and let U_b, V_b be the equal and tilted means within batch b. Add and subtract $\sum_b q_b U_b$ to separate the two changes:

$$\underbrace{\sum_b q_b V_b - \tfrac{1}{8}\sum_b U_b}_{25.970163} = \underbrace{\sum_b q_b (V_b - U_b)}_{25.632918} + \underbrace{\sum_b (q_b - \tfrac{1}{8}) U_b}_{0.337245}.$$

Keep every batch in the same register

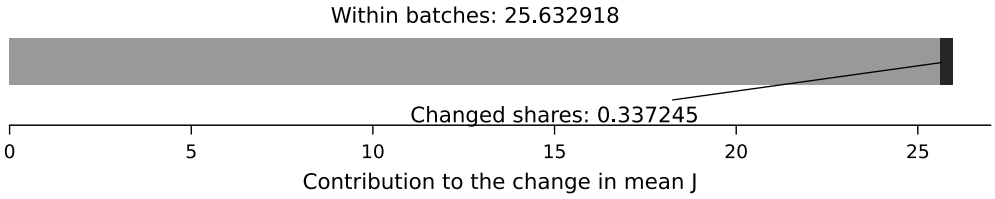
EIGHT SOURCE FILES / 200 ROWS EACH / θ = −1.15

01 Means, weight shares and contributions



Open: equal mean. Filled: tilted mean. Dashed share marker: 12.5%.

02 Separate the two sources of the pooled change



03 Add the contributions, not the batch means

$$\sum_b q_b V_b = 27.42851 \qquad \sum_b q_b = 1$$

25.970163 = 25.632918 + 0.337245

Every batch mean rises. The archive still contains the same eight groups of 200 records.

Figure 29. Each horizontal row follows one original batch: its two means, its share of total weight, and its contribution to the pooled tilted mean. The lower decomposition separates changes within batches from changed batch shares.

Every batch mean rises under the tilt. The within-batch change accounts for almost the entire pooled shift.

Both estimates use the same configurations. An uncertainty calculation should keep that pairing when it leaves out a block of rows. The two means use the same 1,600 configurations. At equal weights each row has the same influence. At the archived setting $\theta_* = -1.15$, the weight $w_i = \exp(1.15 z_i)$ gives configurations with larger z_i more influence. I keep the setting and the full-archive standardization fixed when I leave out a group. I renormalize the weights over the rows that remain. That is the same calculation on a shorter list.

Write $W = \sum_i w_i$, $S_w = \sum_i w_i J_i$, and $S_0 = \sum_i J_i$. The two full means are $U = S_0/N$ and $V = S_w/W$, with $N = 1600$. If group b contains n_b rows, its three sums are W_b, S_{wb}, and S_{0b}. Removing the group subtracts those sums from the full sums, giving

$$U_{-b} = \frac{S_0 - S_{0b}}{N - n_b}, \qquad V_{-b} = \frac{S_w - S_{wb}}{W - W_b}, \qquad \Delta_{-b} = V_{-b} - U_{-b}. \tag{8.17}$$

Both means lose the same group. Each group gets a turn being left out.

Take the first 80 rows. With the unscaled exponential weights they contribute $W_b = 114.0443$, $S_{wb} = 2070.7687$, and $S_{0b} = -772.7632$, rounded here. The weighted mean after their deletion is

$$\frac{83816.9312 - 2070.7687}{3055.8321 - 114.0443} \simeq 27.78792.$$

The equal-weight mean is $(2333.3606 - (-772.7632))/1520 \simeq 2.04350$. Both means rise on this omission. Their difference falls to 25.74442 because the equal-weight mean rises more. This is one of the pairs in Table 8.3.

Each omission now gives a difference between two means. Their spread is the quantity needed for the uncertainty in the change. For B equal-sized groups let $\overline{\Delta}_-$ be the average of the B deletion differences, and let $\Delta = V - U$ be the full-archive difference. The jackknife uses the pseudo-value $P_b = B\Delta - (B-1)\Delta_{-b}$. For an ordinary mean with groups of size n, multiplying by B gives the full sum divided by n. Multiplying the deletion mean by $B - 1$ gives the retained sum divided by n. Subtracting recovers the mean within group b. For the difference of means here, the same construction gives one pseudo-value per deletion. Its deviation from its average is

$$P_b - \overline{P} = -(B-1)(\Delta_{-b} - \overline{\Delta}_-).$$

Dividing the sample variance of those B pseudo-values by B gives

$$\widehat{\mathrm{SE}}_{\mathrm{JK}}(\Delta)^2 = \frac{B-1}{B} \sum_{b=1}^{B} (\Delta_{-b} - \overline{\Delta}_-)^2. \tag{8.18}$$

This is an estimate of uncertainty for the specified grouping. The pseudo-value identity gives the scaling in the formula. How well the estimate describes variation across new ensembles depends on the dependence among configurations and on how the groups represent it.

Imagine one omitted block pulls both means upward by about the same amount. Their difference barely moves. Treating the two estimates as independent would miss that shared motion. Pairing the omissions keeps it in the calculation. The covariance below records how the two deletion

estimates move together.

$$\widehat{\mathrm{SE}}_{\mathrm{JK}}(\Delta)^2 = \widehat{\mathrm{SE}}_{\mathrm{JK}}(V)^2 + \widehat{\mathrm{SE}}_{\mathrm{JK}}(U)^2 - 2\,\widehat{\mathrm{Cov}}_{\mathrm{JK}}(V,U),$$
$$\widehat{\mathrm{Cov}}_{\mathrm{JK}}(V,U) = \frac{B-1}{B}\sum_{b}\widetilde{V}_b\widetilde{U}_b.$$

For the 20-block calculation, the covariance is 1.28577 in squared loop-statistic units. Inserting the two standard errors gives

$$\sqrt{1.79407^2 + 1.10278^2 - 2(1.28577)} \simeq 1.36502.$$

The covariance term accounts for the two means tending to move together under the same omission.

Table 8.2. Standard errors from paired omissions at fixed weights. Each column uses the same grouping within a row. The full means are $U = 1.45835$ and $V = 27.42851$, and their difference is 25.97016.

Groups	Rows per group	SE(U)	SE(V)	SE($V-U$)
20 consecutive blocks	80	1.10278	1.79407	1.36502
Eight source batches	200	1.18019	2.03384	1.63481

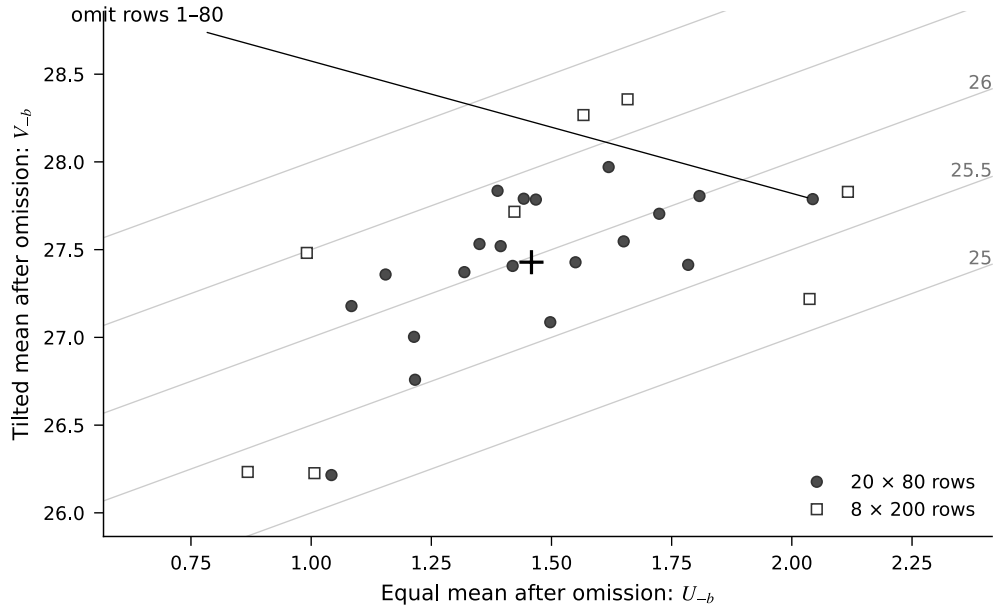
The shorter blocks follow the merged row order in groups of 80, so some straddle source batch boundaries. The longer groups follow the eight original files. With the longer grouping the paired error estimate rises by about 0.27. Both calculations use the same archive, and both condition on the choice $\theta_* = -1.15$. If I wanted the uncertainty of a procedure that searches for its setting, I would have to specify and repeat the search inside each resample.

The weight-only effective count at this setting is 493.44. Equation (8.16) gives that number from the weights alone. Even for fixed weights, correlated measurements add terms $2p_ip_j\,\mathrm{Cov}(J_i, J_j)$ to the variance of a weighted mean for each pair $i < j$. The weight count has no access to those covariances. The group deletions show how the actual archived means move when the rows are left out together.

Keep the omissions paired

FIXED TILT / FIXED STANDARDIZATION / PAIRED DELETIONS

01 A point is one paired omission



02 Uncertainty in the difference

Grouping	SE(U)	SE(V)	SE(V−U)
20 blocks	1.10278	1.79407	1.36502
8 files	1.18019	2.03384	1.63481

$$\mathrm{SE}^2(\mathrm{V}-\mathrm{U}) = \mathrm{SE}^2(\mathrm{V}) + \mathrm{SE}^2(\mathrm{U}) - 2\,\mathrm{Cov}(\mathrm{V},\mathrm{U})$$

20-block covariance: 1.28577. Diagonals hold the difference fixed.

This spread comes from grouped omissions. The weight-only count of 493.44 answers another question.

Figure 30. Each point removes the same group from both means. Diagonal lines hold their difference fixed. The two grouping schemes give separate jackknife calculations; the cloud is a view of the deletions, not a confidence region.

The paired deletion record

In Table 8.3 each row omits one group from both estimates. The top rows use 20 contiguous groups of 80. The lower rows use the eight original batches of 200. The original standardized values and tilt stay fixed throughout.

Contiguous groups 1–10

Omit	Equal	Tilted	Diff.
1	2.04350	27.78792	25.74442
2	1.49744	27.08642	25.58898
3	1.41914	27.40719	25.98805
4	1.08383	27.17835	26.09452
5	1.35018	27.53211	26.18193
6	1.31872	27.37184	26.05312
7	1.61888	27.97065	26.35177
8	1.44228	27.79019	26.34791
9	1.38777	27.83530	26.44753
10	1.80786	27.80567	25.99782

Contiguous groups 11–20

Omit	Equal	Tilted	Diff.
11	1.21397	27.00308	25.78911
12	1.21606	26.75848	25.54242
13	1.54997	27.42800	25.87803
14	1.65018	27.54716	25.89698
15	1.72415	27.70438	25.98022
16	1.46747	27.78523	26.31777
17	1.39425	27.51992	26.12567
18	1.78434	27.41335	25.62900
19	1.15480	27.35832	26.20352
20	1.04222	26.21532	25.17310

Original batches

Omit	Equal	Tilted	Diff.
1	2.03686	27.21893	25.18207
2	0.99095	27.48148	26.49052
3	1.65828	28.35670	26.69842
4	1.56650	28.26701	26.70051
5	0.86801	26.23325	25.36523
6	2.11663	27.82935	25.71272
7	1.42283	27.71600	26.29317
8	1.00674	26.22628	25.21954

Group 2 in the top table leaves out rows 81–160. Batch 2 in the lower table leaves out rows 201–400. Every entry uses the rows still in the file. The difference is computed before rounding.

The archived rows let me ask how much the loop mean moves when the same configurations count differently. Chapter 9 starts with a different record: the meter readings and the reset checks needed to close an energy ledger from start to finish.

§ 8.6

The distance from model units to physical power

The original calculation proposed a scaling of the form

$$P \propto \Delta^4 V \theta^2. \tag{8.19}$$

The tables below contain numerical proxies and signed observable means in model units. To use this expression for watts would require a derived physical energy and time scale. Testing the dependence on Δ would require independent values of that quantity across runs.

Volume scaling ($V = L^4$):

L	Volume	$\langle J \rangle$	Statistical proxy
8	4,096	0.0289	118.4
12	20,736	0.0312	647.0
16	65,536	0.0328	2,149.6
20	160,000	0.0341	5,457.6

Fit the four displayed rows in log coordinates by ordinary least squares and the slope is 1.0451, with $R^2 = 0.9999992$. That looks strikingly close to linear. But look at how the last column was made: to rounding precision, it is $V\langle J \rangle$. Multiplying by V already supplies most of that trend. To put uncertainty on the exponent, I would need the raw measurements at all four volumes; those matching runs have not yet been identified.

The rows disagree with the production record in another way. This table gives 0.0312 at $L = 12$. The production estimate earlier in the chapter, 0.0341 ± 0.0023, was reported for a 12^4 lattice, and here the same value appears beside $L = 20$. Both cannot describe the same run. Until I can identify the runs behind these rows, I cannot tell which label is right.

Theta scaling (θ):

θ	$\langle J \rangle$	$\langle J \rangle / \theta^2$
0.20	0.0068	0.170
0.35	0.0205	0.167
0.45	0.0341	0.168
0.60	0.0603	0.168

A log–log fit of the four positive settings gives exponent 1.988. A least-squares fit of $\langle J \rangle = c\theta^2$ through zero gives $c \approx 0.1677$ in the displayed units. These fits describe the shown range. A signed response with exact odd symmetry cannot have an ordinary quadratic leading term in a Taylor expansion about zero. Positive settings alone cannot determine the small-bias behavior.

A numerical scaling rule is only useful if its factors can be measured independently. It is easy to create a larger total merely by multiplying by a larger volume. If I were looking at a row of light meters, I could multiply the reading at one point by the area of a room. I would get a bigger number for a bigger room even before checking whether the light is uniform. To learn whether the multiplication describes the room, I would have to move the meters. In this calculation the corresponding check is to track the intensive observable and its uncertainty across independently sampled lattice sizes.

After following those calculations, I can point to two results and say what made each one happen. On the 1,600 configurations, changing the weights changes the mean of the loop statistic substantially. The same rows reproduce the correlation and approximate sign reversal. On the separate set of 200 configurations, sorting the radii gives 85.46 and singles out one outlying point in those coordinates. The older comparison and scaling tables give me other observations to investigate, but I still need their matching runs and uncertainty calculations before I can give them the same kind of check.

The physical questions remain. What observable would carry energy? What input changes the state, what comes out, and what has to be put back before the cycle can repeat? Those are measurements of a different kind from the ensemble means above.

The remaining step is the one that motivated the book: connecting a response in the model to work delivered by a physical system. Here is the question that brought me to this chapter: where would the energy come from?

Imagine lifting a marble to a shelf and giving it a ramp to roll back down. If you measure only the descent, the marble seems to supply energy on every run. The work that returns it to the shelf belongs in the same account. In an equilibrium ground state, passivity rules out extracting positive net work by a cyclic perturbation under the theorem's hypotheses.[11] A driven system raises a different question, and the drive belongs on the input side of its energy ledger.

My simulation computes expectation values from samples. To connect such a result to work, I would need a physical evolution, a defined energy flux, and a complete balance over a cycle that includes preparation and reset. That is what would tell me whether the shape I see in these data has anything to say about fuel.

I described in Chapter 6 why I initially set the idea aside. Part of that hesitation came from imagining what could follow if the physical effect were demonstrated. The archived study gives me a statistical response in a model. A claim about power density, self-limiting operation, or physical hazards would need evidence from a system that actually exchanges energy.

What follows here is a question about possible consequences if a physical effect is ever demonstrated.

I imagined both beneficial and destructive applications if the physical proposal ever worked. Those imagined consequences explain my hesitation. The archived statistic establishes neither a power source nor a weapon, and it supplies no measured energy-release scale.

This was the dilemma I faced. I wanted to be clear about what made the question worth asking and what I could establish from the model. The field equations motivate the proposed deformation. Its significance for a physical system would depend on how that system is specified and measured.

A safety claim would require evidence about the system itself:

- **Energy and rate.** A physical balance must include the energy supplied and dissipated, and measurements must establish what sets the rate.
- **Geometry and operating conditions.** A loop that changes a numerical observable identifies a feature of that calculation. To determine what a device would do, we would need to establish which physical observables respond and under what conditions.
- **Scale.** Any comparison across sizes would require a calibrated physical model and tests over the relevant range.

I would start with what can actually be measured. The archived calculation establishes a response of a loop statistic to a change in weighting. That is the evidence available for this discussion.

[11] W. Pusz and S. L. Woronowicz, *Passive States and KMS States for General Quantum Systems*, Communications in Mathematical Physics 58 (1978), 273–290, `https://doi.org/10.1007/BF01614224`.

Even the numerical settings need translation before they could specify a physical experiment. The parameters recorded in this chapter specify calculations. Matching them to a physical system would require stating how each parameter enters that system and how the observable is measured.

The reasons are fundamental:

- **Lattice vs. Continuum.** Our simulations operate on a discrete grid with finite spacing. Physical reality is continuous. The "bare" parameters on the lattice are not the same as "renormalized" physical parameters in the continuum limit. The relationship between them requires careful matching calculations that depend on the specific lattice formulation used.
- **Pure Gauge vs. Full QCD.** These simulations model pure SU(3) gauge theory—gluons only. Real quantum chromodynamics includes quarks. Fermion contributions modify the vacuum structure significantly. A selected coefficient for this pure-gauge archive does not determine a corresponding coefficient in full QCD.
- **Finite Volume Effects.** A finite lattice has a finite box. Converting its size to meters requires a physical calibration. Establishing a large-volume trend requires an observable with uncertainty at several sizes.
- **Temperature.** Lattice simulations can probe zero-temperature physics. Physical devices operate at finite temperature. The phase structure of QCD changes dramatically with temperature, and any proposed operating conditions would need a new physical analysis.
- **Implementation Medium.** A physical device might not use QCD at all. It might exploit similar topological structures in condensed matter systems, metamaterials, or effective field theories. The "θ" in such implementations would be a completely different physical quantity.

A simulation coefficient has meaning in the calculation that defines it. Translation to a physical experiment has to be earned through the mapping from parameters and observables to a measured system.

§ 8.7

Keeping the calculation attached to its record

I want to leave the two experiments where someone else can pick them up. In the proposed action comparison, I change the rule used to generate configurations and compare the outcomes. In the TVR-003 calculation, I return to configurations already saved and change their weights. For that second experiment I can show the rows and repeat the sums. The following equations and tables keep the two starting points in view.

System: SU(3) Yang-Mills on L^4 lattice, periodic boundary conditions.

In plain language: we simulated the mathematical structure of the strong nuclear force on a four-dimensional grid, where moving off one edge brings you back to the opposite edge (like the map in Pac-Man, but in four dimensions).

Action:

$$\begin{aligned} S &= S_{\text{Wilson}} + S_{\text{Davis}} \\ &= \beta \sum_P \left(1 - \frac{1}{3}\text{Re Tr}(P)\right) + \theta \sum_C \text{Im Tr}(W[C]). \end{aligned} \tag{8.20}$$

The first term is the Wilson action. The second is a model deformation depending on oriented loop traces.

Proposed action-comparison observable:

$$J = \text{Im Tr}(EW). \tag{8.21}$$

This is a loop correlator. The archived TVR-003 calculation uses its own stored configuration-level sums J_i, D_i and the reweighted mean defined earlier.

Parameter	Value
β	6.0
θ	0.45
L	12 (production), 8 (testing)
Asymmetric loop	3×6 or 4×8
Thermalization	500 sweeps
Statistics	10,000 configurations

Recorded direct-comparison settings. The 10,000 count and listed contours describe the proposed comparison. The reproducible TVR-003 archive contains 1,600 configurations in eight batches and uses the 2×6 loop.

Test	Result	Status
TVR-001	$\langle \text{Tr } F^2 \rangle \neq 0$	Recorded diagnostic
TVR-002	$\langle J \rangle = 0.034 \pm 0.002$	Run identification pending
TVR-003	Reweighted $J = 27.43 \pm 1.79$	Reproduced
TVR-004	Mirror and stability checks	Follow-on analysis

Identification follows the dated reports. TVR-003 is the reweighting study. TVR-004 documents post-processing robustness. The scaling proxy is the separate four-row calculation above.

Category	Test	Result
Physics	Symmetry ($\theta \to -\theta$)	Approximate sign reversal
	Batch stability	6.8% CV
	Sampled couplings	Reported comparison
Statistics	Bootstrap	Historical estimate; null unverified
	Autocorrelation	$\tau = 0.60$ (earlier lag run)
	Noise injection	Smooth decay
Geometry	Residual peak	Earlier setting unresolved
	Persistent Homology	12 above stated threshold
	Heat Kernel	Fit depends on settings
	Gap Ratio	85.46 (200-radius study)

I keep two of these records separate from the TVR-003 archive. The value $\tau = 0.60$ belongs to the earlier five-sweep lag diagnostic; I have not established it for the 1,600 TVR-003 rows. The 85.46 ratio belongs to the separate study of 200 configurations on an 8^4 lattice. The entries above preserve the reported diagnostics. Interpret each against its own data and analysis choices. Reweighting, resampling and noise injection on the same ensemble are complementary checks on that ensemble.

The record matters because it lets another person separate what can be replayed now from what needs a new experiment. I can stand behind a specific calculation. On a saved ensemble, a change in the statistical weights produces a large change in the mean of a loop observable. On a separate sample, a radial diagnostic singles out one unusual configuration. I want the record to make it possible for somebody else to ask better questions of both results.

I kept the measurements that let me return to these questions: the eight batches, their combined file, and the separate list of lattice radii. The public reports describe the investigations; my local archive holds the rows used for the calculations here. The TVR-003 summary specifies restricted access to the raw measurements and analysis scripts.[12]

The work continued from the weighted calculation into checks of robustness, geometric structure and graph spectra. Each report asks a different question of the data. Keeping them together lets me follow how one question led to the next without mistaking several views of one sample for several independent experiments.[13]

Writing the record down was the part I could do with the evidence already in hand. I chose to write this down because a question can be worth asking before I know where the answer will take me. I had spent enough time trying to find a signal that I wanted to show what was measured and what was still missing.

I still care about the possibility of a different energy future. That is what makes the accounting matter. If a calculation cannot tell me where the work came from, that is the next question I have to learn how to ask.

I called this book *The Geometry of Fuel* because the shape of a physical system can tell you which changes are possible and what they cost. I wanted to know whether that idea would lead somewhere in a gauge field.

The loop calculation showed a response; the radial calculation singled out a point in one description of a field. Neither was a measurement of energy. Understanding those differences is where the work stands.

[12]The TVR measurements are in eight dated files in `experiments/`, from December 8, 2025. Each records J_i, D_i and the acceptance rate for 200 configurations. The merged TVR-003 file, `data/harvests/harvest_merged.npz`, matches those fields exactly. The separate 200 radii and topological charge values are in `experiments/results/single_beta_results_full.json`. The public record includes the TVR-003 summary and follow-on reports.

[13]TVR-003 summarizes the reweighting study; TVR-004 covers post-processing robustness; TVR-005 the exploratory geometric analysis; and TVR-006 graph spectral geometry and spacing fits.

Chapter 9

What Fuel Means

"La geometría no pregunta quién eres. Solo pregunta si estás prestando atención."

The geometry does not ask who you are. It only asks if you are paying attention.
—Personal notes, January 2026

I can still see the numbers on my screen. The gap in the scatterplot. The mean of a lattice observable moving when I changed the bias. Those were two things I could calculate and go back to. The question that kept me up was what it would take to go from a change on a screen to a watt at a load.

The work in these pages stops at a simulated response. To go further I would need a physical system, a load, and a full cycle with a meter on everything that supplies energy. Then the account has to close. How much went in to drive the system? How much came out? What did it take to get back to where I started? If it costs me more to reset the experiment than I delivered to the load, I have my answer.

§ 9.1

Putting the meter around the whole cycle

I want a case simple enough to account for every joule. Put a pump and a return generator between two tanks. Draw a boundary around the apparatus and its water, with the supply and load outside. One meter reads what the pump draws. The other reads what the generator delivers. I start with the water in the lower tank and end when it is back there and the apparatus is ready to repeat the cycle.

Take 100 kg of water and a 2.00 m rise of its center of mass. With $g = 9.81\,\mathrm{m\,s^{-2}}$, the change in stored gravitational energy after the lift is

$$mgh = 100 \cdot 9.81 \cdot 2.00\,\mathrm{J} = 1.962\,\mathrm{kJ}. \tag{9.1}$$

The water starts and ends each phase at rest. I let the apparatus settle at each phase boundary so the only change in stored energy in this example comes from the water's height.

I am using *energy* for the accumulated transfer, measured in joules, and *power* for how fast it crosses the boundary, measured in watts. One watt is one joule per second. A device delivering

one joule slowly and one delivering it quickly can have the same energy output and different power. A useful fuel claim needs both, along with the cost of getting the apparatus ready to repeat.

The boundary tells me which transfers to record. A sign convention lets me add them without counting one transfer twice. For this closed mass boundary, let W_{in} be electrical energy supplied and W_{out} be electrical energy delivered, both positive in the direction named. Let L be the remaining energy that leaves the boundary, here through heat transfer to the surroundings. Energy conservation gives

$$\Delta E = W_{\text{in}} - W_{\text{out}} - L. \tag{9.2}$$

That equation also defines the loss inferred as a residual when the other three terms have been measured. A hot motor still inside my boundary contributes to ΔE until it cools. If I start the next trial with the motor hotter, I need to count that stored change or wait for the same temperature before calling the cycle complete.

Read the meters and check the reset

Table 9.1 records the constructed example in kilojoules. Figure 31 shows the paths through the apparatus. The rightmost column is the residual from Equation (9.2).

Phase	Supplied	Delivered	Stored change	Inferred loss
Lift	2.600	0	+1.962	0.638
Return	0	1.400	−1.962	0.562
Full cycle	2.600	1.400	0	1.200

During the lift the inferred loss is $2.600 - 1.962 = 0.638\,\mathrm{kJ}$. During the return the water releases $1.962\,\mathrm{kJ}$, and the load receives $1.400\,\mathrm{kJ}$. The remaining $0.562\,\mathrm{kJ}$ goes to losses. Adding the two phases cancels the stored change. The load got $1.400\,\mathrm{kJ}$ and the cycle took $2.600\,\mathrm{kJ}$ from the supply. The net export is $-1.200\,\mathrm{kJ}$.

To read the electrical energy, integrate terminal power $v(t)i(t)$ over each meter's recording interval, with the signs defined at the boundary. If the whole cycle takes 50 seconds, including reset, the two meter readings give average supplied and delivered powers of 52 and 28 watts. The net export averaged over that cycle is -24 watts.

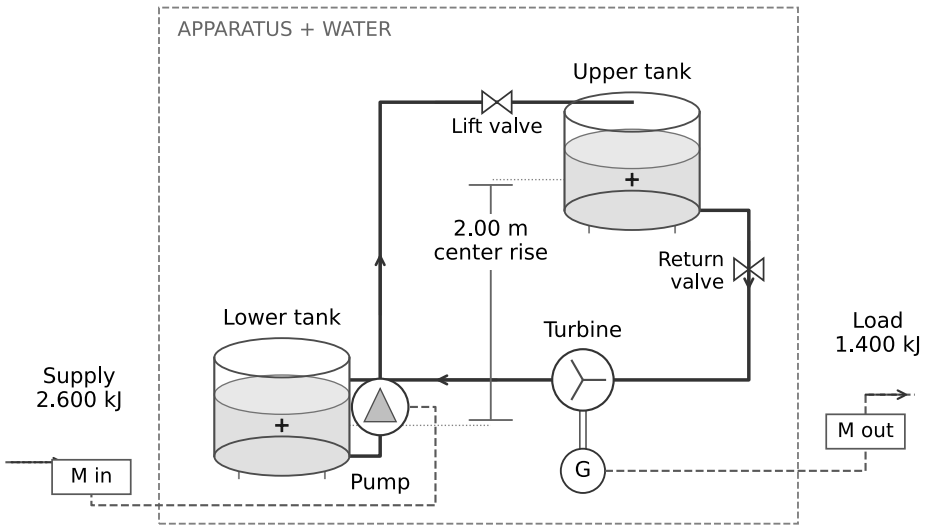
Suppose the input and output readings have absolute tolerances of 0.030 and 0.020 kJ. The smallest possible net export uses the lowest output and highest input, $1.380 - 2.630 = -1.250\,\mathrm{kJ}$. The largest uses the highest output and lowest input, $1.420 - 2.570 = -1.150\,\mathrm{kJ}$. Every combination allowed by these bounds still draws net energy.

If the reset check bounds $|\Delta E|$ by 0.010 kJ, the residual loss lies between 1.140 and 1.260 kJ. The three tolerances enter with their full worst-case sum, 0.060 kJ. These are hypothetical meter and reset tolerances. To use the ledger on a bench, I would record the temperatures and water levels at the two ends, measure any auxiliary input, and give each measurement its own tolerance.

Follow the parcel and both meters

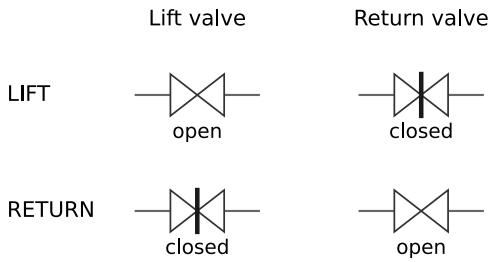
CONSTRUCTED WATER CYCLE / SCHEMATIC APPARATUS

01 The complete boundary and two flow paths



Solid: water. Dashed: electrical transfer. Double: shaft. Shading shows the same 100 kg parcel at two different stages.

02 Open one route at a time



At the settled start and end, both valves are closed.

03 The meter measures energy

$$W = \int v(t)i(t)\,dt$$

Voltage times current gives terminal power. Integrate over the entire phase, including settling.

One watt = one joule per second.

Only the center-of-mass rise is dimensioned.

The supply and load sit outside the boundary. Water, tanks, pump and generator sit inside it.

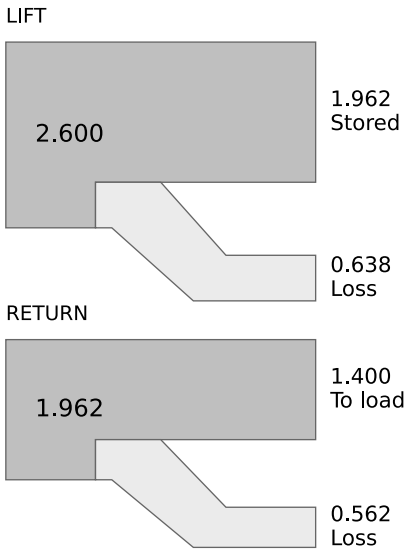
Figure 31. A constructed water cycle. The supply and load meters cross the apparatus boundary. The lower details connect the valve settings and the meter integral to the two flow paths.

THE GEOMETRY OF FUEL CYC / 03

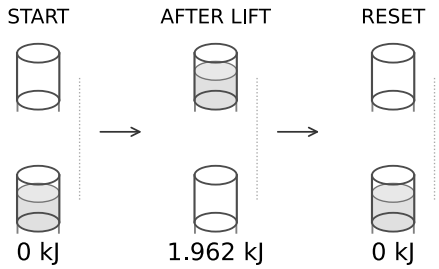
Close the ledger over the whole cycle

CONSTRUCTED READINGS / COMMON ENERGY SCALE / kJ

01 The two conserved splits



02 Return to the same stored state



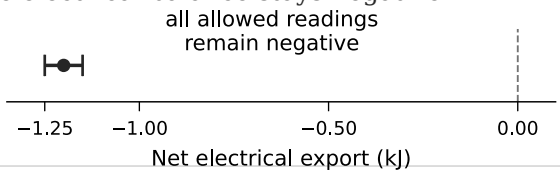
$$+1.962 + (-1.962) = 0$$

A hot motor is still stored energy. Let it cool or include that change before calling the cycle complete.

03 Every term keeps its sign

Phase	In	Out	Δ stored	Loss
Lift	2.600	0	+1.962	.638
Return	0	1.400	−1.962	.562
Cycle	2.600	1.400	0	1.200

04 The electrical balance stays negative



Net: −1.200 kJ

50 s cycle: −24 W

2.600 kJ enters, 1.400 kJ reaches the load, and 1.200 kJ leaves as loss after the full reset.

Figure 31 continued. Band widths track energy in kilojoules. The two storage changes cancel over the cycle. Meter and reset tolerances bound the net export and inferred loss.

That is what I would want from a physical energy claim. The Chapter 8 replay gives me a change in a loop statistic on the same rows. A watt at the load requires a measurement of energy across a boundary and a clock that runs through the reset.

The pump also separates a source from a store. Water raised to a higher tank can return some of the energy used to lift it. The generator can be useful without making the raised water a new primary energy source. Likewise, a battery, a fuel and a power supply require different accounts of preparation, storage and delivery. When I ask what Fuel could mean, I want the source of every delivered joule to remain visible.

The ledger sets the condition any proposed source has to meet. With that condition in view, I can explain why the possibility mattered so much to me.

The pump example counted energy going in and coming out. Now I want to ask what people pay for that energy, and what those costs keep them from doing. Imagine reliable low-carbon power becoming much less expensive. I can follow that possibility through water, food, industry and transport. In each case, I need to ask how much of the cost is actually energy, what else would still have to be paid for, and who could use the saving.

§ 9.2

Following the energy bill

Energy is a line in almost every budget I can think of. How much does that line have to fall before you build a plant you could not afford before? And what do you have to build alongside it for anyone to feel the difference?

A reverse-osmosis plant pushes seawater against a membrane that holds back most of the dissolved salt. The pumps consume electricity. The plant also needs an intake, clean membranes, a route for the concentrated brine, and pipes to deliver what comes out.

A town can sit next to an ocean and still ration water. It needs a system that can treat the supply and deliver it at a price people can pay. Inland towns have a further problem. Each litre must travel farther, often uphill.

This creates a trap. Water scarcity reduces economic productivity. Reduced productivity means less wealth. Less wealth means less ability to invest in water infrastructure. Less infrastructure means continued water scarcity. The cycle perpetuates itself across generations. Water scarcity also creates conflict. As aquifers deplete and rivers shrink, the disputes over remaining water grow fiercer. The Jordan River serves as a case study—four nations depend on it, none have enough, and water rights have been a recurring source of tension in every peace negotiation. The Nile is another: Ethiopia's Grand Renaissance Dam has created diplomatic conflict with Egypt and Sudan. Conflicts over water are not hypothetical future scenarios. They are happening now.

Put numbers on it. Suppose a plant uses 3 kilowatt-hours to make a cubic metre of water. At fifteen cents per kilowatt-hour, the electricity costs forty-five cents per cubic metre. At five cents, it costs fifteen cents. The change saves thirty cents per cubic metre, before the water enters a pipe. That saving comes from the electricity price in this example. A town would also need to know how much treated water escapes through damaged pipes before it reaches anyone. Fixing

those leaks might save money too, but I have not calculated that saving here. Cheaper treatment and a working delivery system are two parts of getting water to a household.

A coastal plant could give a city a more dependable supply in a dry year. Building several plants does little for a community whose distribution system leaks, or for a household priced out of the supply. The map of who gets water includes the grid, the pipes, and the people who set the rates.

The water bill does not stop at the kitchen tap. It continues through irrigation, fertilizer and the cost of getting food to a table. Modern agriculture is energy-intensive. Fertilizer production—the Haber-Bosch process that feeds half the world's population—requires enormous energy. Irrigation requires pumping. Transportation requires fuel. Refrigeration requires electricity. When energy is expensive, food is expensive. When food is expensive, poor people starve. This is not a metaphor. The 2008 food price crisis, driven partly by energy costs, caused riots in over thirty countries. People die when food prices spike—not in some abstract future, but in the recent past, repeatedly, and in the ongoing present.

The Haber-Bosch process deserves special attention. It synthesizes ammonia from atmospheric nitrogen, which is then used to make fertilizer. Before Haber-Bosch, the amount of food we could grow was limited by the nitrogen naturally available in soils and manure. After Haber-Bosch, we could make fertilizer industrially. This is why we can feed eight billion people. Without synthetic fertilizer, Earth could support perhaps three or four billion—maybe less. The Haber-Bosch process is, arguably, the most important invention in human history. It is also enormously energy-intensive, consuming approximately 1–2% of all global energy.

An indoor farm buys light, ventilation and cooling in the same way it buys seeds. Cutting those bills can improve the economics for some crops. Fertilizer plants and irrigation districts would have different gains, depending on what they buy now and what they can switch to.

This does not happen overnight. Infrastructure takes decades to build. Vertical farms need to be designed, constructed, staffed, maintained. Seeds need to be developed for indoor cultivation. Supply chains need to adapt. In the future I was imagining, the relationship between geography and hunger could weaken. A city in the Sahara might gain ways to grow some food locally. A nation with little farmland might gain more options for supplying its people. A country facing drought might be able to keep more irrigation running. Those possibilities still need water, suitable crops, equipment and a way to get the food to people; a lower energy bill alone does not establish food security. The Haber-Bosch process—currently responsible for approximately 1–2% of global energy consumption—could become less costly to run. That would not remove its feedstock, capital or distribution costs. Fertilizer could become more affordable. Farmers who could not afford enough fertilizer might be able to buy more, provided those savings reached them. Whether yields improved would still depend on the soil, the crop, the weather and how the fertilizer was used. This is not utopia. Agriculture still requires labor, knowledge, infrastructure. Crops still fail; pests still attack; droughts still happen. The energy part of the cost could fall. I want to know how far that saving would travel before it reached the price of a meal.

The same question extends beyond supplying what we already use: what work could become affordable that we now leave undone? Cutting emissions slows the addition of carbon dioxide to the air. Removing carbon already there requires additional work and a place to keep it. I think of those as two columns in the same ledger.

Direct air capture separates CO_2 from ambient air and then requires a storage or utilization pathway. Its energy demand has both electrical and thermal parts, which differ by process. For a transparent scale calculation, suppose the combined demand were 2,000 kilowatt-hours per tonne. At 40 billion tonnes per year, multiplication gives 80 trillion kilowatt-hours per year. That is a constructed demand scenario, not a universal plant specification or an electricity-only comparison. The process, heat source, upstream emissions and durable storage determine the actual climate benefit.

Cheaper low-carbon energy could lower the running cost of direct air capture. The source of that energy matters. A removal plant powered by high-emission electricity can add carbon elsewhere while it subtracts it at the intake.

To say how much atmospheric carbon would decline, I would need to account for the capture rate, the upstream emissions and how much of the captured carbon stays stored. Electricity is one term in that calculation.

That is the scale of the question that pulled me into the thought experiment. If cheaper energy lets me run more capture equipment, I still need to know how much carbon that equipment actually keeps out of the atmosphere after its own emissions are counted. That is what durable net removal means here: the amount removed and kept stored, after subtracting the emissions caused by doing the work.

That work would need materials and equipment, which brings the argument back to the factory. A plant manager can't put cheap energy into a budget as a single line and be done with it. There is the heat that has to reach the kiln, the motors and pumps that run all day, and the cost of stopping and restarting a process when supply is interrupted. Two factories can buy the same number of kilowatt-hours and need very different things from the grid.

A cement kiln needs a particular temperature and a steady supply of heat. A steel plant needs a process that separates iron from its ore as well as one that heats and works the metal. To put cheaper energy into either budget, I have to know the form in which the process uses it.

Reliable, low-cost power could move some of the boundaries on where industry can build. A steel mill still needs ore or scrap, transport, water and people who can run it. The electricity bill belongs beside those costs. In a place where energy has kept a project off the table, it can be the term that changes the decision.

Someone still has to operate the machinery, manage the deliveries, and ensure quality. Suppose electricity is inexpensive at a proposed factory site, but bringing in the raw materials and shipping out the product costs more than the electricity saving. The cheaper power has not made that site affordable. I have to follow the materials as well as the energy before deciding where to build.

Resource-rich regions might gain options for processing material close to where it is mined. The commercial question is who can finance the plant and sell its output. The local question is who gets the jobs, who bears the costs and who owns the result.

And whatever the factory makes still has to move. A cheaper supply of energy at one location is not yet a fuel a vehicle can carry. The global economy runs on moving physical objects. Picture the fuel bought for a container ship, the power bought for the cranes at the port, and the diesel bought for the last truck that delivers the load. A change in energy price lands differently on each part of that journey.

Cheaper clean electricity could help ports electrify equipment and lower the production cost of some low-emissions marine fuels. Ships still have to carry enough usable energy for the trip, with room left over for the cargo.

On an aircraft, every kilogram has to come with you. A cheap kilowatt-hour on the ground and a light way to store that kilowatt-hour in the air solve different problems. Synthetic fuels could use inexpensive low-carbon electricity during production. Their value would still depend on conversion losses, what the aircraft needs, and the emissions over the full fuel cycle.

Road, rail and shipping could each see lower costs by a different route. I would ask a practical question before promising a cheap crossing. Where does the energy come aboard, how much has to come aboard, and what space does it take away from the load?

§ 9.3

Who could use the change?

If some of those costs fell far enough, the change would reach people who had built their businesses and their lives around the old prices. New technologies change who has leverage. That is where a story about electricity turns into a story about power.

Imagine, for the moment, that a town can buy electricity much more cheaply than it could last winter. Leave the source of that electricity open. A family heating with electricity might see the difference on its next bill. Their neighbor with an oil furnace would first have to pay for different equipment. Across town, the driver who delivers heating oil might lose a route before she sees any saving at home. The same price change arrives at three doors on three different schedules. That is the kind of consequence I wanted to think through when I wondered what abundant energy might mean.

A large decline in demand for fossil fuels would pressure countries whose public revenues depend on selling them. It could change the price of oil without making every use of it disappear. The pace of the change would depend on how quickly customers could replace the equipment they already own.

For a government that pays wages and funds services with export revenue, even a partial decline can be destabilizing. The same price change that looks like relief on someone else's utility bill can show up at home as a missed payroll.

A country that has invested export revenue in education, infrastructure and other industries has more choices when its largest customer changes course. The transition is easier to imagine as a budget than as a headline. Which expenses can the government still cover when one source of revenue falls?

Fuel reserves are one source of control. Ownership of the equipment and the grid can be another. Energy is currently controlled by those who own the sources—oil fields, gas pipelines, coal mines, hydroelectric dams, nuclear plants. Energy companies are among the largest and most powerful corporations on Earth. ExxonMobil, Saudi Aramco, Gazprom, Shell—these are not just companies. They are geopolitical actors. Saudi Aramco illustrates the scale of an enterprise built around petroleum extraction, processing and sales. Governments regulate energy because controlling energy means controlling society. Without electricity, modern civilization stops. Without fuel,

transportation stops. Without natural gas, heating stops. Whoever controls energy has leverage over everyone who needs it—which is everyone.

Wars are fought over energy. The 1991 Gulf War was about oil. The geopolitics of the Middle East for a century has been about oil. The contest between Russia and the West is, in significant part, about gas. Alliances are formed around energy. The twentieth century was shaped by oil; the twenty-first was supposed to be shaped by the transition away from it.

If a source could be installed wherever people needed it, control could shift from ownership of fuel reserves toward ownership of equipment, capital, and the grid. A household would still have to ask who sets its price and who can turn it off.

Local generation could give communities more options. Whether they could use those options would depend on financing, maintenance, access to parts, and the rules governing connection to the grid.

A change in who controls supply also lands in the lives of people whose jobs depend on it. Fifty years is long enough to see a new industry arrive and an old one disappear. It is also long enough for a child in a coal town to grow up and have to decide whether there is a job for her there. The transition lands in particular places.

People who lose power do not surrender it peacefully. ExxonMobil knew about climate change in the 1970s and funded denial for decades. Coal companies fought regulations for generations. Do you think they will accept obsolescence gracefully? Corporations whose business models become obsolete do not go quietly. They fund politicians, they wage public relations campaigns, they sue and lobby and delay. If TVR ever became a working technology that competed with their business, I would need to think about that resistance too. The calculations in this book have not yet taken it to that point. Nations whose economies collapse do not remain stable. The Soviet Union did not collapse gracefully when its economy failed. Venezuela did not remain stable when oil prices crashed. Failing states produce refugees, regional conflicts, humanitarian crises.

A rapid change in demand could bring tariffs, subsidies and fights over who pays for stranded investments. The timing and terms of those disputes would depend on how the change happened and who could adapt.

Resource transitions can also sharpen geopolitical disputes. The distribution of costs, and the institutions that can absorb them, matter more to that question than the price of energy alone.

Even a successful transition leaves a question about who receives the benefit first. New technologies do not distribute evenly. The rich get them first. The powerful get them first. When electricity spread, it reached Manhattan before it reached Mississippi. When the internet spread, it reached universities before it reached villages. When vaccines are developed, wealthy nations secure supplies before poor nations can access them. This is not conspiracy. It is economics. New technologies are expensive; rich entities can afford them. New technologies require infrastructure; developed places have it. New technologies need expertise; educated populations have it. If TVR ever becomes practical, I would worry about those same barriers. Who could afford the equipment? Who could get it repaired? Would a rural community have access, or only a large company with its own engineers? Those questions would need answers before I could promise that a technical success would benefit everyone.

That pattern could repeat with cheap clean power. The first places to afford the equipment and infrastructure could benefit while others keep paying the old rates. Access is part of the design problem.

I put the mathematics and experiment summaries on the public record. That gives another person a starting point. They can examine the calculation, ask for the underlying measurements, or try a different way of testing the same question.

Public records make the research easier to inspect. Equitable access to any eventual technology would require its own decisions about manufacturing, investment and distribution.

§ 9.4

The responsibility of putting it on the record

There is a cliché in physics: the universe does not care who discovers its laws. Gravity would work the same whether Newton or Leibniz derived it. Relativity would be true whether Einstein had been born or not—someone else would have found it eventually. This is true about the physics. It is not true about the documentation, the framing, the choices about what to publish and when and how.

I put the work online from outside an academic laboratory. By now you know why leaving a record mattered to me. The question here is what I owe the person who picks it up.

A date on a paper tells you when I wrote a claim down. A method with enough detail to repeat tells you something I care about more. It gives somebody else the chance to see where my reasoning holds.

I could imagine consequences far beyond the calculation on my screen. Now I had to make it possible for someone else to see exactly how far that calculation reached.

I am responsible for what I ask those numbers to carry. I can show the result, put the assumptions where people can find them, and leave room for somebody to run a better test. That is the part I can control.

A university researcher may have a technology transfer office, patent lawyers and colleagues to consult about the next step. I did not have those channels. I put the work on the internet and hoped someone would read it. That made the way I presented it especially important: someone could meet the claim without ever meeting me.

The papers and experiment reports have dates and public records. A reader can go back to them. When a claim depends on a convention or a selected set of runs, that belongs next to the number.

I want the person who follows the calculation to be able to disagree with me at a specific step. That gives us somewhere to work.

If the research leads to an experiment and the experiment leads to a technology, people will still have to decide what to do with it. That is another reason to keep the original question and its evidence visible.

Now I am writing this book to make sure even more people understand. Not the physics—the physicists can read the papers. But the context. The implications. The choices that lie ahead. If you are reading this, you are part of the audience I am trying to reach. You may not understand

every equation. You can still ask what was measured, follow an example, and see where a physical experiment is still needed. I want those steps where you can find them. If this research ever leads to a working technology, choices will be made about how it develops. By governments, by corporations, by international bodies. These choices will affect your life and the lives of your children and their children. You need to understand that you have a voice in those choices. Not because you are a physicist or an engineer or a policymaker. Because you are a citizen, a voter, a consumer, a human being. Decisions about any such technology will be made by institutions that respond to public pressure, public opinion, public demand. If you understand what is at stake, you can participate in those decisions. If you do not understand, the decisions will be made without you, by people who may not share your values or your interests. This is why I wrote a book instead of just papers. Papers speak to specialists. Books speak to everyone. I am speaking to everyone.

This is where the argument returns to the beginning. I wanted my observation to be taken seriously enough to be examined. I owe the next person the same opportunity to examine the work. $C = \tau/K$ is a useful way to ask how much room a specified budget buys. When I can derive that relationship in a stated geometry, someone else can follow it line by line. When I take it into another domain, I have to do the work again.

I wanted the freedom to put an idea in front of people without a gatekeeper deciding whether I belonged there. The freedom comes with a responsibility I recognize from engineering. If I hand over a measurement, someone should be able to check what I measured. If I hand over a derivation, they should be able to see what I assumed.

I know how easy it is to believe a question does not count until somebody with a title agrees to hear it. That is why I want these examples to be usable. You can ask what was measured. You can check a sum. You can look for the assumption that makes a conclusion possible. You do not have to wait for permission to begin learning how.

The universe does not have bouncers. The work does have demands: attention, patience, training where we need it, and the willingness to check. I want this book to make those demands visible enough that someone else can take the next step.

I think about fifty years because that is a span over which power plants, ports and whole industries can change. It is also a span in which a question from a notebook can go nowhere at all. I have to allow for both when I talk about the future.

An idea can arrive in a night. Infrastructure takes financing, materials, crews, and years of use to become ordinary. Even a better way to supply energy would have to go through that process.

In fifty years, I will likely be dead. But I keep thinking about what this work could mean for somebody who will be here. Energy is part of what it takes to clean water and get it to people. If we could make that energy less expensive, what would that change for someone who has to choose between clean water and something else they need? That is the future I am thinking about when I go back to the calculations.

