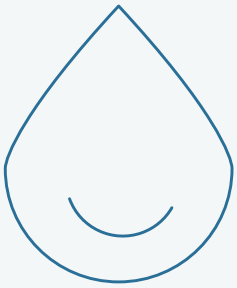
PROBABILISTIC GRAPH MODELING



Probabilistic Modeling of Water Distribution Networks

Leaks, state estimation and where to put the next sensor

386 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Probabilistic Modeling of Water Distribution Networks

Ricardo Marquez

Working draft, September 2026

Preface

This book is written for an engineer who knows conservation laws and linear algebra and has never met a probabilistic graphical model. It takes one idea and follows it through: that the structure of a physical system, the way conserved quantities accumulate at places and move between them and cross the system's boundary, is also the structure of every probabilistic question one can ask about that system. A model of the physics is a graph. A model of what we do not know about the physics is the same graph with probability attached to it. Inference on that graph is how a measurement in one place becomes knowledge about another, how a failure is diagnosed, how a decision is justified.

A water distribution network is a good system to point that idea at. Its physics is settled: a mass balance at every junction and a head-loss law along every pipe, equations engineers have solved by hand and by computer for the better part of a century. So the difficulties the book meets are difficulties of knowing, not of modeling. Its instruments are few: a meter at each district inlet and pumping station, pressure loggers on a handful of hydrants, and thousands of pipes in between that nobody measures. Its parameters age: roughness coefficients were assigned when the pipes were laid and have been changing since. And its most expensive questions are inference questions. Where is the water going that enters a district and never reaches a customer? Which of several places could a burst be? What can the network still deliver with a main out of service?

Every system worth modeling is a piece cut out of a larger one, trading a conserved quantity with an environment it does not model. For a water network the piece is a district, a pressure zone or a whole city. The boundary is where the utility's model stops: the reservoirs and sources that feed it, the service connections that draw from it, and the metered mains that cross into a neighboring district. That boundary is where the modeler's knowledge is thinnest. As the later chapters show, it is also where the mathematics of inference does its most useful work, because the flows and heads at a boundary are exactly what the rest of the network needs to know about the inside. Chapter 1 gives such a system a name, a *boundary-exchanging system*, and the book returns to the boundary at every level.

Nothing in the method of Parts I to V is specific to water. The balances, closures and ports are the same whether the conserved quantity is mass, heat, charge or power, and Table 1.3 lines the four up. This book carries one conserved quantity, mass, from beginning to end. The one exception, the district cooling loop of Example 1.3, puts heat on the same graph to show that nothing in the method assumes a single ledger. Three companion editions take the same method text through electric power transmission, district heating and cooling, and data centers. Each is complete in itself, and a reader of this one needs none of them.

The Introduction says why the book exists, what it inherits from four older traditions, what it adds, and how to read it. Part I is deterministic and establishes the objects. Part II attaches probability to them and states, rather than skips, the places where that attachment needs care.

Part III is inference. Part IV asks the questions an operator actually asks. Part V handles time, hierarchy and scale. Part VI is four complete studies. They run from a loop small enough to check by hand, through the EPA's Net3 and a leak-detection benchmark, to a year of recorded operating data from the L-Town network, scored against the benchmark's published entries. Each gives up one of the modeler's advantages that the one before it had, and each can be reproduced from the material in the earlier parts. Each chapter ends with a short section of notes that says which of its ideas are inherited and from where, and which are the book's own.

Contents

Preface **ii**

0 Introduction **1**
0.1 The questions a water engineer asks . . . 1
0.2 Where the graph comes from . . . 2
0.3 One subject, not four . . . 3
0.4 Probability in the structure, not around the solver . . . 4
0.5 The boundary . . . 6
0.6 From "what is the state" to "what can I know" . . . 6
0.7 What is inherited and what is the book's . . . 7
0.8 How to read this book . . . 7
0.9 What this book does not do . . . 9
Notes and further reading . . . 9

I Physical systems as graphs **10**

1 Conservation, closure, and the boundary **11**
1.1 Three questions about any system . . . 11
1.2 Nodes conserve . . . 12
1.2.1 The balance law . . . 12
1.2.2 Node roles . . . 13
1.2.3 Incidence . . . 13
1.3 Edges close . . . 14
1.3.1 Closure relations . . . 14
1.3.2 A closure is a relation, not an assignment . . . 14
1.3.3 Parameters . . . 15
1.4 The boundary . . . 15
1.4.1 Reservoirs and imposed conditions . . . 15
1.4.2 Exactness at a cut . . . 16
1.4.3 Why the boundary matters most . . . 17
1.5 The residual system . . . 17
1.5.1 Well-posedness . . . 18
1.5.2 Sparsity . . . 18
1.6 Three worked examples . . . 19
1.7 In practice . . . 27

1.8 What is missing . . . 28
Notes and further reading . . . 29
Summary . . . 29
Exercises . . . 30
Solutions to selected exercises . . . 30