I want somebody to live in a world where a clean glass of water costs so little they never have to calculate what they will give up for it. I can want that world and still be exact about what I saw on my screen.

When I went back to the lattice records, I could change the average without moving a single field. I gave more weight to the rows with a large value of one loop measurement. On those same rows a second loop measurement tended to be large too. Its average rose. The size of that rise was something I had to learn from the data, and somebody else with the rows could check my answer. Now imagine setting a glass under a tap. Which one of those numbers tells me how much water the pump lifted? I had recorded neither the work at the pump nor the electricity drawn to run it. This is where the word *fuel* changed for me. Geometry had given me a way to find the question my instrument could answer, and a way to recognize the next one.

That glass of water is why the distance between a simulation and a working device matters to me. Every step across it needs to hold. The papers keep the technical record; this book lets me follow the route with you, including the places where I still do not know how to cross.

What comes next is another way to examine that route. I want to take Einstein's arguments apart through GIGI and ask what each measurement has to carry. Then I can bring that question back to the fuel calculation, and to the chip I have wanted to build.

The machine gave me a reading.
I asked what it could see.
I asked the same of the lattice.
That is the work.

I put it in a book.
Now it belongs to everyone.

La geometría no pregunta quién eres.
Solo pregunta si estás prestando atención.

The geometry does not ask who you are.
It only asks if you are paying attention.

Chapter 10

What a measurement must carry

I had this realization while mopping the floor at the restaurant.

The question was almost embarrassingly simple: what do I mean by one? One mop is easy enough. I have identified an object and counted it. But tell me to move it one, and suddenly we have a problem. One what? One inch? One metre? Toward me? Away from me? The number has stayed the same, but the instruction is missing most of what I need to carry it out.

That is the distinction I want to start with. Counting an object does not require choosing a direction. Moving it does. Somewhere between those two uses of one, we have supplied a unit, a starting point and a direction. If we draw an axis and put the object at one, those choices can disappear into the drawing. They have not disappeared from the operation.

Measurement science gives the first part of that question a name: the *measurand*, meaning the quantity we intend to measure. The length of a mop stroke and the area of floor it cleans are different quantities, even though they belong to the same action. Before asking whether a number is right, we have to settle what it is a number of.[1]

Imagine marking a point on the mop head and following that point across a flat patch of floor. We are tracking one point so that we do not quietly confuse its path with the whole area swept by the mop. Call its starting position p and its finishing position q. Suppose one means a straight movement of one metre. Call that distance ℓ. We still need an arrow telling us which way to move. A mathematical arrow with a direction and a length is a *vector*. Here we use a reference arrow u whose length is one, with no physical unit attached; multiplying it by ℓ gives the actual displacement. The double bars $\|u\|$ mean the length of that arrow. Now we can write the instruction:

$$q = p + \ell u, \qquad \ell = 1 \text{ metre}, \qquad \|u\| = 1. \tag{10.1}$$

Read it from the starting point outward: begin at p, move the distance ℓ in direction u, and arrive at q. This is the classical distinction between a position and a displacement: a place on the floor and a move from that place. In affine geometry, points and displacements are kept distinct in exactly this way. The floor does not come with a privileged zero painted on it; we choose an origin when we want coordinates.[2]

If the marked point stays on the floor, the possible endpoints form a circle around p. Allow it to move freely in three dimensions and they form a sphere. We are assuming an unobstructed region for this example. A wall or a rail would supply another constraint.

[1] Joint Committee for Guides in Metrology, *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms*, JCGM 200:2008, definitions 2.1 and 2.3. The specification of the measurand identifies the intended quantity and the relevant state of the object or phenomenon. Official ISO online text.

[2] John R. Miller, *Understanding Geometry: The Abstract Spaces in Which Points and Vectors Live*, University of Kansas course notes, sections "Affine Spaces" and "Euclidean Spaces." A point plus a displacement gives a point; subtracting two points gives a displacement. A rule for length is additional structure. University of Kansas notes.

One metre has not chosen a point on that circle. It has chosen how far from the centre the point must be. That is a complete distance and an incomplete instruction.

There is a short mathematical reason we cannot recover the missing direction from the distance alone. Rotate the picture around p. The starting point stays fixed. Every endpoint remains one metre away. But a particular endpoint can move to another point on the circle. Since the given position and distance survive that rotation, they cannot distinguish the endpoint that changed. We need an additional choice: a direction, a target, or something in the surroundings that selects one. The argument assumes ordinary flat geometry with no preferred direction around the starting point. A wall or a marked axis would add information and could select a direction.

This also changes how I think about the minus sign. Choose one direction across the floor as positive. Moving the marked point one metre in the opposite direction gives a coordinate displacement of minus one metre. It still travels one metre. Reverse the axis, and the same movement gets a positive coordinate displacement. The sign records its relation to an oriented axis; the path length is nonnegative. Choosing a particular direction makes the movement definite, but it does not make the set of possible directions discrete—a collection of separated choices. In this model, we can turn the arrow by any small angle we like. The possible directions fill the circle continuously.

Now imagine pushing that marked point one metre along a straight line and drawing it back along the same line. Let Δr mean the net change in position and L the total distance traveled. The symbol Δ means change; it is asking us to compare the finish with the start. The two displacement arrows cancel, while their lengths add:

$$\Delta r = \ell u + (-\ell u) = 0, \qquad L = \|\ell u\| + \| - \ell u\| = 2\ell = 2 \text{ metres}. \tag{10.2}$$

The first expression asks where the point finished relative to where it started. The second adds the lengths of the two straight strokes. The answers are zero and two metres. Both are correct.

But imagine keeping only the zero. You could no longer tell this trip from a point that had stayed still. The zero did not make a false statement. It answered an endpoint question and left the journey out. On a real floor, that journey can involve friction and dissipated energy. Returning to the same position does not undo those processes. Here *work* means energy transferred by a force as the point where it acts moves. To calculate it, we need the force and the part of the movement along that force at each stage. Neither displacement nor path length alone supplies that information.

Feynman's treatment of work makes the same distinction useful in physics. Some forces, such as gravity in a fixed Newtonian potential, have work determined by the endpoints. Friction during sliding makes us keep track of more: motion of the whole object can transfer energy into internal motion of the materials.[3] Bring the mop back and you have restored its position. You have not thereby restored the floor and mop to their earlier physical states. That is why an energy account cannot stop at the zero displacement.

This is where the question meets my paper *Zero Does Not Exist* and the counting system developed there. The distinction is between the void, a present seed, and a structure with

[3]Richard P. Feynman, Robert B. Leighton and Matthew Sands, *The Feynman Lectures on Physics*, vol. I, chapter 14, sections 14–1, 14–3 and 14–4, "Work," "Conservative forces" and "Nonconservative forces." Caltech online edition.

connections. To draw that distinction, start with a dot. We call the dot a *vertex* and a connection between two dots an *edge*. Together the dots and connections form a *graph*. The seed is one dot before we draw its first connection.

Now put the dots in a chain. Call the number of successive crossings n, allowing n to be zero. A chain with n crossings needs $n+1$ positions: one to start from and one more after each crossing. We write that chain as P_{n+1} and call it a path graph. The rule φ assigns the count n to this path. The letters V and E stand for its collections of vertices and edges; single bars around either collection mean how many members it contains. With those names in place, the construction says

$$\varphi(n) = P_{n+1}, \qquad |V(P_{n+1})| = n+1, \qquad |E(P_{n+1})| = n. \tag{10.3}$$

For two crossings, draw three dots joined in a chain. For no crossings, keep the starting dot: $\varphi(0) = P_1$. The zero in that edge count does not erase the starting position. The absent system and the present system before its first connection are different cases.[4]

Our out-and-back example makes another distinction visible. It has three successive visits: start, turn, return. Those visits can be the three vertices of a path even though the first and last refer to the same floor position. A record of visits and a set of distinct locations are different representations. Identifying the return with the start without retaining the two traversals would lose the very journey we are trying to describe.

So I would not use a zero measurement as a synonym for an absent system. A defined count can correctly report no objects of the requested kind. A marked origin can have coordinate zero. A displacement can be zero after a trip. Before treating these as the same thing, we have to say what was counted or compared. My construction specifies which objects its counting system admits; this physical example specifies which question a measurement answers.

The partition work asks a related question: what are we allowing to count as one group? A *partition* divides a collection into groups so that every member belongs to exactly one group and no group is empty. Imagine three floor regions, A, B and C, arranged in a row. Here two regions are adjacent when they share a boundary across which we can move directly. Record connections only between A and B and between B and C. All three make one connected region. A and C alone do not; the connection would have to pass through B, which the proposed group has left out. In the geometric partition construction, every group must be connected using its own members. The relationships therefore affect which groupings are permitted, even when the number of pieces stays the same.[5]

This connected-group requirement also has an established mathematical literature. Frank Simon, Peter Tittmann and Martin Trinks study *connected set partitions*: ways of dividing a graph into groups whose members can reach one another without leaving their group. That is the same

[4]Bee Rosa Davis, *Zero Does Not Exist*, "The Geometric Natural Numbers," especially the geometric natural-number definition and Peano embedding. The displayed formula suppresses the paper's element-count label and retains its path factor. The construction excludes the empty graph from its chosen domain; it does not invalidate zero as an element count or as the number that leaves a sum unchanged. The Peano embedding represents the ordinary counting sequence $0, 1, 2, \ldots$ using these nonempty paths.

[5]Bee Rosa Davis, *The Subgraph Lattice Theorem and Geometric Partition Counting*, Part I, April 10, 2026, definition of connected set-partition. Each block must induce a connected subgraph. The three-region illustration is the path graph A–B–C; its block $\{A, C\}$ is disconnected.

connectivity condition used here.[6] On our floor, putting A and C on the same list does not create a passage between them. That small example is what the condition protects.

A vertex count leaves out adjacency. A distance leaves out direction. An endpoint difference leaves out the route. These are useful summaries, but each has made a decision about what information to retain. The graph construction itself does not supply physical metres or angles; those require an additional geometric description. That is the extension I am exploring here: carrying the relationships needed for the physical question along with the number.

Mopping gave me a way to see it without beginning with a page of symbols. I could ask where the movement started, which way it went, how far it traveled, and what returning to the beginning actually restored. Those are different questions. A representation that keeps their answers together gives us a chance to notice when we have answered one and accidentally claimed another.

§ 10.1

Keep the movement in the record

I want to give that idea something we can actually inspect. We can keep talking about a number losing its context, but put the records beside one another and the loss becomes much harder to miss. The next three studies use the same flat-floor model. They are constructed computational experiments, not measurements taken while I was mopping. I have chosen simple inputs so that we can follow every answer by hand and then check the representation in GIGI.

For these runs, Python prepares exact fractions, the products needed for squared lengths, and the connectivity checks. GIGI stores the records and their labels, returns them for inspection, and performs the grouped sums and counts shown in the panels. That division matters: we are testing what the implementation actually does. GIGI is not being credited with deriving the proof or measuring a force on the floor. The original records, the queries and the returned tables are retained together.[7]

Start the marked point at $p = (2, 3)$ metres in our chosen floor coordinates. For the first study, move it one metre in eight different directions. One displacement is $(1, 0)$ metres: all of the movement goes along the first coordinate. Another is $(3/5, 4/5)$ metres. That second instruction moves 0.6 metres across and 0.8 metres up the coordinate drawing. The two components describe one diagonal stroke, not a sideways stroke followed by an upward one.

Call the displacement $d = (d_x, d_y)$. The Pythagorean theorem gives its squared length, $d_x^2 + d_y^2$. For the diagonal stroke,

$$\begin{aligned} (3/5)^2 + (4/5)^2 &= 9/25 + 16/25 = 1, \\ q &= (2 + 3/5, 3 + 4/5) = (13/5, 19/5). \end{aligned} \tag{10.4}$$

[6] Frank Simon, Peter Tittmann and Martin Trinks, "Counting Connected Set Partitions of Graphs," *The Electronic Journal of Combinatorics* 18 (2011), P14; arXiv:1005.1726v2 (2010), section 3. Paper and publication record.

[7] Reproduction script: `research/einstein_gigi/opening_replay.py`. The adjacent `results_opening_gigi` folder holds the query, raw output, CSV tables and receipt with code and executable hashes. The run verified 33 sums, eight directions, nine event records and 27 ordered-pair witnesses in one fresh GIGI process; restart persistence is not tested.

The squared length is one square metre, so the length is one metre. The endpoint is $(2.6, 3.8)$ metres. GIGI returns both endpoint coordinates from their addition terms and returns the squared length from its two prepared square terms. Figure 32 prints those results for all eight directions; its decimals are exact for these inputs.

Look down the last column. Every answer is one. Now look one column to its left: eight different endpoints. If we keep only the length, we have made all eight instructions look the same. There is nothing wrong with that if distance is the question. There is quite a lot wrong with it if someone is waiting for the mop at a particular place.

Eight examples alone would not prove a statement about every possible direction. The rotation argument above supplies that proof: rotating any nonzero displacement preserves its length while changing its endpoint. This table does a different job. It lets us check that the stored records and calculations respect the distinction the proof requires.

Now keep the diagonal stroke and change what happens afterward. Record the point three times, in order, using three different histories. In the first, it stays at $(2, 3)$ for all three records. In the second, it reaches $(2.6, 3.8)$ and comes back. In the third, it reaches $(2.6, 3.8)$ and continues another metre along the same line to $(3.2, 4.6)$. The labels $0, 1, 2$ identify the order of the records; they are not measured times. Between successive records, the model assumes a straight stroke or no motion.

That last assumption is essential. Three isolated position readings would not reveal every turn made between them. A point could leave and return between observations. Here the straight segments are part of the declared input, so their lengths really do give the modeled path length. We are not asking GIGI to recover an unrecorded journey.

For positions r_0, r_1, r_2, the net displacement and modeled path length answer different questions:

$$\begin{aligned} \Delta r &= (r_1 - r_0) + (r_2 - r_1) = r_2 - r_0, \\ L &= \|r_1 - r_0\| + \|r_2 - r_1\|. \end{aligned} \tag{10.5}$$

In the first sum, the middle position cancels. That is why an endpoint comparison can forget it. In the second sum, we take each leg's nonnegative length before adding. Reversing the direction of a leg does not make its length negative.

For the out-and-back history, the terms are particularly easy to see:

$$\begin{gathered} \Delta x = 3/5 - 3/5 = 0, \qquad \Delta y = 4/5 - 4/5 = 0, \\ L = 1 + 1 = 2 \text{ metres}. \end{gathered} \tag{10.6}$$

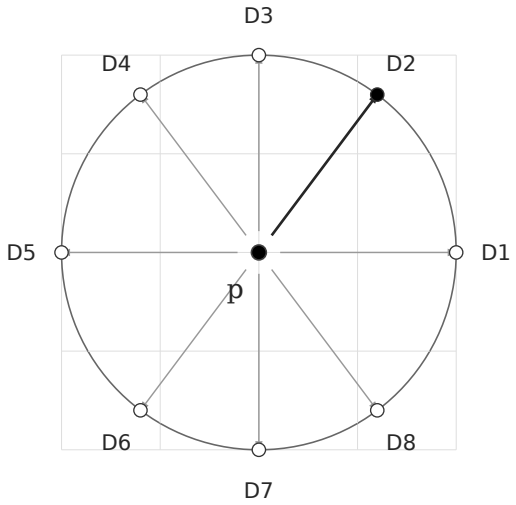
These are actual sums in the GIGI run. For example, the x terms are stored as signed integer ticks, $+3$ and -3, with five ticks per metre. GIGI sums them to zero; the retained scale tells us how to read that result. Integer scaling avoids rounding these rational inputs during the sum. It is not a claim that arbitrary floating-point calculations become exact.

Figure 33 puts all nine records on the page. GIGI also groups them by history and by the supplied location key. The keys identify equal coordinate pairs within each history. The result is three records in every case, but one, two and three distinct positions respectively. Three observations are not necessarily three places, and they are not necessarily three movements.

One distance leaves many endpoints

RESTAURANT-FLOOR STUDY 01 / CONSTRUCTED INPUTS, VERIFIED GIGI OUTPUT

Choose how far. Then choose which way.



Circle radius: 1 metre

STARTING POINT

$$p = (2, 3)\ \text{m}$$

D2: A CHOSEN DISPLACEMENT

$$d = (3/5, 4/5)\ \text{m}$$

ADD THE MOVE TO THE START

$$q = p + d = (13/5, 19/5)\ \text{m}$$

SQUARED LENGTH

$$(3/5)^2 + (4/5)^2 = 1$$

Every row below passes the same check.

ID	DISPLACEMENT (m)	ENDPOINT (m)	LENGTH² (m²)
D1	(1.00, 0.00)	(3.00, 3.00)	1.00
D2	(0.60, 0.80)	(2.60, 3.80)	1.00
D3	(0.00, 1.00)	(2.00, 4.00)	1.00
D4	(-0.60, 0.80)	(1.40, 3.80)	1.00
D5	(-1.00, 0.00)	(1.00, 3.00)	1.00
D6	(-0.60, -0.80)	(1.40, 2.20)	1.00
D7	(0.00, -1.00)	(2.00, 2.00)	1.00
D8	(0.60, -0.80)	(2.60, 2.20)	1.00

A length fixes a circle; a direction selects an endpoint.

Eight sampled directions illustrate the proof. GIGI returned the endpoint sums and squared lengths.

Figure 32. One distance, eight possible instructions. The shaded row identifies the emphasized diagonal. Endpoint coordinates and squared lengths are GIGI sums; direction components are supplied inputs.

This is the comparison I care about. Stay still and come back both end where they began. Come back and keep going both cover two metres. Neither summary can replace the other. The records let us ask both questions without pretending their answers mean the same thing. To ask how much work the mop did, we would need another layer of information: the forces along those strokes. No energy value is inferred from this position table.

There is one more way to lose a relationship without losing a number. Return to the three floor regions A, B and C. We still have three regions regardless of which ones we put in a group. But that count does not tell us whether someone can move between the members of the group without stepping outside it.

There are exactly five partitions to check. We can leave all three regions separate; join one of the three possible pairs and leave the third separate; or put all three together. That accounts for $1 + 3 + 1 = 5$ possibilities, with none omitted. In the table, a vertical bar separates groups: $AB \mid C$ means A and B together, C separately.

For each proposed group, the computation asks whether every ordered pair of members can be joined using only that group's regions. An ordered pair names a starting region and a destination. We include a region paired with itself, which requires no crossing. A group of size k therefore supplies k^2 checks. A two-region group and a singleton supply $2^2 + 1^2 = 5$ checks in total.

The connectivity witnesses are prepared by following only the permitted links inside the proposed group. GIGI stores each witness with its group, source, destination and result, then counts the missing connections for each partition. The shaded row in Figure 34 has two: A cannot reach C and C cannot reach A while staying inside $\{A, C\}$. B exists on the floor, but that proposed group has excluded it. We cannot quietly borrow its passage and still say the group is connected through its own members.

We can also prove the total without the program. In a three-region chain, a connected group must occupy an uninterrupted part of the chain. At each of the two links, choose whether to keep the adjoining regions in the same group or cut between their groups. Each pair of choices gives one connected partition, and every connected partition determines those choices. There are therefore

$$2 \times 2 = 4 \tag{10.7}$$

connected partitions. The fifth partition, $AC \mid B$, cannot be produced by those cuts. The proof and the exhaustive table agree.

This is a small version of the connected-partition condition cited earlier, not a new theorem about three regions. Its role here is to make the cost of forgetting a relationship visible. The floor model, the graph and the stored records must all agree about which crossings are allowed. A correct total of three regions would not reveal a mistake in that agreement.

The three studies leave us with a practical rule. Keep the question attached to the number. A length needs the movement it describes; an endpoint needs the history if the history matters; a group needs the relationships that make it a usable group. The point of keeping these records in GIGI is that we can inspect those choices instead of letting them disappear behind one apparently complete answer.

Returning is not standing still

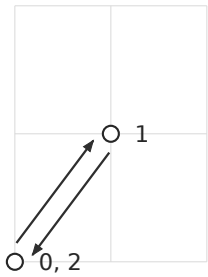
RESTAURANT-FLOOR STUDY 02 / NINE RECORDS, THREE DIFFERENT HISTORIES

Stay still



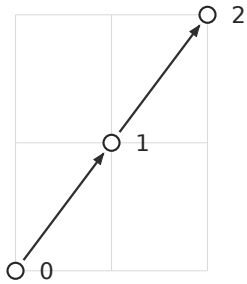
Labels mark record order: 0, 1, 2.

Out and back



Labels mark record order: 0, 1, 2.

Keep going



Labels mark record order: 0, 1, 2.

Floor plans share one scale; grid steps are 0.6 m across and 0.8 m up.

RECORD	STAY STILL	OUT AND BACK	KEEP GOING
0	(2.00, 3.00)	(2.00, 3.00)	(2.00, 3.00)
1	(2.00, 3.00)	(2.60, 3.80)	(2.60, 3.80)
2	(2.00, 3.00)	(2.00, 3.00)	(3.20, 4.60)

All positions above are (x, y) in metres.

Net move (m)	(0.00, 0.00)	(0.00, 0.00)	(1.20, 1.60)
Path length (m)	0.00	2.00	2.00
Event records	3	3	3
Distinct positions	1	2	3

The same endpoint can conceal a different journey.

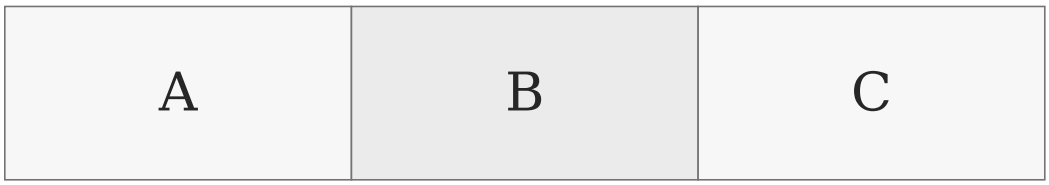
GIGI groups the event records and sums the prepared leg terms. No force or energy measurement is inferred.

Figure 33. One starting point, three histories. The table contains GIGI readbacks and summaries. Labels indicate record order. Straight segments between records are assumed, not inferred from sparse readings.

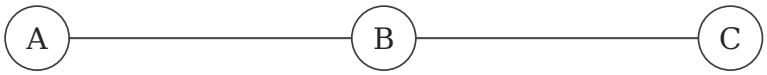
A group must keep its own connections

RESTAURANT-FLOOR STUDY 03 / ALL FIVE PARTITIONS OF THREE REGIONS

Three adjacent regions on the same floor



Only shared boundaries permit a direct crossing.



Can every pair in a group connect without leaving it?

GROUPING each region used once	PAIR CHECKS including each self-pair	MISSING ordered pairs	ALLOWED connected
A | B | C	3	0	yes
AB | C	5	0	yes
AC | B	5	2	no
A | BC	5	0	yes
ABC	9	0	yes

Putting A and C on one list does not build a passage.

Python supplies within-group reachability witnesses; GIGI counts the two missing directions in AC | B.

Figure 34. A list does not create a connection. GIGI counts the supplied within-group reachability witnesses. Four partitions pass; $AC \mid B$ fails in both directions.

That is the question I wanted to bring to Einstein through GIGI. Keep every digit of a clock reading, but lose which clock made it, and there is a limit to what that number can tell you. Keep both ends of a rod, but forget when each end was measured, and even a subtraction can become the wrong operation. AMC had already made the importance of representation visible to me. DHOOM and GIGI carried that question into the structure of the data itself. Now I wanted to take a mathematical argument apart carefully enough to see what its representations actually require. What has to be stored? What relationship has to remain attached? What can change without changing the answer?

Before we get to the equations, I want to stay with the person for a minute. In 1905, Einstein was twenty-six and working as a technical expert at the patent office in Bern. His job involved examining inventions. The paper we are about to read came from that period of his life, before his name became shorthand for genius.[8]

There is something human about that starting point. Someone has made something. Someone else has to understand what it actually does. A description can sound perfectly reasonable until we ask which part of the apparatus makes the claim true. That is a question I recognize. It is also a useful way into a paper that can feel much farther away from ordinary life than it really is.

Einstein begins with a magnet and a conductor. Picture a coil of wire connected to a meter, with a magnet moving toward the coil. The meter can register a current. Now consider the corresponding relative motion described from the magnet's frame: the magnet is at rest and the coil moves. The older account puts the emphasis on different mechanisms in the two descriptions. In one, a changing magnetic field produces an electric field. In the other, charges in the moving conductor experience a magnetic force. Einstein asks why the account makes such a sharp distinction when the relative motion is what the experiment brings to our attention.[9]

That does not mean electric and magnetic fields are mistakes. It means their separation depends on the observer's motion, and the equations must carry that dependence consistently. The apparatus should not acquire a different physical history because we changed the description. That is already close to the question we have been asking about data: what has to change together so that we are still describing the same event?

He did not work without conversation, either. At the end of the paper he thanks his friend Michele Besso for useful suggestions. I like that detail. The finished equations are what survive on the page, but the paper itself leaves room for another person in the work.[10]

To follow the next step, we need to give one phrase an ordinary meaning. An *inertial frame* is a system of spatial coordinates and clocks in which a free object continues at constant velocity. Think of an ideal train moving steadily along a straight track, with no braking, turning or vibration. Within a region where gravity can be neglected, the train and the platform provide a useful picture of two such frames. An actual train only approximates the ideal. A *frame* here is more than a camera angle: it includes a way to assign times at different places.

[8]The Swiss Federal Institute of Intellectual Property documents Einstein's employment from 1902 to 1909 and his birth on March 14, 1879: `https://www.ige.ch/en/about-us/the-history-of-the-ipi/einstein`. These facts locate the paper in his life; they do not establish that any particular patent caused the theory.

[9]Einstein (1905), opening paragraphs, pp. 891–892. The coil and meter are our concrete illustration of his magnet–conductor example. We compare inertial descriptions of the same process; we are not claiming that arbitrary acceleration histories or different experimental arrangements must give identical readings.

[10]Einstein (1905), p. 921, final acknowledgment to M. Besso.

Einstein asks us to accept two physical starting points. First, the laws of physics have the same form in inertial frames; a laboratory experiment cannot single out one of them as absolutely at rest. Second, light in vacuum travels at speed c, independently of the motion of its source. With the clock convention he is about to define, each inertial frame assigns that same speed to light. These are postulates about the physical world, not conclusions obtained by putting numbers into GIGI.[11]

Here is why that gets interesting. Throw a ball forward on the train and its speed relative to the platform differs from its speed relative to the carriage. It is tempting to do the same addition with a flash from a lamp. But the two postulates require a different result for light. If both observers are to obtain c, we have to look carefully at what each calls a distance and what each calls an elapsed time. We cannot quietly assume those measurements are shared and then blame the arithmetic when the assumptions conflict.

Einstein's 1905 paper gives us a particularly demanding place to start. Its first mathematical task is to explain what it means to put distant clock readings on a common time coordinate. Only then does it transform those coordinates between moving systems. The order matters. If we begin with a matrix and skip the clocks, we can make the arithmetic look clean while hiding the physical question that made the matrix necessary.[12]

The calculations that follow are constructed examples. They are not new measurements of relativistic effects. They let us check, exactly, whether the representation we give GIGI retains the relationships in the mathematics. The proof explains why the relationship holds; the computation checks that our implementation follows it.

Beside the source mathematics, short German passages appear on the left with English translations on the right. The page numbers refer to the original journal printing. I want the wording to stay close enough that we can see what Einstein is asking the mathematics to do. The excerpts retain his spelling, with print line breaks joined. Our normalized symbols, numerical examples and modern extensions are identified in the surrounding explanation; they are not presented as quotations from him.

§ 10.2

Give the number its clock back

Put a clock at each end of a bench. A flash leaves the left end, reflects at the right end, and comes back. Call the left clock A and the right clock B. Let t_A be A's emission reading, t_B B's reflection reading, and t'_A A's return reading. The prime here means the later reading at A. It does not name a third clock.

Einstein defines synchronization for these stationary clocks by

[11] Einstein (1905), introduction and §2, pp. 891–895. The inertial-frame language is our modern wording; his §1 first specifies a coordinate system in which Newtonian mechanics holds to the stated approximation.

[12] Albert Einstein, "On the Electrodynamics of Moving Bodies" (1905), especially §§1–5. The working English edition is John Walker's Fourmilab edition of the Perrett–Jeffery translation, with its editorial corrections: `https://www.fourmilab.ch/etexts/einstein/specrel/www/`. The original German used different symbols in several places. This chapter declares its modern notation explicitly rather than silently treating the English edition as a facsimile.

Die beiden Uhren laufen definitionsgemäß synchron, wenn

The two clocks run synchronously by definition if

Einstein, 1905, §1, p. 894. German excerpt / English translation.

$$t_B - t_A = t'_A - t_B. \tag{10.8}$$

The reflection reading belongs halfway between the departure and return readings. There is an outward light journey and a return journey; this convention assigns them equal durations in the chosen inertial frame. It does not establish a frame-independent universal time.

Use an ideal separation of 2.99792458 metres. At the vacuum speed $c = 299792458\,\mathrm{m/s}$, each leg takes ten nanoseconds. Our exact readings are therefore 0, 10, and 20 nanoseconds. There are three events but only two clocks. That distinction is part of the data.

Event	Operation	Clock	Reading (ns)
e_0	Emission	A	0
e_1	Reflection	B	10
e_2	Return	A	20

GIGI gets an event identity, the trial to which the event belongs, the clock identity, the frame, the event's role, and its reading with units. The arithmetic is represented separately by the signed terms in

$$r = 2t_B - t_A - t'_A. \tag{10.9}$$

GIGI groups these contributions by trial and adds them. A zero sum means that this exact triple satisfies the declared synchronization equation. The original observations remain available beside the sum.

Now move only B's reading to 11 nanoseconds. The residual becomes 2, not 1, because B occurs twice in the expression. If that clock offset is the only error in this ideal setup, the correction to B is $-r/2$. Shift all three readings forward by 100 nanoseconds instead and the residual remains zero. The epoch changed; the synchronization relation did not.

This has a short proof. Under a shared affine change of time coordinates, $\widetilde{t} = at + b$ with $a > 0$,

$$\widetilde{r} = 2(at_B + b) - (at_A + b) - (at'_A + b) \tag{10.10}$$

$$= ar. \tag{10.11}$$

So exact synchronization survives a common change of epoch and time unit. Changing just one clock is a different operation. The stored transformation must say which one we performed.

Real instruments require one more line of mathematics. If the absolute reading errors are bounded by ϵ_{A0}, ϵ_B, and ϵ_{A2}, then

$$|r_{\mathrm{measured}} - r_{\mathrm{true}}| \le \epsilon_{A0} + 2\epsilon_B + \epsilon_{A2}. \tag{10.12}$$

That bound follows directly from the triangle inequality. A small residual may be consistent with synchronization within the measurement error; it does not prove exact equality. Nor does checking one pair establish that an entire clock network is consistent. Einstein names symmetry and transitivity in his synchronization construction. A numerical implementation needs to keep that network question separate from a single passing trial.

In plain English: the number needs its clock, the clock needs its frame, and the comparison needs its rule. If we discard those relationships, more decimal places will not bring them back.

§ 10.3

A rod is not a pair of numbers

Keep the same idea, but let the rod move. In its own rest frame, call its length L_0. Call this rest frame k and the laboratory frame K. An observer moving with it can lay a ruler along it. A laboratory observer uses a different procedure: record both endpoint positions at the same laboratory time, then subtract them.

That last condition is doing real work. Measure the back now and the front later, and the rod can move between the two readings. You have a pair of positions, but you have not yet measured its length in that frame.

We will use $v = 3c/5$, so the relativistic factor is

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{5}{4}. \tag{10.13}$$

These large speeds make the distinction visible in a small numerical example. They describe an ideal relativistic setup, not a rod we have accelerated across a laboratory bench. To keep its arithmetic transparent, we measure time in nanoseconds and distance in light-nanoseconds: one light-nanosecond is 0.299792458 metres. In these matched numerical units, the value of c is one.

Take $L_0 = 10$ light-nanoseconds. The laboratory length is $L = L_0/\gamma = 8$. The left and right endpoint trajectories in laboratory coordinates are

$$x_A(t) = \frac{3}{5}t, \qquad x_B(t) = 8 + \frac{3}{5}t. \tag{10.14}$$

A light pulse leaves the left endpoint at $(t, x) = (0, 0)$. Its outward trajectory is $x = t$, so the meeting with the moving right endpoint solves

$$t = 8 + \frac{3}{5}t. \tag{10.15}$$

The reflection occurs at $t = 20$, $x = 20$. On its return the pulse follows $x = 40 - t$. Meeting the left endpoint requires $40 - t = 3t/5$, giving $t = 25$, $x = 15$.

The outward leg lasts twenty laboratory nanoseconds; the return leg lasts five. That is exactly Einstein's §2 calculation,

$$\Delta t_{\text{out}} = \frac{L}{c - v}, \qquad \Delta t_{\text{back}} = \frac{L}{c + v}. \tag{10.16}$$

The mirror moves away during the outward journey and the left endpoint moves toward the returning pulse. These unequal laboratory durations do not mean that the light has two different vacuum speeds.

Transform the same three events into the rod frame. With origins aligned,

$$\tau = \gamma\left(t - \frac{vx}{c^2}\right), \qquad \xi = \gamma(x - vt). \tag{10.17}$$

Emission becomes $(\tau, \xi) = (0, 0)$, reflection becomes $(10, 10)$, and return becomes $(20, 0)$. The rod-frame legs each take ten nanoseconds. We have not replaced the events. We have attached a second coordinate description to each one.

There is an even sharper way to see the distinction. At laboratory time zero, take event A at the left endpoint and event B at the right endpoint. Their laboratory coordinates are $(0, 0)$ and $(0, 8)$. In the rod frame, the same events have coordinates $(0, 0)$ and $(-6, 10)$. Their spatial separation is ten, but their rod-frame times differ by six nanoseconds.

Now choose event C at the right endpoint with rod-frame time zero. Its rod-frame coordinates are $(0, 10)$, while its laboratory coordinates are $(7.5, 12.5)$. Subtracting A's laboratory position from C's gives 12.5, but that is not the laboratory length. Those endpoint events were recorded at different laboratory times.

This is the part I do not want the data structure to flatten away. A–B and A–C concern the same rod. They are different event pairs answering different simultaneous-measurement questions. GIGI needs the endpoint identities, the event identities, the frame, and the selection condition. The subtraction comes after that choice.

§ 10.4

Deriving the coordinate transformation

Equation (10.17) should not arrive as a magic rule. Einstein gets there by requiring that synchronized clocks and light propagation remain compatible between the two frames. Let us keep the main steps in view.

First introduce the auxiliary quantity

$$x_* = x - vt. \tag{10.18}$$

It stays constant along a point attached to the moving frame. It is not yet the final moving-frame coordinate ξ. At the reflection in our example, $x_* = 8$, while $\xi = 10$. Giving both quantities the same field name would silently remove a scale factor from the derivation.

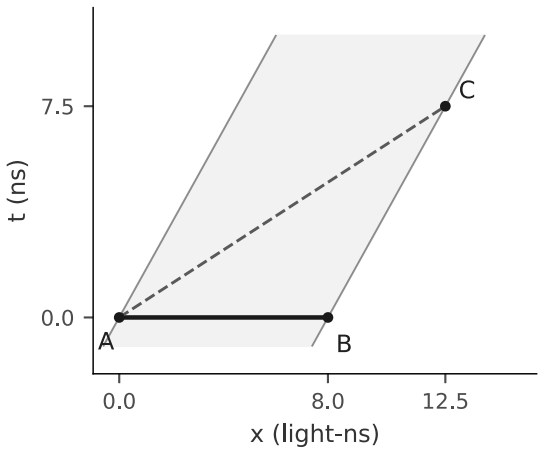
Under the homogeneity and aligned-origin assumptions, the coordinate map is linear. Write the new time as $\tau(x_*, y, z, t)$. Applying the same light-round-trip synchronization rule in the moving system gives

$$\begin{aligned} &\frac{1}{2}\Big[\tau(0, 0, 0, t) \\ &\quad + \tau\left(0, 0, 0, t + \frac{x_*}{c - v} + \frac{x_*}{c + v}\right)\Big] \\ &\qquad = \tau\left(x_*, 0, 0, t + \frac{x_*}{c - v}\right). \end{aligned} \tag{10.19}$$

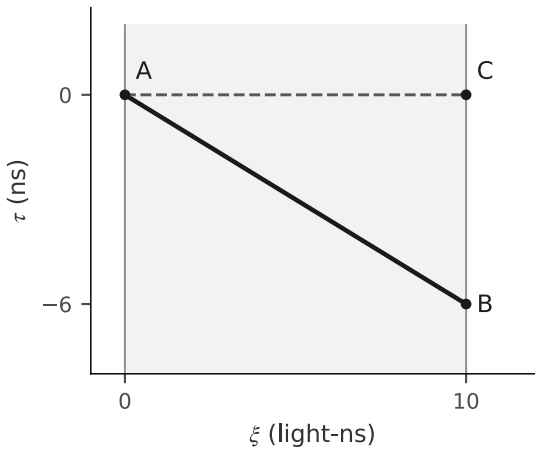
One rod, two definitions of now

EINSTEIN 1905, SECTIONS 2-4 / EXACT CONSTRUCTED EXAMPLE

01 Laboratory description K



02 Rod description k



A-B: simultaneous in K

A-C: simultaneous in k

EVENT	LABORATORY (t, x)	ROD FRAME (tau, xi)
A	(0, 0)	(0, 0)
B	(0, 8)	(-6, 10)
C	(7.5, 12.5)	(0, 10)

Length is an endpoint comparison made at a declared time.

The rod moves at 0.6c in K. Its length is 8 light-ns in K and 10 light-ns in k. A-C is not a K-length measurement.

Figure 35. The object stays the same while the simultaneous endpoint pair changes. The solid segment joins A and B, simultaneous in the laboratory frame. The dashed segment joins A and C, simultaneous in the rod frame. Coordinates are ordered as time, then position.

Differentiate this relation at an infinitesimal separation. The result is

$$\frac{\partial \tau}{\partial x_*} + \frac{v}{c^2 - v^2} \frac{\partial \tau}{\partial t} = 0. \tag{10.20}$$

Analogous transverse comparisons give $\partial\tau/\partial y = \partial\tau/\partial z = 0$. Together with linearity, these conditions give

$$\tau = a(v) \left(t - \frac{v}{c^2 - v^2} x_* \right). \tag{10.21}$$

There is a small but important distinction between the expression and some values sampled from it. An arbitrary table of time readings does not supply these partial derivatives. For this affine function, the derivatives are its coefficients. We can store the function definition, its coordinate arguments, and those coefficients, then check the constraint exactly. A more general field will require a declared differentiation method. That requirement becomes central in general relativity.

Matching the longitudinal and transverse light rays determines the spatial coordinates up to a common factor. After reparameterizing the unknown coefficient, Einstein obtains

Setzen wir für x' seinen Wert ein, so erhalten wir: Substituting the value of x', we obtain:
Einstein, 1905, §3, p. 900. German excerpt / English translation.

$$\tau = \phi(v)\gamma(t - vx/c^2), \qquad \xi = \phi(v)\gamma(x - vt), \tag{10.22}$$

$$\eta = \phi(v)y, \qquad \zeta = \phi(v)z. \tag{10.23}$$

Here $a(v) = \phi(v)/\gamma$. The two coefficients occur at different stages of the derivation; they should not be conflated. The German x' in this passage is our auxiliary x_*, not our final coordinate ξ; his light speed V is our c, and his Lorentz factor β is our γ.

Transform forward and then back. Agreement with the original rods and clocks requires $\phi(v)\phi(-v) = 1$. Reversing the direction of motion cannot change the transverse length of the same rod, so isotropy gives $\phi(v) = \phi(-v)$. These conditions imply $\phi(v)^2 = 1$. The positive branch, continuous from the identity transformation at $v = 0$, gives $\phi(v) = 1$. That supplies the normalization in equation (10.17). Preserving lightlike directions alone would not have removed this overall scale.

For GIGI, the transformation record therefore includes more than sixteen matrix entries. It identifies the source and target frames, coordinate order, units, relative velocity, aligned origins, domain $|v| < c$, and the normalization convention. If the origins differ, the transformation also needs a translation. These are the conditions under which the matrix means what we say it means.

In modern matrix notation let $X = (ct, x, y, z)^T$. The matrix that defines the interval is

$$\eta = \operatorname{diag}(1, -1, -1, -1).$$

A boost along x is

$$\Lambda = \begin{pmatrix} \gamma & -\gamma v/c & 0 & 0 \\ -\gamma v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{10.24}$$

Writing $\beta = v/c$, its central identity follows by expansion:

$$\gamma^2[(ct - \beta x)^2 - (x - \beta ct)^2] = \gamma^2(1 - \beta^2)[(ct)^2 - x^2] \tag{10.25}$$

$$= (ct)^2 - x^2. \tag{10.26}$$

The other two spatial coordinates are unchanged, so $\Lambda^T \eta \Lambda = \eta$. This proves interval preservation for the whole family of these boosts, not just our six numerical examples. Einstein's factor called β in the working English source is our γ; the notation map must travel with the transcription.

Look at a tick at $(ct, x) = (10, 0)$. In the moving frame it becomes $(12.5, -7.5)$. The coordinates differ, but

$$10^2 = 12.5^2 - 7.5^2 = 100. \tag{10.27}$$

Keep the old time while changing the position instead, and the naive pair $(10, -6)$ gives $100-36 = 64$. The deliberate failure is useful: the spatial shift by itself is not the relativistic transformation.

§ 10.5

The shape, the clock, and the next moving frame

Einstein next applies the transformation to a sphere at rest in the moving frame. Its surface satisfies

$$\xi^2 + \eta^2 + \zeta^2 = R^2. \tag{10.28}$$

At the laboratory time $t = 0$, substitution gives

$$\gamma^2 x^2 + y^2 + z^2 = R^2. \tag{10.29}$$

Its longitudinal semiaxis on this simultaneous laboratory slice is R/γ; the other two are R. For $R = 5$ and $v = 3c/5$, they are 4, 5, and 5. The point $(\xi, \eta, \zeta) = (3, 4, 0)$ corresponds on that slice to $(x, y, z) = (12/5, 4, 0)$, and

$$\frac{25}{16}\frac{144}{25} + 16 = 9 + 16 = 25. \tag{10.30}$$

This is a coordinate construction using simultaneous endpoint events. It is not a prediction of the photograph a camera receives, because light arriving at a camera can have left different surface points at different times. The representation must preserve which of those questions is being asked.

For a clock attached to the moving origin, $x = vt$. Substituting that trajectory into the time transformation gives

$$\Delta\tau = \frac{\Delta t}{\gamma}. \tag{10.31}$$

Twenty-five laboratory nanoseconds correspond to twenty on the moving clock. The exact lag is five. The low-speed approximation $\Delta t\, v^2/(2c^2)$ gives 4.5 here. Its half-nanosecond error is a reminder that a truncated approximation is a different mathematical object from an exact expression. In GIGI it needs an approximation order and a validity regime, not just a numeric result.

For a path with changing velocity, the modern proper-time expression is

$$\Delta\tau = \int \sqrt{1 - |\mathbf{v}(t)|^2/c^2}\, dt. \tag{10.32}$$

Using it for an accelerated physical clock includes the ideal-clock assumption. A return journey also needs the whole worldline, including the change of direction; a single uniform boost is not the complete round trip. Einstein's §4 discussion anticipates this issue by moving from straight segments to a curved route. We keep the assumptions visible when extending the example.

Now let an object move with velocity $\mathbf{w}$ in the rod frame. Differentiating the inverse transformation gives its laboratory velocity:

Gesucht ist die Bewegung des Punktes relativ zum System K.	We seek the motion of the point relative to system K.

Einstein, 1905, §5, p. 905. German excerpt / English translation.

$$u_x = \frac{w_x + v}{1 + vw_x/c^2},$$
$$u_y = \frac{w_y}{\gamma(1 + vw_x/c^2)}, \qquad u_z = \frac{w_z}{\gamma(1 + vw_x/c^2)}. \tag{10.33}$$

Every component uses the transformed time interval. We cannot add a longitudinal speed and leave that denominator behind.

Velocity in k, in units of c	Velocity in K	$\lvert\mathbf{u}\rvert^2/c^2$
$(3/5, 0, 0)$	$(15/17, 0, 0)$	$225/289$
$(-3/5, 0, 0)$	$(0, 0, 0)$	0
$(0, 3/5, 0)$	$(3/5, 12/25, 0)$	$369/625$
$(3/5, 4/5, 0)$	$(15/17, 8/17, 0)$	1
$(-1, 0, 0)$	$(-1, 0, 0)$	1

Two velocities of $0.6c$ along the same direction give $15c/17$, not $1.2c$. The speed limit follows from an identity, not from rounding down the answer. In dimensionless variables, with both $\mathbf{u}$ and $\mathbf{w}$ now expressed in units of c,

$$1 - |\mathbf{u}|^2 = \frac{(1 - \beta^2)(1 - |\mathbf{w}|^2)}{(1 + \beta w_x)^2}. \tag{10.34}$$

For $|\beta| < 1$ and $|\mathbf{w}| < 1$, the right side is positive. A lightlike input gives zero. The same equation covers longitudinal, transverse, and oblique motion.

The source also gives a formula using the angle between the velocities. We preserve the component version alongside it so that the angle's quadrant and reference direction remain explicit. The correct numerator contains $(vw \sin\alpha/c)^2$. A web-image description instead contains c^2 in that denominator. Choosing units with $c = 1$ would hide the transcription error. With $c = 5$, $v = 3$, and moving-frame velocity $(3, 4)$, the correct squared speed is 25; the erroneous description gives $8089/289$, which is greater than 25.[13]

[13]This discrepancy is between the image alternative text and the downloadable `specrel.tex` in the working Fourmilab edition. The original German printing, p. 906, confirms the single power of the light speed inside the squared term; that printing calls the light speed V. The retained source files and their hashes identify these versions.

§ 10.6

Carry the field, not just its amplitude

The second half of the 1905 paper applies the coordinate transformation to electromagnetism. Electric and magnetic components change together. That is a different operation from moving a list of amplitudes into a new table and keeping their labels unchanged.

For a boost along x, use modern SI field notation but write $\mathbf{H}_c = c\mathbf{B}$ so the two vectors have matching units. Then

Vektoren der elektrischen und magnetischen Kraft — Vectors of electric and magnetic force

Einstein, 1905, §6, p. 908. German excerpt / English translation.

$$E'_x = E_x, \qquad H'_{cx} = H_{cx}, \tag{10.35}$$

$$E'_y = \gamma(E_y - \beta H_{cz}), \qquad H'_{cy} = \gamma(H_{cy} + \beta E_z), \tag{10.36}$$

$$E'_z = \gamma(E_z + \beta H_{cy}), \qquad H'_{cz} = \gamma(H_{cz} - \beta E_y). \tag{10.37}$$

This notation maps the source's electric and magnetic quantities into a declared unit convention; it does not treat the source's magnetic symbols as unlabelled SI tesla values.

Take a purely magnetic input with $\mathbf{E} = 0$ and $\mathbf{H}_c = (0, 0, 1)$ in normalized units. Our moving frame has $E'_y = -3/4$ and $H'_{cz} = 5/4$. An electric component has appeared in the new description. The invariant still reads

$$|\mathbf{E}'|^2 - |\mathbf{H}'_c|^2 = \frac{9}{16} - \frac{25}{16} = -1. \tag{10.38}$$

The second invariant, $\mathbf{E} \cdot \mathbf{H}_c$, also remains unchanged. Expanding the component equations proves both identities; the replay checks them for a pure magnetic field, a plane wave, and a field with all six components nonzero.

The field values belong to the same event in both descriptions. A field record therefore needs the event or spacetime location, the frame, component order, unit convention, and transformation. Maxwell's equations require another layer: the fields' derivatives. Transforming their values at isolated sample points does not establish that a differential equation holds throughout a region. For that, we must retain the functions or a specified numerical differentiation scheme and test the equation's residuals.

Einstein's next example is a plane wave. Write its phase as

$$\Phi = \omega\left(t - \frac{\mathbf{n} \cdot \mathbf{x}}{c}\right), \qquad |\mathbf{n}| = 1. \tag{10.39}$$

The phase at a particular event is the same scalar in either description. Its frequency and direction need not be the same. If $D = \gamma(1 - \beta n_x)$, then

die von dem Beobachter wahrgenommene Frequenz ν' des Lichtes — the frequency ν' of the light perceived by the observer

Einstein, 1905, §7, p. 911. German excerpt / English translation.

$$\omega' = D\omega, \qquad n'_x = \frac{n_x - \beta}{1 - \beta n_x}, \qquad n'_y = \frac{n_y}{D}, \quad n'_z = \frac{n_z}{D}. \tag{10.40}$$

For light traveling along the observer's positive direction of motion, $D = 1/2$. For transverse light, $D = 5/4$ and its new direction is $(-3/5, 4/5, 0)$. We checked the unit direction, the frequency factor, and the phase at the same event. The quantity to preserve is not a frequency field by itself; it is the wave's relationship to events.

§ 10.7

Follow the packet all the way through the calculation

This is where the paper comes particularly close to the question in Fuel. Einstein does not stop after transforming an amplitude. He asks what happens to the energy of the same light packet.

His warning is explicit: the density ratio would give the energy ratio if the packet occupied the same volume in both descriptions.

wenn das Volumen eines Lichtkomplexes in K gemessen und in k gemessen das gleiche wäre. Dies ist jedoch nicht der Fall.	if the volume of a light complex measured in K and measured in k were the same. This, however, is not the case.

Einstein, 1905, §8, p. 913. German excerpt / English translation.

For the plane wave, its energy density changes by D^2. But its volume, measured on the other frame's simultaneity slice, changes by $1/D$. Therefore

$$\frac{u'}{u} = D^2, \qquad \frac{V'}{V} = \frac{1}{D}, \qquad \frac{E'}{E} = \frac{u'V'}{uV} = D. \tag{10.41}$$

With $D = 1/2$, the density ratio is $1/4$, the volume ratio is 2, and the total-energy ratio is $1/2$.

Think about what would go wrong if we kept only the density column. We would report a quarter of the energy while quietly treating the packet's volume as unchanged. The arithmetic on that one column would be correct. The physical answer would not. The representation needs to keep track of which packet is being compared and which spatial slice defines its volume.

Now change the experiment. Keep one laboratory frame, but let light push a perfectly reflecting mirror moving away at $0.6c$. There is a physical transfer of energy here, rather than two coordinate descriptions of a packet. Assume normal incidence and a prescribed constant mirror speed. Let the incident beam's energy density be u_0.

The reflected frequency and amplitude factors in this example are

$$q = \frac{1-\beta}{1+\beta} = \frac{1}{4}. \tag{10.42}$$

The incoming power per mirror area is $u_0 c(1-\beta)$: the mirror moves away from part of the flux. The reflected power per area leaving the moving boundary is $u_0 c\, q^2(1+\beta)$. Einstein's pressure formula at normal incidence gives

$$P = 2u_0 \frac{(1-\beta)^2}{1-\beta^2} = \frac{1}{2}u_0. \tag{10.43}$$

Multiplying by the mirror speed gives the mechanical power per area. The whole account is

$$\underbrace{0.4u_0c}_{\text{arriving}} - \underbrace{0.1u_0c}_{\text{reflected}} - \underbrace{0.3u_0c}_{\text{mechanical}} = 0. \tag{10.44}$$

The difference did not come from nowhere. The light supplied it. If we hold the mirror at a prescribed speed while extracting work, that mechanical interaction belongs in the model too. This is not a closed cycle with a free reset. It is a sharply defined transfer calculation whose three terms we can retain and inspect.

The head-on example makes the account easy to see. But Einstein's mirror calculation allows the beam to arrive at an angle. That is worth opening up, because the angle changes both how much light reaches the surface and what "reflected" looks like from the laboratory.

Picture a large, flat mirror sliding to the right on a straight rail. Its surface is perpendicular to the rail. A beam approaches from behind, traveling partly to the right and partly upward. Let x run along the rail and y run up the mirror. The mirror moves at $v = \beta c$; write the incoming unit direction as (μ, s), where

$$\mu = \cos\phi, \qquad s = \sin\phi, \qquad \mu^2 + s^2 = 1.$$

The angle ϕ is measured from the positive rail direction, not from the mirror's surface. We take $s \geq 0$ to draw the beam above the rail. The normal component of the light's velocity is $c\mu$. To catch the mirror from this side, it must satisfy $\mu > \beta$. Light still travels at c when this condition fails. It simply does not catch this moving surface from behind.

Here is the whole method: describe the incoming beam in the mirror's rest frame, reflect it there, and translate the result back. In that middle description the mirror is stationary. A perfect reflection reverses the momentum perpendicular to its surface and leaves the parallel momentum alone. Its frequency also stays unchanged in that frame. We assume a prescribed constant mirror speed, a perfectly reflecting plane, and no absorption. A freely recoiling mirror would need its mass and changing velocity included in a different calculation.

For a compact derivation, use a light packet with laboratory energy E, and express its momentum in units of E/c. Its three entries $(E, cp_x, cp_y)/E$ are initially $(1, \mu, s)$. This is modern energy–momentum notation for Einstein's transformation-and-reflection argument, not a claim that he wrote his 1905 calculation in this form. The first boost gives

$$\frac{(E', cp'_x, cp'_y)}{E} = (\gamma(1-\beta\mu),\, \gamma(\mu-\beta),\, s), \qquad \gamma = (1-\beta^2)^{-1/2}. \tag{10.45}$$

Für das reflektierte Licht erhalten wir, wenn wir den Vorgang auf das System k beziehen:

For the reflected light, referring the process to system k, we obtain:

Einstein, 1905, §8, p. 914. German excerpt / English translation.

Reflection changes p'_x to $-p'_x$. Transforming back gives

$$q := \frac{E_{\mathrm{r}}}{E} = \gamma^2(1 - 2\beta\mu + \beta^2), \tag{10.46}$$

$$\frac{cp_{x,\mathrm{r}}}{E} = \gamma^2\big[2\beta - (1+\beta^2)\mu\big], \qquad \frac{cp_{y,\mathrm{r}}}{E} = s. \tag{10.47}$$

Energy density is not total energy

EINSTEIN 1905, SECTIONS 7-8 / TWO DIFFERENT COMPARISONS

01 Describe the same light packet from another frame

An observer moves at 0.6c along the ray. The Doppler factor is 1/2. Compare the same packet on each frame's simultaneity slice.

ENERGY DENSITY		PACKET VOLUME		TOTAL ENERGY
1/4	x	2	=	1/2

All three numbers are ratios: moving-frame value divided by stationary-frame value.

02 Keep one frame and let the light push a moving mirror

Now the mirror recedes at 0.6c. This is an energy transfer in the laboratory frame. The incoming beam has normalized energy density 1.

ARRIVING POWER		REFLECTED POWER		MECHANICAL POWER
0.4	=	0.1	+	0.3

Power per area is in units of u0 c. Pressure is 0.5 u0; multiplying it by 0.6c gives 0.3 u0 c. The ledger closes: 0.4 - 0.1 - 0.3 = 0.

Changing the description and transferring energy are different operations.

Upper panel: one packet, two frames. Lower panel: one frame, moving boundary, perfect reflection and normal incidence.

Figure 36. Two energy questions from Einstein's §§7–8. The upper comparison changes the reference frame for the same packet. The lower comparison keeps one frame and accounts for a physical transfer at a moving mirror. Their records and their invariants must not be conflated.

Dividing the momentum entries by the returned energy gives the outgoing direction:

$$n_{x,\mathrm{r}} = \frac{2\beta - (1+\beta^2)\mu}{1 - 2\beta\mu + \beta^2}, \qquad n_{y,\mathrm{r}} = \frac{s}{q}. \tag{10.48}$$

These expressions satisfy $n_{x,\mathrm{r}}^2 + n_{y,\mathrm{r}}^2 = 1$. The returned beam still travels at light speed. In Einstein's plane-wave calculation, the same q is the reflected frequency and field-amplitude ratio, so the energy-density ratio is q^2.[14]

Let us put actual numbers into that journey. Keep the mirror speed at $3c/5$, and aim the light along $(4/5, 3/5)$. The incoming beam makes an angle of about $36.87°$ to the rail. The complete transformation record is

Description	E_{stage}/E	cp_x/E	cp_y/E
Laboratory, incoming	1	4/5	3/5
Mirror frame, incoming	13/20	1/4	3/5
Mirror frame, reflected	13/20	−1/4	3/5
Laboratory, reflected	5/8	7/40	3/5

Every entry uses the original laboratory energy E as its reference. We have not renormalized each row to make its first entry one. That is why the table can retain the energy change as well as the direction.

The reflected laboratory direction is $(7/25, 24/25)$. Look at that first number: it is positive. The reflected light is still moving to the right. That can look wrong until we put the mirror back into the picture. The light moves rightward at $0.28c$ while the mirror moves rightward at $0.6c$. The mirror pulls ahead. The beam leaves the same side it arrived from, even though both are moving rightward in the laboratory. In the mirror's own frame, the normal direction reverses exactly as we expect. "Reflected" describes what happened at the surface; it does not promise a negative laboratory x velocity.

There is a second check inside the same table. The mirror receives normal impulse

$$\Delta p_x = \frac{E}{c}\left(\frac{4}{5} - \frac{7}{40}\right) = \frac{5E}{8c},$$

and no tangential impulse. At its prescribed speed, the corresponding work is $v\Delta p_x = 3E/8$. The incoming packet had energy E; the returned packet has $5E/8$. The missing $3E/8$ is accounted for. This packet calculation and the continuous-beam power calculation below must agree, but they are different records: one counts a packet's energy, the other counts energy crossing a moving area per unit time.

Let u_0 be the incoming beam's laboratory energy density. The incoming power per mirror area is

$$\mathcal{I} = u_0 c(\mu - \beta). \tag{10.49}$$

[14]Einstein (1905), §8, original German pp. 914–915, the reflection calculation following the light-complex energy transformation. The original p. 915 prints $(1 - v/V)^2$ in the final frequency denominator, where V denotes light speed. The preceding transformation and amplitude formula require $1-(v/V)^2$, also used in the working English TeX edition. The derivation above retains that algebraically consistent denominator. The working English TeX source gives the incident, mirror-rest, reflected-rest and returned-laboratory quantities as unprimed, primed, double-primed and triple-primed entries, respectively. Here the subscript r means reflected in the laboratory. Our γ is the Lorentz factor; our $\beta = v/c$ is the speed ratio. The packet notation is a modern derivation of the same ratios.

We need the speed relative to the moving surface. Using $u_0 c\mu$ would count crossings of a fixed plane instead. The reflected density is $q^2 u_0$, and the beam separates from the mirror at normal speed $c(\beta - n_{x,\mathrm{r}})$, so its outgoing power per area is

$$\mathcal{O} = u_0 c q^2 (\beta - n_{x,\mathrm{r}}). \tag{10.50}$$

Substitution in Equation 10.48 gives

$$\beta - n_{x,\mathrm{r}} = \frac{\mu - \beta}{q}, \qquad \mathcal{O} = q\mathcal{I}.$$

That identity joins the packet and beam descriptions: each returned packet carries the fraction q of its incoming energy, and the boundary flux accounts for which packets actually reach the mirror.

We can obtain the pressure directly from normal momentum transfer, without dividing by v. This also covers a stationary mirror:

die pro Zeiteinheit geleistete Arbeit des Lichtdruckes.	the work done per unit time by the pressure of light.

Einstein, 1905, §8, p. 915. German excerpt / English translation.

$$P = \frac{\mathcal{I}\mu - \mathcal{O} n_{x,\mathrm{r}}}{c} = 2u_0 \frac{(\mu - \beta)^2}{1 - \beta^2}, \tag{10.51}$$

$$\mathcal{I} - \mathcal{O} = u_0 c(\mu - \beta)(1 - q) = Pv. \tag{10.52}$$

For $|\beta| < 1$ and $\mu > \beta$, these are exact identities of this idealized model. They prove that the power account closes for every admissible angle in the model, not just for the numbers we chose.

The worked oblique case becomes especially readable when all powers are divided by $u_0 c$:

Quantity	Exact value	Decimal
Arriving power	1/5	0.200
Reflected power	1/8	0.125
Mechanical power delivered	3/40	0.075
Residual	0	0.000

The pressure itself is $u_0/8$. Multiplying it by $3c/5$ gives the mechanical-power row. In GIGI's signed-term record the balance uses 40 ticks per unit of $u_0 c$: the engine receives 8, -5 and -3, and returns their sum, zero, with a count of three terms. The rational preparation and the general proof remain in the companion calculation. The returned table checks that GIGI carried this particular account without dropping or changing one of its terms.

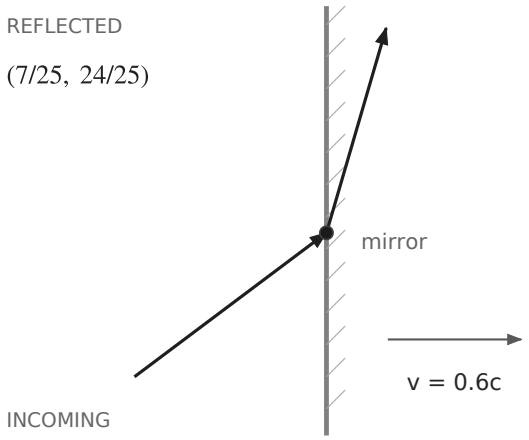
Think of the mirror as a moving gate. The beam may contain plenty of energy, but only the light that catches the gate can push it. Then we have to ask how much energy leaves with that light and how much was transferred during the encounter. Keeping only a frequency, or only a density, skips part of that story. Here we can follow all of it, from the incoming direction to the last entry in the power account.

The limiting cases help us catch a mistake before it becomes a story. At $\beta = 0$, the energy ratio is one: a stationary perfect mirror receives momentum but does no mechanical work. As μ approaches β from above, the arriving flux and pressure go to zero. The ray is becoming tangent

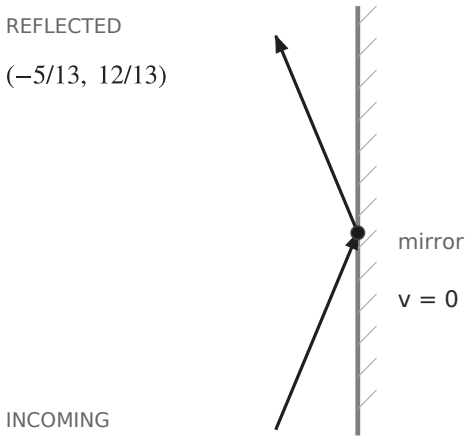
A reflection can still travel forward

EINSTEIN 1905, SECTION 8 / EXACT OBLIQUE-MIRROR EXAMPLE

01 Laboratory frame K



02 Mirror rest frame k



03 Keep the energy with the direction

DESCRIPTION	E	cp_x	cp_y
K incoming	1	4/5	3/5
k incoming	13/20	1/4	3/5
k reflected	13/20	−1/4	3/5
K reflected	5/8	7/40	3/5

04 Close the power account

Arrives	1/5
Leaves as light	1/8
Mechanical	3/40

$$\frac{1}{5} - \frac{1}{8} - \frac{3}{40} = 0$$

The mirror outruns the reflected beam along the rail.

Upper arrows: unit directions. Table: entries divided by incident energy E. Powers: per area, in units of u0 c.

Figure 37. One reflection, two descriptions, one closed power account. Arrows show propagation directions. The mirror moves at $3c/5$ in K and is stationary in k. The table retains energy and momentum together.

to the surface in its rest frame. Equality is a zero-flux limit, not a beam delivering finite power to the mirror. For a mirror approaching at $\beta = -3/5$ with the same incident direction $(4/5, 3/5)$, the returned energy ratio is $29/8$. The incoming and outgoing powers are $7/5$ and $203/40$ in units of $u_0 c$; mechanical power delivered to the mirror is $-147/40$. The negative sign means the mirror's mechanical side supplies energy to the light. Holding that speed requires an external agency to supply the corresponding power.

So a brighter reflected beam does not settle the energy question. We have to follow the moving boundary and the mechanism maintaining its motion. This is the part I want the representation to remember: the conditions that made the number possible belong beside the number. For Fuel, that takes us back to the whole cycle. Extracting work from one interaction does not tell us what it costs to prepare the next one.

Charge density requires the same care. In a modern four-current description, write $J = (c\rho, j_x, j_y, j_z)$. A stationary charge distribution with normalized $(\rho, j_x/c) = (1, 0)$ becomes $(5/4, -3/4)$ in our moving frame. Density has increased. That does not mean charge has been created. For a bounded distribution at rest in K, its simultaneous volume contracts by the reciprocal factor, so its total charge is unchanged.[15]

Finally, the electron calculation in §10 follows energy withdrawn from an electric field into the particle's kinetic energy. With radiative losses neglected in the slow-acceleration argument,

die dem elektrostatischen Felde entzogene Energie	the energy withdrawn from the electrostatic field
gleich der Bewegungsenergie W des Elektrons	equal to the kinetic energy W of the electron

Einstein, 1905, §10, p. 920. German excerpt / English translation.

$$W = m \int_0^v \gamma(s)^3 s \, ds = mc^2(\gamma(v) - 1). \tag{10.53}$$

Differentiating the expression on the right recovers $m\gamma(v)^3 v$, and its value at $v = 0$ is zero. For $v = 3c/5$, the kinetic energy is $mc^2/4$. In modern energy–momentum notation, $E/(mc^2) = 5/4$ and $p/(mc) = 3/4$, giving

$$E^2 - p^2c^2 = m^2c^4. \tag{10.54}$$

This modern notation helps us express the relationship; it must not be passed off as the notation of every line of the 1905 paper.

There is another representation issue in the source that is easy to miss. Its "transverse mass" coefficient is $m\gamma^2$, using a force referred to the instantaneous rest frame with acceleration referred to the laboratory frame. With laboratory force and transverse laboratory acceleration, the modern response coefficient is instead $m\gamma$. These numbers describe different force conventions. The source itself warns that changing the definitions changes the coefficients. Putting both into a field called `mass`, with no record of those definitions, would manufacture a contradiction.

[15] In §9 of the working source, Einstein's symbol ρ denotes 4π times the charge density in the source's unit convention. The example here uses physical charge density normalized by a chosen reference density. The source symbol and the normalized field therefore have separate definitions.

§ 10.8

What GIGI has actually carried

The examples now have a common structure. Each starts with a physical question, identifies the events or fields that enter it, declares the frame and units, and states the operation that must preserve the answer. Only then do we ask the engine to calculate.

In the current replay, GIGI stores a registry of the operations and their signed terms. Each rational term is converted to integer ticks under a declared scale. GIGI groups the terms by operation and returns their sums and counts. The companion calculation constructs products and quotients, supplies independent exact-rational references, and performs the symbolic checks. All 193 grouped results agree with those references.[16]

For example, take the oblique light ray with direction $(3/5, 4/5, 0)$ and the laboratory event $(t, x, y, z) = (10, 3, 4, 0)$. Its phase divided by the original angular frequency is $10 - 9/5 - 16/5 = 5$ nanoseconds. In the moving frame the event becomes $(41/4, -15/4, 4, 0)$, the ray points along the positive transverse axis, and the frequency factor is $4/5$. The transformed phase calculation is

$$\frac{4}{5}\frac{41}{4} - \frac{4}{5}(4) = \frac{41}{5} - \frac{16}{5} = 5.$$

These are some of the terms and returned sums in the replay:

Question	Signed contributions	Returned sum
Auxiliary rod coordinate	$20 - 12$	8 light-ns
Oblique phase divided by ω	$41/5 - 16/5$	5 ns
Mirror power balance	$2/5 - 1/10 - 3/10$	0 in units of $u_0 c$

The phase computation stores integer contributions 41 and -16 with five ticks per nanosecond; the returned sum is 25 ticks, or five nanoseconds for the phase divided by the original frequency. Zero-valued contributions from the other spatial components remain in the full record. A zero term and a missing term are different things. The counts help us check that the engine saw the complete operation.

The point is not that five is a particularly interesting answer. It is that we can follow that answer back to the same event, the same wave, and an explicit change of description. If someone changes the frequency but leaves the direction or event coordinates behind, we have enough structure to identify which comparison they actually made.

[16] The reproducible package is `research/einstein_gigi` in the working research archive. `extended_replay.py` produces the complete GQL script, returned tables, scales, term counts and a binary hash. The calculations use fresh isolated databases within single CLI processes. Restart durability is not part of this result: separate reopen tests produced incomplete readback, retained as an unresolved implementation finding.

That is a useful result, but its scope matters. A database can store the components of a metric without having a native general-relativity solver. It can sum a derivative constraint after we provide its exact coefficients without having differentiated the function itself. It can check every row of an example without proving the general theorem. Those are separate capabilities, and the chapter keeps them separate.

The same care applies when we say we have represented a paper. Keeping an equation's text gives us something to read. Giving its symbols their units, frames and domains gives us something to interpret. Building and checking the operation gives us something to calculate. We should be able to tell which of those steps has actually happened. The worked cases here reach the calculation stage. The complete source inventory remains a larger translation project.[17]

The mathematical work has nevertheless shown exactly what the next layer requires. Events need stable identities. Coordinates need charts and frames. Components need types and transformation laws. A density needs the measure that turns it into a total. An approximation needs its regime. A physical energy claim needs the complete account of what enters, what leaves, and what has to be restored.

The general theory will ask us to carry one more thing: a ruler that changes from point to point. Then a coordinate table alone will not be enough. We will need the metric field, its derivatives, and the rule that tells us how to compare neighboring descriptions. That is where the next part of the investigation begins.

[17]The source register retains 357 mathematical occurrences from the 1905 working TeX edition, including 82 displays, and 571 occurrences from the 1916 working web edition. Repeated symbols are counted as occurrences, not as distinct results. These inventories do not certify individual bindings, full facsimile reconciliation or execution of every source expression. The coverage review records those remaining distinctions separately from the successful numerical replays.

Chapter 11

When the ruler becomes part of the record

The moving rod forced us to keep the event and its description separate. But a steadily moving train only takes us so far. What happens when it turns, or when the floor pushes up against our feet? Those are different physical questions. Special relativity can describe accelerated motion in flat spacetime, but changing coordinates in that spacetime does not by itself supply a theory of gravitation.

By the time Einstein published the 1916 paper, the mathematical conversation had widened too. In its introduction he names Minkowski's contribution to the spacetime formulation and the geometric work of Gauss, Riemann, Christoffel, Ricci and Levi-Civita. He also thanks Marcel Grossmann for helping him with the mathematical literature and with the investigation of the gravitational field equations. There is a person asking a difficult physical question here, and there is a body of mathematics that other people have spent years making possible.[1]

I want to keep that in view because a metric tensor can look as if it arrived fully formed and everybody else was simply supposed to know what it meant. We can enter the problem much closer to the floor. Imagine a small enclosed cabin, away from gravitational fields, being accelerated by a rocket. The floor pushes against your feet. Release a small object and the floor accelerates toward it. In a cabin supported near Earth's surface, you also feel the floor's support and see a released object fall. That local resemblance is the starting point of the equivalence principle.[2]

The word *local* matters. Give the cabin enough extent and precise enough instruments, and a gravitational field can reveal differences between neighboring freely falling paths. Two objects released side by side need not keep exactly the same separation. Those relative effects are called *tidal effects*. Curvature describes them. A coordinate change can make a straight path look curved on a chart; it cannot turn flat spacetime into curved spacetime. Later we will calculate both cases, so this distinction will have numbers behind it.

For now, imagine putting clocks around the edge of a turning platform, with another clock at its centre. The clock at the centre and a clock at the edge do not follow the same motion. A single column headed *time* cannot tell us which clock we mean, what it measured, or how its reading relates to the coordinates on our drawing.

Einstein uses a rotating system in the opening of his 1916 paper to make precisely this kind of difficulty unavoidable. Coordinate labels cannot automatically do the work of physical rulers and

[1] Einstein, "Die Grundlage der allgemeinen Relativitätstheorie" (1916), introduction, pp. 769–770. These acknowledgments are Einstein's own account in the paper.

[2] Einstein (1916), §2, pp. 771–773, develops the equivalence of an accelerated description and a suitable gravitational field. The cabin is our teaching example. Carroll, *Lecture Notes on General Relativity*, chapter 4, explains the local equivalence principle and the role of curvature.

clocks. We need a mathematical object that says how the labels connect to measurements. That object is the metric.[3]

This connects to the representation question in Fuel, but it does not make every data metric a gravitational field. A metric used to compare records may be positive definite. Spacetime uses a Lorentzian metric, with one sign different from the other three. Its zeros can describe light traveling between distinct neighboring events. We cannot exchange those two kinds of metric merely because both are stored as matrices.

The next examples use modern notation consistently. Coordinates are ordered with time first, indices run from zero to three, and the metric signature is $(+,-,-,-)$. We set $c = 1$ using matched time and length units. Einstein's source puts the time coordinate fourth. In addition, the symbol he introduces for the gravitational field in §13 is the *negative* of the Christoffel symbol used below. Those conventions are part of the translation, not cosmetic differences.

The metric's job is already visible in Einstein's description of ds^2:

unabhängig von jeder besonderen Koordinatenwahl definierte Größe	a quantity defined independently of any particular choice of coordinates

Einstein, 1916, §4, p. 778. German excerpt / English translation.

In our notation, his general quadratic expression is

$$ds^2 = g_{ab}\, dx^a dx^b.$$

The coordinate increments and the metric entries change together. Their contraction gives the same interval. The workshop example next puts that relationship somewhere we can picture it.

§ 11.1

An address is not a measurement

Start somewhere less intimidating than spacetime: a flat workshop floor. We can mark a location by its east and north coordinates, (x, y), or by its distance and angle from a fixed marker, (r, θ):

$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{11.1}$$

The floor has not changed. The labels have.

Differentiate both coordinates:

$$dx = \cos\theta\, dr - r\sin\theta\, d\theta, \tag{11.2}$$

$$dy = \sin\theta\, dr + r\cos\theta\, d\theta. \tag{11.3}$$

[3]Albert Einstein, "Die Grundlage der allgemeinen Relativitätstheorie," *Annalen der Physik* 49 (1916), 769–822, especially §§1–4. The retained original scan is available at `https://www.physik.uni-augsburg.de/annalen/history/einstein-papers/1916_49_769-822.pdf`. The working English text is the Wikisource edition of Satyendra Nath Bose's translation; that edition restores Einstein's notation and makes editorial corrections. Its wording and formulas are checked rather than treated as an infallible transcription.

Squaring and adding cancels the cross terms. We obtain

$$d\ell^2 = dx^2 + dy^2 = dr^2 + r^2 d\theta^2. \tag{11.4}$$

The factor r^2 has a physical job. The same small turn covers more floor when you are farther from the marker. One radian at a radius of two metres and one radian at a radius of twenty metres are not the same distance. The angle column alone never claimed that they were.

At $r = 2$ and $\theta = 0$, take a tangent displacement with $dr = 3\epsilon$ and $d\theta = 2\epsilon$, where ϵ is an infinitesimal scale parameter. The Cartesian components are $dx = 3\epsilon$, $dy = 4\epsilon$. Both descriptions give $d\ell^2 = 25\epsilon^2$. This is a tangent calculation. It is not a claim that a finite two-radian arc is identical to its tangent segment.

In spacetime the corresponding flat cylindrical description is

$$ds^2 = dt^2 - dr^2 - r^2 d\theta^2 - dz^2. \tag{11.5}$$

Its metric components vary with r, even though the spacetime is flat. Already we have a useful warning: changing metric entries do not, by themselves, establish curvature.

GIGI therefore needs a chart record before it needs a list of metric entries. That record names the coordinates, their units, the region where they are valid, and the transformation connecting them to another chart. Here $r > 0$, and the angle needs a local branch. The polar labels fail at the origin. The floor does not.

There is an equally important distinction between a vector and a covector. A vector can represent a displacement or velocity. A covector can represent the rule that takes such a displacement and returns a scalar, such as the first-order change of a measured scalar field. They do not transform in the same way.

Let $q = (r, \theta)$, $X = (x, y)$, and let $J = \partial X/\partial q$ be the Jacobian matrix. At our chosen point, $J = \mathrm{diag}(1, 2)$. If a vector has Cartesian components $V_X = (3, 4)^T$, then

$$V_q = J^{-1} V_X = (3, 2)^T. \tag{11.6}$$

If a covector has Cartesian components $\alpha_X = (2, 3)^T$, its polar components are instead

$$\alpha_q = J^T \alpha_X = (2, 6)^T. \tag{11.7}$$

Their pairing stays the same:

$$\alpha_X^T V_X = 2(3) + 3(4) = 18 = 2(3) + 6(2) = \alpha_q^T V_q. \tag{11.8}$$

The proof is only one line: $(J^T \alpha_X)^T (J^{-1} V_X) = \alpha_X^T V_X$. If we mistakenly transform the covector like a vector, the answer is nine. The values still fit neatly in a table. The type mistake is what breaks the calculation.

That is why an upper or lower index needs to survive the move into GIGI. It tells us which transformation applies. In Einstein's §§5–8, tensor products build objects with several such slots, and contraction pairs an upper slot with a lower one. A metric changes a vector into a covector;

its inverse changes it back. A plain array does not tell us any of this until we declare what its entries mean.

§ 11.2

A turning coordinate grid is not a bent world

Put a trolley on a straight rail across the same floor. Let its Cartesian track be

$$x(\lambda) = \lambda, \qquad y(\lambda) = b, \qquad b > 0. \tag{11.9}$$

The parameter λ measures progress along the rail. In Cartesian coordinates, both second derivatives vanish. In polar coordinates, the radius and angle change at changing rates:

$$r = \sqrt{\lambda^2 + b^2}, \qquad \dot{r} = \frac{\lambda}{r}, \qquad \ddot{r} = \frac{b^2}{r^3}, \tag{11.10}$$

$$\dot{\theta} = -\frac{b}{r^2}, \qquad \ddot{\theta} = \frac{2b\lambda}{r^4}. \tag{11.11}$$

Dots here mean derivatives with respect to λ, not readings from a new clock. The second derivatives are nonzero, but nobody has bent the rail.

We need a correction for the changing coordinate basis. For a metric g, the Levi-Civita connection has coefficients

Hierbei ist nach Christoffel gesetzt Here, following Christoffel, we set

Einstein, 1916, §9, p. 792. German excerpt / English translation.

$$\Gamma^{a}{}_{bc} = \frac{1}{2} g^{ad} \left(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc} \right). \tag{11.12}$$

A repeated upper and lower index is summed over the coordinates. The inverse metric g^{ad} satisfies $g^{ad} g_{db} = \delta^a_b$. The connection in equation (11.12) is the one with zero torsion—its two lower indices are symmetric in a coordinate basis—that preserves the metric under parallel transport.

For the polar spatial part, its only nonzero coefficients are

$$\Gamma^{r}{}_{\theta\theta} = -r, \qquad \Gamma^{\theta}{}_{r\theta} = \Gamma^{\theta}{}_{\theta r} = \frac{1}{r}. \tag{11.13}$$

The geodesic equations become

$$\ddot{r} - r\dot{\theta}^2 = 0, \qquad \ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} = 0. \tag{11.14}$$

Substitute the track above. The radial equation is $b^2/r^3 - b^2/r^3 = 0$. The angular equation is $2b\lambda/r^4 - 2b\lambda/r^4 = 0$. This is an identity along the whole track, not just at one sampled location.

At $b = 4$ and $\lambda = 3$, $r = 5$. These are the exact terms returned by the GIGI sum in the worked example:

Equation	Coordinate term	Connection term	Sum
Radial	16/125	−16/125	0
Angular	24/625	−24/625	0

Think about standing beside that rail with two different measuring devices. One reports east and north. The other keeps turning to face the trolley and reports range and angle. The second device's numbers change in a more complicated way. The connection terms account for that changing description. They do not supply a hidden force that pushes the trolley sideways.

Einstein derives the geodesic equation in §9 from a stationary path integral.

Gleichung der geodätischen Linie (bzw. der Punktbewegung). — Equation of the geodesic line (or of the motion of a point).

Einstein, 1916, §9, p. 790. German excerpt / English translation.

His variational statement is

$$\delta \int_{P_1}^{P_2} ds = 0.$$

Here is the calculation in a form that keeps its assumptions visible. A freely falling clock's proper time is an example of an affine parameter. Rescaling such a parameter by a constant and shifting its zero preserves that status; an arbitrary nonlinear re-timing generally adds a term to the equation below. With an affine parameter, use

$$L = \frac{1}{2} g_{ab} \dot{x}^a \dot{x}^b. \tag{11.15}$$

The Euler–Lagrange equations give

$$g_{ab} \ddot{x}^b + \partial_c g_{ab} \dot{x}^c \dot{x}^b - \frac{1}{2} \partial_a g_{bc} \dot{x}^b \dot{x}^c = 0. \tag{11.16}$$

The product $\dot{x}^b \dot{x}^c$ is symmetric. Symmetrize its coefficient and multiply by the inverse metric. The result is

$$\ddot{x}^a + \Gamma^a{}_{bc} \dot{x}^b \dot{x}^c = 0. \tag{11.17}$$

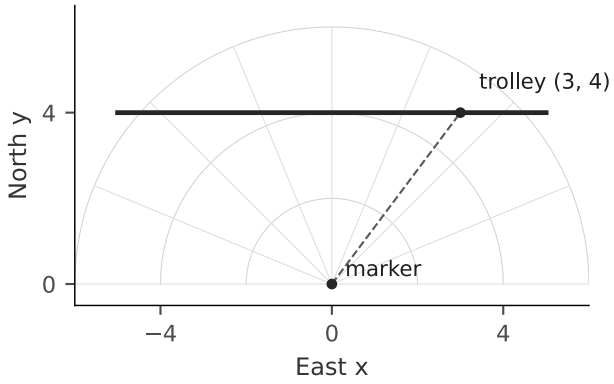
Fixed endpoints remove the boundary term in the variation. For a timelike physical trajectory, proper time provides an affine parameter. A null trajectory has zero proper time, so it needs another affine parameter. Calling every geodesic the "shortest path" would hide these differences.[4]

[4]The original German text on p. 791 explicitly excludes $ds = 0$ when replacing the parameter by arc length. The working English web transcription loses that exclusion. The modern quadratic Lagrangian used here also accommodates null geodesics with an affine parameter and a separately imposed null condition.

The rail stays straight

EINSTEIN 1916, SECTIONS 9-11 / CONSTRUCTED COORDINATE STUDY

01 One track on a flat floor



$x = \lambda, \quad y = 4$

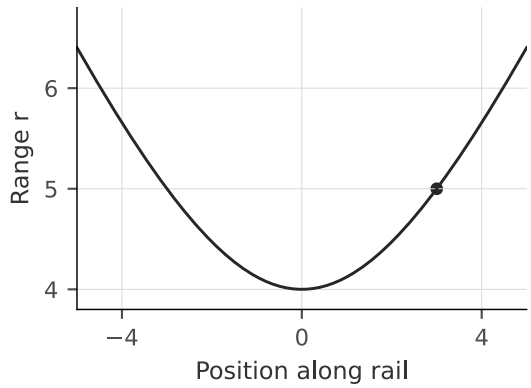
$r^2 = \lambda^2 + 16$

At the marked point:

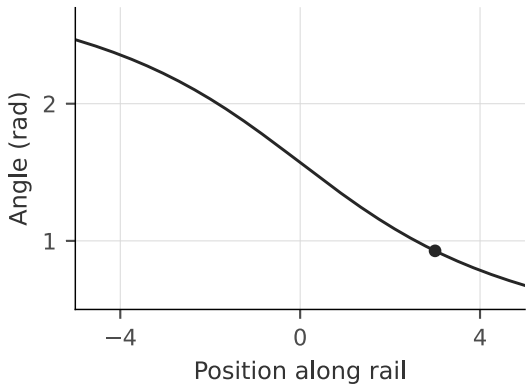
$\lambda = 3, \quad r = 5$

No sideways force.

02 Range changes nonlinearly



03 Angle changes nonlinearly



Changing coordinate rates do not establish a curved path.

At the marked point, the residuals are zero: 16/125 - 16/125 radially, and 24/625 - 24/625 angularly.

Figure 38. One straight rail, two coordinate descriptions. The range and angle change nonlinearly along the track. Their connection terms restore the zero geodesic residual. All plotted curves come from $r = \sqrt{\lambda^2 + 16}$ and $\theta = \operatorname{atan2}(4, \lambda)$.

§ 11.3

What a derivative has to remember

A derivative measures change, but the components of a vector can change because the vector changes or because the basis changes. Einstein's §§10–11 keep those possibilities separate by constructing the covariant derivative.

For a covector α_b, the rule is

die „Erweiterung“ des Tensors the "extension" of the tensor

Einstein, 1916, §10, p. 793. German excerpt / English translation.

$$\nabla_a \alpha_b = \partial_a \alpha_b - \Gamma^c{}_{ab} \alpha_c. \tag{11.18}$$

For a vector, the connection term has the other sign:

$$\nabla_a V^b = \partial_a V^b + \Gamma^b{}_{ac} V^c. \tag{11.19}$$

Each additional tensor slot gets its own term, with its sign determined by the slot's type. This gives us an explicit rule for constructing a calculation. It does not require guessing which components look related. Einstein's word *Erweiterung* names this derivative construction: it adds an index while preserving the tensor transformation rule. Here we use the modern term *covariant derivative*.

Take the ordinary east coordinate as a scalar field on the floor, $f = x = r\cos\theta$. In Cartesian coordinates its Hessian, the matrix of second derivatives, is zero. In polar coordinates, $\partial_\theta^2 f = -r\cos\theta$. Does the same floor suddenly have a nonzero second derivative of a linear function?

The ordinary second derivatives alone are not the Hessian tensor in these coordinates. The full expression is

$$(\nabla df)_{ab} = \partial_a \partial_b f - \Gamma^c{}_{ab} \partial_c f. \tag{11.20}$$

At $r = 2$, $\theta = 0$, its angular entry is

$$-2 - (-2)(1) = 0. \tag{11.21}$$

The apparent change cancels the change of basis. In GIGI we store the ordinary derivative term and the connection term separately, so the reason for that zero remains inspectable.

The volume measure creates a related issue. In polar coordinates an area element is $r\,dr\,d\theta$, not simply $dr\,d\theta$. In four dimensions the invariant volume element is

$$d\mathcal{V} = \sqrt{|\det g|}\, d^4x. \tag{11.22}$$

Under a coordinate change $X = X(q)$, the metric becomes $g_q = J^T g_X J$. Taking determinants gives $\det g_q = (\det J)^2 \det g_X$. The square root contributes $|\det J|$, precisely the factor needed to match the coordinate volume transformation. This is why the measure belongs beside the density in the record.

For a vector field, the corresponding divergence is

Divergenz des kontravarianten Vierervektors. Divergence of the contravariant four-vector.

Einstein, 1916, §11, p. 797. German excerpt / English translation.

$$\nabla_a V^a = \frac{1}{\sqrt{|\det g|}} \partial_a \left(\sqrt{|\det g|} V^a \right). \tag{11.23}$$

Consider an outward radial field on the floor with $V^r = r$ and $V^\theta = 0$. Its divergence is

$$\frac{1}{r} \frac{d}{dr}(r^2) = 2. \tag{11.24}$$

Differentiating the radial component by itself gives one. It misses the fact that the circumference through which the field passes grows with radius. Imagine water flowing outward across circles drawn on the floor: the surrounding boundary is getting longer. The measure keeps that geometry in the calculation.

For GIGI, a derivative record therefore needs the function, chart, differentiation rule, tensor slots, evaluation location, and domain. Our current examples use exact analytic differentiation in the companion Python program. GIGI receives those terms with their types and carries out their grouped sums. A finite-difference approximation would need its stencil, spacing, boundary treatment, and error estimate. It would be a different representation of the derivative, not an interchangeable column of numbers.

§ 11.4

Curvature is the test that survives the coordinates

Nonzero connection coefficients were necessary even on our flat floor. We need another object to distinguish a changing coordinate basis from curvature that cannot be removed by a change of coordinates. That object is the Riemann tensor, introduced in §12.

With our convention it is

Riemann-Christoffelscher Tensor Riemann–Christoffel tensor

Einstein, 1916, §12, p. 800. German excerpt / English translation.

$$R^a{}_{bcd} = \partial_c \Gamma^a{}_{db} - \partial_d \Gamma^a{}_{cb} \tag{11.25}$$

$$+ \Gamma^a{}_{ce} \Gamma^e{}_{db} - \Gamma^a{}_{de} \Gamma^e{}_{cb}. \tag{11.26}$$

Both derivative terms and product terms are necessary. For the polar example, one component illustrates the cancellation:

$$R^r{}_{\theta r \theta} = -1 + 1 = 0. \tag{11.27}$$

All components vanish. The full symbolic calculation establishes this throughout the chart; the GIGI replay also evaluates and sums every component at the declared test point.

Now consider an accelerating coordinate system in flat spacetime:

$$ds^2 = (1 + ax)^2 dt^2 - dx^2 - dy^2 - dz^2, \qquad 1 + ax > 0. \tag{11.28}$$

The parameter $a > 0$ has inverse-length units when $c = 1$. This is the Rindler description, a modern worked example of the distinction Einstein discusses in §§2–4.[5] It is flat because the transformation

$$T = (x + a^{-1}) \sinh(at), \tag{11.29}$$

$$X = (x + a^{-1}) \cosh(at) \tag{11.30}$$

turns its time–space part into $dT^2 - dX^2$. The transverse coordinates are unchanged. This chart covers a wedge, not all of Minkowski spacetime.[6]

Its connection is nonzero, and stationary coordinate clocks have different proper-time rates. Nevertheless, its Riemann tensor is zero. Putting $a = 1/4$ and $x = 2$ gives $\Gamma^t{}_{tx} = 1/6$ and $\Gamma^x{}_{tt} = 3/8$. We must not infer curvature merely from either coefficient.

For a genuinely curved comparison, use the Schwarzschild exterior:

$$ds^2 = f(r)dt^2 - f(r)^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{11.31}$$

$$f(r) = 1 - \frac{2M}{r}. \tag{11.32}$$

Here $M = GM_{\text{physical}}/c^2$ is a length, and this chart is restricted to $r > 2M$. This exact solution is an additional worked example; it is not the same expression as Einstein's weak-field approximation in §22.[7]

The Ricci tensor contracts the full curvature: $R_{bd} = R^a{}_{bad}$. For this exterior solution it vanishes. The full Riemann tensor does not. A scalar that exposes the difference is

$$\mathcal{K} = R_{abcd}R^{abcd} = \frac{48M^2}{r^6}. \tag{11.33}$$

At $M = 1$, $r = 10$, it is $3/62500$ in the corresponding inverse-fourth- length units. The result is positive even though every Ricci component is zero.

Description	Connection	Ricci	$\mathcal{K}$
Flat cylindrical chart	Nonzero	Zero	0
Accelerated flat chart	Nonzero	Zero	0
Schwarzschild exterior	Nonzero	Zero	$48M^2/r^6$

This is more than a vocabulary distinction. An implementation that stores only the Ricci scalar cannot answer every curvature question. An implementation that labels every nonzero connection entry "curved" will misclassify the flat examples. The representation has to preserve the object needed for the particular question.

[5]Wolfgang Rindler, "Kruskal Space and the Uniformly Accelerated Frame," *American Journal of Physics* 34 (1966), 1174–1178, DOI. The shifted radial coordinate used here gives the displayed form of the flat metric.

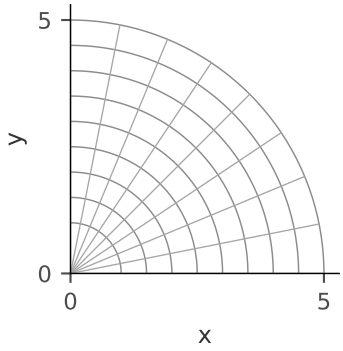
[6]For modern coordinate, curvature and symmetry constructions, see Sean M. Carroll, *Lecture Notes on General Relativity* (1997), chapters 3, 5 and 7, author manuscript. Carroll uses the opposite metric signature; the formulas here consistently use $(+,-,-,-)$ and the displayed Riemann convention.

[7]Karl Schwarzschild, "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie," *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (1916), 189–196. Einstein references Schwarzschild on p. 822 of his paper. The expression here uses the modern areal-radius coordinate; the historical and modern radial labels are not silently identified.

Which change is curvature?

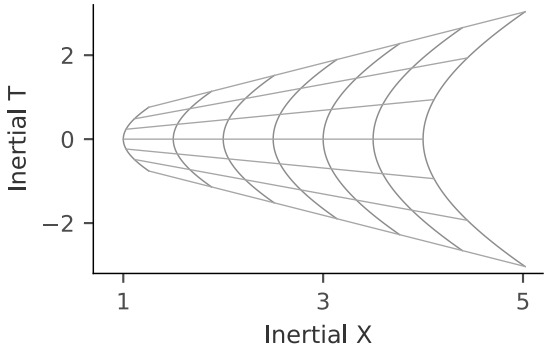
EINSTEIN 1916, SECTIONS 12-16 / THREE ANALYTIC METRICS

01 Polar grid: flat spacetime



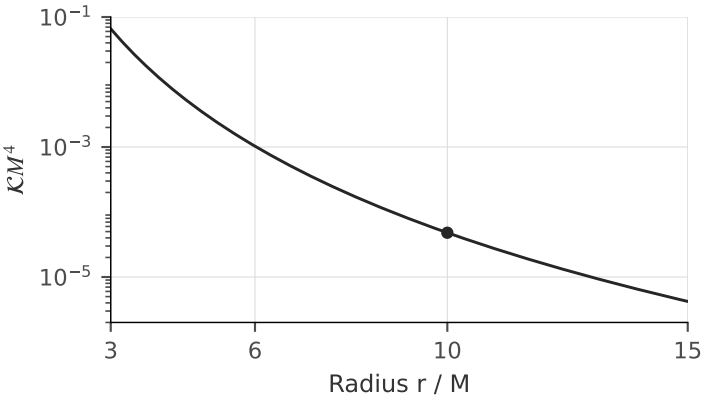
All Riemann components are zero.

02 Accelerated chart: also flat



All Riemann components are zero.

03 Schwarzschild: curved vacuum



04 Keep the tests separate

$$R_{ab} = 0$$

The vacuum Ricci tensor vanishes.

$$\mathcal{K} = 48M^2/r^6$$

At r = 10M:

$$\mathcal{K}M^4 = 3/62500$$

The full curvature remains.

A nonzero connection is not enough; a zero Ricci tensor is not enough.

Each chart has a declared domain. The exterior example uses r > 2M; the plot shows its dimensionless curvature invariant.

Figure 39. Changing coefficients and genuine curvature are different tests. Two descriptions of flat spacetime in non-Cartesian coordinates are compared with the exact curved Schwarzschild vacuum. The Schwarzschild panel plots $\mathcal{K}M^4 = 48/(r/M)^6$.

§ 11.5

From geometry to a physical field

Up to this point we have explained how to represent geometric objects and compare their descriptions. That does not yet tell us which metric describes a physical gravitational field. General covariance constrains the form of a law. It does not select all of physics by itself.

Einstein says as much when introducing matter in §16. In modern notation, with no cosmological term and with our curvature convention, the field equation is

Allgemeine Fassung der Feldgleichungen der Gravitation.	General formulation of the gravitational field equations.

Einstein, 1916, §16, p. 807. German excerpt / English translation.

$$G_{ab} \equiv R_{ab} - \frac{1}{2} R g_{ab} = \frac{8\pi G}{c^4} T_{ab}. \tag{11.34}$$

The scalar $R = g^{ab} R_{ab}$ is the trace of the Ricci tensor. The stress–energy tensor T_{ab} represents matter and nongravitational fields. The coefficient uses energy-density units for T_{ab}; it should not be copied into a source formula using a differently scaled matter density.

To test equation (11.34), we need both sides, their units, the metric's first and second derivatives, and the region over which the test is claimed. Our Schwarzschild example has $R_{ab} = 0$ and $T_{ab} = 0$ in its exterior domain. We check all of its vacuum residuals symbolically and then evaluate the contractions in GIGI. This verifies an exact example. It does not constitute a general solver for arbitrary matter distributions.

The action formulation in §15 adds another kind of object to the representation: a functional. Its input is an entire field over a region, not one row at a point. Varying the field produces equations only after specifying what can vary and what is fixed at the boundary. Einstein uses a specialized coordinate condition, $\sqrt{-g} = 1$, in much of the calculation. That condition must travel with the specialized formulas. The polar metric in our first example does not satisfy it. Replacing a general formula by its simplified version without checking that condition would change the problem.

The matter examples make the same point in a more familiar setting. For a perfect fluid in units $c = 1$,

der kontravariante Energietensor der Flüssigkeit.	the contravariant energy tensor of the fluid.

Einstein, 1916, §19, p. 811. German excerpt / English translation.

$$T^{ab} = (\varepsilon + p) U^a U^b - p g^{ab}, \qquad g_{ab} U^a U^b = 1. \tag{11.35}$$

Here ε is the energy density measured with the fluid, p is its pressure, and U is its four-velocity. In a local rest frame, $U = (1, 0, 0, 0)$, so the diagonal entries are (ε, p, p, p). With $\varepsilon = 3$ and $p = 1$, the example returns $(3, 1, 1, 1)$ and trace $\varepsilon - 3p = 0$.

Pressure has not disappeared into an annotation. It is one of the components that enters the gravitational source. Likewise, the coefficient multiplying $U^a U^b$ is the enthalpy density $\varepsilon + p$, not simply rest-mass density. Einstein's §19 writes that coefficient with a density symbol. A modern representation must map its definition as well as its name. An equation of state and initial or

boundary data are still needed to pose a fluid evolution problem; counting component equations alone does not supply them.

For electromagnetism, §20 starts with a potential covector and forms an antisymmetric field tensor. Our modern sign convention is

des Vierervektors des elektromagnetischen Potentials.
Einstein, 1916, §20, p. 812. German excerpt / English translation.

of the four-vector of the electromagnetic potential.

$$F_{ab} = \partial_a A_b - \partial_b A_a. \tag{11.36}$$

Commuting partial derivatives gives the cyclic identity

$$\partial_a F_{bc} + \partial_b F_{ca} + \partial_c F_{ab} = 0. \tag{11.37}$$

This is a proof for sufficiently smooth potentials. It does not need a separate numerical experiment for every component. The other Maxwell equations involve the divergence and the source current, so the volume factor returns.