2 From equations to factor graphs 32
2.1 The rows as a graph . . . 32
2.2 Three languages . . . 33
2.2.1 Directed graphs . . . 33
2.2.2 The directionality problem . . . 34
2.2.3 Undirected graphs . . . 36
2.2.4 Factor graphs . . . 36
2.3 The physical system as a factor graph . . . 37
2.3.1 A junction is not a variable . . . 38
2.3.2 Two domains on one graph . . . 38
2.4 The factor graph is not unique . . . 39
2.5 Vocabulary . . . 41
2.6 In practice . . . 42
2.7 What is missing . . . 42
Notes and further reading . . . 43
Summary . . . 43
Exercises . . . 44
Solutions to selected exercises . . . 44

II Probability on a physical graph 46

3 Where probability enters 47
3.1 A probability on the state . . . 47
3.2 Five kinds of factor . . . 48
3.2.1 Priors . . . 48
3.2.2 Closure factors . . . 49
3.2.3 Conservation factors . . . 49
3.2.4 Measurement factors . . . 50
3.2.5 Failure factors . . . 50
3.2.6 The complete list . . . 51
3.3 The joint as an objective . . . 51
3.3.1 Residuals in units of their widths . . . 52
3.3.2 Conditioning a Gaussian on one reading . . . 52
3.4 A worked example: one sensor, then two . . . 55
3.5 A second example: the triangle with a flow meter . . . 58
3.6 Reading the graph . . . 60
3.7 In practice . . . 61
3.8 What is missing . . . 62
Notes and further reading . . . 62
Summary . . . 63

Exercises . . . 63
Solutions to selected exercises . . . 64

4 Hard constraints, soft constraints, and the manifold 65
4.1 The hard joint and its manifold . . . 65
4.2 Conditioning on the manifold . . . 66
4.3 The soft joint . . . 67
4.4 Why soften . . . 68
4.5 Choosing the widths . . . 69
4.6 Worked example: the chain with soft physics . . . 71
4.7 A second example: the loop with an unmodeled leak . . . 73
4.8 In practice . . . 76
4.9 What is missing . . . 76
Notes and further reading . . . 77
Summary . . . 77
Exercises . . . 78
Solutions to selected exercises . . . 78

5 Factorization, separators, and treewidth 80
5.1 Factorization is not independence . . . 80
5.2 Reading independence off the graph . . . 81
5.3 Explaining away, precisely . . . 82
5.4 Ports are separators . . . 83
5.5 Elimination and fill . . . 84
5.5.1 One elimination . . . 84
5.5.2 Order and width . . . 85
5.5.3 What this means for a physical graph . . . 86
5.6 Worked examples: counting width and fill . . . 87
5.7 Why separators matter for cost . . . 89
5.8 In practice . . . 91
5.9 What is missing . . . 92
Notes and further reading . . . 92
Summary . . . 92
Exercises . . . 93
Solutions to selected exercises . . . 93

6 The linear-Gaussian case 95
6.1 The Gaussian in information form . . . 95
6.2 The most probable state . . . 96
6.2.1 Gauss–Newton against Newton . . . 97
6.3 Elimination is Cholesky . . . 97
6.3.1 What the factorization costs . . . 98
6.4 The Laplace posterior . . . 100
6.5 Worked example: the chain in matrix form . . . 100
6.6 Second worked example: the water loop in matrix form . . . 103
6.7 Filters, fields, and state estimators . . . 106
6.8 In practice . . . 106

6.9 What is missing . . . 107
Notes and further reading . . . 107
Summary . . . 108
Exercises . . . 108
Solutions to selected exercises . . . 109

7 Beyond Gaussian: hybrid states and nonlinearity 111
7.1 Four ways out of the Gaussian world . . . 111
7.2 Hybrid states: the mixture over branches . . . 112
7.2.1 Worked example: a pipe that may be shut . . . 113
7.2.2 Many discrete variables . . . 115
7.3 Piecewise closures and regimes . . . 115
7.3.1 Worked example: a pressure-reducing valve . . . 116
7.4 Strong nonlinearity . . . 117
7.5 Bounded parameters . . . 119
7.6 What survives . . . 119
7.7 In practice . . . 120
7.8 What is missing . . . 121
Notes and further reading . . . 121
Summary . . . 122
Exercises . . . 122
Solutions to selected exercises . . . 123

III Inference 126

8 Exact inference: elimination, junction trees, Schur complements 127
8.1 What exact means . . . 127
8.2 Messages on a tree . . . 127
8.3 Loops: the junction tree . . . 129
8.4 Regions and ports: the Schur complement as a message . . . 130
8.5 Reuse . . . 132
8.6 Worked example: two districts, one transfer main . . . 133
8.7 Second worked example: two districts and a boundary main . . . 135
8.8 Discrete variables by region . . . 137
8.9 In practice . . . 138
8.10 What is missing . . . 139
Notes and further reading . . . 139
Summary . . . 140
Exercises . . . 140
Solutions to selected exercises . . . 141

9 Approximate inference 143
9.1 When exact is too expensive . . . 143
9.2 Loopy belief propagation . . . 144
9.2.1 On the book's models . . . 144
9.3 Variational methods and expectation propagation . . . 147