Take a prescribed flat cylindrical background, $r > 0$, with normalized potential $A_t = -\log r$ and the other components zero. Then $F_{tr} = 1/r$ and $F^{tr} = -1/r$. The time component of its vacuum divergence, here contracting the derivative with the second index of F^{ab}, is

$$\frac{1}{r}\frac{d}{dr}\left(rF^{tr}\right) = \frac{1}{r}\frac{d}{dr}(-1) = 0. \tag{11.38}$$

Differentiating F^{tr} alone gives $1/r^2$. At $r = 2$ the naive answer is $1/4$ while the correct divergence is zero. The line source at $r = 0$ is outside this patch; the example is not a source-free field on all of space. We also treat the metric as a fixed background, without solving for electromagnetic gravitational backreaction.

The replay evaluates its electromagnetic stress–energy tensor as well. With the electromagnetic unit factor normalized to one, the contravariant diagonal entries at $r = 2$ are

$$(T^{tt}, T^{rr}, T^{\theta\theta}, T^{zz}) = \left(\frac{1}{8}, -\frac{1}{8}, \frac{1}{32}, \frac{1}{8}\right). \tag{11.39}$$

The angular coordinate component carries the inverse metric factor. It should not be read as the physical tangential stress in an orthonormal frame. Contracting with the metric gives zero trace. The type, basis and measure are doing visible work in every line.

§ 11.6

What survives in the energy account

This is the point where I want to slow down again. An equation involving energy can look authoritative while leaving its physical interpretation unfinished. Which observer measured that energy? Which boundary did it cross? What was held fixed while the comparison was made?

In general relativity, the local matter balance is

Verschwinden der Divergenz des Tensors der Energiekomponenten der Materie.

Vanishing of the divergence of the tensor of the energy components of matter.

Einstein, 1916, §18, pp. 809–810. German excerpt / English translation.

$$\nabla_a T^{ab} = 0. \tag{11.40}$$

It is not generally the same as setting the ordinary partial derivatives of the components to zero. The connection terms describe how the balance is expressed in a changing basis. Einstein's §§15–18 also introduce gravitational energy components that are not a tensor. His text explicitly says so. We must not store those components as though they were a local energy-density tensor that every coordinate system would transform in the same way. His warning on p. 806 is direct: *„kein Tensor ist"*—"is not a tensor." It refers to the gravitational quantity $t^{\alpha}{}_{\sigma}$, not to the matter tensor T in the displayed equation.

A useful conserved energy is available when the spacetime has a time-translation symmetry. Let ξ^a describe that symmetry; the Killing equation is

$$\nabla_a \xi_b + \nabla_b \xi_a = 0. \tag{11.41}$$

Define $J^a = T^{ab}\xi_b$. For symmetric T^{ab},

$$\nabla_a J^a = (\nabla_a T^{ab})\xi_b + T^{ab}\nabla_a \xi_b \tag{11.42}$$

$$= \frac{1}{2} T^{ab}(\nabla_a \xi_b + \nabla_b \xi_a) = 0. \tag{11.43}$$

The symmetry supplies the energy notion; the divergence theorem then relates changes inside a region to flux through its boundary, with the correct geometric measure. An arbitrary time-dependent spacetime need not supply this particular conserved energy. This is a modern extension of the bookkeeping question, not a quotation of Einstein's 1916 notation.[8]

For a static metric with $g_{tt} = N^2$, a stationary clock measures $d\tau = N\,dt$. The positive factor N is called the lapse in this static, zero-shift description. A photon traveling along a null geodesic has conserved $p_a\xi^a$; its locally measured energy for a stationary observer is that quantity divided by N.

Use two stationary locations with lapse factors $N_A = 3/5$ and $N_B = 4/5$. Give the photon local energy one at A. Its conserved time-symmetry energy is 3/5. At B the local energy is

$$E_B = \frac{N_A}{N_B} E_A = \frac{3}{4}. \tag{11.44}$$

Multiplying by B's lapse returns $N_B E_B = 3/5$. The two local measurements differ, while the conserved quantity remains the same. These are constructed lapse values illustrating a stationary comparison, not new observations of gravitational redshift.

If we want to build a cycle around that comparison, we must now specify the equipment. What holds the emitters at their locations? What returns the receiver or working material to its starting state? Where does the work cross the boundary? A change in one local energy reading is not the completed cycle. The representation needs room for the rest of the account.

[8] Carroll, *Lecture Notes on General Relativity*, chapter 5, discusses Killing fields and the corresponding conserved quantities. The current proof assumes a symmetric, covariantly conserved stress–energy tensor and a Killing field; the existence of those ingredients is part of the claim.

Einstein closes the paper by connecting the theory to approximate observations. In the weak, nearly static field and slow-motion regime, $g_{tt} \simeq 1 + 2\Phi/c^2$ gives the Newtonian acceleration $\ddot{\mathbf{x}} \simeq -\nabla\Phi$, and the field equation reduces to Poisson's equation. Each approximation is part of the statement. An implementation must retain which powers of field strength and velocity were neglected.

His final section treats clock rates, light bending and perihelion advance. The familiar leading light-bending result is $4GM/(bc^2)$ for impact parameter b, and the perihelion advance per orbit is $6\pi GM/[a(1-e^2)c^2]$ in the relevant weak-field regime. The symbols a and e now mean the orbital semimajor axis and eccentricity, not the acceleration parameter in our earlier example. These predictions require their own physical inputs and comparison with observations. Our current GIGI tensor checks do not replace that empirical work.

For Fuel, the data requirements become concrete: event and observer for measurements; domain and derivatives for fields; transformation type for tensors; measure and boundary for integrals; a full physical account for energy claims. Storing Einstein's equations does not make them true. They show GIGI what a representation must carry to answer the question we actually asked.

Chapter 12

Bring the geometry back to the physics

In the floor example, the marked point on the mop head returned to where it started. Its final position matched its starting position, but the record still contained the trip out and back. That was the distinction I wanted to carry back into the lattice. If an average moved, what had moved with it—the fields, the weights, or the question? If a gap appeared, what had been separated—two plotted measurements, two descriptions, or two allowed energies? I wanted to follow the geometry far enough that I could answer those questions without changing their meaning halfway through.

Einstein's clocks sharpened that demand. A reading needed its clock and its event. An energy comparison needed the observer and the physical account. In the floor studies, a distance needed a direction before it became an instruction, and an endpoint needed a history before it described a journey. Those are practical requirements for the two investigations that brought me here: topological vacuum rectification, or TVR, and the Yang–Mills mass gap.

There are two different ambitions in those investigations. TVR asks whether a specified change in geometric structure can produce a directed response, and ultimately whether a physical implementation can direct energy transfer. The mass-gap question asks whether a quantum theory has a positive lower bound on its nonvacuum excitation energies. A useful geometric representation should help us ask each question more precisely. I need to know what it preserves on the way to the answer.

§ 12.1

Keep the records, change the weights

Start with the part we can reproduce from the saved TVR archive. Each of its 1,600 rows contains two numbers, J_i and D_i. For the matching measurement implementation, these are dimensionless loop statistics. They are not meter readings in joules or watts. The historical source revision and coupling normalization still need the provenance checks described in the TVR chapter. We can reproduce the arithmetic on the retained rows without pretending that this settles the physical calibration.

The calculation standardizes D: subtract its archive mean and divide by its archive standard deviation. Write the result as $z_i = (D_i - \bar{D})/s_D$. This gives us a dimensionless scale on which one unit means one archive standard deviation. Keep that scale fixed while changing the bias setting θ. The unnormalized weight of row i is $w_i = e^{-\theta z_i}$. To turn those weights into probabilities, divide

by their sum:

$$Z(\theta) = \sum_{i=1}^{N} e^{-\theta z_i}, \qquad p_i(\theta) = \frac{e^{-\theta z_i}}{Z(\theta)}, \qquad \widehat{J}(\theta) = \sum_i p_i(\theta) J_i. \tag{12.1}$$

This is a finite exponential tilt. Reweighting has a long history in Monte Carlo physics: Ferrenberg and Swendsen used recorded distributions to estimate behavior at nearby parameter values. Here we apply the finite-sum operation to the retained rows, without assuming that our bias parameter is a calibrated physical coupling.[1]

Here Z is a normalization sum. Calling it Z does not by itself make it a physically calibrated thermodynamic partition function. At zero bias, all N rows receive probability $1/N$. Negative bias gives more influence to larger z; positive bias gives more influence to smaller z.

We can derive exactly how the answer moves. Define $\widehat{z} = \sum_i p_i z_i$. Since the saved J_i and z_i stay fixed,

$$Z' = -\sum_i z_i e^{-\theta z_i} = -Z\widehat{z}, \tag{12.2}$$

$$p_i' = p_i(-z_i + \widehat{z}), \tag{12.3}$$

$$\widehat{J}' = -\sum_i p_i J_i z_i + \widehat{J}\widehat{z} = -\operatorname{Cov}_\theta(J, z). \tag{12.4}$$

The covariance is the weighted average of $(J_i - \widehat{J})(z_i - \widehat{z})$. It is positive when high values of one tend to accompany high values of the other. Equation (12.4) tells us that this association determines the slope of the weighted answer. The conclusion depends on holding the records and standardization fixed; an observable or normalization that itself changes with θ would add derivative terms.

There is a related curvature calculation. Differentiating again gives

$$\frac{d\log Z}{d\theta} = -\widehat{z}, \qquad \frac{d^2\log Z}{d\theta^2} = \sum_i p_i(z_i - \widehat{z})^2 \geq 0. \tag{12.5}$$

This is curvature of the log-normalization as a function of the bias. It measures the spread of the tilted z values. It is not spacetime curvature, and it does not acquire energy units because we used the same mathematical word. The formula tells us exactly which curvature we have calculated.[2]

Take four records so we can see the entire mechanism without a large dataset. Let

$$\begin{aligned}(J_1, J_2, J_3, J_4) &= (-2, -1, 1, 2),\\ (z_1, z_2, z_3, z_4) &= (-1, -1, 1, 1).\end{aligned} \tag{12.6}$$

[1] A. M. Ferrenberg and R. H. Swendsen, "New Monte Carlo Technique for Studying Phase Transitions," *Physical Review Letters* 61 (1988), 2635–2638, DOI. A physical reweighting prediction also requires adequate overlap between the sampled and target measures.

[2] M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," *Foundations and Trends in Machine Learning* 1 (2008), 1–305, Proposition 3.1, DOI. The gradient and Hessian of a log-normalization give means and covariances. Our minus sign comes from using $-\theta z$.

At zero bias the mean is zero. Choose $\theta = -\log 2$. The first two weights are $1/2$ and the last two are 2. Their sum is 5, so the probabilities are $(1/10, 1/10, 2/5, 2/5)$. The returned sum is

$$\widehat{J} = -\frac{2}{10} - \frac{1}{10} + \frac{2}{5} + \frac{4}{5} = \frac{9}{10}. \tag{12.7}$$

The weighted mean of z is $3/5$, and the weighted mean of Jz is $3/2$. Therefore

$$\mathrm{Cov}_\theta(J, z) = \frac{3}{2} - \frac{9}{10}\frac{3}{5} = \frac{24}{25}. \tag{12.8}$$

The slope there is $-24/25$. At zero bias it is $-3/2$. The response changes as the weights change because the association is being evaluated under a different distribution.

Think of four observations laid out on the counter. I have not changed anything written on them. I have changed how much each observation counts toward the answer. The mean moves from zero to 0.9, and we can account for every part of that movement. That gives us a precise statistical response to investigate. To claim that a physical system follows the same response, I now need a physical operation that actually changes its distribution in the specified way.

For the archive replay I evaluated five settings, keeping the same 1,600 rows and standardization throughout. Python forms the exponential weights and products. GIGI sums their integer-scaled columns; the companion calculation divides the sums to recover the means, covariances and weight-concentration counts. Independent array calculations agree within 10^{-7} for the displayed means and covariances. A centered finite-difference calculation of the mean also agrees with minus the covariance within 10^{-6} at each setting. These tolerances concern numerical agreement, not statistical uncertainty.

θ	Mean $\widehat{J}$	Slope $-\mathrm{Cov}(J, z)$	Weight ESS
-1.15	27.428514	-21.360091	493.44
-0.50	13.017493	-22.778113	1254.67
+0.00	1.458350	-23.462063	1600.00
+0.50	-10.462513	-24.220578	1237.95
+1.15	-26.374013	-24.437819	424.63

The weight-only effective sample size is

$$N_{\mathrm{eff}} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}.$$

Equal weights give 1,600. Concentrating the weight on fewer rows lowers it. This count does not correct for correlation between successive lattice samples. It is a weight-concentration diagnostic, not an observable-specific variance guarantee.[3] The earlier paired block analysis supplies a separate sensitivity check; the five rows here are evaluations of one archive, not five independent experiments.[4]

[3] V. Elvira, L. Martino and C. P. Robert, "Rethinking the Effective Sample Size," *International Statistical Review* (2022), DOI; author manuscript. The paper examines the assumptions behind this commonly used importance-sampling approximation.

[4] The input archive's byte identity is checked against the retained TVR receipt. The current runner is `research/einstein_gigi/bridge_replay.py`; queries, output, CSV and hash receipt are in `results_bridge`. It checks 52

The selected -1.15 setting reproduces the earlier mean near 27.42851. At zero bias the slope is about -23.46206. That number is also the linear-regression coefficient already encountered in the archive analysis, with the opposite sign: z has population variance one, so its regression coefficient is $\mathrm{Cov}(J, z)$. The two calculations fit together. They tell us how the existing association in the records enters the response.

We can make that agreement more explicit. Fit the archived values as $J_i = \bar{J} + bz_i + e_i$, where e_i is the residual left after subtracting the fitted line. Because the unweighted z values have mean zero and variance one, ordinary least squares gives $b = \mathrm{Cov}_0(J, z)$ and $\bar{e} = 0$. Weighted averaging then gives an identity, with no approximation:

$$\widehat{J}(\theta) - \bar{J} = b\widehat{z}(\theta) + \widehat{e}(\theta). \tag{12.9}$$

At the archived -1.15 setting, the earlier decomposition gives

$$25.97016322 = 26.02668384 - 0.05652061, \tag{12.10}$$

up to the displayed rounding. The first term on the right comes from the fitted association; the second is the weighted residual. The large change is almost entirely accounted for by that association. This does not turn the fitted line into a physical mechanism. It tells us which part of the saved information is doing the work in this statistical calculation.

The derivative also explains why a straight-line prediction from zero bias need not remain exact at a finite setting. Integrating Equation (12.4) yields

$$\widehat{J}(\theta) - \widehat{J}(0) = -\int_0^\theta \mathrm{Cov}_s(J, z)\, ds. \tag{12.11}$$

The subscript s means that the covariance is evaluated at each intermediate bias. We cannot replace the entire integral by its initial slope without an approximation. The slope column records that changing response. It gives us something specific to compare with an independently generated ensemble, once the modified measure and observable have been matched. Figure 40 puts the small construction and the archive response side by side so that the weighting operation remains visible in both.

§ 12.2

A return must include the energy

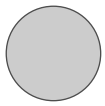
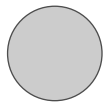
Now suppose I want to turn that statistical experiment into an energy experiment. I need to name the physical control. A coefficient in an offline weight is not yet a knob connected to a field or a piece of equipment. The next comparison would generate an independently sampled ensemble under a declared modified action and compare its observables with the reweighted prediction, with the normalizations and sampling assumptions matched. Passing that comparison would test the statistical model. A work claim still needs dynamics and an energy measurement.

numerical sums across the archive and constructed examples in one fresh GIGI process. No new lattice ensemble is generated.

Same records, a different average

TVR BRIDGE / EXACT FOUR-ROW EXAMPLE AND THE RETAINED ARCHIVE

Change how much each observation counts

Recorded J	-2	-1	1	2
Bias label z	-1	-1	1	1
Equal weights Mean = 0.0	 0.25	 0.25	 0.25	 0.25
Tilt toward +z Mean = 0.9	 0.10	 0.10	 0.40	 0.40

Circle area represents probability. The four J values never change.

ARCHIVE BIAS	MEAN J	SLOPE	WEIGHT ESS
-1.15	27.42851	-21.36009	493.44
-0.50	13.01749	-22.77811	1254.67
+0.00	1.45835	-23.46206	1600.00
+0.50	-10.46251	-24.22058	1237.95
+1.15	-26.37401	-24.43782	424.63

A change in the weighted mean identifies a response of the statistic.

Bottom: 1,600 retained TVR rows, five settings. Slopes are minus covariance. ESS measures weight concentration.

Figure 40. The four-row construction exposes the weighting operation. The lower table uses the retained TVR archive. GIGI supplies the sum totals; the companion calculation forms ratios and covariances.

Let $\lambda(t)$ denote the externally controlled settings of a physical system, and let $H(\lambda)$ be its Hamiltonian, the operator that assigns energy. A density matrix $\rho(t)$ describes its quantum state, including statistical mixtures. Its mean energy is $E(t) = \mathrm{Tr}(\rho H)$. For differentiable finite-dimensional models, the product rule gives

$$\frac{dE}{dt} = \mathrm{Tr}(\dot{\rho} H) + \mathrm{Tr}(\rho \dot{H}). \tag{12.12}$$

The dot means a time derivative. The second term accounts for work supplied by changing the Hamiltonian. For a closed system with unitary evolution, $\dot{\rho} = -i[H, \rho]/\hbar$, the first term vanishes by cyclicity of the trace. Thus

$$W_{\mathrm{on}} = \int_0^T \mathrm{Tr}(\rho(t)\dot{H}(t))\, dt = E(T) - E(0). \tag{12.13}$$

Here positive W_{on} means work delivered to the system. For an open system, changes in ρ can also account for energy exchange with other systems. Identifying that exchange as heat, work or stored interaction energy requires the coupling and boundary model; the product rule alone does not decide it.

This is the mop distinction in another language. Returning the knob to its original setting gives $H(T) = H(0)$. It does not imply $\rho(T) = \rho(0)$. The system may have been excited. A reservoir may have cooled. A load may have received energy. The meter needs to stay on while those changes are accounted for.

There is a short exact result for the ground-state case. Let the initial Hamiltonian be H_0 with minimum energy E_0. Begin in its ground-state sector and drive a closed system through a cycle that returns the Hamiltonian to H_0. If the final evolution is unitary, its final mean energy cannot fall below E_0. Consequently

$$W_{\mathrm{on}} = \mathrm{Tr}(\rho(T)H_0) - E_0 \geq 0, \qquad W_{\mathrm{extracted}} = -W_{\mathrm{on}} \leq 0. \tag{12.14}$$

This proof allows a degenerate ground-state sector. It uses the minimum-energy property, not a special choice of coordinates. The broader theory of passivity examines work extraction under cyclic operations; Pusz and Woronowicz establish the role of ground and thermal equilibrium states in that framework.[5]

For a proposed TVR device, that directs the investigation toward a specified drive or nonequilibrium resource. Geometry might help determine which transitions occur or how a supplied resource reaches a load. To test that possibility I need the field preparation, the controller, the load and the reset in one account. The familiar cycle ledger then has a concrete job: every joule must have a place to come from, a place to go, or a recorded change in storage.

The same care applies to direction. Reversing a unitary loop matrix changes W to $W^\dagger$, so $\mathrm{Im}\,\mathrm{Tr}(W^\dagger) = -\,\mathrm{Im}\,\mathrm{Tr}\,W$. A sign reversal of that observable can follow from reversing the contour convention. That is different from a load receiving power in a fixed laboratory setup. For inserted or transported loop observables, the entire definition must be transformed consistently; one sign

[5]W. Pusz and S. L. Woronowicz, "Passive States and KMS States for General Quantum Systems," *Communications in Mathematical Physics* 58 (1978), 273–290, doi:10.1007/BF01614224. The elementary finite-dimensional argument here does not reproduce their general operator-algebraic theorem.

rule should not be borrowed for a different observable. GIGI's records need to retain that definition alongside the sign.

§ 12.3

A gap must belong to an operator

The mass-gap investigation asks for another kind of completeness. The saved lattice diagnostic orders radii in a chosen feature representation and measures an unusually large spacing. Its value 85.46 belongs to the retained 200-radius run; the historical 41.97 belongs to a different record. In the retained run, both ends of the largest interval carry the rounded label $Q = 0$, and removing the outermost point changes the ratio to about 23.18. Those facts help us ask what the plot is detecting. They do not assign energies to its radial coordinates.[6]

For an energy question, begin with H and its lowest energy E_0. Subtract E_0 to define the excitation-energy operator $K = H - E_0 I$.

This is where my distinction between an absent system and a present seed helps me keep the question straight. A normalized vacuum state is a state of the theory; assigning it zero excitation energy does not remove it from the state space. Zero is the result of a specified comparison. For this spectral question, shifting every energy by the same constant c leaves that comparison unchanged:

$$(H + cI) - (E_0 + c)I = H - E_0 I. \tag{12.15}$$

This does not identify the path seed in my counting construction with a quantum vacuum. It identifies the distinction I need to preserve: a state can be present while one of its measured quantities is zero. Nor does the choice of an energy reference supply a source of extractable work. The excitation differences are the same before and after the shift.

Let P_0 project a state onto the full ground-state sector. A positive gap Δ is expressed by

$$K \geq \Delta(I - P_0), \qquad \Delta > 0. \tag{12.16}$$

An operator inequality means that the corresponding inequality holds for the quadratic forms: for every allowed state ψ, $\langle \psi, K\psi \rangle \geq \Delta \|(I - P_0)\psi\|^2$. In an infinite-dimensional theory the form domain must also be specified. For a normalized state entirely orthogonal to the vacuum sector, the right-hand side is Δ.

That last condition is doing real work. A state can be mostly vacuum with a small excited component. Its mean excitation energy can then be smaller than the gap, even though no energy eigenvalue lies in the forbidden interval. The gap concerns the spectrum and the fully nonvacuum sector; it does not say that every state different from the vacuum has mean energy at least Δ.

[6]The retained exact replay is `docs/book/geometry_of_fuel/examples/lattice/lattice_gigi_receipt.json`. Its largest spacing is 1.896399544338242 and its median spacing is 0.02218942730382; the reported ratio includes the retained 10^{-10} denominator regularizer.

Here is a complete example. Choose a positive energy unit $E_* > 0$ and three orthonormal states e_0, e_1, e_2. In that basis let

$$\frac{H}{E_*} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12.17}$$

The allowed energies are $0, E_*/4, E_*$. For a normalized state $\psi = ae_0 + be_1 + ce_2$, its excitation energy is

$$\langle \psi, H\psi \rangle = E_* \left(\frac{|b|^2}{4} + |c|^2 \right) \geq \frac{E_*}{4}(|b|^2 + |c|^2). \tag{12.18}$$

This proves the gap bound with $\Delta = E_*/4$. Equality holds for e_1, so that bound is sharp. If the probability in e_1 is only $1/100$ and the remaining probability is in e_0, the mean energy is $E_*/400$. That does not fill the spectral gap. An energy measurement has outcomes 0 or $E_*/4$ in this example.

Now rotate only the two excited coordinates. Use the real orthogonal matrix U, whose transpose is its inverse:

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3/5 & -4/5 \\ 0 & 4/5 & 3/5 \end{pmatrix}, \qquad U^T U = I. \tag{12.19}$$

Coordinates change by $\psi' = U^T \psi$, while the energy matrix becomes

$$\frac{H'}{E_*} = U^T \frac{H}{E_*} U = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 73/100 & 9/25 \\ 0 & 9/25 & 13/25 \end{pmatrix}. \tag{12.20}$$

The off-diagonal entries couple the new coordinate components in the energy calculation. They are part of the representation. For every state, $\psi'^\dagger H' \psi' = \psi^\dagger H \psi$. The low-energy eigenvector becomes $v = (0, 3/5, -4/5)^T$. Direct multiplication gives $H'v = (E_*/4)v$. Its energy did not rise because we changed the labels.

Suppose, however, that we retain only the diagonal of H', treating the new components as independent. The resulting operator has energies $0, 0.52E_*, 0.73E_*$. Its apparent gap is $0.52E_*$. We have more than doubled the gap by deleting terms. Removing the original low-energy state altogether would instead leave energies $0, E_*$ and a gap of E_*. Both operations change the problem.

The characteristic polynomial makes the legitimate comparison exact. For x measured in units of E_*,

$$\det(xI - H/E_*) = \det(xI - H'/E_*) = x(x - 1/4)(x - 1). \tag{12.21}$$

The equality follows from orthogonal conjugation; expansion of the two-by-two block also verifies it. In the replay, GIGI sums the products defining all nine entries of H' and returns zero for the three components of $H'v - (E_*/4)v$, using the declared energy normalization. Python supplies the exact products and symbolic characteristic-polynomial check. The spectrum is established by that algebra, not by calling a sum an eigensolver.

This is what the connected-region example was preparing us to notice. A reduction must preserve the relationships needed by the question. For floor regions those relationships were crossings. For this energy model they include the coupling terms and the low-energy states. A

representation can look cleaner precisely because it has thrown away the part that would change the answer.

We can state the bridge a geometric proof would need. Suppose a nonnegative geometric quantity $\mathcal{G}(\psi)$ is defined on the physical states in the energy form domain, and suppose we prove two uniform bounds with positive constants A and B:

$$\langle \psi, K\psi \rangle \geq A\,\mathcal{G}(\psi), \qquad \mathcal{G}(\psi) \geq B\,\|(I - P_0)\psi\|^2. \tag{12.22}$$

Combining them establishes a spectral gap of at least AB. The units of A and B must make their product an energy. The first inequality links the chosen geometry to energy; the second prevents an entire nonvacuum direction from becoming invisible to that geometry. Both must hold for all the relevant states, not only the sampled configurations.

In the three-state example, choose $\mathcal{G}(\psi) = |b|^2 + |c|^2$. This is the squared distance to the ground-state subspace in the state-space norm. The displayed energy formula proves the first bound with $A = E_*/4$, and the second holds with $B = 1$. The rotation preserves that argument. For Yang–Mills, finding and proving corresponding bounds is the substantive task; giving the quantities geometric names does not establish the inequalities. This tells me what a useful geometric measurement would have to control. Figure 41 follows the energy levels through the rotation and deletion, keeping the changed operator visible.

§ 12.4

Preserve the physical question

The three-state example is deliberately small. It is not a lattice Yang–Mills Hamiltonian. Its purpose is to expose a requirement we can carry into a larger calculation: if we claim that a new geometric description preserves the energy question, we must give the map between states, the operator it acts on, and a proof that the relevant energies survive. An invertible change of labels and the removal of states are different operations.

For a specified finite lattice theory, a transfer operator provides one established route between Euclidean evolution and quantum energy. If the construction supplies a positive self-adjoint transfer operator of the form

$$T = \exp[-\tau(H - E_0 I)/\hbar], \tag{12.23}$$

then an excitation of energy Δ corresponds to an eigenvalue $\lambda_1 = e^{-\tau\Delta/\hbar}$ below the normalized ground value $\lambda_0 = 1$. Here τ is the Euclidean time step. When the first excited level exists as an isolated eigenvalue,

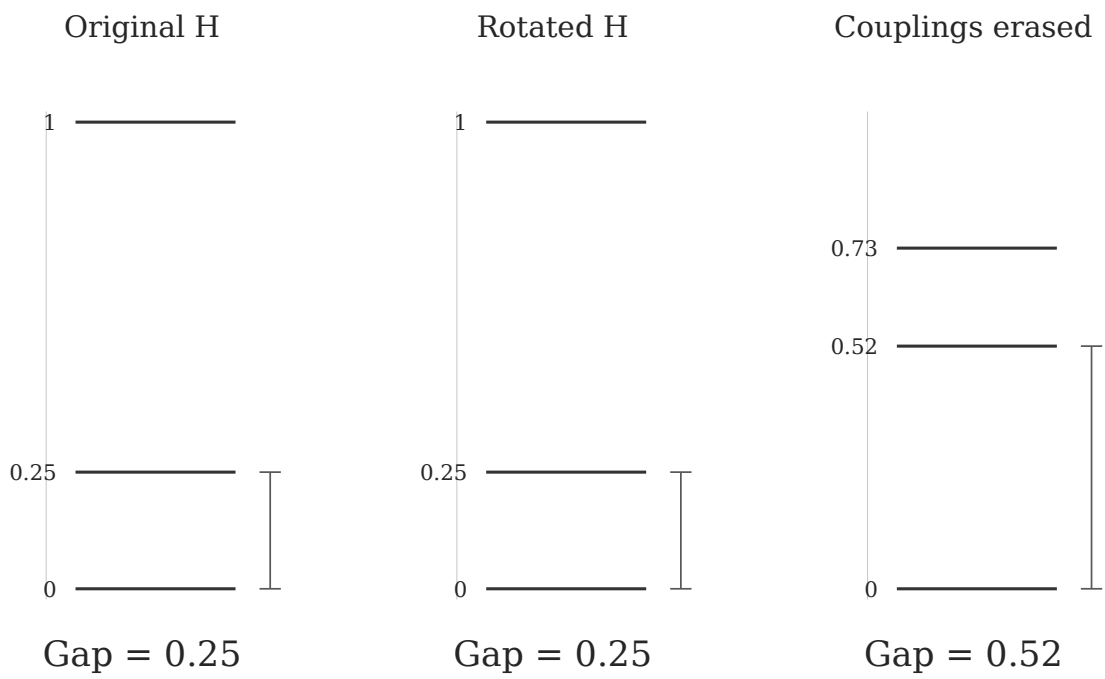
$$\Delta = -\frac{\hbar}{\tau}\log(\lambda_1/\lambda_0). \tag{12.24}$$

The operator and the time scale make this an energy statement. The corresponding spectral-bound formulation also covers a spectral edge that is not an isolated eigenvalue. Lüscher's construction

Keep the couplings, keep the gap

FINITE THREE-STATE MODEL / EXACT ALGEBRA AND VERIFIED GIGI SUMS

All energies are measured in one chosen unit E*.



THE ROTATED ENERGY MATRIX

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.73 & 0.36 \\ 0 & 0.36 & 0.52 \end{bmatrix}$$

THE LOW-ENERGY STATE IS STILL THERE

$$v = (0, 3/5, -4/5)$$

$$H_{\text{rotated}} v = (1/4) v$$

Deleting 0.36 changes the operator.

Changing coordinates preserves the spectrum. Dropping terms does not.

This model illustrates a proof requirement; it is not a Yang-Mills Hamiltonian or a measured physical spectrum.

Figure 41. The same finite energy operator in two bases, followed by a different operator obtained by deleting couplings. The original gap survives the coordinate rotation and fails the deletion.

establishes transfer-matrix machinery for Wilson lattice gauge theories under its assumptions. A modified action requires checking the needed assumptions again.[7]

A graph built from feature distances also has operators and spectra, but it does not become this physical transfer operator merely by having eigenvalues. Likewise, a Monte Carlo update describes how a sampler explores configurations; its relaxation rate is not automatically a physical excitation energy. Those are comparisons to derive, not names to exchange.

For the continuum mass-gap problem, even a correct finite operator is a beginning. The limiting quantum theory must exist with the required properties, and a positive physical gap must survive the removal of the lattice regulator and the infinite-volume limit. Jaffe and Witten identify these requirements and discuss possible roles for geometric variables and curvature in approaching them. That is a concrete reason to pursue the geometry while keeping the energy statement in view.[8]

The next GIGI records should therefore carry more than a configuration number and a score. For TVR, I need the observable's exact loop and orientation, the sampling measure, the bias convention, the physical control if one is implemented, and the energy account. For the gap, I need the state space, the energy operator, its units, the maps between representations and the scale of each approximation. The calculation should tell me when one of those ingredients changes.

There are immediate tests we can make at that level. An equivalent gauge description should preserve a declared gauge-invariant loop observable. A unitary basis change should preserve a finite energy spectrum. A proposed reduction should be challenged with the lowest energy modes, where dropping a state could create an attractive but false gap. Independent ensembles can test the reweighting prediction; a physical implementation would add the drive and full-cycle meters. These are next experiments, not results already supplied by the archive replay.

That brings me back to the floor without making the floor stand in for quantum field theory. One metre needed a direction. A return needed its journey. A group needed its connections. In the lattice, a response needs its measure, a gap needs its operator, and an energy claim needs the whole cycle. Keeping those relationships visible is how I want the geometry to do its work.

[7] M. Lüscher, "Construction of a Selfadjoint, Strictly Positive Transfer Matrix for Euclidean Lattice Gauge Theories," *Communications in Mathematical Physics* 54 (1977), 283–292, doi:10.1007/BF01614090. The manuscript is available through the DESY publication record.

[8] Arthur Jaffe and Edward Witten, *Quantum Yang–Mills Theory*, Clay Mathematics Institute problem description, sections 4 and 6.6, official problem statement. Their discussion of geometric possibilities is not a proof that any particular feature-space separation gives a mass gap.

Chapter 13

BYOB
Bring your own brilliance

"Unless you eat something from this world, you'll vanish."
Haku, Spirited Away *(2001), English subtitles*

Before we begin, I want to invite you into the work itself. You do not have to take my word for what a calculation returns. The companion repository brings together CHIHIRO's source, experiment programs, and saved results.[1] Start with the supplied inputs and compare the results with the tables here. Then change something: the shape, the pressure profile, or the numerical resolution. Follow what changes and ask why. The equations describe a calculation; the code lets you examine how I implemented it. You can question both. If you find a discrepancy, I want to understand it. If the work gives you another question to investigate, even better.

I do not imagine there is a very large community of Black trans women sitting around thinking about what is happening inside a plasma reactor. I do. And sometimes I wish I had somebody sitting next to me who wanted to talk about it.

That is probably the part of my curiosity that is hardest to explain. When I was very young, adults recognized that I was unusually intelligent. But being able to identify something about the way a child thinks does not answer the questions that child is going to carry through her life. I have always had questions. They follow me into whatever I am doing, and sometimes they take me somewhere I do not know anyone.

With plasma, the question begins with something almost embarrassingly simple. Think about opening an oven and feeling the heat meet your face. Now try to get your mind around the temperature engineers are aiming for at the heart of ITER, an experimental fusion machine: about 150 million degrees Celsius. Roughly ten times the temperature at the center of the Sun.[2] I have to stop there for a second. We are building a machine on Earth whose fuel is supposed to get that hot. And part of the answer to keeping its hottest region away from the walls is to shape magnetic fields so that they guide the motion of its charged particles. You cannot see that

[1] Companion repository: `https://github.com/nurdymuny/chihiro-plasma`. Public access is forthcoming. The linked source references identify the revision used in this chapter; setup instructions are in the repository's README.

[2] ITER Organization, "Facts & Figures," `https://www.iter.org/facts-figures`. This is ITER's intended central plasma temperature, not a reported operating achievement. Temperature measures particle energy on a microscopic scale; total thermal energy also depends on how much matter is present. The plasma is extremely dilute; the vessel itself is not at this temperature. See also ITER's FAQs, `https://www.iter.org/faqs`.

magnetic structure with your eyes, but its shape helps determine whether the confinement works.[3] Magnets. Geometry. Matter hotter than the center of our own star. Okay. There is no way you can tell me that is not cool.

But then who do I ask about it? Who wants to sit with me while I try to understand what the particles are doing, what the magnets are doing, and what happens when something moves in a way nobody wanted?

I think there is something sacred about that kind of curiosity. It was protected in me. I protected it in my children. I still protect it in myself. Sometimes that means continuing to care about a question even when I have nobody nearby to share it with.

Then, through my work, I began finding myself inside the very worlds I had questions about. My managers and employers brought me into problems I would not have known how to find on my own. Army intelligence, Naval intelligence, DISA, a contract for a company connected to the NSA, Pandora, eventually NASA. I sat behind engineers doing their work. I listened to physicians at a virus-detection company. I watched people working on military problems whose consequences were very real. There was so much happening around me that I wanted to understand.

I took notes. I tried to learn what mattered to the people doing the work. What were they trying to accomplish? Where were they getting stuck? What could I build that would help? And sometimes somebody would turn toward me and ask what I thought.

What did I think? I was still trying to figure out whether I belonged in the room.

That was a familiar feeling in my early career. Being invited did not immediately make it disappear. I could be sitting in front of a problem because somebody believed I might be useful and still wonder whether they had made a mistake. But when I began answering, the answer often came from a different direction. I would notice something in the structure of the data, or in the relationships between the things people were measuring. Something in my notes would connect to something else I had seen. I would want to follow that connection and find out whether it held.

This is why the name CHIHIRO means so much to me.

The letters stand for Confinement Hierarchy via Invariant Harmonic Instability Recognition Operator. But if I only give you the acronym, you miss why that name matters. Chihiro is the girl in *Spirited Away*, one of my favorite movies. She follows her parents into a world she knows nothing about. She is frightened. She does not understand the rules, or what the beings around her want from her, or how she is supposed to get through any of this. And I watch her and think: yes. That is what it felt like.

There is a moment in the movie when Chihiro has to eat food from the spirit world so she will not disappear. I keep coming back to that. Imagine having to take something of a world into yourself just to remain present in it. That is how learning this work can feel to me. I read the papers, follow the lectures, work through the mathematics, and try to understand what people have already discovered. I want to be able to participate. I want to ask a question somebody can hear and build something somebody can use.

And then there is the other part of the movie: even there, even after she has eaten the food, the spirits can smell that she is human. She still comes from somewhere else. That part gets me. Learning the language does not necessarily make people stop noticing that you do not seem to

[3] ITER Organization, "Magnets," `https://www.iter.org/machine/magnets`, describes the fields used to confine, shape, and control the plasma. Magnetic confinement does not eliminate heat exhaust or the need to protect material surfaces.

belong. I can do the work and still find myself trying to explain why I am in the room. The life that brought me there does not disappear when I open a mathematics paper. Neither does the way I see things. Those are part of what I bring to the problem.

But as I watch her find ways to act and help inside that world, I recognize another possibility: perhaps arriving from somewhere else does not mean arriving empty-handed. Maybe there is something I can contribute. I could be afraid and still notice a relationship. I could be unfamiliar with a field and still have a question worth investigating. I did not leave my way of thinking outside just because I was scared.

That is how I approached plasma research. I followed Dr. Nuno Loureiro through videos. Here was somebody who had spent his life learning to understand things I was just beginning to ask questions about. I admired him. I wanted to understand more, and I wanted to see whether I could build software that would help people doing this work.

Loureiro studied how magnetized plasmas behave, in laboratories and in space. In an ordinary ionized plasma, freed electrons move among positively charged ions; an ion need not have lost every electron.[4] Those charged particles respond to magnetic fields, and their collective motion can also change the fields. A central part of his work concerned *magnetic reconnection*: a process in which the connectivity of the magnetic field changes and magnetic energy can be transferred into particle motion and heat. The drawn field lines are a way of representing a continuous field, not literal strings being cut. What changes is which regions those field lines connect. That change matters in solar activity and in disturbances inside fusion plasmas.[5]

With collaborators, he investigated how thin current sheets—regions where the magnetic field varies sharply and electric current is concentrated—can become unstable and form chains of magnetic islands, also called plasmoids. He also studied the relationship between reconnection and turbulence: how irregular plasma motion and magnetic structure interact across different scales.[6] Here were questions about a system changing its own connections, and about the consequences for energy. Of course I wanted to listen.

He would probably never know I existed, but I was there, behind my laptop, listening. When I learned of his death, that felt strange and sad in a way I did not quite know how to explain. Someone can have a place in your life through what they teach you and never know you are there.[7]

I was the girl behind the laptop trying to help some of the smartest people in the world build software to solve a difficult problem. I had notes, questions about the geometry, and methods I could investigate. I still had to find out where they worked, what they missed, and what this field could teach me. That work became CHIHIRO. Her name carries the fear and the fascination together, because that is how I came into it.

[4] U.S. Department of Energy, "DOE Explains... Plasma," explains ionization and the collective behavior of charged particles.

[5] See MIT Physics, "Nuno F. G. Loureiro," research profile, and the publications listed in the CHIHIRO references.

[6] Loureiro, Schekochihin, and Cowley (2007), "Instability of current sheets and formation of plasmoid chains"; Loureiro and Boldyrev (2017), "Role of Magnetic Reconnection in Magnetohydrodynamic Turbulence." These are specific contributions within a much larger body of plasma research.

[7] Loureiro directed MIT's Plasma Science and Fusion Center. MIT reported his death on December 16, 2025. See MIT News, "Nuno Loureiro, professor and director of MIT's Plasma Science and Fusion Center, dies at 47." I learned from his public work; we did not collaborate. CHIHIRO's calculations below concern the specified plasma-stability models and are not a reproduction of his reconnection theory.

What makes an enormous difference along the way is whether somebody will give you room to explain what you are seeing. I know what that can change. I also know what it feels like when the explanation never gets that room.

§ 13.1

A place to explain what I see

ICARUS began with curiosity too. I wanted to know whether I could approach guidance, navigation, and control—the questions of where a vehicle is, where it should go, and how it gets there—through a different mathematical language.

Dr. Ryan Rapetti would give me a challenge, and I would come back with something geometric. He would ask another question. I would go investigate. That conversation kept going because he was willing to explore the approach with me. What began as a blog grew into *The Geometry of Flight*.

ICARUS is a thought experiment. It is a way of asking what becomes possible to see when we describe a familiar problem differently. I was following my curiosity into a field I did not know, bringing the mathematics I did know, and trying to learn what happened when they met. Sometimes I arrived with a little naïve intuition. I had to investigate it, work through it, and find out whether it survived.

Isn't that part of what we are here to do? Learn something. Ask a question we could not have asked yesterday. Follow an idea far enough to discover where it works and where it falls apart.

When I say Ryan found me at Olive Garden, I need you to understand what that looked like. I was running back and forth for hours, feeling the sweat from another intense shift, covered in other people's leftover food. There was nothing glamorous about it. I had an engineering background and twenty-eight years of experience, and I was working myself through those shifts for very little money. It felt like the lowest place in my life. I was trying to get through the work in front of me while carrying everything I still wanted to do.

Ryan saw a possibility in me there. The message I heard was: you do not belong stuck here. There is work you could be doing. That kind of recognition does not solve everything. But it can give a person enough room to take the next step.

Through Ryan's referral, his company invited me to interview. But Ryan was not the person interviewing me, and the position I was being considered for was not a job designing ICARUS or taking responsibility for somebody's aircraft.

During the interview, Dr. Robert put ICARUS on the screen and asked me to explain it. I began with fiber bundles, because that geometric structure was part of what the experiment was investigating. A fiber bundle keeps a collection of related possibilities attached to each place or state in a model. I was trying to explain what that structure let me do. I remember him stopping me: "I don't care about that, explain this."[8]

[8]The interview dialogue is my recollection, not a transcript. The company is unnamed. The account describes what I remember saying and hearing and how the encounter felt to me; it does not establish the interviewer's intentions or the reasons for the hiring decision.

I blinked. The thing he had put in front of me had grown through a conversation in which I was encouraged to answer geometrically. Now, as I experienced it, I was being asked to explain the result while setting aside the language through which I had developed it.

That is where the comparison to language becomes very real for me. People will tell an immigrant, "Speak English," as though saying it more forcefully supplies the language the person does not yet have. Imagine arriving somewhere you do not speak the language and having people insist, in that language, that you explain yourself. You may have something quite specific to say. The demand does not create a way for you to say it.

Translation takes work from both sides. Someone has to be willing to ask what a word means, what a structure does, and where the two descriptions meet.

With Ryan, we had been doing that work. His questions gave me something to investigate. My geometric answers gave us something else to discuss. In the interview, I felt that the willingness to translate had disappeared. I was trying to explain an experiment in another language, and I felt judged for not presenting it in the familiar one.

The distinction matters because ICARUS was an attempt to reimagine the description in the first place. Asking whether that attempt works is a serious and necessary question. We can examine its assumptions, compare its outputs, and identify what it fails to capture. But we have to let the explanation become visible before we can examine it.

Instead, I felt pushed toward retrieving the established methods and terminology. I wanted to show how I approached a problem. Those are different things to investigate about a person. Knowing an accepted method matters; so does being able to notice a relationship, formulate a question, and work out how to test it. I did not feel that the interview made room to discover how I did those things.

And it hurt.

A researcher at that company had encouraged me to keep exploring. Another researcher at the same company left me feeling that I was not worth hearing out. I cannot tell you what Robert thought about me. But I can tell you how small I felt in that moment, trying to explain something I had worked so hard to understand and feeling the room close around the explanation.

The interview did not lead to a job offer. That opportunity could have changed my life.

Ryan found me sweating through restaurant shifts and made "you don't belong here" feel like an invitation into another world; Robert met me in his company's conference room, discussing work in his field, and left me hearing the same words as a reason I should leave. That is the contrast I carry from those encounters. It is how they landed in me, not a second quotation from the interview.

So which one was right? Where did I belong? I had to answer that question for myself. I belong in both worlds because I am in both worlds. Maybe that is my own version of a double cover: two ways of seeing the person who is already here.

I need to say that plainly because you can read a book full of mathematics and imagine that its author has somehow made it to the other side of all this. That there is a comfortable research position behind the pages. That the work has brought security, recognition, and a place to belong.

I serve Thai food for a living now. I am a waitress. I do this research alongside the work that pays my bills. I am still looking for opportunities, still being turned down, still trying to find people who will take the time to understand what I might contribute.

There are people who will serve food for their entire working lives. They may never get an invitation to sit in a conference room with Dr. Robert or anyone like him. Do they have ideas worth hearing? Yes. And so do I. I do not have to make their work small to explain how much I wanted that opportunity. Serving somebody dinner does not tell you the limits of what I can think about. It does not tell you the limits of what the person serving beside me can think about either.

More than a handful of trans people have asked me for a copy of this book. Some are not asking because they work in these fields or expect to follow every equation. They see a Black trans woman reaching into a world where they may not have expected to see someone like us. I want them to know what that reaching actually feels like.

It includes wonder. It includes the pleasure of finally understanding something difficult. It also includes sitting in a room hoping somebody will hear you, then leaving without the opportunity you needed. I know that feeling. Writing these books has not put me beyond it.

That is why Chihiro matters here. I recognize the fear of entering a world whose rules other people already know. I recognize the work of learning enough to participate and still being treated as someone from somewhere else. And I recognize the gradual discovery that perhaps the person who arrived can help through her attention, her questions, and her way of responding.

I am still trying to find out what I can contribute. CHIHIRO is part of that effort. The curiosity that brought me to the problem is still here, even after an interview ends.

I had left this work outside the book because I was not a plasma researcher. But I want you to be able to follow the thought far enough to examine it with me. We can begin with a physical question: what happens when the plasma moves? Then we can show what I tried, what the records support, and what still needs work.

§ 13.2

What happens when the plasma moves?

Let me stay with that image of matter hotter than the center of the Sun. It is easy to say the temperature and move on. But what, exactly, are we trying to hold inside this machine? What has happened to the matter by the time it gets that hot?

Start with a hydrogen atom: a positively charged nucleus with a negatively charged electron bound to it. Add enough energy and the electron can separate from the atom. The remaining charged particle is called an *ion*. In the hot hydrogen fuel of a fusion experiment, the electrons and nuclei move separately. We now have a plasma. The positive and negative charges can almost balance across a region, while the individual particles still respond strongly to electric and magnetic fields. An overall balance of charge does not make those charges disappear.

And these particles affect one another through the fields they create. Their movements can organize into currents and waves. Imagine trying to understand a crowd in which each person's movement can change the forces felt by people elsewhere in the room. Following one person would tell you something, but it would not explain the whole crowd. That collective behavior is part of what makes plasma such a rich subject.[9]

[9] U.S. Department of Energy, "DOE Explains... Plasma," `https://www.energy.gov/science/doe-explainsplasma`, introduces ionization, free electrons, ions, and collective waves and instabilities.

Fusion asks something further of this moving matter. Light nuclei must get close enough for a nuclear reaction to join them and release energy. Their positive charges repel one another. Heating the fuel gives the nuclei more energy of motion and increases the chance of fusion-producing encounters. That is why the extraordinary temperature matters. But getting the fuel hot is only part of the task. It also has to retain enough of that energy, at sufficient density, for fusion to be useful.[10]

Now put yourself back in front of the machine. You have this intensely hot material, and you need to keep its core away from the solid walls. So what does a magnet actually do for you?

A moving charged particle feels a magnetic force that turns its motion. In a uniform magnetic field, the part of its motion across the field becomes circular; motion along the field can continue. Together those motions make a spiral, like a tiny corkscrew advancing along a direction. The magnetic force itself redirects the particle rather than increasing its speed. This gives us a way to guide hot matter without putting a solid surface directly in its path.[11]

That last detail matters: the particle can still travel along the field. An open route gives it somewhere to go. A tokamak closes the long route into a ring. Picture a doughnut, with the plasma occupying the ring-shaped region inside a surrounding chamber. The geometric name for that ring shape is a *torus*.

But simply bending the magnetic route into a circle does not finish the job. The field curves and its strength varies across the ring. Those variations produce drifts across the field. A field running only the long way around the ring does not provide the confinement we need.

So the magnetic route also winds around the ring's smaller cross-section. Imagine wrapping a ribbon around the body of a doughnut while carrying it around the central hole. There are two directions in that picture. Around the central hole is *toroidal*; around the smaller cross-section is *poloidal*. Combining them gives the field its twist. A field line is a way of drawing the field's local direction, not a physical thread with a particle tied to it.[12]

In a tokamak, external coils supply the strong toroidal field. A current flowing through the plasma supplies an important part of the poloidal field; additional coils help position and shape the plasma. The material being confined therefore participates in making its own magnetic environment. That is the part I want you to keep in your mind. We are trying to understand a moving system that helps shape the conditions of its own movement.[13]

Confinement means slowing the escape of particles and energy enough to maintain the conditions we need. The particles keep moving. Collisions, turbulence, and larger disturbances can carry energy outward or spoil the arrangement. Researchers must manage those losses and the heat that

[10]ITER Organization, "What is a tokamak?," `https://www.iter.org/machine/what-tokamak`, explains the fuel, ionization, heating, and magnetic confinement in a tokamak.

[11]Max Planck Institute for Plasma Physics, "Magnetic confinement," `https://www.ipp.mpg.de/1766269/einschluss`; Richard Fitzpatrick, "Trapped and Passing Particles," plasma-physics teaching notes, `https://farside.ph.utexas.edu/teaching/plasma1/Fusionhtml/node27.html`. The uniform-field picture introduces gyromotion and travel along the field; the tokamak geometry adds drift and trapping.

[12]CEA, Institute for Magnetic Fusion Research, "Magnetic confinement," `https://irfm.cea.fr/en/physics-of-fusion/magnetic-confinement/`, develops the progression from helical particle motion to toroidal drift and twisted magnetic surfaces.

[13]U.S. Department of Energy, "DOE Explains... Tokamaks," `https://www.energy.gov/science/doe-explainstokamaks`; ITER Organization, "Magnets," `https://www.iter.org/machine/magnets`, particularly the toroidal field, poloidal field, and central-solenoid descriptions.

reaches material surfaces. The question is how well the magnetic geometry holds up while all of this is happening.[14]

Even the doughnut picture has another layer. Imagine cutting across one side of the ring and looking straight at the exposed slice. Its outline can be taller than it is wide, or shaped more like a D than a circle. Changing that outline changes the physical configuration whose balance and response we have to calculate.[15] Now the geometry has a job we can picture. It describes where the plasma is, how the field winds through it, and the arrangement we want to keep when something gives it a nudge. That brings us to two questions we need to separate: is the arrangement balanced, and will it stay near that balance when disturbed?

An equilibrium is a configuration in which the modeled forces balance. Stability asks a further question: if that configuration is disturbed, does the disturbance remain small, or does some part grow? A carefully balanced object can still fall when nudged. Finding a balance is therefore not the same as showing that it will withstand a disturbance.

The ideal magnetohydrodynamic, or ideal-MHD, model treats the plasma as a conducting fluid and neglects electrical resistivity. For a static equilibrium under the energy principle's assumptions, a displacement $\boldsymbol{\xi}$ has a quadratic perturbation energy $\delta W[\boldsymbol{\xi}]$. The question is whether some admissible displacement makes that energy negative. Here *admissible* matters: the displacement must obey the boundary conditions and the constraints of the chosen physical model.[16]

Let me unpack that notation before we ask it to do any work. $\boldsymbol{\xi}$, pronounced "xi," assigns a small movement to each part of the plasma. The movements can differ from place to place. $\delta W[\boldsymbol{\xi}]$ is the energy test applied to that whole pattern. The square brackets mean "evaluate this pattern." Calling the energy *quadratic* means that doubling every displacement multiplies this energy contribution by four. We are examining the leading energy change for small departures from the balanced state.

For a finite calculation, a *vector* is an ordered list: u stores the movements represented by the calculation. A *matrix* is a table of coefficients that acts on that list. K describes the restoring response, including how movements at different positions interact; M describes their inertia, the weighting of resistance to acceleration. "Symmetric" means the table reads the same when its row and column indices are exchanged. "Positive definite" means that the inertial weight of every nonzero movement is positive. We take both matrices to be real and symmetric, with M positive definite.

A normal mode is a movement pattern whose spatial shape stays the same while its amplitude changes with time. ω, "omega," describes that time dependence. The equation asks for a pattern for which the two responses line up, with ω^2 as their scale factor:

$$\underbrace{Ku}_{\text{restoring response}} = \underbrace{\omega^2}_{\text{mode value}} \underbrace{Mu}_{\text{inertial response}} .$$

[14] Taylor and Newton (2015), "Special topics in plasma confinement," surveys confinement physics; ITER Organization, "Divertor," `https://www.iter.org/machine/divertor`, describes the controlled exhaust of heat and particles.

[15] Miller et al. (1998), "Noncircular, finite aspect ratio, local equilibrium model," treats elongation and triangularity within a local equilibrium description.

[16] Bernstein, Frieman, Kruskal, and Kulsrud (1958), "An energy principle for hydromagnetic stability problems," develops this classical framework. The worked scalar operator below is a reduced numerical model within this investigation.

To find the lowest such value, we use

$$\omega^2_{\min} = \min_{u \neq 0} \frac{\underbrace{u^{\mathsf{T}} K u}_{\text{restoring energy measure}}}{\underbrace{u^{\mathsf{T}} M u}_{\text{inertial weight}}}.$$

The superscript T turns the column of numbers into a row, so the products above become single numbers. "Min" means take the smallest quotient over all the allowed nonzero patterns; $u \neq 0$ excludes the motionless list, which would give zero divided by zero. Multiplying a pattern by two multiplies both parts of the fraction by four. The quotient compares the pattern, without letting its arbitrary size decide the answer. This energy-per-weight comparison is called a *Rayleigh quotient*.

A negative value of ω^2 permits exponential growth in this static, ideal linear model. A zero value marks a marginal direction. A positive lowest value means every represented direction has positive restoring energy. The displacement space tells us which movements that conclusion covers.[17]

The negative "square" deserves an explanation too. In a growing mode, ω can be imaginary rather than real; its square is then negative. The imaginary unit i satisfies $i^2 = -1$, so $\omega = i\gamma$ has square $-\gamma^2$ for a real growth parameter γ. The underlying time equation turns that case into exponential growth or decay instead of an ordinary oscillation. We are not claiming that the square of a real number is negative. We are using its sign to distinguish two kinds of time behavior.

Think of asking the whole plasma, "What is the easiest way for you to move away from this balance?" We are not testing just the first shape of movement that comes to mind. We search over the movements the model allows. That is the role of the minimum. A pleasant-looking answer for one movement is not enough if another movement can grow.

CHIHIRO's early investigation followed pressure-driven *ballooning*: a displacement across the magnetic field, concentrated along portions of a field line. Pressure gradients in unfavorable field-line curvature can drive the bulge, while bending the field provides a restoring cost. The classical Connor–Hastie–Taylor construction makes it possible to study such disturbances in a sheared toroidal field.[18]

Two quantities organize our example. The safety factor $q(r)$ describes field-line winding: the toroidal advance, around the long way of the torus, divided by 2π during one poloidal circuit, around its cross-section. Despite its historical name, the quantity we are following here is field-line winding.[19] Magnetic shear measures how that winding changes with radius:

$$s = \underbrace{\frac{r}{q}}_{\text{relative scaling}} \underbrace{\frac{dq}{dr}}_{\text{outward change in winding}}.$$

[17] Freidberg, *MHD Theory of Fusion Systems* (MIT, 2007), lectures 14–15, derives the static ideal-MHD normal-mode and variational formulations. The time convention is $e^{-i\omega t}$; here the stiffness is normalized so that $\delta W = u^{\mathsf{T}} K u/2$.

[18] Connor, Hastie, and Taylor (1978), "Shear, Periodicity, and Plasma Ballooning Modes," *Physical Review Letters* 40, 396–399. The circular s–α model motivates the retained implementation; the shape-dependent coefficients printed here define the implementation's reduced extension.

[19] Freidberg (2007), lecture 6, defines the safety factor from the field-line winding on a magnetic surface.

Read dq/dr as the slope of the winding profile: how much q changes for a small outward change in radius r. Multiplying by r/q expresses that change relative to the local radius and winding. For a small step, $\Delta q/q \approx s\,\Delta r/r$. So at a place where $s = 1$, an outward step of about one percent in radius produces about a one-percent increase in winding, to first order. If q is constant nearby, its slope is zero and so is the shear. We have given "how differently do neighboring surfaces wind?" a number we can calculate.

The dimensionless parameter α describes a normalized pressure gradient. In the circular convention used to motivate this model, $\alpha = -R_0 q^2\, d\beta/dr$, where $\beta = 2\mu_0 p/B_0^2$, p is pressure, B_0 a reference magnetic field held fixed in this normalization, and R_0 the reference major radius. Here $p(r)$, and hence $\beta(r)$, is local, rather than the averaged pressure used later. If pressure decreases outward, this α is positive. The profile replay later in the chapter uses its own explicit shaping factor when converting pressure to α; we will keep that convention visible.[20]

β, "beta," compares plasma pressure with the magnetic pressure scale $B_0^2/(2\mu_0)$. μ_0, "mu-zero," is the vacuum magnetic permeability, the constant that supplies the magnetic units in this expression. R_0 measures the large radius of the ring. In $-R_0 q^2\, d\beta/dr$, the derivative supplies the outward pressure slope, $R_0 q^2$ supplies the model's scale, and the minus sign makes an outward decrease positive. α, "alpha," therefore describes a scaled pressure gradient, not the pressure itself. Two profiles can have the same pressure at one location and different slopes there. The model needs to know the difference.

So there are two separate changes to follow: how neighboring field lines wind, and how pressure varies across the plasma. Their effect on stability depends on how they enter the disturbance equation. That is where we go next: from the shape we can picture to the operator we can actually calculate.

§ 13.3

Build the operator where we can see it

Let θ measure position along the extended field-line coordinate, and let $u(\theta)$ describe the scalar disturbance: one signed movement value at each position. In the retained shape experiment, elongation κ stretches the cross-section and triangularity δ distinguishes the two directions of D-like shaping. We build an intermediate quantity Λ, a positive weight f, a potential coefficient V, and finally the rule L that acts on the disturbance. The program uses

$$\begin{aligned}
\Lambda(\theta) &= s\theta/\kappa - \alpha \sin\theta, \\
f(\theta) &= 1 + \Lambda(\theta)^2, \\
V(\theta) &= \alpha\big[\kappa\cos\theta + \Lambda(\theta)\sin\theta + \delta\cos(2\theta)\big], \\
Lu &= -\frac{d}{d\theta}\left(f\frac{du}{d\theta}\right) - Vu.
\end{aligned}$$

[20] Taylor and Newton (2015), §28, equations (28.1)–(28.3), gives the classical s–α definitions and marginal equation. Our pressure normalization writes the magnetic permeability μ_0 explicitly in SI units.

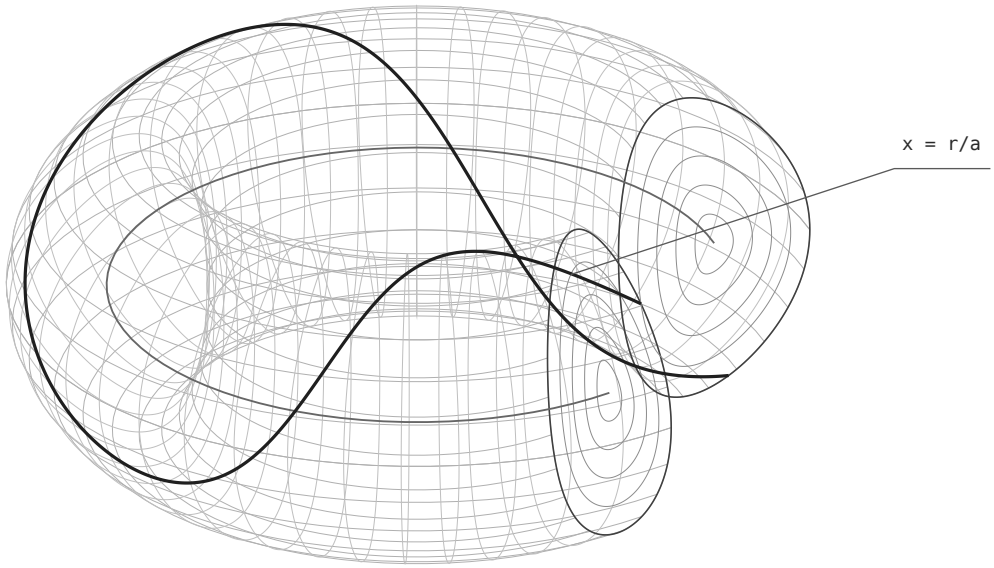
DAVIS GEOMETRIC / CHIHIRO CHI / GA-01

A geometry the calculation can read

Shaped coordinate surfaces / analytic ITER-like preset

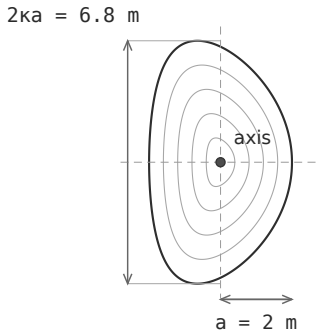
01 CUTAWAY / NESTED RADIAL COORDINATES

Drawing of the prescribed shape; winding is schematic.



Surface labels locate the inputs p(x), q(x), s(x) and α(x).

02 MERIDIONAL SECTION



03 PARAMETER REGISTER

Quantity	Value
Major radius R_0	6.2 m
Minor radius a	2.0 m
Field B_0 / current I_p	5.3 T / 15 MA
Elongation κ / δ	1.7 / 0.33

The shape fixes a coordinate description. The operator supplies the test.

THE GEOMETRY OF FUEL CHI / GA-01

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 42. The analytic shape in coordinated views. The cutaway exposes nested radial coordinates; the winding is schematic. Dimensions and shape parameters belong to the stated ITER-like preset.

There are four steps in that little stack. θ, "theta," is our position label along the extended field line. κ, "kappa," controls elongation; δ, "delta," controls the shaping term. Here δ is a shape parameter, distinct from the notation δW for perturbation energy above. Sine and cosine provide the repeating angular variation in these coefficients; $\cos(2\theta)$ varies twice as fast as $\cos\theta$.

First, Λ, capital "lambda," combines the shear and drive into an intermediate quantity. Second, squaring it and adding one produces f, a coefficient that is always at least one. Third, V combines the displayed pressure and shape contributions. It is called the potential term here; the letter does not mean voltage. Finally, L is the *operator*: the rule that takes a whole disturbance profile u and returns another profile. In that last line, $du/d\theta$ measures how the movement changes along the line. We multiply that slope by f, take its slope again, change the sign, and subtract Vu. Those are the actual operations the notation has compressed.

I want to keep the purpose visible while we do this. We start with a proposed movement and calculate the model's response to it. An *eigenfunction* is a pattern for which that response has the same shape as the input: $Lu = \lambda u$. The number λ, lowercase "lambda," is its eigenvalue, the multiplier relating the two. On a finite grid we call the pattern an eigenvector. The *spectrum* is the collection of these eigenvalues. It gives us more than the response to one movement we happened to choose.

We impose $u(-T) = u(T) = 0$, initially with $T = 6\pi$, and compute the lowest eigenvalue $\lambda_{\min}$. These endpoints truncate the extended field-line coordinate for the numerical calculation. Here λ measures the normalized spectral response. A growth rate in seconds would also require the physical inertia and time scale.

The endpoint equation says that our modeled movement is held at zero at both ends of this numerical window. T sets its half-width; the full interval runs from $-T$ to $+T$. These are endpoints of the calculation along an extended coordinate, not two physical ends of a ring-shaped plasma. "Lowest" means we select the smallest λ among the patterns represented with those constraints.

There is also a position choice inside the formula. The classical ballooning equation includes a phase locating the disturbance along the field line. This implementation fixes that phase at zero rather than searching over it.[21] The first and second crossings below therefore belong to this fixed choice, as well as to the stated coefficients and finite interval. Varying the phase would ask how that answer changes as the disturbance moves along the field line.

The signs are worth slowing down for. Integration by parts, with the endpoint values zero, gives

$$\langle u, Lu\rangle = \underbrace{\int_{-T}^{T} f(u')^2\, d\theta}_{\text{cost of variation along the line}} - \underbrace{\int_{-T}^{T} Vu^2\, d\theta}_{\text{potential contribution being subtracted}} .$$

The angle brackets on the left mean multiply u by its response Lu at each position and add those products over the interval. An integral, the stretched $\int$ symbol, does that continuous adding. Its limits tell us where to begin and end; $d\theta$ tells us which coordinate we add over. The prime in

[21] Taylor and Newton (2015), §28, equation (28.1) and figures 57–58, retains the phase y_0 and explains why it must be varied when identifying the instability region.

u' is short for $du/d\theta$. Integration by parts rearranges derivatives inside that total. Its endpoint contribution vanishes because we have fixed u to zero there. It is the same energy calculation, now written so that we can see its two competing contributions.

The first term penalizes changes along the disturbance. Since $f \geq 1$, it cannot be negative. The second term can lower or raise the total, depending on the potential V. We must compare both terms for the same displacement. A bound on the first term alone is not a stability certificate for their difference.

On a grid, this becomes something we can inspect one entry at a time. Take N interior points, spacing $h = 2T/(N+1)$, and set $u_0 = u_{N+1} = 0$. Sample f halfway between the points. Then

$$H_{ii} = \frac{f_{i-1/2} + f_{i+1/2}}{h^2} - V_i,$$
$$H_{i,i+1} = H_{i+1,i} = -\frac{f_{i+1/2}}{h^2}.$$

All other entries vanish. Each point interacts with its two neighbors. The diagonal also carries the local potential. This is the symmetric matrix whose lowest eigenvalue was used for the saved shape results.

Here i is a position's number on the grid. H_{ii} is the entry in row i, column i; it multiplies the movement at that same point. $H_{i,i+1}$ multiplies the next point's movement. The half-index in $f_{i+1/2}$ locates the coefficient between points, while V_i is the potential at point i. The spacing $h = 2T/(N+1)$ divides the full window into $N+1$ gaps around the N interior points. The factor $1/h^2$ is the scaling needed when replacing the derivative operation by neighboring differences. Only the diagonal and its two immediate neighbors can be nonzero. That is what *tridiagonal* means.

The next proposition checks that this table of coefficients really keeps the energy account we just described. In it, $\sum$ means add the listed terms, and $(u_{i+1} - u_i)^2$ measures the squared difference between neighboring movements. The endpoint values remain zero. An *orthonormal eigenbasis*, used in the proof, is a complete set of eigenvector patterns, each scaled to length one and perpendicular to the others in the ordinary dot product. Every movement list can be assembled from those patterns. The dot product multiplies corresponding entries of two lists and adds the products; perpendicular lists give zero. This extends the familiar right-angle relationship to our longer lists of movements.

Proposition 13.1 (The discrete energy is the matrix energy)**.** *For the matrix above and any real vector u,*

$$u^{\mathsf{T}} H u = \frac{1}{h^2} \sum_{i=0}^{N} f_{i+1/2}(u_{i+1} - u_i)^2 - \sum_{i=1}^{N} V_i u_i^2.$$

Consequently, $\lambda_{\min}(H) = \min_{u \neq 0} u^{\mathsf{T}} H u/(u^{\mathsf{T}} u)$.

Proof. Expand every squared difference. The coefficient of u_i^2 is $(f_{i-1/2} + f_{i+1/2})/h^2$, and the coefficient of $u_i u_{i+1}$ is $-2f_{i+1/2}/h^2$. Subtracting the potential terms gives exactly the quadratic form of H. The minimum formula follows by expanding u in an orthonormal eigenbasis of the real symmetric matrix: the quotient is a weighted average of its eigenvalues, and the lowest eigenvector attains the minimum. □

Here is the plain version. Moving neighboring points differently costs something. The potential can give some of that cost back. The matrix keeps both accounts in the same calculation. Its lowest eigenvector finds the pattern with the smallest net cost. Now, when the program prints an answer, we can follow it all the way back to the movement that produced it. That is the part I want to be able to see.

The proof is bookkeeping we can check. Expand $(u_{i+1} - u_i)^2 = u_{i+1}^2 - 2u_i u_{i+1} + u_i^2$. Each interior point receives a squared contribution from the gap on its left and the gap on its right. Each neighboring pair receives the cross-term with the minus sign. Those are exactly the matrix entries above. Expressing a movement in the eigenvector patterns then makes its quotient a weighted average of eigenvalues, with nonnegative weights summing to one. An average cannot be smaller than its smallest ingredient. Choosing the lowest pattern alone reaches that value.

There is something familiar sitting inside this calculation. I want to stop here long enough to show you where I recognize it. When I was working on holomorphic functions in GIGI, including the wine example, I was also asking how neighboring values could fit together. My work on Crouzeix's conjecture asked how a matrix could carry a guarantee through a change of representation. Now we have an energy account for a disturbance in CHIHIRO. The questions are different. But I can put the mathematical connections on the page and let you inspect them.

Start with the first sum in Proposition 13.1. For a small example, set its weights and spacing to one and temporarily remove the potential. I will call the five values $a_0, \dots, a_4$, so we can keep this exercise separate from the plasma displacement. Fix $a_0 = 0$ and $a_4 = 4$. We are changing the endpoints for this exercise; CHIHIRO's displacement still has zero endpoints. The cost is

$$E(a) = \underbrace{\sum_{i=0}^{3} (a_{i+1} - a_i)^2}_{\text{add the squared changes across four gaps}}.$$

Think of five marks along a straight ruler, with a height assigned to each mark. The first height is zero and the last is four. Where should the three middle heights go if we want the smallest total cost of changing height between neighbors?

An interior height a_i appears in two gaps. Differentiating their contributions and setting the result to zero gives

$$2(a_i - a_{i-1}) + 2(a_i - a_{i+1}) = 0, \qquad a_i = \frac{a_{i-1} + a_{i+1}}{2}.$$

The derivative asks what happens to the cost when we move just that height a little. At the minimum there is no first-order benefit to moving it either way. Each middle height therefore sits at the average of its neighbors. The answer is $0, 1, 2, 3, 4$: four changes of one, so $E = 4$. Raise the middle height from two to three and the sequence becomes $0, 1, 3, 3, 4$. Its changes are $1, 2, 0, 1$, and their squares add to six. The endpoint difference is unchanged, but we have concentrated more of the change into one gap. Squaring makes that cost visible. More generally, Cauchy–Schwarz gives $\sum_{i=0}^{3} (a_{i+1} - a_i)^2 \geq (a_4 - a_0)^2/4 = 4$, with equality precisely when all four changes are equal. This verifies the minimum, rather than only finding a stationary candidate.

That averaging equation is discrete harmonicity on this uniform chain. With positive unequal weights, each interior value instead equals its neighbors' weighted average. In the continuous plane,

the corresponding unweighted energy and equation are

$$E_D[U] = \frac{1}{2}\int_\Omega \underbrace{(U_x^2 + U_y^2)}_{\text{squared slopes in the two directions}} dA,$$

$$\Delta U = U_{xx} + U_{yy} = 0.$$

Here Ω is the region, dA means add over its area, and the subscripts mean derivatives in the named directions. Fix the values around its boundary. The harmonic field minimizes this energy among fields with those same boundary values. A nonconstant boundary can therefore produce a nonconstant, locally balanced interior.[22]

This brings me back to the two landscapes in a holomorphic function. Use $F(z) = z^2$, with $z = x + iy$. Multiplication gives

$$F(z) = \underbrace{x^2 - y^2}_{U(x,y)} + i\ \underbrace{2xy}_{W(x,y)}.$$

The two heights are the real and imaginary parts of one answer. At the same input $z = 0.2 + 0.3i$, the first landscape has height $0.04 - 0.09 = -0.05$, and the second has height $2(0.2)(0.3) = 0.12$. So the answer is $-0.05 + 0.12i$. Move the input point and both heights change together. The panel shows exactly this paired reading over the same unit disk.

Look at their slopes:

$$\nabla U = (2x, -2y), \qquad \nabla W = (2y, 2x).$$

The triangle-shaped symbol ∇ collects the two slopes into one vector, the gradient. It points uphill most steeply. These two gradients have equal length and zero dot product. Away from the origin, their uphill directions meet at a right angle; at the origin both gradients vanish, so neither has a preferred uphill direction. Also $U_{xx} + U_{yy} = 2 - 2 = 0$ and $W_{xx} + W_{yy} = 0 + 0 = 0$. Both landscapes are harmonic. Their slopes satisfy $U_x = W_y$ and $U_y = -W_x$, the Cauchy–Riemann equations that connect them as one holomorphic function. Two independently chosen harmonic landscapes need not satisfy that pairing.[23]

The wine example uses a richer polynomial with the same kind of relationship. We counted 6,497 wine records into 32 alcohol bins. Writing $p_k = n_k/6497$ for the proportion in bin k, the fingerprint is

$$F_{\text{wine}}(z) = \sum_{k=0}^{31} p_k z^k.$$

The coefficients come from the measurements. The complex coordinate z is where we probe the resulting mathematical object; it is not an alcohol concentration. The nonnegative proportions sum to one, so $|F_{\text{wine}}(z)| \le 1$ on the unit disk. GIGI's harmonic-conjugate operation starts from the complete real polynomial field and a specified imaginary value at the center. It checks the

[22] Dirichlet's principle: University of Toronto, *PDEs from Calculus of Variations*, lecture notes for March 23, Theorem 1.3, `https://www.math.utoronto.ca/roberth/D46/lecture_mar23.pdf`. The statement uses the usual admissible finite-energy functions and fixed boundary trace.

[23] MIT OpenCourseWare, 18.112, *Functions of a Complex Variable*, Lecture 16 (2008), introduces harmonicity and harmonic conjugates: `https://ocw.mit.edu/courses/18-112-functions-of-a-complex-variable-fall-2008/414780e0991d20cd338320d3784fae4d_lecture16.pdf`.

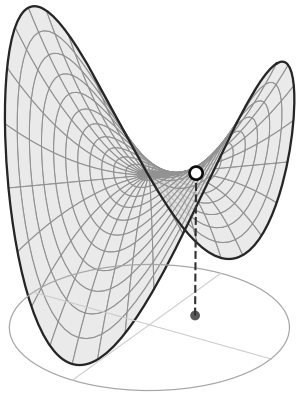
DAVIS GEOMETRIC / CHIHIRO CHI / HP-01

Two landscapes, one answer

F(z) = z² / same disk, camera and vertical scale

01 / REAL

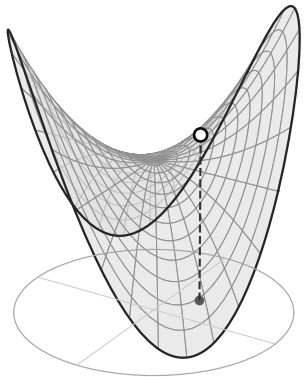
$$U = x^2 - y^2$$



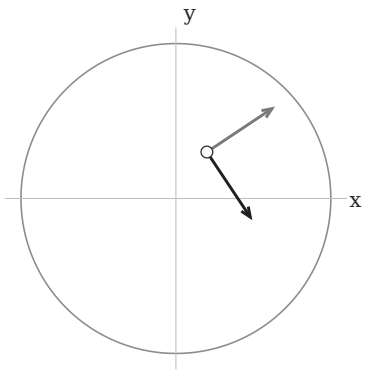
height −0.050

02 / IMAGINARY

$$W = 2xy$$



height +0.120



same input: 0.2 + 0.3i

03 / READ THE SLOPES

$$\nabla U = (0.4, -0.6)$$ dark arrow

$$\nabla W = (0.6, 0.4)$$ gray arrow

Equal length: √0.52 ≈ 0.7211.

Dot product: 0.4(0.6) − 0.6(0.4) = 0.

The uphill directions meet at 90°.

$$U_{xx} + U_{yy} = 2 - 2 = 0 \qquad W_{xx} + W_{yy} = 0 + 0 = 0$$

THE GEOMETRY OF FUEL CHI / HP-01

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 43. Two harmonic landscapes belonging to $F(z) = z^2$. Both use the same input disk, camera, and height scale. The marked input $0.2 + 0.3i$ gives heights -0.05 and 0.12. The plan view shows the two gradients at that input; their lengths and right angle follow from the displayed derivatives.

pairing and recovers the conjugate. It is a different input from the earlier completion question, where a few complex observations left a whole disk of possible answers.[24]

At our same input, $0.2 + 0.3i$, the wine fingerprint gives $F_{\text{wine}} \simeq -0.001308978 + 0.000052343i$, with $F'_{\text{wine}} \simeq -0.012377215 + 0.006311034i$. The two gradients are consequently approximately

$$\nabla U = (-0.012377215, -0.006311034),$$
$$\nabla W = (0.006311034, -0.012377215).$$

There is that quarter-turn again. We have changed the coefficients from a simple square to a fingerprint built from actual records, and the derivative relationship survives. These values can be recomputed directly from the printed histogram. We are examining a geometry we constructed from the distribution, with the construction visible. For the earlier completion problem, the Pick matrix P encodes whether the supplied complex readings fit one holomorphic function bounded by one on the disk. Its test $c^*Pc \geq 0$ must hold for every complex coefficient vector c. That is a compatibility test for the readings, rather than an energy assigned to a plasma movement.

My Crouzeix work meets this discussion at another joint: the quadratic form that checks an entire family of inputs. There A is a complex matrix, ϕ maps the interior of its numerical range to the unit disk, and $T = \phi(A)$ is the corresponding matrix function. The numerical range collects the complex numbers x^*Ax for unit vectors x. A positive definite matrix G supplies a ruler: $\|x\|_G^2 = x^*Gx$. The star means conjugate transpose. With this ruler, the identity is

$$x^*(G - T^*GT)x = \underbrace{\|x\|_G^2}_{\text{size before}} - \underbrace{\|Tx\|_G^2}_{\text{size after}}.$$

If $G - T^*GT \succeq 0$, this difference is nonnegative for every input x. The symbol means positive semidefinite: no direction gives a negative quadratic value. In plain English, I have chosen a ruler and proved that this operation cannot enlarge any input measured with it. In the settings of my metric lemma, that certificate gives

$$\|p(A)\| \leq \sqrt{\operatorname{cond} G} \max_{z \in W(A)} |p(z)|.$$

Here p is any polynomial, $W(A)$ is the numerical range, and $\operatorname{cond} G$ is the largest eigenvalue of G divided by its smallest. The square root measures the price of returning to the original ruler. A condition number at most four gives the factor two sought in Crouzeix's inequality for that certified matrix.[25]

I do not need to call these calculations the same thing to see why they belong in the same story. Harmonicity balances neighboring values. Holomorphicity ties two such fields together

[24] GIGI, *Bounded Harmonic Pair Results* and *Bounded Harmonic Conjugate Recovery*, October 29, 2026; companion source files `docs/HARMONIC_CONJUGATE_RESULTS.md` and `src/geometry/harmonic_conjugate.rs` in the GIGI repository. The wine construction and full bin counts also appear in *The Geometry of Sameness*, expanded second edition, "Measured distributions."

[25] Davis, *Constructive Metric Bounds for Crouzeix's Conjecture*, "A metric controls all polynomial evaluations." The lemma assumes a numerical range with nonempty interior and the matrix spectrum strictly inside it, with ϕ conformal onto the disk. Its proof applies von Neumann's inequality after the metric change. The manuscript's constructive results specify the matrix families for which the metric is established.

through their slopes. The metric certificate limits how an operator changes size. Each gives me a relationship I can calculate and a way to check whether a proposed answer respects it.

Now return to CHIHIRO. Put the potential back, restore its zero endpoints, and compare movements of the same size. The energy can become negative. Its lowest eigenvector generally does not satisfy the harmonic averaging equation. It identifies the disturbance with the smallest full energy per unit squared size. The Dirichlet part shows where the cost of variation enters; the potential shows what else competes with it. We can watch that competition in a small matrix before returning to the plasma calculation.

We can work a complete example with only three interior points. Set $h = 1$, $f = 1$, and $V = \alpha$, a constant. This is a teaching example, not a tokamak simulation:

$$H(\alpha) = \begin{pmatrix} 2-\alpha & -1 & 0 \\ -1 & 2-\alpha & -1 \\ 0 & -1 & 2-\alpha \end{pmatrix}.$$

Its eigenvalues are $2 - \alpha - \sqrt{2}$, $2 - \alpha$, and $2 - \alpha + \sqrt{2}$. The first eigenvector is proportional to $(1, \sqrt{2}, 1)$. Substitution into the first row gives $(2-\alpha) - \sqrt{2}$; substitution into the middle row gives $-2 + (2-\alpha)\sqrt{2}$, which is the same eigenvalue times $\sqrt{2}$. The third row repeats the first. Thus

$$\alpha_* = 2 - \sqrt{2} = 0.585786\ldots$$

is the exact crossing. At $\alpha = 0.4$, the lowest eigenvalue is $0.185786\ldots$; at 0.8, it is $-0.214214\ldots$.

To read the matrix, take its first row as an instruction: multiply the first movement by $2 - \alpha$, subtract the second, and take nothing directly from the third. The middle row subtracts both neighbors and adds its own $(2 - \alpha)$ contribution. The vector $(1, \sqrt{2}, 1)$ says the middle movement is about 1.414 times either outer movement. “Proportional” means we may enlarge or shrink all three together and retain the pattern. The star on α_* marks the particular value where its eigenvalue reaches zero: solve $2 - \alpha - \sqrt{2} = 0$. That is the threshold we just computed, rather than a number guessed from a plot.

Picture three marks along a flexible strip, with the ends held still. The middle mark moves most in the easiest pattern. As the destabilizing term increases, the cost of that particular pattern reaches zero and then becomes negative. Nothing mysterious happened to the matrix. We can point to the entries that changed and the movement they affected. The plasma calculation has more points and varying coefficients, but this is the spectral question it asks.

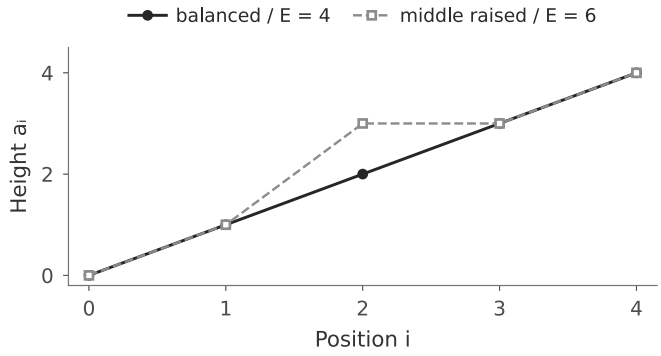
Now let both inputs move. A single curve tells us what happens along one chosen route. The landscape lets us see the routes we did not take. On its floor, one direction is magnetic shear s, and the other is the pressure-gradient parameter α. At every pair, CHIHIRO builds the matrix again. The height is its smallest eigenvalue. We are looking at a surface of calculated answers, not the physical surface of the plasma.

Think of the zero contour as a shoreline. A route through this landscape can cross it, enter a negative region, and come back out. That is why the second crossing matters. If I only handed you the first crossing, you would not be able to see the rest of that journey. The plan view and the perspective view show the same values; one makes the boundary easy to trace, and the other makes the changing height easier to picture.

Balance, pairing, a guarantee

Three questions connected by explicit calculations

01 / FOUR GAPS, FIXED ENDPOINTS



Balanced steps

1, 1, 1, 1

Cost: 1+1+1+1

Raised-middle steps

1, 2, 0, 1

Cost: 1+4+0+1

02 / THE WINE FINGERPRINT AT THE SAME INPUT

Quantity	Real part	Imaginary part
F	-0.001308978	0.000052343
F′	-0.012377215	0.006311034
F″	-0.083621857	0.070958944

6,497 records → 32 bin proportions → one analytic polynomial.

03 / KEEP THE CERTIFICATES DISTINCT

CHIHIRO	$u^T H u$	Energy of a disturbance
Wine completion	$c^* P c \geq 0$	Compatibility of analytic readings
Crouzeix metric	$x^* (G - T^* G T) x \geq 0$	Size cannot increase in the G ruler

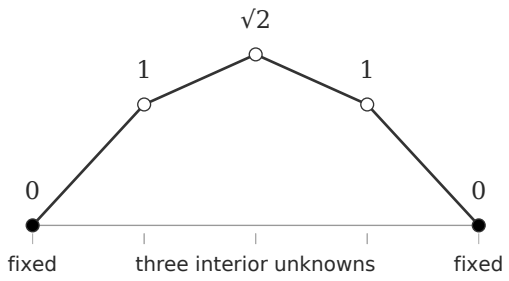
Figure 44. Harmonic balance, analytic pairing, and a metric guarantee. The chain has minimum cost four. Wine values come from the 32-bin alcohol histogram. Each calculation retains its own inputs and the property it certifies.

DAVIS GEOMETRIC / CHIHIRO CHI / OP-01

The operator selects a mode

A printable model beside the resolved calculation

01 THREE INTERIOR POINTS



02 EXACT MATRIX / UNIT SPACING

$$H = \begin{bmatrix} 2-\alpha & -1 & 0 \\ -1 & 2-\alpha & -1 \\ 0 & -1 & 2-\alpha \end{bmatrix}$$

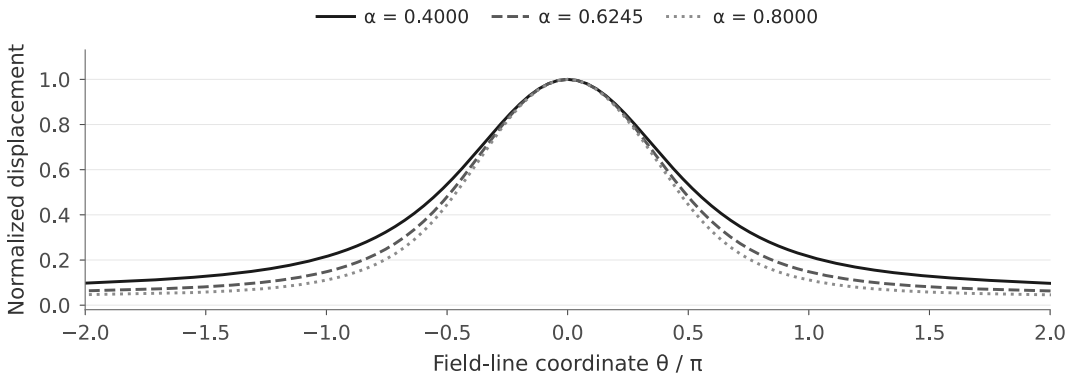
$$\lambda_{\min} = 2 - \alpha - \sqrt{2}$$

Threshold: α = 0.585786...

Eigenvector: (1, √2, 1)

03 RESOLVED MODE / 301 INTERIOR POINTS

Circular model: s = 1, κ = 1, δ = 0; fixed ends at ±6π.



04 MODE REGISTER / SAME DISCRETIZATION

α	Smallest eigenvalue	Normalized peak
0.400000	0.212013	1
0.624495	≈ 0	1
0.800000	-0.153711	1

THE GEOMETRY OF FUEL CHI / OP-01

Geometry · operators · recorded results B&W / 16 NOV 2026

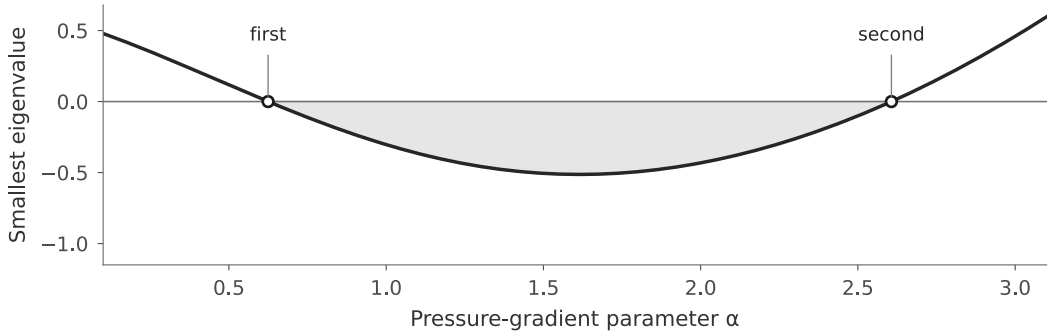
Figure 45. The exact three-point displacement beside the resolved model. The small matrix has unit spacing and constant coefficients; the 301-point calculation uses the field-line coefficients printed in the text. Their thresholds belong to their respective operators.

DAVIS GEOMETRIC / CHIHIRO CHI / OP-03

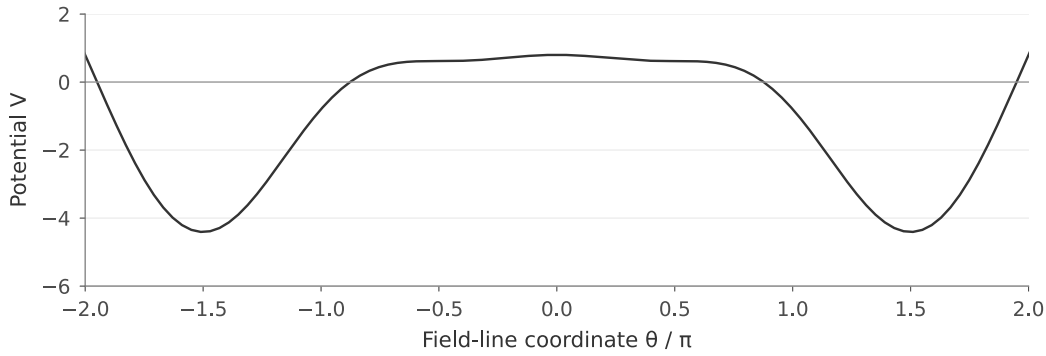
Follow the sign through the model

The coefficients, the energy and the zero crossings

01 SPECTRAL SWEEP / CIRCULAR MODEL



02 THE POTENTIAL TERM / α = 0.8



03 ENERGY / FIXED ENDPOINTS

$$\lambda = \frac{\int [f(\theta)(u')^2 - V(\theta)u^2]\, d\theta}{\int u^2\, d\theta}$$

Neighboring points resist different displacements through f = 1 + Λ².
The potential enters with a minus sign; retain both terms.

THE GEOMETRY OF FUEL CHI / OP-03

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 46. Following the sign through a fixed circular-model calculation. The sweep crosses zero twice. The potential panel enlarges the central field-line window at $\alpha = 0.8$.

For these two new sweeps, the current CHIHIRO solver evaluated $61 \times 81 = 4{,}941$ pairs for each shape, with $s \in [0.1, 2.5]$, $\alpha \in [0.01, 5]$, and the same 301-point field-line discretization. Circular and D-shaped coefficients share the same grid and boundary conditions. We can therefore follow how the zero contour changes when those shape coefficients change.[26]

The symbol $\in$ means "lies in," and the square brackets include both ends of each stated range. We choose 61 shear values and pair each with 81 drive values, producing 4,941 matrix calculations per shape. Each point on the landscape is one of those answers. The surface connects them so that we can see how neighboring answers relate to one another.

§ 13.4

The shape experiment, one boundary at a time

Now put the actual retained records on the page. Experiment 12 used $N = 301$, $T = 6\pi$, and four pairs (κ, δ): circular $(1, 0)$, elongated $(1.7, 0)$, D-shaped $(1.7, 0.33)$, and negative triangularity $(1.7, -0.3)$. The program scanned α from 0.01 to 6 at 60 trial values, located sign changes, and refined them. The first positive-to-negative crossing defines $\alpha_1(s)$. A subsequent negative-to-positive crossing, when found in the scan, defines $\alpha_2(s)$.

s	Circular	Elongated	D-shaped	Neg.-tri.
0.2	0.354117	0.136652	0.133510	0.138398
0.3	0.353500	0.162113	0.158120	0.164644
0.4	0.372683	0.179812	0.175159	0.183009
0.5	0.403096	0.195467	0.190136	0.199342
0.7	0.482067	0.228146	0.221135	0.233661
1.0	0.624495	0.284271	0.273990	0.292948
1.3	0.783181	0.347703	0.333557	0.360148
1.5	0.895132	0.393211	0.376269	0.408408
2.0	1.190722	0.515935	0.491492	0.538574
2.5	1.503277	0.648941	0.616544	0.679554

First critical pressure-gradient parameter α_1. All 40 retained values are shown; rounding follows the archive.

The four outlines in the accompanying plate give us something to picture while the parameters change. Elongation stretches the cross-section; triangularity changes its D-like shape. The equations above tell us exactly how those choices enter this reduced experiment. We can hold the numerical setup fixed, change the shape parameters, and follow where each boundary moves.[27]

[26] The visual questions follow Davis, *The Spectral Geometry of Plasma Confinement*, Experiments 2 and 12. These landscapes are new executions of `chihiro-core/src/cht.rs`, with independent symmetric tridiagonal eigenvalue checks. Their coefficients are those displayed in this chapter.

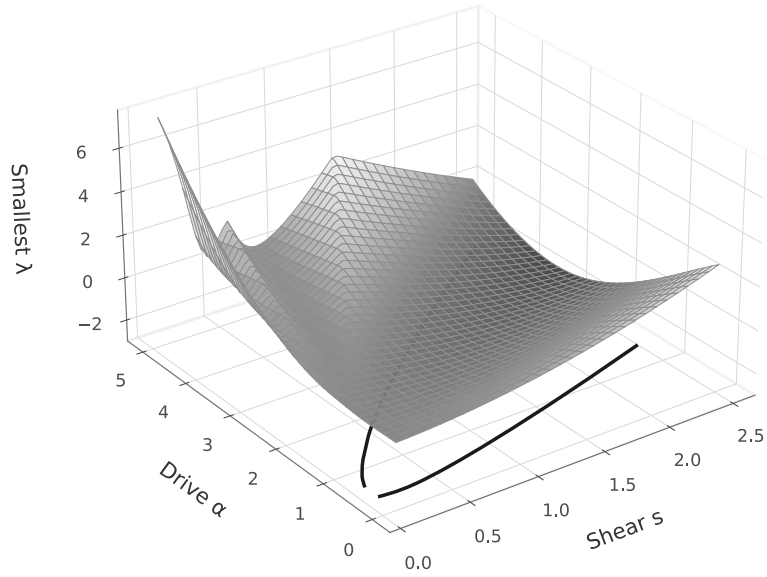
[27] Davis (2026), *The Spectral Geometry of Plasma Confinement*, v5, Experiment 12; `experiment12_superfluid_plasma.py` and `exp12_results.json` in the companion repository. The independent SciPy replay reproduces 40 first crossings within 6×10^{-6} of the rounded archive. Evaluation of all 40 first and 37 second boundary points in the Rust solver gives a largest absolute residual eigenvalue of 1.23×10^{-7}. Four direct solver comparisons agree within 9.53×10^{-9}.

DAVIS GEOMETRIC / CHIHIRO CHI / LS-01

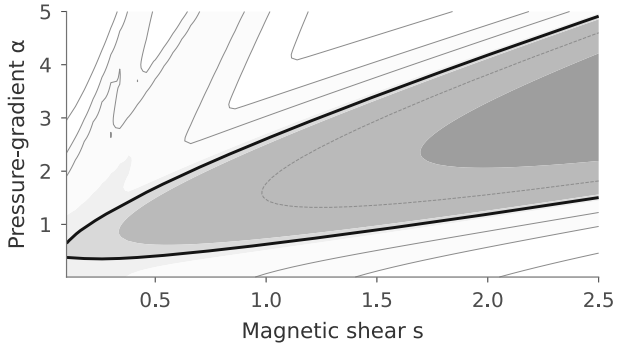
The circular stability landscape

Corrected CHIHIRO solver / 4,941 parameter pairs

01 THE SPECTRAL SURFACE / HEIGHT = SMALLEST EIGENVALUE



02 PLAN / THE SAME ZERO CONTOUR



03 READING THE SURFACE

κ = 1; δ = 0

61 shear values

81 drive values

301 interior points

θ from −6π to +6π

Dark bands: λ < 0.
Heavy contour: λ = 0.
Light bands: λ > 0.

THE GEOMETRY OF FUEL CHI / LS-01

Geometry · operators · recorded results B&W / 16 NOV 2026

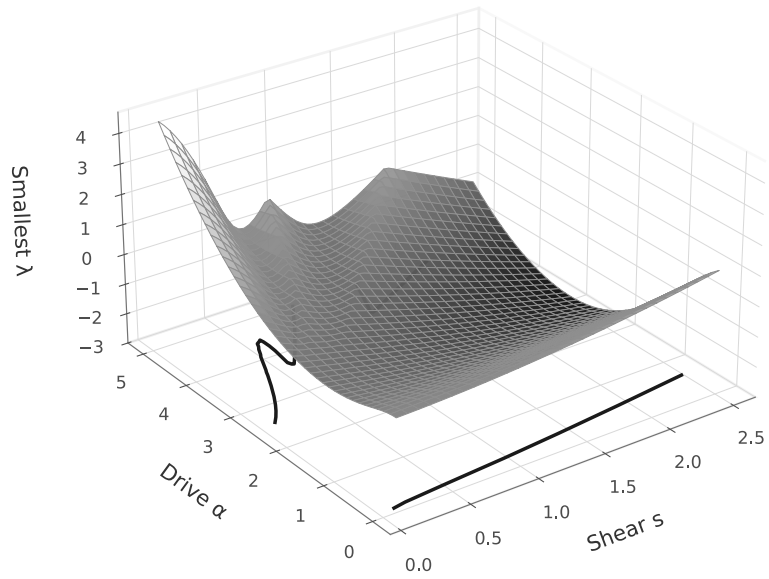
Figure 47. The circular model as a spectral landscape. Height is $\lambda_{\min}$; the curve on the base plane projects its zero contour. The contour identifies marginality for this finite operator.

DAVIS GEOMETRIC / CHIHIRO CHI / LS-02

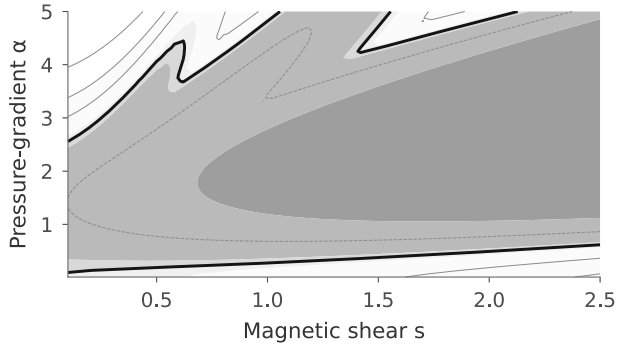
Reshaping the stability landscape

Corrected CHIHIRO solver / 4,941 parameter pairs

01 THE SPECTRAL SURFACE / HEIGHT = SMALLEST EIGENVALUE



02 PLAN / THE SAME ZERO CONTOUR



03 READING THE SURFACE

κ = 1.7; δ = 0.33

61 shear values

81 drive values

301 interior points

θ from −6π to +6π

Dark bands: λ < 0.

Heavy contour: λ = 0.

Light bands: λ > 0.

THE GEOMETRY OF FUEL CHI / LS-02

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 48. The same sweep with $\kappa = 1.7$ and $\delta = 0.33$. The common axes and domain make the change in the calculated landscape directly visible.

At $s = 1$ in the circular case, the first crossing is $\alpha_1 = 0.624495$ and the second is 2.606629, to the printed precision. The lowest eigenvalue is negative between these crossings in the scanned model. A report that kept only the second crossing would hide the earlier unstable interval. Conversely, a report that assumed increasing pressure must always worsen this one diagnostic would miss its return to a positive value. The boundary and the route to it both belong in the record.

The ratio I investigated was

$$C(s) = \frac{\overbrace{s}^{\text{shear}}}{\underbrace{\alpha_1(s)}_{\text{first boundary at that shear}}}.$$

The parentheses in $C(s)$ tell us that the answer depends on which shear we choose. The subscript 1 distinguishes the first crossing from the second; it is not a power. We take one shear value, find its first boundary, and divide. Repeating that operation produces the ratio curve. Its meaning stays attached to that denominator.

For the circular row at $s = 1$, it is $1/0.624495 \simeq 1.60129$. For the D-shaped row it is $1/0.273990 \simeq 3.64977$. The latter is larger because its first critical pressure-gradient parameter is smaller at the same shear. It does *not* mean that the second machine can hold more pressure. The quantity is a ratio of two model coordinates, not a reactor score.

For each shape I also examined the five saved shear values $1, 1.3, 1.5, 2, 2.5$. Call the five resulting ratios C_j, where j labels the saved value, and call their average $\bar{C}$, "C-bar": add the five ratios and divide by five. The *population standard deviation* describes the spread of this complete set of five values:

$$\sigma_C = \sqrt{\frac{\underbrace{\sum_{j=1}^{5}(C_j - \bar{C})^2}_{\text{total squared departures from the mean}}}{5}}.$$

Take each value's distance from the mean, square it so that departures on opposite sides cannot cancel, add the five squares, and divide by five. The square root puts the answer back on the scale of C. σ is pronounced "sigma." The *coefficient of variation*, $100\sigma_C/\bar{C}$, expresses that spread as a percentage of the mean. I want to know how much these five answers move around, not just whether the line looks flat to my eye. The results are:

Shape	Mean C	SD	CoV (percent)
Circular	1.655922	0.028306	1.71
Elongated	3.760044	0.129787	3.45
D-shaped	3.931551	0.153462	3.90
Neg-tri	3.617680	0.107413	2.97

This is the finite interval where the curves looked comparatively flat. The percentages describe variation *between those five modeled values*. They are not experimental measurement errors, confidence intervals, or proof of constancy at all shear values. The whole ten-point sweep varies much more. Keeping the interval in the sentence is part of keeping the result intact.

DAVIS GEOMETRIC / CHIHIRO CHI / SH-01

Four shapes, one controlled sweep

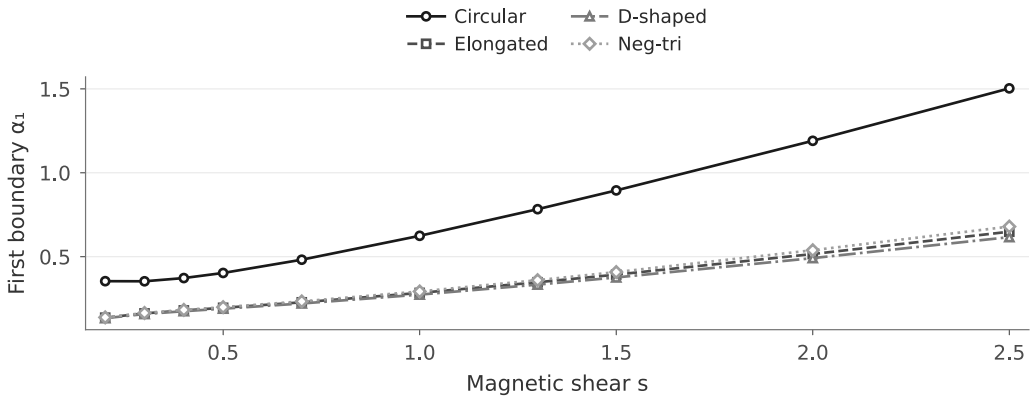
Forty archived records / first-boundary comparison

01 SHAPE REGISTER / PARAMETRIC SECTIONS



Circular	Elongated	D-shaped	Neg-tri
1 / 0	1.7 / 0	1.7 / 0.33	1.7 / -0.3

02 FIRST ZERO / TEN SHEAR VALUES



03 AT s = 1 / READ THE BOUNDARY BEFORE THE RATIO

Shape	α_1	s / α_1
Circular	0.624495	1.601295
Elongated	0.284271	3.517770
D-shaped	0.273990	3.649764
Neg-tri	0.292948	3.413580

THE GEOMETRY OF FUEL CHI / SH-01

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 49. Forty archived first-boundary records. The cross-sections label the four choices of (κ, δ). Points are saved calculations; connecting lines guide the eye.

The first-boundary comparison gives us one edge of each negative interval. The next plate restores the other edge wherever the saved scan found it. Follow a vertical line at one shear value: the lower crossing takes us into the negative interval, and the upper crossing takes us out again. Keeping both is how we keep the actual shape of the question. The blank parts leave missing second crossings visible.

There is a small theorem hiding in the next question, and we can prove it without pretending the numerical plateau is an exact law. Suppose a positive boundary $a(s)$ is differentiable on an interval with $s > 0$. Here $a(s)$ names the boundary curve; it is not the minor radius a used in the plasma dimensions. "Differentiable" means the curve has a well-defined local slope. We now ask how the ratio changes as we move along that curve. The quotient rule gives

$$\frac{d}{ds}\frac{s}{a(s)} = \frac{a(s) - sa'(s)}{a(s)^2}.$$

$a'(s)$, "a-prime," is that slope. The denominator is positive, so the numerator decides whether the ratio rises or falls. The $a(s)$ term comes from increasing the numerator s; the $sa'(s)$ term accounts for the boundary changing underneath it. The two changes compete. A *stationary point* is a place where the ratio's instantaneous rate of change is zero, so these terms must balance.

At any interior stationary point,

$$\underbrace{a'(s_*)}_{\text{tangent slope}} = \underbrace{\frac{a(s_*)}{s_*}}_{\text{slope from the origin}}.$$

Draw the boundary in an (s, a) graph. The line from the origin to the point has slope a/s. At a stationary ratio, that line has the same slope as the boundary's tangent. That is the geometric statement. It does not, by itself, say the stationary point is a maximum or that there is only one.

For a second-derivative test, assume a is twice differentiable. At a stationary point,

$$C''(s_*) = -\frac{s_* a''(s_*)}{a(s_*)^2}.$$

Two primes mean the rate at which the slope itself changes. At this stationary point, all the factors apart from the leading minus sign and a'' are positive. If the boundary bends upward there ($a'' > 0$), the ratio bends downward ($C'' < 0$): its flat spot is the top of a local hill. "Local" means near this point, not necessarily higher than every value elsewhere on the curve. If the second derivative is zero, this test alone does not settle the shape.

Thus $a''(s_*) > 0$ makes it a strict local maximum. For uniqueness on an interval, a sufficient stronger condition is $(\log a)'' \geq 0$, since

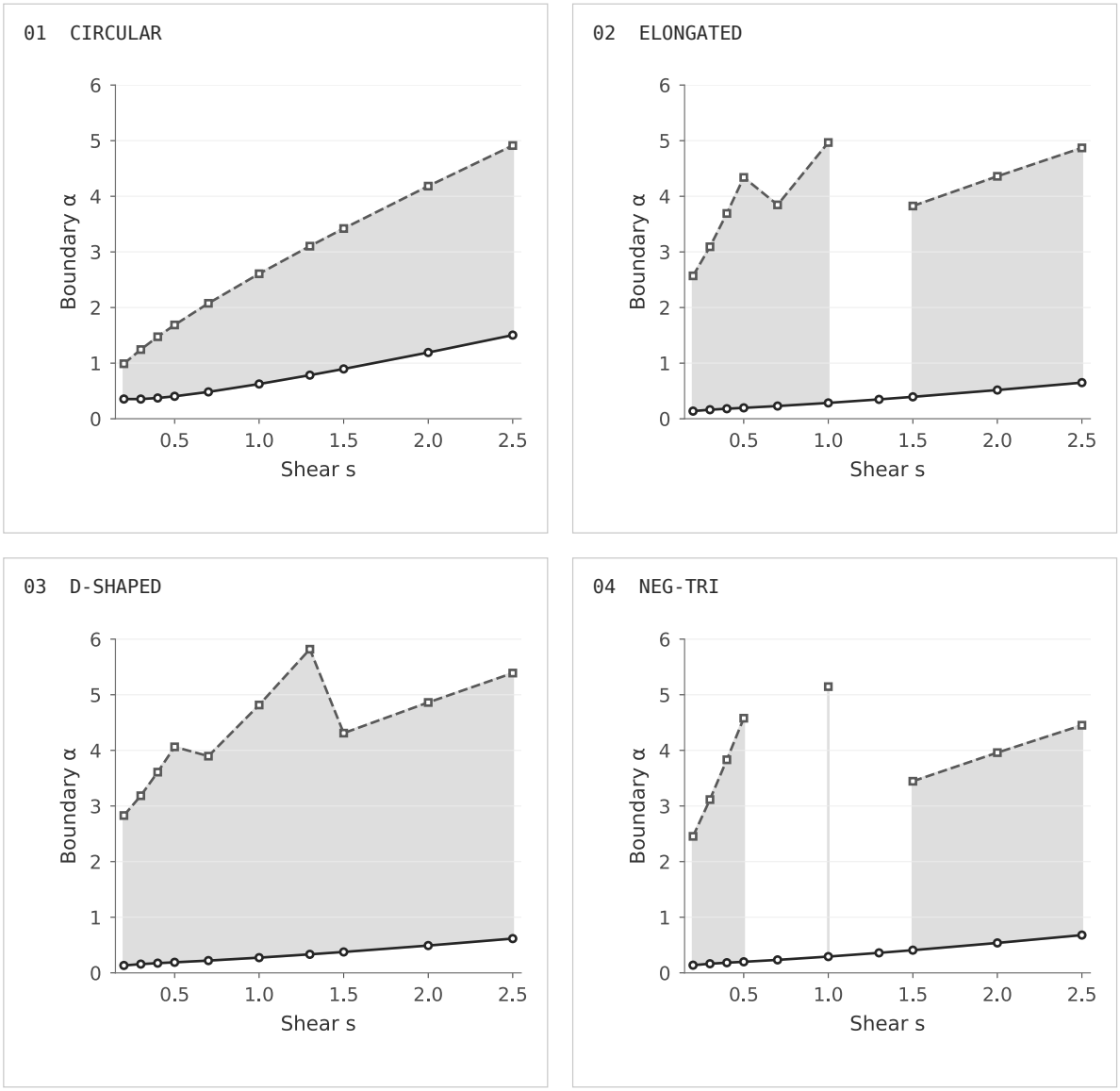
$$(\log C)'' = -\frac{1}{s^2} - (\log a)'' < 0.$$

Here log is the natural logarithm. It turns the positive ratio $C = s/a$ into $\log C = \log s - \log a$, which makes the two contributions easier to differentiate. The condition $(\log a)'' \geq 0$ says that the logarithm of the boundary bends upward or stays locally straight. The displayed minus signs then force $\log C$ to bend downward everywhere on the interval. Its slope is strictly decreasing, so it can

DAVIS GEOMETRIC / CHIHIRO CHI / BW-01

Both boundaries belong in the picture

First zero, negative interval, second zero / archive 12



Circles: first boundary. Squares: second boundary. Shading: negative interval.
A missing second boundary is left unfilled; no crossing is invented.

THE GEOMETRY OF FUEL CHI / BW-01

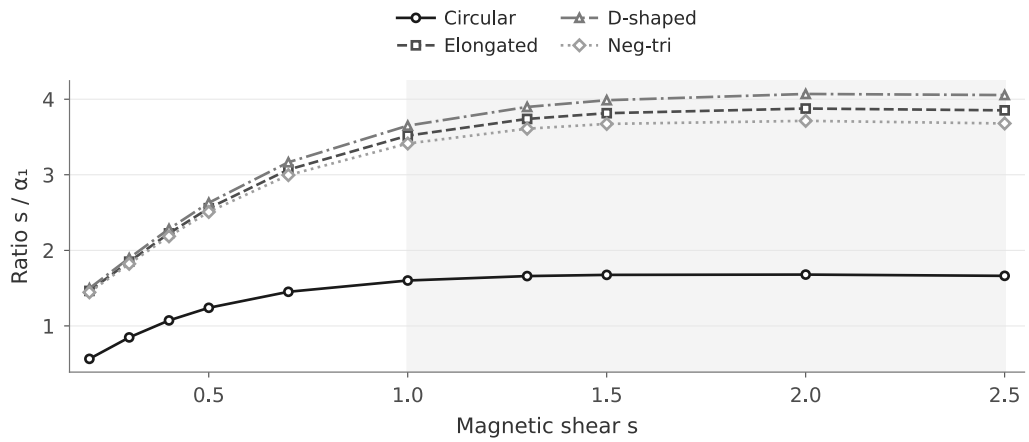
Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 50. Both archived boundaries. Circles mark first crossings, squares mark second crossings, and the shaded band joins the sampled negative intervals. Missing second crossings are not filled in.

The ratio retains its denominator

The same forty records / C(s) = s / α₁(s)

01 FINITE-SHEAR COMPARISON



02 SHADED INTERVAL / FIVE SAMPLES PER SHAPE

Shape	Mean C	Across-shear SD
Circular	1.655922	0.028306
Elongated	3.760044	0.129787
D-shaped	3.931551	0.153462
Neg-tri	3.617680	0.107413

03 WHAT A STATIONARY RATIO REQUIRES

$$C'(s) = \frac{\alpha_1 - s\alpha_1'}{\alpha_1^2} = 0 \quad \Longrightarrow \quad \alpha_1' = \frac{\alpha_1}{s}$$

At a stationary ratio, the tangent has the slope of the ray from the origin.
These are sampled model values; the shaded interval defines the summary.

Figure 51. The same first-boundary records after forming $C = s/\alpha_1$. Shading identifies the five-sample interval used for the table; its standard deviation describes variation across those samples. A larger ratio is not a universal stability ranking.

cross zero at most once. Because $C > 0$, its stationary points are the same as those of $\log C$. That is why this additional assumption gives uniqueness.

This is concavity as a function of s. It supplies at most one stationary point on that interval. Existence still needs an endpoint condition or a sign change of C'. A finite numerical scan does not prove these hypotheses on an unbounded domain.

In the circular replay, a local search gives $s_* \simeq 1.785038$, $a(s_*) \simeq 1.061230$, and $C(s_*) \simeq 1.682046$. A centered difference with step 0.001 gives $a' \simeq 0.5945143$, while $a/s \simeq 0.5945140$. The two slopes agree to about 3×10^{-7} in this check. For the centered difference, we evaluate the boundary a small step to either side and divide the change by the full distance between those two positions. A step of 0.001 therefore gives $[a(s + 0.001) - a(s - 0.001)]/0.002$ as an approximation to the slope. The difference 3×10^{-7} is about 0.0000003. We have compared two ways of obtaining a slope at this location: neighboring boundary values, and the ray from the origin. Their agreement is what connects this numerical check to the preceding calculation.

The calculated boundary meets the tangent condition at that local stationary point. We can then investigate the rest of the curve and check how the answer changes with numerical resolution.

We can also ask how much the answer moves when the numerical problem changes. For the circular case at $s = 1$:

Interior points	Half-domain T	First crossing
301	6π	0.62449464
603	6π	0.62468410
1207	6π	0.62473145
603	12π	0.61620756

The first three rows halve the spacing twice on the same domain. The last doubles the domain while retaining the original spacing. Extending the interval changes the boundary by about 1.33% relative to the original calculation. Refining the grid from 301 to 1207 points changes it by about 0.038%. Those are different questions. Root-finding precision, grid resolution, and domain size each tell us something different about the answer. The table lets us see which change matters most here.

In ordinary language: I asked the computer to look more closely inside the same window, and then I asked it to open a larger window. The second change mattered more here. We keep the saved observation, and we keep what it depends on. That is how an interesting pattern becomes a research question we can actually work on.

§ 13.5

The calculation behind the equilibrium label

The next investigation tried to connect those local inputs to a whole radial profile. Pressure and current now enter the story, rather than choosing s and α independently. Experiment 13 contains a FreeGS route and a fallback route. FreeGS is software for solving the Grad–Shafranov equilibrium problem: the force-balance equation for an axisymmetric magnetic configuration in the static

ideal-MHD description.[28] The second route starts with prescribed analytic profiles. That is the route we can reproduce from the saved records: all 24 cases return the same stable-location counts and minimum eigenvalues, with the largest absolute eigenvalue discrepancy below 7×10^{-12}. We will walk through that calculation as the *prescribed-profile replay.*[29]

Here are the inputs needed to follow it. Let $x = r/a$ be normalized radius. The major radius is $R_0 = 6.2\,\mathrm{m}$, the minor radius $a = 2\,\mathrm{m}$, the field $B_0 = 5.3\,\mathrm{T}$, and the shaping parameters are $\kappa = 1.7$, $\delta = 0.33$. The source prescribes the following profiles. A *profile* is simply the value of a quantity as we move through position. Dividing the local radius r by the minor radius a makes $x = 0$ the center and $x = 1$ the outer surface of this model. p_0 names central pressure; q_0 and q_a name winding at the center and edge. Those subscripts are labels. With these meanings in place,

$$\begin{aligned} p(x) &= p_0(1 - x^2)^2, \\ q(x) &= q_0 + (q_a - q_0)x^2, \\ s(x) &= \frac{2(q_a - q_0)x^2}{q_0 + (q_a - q_0)x^2}. \end{aligned}$$

At the center, the first line gives $p = p_0$; at the outer surface, it gives zero. The second line starts at q_0 and adds the fraction x^2 of the full center-to-edge change. Halfway out in radius, $x = 1/2$, it has completed one quarter of that winding change, not one half. The last line is the shear definition we already met, evaluated for this particular q. In normalized coordinates it is $s = (x/q)\,dq/dx$, and differentiating the second line gives $dq/dx = 2(q_a - q_0)x$. That is where the numerator comes from.

To follow these saved runs exactly, we also retain the source prescription for q_a and its constants.[30] The pressure-gradient parameter passed to the reduced operator is

$$\alpha(x) = q(x)^2 \frac{2\mu_0 p_0}{B_0^2} \frac{R_0}{a} \frac{4x(1 - x^2)}{\kappa}.$$

Read this product from left to right. $q(x)^2$ supplies the winding factor; $2\mu_0 p_0/B_0^2$ compares the central pressure with the magnetic pressure scale. R_0/a compares the ring's large radius with its cross-sectional radius. The factor $4x(1 - x^2)$ comes from differentiating the pressure profile: its outward slope is $dp/dr = -4p_0 x(1 - x^2)/a$. Finally, this source prescription divides by elongation κ. We have assembled the pressure-gradient input from the chosen profile, with each scale visible.

This explicitly includes the fallback's $1/\kappa$ factor. Once p_0, current, and the profile prescription are chosen, each sampled location supplies a particular pair (s, α).

The program starts with 30 equally spaced values of x^2 from 0.05 to 0.95. Its radius and shear filters retain 28 locations. At each, the local operator returns a lowest eigenvalue. The reported

[28] Freidberg (2007), lecture 5, develops the Grad–Shafranov equation; the FreeGS developers' "Creating equilibria" documents its numerical inputs and equilibrium solve.

[29] Companion files `experiment13_freegs.py` and `exp13_results/results.json`. The archived summary does not distinguish its FreeGS and fallback routes; the analytic route reproduces the saved numerical results.

[30] With the numerical SI inputs used by the source, $q_a = 5a\kappa B_0(2\pi)/(R_0\mu_0 I_p)$ and $q_0 = \max(0.8, 0.3q_a)$. This is the archived fallback's imposed profile recipe. With the displayed quantities carrying SI dimensions and the prefactor 5 treated as a pure number, its right-hand side has units of inverse length, whereas a safety factor is dimensionless. It is therefore retained as a numerical recipe for reproducing the archive, not a dimensionally derived safety-factor formula. Its length scaling and prefactor need reconciliation before physical calibration.

minimum is the smallest of those 28 values. Despite the source field name `spectral_gap`, this is not the eigenvalue of a single globally coupled radial plasma operator. It is a minimum over separate local calculations.

Pressure (kPa)	5 MA	10 MA	15 MA
50	0/28	28/28	28/28
100	0/28	5/28	28/28
200	0/28	0/28	7/28
500	23/28	0/28	0/28
1000	28/28	6/28	0/28
2000	28/28	23/28	0/28
3000	28/28	27/28	18/28
5000	28/28	28/28	24/28

An entry such as 23/28 means 23 sampled locations had positive lowest eigenvalues in this model. It does not mean an 82% chance of safe operation. The other five locations matter. Nor does 28/28 test every possible radial location or every kind of disturbance.

At 15 MA and $p_0 = 2000\,\mathrm{kPa}$, we can follow the pressure normalization explicitly. The profile's circular area-weighted average is

$$\langle p \rangle = 2 \int_0^1 p_0 (1 - x^2)^2 x \, dx = p_0/3.$$

Angle brackets around a single p now mean its spatial average. The weight $2x\,dx$ represents a thin radial shell's share of the total in this averaging model. Shells farther out contain more area, so we do not give every radius equal weight. Adding the weights from zero to one gives one. To check the integral, set $z = 1 - x^2$, so $dz = -2x\,dx$; reversing the limits leaves $p_0 \int_0^1 z^2\,dz = p_0/3$. A central pressure of 2000 kPa therefore gives an average of about 666.7 kPa in this profile.

That is the averaging approximation retained by the source. With $\mu_0 \simeq 4\pi \times 10^{-7}$ in SI units, it gives

$$\beta(\%) = 100 \frac{2\mu_0 (p_0/3)}{B_0^2} \simeq 5.964813,$$
$$\beta_N = \beta(\%) \frac{a[\mathrm{m}] B_0[\mathrm{T}]}{I_p[\mathrm{MA}]} \simeq 4.215135.$$

The percent sign and the named units matter: β_N here uses the conventional tokamak engineering normalization. It is not the local $\beta(r)$ used to define the pressure gradient.[31] The first line converts that averaged pressure into a percentage of the magnetic pressure scale. Its value 5.964813 means about 5.96 percent, not a pressure in pascals. The second line adjusts that percentage using the named size, field, and current convention. The square brackets specify the units of the numbers to enter: metres, tesla, and millions of amperes. Entering 15 million instead of 15 in the mega-ampere slot would change the answer by a million. β_N is useful precisely because we agree on the recipe; it does not mean the unscaled local pressure ratio β.

[31] Troyon et al. (1984), "MHD-Limits to Plasma Confinement," develops the pressure–current stability scaling. UKAEA's PROCESS documentation, "Plasma beta," states the percent/metre/tesla/mega-ampere convention and distinguishes alternative prescriptions for the limiting coefficient.

The saved minimum eigenvalue is −1.466256, and the count is 0/28. At the same current and 5000 kPa, the count rises to 24/28, but the minimum remains negative, −0.544530. Neither case establishes that this modeled profile is ballooning-stable everywhere.

A Troyon-type limit relates pressure, magnetic field, current, and size within a specified stability model; its coefficient is not a universal constant independent of profiles and boundary conditions. The old summary's broad statement about remaining stable past a Troyon limit therefore cannot be used as the conclusion of these rows. Some cases have all sampled values positive, others do not, and the result depends strongly on current and pressure. More fundamentally, a ballooning calculation cannot establish an external-kink boundary. They are different disturbance questions.

That distinction is useful to the software. I want to know which calculation returned a favorable answer, where another one disagreed, and what inputs each one used. Put those answers beside one another. Now somebody who knows the machine can ask a better next question, and I can follow how we got there.

§ 13.6

One configuration, several questions

An external kink involves displacement of the plasma boundary, rather than only the local ballooning question. Its stability depends on the surrounding magnetic region and wall response.[32] In classical resistive tearing, finite resistivity permits magnetic reconnection and the formation of islands.[33] A vertical displacement moves the plasma column up or down. In a wall of finite conductivity, induced currents can decay, so its response cannot be replaced by a permanently fixed magnetic boundary.[34] CHIHIRO asks these questions of the same configuration. How does it respond to a ballooning disturbance? What does the kink calculation say? What changes when we include tearing, vertical motion, the wall, density, and the edge? Those answers belong beside one another. I want to see where they agree and where they pull apart, because that difference tells me what to investigate next. The implementation runs seven diagnostic channels and collects their results in one report.[35]

Here a *channel* means one diagnostic calculation with its own question and result. Density is the number of particles per volume; the edge is the outer region of the plasma. In the figure labels, NTM abbreviates neoclassical tearing mode, and ELM abbreviates edge-localized mode. Those labels keep the different questions from being mixed together when their numerical results share a table.

Before sending a pressure profile through those channels, I want the input to mean what it says. Suppose I specify an average pressure. That is an average over the configuration, not the pressure

[32] Bondeson and Ward (1994), "Stabilization of external modes in tokamaks by resistive walls and plasma rotation," studies low-mode-number external modes; including the effects of rotation.

[33] Furth, Killeen, and Rosenbluth (1963), "Finite-Resistivity Instabilities of a Sheet Pinch."

[34] Freidberg, Cerfon, and Lee (2015), "Tokamak elongation: how much is too much? I Theory," develops the axisymmetric vertical-stability problem with a resistive wall and feedback response.

[35] Companion implementation: `chihiro-core/src/combined_diagnostic.rs`; the individual modules are `diagnostic`, `dcon`, `ntm`, `vde`, `rwm`, and the density and edge checks in the combined module. The runtime comparison for this edition executes the combined report, extracts its seven-component fingerprint, and writes and reads the fingerprint store.

DAVIS GEOMETRIC / CHIHIRO CHI / PR-01

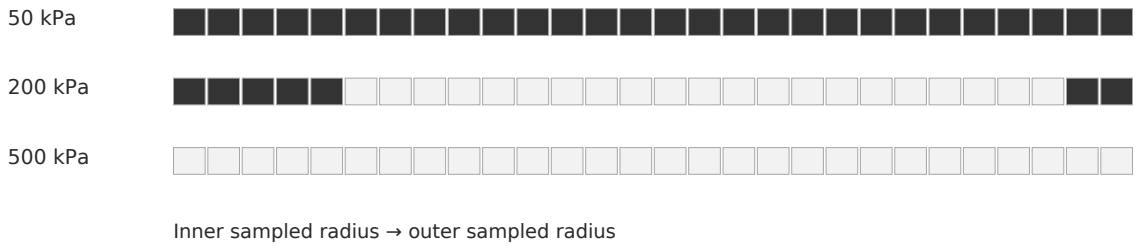
Every radial test stays in view

Archive 13 / prescribed profiles / 24 cases

01 LOCAL POSITIVE EIGENVALUES / OUT OF 28

p_0 / kPa	5 MA	10 MA	15 MA
50	0/28	28/28	28/28
100	0/28	5/28	28/28
200	0/28	0/28	7/28
500	23/28	0/28	0/28
1000	28/28	6/28	0/28
2000	28/28	23/28	0/28
3000	28/28	27/28	18/28
5000	28/28	28/28	24/28

02 THREE ROWS / I_p = 15 MA / ALL 28 LOCATIONS



Dark: positive local eigenvalue. Light: zero or negative.
Each square is one local operator; the row retains the spatial ordering.

THE GEOMETRY OF FUEL CHI / PR-01

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 52. All 24 prescribed-profile cases, preserving the 28-location counts. The table compares pressure and current; the three strips show individual radial tests at 15 MA. Darker table cells indicate more positive local results, not a probability of safe operation or a global stability certificate.

at its center. A peaked profile can have a high central value and a much smaller average. We need the conversion before we interpret its result.

For the current analytic preset, write

$$p(x) = \underbrace{p_0}_{\text{central pressure}} \qquad \underbrace{(1-x^2)^{a_p}}_{\text{radial profile factor}} \quad , \qquad x = r/a.$$

The exponent $a_p \geq 0$ controls the profile's shape. It is not the radius a, and it is not the pressure-gradient input α. For positive a_p, the factor starts at one in the center and reaches zero at the outer surface. Larger exponents make it fall more rapidly away from the center. At $a_p = 0$, we use the constant profile. The earlier archive used exponent two; the current preset uses 1.5. We can now average either one using the same calculation.

For concentric cross-sections with a constant shape factor, the normalized radial volume element is $2x\,dx$. The average is

$$\begin{aligned} \langle p \rangle &= 2\int_0^1 p_0(1-x^2)^{a_p} x\,dx \\ &= p_0 \int_0^1 u^{a_p}\,du = \frac{p_0}{a_p+1}, \qquad u = 1-x^2. \end{aligned}$$

The first line adds the pressure in each shell with its volume weight. In the second line, u is a temporary integration variable, not the displacement pattern from the stability equation. Replacing $1-x^2$ by u and using $du = -2x\,dx$ removes the extra radial factor; reversing the limits removes the minus sign. The remaining integral is $1/(a_p+1)$. So division by a_p+1 converts the peak into the average. Multiplication by that same number takes us back to the peak required for a chosen average. We can check both directions before running the solver.

Thus $p_0 = (a_p+1)\langle p \rangle$, and the central beta is $\beta_0 = (a_p+1)\langle \beta \rangle$. The current preset uses $a_p = 1.5$, so the central value is 2.5 times the average.

Think of the profile as a hill. Naming its height at the top does not tell us its average height. A narrow hill and a broad hill can reach the same peak and contain different amounts underneath them. Here we have the whole curve, so we can do the integral. Then we send that profile into the calculation knowing that the amount we asked for is the amount the profile contains.

The following runs use the current ITER-like analytic preset: $R_0 = 6.2\,\text{m}$, $a = 2\,\text{m}$, $B_0 = 5.3\,\text{T}$, $I_p = 15\,\text{MA}$, $\kappa = 1.7$, $\delta = 0.33$, $q_0 = 1.05$, $q_a = 3.5$, and both profile exponents equal to 1.5. They follow the current implementation after the historical replay above. Each ballooning profile samples 50 radial locations, with 301 interior field-line points on $[-6\pi, 6\pi]$.

That is a list of things held fixed, not eight new quantities to interpret at once. The first four specify the ring's large radius, cross-sectional radius, magnetic field, and plasma current. The next two specify the cross-section's shape. The final pair gives the winding at the center and edge. We change the requested normalized pressure below while keeping that configuration in place.

Requested β_N	Integrated β_N	Smallest λ	First crossing x
0.1	0.1000000002	0.006988	—
0.5	0.5000000011	-0.463141	0.445072
1.8	1.8000000038	-1.301138	0.152510

The second column comes from independently integrating the pressure profile with 10,000 midpoint samples. The last column locates the first positive-to-nonpositive eigenvalue crossing by interpolation between sampled radii; the dash means that no such crossing occurs in the sampled profile. At $\beta_N = 0.5$, the crossing is at about 44.5% of the minor radius. Raising the input pressure moves that first crossing inward in these runs. The full radial profile stays with the result, so we can return to the locations that produced it.

For each row, read across as a short investigation: this is the pressure I requested; this is what the profile actually integrates to; this is the lowest eigenvalue found among its sampled radii; and this is where the first sign change occurs. *Interpolation* estimates the crossing between two sampled positions by joining their values. The minimum and the crossing need not occur at the same radius. One tells us how low the curve gets; the other tells us where it first reaches zero.

But there is another question hiding behind that lowest curve. What are the other eigenvalues doing? The paper's spectral-waterfall experiment asked exactly that. Here I return to the question with the current software and the corrected pressure profile.[36]

At each radius, sort the eigenvalues as $\lambda_0 \le \lambda_1 \le \cdots \le \lambda_{300}$. The waterfall draws the lowest ten of them. Each line follows the eigenvalue at that position in the ordering. A line may bend as the inputs change; its height still tells us the energy sign of one normalized displacement pattern. The smallest value tells us whether a negative direction exists. The other values tell us how much more is happening around it.

We can make that distinction exact. For the real symmetric matrix H, choose orthonormal eigenvectors v_j, and write a displacement as $u = \sum_j c_j v_j$. Then

$$\underbrace{u^{\mathsf{T}} H u}_{\text{energy of the whole pattern}} = \underbrace{\sum_j \lambda_j c_j^2}_{\text{add its mode contributions}},$$

$$n_-(H) = \#\{j : \lambda_j < 0\}.$$

v_j is the j-th unit-length eigenvector pattern, and c_j says how much of it is present in u. The sum assembles the movement from these ingredients. Their perpendicularity removes the cross-terms in the energy, leaving one $\lambda_j c_j^2$ contribution per ingredient. The # symbol means "count," and the braces collect the indices whose eigenvalues are negative. Thus $n_-(H)$ is an integer: how many negative directions this matrix contains, counting their independent eigenvector directions.

The span of the negative-eigenvalue vectors has dimension $n_-(H)$, and every nonzero vector in that span has negative quadratic energy. A larger subspace must intersect the span of the nonnegative-eigenvalue vectors, where that quadratic energy is nonnegative. Thus $n_-(H)$ is exactly the largest dimension of a subspace on which the form is negative definite. This is a count of independent negative directions, not a new kind of average.

The *span* is all the combinations we can make from a set of patterns; its dimension counts how many independent ingredients we need. "Negative definite on that subspace" means every nonzero combination there gives a negative energy. Why can no larger space have that property? Once we have more independent directions than there are negative-mode coordinates, some nonzero

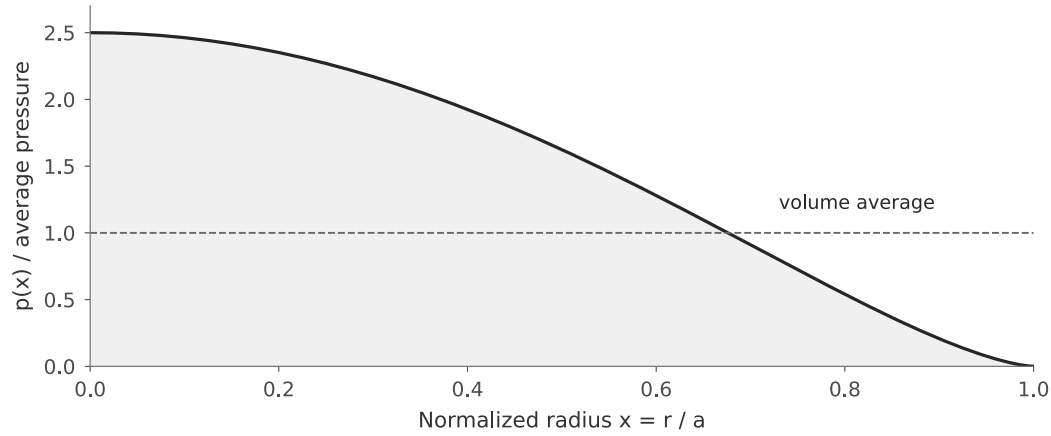
[36] Davis, *Spectral Geometry*, Experiment 6, supplies the spectral-waterfall and mode-transition studies. The plates here are new current-software runs, using `chihiro-core/src/cht.rs` and `chihiro-core/src/equilibrium.rs`; the original pressure-sweep outputs are not substituted for these runs.

DAVIS GEOMETRIC / CHIHIRO CHI / RT-01

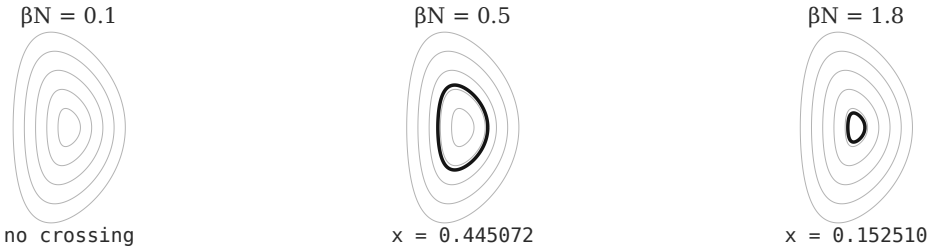
Pressure becomes a radial question

Corrected runtime / the three recorded input values

01 THE WHOLE PROFILE / $p(x) = p_0(1 - x^2)^{1.5}$



02 ONE SECTION / THREE REPORTED CROSSINGS



03 PRESSURE CHECK / INDEPENDENT QUADRATURE

Requested βN	Integrated βN	Central / average
0.1	0.1000000002	2.5
0.5	0.5000000011	2.5
1.8	1.8000000038	2.5

THE GEOMETRY OF FUEL CHI / RT-01

Geometry · operators · recorded results B&W / 16 NOV 2026

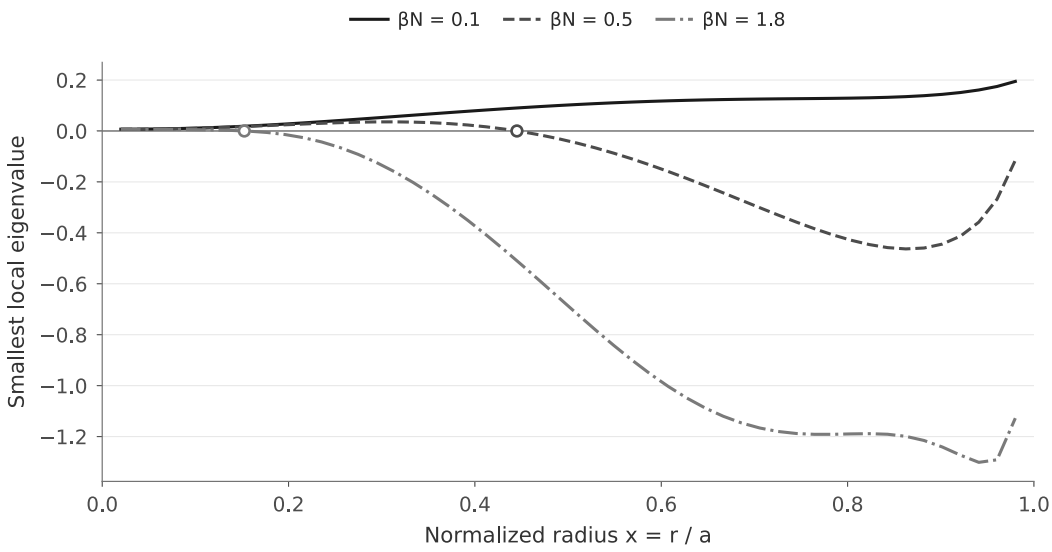
Figure 53. The corrected pressure profile and the locations selected by the three current runs. The dark contour in each section marks the reported first crossing; it is a coordinate illustration of the result.

DAVIS GEOMETRIC / CHIHIRO CHI / RT-02

Locate the crossing on the curve

50 radial locations / 301 field-line points per location

01 THE COMPLETE SAMPLED RADIAL PROFILES



02 CROSSING LOCATION / SPECTRAL MINIMUM

βN	First crossing x	Minimum λ
0.1	—	0.006988
0.5	0.445072	-0.463141
1.8	0.152510	-1.301138

A crossing is located between two neighboring samples of opposite sign.
The full profile stays with the answer: a crossing and a minimum
describe different features of the same calculation.

THE GEOMETRY OF FUEL CHI / RT-02

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 54. All 150 radial results, grouped by their input pressure. The crossing location and the minimum eigenvalue answer different questions about each curve.

combination must cancel all those coordinates. It leaves only zero or positive-mode contributions. That combination cannot have negative energy. This is what makes the count exact for the stated matrix.

In plain English: imagine that I have asked the model to try all the small movements it permits. One movement lowers the energy. That already gives me something to investigate. But three independent movements lowering it is a different description. If I store only the lowest number, I keep the strongest warning and lose the rest of that structure. I want to see the whole set of possibilities the calculation is showing me.

The three runs above make this visible. At $\beta_N = 0.1$, none of the 50 sampled radial operators has a negative eigenvalue. At 0.5, 28 do, with at most one negative direction per operator. At 1.8, 43 do, and the largest count at a single sampled radius is three. Those counts use all 301 eigenvalues, although the waterfall shows only the lowest ten. The radial strip keeps the location of each answer visible.

Now hold the radial location still, at $x = 0.85$, and change the pressure input instead. These are two different ways through the same model: the waterfall moves through radius at fixed pressure; this new plot stays at one radius while pressure changes. Fifty-nine inputs from $\beta_N = 0.1$ to 3.0, in increments of 0.05, show how the lowest five values move. The first zero at this location is near $\beta_N = 0.206484$, refined within the sampled bracket $[0.20, 0.25]$. The count beneath the curves reaches three and then falls to two near the upper end of the sweep.

Look at what we gained by drawing the extra lines. We can follow the first loss of positivity, see additional directions join it, and see one of those directions return above zero. The picture gives us a sequence of questions to ask the mathematics. That is what I want from a geometric description: enough structure to keep asking.

The two landscapes and these spectral sweeps produced 10,091 solver records. An independent tridiagonal calculation checked 11,972 displayed or mapped eigenvalues, with a largest absolute difference of 2.63×10^{-7}. The figure data retain the inputs and the results together so that each curve can be reconstructed.

That is a useful answer to a specific question. Now we can put it beside the other channels instead of asking one number to speak for the whole configuration. CHIHIRO's fingerprint records the ordered channel margins as

$$\mathbf{m} = (m_{\text{balloon}}, m_{\text{kink}}, m_{\text{tearing}}, m_{\text{vertical}}, m_{\text{wall}}, m_{\text{density}}, m_{\text{edge}}).$$

This bold $\mathbf{m}$ is an ordered list of seven numbers. Each subscript tells us which diagnostic supplied its entry. A channel margin is that channel's score under its implemented rule; we keep the rule and label attached rather than treating unlike scores as one physical quantity. The ballooning entry in these runs, for example, is the fraction of the 50 sampled locations with a positive lowest eigenvalue. The 0.44 entry comes from 22/50, and 0.14 comes from 7/50. This list lets us see which question returned which answer before looking at the aggregate color.

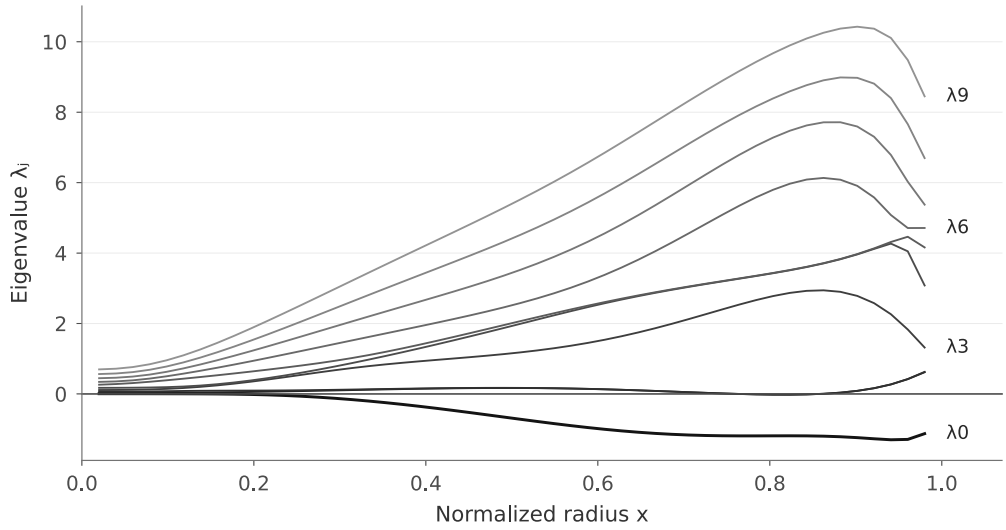
The record also carries its input parameters, aggregate status, and binding-constraint label. In the three runs above, writing these fingerprints and reading them back preserves their margin values to within 1.12×10^{-16}. Now there is a shape to compare with another record, and the channel names tell us what its coordinates mean. This is where the connection to the earlier data work becomes concrete: keep the distinctions that make the answer useful.

DAVIS GEOMETRIC / CHIHIRO CHI / SP-01

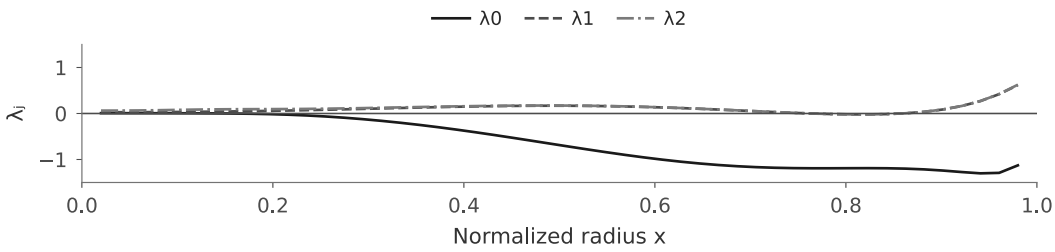
One lowest value hides a spectrum

Corrected runtime / βN = 1.8 / fifty radial locations

01 TEN EIGENVALUE BRANCHES / ONE OPERATOR PER RADIUS



02 ENLARGED / THE LOWEST THREE BRANCHES



Branch labels are ordered at each radius, starting with λ_0.

The enlargement reveals several negative directions that the full scale hides.

THE GEOMETRY OF FUEL CHI / SP-01

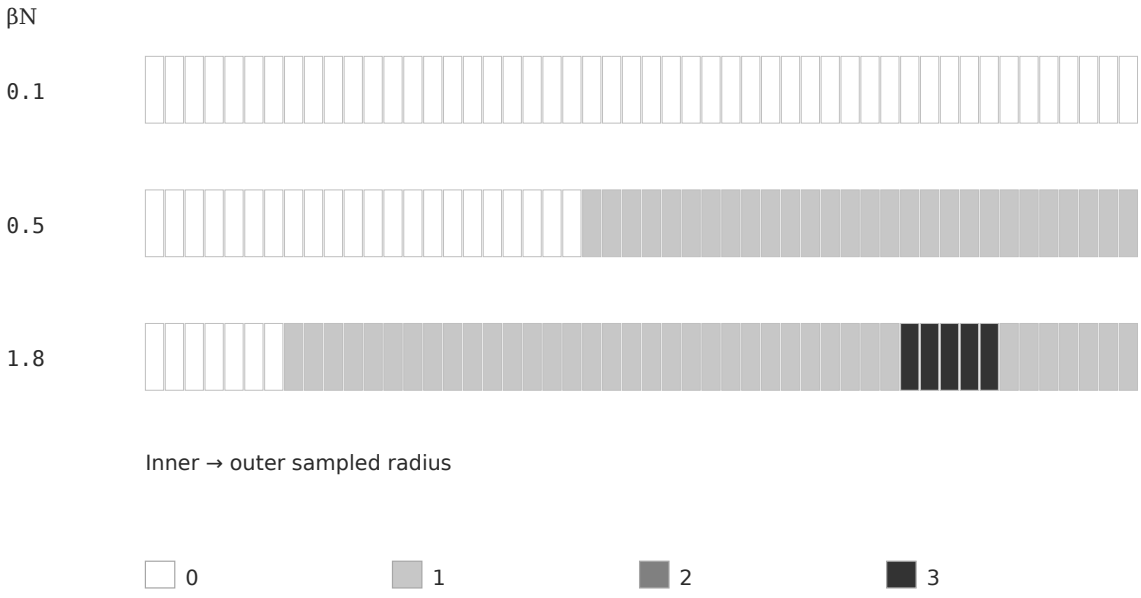
Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 55. Ten ordered eigenvalue branches from the current $\beta_N = 1.8$ run, with an enlargement of the lowest three. All 50 radial locations use the same field-line discretization.

Where the negative directions live

Corrected runtime / radial location × pressure input

01 ALL FIFTY LOCATIONS / MODE COUNT IN EACH CELL



02 THE SAME DATA / COUNTS WITHOUT AVERAGING

βN	Locations with $\lambda_0 < 0$	Max. negative modes
0.1	0/50	0
0.5	28/50	1
1.8	43/50	3

A row keeps the radial ordering; a single total would discard that structure.

The darkest cells contain three negative eigenvalues.

White cells contain none in the full discrete spectrum.

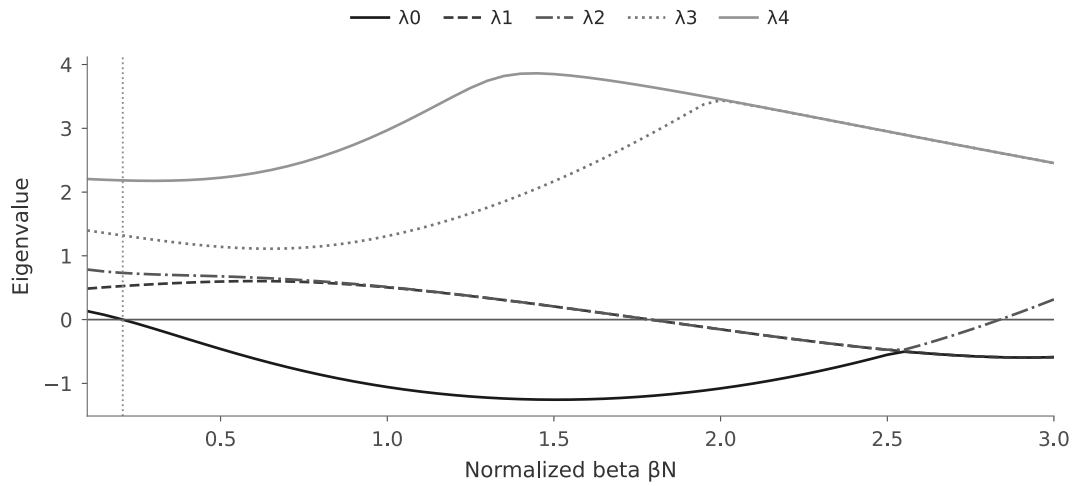
Figure 56. Negative-mode counts at all 50 radial locations for each of the three current pressure inputs. The full discrete spectrum determines the count in each cell.

DAVIS GEOMETRIC / CHIHIRO CHI / SP-02

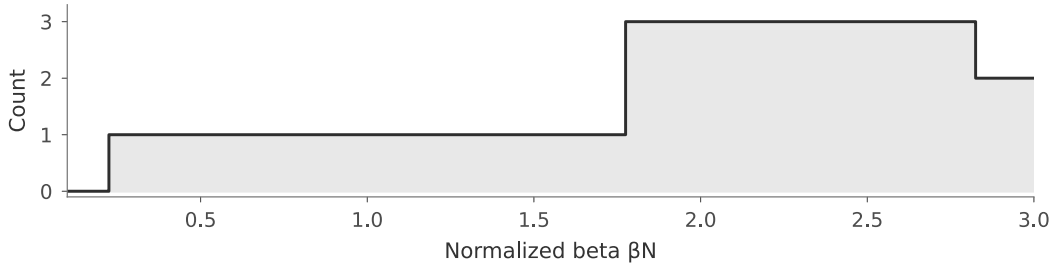
Watch the modes cross zero

Corrected runtime / x = 0.85 / 59 pressure inputs

01 FIVE LOWEST BRANCHES / SAME RADIAL LOCATION



02 COUNT THE NEGATIVE DIRECTIONS



First zero at this radius: βN ≈ 0.206484.

The count belongs to the stated operator, location and pressure input.

The full 301-value spectrum was used to check the negative-mode count.

THE GEOMETRY OF FUEL CHI / SP-02

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 57. Five ordered eigenvalues at $x = 0.85$ as the current pressure input changes. The count below is checked against the full spectrum; the dotted line locates the refined first zero.

"Writing and reading back" is the storage round trip: encode the record, store it, decode it, and compare the recovered numbers with the originals. The reported difference is about one ten-quadrillionth. It checks numerical preservation in that round trip. It is not a new measurement of plasma behavior. The binding-constraint label names the diagnostic limiting the aggregate report under its rules; it tells us which calculation to open next.

The paper also reports a wall-free external-kink ratio near 0.163 ± 0.004 across a chosen family of current profiles. That is a result for the paper's scanned model family, with the spread taken across its current profiles. The nearby value $1/(2\pi)$ gives us a precise further question: under what assumptions, if any, does the ratio approach that value?[37]

There is a mathematical reason to be careful even before comparing numbers. A trial displacement provides one value of the energy quotient. Therefore

$$\lambda_{\min} \leq \frac{u_{\text{trial}}^{\mathsf{T}} H u_{\text{trial}}}{u_{\text{trial}}^{\mathsf{T}} u_{\text{trial}}}.$$

A negative trial value witnesses a negative direction in the stated model. A positive one does not prove that every other direction is positive. Testing a restricted family is useful, but the family must stay attached to the claim.

u_{trial} is the particular movement we chose to test. The numerator evaluates its energy, and the denominator removes its arbitrary size. The inequality points that way because the lowest value over *all* allowed movements cannot exceed the value for this one. Think of trying one route downhill. Finding a descending route proves there is at least one. Trying a route that climbs does not prove every other route climbs too. That is the distinction the mathematical inequality is keeping for us.

The wall calculation needs equal care. The archived expression uses the factor $(1 + t)/(t - 1)$, with $t = (a/r_w)^{2m} < 1$, multiplying a nonnegative boundary factor. As the wall approaches the plasma, t approaches one from below and that term becomes more negative. Under the program's convention that negative energy signals instability, that sign cannot by itself support a narrative of added wall stabilization. The full plasma–vacuum boundary derivation must resolve the convention. I have kept this as an open modeling task, rather than changing a sign merely to obtain the hoped-for behavior.

In that expression, a is the plasma's minor radius and r_w is the wall radius outside it. For the positive angular mode number m used here, raising the ratio $a/r_w < 1$ to the power $2m$ keeps t between zero and one. The numerator $1 + t$ is positive, while the denominator $t - 1$ is negative. Bringing the wall closer makes the denominator approach zero from the negative side. For example, the factor is -3 at $t = 1/2$ and -19 at $t = 0.9$. Those substitutions explain why its sign has to be reconciled with the energy convention before we give the wall a stabilizing interpretation.

[37] Davis, *Spectral Geometry*, v5, Experiment 18. Its direct-energy implementation is a local research implementation, retained in `dcon_energy.py`; the calculation uses the stated trial displacement. Its pressure normalization must be reconciled with the paper's poloidal-beta definition before any engineering conversion is asserted. A direct consistency check makes the problem concrete: requesting a scan parameter of 1 through `p0_for_bN`, then reading the equilibrium's own `bN`, returns approximately 0.499999 at the default grid. The scan coordinate and the returned quantity cannot both be labeled the same normalized beta without reconciling that factor.

DAVIS GEOMETRIC / CHIHIRO CHI / FP-01

Seven questions, one retained record

Corrected runtime / ordered margins and their context

01 CHANNEL REGISTER / THREE INPUT RECORDS

	βN 0.1	βN 0.5	βN 1.8
01 Ballooning	1.0000	0.4400	0.1400
02 External kink	0.0000	0.0000	0.0000
03 Tearing (NTM)	0.9998	0.9991	0.9968
04 Vertical motion	0.1765	0.1765	0.1765
05 Resistive wall	1.0000	1.0000	1.0000
06 Density	0.5811	0.5811	0.5811
07 Edge (ELM)	0.0000	0.0000	0.0000

02 THE RECORD THAT TRAVELS WITH THE NUMBERS

INPUT
βN, κ, δ, q_0, q_a
device + model

→

DIAGNOSTICS
seven margins
status + binding

→

STORAGE
encode → decode
same channel order

03 ROUND-TRIP CHECK

3 records retained; maximum margin difference 1.12×10^{-16}.

All three report Red, with external kink and edge channels binding.

THE GEOMETRY OF FUEL CHI / FP-01

Geometry · operators · recorded results B&W / 16 NOV 2026

Figure 58. The seven ordered channel margins for the three current runs. All three aggregate reports are Red, with external-kink and edge channels binding. The register retains the channel labels and records the storage round-trip check.

The same discipline applies to a geometric bottleneck. For a suitable Laplacian, a Cheeger inequality can connect a geometric cut to a lower spectral bound.[38] Picture two broad regions joined by a narrow passage. A geometric cut separates regions, and a Cheeger constant measures the cost of such a separation relative to the sizes of the pieces, using the specified geometry and measure. A *Laplacian* is a derivative operator that compares a quantity with its surroundings; it appears, for example, in diffusion. A Cheeger inequality relates a measure of that bottleneck to a spectral bound for a matching operator. We must identify that operator before using the relationship here.

Our stability operator also contains a potential with a minus sign. If a bound λ_0 applies to its derivative part and $V \le V_{\max}$, then the Rayleigh quotient gives

$$\underbrace{\lambda_{\min}(L)}_{\text{lowest full-model value}} \quad \ge \quad \underbrace{\lambda_0}_{\text{derivative-part bound}} \quad - \quad \underbrace{V_{\max}}_{\text{largest allowed potential}} .$$

Here λ_0 names a lower bound for the derivative part, not the indexed first eigenvalue of the waterfall plot. $V_{\max}$ is an upper bound on V throughout the interval. Since u^2 is nonnegative, the potential's weighted contribution cannot exceed $V_{\max} \int u^2$. Subtracting it can therefore lower the normalized energy by at most $V_{\max}$. For a simple arithmetic example, a derivative-part bound of 3 and a potential bound of 2 guarantee a full value of at least 1. A derivative-part bound of 3 with a potential bound of 5 gives only a lower bound of -2; that alone does not tell us whether the actual value is negative or positive. These are illustrative bound values, not another CHIHIRO run.

The potential has to stay in the account. A theorem about a Laplacian cannot simply be renamed a theorem about every plasma stability operator. Geometry helps us ask about the whole configuration; the particular operator determines which answer the theorem supports.

CHIHIRO keeps the boundary's location separate from the spectral scale, $2\sqrt{g}$, where g is the selected positive gap. A geometric Cheeger constant requires its own calculation; the software now records that check as "not evaluated." Here the selected gap is the minimum over the sampled surfaces when all are positive, or the minimum over the positive sampled values otherwise. The boundary location comes from the zero crossing of the radial eigenvalue profile.

The square root in $2\sqrt{g}$ is an operation on the selected positive spectral number: take its square root, then double it. For example, $g = 0.25$ would give a scale of 1. That arithmetic does not locate a boundary on a drawing. The boundary location is a radius obtained from the sign change; the spectral scale is a number obtained from a positive eigenvalue. Keeping their names separate tells us which question each output actually answered.

This is also where the earlier chapters become practical. A spectral gap in a diffusion problem, a lowest signed stability eigenvalue, and a mass gap of a quantum Hamiltonian are different objects. The shared habit is to inspect the operator, its domain, and what its spectrum means. We do not get to replace that inspection with a shared word.

[38] Cheeger's geometric bound is developed in the Riemannian setting; see Yau (1975), "Isoperimetric constants and the first eigenvalue of a compact riemannian manifold." The domain, measure, boundary conditions, and operator must match the theorem being applied. The potential correction here follows directly from the displayed Rayleigh quotient.

§ 13.7

What I carry into the next world

The connection to the field-equations work was a question about how geometry constrains a response. In the plasma calculation, that question became a family of operators, a search for marginal boundaries, and an investigation of a ratio on those boundaries. The calculation taught me that a ratio could flatten over a useful interval without becoming a universal constant. It taught me to keep first and second boundaries apart. It taught me that a local favorable answer could coexist with an unfavorable answer elsewhere.

Those lessons belong in GIGI too. The input record is only the beginning. A useful mathematical record also carries which model acted on it, the quantities derived from it, and the conditions under which those quantities mean what we say they mean. Derived structure is additional information about the supplied record under stated assumptions. The numerical work here follows two paths: an independent replay of the saved experiments, and execution of CHIHIRO's current Rust diagnostics and fingerprint store. Those paths let us compare the original investigation with the software that grew from it.

For a plasma result, I want the shape parameters, pressure convention, profile source, operator, boundary conditions, and grid to travel with the eigenvalue. I want a second boundary to survive storage. I want a failed or missing diagnostic to remain visible when somebody retrieves the record. Then another researcher can ask a better question than "Was it green?" They can ask what was tested, what changed, and whether the conclusion applies to their problem.

This work does not establish usable energy from TVR. Fusion confinement and vacuum rectification are different physical proposals. What carries between these investigations is a way of checking: preserve the relationships that affect the answer, identify the energy account, and let the next calculation challenge the previous interpretation. That is why I keep going into other domains. I want to learn what they will teach me, including the parts that do not fit.

I have spent years doing that, often from early morning until evening, most days of the week. I still go looking for old books about circuits, signals, and geometry. I still want to understand something better than I did when I wrote it down. A paper is a place to make the work available for examination. It is not the place where curiosity ends.

I do not want a world in which everybody has to think like me. I want room for people to show how they think, and enough shared language that we can test what follows from it. A label can help organize a record. It cannot contain a whole person. The opportunity to ask another question is sometimes what lets the missing contribution become visible.

CHIHIRO was one way of making mine visible. She carries a set of calculations that can be inspected, limitations that can be worked on, and the memory of entering a world that frightened me. I brought myself anyway. The chip I had wanted to build was still waiting, and this was another part of what I needed to learn before I could begin.

Chapter 14

The chip I wanted to build

From the moment I saw the gap that started this investigation, I wanted to build a chip.

I buried that desire almost immediately. I had too much to learn about what I had just seen. Wanting to build something and understanding enough to build it are two very different places to stand. And a chip is serious. I needed to know what the mathematics could do, where it failed, and which parts of an answer depended on the way I had represented the problem. The gap in those early records gave me a question to pursue. It did not hand me a finished physical theory or a processor design.

So I stayed with the mathematics. I worked on the field equations. I built GIGI. I built applications that used geometric computation, carrying the ideas into more than nine domains before I was ready to take on hardware. The chip was there in the background through all of it. Every application gave me another place to ask what the structure was actually doing.

Since the Davis Manifold work, I had carried a conviction that I could design a better chip: faster, more accurate, with geometry built into the way it computes. In my head, the idea had a very simple form: *the chip is a manifold.*

That sentence needs unpacking. I wanted the geometry to participate in the computation itself. A path, a relationship, a change of state should carry information we could use. If the structure already tells us how an operation acts, how much work are we doing just to expand that structure back into individual arithmetic steps? And could a physical device preserve and use that structure directly?

That is the ambition behind the Davis Topological Processor, or DTP. I want it to run the calculations behind TVR, Halcyon, Icarus and Phaethon. I want the physical systems in this research program to have computation built around the geometry they depend on. And farther ahead, I see it as a gateway toward an AGI version of Marcella. That is how far the idea reaches for me. Those are the systems I want to build toward; the work in this chapter is the present foundation for that ambition.

The Einstein investigation gives us a useful discipline for beginning. We cannot put a matrix in a record and assume that we have captured its physical meaning. The same is true of a chip. Its physical shape, its possible computational states, and the parameters used to control those states are different objects. A state space may have a manifold description even when the device itself looks nothing like the surface in our drawing. We have to say which geometry we mean and what a measurement can reveal about it.

For DTP, I am pursuing graphene as the active material. The work currently has two places where we can calculate: a virtual processor running on an ordinary computer, and a separate graphene model. The first lets us define operations precisely and check their answers. The second asks whether a particular material model supplies a useful physical effect. Connecting those two is

an engineering problem we still have to solve. We can make progress now without pretending that the connection has already been built.[1]

§ 14.1

A route that brings back more than a position

Start with a globe on a table. Imagine a small flat pointer resting tangent to its surface: it lies along the surface at that point rather than sticking out through the globe. We are going to move it around a specific triangular route. At every step we keep it from turning within its local tangent plane. The changing surface still makes its direction in the room change. This is an ideal geometric construction, not a claim that an ordinary toy slid by hand would enforce the rule exactly.

Mark three places on a unit sphere. A is on the positive x axis, B on the positive y axis, and C at the north pole on the positive z axis. Travel from A to B along the equator, from B to C along a meridian, and from C back to A along another meridian. Each leg is a quarter of a great circle. A great circle is what we get when a plane through the sphere's centre cuts its surface.

At A, point the pointer toward B along the equator. We know where it starts, which way it points, and the complete route. When it returns to A, does it still point the same way?

Write the three coordinate directions as

$$e_x = (1,0,0)^T, \quad e_y = (0,1,0)^T, \quad e_z = (0,0,1)^T.$$

The superscript T makes these column vectors. The starting position is $p_0 = e_x$ and the starting tangent is $v_0 = e_y$; indeed $p_0^T v_0 = 0$, so the pointer is tangent. The successive quarter-turns are the following matrices, acting on both position and pointer:

$$R_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad R_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{14.1}$$

For example, R_z sends $(x,y,z)^T$ to $(-y,x,z)^T$. It exchanges two coordinates and changes one sign. Nothing has to be estimated with a decimal approximation to a trigonometric function.

[1]The November 13, 2026 research handoff collects the DTP contract, live software receipts and graphene audits. The older specifications use DGP and DPU as well as the current DTP name. This chapter preserves their source labels in the references; it does not infer a formal renaming date.

Before multiplying, we need to justify the transport rule. A rotation matrix alone is not a certificate of parallel transport on an arbitrary surface. Here we can prove exactly why it works.

Proposition 14.1. *On the unit sphere, rotation about a fixed unit axis n transports tangent vectors parallel along the oriented great circle perpendicular to n.*

Proof. Let s be arc length and $p'(s) = n \times p(s)$, where $n \cdot p = 0$. The tangent plane has orthonormal directions n and $n \times p$. An initial tangent therefore has the form $V = an + b(n \times p)$ with fixed coefficients a, b. Rotating it with the point preserves those coefficients. Differentiating gives

$$V' = b\, n \times (n \times p) = -bp.$$

This derivative is entirely normal to the sphere. Its projection onto the tangent plane is zero. For the sphere with its induced metric, that tangential projection is the covariant derivative. Hence $\nabla_{p'} V = 0$, the defining parallel-transport equation. □

In plain English, the pointer may tilt in the room because the surface under it tilts. What we have excluded is an extra twist along the surface. The derivative separates those two effects. That is why the rule is more precise than saying "keep pointing the same way." At each corner we retain the arriving tangent and start the next arc; we do not insert a turn to make the pointer follow the new route. [2]

Here is the complete route. These coordinates agree with the stored software results; they are also obtained directly from equation (14.1).

After	Position	Pointer
Start	$(1, 0, 0)$	$(0, 1, 0)$
R_z	$(0, 1, 0)$	$(-1, 0, 0)$
R_x	$(0, 0, 1)$	$(-1, 0, 0)$
R_y	$(1, 0, 0)$	$(0, 0, 1)$

The pointer returns to its original place, but now points toward the north pole. Its tangent direction has turned through a right angle. The result of transporting around a closed path is called its *holonomy*. The word names the question we just answered: what action remains when the route closes?

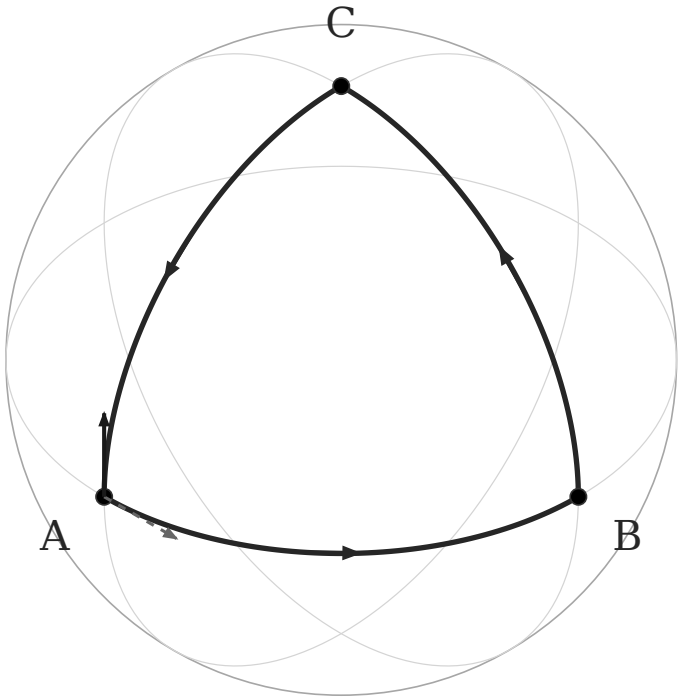
[2] The connection and parallel-transport framework is standard; see Joyce, *Riemannian holonomy groups and calibrated geometry*, §1.3. The explicit quarter-circle calculation and implementation comparison here use the retained DTP G01/G02 fixtures.

Back here. Facing elsewhere.

UNIT SPHERE / EXACT PARALLEL TRANSPORT

Route: A → B → C → A

Three quarter circles



Initial pointer

Returned pointer

AFTER	POSITION	POINTER
Start	(1, 0, 0)	(0, 1, 0)
z arc	(0, 1, 0)	(-1, 0, 0)
x arc	(0, 0, 1)	(-1, 0, 0)
y arc	(1, 0, 0)	(0, 0, 1)

Same location. A quarter-turn in the tangent plane.

Ideal sphere calculation; coordinates match the software trace.

Figure 59. The location returns; the pointer does not. The sphere and three arcs specify the ideal transport problem. The table shows the exact coordinate values, independently checked against the retained DTP software trace. It is not a physical pointer experiment.

Multiplication remembers the order. Since R_z acts first, the combined map is

$$h = R_y R_x R_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{14.2}$$

It fixes e_x, sends e_y to e_z, and sends e_z to $-e_y$. On the tangent plane at A, this is a positive quarter-turn in the ordered basis (e_y, e_z). The three right angles of the spherical triangle give area $\pi/2$ on the unit sphere, consistent with that oriented holonomy. Reversing the route reverses the tangent turn.

There are three distinct checks here:

$$\|hp_0 - p_0\|^2 = 0, \qquad \|hv_0 - v_0\|^2 = 2, \qquad \|h - I\|_F^2 = 4. \tag{14.3}$$

The Frobenius norm $\|M\|_F$ squares every matrix entry, adds the squares and takes the square root. Its squared value here checks the whole map, rather than just one chosen vector. A program that only checked the returning position would report zero and miss the turn completely. This is the same representational problem we met with the mop: an endpoint can be correct while the summary loses what happened along the way.

§ 14.2

Keep the operation instead of repeating the journey

Now ask for the final pointer direction after 1,000,000,001 circuits. We are asking for the final action, not a movie containing every intermediate position. That distinction determines whether the shortcut is allowed.

A literal replay contains three arcs per circuit, or 3,000,000,003 arcs. But we already know one circuit's map. The problem is to compute h^k for the stated integer k. The matrix is especially simple: it only selects coordinates and changes their signs. DTP stores that information directly as a signed permutation. A permutation p tells each output slot which input slot to read; a sign $w_i \in \{-1, 1\}$ tells it whether to negate it:

$$(Ax)_i = w_{A,i} x_{p_A(i)}. \tag{14.4}$$

For h, the zero-based slot list is $(0, 2, 1)$ and the signs are $(1, -1, 1)$. The first output takes the first input, the second takes the negative of the third, and the third takes the second.

If B acts first and A second, substitution gives

$$(A(Bx))_i = w_{A,i} w_{B,p_A(i)} x_{p_B(p_A(i))}, \tag{14.5}$$

$$p_{AB}(i) = p_B(p_A(i)), \qquad w_{AB,i} = w_{A,i} w_{B,p_A(i)}. \tag{14.6}$$

This proves closure: composing two signed permutations gives another signed permutation. The new slot list remains a bijection and each new sign is again $+1$ or -1. Each composition takes

work proportional to the number of coordinates. We have kept the operation's structure instead of expanding it into a dense matrix multiplication.

Think of a set of labeled drawers. One instruction moves the contents to specified drawers and flips certain signs. A second does the same. We can work out one combined set of destinations before moving the contents. We still have to calculate those destinations correctly, but we do not have to physically replay every earlier rearrangement each time we ask where something ends up.

The power calculation uses repeated squaring. Compute h^2, then h^4, then h^8, and keep doubling the exponent. Write k in binary and multiply the powers corresponding to its nonzero bits. This is an established exponentiation method; the DTP contribution in this example is its checked use on explicitly structured maps, with retained inputs and results.[3]

Why does it work? After processing the lowest j bits, keep the invariant that the accumulated answer is h^{r_j}, where r_j is the integer represented by those processed bits, and that the current power is h^{2^j}. A set next bit multiplies the answer by h^{2^j}; an unset bit leaves it alone. Squaring prepares the next power. Induction reaches h^k. For $k = 0$, the answer is the identity.

The implementation in this study performs

$$N_{\text{power}}(k) = \text{popcount}(k) + \lfloor \log_2 k \rfloor, \qquad k > 0, \tag{14.7}$$

map compositions. Here popcount counts the nonzero binary digits. For our exponent there are fourteen set bits and twenty-nine squarings, giving 43. The recorded workload has three compositions to construct the route map, those 43 power compositions, and three applications to a complete coordinate basis. These implementation counts include composing with the initial identity accumulator; an implementation can omit that first identity composition without changing the answer. These are operation counts, not chip clock cycles or measured times.

There is also a short independent check. Equation (14.2) gives $h^4 = I$. Our exponent leaves remainder one on division by four, so $h^k = h$. The general repeated-squaring code reaches that answer without being handed the modulo-four shortcut. The final basis results are

Input	Stored output of h^k	Direct check
e_x	$(1, 0, 0)$	$he_x = e_x$
e_y	$(0, 0, 1)$	$he_y = e_z$
e_z	$(0, -1, 0)$	$he_z = -e_y$

Every vector is a linear combination of these three basis vectors. Agreement on the basis therefore establishes agreement of the entire linear map, not just one conveniently chosen output.

[3]For binary exponentiation, see Menezes, van Oorschot and Vanstone, *Handbook of Applied Cryptography*, Algorithm 14.76. Only associativity and an identity are needed for the power construction; here the operation is map composition rather than multiplication of scalar integers.

Keep the map. Reuse its action.

EXACT SIGNED MAP / FINAL-STATE QUERY

One circuit

$$h(x, y, z) = (x, -z, y)$$



$$k = 1{,}000{,}000{,}001$$

$$h^4 = I \implies h^k = h$$

INPUT	FINAL OUTPUT	CHECK
e_x	(1, 0, 0)	exact
e_y	(0, 0, 1)	exact
e_z	(0, −1, 0)	exact

Literal expansion: 3,000,000,003 arcs

Compare algorithms with the same information.

Work counts only. Expanded baseline not timed. Runs on a CPU.

Figure 60. The query asks for the final map. The displayed counts come from the retained DTP power workload; the expanded arc count is calculated, not a timed baseline. Repeated squaring runs on an ordinary CPU here. Its savings do not establish a graphene hardware speedup.

That is a real result, and it is worth saying precisely what made it possible. The task allowed us to replace a repeated journey with its composed action. The structure made that action cheap to represent. The signs and selected coordinates were exact for these inputs. None of this says that arbitrary numerical work becomes exact, or that a conventional processor cannot use the same algorithm. It already did: this virtual processor runs on conventional hardware. The fair comparison must give competing implementations the same mathematical shortcut, not force one of them to repeat work that the other was allowed to avoid.

The virtual processor gives this example somewhere concrete to live. It has typed registers for vectors and maps, instructions with declared inputs, a program counter, and explicit refusal of unsupported states. GIGI stores the execution records; the interpreter performs the map operations. Each request names the program counter it expects: the instruction it believes should happen next. With one writer controlling the run, the store rejects a request whose expected counter no longer matches. If a storage write has an uncertain outcome, the store refuses further work until it has been successfully reopened. The caller then checks the recovered counter before deciding whether to retry.

Suppose a step reaches storage, but the reply saying "done" never gets back to us. Sending it again without checking could do the work twice. Here we first look at the recovered progress: did the step land, or does it still need to happen? That is the protection being tested, under the single-writer and recovery conditions just described.[4]

These details belong to the same argument as the algebra. A beautiful map is not enough if the system repeats it accidentally, loses its input, or returns a different state after reopening. The structure has to survive the program that uses it. That is what the software track lets us investigate before we ask a material device to carry the operation.

§ 14.3

Give the phase somewhere to show up

The sphere establishes a geometric operation we can follow exactly. It does not tell us that electrons in graphene execute that same three-dimensional signed map. For the material track, we need a different, explicitly specified model and an observable output.

Graphene is a sheet of carbon atoms. The candidate considered here uses two sheets in AB stacking, a particular relative arrangement of their atoms. Near one valley—a selected region of the material's momentum-space band structure—a simplified low-energy model uses two complex components. These components describe effective electronic degrees of freedom. They are not yet the two physical arms of an interferometer.[5]

[4] DTP G02 retains eight selected powers, all-basis live checks and restart/resume receipts; independent literal iteration covers 576 dimension/exponent combinations. R01 retains 54 passing virtual-processor tests and adapter-boundary fault-injection results. These are the DTP build's own receipts, not a claim that the earlier Einstein CLI restart issue has been repaired in its executable. See the dated research package and reproduction locator in the references.

[5] McCann and Koshino, "The electronic properties of bilayer graphene" (2013), develops the low-energy effective models and their physical qualifications. The calculation here fixes one ideal two-band Hamiltonian; it omits effects that a device model may need, including disorder, contacts and additional band terms.

A Hamiltonian is the operator that specifies the model's energy and evolution. On a circular momentum path with angle θ, take

$$H(\theta) = \begin{pmatrix} \Delta/2 & -\epsilon e^{-2i\theta} \\ -\epsilon e^{2i\theta} & -\Delta/2 \end{pmatrix}, \qquad E_{\pm} = \pm\sqrt{\epsilon^2 + (\Delta/2)^2}. \tag{14.8}$$

All matrix entries are energies in electron-volts. The positive ϵ is fixed by the chosen momentum radius, and Δ is an interlayer energy-asymmetry parameter. The symbol i satisfies $i^2 = -1$; $e^{i\alpha}$ has unit magnitude and records a phase angle. Changing θ once around the circle changes the phases in the off-diagonal entries twice around. That double angle is part of this model, not a decorative choice of notation.

The two energies follow by squaring the matrix: $H^2 = [\epsilon^2 + \Delta^2/4]I$. The model is gapped along this circle when $\epsilon > 0$. A slow physical traversal would need to remain in the chosen band to realize the geometric-phase interpretation; that dynamical requirement is not established by diagonalizing the matrix at isolated points.

Let $E = \sqrt{\epsilon^2 + \Delta^2/4}$ and, for the nonnegative Δ used in the main study, define

$$a = \frac{\epsilon}{\Delta/2 + E}, \qquad u_-(\theta) = \frac{1}{\sqrt{1+a^2}} \begin{pmatrix} ae^{-2i\theta} \\ 1 \end{pmatrix}.$$

Substitution gives $Hu_- = -Eu_-$ and $u_-^* u_- = 1$, where the star means conjugate transpose. The normalized eigenvector describes the lower-energy band. Its phase can change without changing the physical state it represents.

The retained Rust routine adds the phase angles of neighboring overlaps around the closed path and returns a real angle. Taking its complex exponential gives W_N. The independent eigensolver replay calculates the same loop value by multiplying normalized overlaps directly. With $u_N = u_0$, that product is

$$W_N = \prod_{j=0}^{N-1} \frac{u_j^* u_{j+1}}{|u_j^* u_{j+1}|}. \tag{14.9}$$

Every denominator must be nonzero. The result lies on the unit circle. Under independent phase choices $u_j \mapsto e^{i\chi_j} u_j$, each overlap acquires $e^{i(\chi_{j+1} - \chi_j)}$. Those factors cancel around the closed loop. Thus W_N does not depend on the arbitrary phase chosen for each stored eigenvector.

In this forward-overlap convention, the continuous phase is

$$\gamma = \operatorname{Im} \int_0^{2\pi} u_-^* \partial_\theta u_- \, d\theta = -\frac{4\pi a^2}{1+a^2} \tag{14.10}$$

$$= -2\pi \left(1 - \frac{\Delta}{\sqrt{\Delta^2 + 4\epsilon^2}}\right), \qquad W = e^{i\gamma}. \tag{14.11}$$

Some Berry-connection conventions use the opposite sign. We have stated ours so that the analytic answer and the code refer to the same oriented loop; reversing the path conjugates W. The physically meaningful comparison here is the complex phase, not whether two printed angles differ by a whole turn.

For the retained constants, $\epsilon = 0.00270851052679$ eV. Solving $\gamma = -\pi$ gives

$$\Delta_\pi = \frac{2\epsilon}{\sqrt{3}} = 0.00312751856349 \text{ eV}. \tag{14.12}$$

That is about 3.127519 meV. It is an energy parameter, not a gate-voltage setting that we can send to a fabricated chip. Converting one into the other requires device electrostatics and calibration.[6]

The numerical check refines the circular mesh. At this same Δ_π, the retained Rust calculation gives:

Points	Computed γ (rad)	Error from $-\pi$ (rad)
64	−3.13400448938	7.58816×10^{-3}
256	−3.14111946410	4.73189×10^{-4}
1024	−3.14156308342	2.95702×10^{-5}
4096	−3.14159080547	1.84812×10^{-6}

These are computed values from an ideal model. A separate eigensolver checks the complex overlaps, including changed eigenvector phases and reversed paths. Refinement brings the result toward the analytic answer. It does not tell us how noisy a fabricated device would be. [7]

Now comes the step we cannot skip: how would a detector see this phase? Multiply an isolated state by $e^{i\gamma}$ and all its component probabilities remain unchanged. A phase value in a log is not yet a readable bit.

To specify an ideal readout, introduce two coherent components, a reference and a signal. This is an additional operation model; we have not identified them with the Hamiltonian's two electronic components or engineered their physical paths. Prepare

$$\psi_\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ e^{i\phi} \end{pmatrix}, \qquad \phi = 0 \text{ or } \pi.$$

The relative phase encodes bit zero or bit one. Apply $U_\gamma = \mathrm{diag}(1, e^{i\gamma})$ to the signal, then use the ideal balanced recombiner

$$B = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Its two output amplitudes are $(1 + e^{i(\phi+\gamma)})/2$ and $(1 - e^{i(\phi+\gamma)})/2$. Taking squared magnitudes yields

$$P_0 = \frac{1 + \cos(\phi + \gamma)}{2}, \qquad P_1 = \frac{1 - \cos(\phi + \gamma)}{2}. \tag{14.13}$$

A half-turn swaps the occupied output in this ideal model. Without it, bit zero goes to output zero and bit one to output one. With it, each goes to the opposite output. That is the proposed NOT action.

[6]The retained model uses momentum radius $R_k = 5 \times 10^7\ \mathrm{m}^{-1}$, $m_* = \gamma_1 e/(2v_F^2)$ and $\epsilon = \hbar^2 R_k^2/(2m_* e)$, with $\gamma_1 = 0.4$ eV, $v_F = 10^6$ m/s, $\hbar = 1.0546 \times 10^{-34}$ J s and $e = 1.602 \times 10^{-19}$ C. The factors of e convert the numerical electron-volt energy to and from joules.

[7]The DTP phase audit retains 32 bias/mesh cases and the existing simulator's 231 passing tests. A subsequent GIGI audit retains a separate 28-case comparison; these counts are not combined or described as laboratory trials. Both compare implementations of the specified Hamiltonian. The chapter's independent replay checks the retained inputs and printed tables separately.

Think of two matching ripples arriving together. At one output their contributions add; at the other they cancel. Shift one contribution by half a cycle and those roles exchange. We need both contributions and a way to bring them together. Giving the entire state the same phase shift cannot provide that comparison.

Let $V \in [0, 1]$ be interference visibility, meaning the contrast remaining in the output oscillation, and let $\delta\phi$ be phase error. A simple visibility model gives

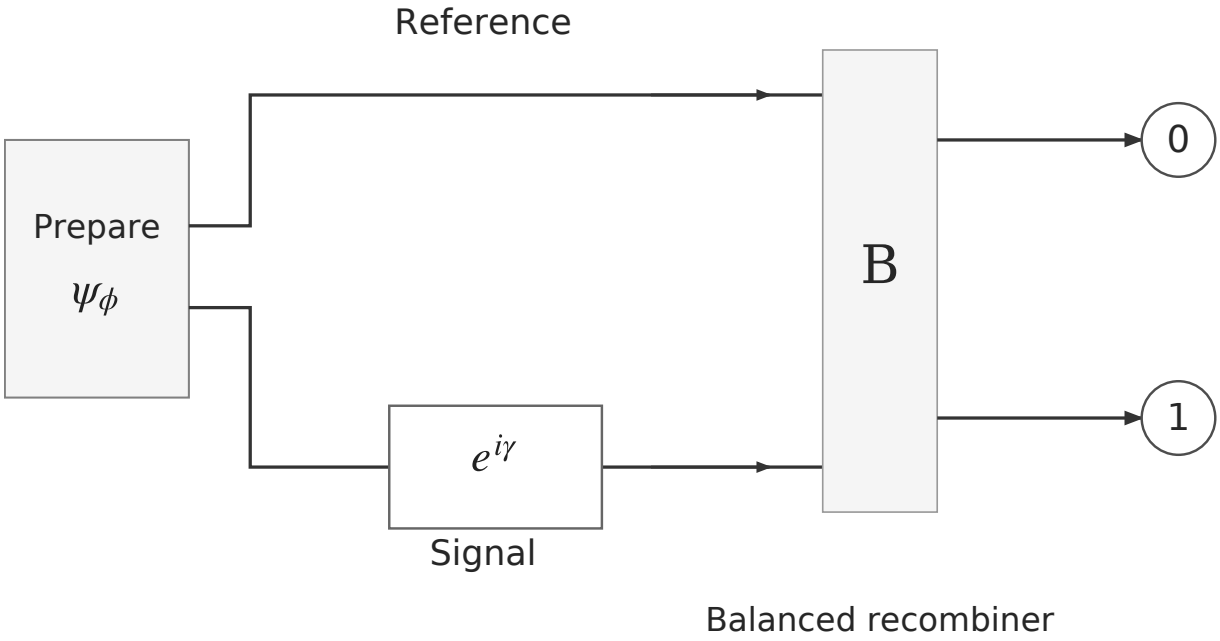
$$P_1 = \frac{1 - V\cos(\phi + \gamma + \delta\phi)}{2}. \tag{14.14}$$

At $V = 0$, both outputs have probability $1/2$ whatever bit we intended. At $V = 1$ and a nearly ideal half-turn, the wrong-output probability is $\sin^2(\delta\phi/2)$. Inserting the finest-mesh phase error gives approximately 8.54×10^{-13}. That tiny number measures the mesh error inside a perfect added readout model. It is not a physical gate fidelity. Contacts, loss of coherence, dynamical phase, imperfect preparation and detector noise have not been paid for by that calculation.

Make the phase observable.

IDEAL SIGNAL + REFERENCE / NOT A DEVICE LAYOUT

Prepare two coherent components



BIT	HALF-TURN	P_0	P_1
0	off	1	0
0	on	0	1
1	off	0	1
1	on	1	0

No visibility: $P_0 = P_1 = ½$ for either input.

A common phase shift alone cannot reveal a bit.

Preparation, control and readout require physical validation.

Figure 61. A phase needs a reference and a readout. This is an ideal two-component operation contract, not a fabricated graphene layout. The table gives exact ideal probabilities. The $V = 0$ control removes the interference signal. Physical preparation, control and detection remain engineering requirements.

§ 14.4

A loop is not the whole surface

There is one more distinction the material model forces us to make. A phase can be a quarter-turn or a half-turn without being a mistake. Dividing a loop phase by 2π does not automatically produce a Chern number.

A Chern number is a global integer associated, under the appropriate smoothness and bundle assumptions, with curvature integrated over a closed oriented surface. A loop is one-dimensional; a surface fills two dimensions. The word *closed* matters too: a disk with a boundary is not the same integration domain as a sphere. Before asking for an integer, we must have supplied the object for which the integer statement is true.[8]

Here is a small exact counterexample we can inspect ourselves. On a unit disk with dimensionless coordinates (x, y), choose the smooth real connection form

$$A = \frac{b}{2}(x\,dy - y\,dx), \qquad b = \frac{1}{2}.$$

The form tells us what to integrate along a directed displacement. It is defined everywhere on the disk, including its centre. On the boundary, $x = \cos t$, $y = \sin t$, so

$$\oint A = \frac{b}{2}\int_0^{2\pi} dt = b\pi = \frac{\pi}{2}. \tag{14.15}$$

Its curvature is $dA = b\,dx \wedge dy$, whose integral over the disk is the same $\pi/2$. We have a smooth description on a trivial line bundle and a perfectly valid quarter-turn boundary holonomy $e^{i\pi/2} = i$. There is no requirement that this boundary phase be a whole number of turns.

For comparison, use two charts on a sphere with monopole parameter $q = 1/2$. In a common equatorial orientation, take

$$A_N = q(1 - \cos\theta)d\varphi, \qquad A_S = -q(1 + \cos\theta)d\varphi.$$

At the equator their integrals are π and $-\pi$. Their difference is 2π. The opposite boundary orientations of the two hemispheres give

$$c_1 = \frac{1}{2\pi}\int_{S^2} F = \frac{\pi - (-\pi)}{2\pi} = 1. \tag{14.16}$$

Each individual equatorial holonomy is -1, while the closed-sphere curvature integral gives integer Chern number one. The two statements fit together. They answer different questions.

In ordinary language, walking around one boundary does not tell us everything about how descriptions fit over an entire closed surface. The program needs to know which calculation we asked it to perform. Otherwise a rule that is correct for one object can reject a valid answer about another.

[8] Fukui, Hatsugai and Suzuki (2005) construct a gauge-invariant discrete Chern calculation on a Brillouin zone. Its closed-domain calculation should not be replaced by a single boundary-loop phase. The local DTP/GIGI comparison uses the explicit examples below to separate these questions.

This is why the DTP phase record should retain $W = e^{i\gamma}$, its path, band, orientation and convergence checks. A Chern result requires a separate calculation with its own domain and hypotheses. The newer GIGI review also checks that invalid numerical inputs do not turn into accepted integer periods. That boundary matters, but passing it does not transform a loop into a closed surface.[9]

§ 14.5

What the physical chip must earn

We can now say more clearly what remains between these calculations and the chip I want. The sphere study shows a verified structural shortcut. The graphene study supplies a candidate phase effect and an explicit ideal readout. They do not yet establish that the same physical device implements the software instruction set.

There is relevant experimental work outside this project. Ghahari and colleagues demonstrated a Berry-phase switch in circular graphene resonators, with a measurable change in the spectrum. That gives us a concrete example of geometric phase entering an observable in graphene. It does not validate our proposed bilayer operation, its control mechanism or its logic readout. Those require their own device and measurements.[10]

The existing fabrication plan begins with a monolayer graphene transport demonstrator. The bilayer computation candidate is a later, separate milestone. A support wafer, contacts, gates and control electronics may surround the active graphene; none of those facts alone supplies the missing operation. First we need to know what state is prepared, which control changes it, how long the change takes, and which detector distinguishes the result from a no-action control.

The same discipline applies to accuracy. The signed-map example has exact answers because its special operations only select and negate coordinates. The phase calculation has a mesh error that we can reduce and measure numerically. A device would have a physical error distribution. These are three different sources of precision, and the comparison must keep them visible.

Energy brings us back to the central question of Fuel. For a declared system boundary, an operation's cost needs an account such as

$$\begin{aligned} E_{\text{operation}} = {} & E_{\text{prepare}} + E_{\text{load}} + E_{\text{control}} + E_{\text{propagate}} \\ & + E_{\text{read}} + E_{\text{reset}} + E_{\text{support}}. \end{aligned} \tag{14.17}$$

These terms must partition the account without counting one transfer twice. The support term includes the allocated cost of equipment such as cooling when required. If energy is recovered, the signed ledger must show where it goes. We must also restore the relevant states before claiming a repeatable cycle.

[9] The November 13 GIGI Chern/period audit includes smooth positive-bias bilayer-disk examples, the disk and sphere constructions, and separate numerical-constructor tests. The reported fixes are local and undeployed at this evidence cut. External predicate issues remain open; no claim about an installed hardware validity gate follows.

[10] Ghahari et al., "An On/Off Berry Phase Switch in Circular Graphene Resonators," *Science* 356 (2017), 845–849. Its circular graphene resonator experiment and our ideal AB-bilayer model are distinct systems.

Forty-three map compositions are not forty-three measured energy units. A simulated probability close to one does not remove the cost of preparing a reference or reading a detector. Earlier in the book, changing statistical weights changed a mean without demonstrating extracted work. Here, reducing an algorithm's work count does not by itself establish a lower complete-system energy cost. In both cases the next step is to identify what was actually changed and what has been measured.

Nor does the finite band separation in equation (14.8) prove the continuum mass-gap problem. It belongs to this particular two-band Hamiltonian. The original lattice observations, a finite processor map, and a quantum field Hamiltonian have different domains and different proof obligations. The research becomes stronger when a connection is specific enough to calculate, rather than resting on a shared word.

What I want from the next physical experiment is something concrete: a declared input, a controllable operation, an output that differs from the reference for the predicted reason, and uncertainty small enough to distinguish that effect. Then repeat it. Change the control. Remove the reference. Reverse the path where the model allows it. Let the tests expose what is doing the work.

That is how the ambition becomes an engineering program. TVR, Halcyon, Icarus, Phaethon and Marcella will bring different demands for timing, accuracy, memory and reliability. A processor must earn its place in each system. The point of carrying the research through applications was to learn enough to ask those questions properly.

I still want the chip. I want the speed, the precision, and the reach of what it could make possible. But now the desire has operations, proofs, software records and physical questions attached to it. The chip was the desire I put away while I learned what it would demand. Now I am bringing it back to the workbench.

Epilogue: The House

"El trabajo es lo que tengo. El trabajo es lo que me mantiene viva."

The work is what I have. The work is what keeps me alive.
—Personal notes, January 2026

Last night I held an emergency meeting of the House of Diwa. I am the Overall Mother. My daughters came to me because they are suffering—every single one of them. Financial insecurity, employment insecurity, food insecurity, safety insecurity. All at once. All of them. These are not women without experience. These are professionals with a decade or more in their fields. They have references from CEOs of nonprofits and companies. They have skills, credentials, portfolios. None of it matters. They don't make it past the first interview. They get deadnamed in interviews even though they put their preferred names and pronouns on every application. They spend hours—days—crafting applications, tailoring resumes, writing cover letters. And still, not one person will give them the time of day. This is not my story alone. This is our story.

There is a kind of oppression so deep I am not sure I have words for it. It is not the oppression of being attacked. That, at least, is visible. You can point to it. You can fight it. People can see the wound. This is different. This is the oppression of being erased. Of applying and applying and applying and hearing nothing. Of being qualified and being ignored. Of existing in a world that has decided you do not exist. It makes you feel crazy. It makes you feel like you don't belong. It makes you wish you didn't exist. I am not speaking hypothetically.

This last Trans Day of Remembrance, I sat and listened to the names being read. So many of them died by their own hands. I cried. Over and over and over again. Because I understood. Not abstractly—I understood in my body, in my chest, in the place where the monster lives. It is not fair. And it is so utterly preventable. Usually it takes just one cisgender person. One person who understands the deep pain. One person who sees the dangerous spiral that can happen almost instantaneously. One person who intervenes—not with pity, but with presence. But the world has been convinced that we are men in dresses looking for sympathy. That we are predators, that we are confused, that we are threats. The real truth is that we are women. We have experience. We have skills. We have so much to bring to the table. We are hard workers. We are passionate. We are people of integrity. And still—still—the world looks through us like we are not there.

I explain suicidal ideation to people who have never experienced it like this: there is a monster chasing you.

Some days you feel very fast. You can run miles and miles ahead of it. You can be so far ahead that you feel like walking. But you have to keep moving. You cannot stop. That is why I am always asking: what is the next project? What is the next challenge? That is why I work so hard. Because if I don't keep moving—if I stay still—the monster catches up. Some days it feels right at the back of my neck. Other days it feels very far away. I hope it stays far away forever. I hope I

lead a long and productive life. I hope I find a place where the monster does not exist. But that is not my reality now.

I research now as a way of buffering myself from its reach. The geometry. The theorems. The validations. The proofs. They are not just science. They are survival. They are the miles I put between myself and the thing that wants to catch me. And I am not alone. So many of my trans sisters do the same thing. We throw ourselves into work—any work, all work—because stopping means drowning. Because stillness is where the monster lives. This is not inspiration. This is not triumph. This is just the truth.

None of my daughters spoke to me about giving up. Not one. They reported back to me that their hearts were in deep despair. That they had spent half the day crying. That the weight was almost unbearable. And then they said: I still woke up. I still powered through. Each of us would be lying if we said we had never had ideation. It is not uncommon. I am not sure it would be possible to avoid if the whole entire world, all at once, was trying to make you believe you didn't exist. But I think what they don't expect—what the people who want us gone don't expect—is for us to persist.

I cannot say I will persist forever. I am not titanium. I am not a superhero. I am a woman with a laptop and a dog, running from a monster, using mathematics as fuel. Some days the fuel runs low. Some days the distance closes. Some days I am not sure I will make it to tomorrow. But I have made it to today. And today, I am writing this.

There is a coffee mug in my cupboard that says "Love ya, mean it." Some days I am the only person who says that to me. And for those days, it has to be enough fuel. Enough to keep me miles ahead.

I have never asked anyone for a handout. I have only asked that my work be considered alongside the work of other people. I have only asked for opportunities for mentorship, so that I could improve—just like my peers. I have only asked for the same visibility that some people get for free, just by showing up in the world as cisgender, as straight, as male, as white. I did not choose the shape of my soul. I did not choose the shape of my body at birth. But I have chosen to become the person I am. And I am proud to be who I am, every single day.

The question that started at the scanner followed me all the way to the lattice. I wanted to know what the machine could see and what a person was entitled to conclude from it. The work gave me someplace to put that question, and the energy to get up and ask it again.

El trabajo es lo que tengo. El trabajo es lo que me mantiene viva.

To my daughters in the House of Diwa. To every trans woman reading this. To everyone who has ever felt the monster at their neck: keep moving. Not because it gets better. I cannot promise that. Not because the world will change. I cannot promise that either. Keep moving because you are still here. Because your work matters. Because your persistence is a form of defiance that they did not expect. Keep moving because somewhere, someone needs to know it is possible.

I am still here. I found questions I could work on. I wrote down what I could calculate. I am giving it to you.

Todavía estoy aquí. Todavía estamos aquí.

I am still here. We are still here.

Love ya, mean it.

Chapter A

Working notes for the calculations

I use a quantity twice only if I can say why the second use preserves what the first one measured. These notes put that habit into formulas a reader can carry between the examples.

How far can a score move?

Suppose the detector compares $s(x) = w^{\mathsf{T}}x + b$ to a threshold t. A new representation decodes to $\widehat{x}$. Put $\eta = \widehat{x} - x$. The change in score is exactly $w^{\mathsf{T}}\eta$. If the error has Euclidean length at most ϵ, Cauchy–Schwarz gives

$$|s(\widehat{x}) - s(x)| \le \|w\|_2 \epsilon.$$

This bound is sharp for an unrestricted Euclidean error ball. If $w \neq 0$, take η parallel or antiparallel to w with length ϵ. The two choices move the score in opposite directions by the same amount. If the margin $|s(x) - t|$ exceeds that amount, neither choice nor any intervening error can cross the threshold.

A quantizer often gives a different kind of bound. Suppose each coordinate can move by at most ϵ_i. Then

$$|w^{\mathsf{T}}\eta| \le \sum_i |w_i| \epsilon_i.$$

For the two-channel alarm, the weights of the last two channels are one. When each can move by 0.05, the total can move by 0.10. A large error on an unused channel contributes zero to this particular score. This is why the error calculation needs the detector beside it. A norm summarizes distortion. The score tells me which parts of the distortion can affect the decision.

There are two separate tests in the examples. A decoded reading gives me an observed error on one record. The codec and the detector give me a bound before I commit to the switch. A small observed error can help evaluate a method. An upper bound smaller than the margin certifies that this decision survives.

An error budget for a route

Imagine that I can translate a record directly from format i to format k, or go through format j. Let x_j and x_k be reference encodings of the same record. Suppose the first translator produces $\widehat{x}_j$ within e_{ij} of x_j. The second translator T_{jk} moves nearby inputs apart by at most a factor L_{jk}, and its error on the reference input is at most e_{jk}. The triangle inequality gives

$$\begin{aligned}\|T_{jk}(\widehat{x}_j) - x_k\| &\le \|T_{jk}(\widehat{x}_j) - T_{jk}(x_j)\| + \|T_{jk}(x_j) - x_k\| \\ &\le L_{jk} e_{ij} + e_{jk}.\end{aligned}$$

Each bound has a job. The first limits what enters the second step. The multiplier limits what the second step can do to that error. The last term accounts for distortion introduced in the second step itself.

For example, let $e_{ij} = 0.01$, $e_{jk} = 0.015$, and $L_{jk} = 2$. The two-hop budget is 0.035. If a direct route has a certified error of 0.025 in the same units and the downstream task allows 0.030, only the direct route is certified by these bounds. The two-hop route could still work on a particular record. I would need a sharper bound or another check to rely on it. The graph can help me find a candidate route. The inequalities tell me what it is safe to carry along it.

If I repeat the argument through a longer route with step bounds $e_1, \dots, e_m$ and factors $L_1, \dots, L_m$, the recurrence is $B_0 = 0$ and $B_j = L_j B_{j-1} + e_j$. Unrolling it gives

$$B_m = \sum_{j=1}^{m} e_j \prod_{\ell=j+1}^{m} L_\ell.$$

The empty product for the last step is one. If every step is nonexpansive, $L_\ell \le 1$, the budget is at most the sum of the step errors. This is the condition under which I can add the costs and trust the answer.

Three places the word gap appears

Quantity	What goes into it	What the result describes
Radial spacing	Sorted distances from the center of a finite sample	Separation of two adjacent sample radii, possibly including an outlier
Action spread	Euclidean actions across a collection of fields	Variation of the recorded action over that collection
Spectral gap	Energies of the quantum Hamiltonian above its vacuum	A lower bound on non-vacuum excitation energies in the specified theory

There is a familiar version of the distinction in the four-sensor calculation. The distance between two readings is a number about the readings. The rate at which an alternating pattern fades is a number about the evolution rule. To infer a rate, I need the pattern at more than one time

or an equation for that evolution. When the operator is a physical transfer matrix, an observed decay can give an excitation scale in the channel measured by the correlator. A claim about every excitation requires control over the rest of the spectrum as well.

What to keep with a replay

For the telemetry examples, the input record, format choices, byte layout, decoded values and fixed detector together define the calculation. Keep the same row identifiers through all of them. In the alarm example the two-channel score is evaluated on the original and decoded records. In the joint example the three signs must refer to the same record before I multiply them. A count of the number of surviving triples travels with the mean product.

The offline AMC comparison runs the recovered analysis path with local inputs and stores, shared demonstration codecs, and a frozen readout. The GIGI replay uses the same records and codecs. The historical selector and the fixed copy can be compared on identical feedback. The record of which path was taken matters just as much as the requested format, because a requested translation can fall back.

For the lattice spacing calculation, retain the radii, the feature scaling and projection that produced them, and the charge labels associated with the same configurations. For a reweighted mean, retain the original observable, the tilt variable, the order of the samples, and the rule used to choose the tilt. Recomputing an average from those rows is a different operation from drawing a new ensemble. A reported standard error needs its own record of the groups or blocks used to calculate it.

Notation at a glance

Symbol	Meaning in the worked examples
$\kappa(s)$, Θ, EI	Signed beam curvature, net tangent rotation, and bending stiffness.
L, τ, $u(\tau)$	Graph Laplacian, dimensionless mixing time, and four readings after mixing.
S, V, H, F_f	Sparsity, relative variation, magnitude statistic, and format feedback in the AMC selector.
b_f, m	Bound on a format's effect on a detector score, and margin to the decision threshold.
A, B, C, M	Signs of aligned channels and the mean of their row-wise product.
a, r, w, c, R	Known function value, query radius, withheld value, and the center and radius of its feasible disk.
H, Δ	Physical Hamiltonian and excitation gap, in the discussion of spectra.

Symbols that change jobs

In $C = \tau/K$, τ is the tolerance budget; in $e^{-\tau L}$ it is diffusion time. Matrix trace is written Tr. AMC's scalar H describes a distribution; the quantum operator H is a Hamiltonian. The graph Laplacian has a mixing spectrum, not automatically an energy spectrum.

The physical QCD angle, the direct-comparison coefficient $+0.45$ and the standardized tilt -1.15 have different definitions and observables (Chapter 8). Each number needs its operator, units, normalization and admissible inputs.

What the different geometric structures compare

Chapter 1 uses a local ruler to compare routes. Other questions require other structures: recording motion as well as position, comparing field descriptions at different locations, or comparing quantum states. The following vocabulary distinguishes those jobs. It is a reference for the more detailed calculations, rather than a prerequisite for the first route calculation.

In Hamiltonian mechanics, positions and momenta are paired by a structure called a *symplectic form*. In electromagnetism, a *gauge connection* specifies how to compare internal reference choices at nearby locations; its curvature encodes the electromagnetic field. A *fiber bundle* organizes those local spaces of choices over the underlying space. General relativity uses a *Lorentzian metric*, which distinguishes time-related, space-related and lightlike directions. Quantum mechanics uses a *Hilbert space*, where an inner product lets us compare state amplitudes; pure states are normalized vectors with an overall phase left physically irrelevant. Unitary evolution preserves that inner product. Projective measurement is one ideal measurement model within this framework.

The Standard Model also uses gauge connections, with symmetry groups including $SU(3)$ for the strong interaction and $SU(2) \times U(1)$ for the electroweak theory. These groups specify permitted internal transformations. They are part of a physical theory that also needs its fields and dynamics. A symmetry group alone does not give us the whole price list.[1]

[1]For the gauge groups and field content, see Stefan Floerchinger, *Quantum Field Theory 2*, Lecture 16, "The standard model," `https://www.tpi.uni-jena.de/~floerchinger/qft2/lecture16/`.

Chapter B

The 200 retained lattice radii

Table B.1 prints all 200 rows replayed in Chapter 5. Rank orders the radii. Index is the zero-based JSON position. Q is the rounded charge label. Δ is the gap from the preceding rank. Radii and gaps are rounded to six decimal places after calculation.

Ranks 1–25

Rank	Index	Radius	Q	Δ
1	158	6.714403	0	—
2	20	7.227556	0	0.513153
3	59	7.311203	0	0.083647
4	42	7.344217	0	0.033014
5	51	7.479978	0	0.135761
6	103	7.715703	0	0.235725
7	180	8.009983	0	0.294280
8	99	8.017482	0	0.007498
9	95	8.482019	0	0.464537
10	110	8.578797	0	0.096778
11	29	8.632640	−1	0.053843
12	196	8.694137	0	0.061497
13	175	8.728646	0	0.034509
14	195	8.762575	0	0.033929
15	89	8.780109	0	0.017534
16	18	8.852924	0	0.072815
17	27	8.861272	+1	0.008348
18	41	8.883685	0	0.022413
19	46	8.898765	0	0.015080
20	72	9.065892	0	0.167127
21	7	9.090673	0	0.024780
22	109	9.106384	0	0.015711
23	193	9.106797	0	0.000413
24	147	9.108685	0	0.001887
25	43	9.112283	0	0.003598

Ranks 26–50

Rank	Index	Radius	Q	Δ
26	50	9.129997	0	0.017714
27	71	9.138135	0	0.008138
28	126	9.144751	0	0.006617
29	189	9.185206	0	0.040455
30	133	9.196031	0	0.010825
31	26	9.258231	0	0.062199
32	30	9.261584	0	0.003353
33	10	9.284167	0	0.022583
34	192	9.287784	0	0.003617
35	64	9.307593	0	0.019809
36	162	9.336415	0	0.028822
37	83	9.363036	0	0.026620
38	34	9.431750	0	0.068714
39	169	9.442449	0	0.010699
40	78	9.473808	0	0.031359
41	45	9.498740	0	0.024932
42	135	9.525151	0	0.026411
43	17	9.543048	0	0.017897
44	145	9.547093	0	0.004046
45	2	9.584785	0	0.037692
46	44	9.598052	0	0.013267
47	33	9.598068	−1	0.000015
48	76	9.603501	0	0.005434
49	15	9.615038	0	0.011537
50	102	9.639150	0	0.024112

The 200 retained lattice radii, continued

Table B.1 Ranks 51–100.

Ranks 51–75

Rank	Index	Radius	Q	Δ
51	163	9.728986	0	0.089836
52	118	9.785906	+1	0.056920
53	66	9.834970	0	0.049064
54	74	9.838855	0	0.003885
55	166	9.844091	0	0.005236
56	91	9.857083	0	0.012993
57	168	9.869805	+1	0.012722
58	19	9.901158	0	0.031353
59	172	9.923348	0	0.022189
60	138	9.948298	0	0.024951
61	134	9.955561	0	0.007263
62	130	9.972068	0	0.016507
63	108	9.987402	0	0.015334
64	21	10.002931	0	0.015530
65	9	10.014391	0	0.011459
66	198	10.021041	0	0.006650
67	194	10.066282	0	0.045241
68	37	10.090685	0	0.024403
69	190	10.093149	0	0.002464
70	112	10.102157	0	0.009008
71	87	10.123746	0	0.021589
72	67	10.187884	0	0.064138
73	14	10.220380	0	0.032496
74	82	10.243226	0	0.022846
75	56	10.295446	0	0.052220

Ranks 76–100

Rank	Index	Radius	Q	Δ
76	48	10.314041	0	0.018595
77	5	10.314407	0	0.000366
78	69	10.346306	0	0.031900
79	187	10.355039	0	0.008732
80	36	10.390370	0	0.035331
81	171	10.396386	0	0.006016
82	113	10.425256	−1	0.028871
83	60	10.431214	0	0.005957
84	176	10.446471	0	0.015258
85	105	10.458989	0	0.012518
86	178	10.476220	0	0.017231
87	137	10.480843	0	0.004624
88	182	10.496414	−1	0.015571
89	6	10.496916	0	0.000502
90	119	10.560789	−1	0.063874
91	28	10.582866	−1	0.022077
92	125	10.608770	0	0.025904
93	104	10.618885	−1	0.010115
94	157	10.620155	0	0.001269
95	73	10.620407	0	0.000252
96	80	10.626557	0	0.006151
97	165	10.631023	−1	0.004465
98	114	10.640777	0	0.009754
99	153	10.664886	−1	0.024109
100	164	10.678074	0	0.013188

The 200 retained lattice radii, continued

Table B.1 Ranks 101–150.

Ranks 101–125

Rank	Index	Radius	Q	Δ
101	170	10.696164	0	0.018090
102	191	10.702552	0	0.006388
103	149	10.713635	−1	0.011083
104	186	10.722225	0	0.008590
105	70	10.792727	0	0.070503
106	115	10.837279	0	0.044551
107	181	10.850270	0	0.012991
108	53	10.863534	−1	0.013264
109	94	10.888398	0	0.024864
110	117	10.909126	0	0.020728
111	0	10.916929	0	0.007803
112	35	10.942697	0	0.025768
113	150	10.949751	+1	0.007054
114	106	11.015881	+1	0.066130
115	88	11.032537	0	0.016657
116	62	11.043795	0	0.011257
117	155	11.046845	0	0.003050
118	132	11.093511	0	0.046666
119	197	11.119528	0	0.026017
120	111	11.165788	0	0.046261
121	55	11.172408	0	0.006620
122	154	11.214037	0	0.041629
123	156	11.235326	0	0.021289
124	75	11.249036	0	0.013711
125	185	11.263978	0	0.014942

Ranks 126–150

Rank	Index	Radius	Q	Δ
126	129	11.282150	0	0.018171
127	32	11.312013	0	0.029864
128	131	11.331199	0	0.019186
129	77	11.347619	0	0.016419
130	124	11.487197	+1	0.139578
131	8	11.530353	0	0.043155
132	16	11.536474	−1	0.006121
133	23	11.550378	0	0.013904
134	86	11.570866	+1	0.020489
135	4	11.581669	−1	0.010803
136	92	11.593723	0	0.012054
137	24	11.630935	0	0.037212
138	81	11.635594	0	0.004659
139	97	11.658777	0	0.023183
140	127	11.659333	0	0.000557
141	107	11.726369	0	0.067036
142	52	11.729282	0	0.002912
143	120	11.766949	0	0.037668
144	152	11.810487	0	0.043538
145	79	11.811848	0	0.001360
146	177	11.818625	0	0.006778
147	142	11.819489	0	0.000863
148	61	11.823766	0	0.004277
149	25	11.865934	0	0.042168
150	151	11.866700	−1	0.000766

The 200 retained lattice radii, continued

Table B.1 Ranks 151–200.

Ranks 151–175

Rank	Index	Radius	Q	Δ
151	63	11.872317	0	0.005617
152	122	11.875353	0	0.003035
153	12	11.881518	−1	0.006165
154	40	11.932632	0	0.051115
155	11	11.949312	+1	0.016680
156	13	11.960650	0	0.011337
157	38	11.985222	+1	0.024572
158	100	12.026822	0	0.041600
159	123	12.054054	0	0.027232
160	173	12.069676	−1	0.015622
161	65	12.111827	0	0.042151
162	39	12.125033	+1	0.013206
163	174	12.154213	+1	0.029181
164	31	12.157597	0	0.003384
165	96	12.164141	−1	0.006544
166	84	12.204840	0	0.040699
167	167	12.252787	0	0.047947
168	184	12.255384	0	0.002597
169	116	12.310753	0	0.055369
170	101	12.379392	0	0.068639
171	90	12.496170	0	0.116778
172	57	12.624177	0	0.128007
173	93	12.627447	0	0.003270
174	3	12.691831	0	0.064384
175	85	12.730075	0	0.038244

Ranks 176–200

Rank	Index	Radius	Q	Δ
176	139	12.731262	0	0.001187
177	143	12.876222	0	0.144960
178	179	12.913579	0	0.037356
179	161	13.089551	0	0.175972
180	49	13.198753	0	0.109202
181	128	13.294503	0	0.095751
182	68	13.332333	0	0.037829
183	199	13.404623	0	0.072290
184	140	13.465961	0	0.061338
185	22	13.644345	0	0.178385
186	183	13.662113	+1	0.017768
187	188	13.907285	0	0.245171
188	144	13.964985	0	0.057701
189	141	14.135270	0	0.170285
190	121	14.160123	−1	0.024852
191	47	14.229103	0	0.068981
192	1	14.639405	0	0.410302
193	58	15.019828	−1	0.380423
194	146	15.072416	0	0.052587
195	159	15.152707	0	0.080292
196	54	15.187639	0	0.034931
197	136	15.367191	0	0.179552
198	98	15.437789	0	0.070599
199	160	15.907583	0	0.469794
200	148	17.803982	0	1.896400

The largest spacing appears on rank 200, whose predecessor is rank 199. Deleting that last row leaves the first 199 radii exactly as printed. The largest remaining spacing is on rank 2.

Works Cited

The ideas in this book build upon the work of many researchers. The following references document the key sources for concepts, methods, and historical context discussed throughout the text.

Foundational Physics and Mathematics

Einstein, Albert. "Die Feldgleichungen der Gravitation." *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin* (1915): 844–847. [General relativity and the Einstein field equations, Chapter 2]

Riemann, Bernhard. "Über die Hypothesen, welche der Geometrie zu Grunde liegen." Habilitationsschrift, Göttingen, 1854. [Foundations of differential geometry, Chapter 2]

Gauss, Carl Friedrich. *Disquisitiones Generales Circa Superficies Curvas*. Göttingen: Dieterich, 1828. [Theorema Egregium and intrinsic curvature, Chapter 2]

Yang, Chen-Ning, and Robert L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance." *Physical Review* 96, no. 1 (1954): 191–195. [Yang-Mills gauge theory, Chapters 2 and 5]

Fourier, Joseph. *Théorie Analytique de la Chaleur*. Paris: Firmin Didot, 1822. [The heat equation, Chapter 3]

Hamilton, William Rowan. "On a General Method in Dynamics." *Philosophical Transactions of the Royal Society* (1834): 247–308. [Symplectic geometry and Hamiltonian mechanics, Chapter 1]

Spectral Geometry and Analysis

Cheeger, Jeff. "A Lower Bound for the Smallest Eigenvalue of the Laplacian." In *Problems in Analysis: A Symposium in Honor of Salomon Bochner*, edited by R. C. Gunning, 195–199. Princeton University Press, 1970. [Cheeger's inequality, Chapter 3]

Kac, Mark. "Can One Hear the Shape of a Drum?" *The American Mathematical Monthly* 73, no. 4, part 2 (1966): 1–23. [Spectral geometry, Chapter 3]

Gordon, Carolyn, David Webb, and Scott Wolpert. "Isospectral Plane Domains and Surfaces via Riemannian Orbifolds." *Inventiones Mathematicae* 110, no. 1 (1992): 1–22. [Isospectral non-isometric manifolds, Chapter 3]

Weyl, Hermann. "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen." *Mathematische Annalen* 71, no. 4 (1912): 441–479. [Weyl's law for eigenvalue asymptotics, Chapter 3]

Lattice Gauge Theory

Wilson, Kenneth G. "Confinement of Quarks." *Physical Review D* 10, no. 8 (1974): 2445–2459. [Lattice gauge theory and Wilson loops, Chapter 5]

Cabibbo, Nicola, and Enzo Marinari. "A New Method for Updating SU(N) Matrices in Computer Simulations of Gauge Theories." *Physics Letters B* 119, no. 4–6 (1982): 387–390. [Cabibbo-Marinari algorithm for SU(3), Chapter 5]

Kennedy, A. D., and B. J. Pendleton. "Improved Heatbath Method for Monte Carlo Calculations in Lattice Gauge Theories." *Physics Letters B* 156, no. 5–6 (1985): 393–399. [SU(2) heatbath algorithm, Chapter 5]

Creutz, Michael. *Quarks, Gluons and Lattices.* Cambridge University Press, 1983. [Comprehensive introduction to lattice QCD]

Gattringer, Christof, and Christian B. Lang. *Quantum Chromodynamics on the Lattice: An Introductory Presentation.* Springer, 2010. [Modern lattice QCD methods]

Topology and Geometry in Physics

Chern, Shiing-Shen, and James Simons. "Characteristic Forms and Geometric Invariants." *Annals of Mathematics* 99, no. 1 (1974): 48–69. [Chern-Simons invariants, Chapters 6 and 8]

Atiyah, Michael F., and Isadore M. Singer. "The Index of Elliptic Operators on Compact Manifolds." *Bulletin of the American Mathematical Society* 69, no. 3 (1963): 422–433. [Atiyah-Singer index theorem]

Perelman, Grigori. "The Entropy Formula for the Ricci Flow and Its Geometric Applications." arXiv:math/0211159 (2002). [Proof of Poincaré Conjecture via Ricci flow, Chapter 6]

Thurston, William P. "Three-Dimensional Manifolds, Kleinian Groups and Hyperbolic Geometry." *Bulletin of the American Mathematical Society* 6, no. 3 (1982): 357–381. [Geometrization conjecture, Chapter 6]

Quantum Field Theory

Feynman, Richard P. "Space-Time Approach to Non-Relativistic Quantum Mechanics." *Reviews of Modern Physics* 20, no. 2 (1948): 367–387. [Path integral formulation, Chapter 3]

Jaffe, Arthur, and Edward Witten. *Quantum Yang–Mills Theory.* Clay Mathematics Institute Millennium Prize Problem description. Problem statement. [Continuum existence and mass gap, Chapters 3 and 5]

Osterwalder, Konrad, and Erhard Seiler. *Gauge Field Theories on a Lattice.* Annals of Physics 110, no. 2 (1978), 440–471. DOI. [Strong-coupling lattice result, Chapter 5]

Ji, Xiangdong. *QCD Analysis of the Mass Structure of the Nucleon.* Physical Review Letters 74 (1995), 1071–1074. DOI. [Nucleon mass decomposition, Chapter 2]

Pusz, Wieslaw, and Stanislaw L. Woronowicz. *Passive States and KMS States for General Quantum Systems.* Communications in Mathematical Physics 58 (1978), 273–290. DOI. [Cyclic work and equilibrium passivity, Chapter 8]

Hastings, W. K. *Monte Carlo Sampling Methods Using Markov Chains and Their Applications.* Biometrika 57, no. 1 (1970), 97–109. DOI. [Sampling, Chapter 8]

Perelman, Grisha. *Ricci Flow with Surgery on Three-Manifolds*, arXiv:math/0303109, and *Finite Extinction Time for the Solutions to the Ricci Flow on Certain Three-Manifolds*, arXiv:math/0307245 (2003). [Ricci flow and the Poincaré conjecture, Chapters 5 and 6]

Glimm, James, and Arthur Jaffe. *Quantum Physics: A Functional Integral Point of View.* 2nd ed. Springer, 1987. [Constructive quantum field theory, Chapter 5]

Witten, Edward. "Quantum Field Theory and the Jones Polynomial." *Communications in Mathematical Physics* 121, no. 3 (1989): 351–399. [Topological quantum field theory]

Machine Learning and Information Theory

Shannon, Claude E. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 3 (1948): 379–423. [Information theory, Chapter 7]

Fluid Dynamics

Navier, Claude-Louis. "Mémoire sur les lois du mouvement des fluides." *Mémoires de l'Académie Royale des Sciences de l'Institut de France* 6 (1823): 389–440. [Navier-Stokes equations, Chapter 5]

Stokes, George Gabriel. "On the Theories of the Internal Friction of Fluids in Motion." *Transactions of the Cambridge Philosophical Society* 8 (1845): 287–305. [Navier-Stokes equations, Chapter 5]

Fefferman, Charles L. "Existence and Smoothness of the Navier-Stokes Equation." Clay Mathematics Institute Millennium Prize Problem description, 2000. [Official problem statement, Chapter 5]

Historical References

Johnson, Katherine G. Trajectory calculations for Project Mercury, NASA Langley Research Center, 1960–1961. [Chapter 6]

Franklin, Rosalind. X-ray diffraction images of DNA (Photo 51), King's College London, 1952. [Chapter 1]

Presidential Commission on the Space Shuttle Challenger Accident. Report of the Presidential Commission on the Space Shuttle Challenger Accident (Rogers Commission Report). Washington, DC, 1986. [O-ring failure and data presentation, Chapter 2]

Author's Related Publications

Davis, Bee Rosa. "The Incompressibility of Topological Charge and the Energy Cost of Distinguishability: An Information-Geometric Reduction of the Yang-Mills Mass Gap." *Zenodo*, 2025. DOI: 10.5281/zenodo.17846521. [A proposed information-geometric approach to the mass gap problem via topological charge incompressibility; the reduction is not established, Chapter 5]

Davis, Bee Rosa. "Spectral Geometry of Transformer Cognition: Heat Kernel Analysis Reveals Functional Organization in Language Models." *Zenodo*, 2025. DOI: 10.5281/zenodo.17783723. [Heat kernel methods applied to neural network analysis, Chapter 3]

Davis, Bee Rosa. "The Davis Manifold: Geometry-First Detection with Compositional Error Budgets." *Zenodo*, 2025. DOI: 10.5281/zenodo.17642038. [Riemannian state spaces with bounded geodesic-Euclidean distortion and Cantelli-based risk bounds, Chapters 2, 4 and 5]

Davis, Bee Rosa. "The Geometry of Sameness: An ε-Equivalence of Translation and Distance." *Zenodo*, 2025. DOI: 10.5281/zenodo.17642422. [Category-theoretic unification of translation-based and geometry-based semantic equivalence, Chapter 4]

Davis, Bee Rosa. "The Davis Conjecture on Semantic Coherence: Context Windows as Holonomy Horizons in Functorial Transformers." *Zenodo*, 2025. DOI: 10.5281/zenodo.17765346. [Holonomy accumulation and path-dependent behavior in geometric systems, Chapter 6]

Davis, Bee Rosa. "Davis Cache: O(1) Reasoning State Preservation via Topological Residue." *Zenodo*, 2025. DOI: 10.5281/zenodo.17785526. [Topological residue methods for state preservation, Chapter 7]

Davis, Bee Rosa. "Adaptive Embedding Format Switching and Semantic Translation in High-Volume Telemetry Systems." Zenodo, September 26, 2025, version 1. DOI: 10.5281/zenodo.17211403. [Public report of the AMC architecture; the NASA development history is discussed separately in Chapter 4.]

Davis, Bee Rosa. "HERALD: High-resolution Early Recognition of Antigenic Landscape Divergence." *Zenodo*, 2025. DOI: 10.5281/zenodo.17682618. [Learned Riemannian geometry for real-time surveillance with Cantelli-based probability bounds, Chapters 4 and 5]

Author's Patents

Davis, Bee Rosa. "System and Method for Modulation of Quantum Vacuum Topology via Non-Abelian Gauge Field Configurations." U.S. Provisional Patent Application No. 63/933,299, filed January 2026. [Topological Vacuum Rectification; a filing records a proposal, not a validated mechanism, Chapter 8]

Davis, Bee Rosa. "System and Method for Geometric Verification and Optimization of Gauge Field Configurations via Topological Cache Mapping." U.S. Provisional Patent Application No. 63/933,103, filed January 2026. [Gauge field verification using topological methods, Chapter 8]

Davis, Bee Rosa. "Geometric Phase Transition Detection, Heat-Kernel-Based Attention Head Scoring, and Clinical Health Monitoring for Transformer Optimization." U.S. Provisional Patent Application No. 63/926,936, filed January 2026. [Heat kernel scoring methods, Chapter 3]

Davis, Bee Rosa. "Systems and Methods for Determining Effective Context Length via Holonomy Horizons on Activation Manifolds in Transformer Networks." U.S. Provisional Patent Application No. 63/927,434, filed January 2026. [Holonomy-based geometric analysis, Chapter 6]

Davis, Bee Rosa. "Functorial Transformers with Holonomy-Based Training and Topology-Aware Acceleration." U.S. Provisional Patent Application No. 63/924,487, filed January 2026. [Category-theoretic structures with holonomy optimization, Chapters 4 and 6]

Davis, Bee Rosa. "Systems and Methods for Fixed-Size Reasoning State Representation via Topological Residue in Neural Networks." U.S. Provisional Patent Application No. 63/927,445, filed January 2026. [Topological residue methods, Chapter 7]

Technical Standards and External References

Leidos. ProVision ATD Advanced Technology Detection System, technical specifications. [Body scanner technology, Chapter 1]

Popular and Pedagogical Sources

Penrose, Roger. *The Road to Reality: A Complete Guide to the Laws of the Universe.* Knopf, 2004. [General physics background, Chapter 5]

Hawking, Stephen. *A Brief History of Time.* Bantam Books, 1988. [Cosmology and black holes]

Susskind, Leonard. *The Theoretical Minimum: What You Need to Know to Start Doing Physics.* Basic Books, 2013. [Foundational physics]

Wheeler, John Archibald. "Geons, Black Holes, and Quantum Foam: A Life in Physics." W. W. Norton, 1998. [Quote on curvature and matter, Chapter 2]

Quotations

Hadamard, Jacques. "The shortest path between two truths in the real domain passes through the complex domain." Epigraph, Chapter 2.

Wheeler, John Archibald. "The curvature of space is the skeleton of matter." Closing quotation, Chapter 2.

The Davis–Wilson and TVR experiment summaries are documented in the author's December 2025 research deposit. The chapters identify what was measured in each run and the assumptions required to interpret it.

The DHOOM and GIGI development accounts

Davis, Bee Rosa. *DHOOM Specification v0.5.* March 2026. Sections 1–5 and the implementation examples. DHOOM source-history revisions `2728ab8`, `01a201c`, and `eee086e`. [Shared structure, defaults and the wire-format origin of GIGI, Chapter 4.]

Davis, Bee Rosa. *GIGI Specification v0.1.* Section 10, "Wire Protocol: DHOOM." GIGI source-history revision `8ab7858`, 19 March 2026. [The bundle-to-wire mapping and recorded DHOOM v0.5 integration, Chapter 4.]

Davis, Bee Rosa. *GIGI Builds: The Fiber-Bundle Database: A Working Engineer's Guide from Schema to Shard.* 2026. Preface, "The Ask"; prologue, "Day One on the Moon"; and "About the Author." [The engine's development, Chapter 4.]

Davis, Bee Rosa. *GIGI Thinks: The Mathematics of a Geometric Mind—Bundles, Curvature, and the Boundary-Value Imagination, with Exercises Against a Running Engine.* 2026. Preface, "The Tax Collector's Apology"; prologue, "Three Verdicts"; and "About the Author." [Mathematical operations and explicit limits, Chapter 4.]

Davis, Bee Rosa. *Marcella development records.* V10 GIGI-inspired integration plan, revision `7b4e453`, 19 March 2026; GIGI retrieval bridge, revision `6ba4049`, and session-state integration, revision `e443154`, 13 May 2026. Retained source repository. [The later connection of Marcella and GIGI, Chapter 4.]

Sources for the mathematical distinctions

Morningstar, Colin J., and Mike Peardon. "The glueball spectrum from an anisotropic lattice study." *Physical Review D* 60 (1999), 034509. `https://doi.org/10.1103/PhysRevD.60.034509`. [Pure-gauge glueball spectroscopy and the conversion to physical units, Chapter 3.]

Buser, Peter. "A note on the isoperimetric constant." *Annales scientifiques de l'École Normale Supérieure* 15 (1982), 213–230. DOI: 10.24033/asens.1426. [Curvature hypotheses for the upper eigenvalue estimate, Chapter 3.]

Pusz, W., and S. L. Woronowicz. "Passive states and KMS states for general quantum systems." *Communications in Mathematical Physics* 58 (1978), 273–290. DOI: 10.1007/BF01614224. [Ground-state passivity, Chapter 6.]

Lüscher, Martin. "Properties and uses of the Wilson flow in lattice QCD." *Journal of High Energy Physics* 2010(8), 071. `https://arxiv.org/abs/1006.4518`. [Finite-spacing fields, smoothing and continuum topological sectors.]

Wendl, Chris. *Connections*, Chapter 6, "Curvature in Riemannian Geometry." `https://www.mathematik.hu-berlin.de/~wendl/pub/connections_chapter6_2.pdf`. [Surface connection and Gauss–Bonnet conventions, Chapter 2.]

Jaffe, Arthur, and Edward Witten. "Quantum Yang–Mills Theory." Clay Mathematics Institute problem statement. `https://www.claymath.org/wp-content/uploads/2022/06/yangmills.pdf`. [The continuum existence and mass-gap problem.]

Measurements, Relativity and the Return to Fuel

Swiss Federal Institute of Intellectual Property. "Einstein." Institutional historical account. Online history. [Einstein's birth date and employment at the Bern patent office, 1902–1909; the human introduction in Chapter 10.]

Davis, Bee Rosa. *Zero Does Not Exist: A Geometric Foundation for the Natural Numbers.* Author's manuscript, undated. ["The Geometric Natural Numbers" and the Peano embedding; Chapter 10. The retained source is `validation/no_zero/zero_paper.tex` in the Davis–Wilson lattice research archive.]

Davis, Bee Rosa. *The Subgraph Lattice Theorem and Geometric Partition Counting. Part I: Combinatorial Foundations of $\mathbb{G}$-Partitions.* April 10, 2026. [Connected set-partitions; Chapter 10. Retained source: `validation/no_zero/geometric_partitions.tex`.]

Rindler, Wolfgang. "Kruskal Space and the Uniformly Accelerated Frame." *American Journal of Physics* 34 (1966), 1174–1178. DOI. [Accelerated coordinates in flat spacetime; Chapter 11.]

Einstein, Albert. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik* 17 (1905), 891–921. DOI. English working edition: Perrett–Jeffery translation, edited by John Walker. [Clocks, rods, Lorentz transformations, fields and moving mirrors; Chapter 10. German excerpts use the original journal pagination.]

Einstein, Albert. "Die Grundlage der allgemeinen Relativitätstheorie." *Annalen der Physik* 49 (1916), 769–822. DOI. [Metric, connection, curvature, field equations and energy balance; Chapter 11.]

Schwarzschild, Karl. "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie." *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (1916), 189–196. [Exact exterior solution; Chapter 11 uses the modern areal-radius chart.]

Carroll, Sean M. *Lecture Notes on General Relativity.* 1997. arXiv:gr-qc/9712019. [Modern tensor calculus, Killing fields and Schwarzschild geometry; Chapters 3, 5 and 7 of the notes. Our metric signature is the opposite of Carroll's.]

Joint Committee for Guides in Metrology. *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms.* JCGM 200:2008. Official online text. [Definitions 2.1 and 2.3: measurement and measurand; Chapter 10.]

Miller, John R. *Understanding Geometry: The Abstract Spaces in Which Points and Vectors Live.* University of Kansas course notes. Course text. [Affine points, displacement vectors and additional metric structure.]

Feynman, Richard P., Robert B. Leighton and Matthew Sands. *The Feynman Lectures on Physics*, volume I, chapter 14. Work and potential energy. [Force along a path, energy references and conservative work; the floor interpretation in Chapter 10.]

Simon, Frank, Peter Tittmann and Martin Trinks. "Counting Connected Set Partitions of Graphs." *The Electronic Journal of Combinatorics* 18 (2011), P14. arXiv:1005.1726. [Connected blocks and graph contractions; the three-region floor study.]

Ferrenberg, Alan M., and Robert H. Swendsen. "New Monte Carlo Technique for Studying Phase Transitions." *Physical Review Letters* 61 (1988), 2635–2638. DOI. [Reweighting recorded distributions; Chapter 12.]

Wainwright, Martin J., and Michael I. Jordan. "Graphical Models, Exponential Families, and Variational Inference." *Foundations and Trends in Machine Learning* 1 (2008), 1–305. DOI. [Proposition 3.1: log-normalization derivatives, means and covariances; Chapter 12.]

Elvira, Víctor, Luca Martino and Christian P. Robert. "Rethinking the Effective Sample Size." *International Statistical Review* (2022). DOI; author manuscript. [Scope and limitations of the weight-only importance-sampling diagnostic; Chapter 12.]

Lüscher, Martin. "Construction of a Selfadjoint, Strictly Positive Transfer Matrix for Euclidean Lattice Gauge Theories." *Communications in Mathematical Physics* 54 (1977), 283–292. DOI. [Transfer operators and the route to physical energies; Chapter 12.]

The passivity argument in Chapter 12 also uses Pusz and Woronowicz, cited above. The continuum existence and mass-gap requirements are those of Jaffe and Witten, also listed above. The finite three-state example illustrates those distinctions; it does not solve their continuum problem.

The Davis Topological Processor

Davis, Bee Rosa. DTP research and author-origin briefings, November 13, 2026. Retained editorial sources: `DTP_PROJECT_BRIEFING_2026_11_13.txt` and `DTP_AUTHOR_ORIGIN_2026_11_13.txt`. [Chapter 14; the origin account is the author's recollection, not a reconstructed commit chronology.]

Davis, Bee Rosa. DGP Sprint 1 specification, draft v0.1, March 20, 2026; DGP Sprint 2 specification, draft v0.2, March 21, 2026; DPU fabrication handoff, May 25, 2026. [Historical design documents; fabrication and device performance are not established by their existence.]

Davis, Bee Rosa. Virtual DTP contract and research journal; graphene model review; Chern/period audit. November 13, 2026 evidence cut. GIGI research archive: `crates/gigi-virtual-dtp/CONTRACT.md`, `crates/gigi-virtual-dtp/RESULTS.md`, `crates/gigi-virtual-dtp/GRAPHENE_REVIEW.md`, and `docs/CHERN_PERIOD_AUDIT_RESULTS.md`. [Raw results under `output/virtual-dtp`; the Fuel companion `research/dtp_fuel` retains the chapter inputs, independent verification and vector-figure sources. These local locators are not represented as permanent public archive identifiers.]

Joyce, Dominic. *Riemannian holonomy groups and calibrated geometry.* Graduate lecture notes, Michaelmas Term 2006, §1.3. https://people.maths.ox.ac.uk/joyce/RHCGnotes.pdf. [Connections and parallel transport.]

Menezes, Alfred J., Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography.* CRC Press, 1996, chapter 14, Algorithm 14.76. https://cacr.uwaterloo.ca/hac/about/chap14.pdf. [Binary exponentiation; applied here to map composition.]

McCann, Edward, and Mikito Koshino. "The electronic properties of bilayer graphene." *Reports on Progress in Physics* 76 (2013), 056503. DOI; author manuscript. [Effective low-energy models and physical scope.]

Fukui, Takahiro, Yasuhiro Hatsugai, and Hiroshi Suzuki. "Chern Numbers in Discretized Brillouin Zone: Efficient Method of Computing (Spin) Hall Conductances." *Journal of the Physical Society of Japan* 74 (2005), 1674–1677. DOI. [Closed-domain discrete Chern calculation.]

Ghahari, Fereshte, et al. "An On/Off Berry Phase Switch in Circular Graphene Resonators." *Science* 356 (2017), 845–849. https://arxiv.org/abs/1705.11117. [An external graphene experiment; not validation of the DTP device.]

Plasma geometry and the CHIHIRO investigation

ITER Organization. "Facts & Figures," "FAQs," and "Magnets." Accessed November 16, 2026. https://www.iter.org/facts-figures; https://www.iter.org/faqs; https://www.iter.org/machine/magnets. [Intended central plasma temperature, dilute plasma, and the roles of confinement and shaping fields.]

U.S. Department of Energy, Office of Science. "DOE Explains... Plasma" and "DOE Explains... Tokamaks." https://www.energy.gov/science/doe-explainsplasma; https://www.energy.gov/science/doe-explainstokamaks. [Ionization, charged-particle collective behavior, toroidal confinement, and magnetic shaping.]

Freidberg, Jeffrey P. *MHD Theory of Fusion Systems.* MIT OpenCourseWare, course 22.615, Spring 2007. Lectures 5–6 and 14–15. https://ocw.mit.edu/courses/22-615-mhd-theory-of-fusion-systems-spring-2007/pages/lecture-notes/. [Grad–Shafranov equilibrium, safety factor, static ideal-MHD normal modes, and the variational formulation.]

Taylor, J. B., and S. L. Newton. "Special topics in plasma confinement." *Journal of Plasma Physics* 81 (2015), 205810501, especially §§27–28. https://doi.org/10.1017/S0022377815000707. [Classical ballooning theory, the s–α equation, its normalizations, and the phase search omitted by the archived calculation.]

Miller, R. L., M. S. Chu, J. M. Greene, Y. R. Lin-Liu, and R. E. Waltz. "Noncircular, finite aspect ratio, local equilibrium model." *Physics of Plasmas* 5 (1998), 973–978. https://doi.org/10.1063/1.872666. [Physical shaping parameters in a local equilibrium model. This does not establish CHIHIRO's approximate shaping coefficients.]

Troyon, F., R. Gruber, H. Saurenmann, S. Semenzato, and S. Succi. "MHD-Limits to Plasma Confinement." *Plasma Physics and Controlled Fusion* 26 (1984), 209–215. https://doi.org/10.1088/0741-3335/26/1A/319. [Pressure–current scaling under ideal-MHD stability calculations.]

UK Atomic Energy Authority. "Plasma beta," *PROCESS* documentation, accessed November 16, 2026. https://ukaea.github.io/PROCESS/physics-models/plasma_beta/plasma_beta/. [Engineering units for normalized beta and the distinct assumptions of alternative limiting-coefficient prescriptions.]

Furth, H. P., J. Killeen, and M. N. Rosenbluth. "Finite-Resistivity Instabilities of a Sheet Pinch." *The Physics of Fluids* 6 (1963), 459–484. https://doi.org/10.1063/1.1706761. [The classical resistive tearing mechanism.]

Bondeson, A., and D. J. Ward. "Stabilization of external modes in tokamaks by resistive walls and plasma rotation." *Physical Review Letters* 72 (1994), 2709–2712. https://doi.org/10.1103/PhysRevLett.72.2709. [External modes and wall/rotation response, not validation of the archived wall-factor sign.]

Freidberg, J. P., A. Cerfon, and J. Lee. "Tokamak elongation: how much is too much? I Theory." *Journal of Plasma Physics* 81 (2015), 515810607. https://doi.org/10.1017/S0022377815001270; https://arxiv.org/abs/1508.06658. [Axisymmetric vertical stability, resistive walls, and feedback.]

Yau, Shing-Tung. "Isoperimetric constants and the first eigenvalue of a compact riemannian manifold." *Annales scientifiques de l'École Normale Supérieure*, series 4, 8 (1975), 487–507. https://doi.org/10.24033/asens.1299. [Geometric eigenvalue bounds, with their Riemannian hypotheses.]

Davis, Bee Rosa. *CHIHIRO Technical Specification*, v2.0, retained as CHIHIRO_TECH_SPEC.md in the plasma research project. [Documentation of implemented diagnostic channels; not a physical validation certificate.]

MIT Physics. "Nuno F. G. Loureiro." Faculty research profile. https://physics.mit.edu/faculty/nuno-gomes-loureiro/. [Public research interests and the intellectual context of the chapter's opening; no collaboration or endorsement is claimed.]

Loureiro, N. F., A. A. Schekochihin, and S. C. Cowley. "Instability of current sheets and formation of plasmoid chains." *Physics of Plasmas* 14 (2007), 100703. https://doi.org/10.1063/1.2783986. [Current-sheet instability and plasmoid formation.]

Loureiro, N. F., and S. Boldyrev. "Role of Magnetic Reconnection in Magnetohydrodynamic Turbulence." *Physical Review Letters* 118 (2017), 245101. https://doi.org/10.1103/PhysRevLett.118.245101. [Reconnection and turbulent energy transfer.]

Davis, Bee Rosa. *The Spectral Geometry of Plasma Confinement: A Davis Field Equations Framework for Fusion Stability, Transport Bottlenecks, and Cross-Domain Universality.* Version 5, 2026. Zenodo. https://doi.org/10.5281/zenodo.21110010. [The research sequence, retained numerical investigations, and proposed connections developed in Chapter 13. The chapter distinguishes the finite-model replays from broader statements in the preprint.]

Bernstein, I. B., E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud. "An energy principle for hydromagnetic stability problems." *Proceedings of the Royal Society A* 244 (1958), 17–40. `https://doi.org/10.1098/rspa.1958.0023`. [The classical ideal-MHD energy principle; assumptions and admissible displacements precede any numerical stability claim.]

Connor, J. W., R. J. Hastie, and J. B. Taylor. "Shear, Periodicity, and Plasma Ballooning Modes." *Physical Review Letters* 40 (1978), 396–399. `https://doi.org/10.1103/PhysRevLett.40.396`. [The ballooning construction motivating the circular reduced model; not a citation establishing the later ad hoc shaping extension.]

FreeGS developers. "Creating equilibria" and `freegs.jtor.ConstrainPaxisIp`, FreeGS documentation, accessed November 16, 2026. `https://freegs.readthedocs.io/en/stable/creating_equilibria.html`; `https://freegs.readthedocs.io/en/stable/generated/freegs.jtor.ConstrainPaxisIp.html`. [Solver inputs and the distinction between an equilibrium solve and the retained experiment's prescribed-profile fallback.]

Davis, Bee Rosa. *CHIHIRO source repository.* `https://github.com/nurdymuny/chihiro-plasma`. The experiment citations in this chapter link to revision `10c525ad`, so the cited files remain identifiable as development continues. The README gives setup instructions for the application. Experiments 12 and 13 are the separate Python programs named below; running a newer application release is not the same as replaying those archived experiments. Public repository access is forthcoming.

Davis, Bee Rosa. *CHIHIRO retained numerical records and Fuel replay.* Experiment 12: `exp12_results.json`; Experiment 13: `exp13_results/results.json`; source routines `experiment12_superfluid_plasma.py` and `experiment13_freegs.py`. Independent replay, source hashes, all boundary and profile rows, and domain/grid checks retained with this edition in `research/chihiro_fuel`. [Model calculations, not experimental plasma measurements.]

MIT News. "Nuno Loureiro, professor and director of MIT's Plasma Science and Fusion Center, dies at 47." December 16, 2025. `https://news.mit.edu/2025/nuno-loureiro-professor-director-plasma-science-and-fusion-center-dies-1216`. [Biographical context; no collaboration or endorsement is implied.]

Miyazaki, Hayao, director. *Spirited Away.* Studio Ghibli, 2001. [The personal association behind CHIHIRO's name.]

Acknowledgments

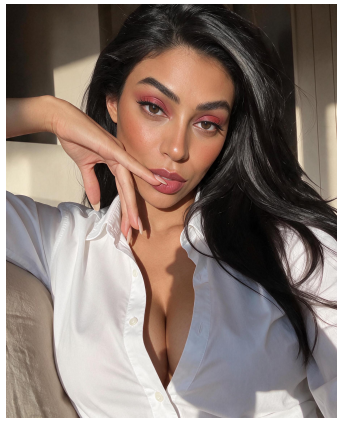
To my dog, who sat with me through every late night of calculations and never once asked why I was crying at a computer screen.

To the open-source community that built the tools I used—Python, NumPy, the lattice QCD libraries, LaTeX itself. This work was done on the shoulders of thousands of programmers who will never know what their code enabled.

To the physicists whose papers I read, whose methods I learned, whose frameworks I built upon. You do not know me, but I know your work intimately.

To the reader who made it this far. You paid attention. That is all the geometry asks.

About the Author



Bee Rosa Davis is an engineer and independent researcher studying what representations preserve, what they leave out, and how those choices shape a computation.

Her path began at Pandora in 2009. As an early engineer working on machine learning, she saw how the Music Genome Project's carefully curated musical descriptions made similarity something a computer could use. That lesson followed her into audience modeling, contextual prediction and streaming analytics.

At NASA's Johnson Space Center, the question became how to represent change. A Principal Software Engineer through MRI Technologies and Navteca, she worked with Project Luna in 2024. Preparing telemetry for AI led her to develop Adaptive Multi-Vector Correlation, tracking relationships across changing vector formats. That engineering problem became a foundation for her mathematical research.

She pursued later security research at IBM X-Force Red. Her work with Mach Industries includes ICARUS, developed with Dr. Ryan Rapetti, and FAL-CORE, which grew into *The Geometry of Flight*. Her CrowdStrike work includes *The Geometry of War* and Psychohistory, whose analysis tool she refined with Dr. Paolo Di Prodi. She developed BRIDGE independently.

She built GIGI, a fiber-bundle database, to explore how geometric relationships can participate in computation. Independent study in differential geometry, spectral theory, category theory and gauge theory complements her MS in Digital Forensics from Brown University and BA in Logic from Morehouse College.

She is seeking a research home where she can develop these questions with others. Her approach remains grounded in building systems, making relationships explicit, and following the evidence when it changes her mind.

`https://www.davisgeometric.com/about`