9.4 Simulation inside a region . . . 148
9.4.1 What a mixture message loses when collapsed . . . 149
9.4.2 Worked example: a treatment works as a simulated region . . . 149
9.5 Pruning by evidence . . . 151
9.6 The hierarchy, and how to choose . . . 153
9.7 In practice . . . 153
9.8 What is missing . . . 154
Notes and further reading . . . 154
Summary . . . 155
Exercises . . . 155
Solutions to selected exercises . . . 156

10 Enumeration, Monte Carlo, and factorized inference 158
10.1 The question, stated honestly . . . 158
10.2 The problem . . . 159
10.3 Enumeration: the product grid . . . 159
10.4 Monte Carlo . . . 161
10.5 Attribution: where the structural difference appears . . . 162
10.6 The factorized route: cuts of the dependency graph . . . 163
10.7 Second worked example: two works on one river . . . 164
10.8 Beyond three inputs . . . 167
10.9 The claim . . . 169
10.10In practice . . . 169
10.11What is missing . . . 170
Notes and further reading . . . 170
Summary . . . 171
Exercises . . . 171
Solutions to selected exercises . . . 172

11 Sequential evidence 173
11.1 Readings arrive one at a time . . . 173
11.2 One reading, one rank-one update . . . 173
11.3 The innovation . . . 174
11.4 Many readings . . . 175
11.5 Worked example: a stream of flow readings . . . 176
11.6 The Laplace posterior and re-linearization . . . 177
11.7 The value of a reading before it is taken . . . 179
11.8 Cost at scale . . . 180
11.9 Worked example: where to put the next pressure logger . . . 181
11.10In practice . . . 182
11.11What is missing . . . 183
Notes and further reading . . . 183
Summary . . . 184
Exercises . . . 185
Solutions to selected exercises . . . 185

IV The operator's questions 187

12 Conditioning, diagnosis, and virtual sensors 188
12.1 Everything is a readout . . . 188
12.2 Virtual sensors . . . 188
12.3 Is the model adequate? . . . 189
12.4 Where is it wrong? . . . 190
12.5 Worked example: bad data on a water network . . . 191
12.6 Which explanation? . . . 192
12.7 Can this parameter be recovered? . . . 194
12.8 Worked example: can tuberculation be seen? . . . 194
12.9 Which sensor next? . . . 197
12.10Should the Gaussian be believed? . . . 197
12.11In practice . . . 199
12.12What is missing . . . 200
Notes and further reading . . . 200
Summary . . . 200
Exercises . . . 201
Solutions to selected exercises . . . 202

13 Interventions and counterfactuals 203
13.1 Why the graph is not enough . . . 203
13.2 Rows are mechanisms; inputs are exogenous . . . 203
13.3 Seeing versus doing, on the chain . . . 204
13.4 Counterfactuals: three steps . . . 206
13.5 Decisions as branches . . . 209
13.6 Worked example: a drought week . . . 210
13.7 Worked example: holding the transfer flow . . . 211
13.8 What the semantics does not decide . . . 213
13.9 In practice . . . 213
13.10What is missing . . . 214
Notes and further reading . . . 214
Summary . . . 214
Exercises . . . 215
Solutions to selected exercises . . . 216

14 Attribution, sensitivity, and certificates 219
14.1 Three questions, one derivative . . . 219
14.2 Sensitivities from the solve . . . 219
14.3 Attribution . . . 220
14.4 A census with gradients . . . 222
14.4.1 What the census answers . . . 223
14.4.2 The three-main problem . . . 223
14.5 The kink . . . 225
14.6 Worked example: convexity lost and recovered . . . 226
14.7 Conditions . . . 228
14.8 In practice . . . 229

14.9 What is missing . . . 230
Notes and further reading . . . 231
Summary . . . 231
Exercises . . . 232
Solutions to selected exercises . . . 232

15 Combinatorial failure **234**
15.1 The question . . . 234
15.2 The prior over failure sets . . . 235
15.2.1 Independent failures . . . 235
15.2.2 Failure as an intervention . . . 236
15.3 The consequence per branch, and how to get it cheaply . . . 236
15.4 Worked example: seven pipes, one district . . . 237
15.5 The posterior over failure sets . . . 240
15.6 Enumeration by regions . . . 241
15.7 The worst k-set as a query . . . 242
15.8 Common cause and correlated failure . . . 242
15.8.1 A shared shock . . . 243
15.8.2 A mixture over the rate . . . 243
15.9 Worked example: a booster station . . . 243
15.10In practice . . . 245
15.11What is missing . . . 246
Notes and further reading . . . 246
Summary . . . 247
Exercises . . . 247
Solutions to selected exercises . . . 248

V Time, hierarchy, and scale **249**

16 Dynamics and temporal factors **250**
16.1 The accumulation term made live . . . 250
16.2 Integration as a sequence of steady solves . . . 251
16.3 The unrolled graph . . . 252
16.4 Process noise . . . 253
16.5 Filtering, smoothing, and prediction as eliminations . . . 254
16.6 Worked example: tracking a drifting source . . . 255
16.7 Detecting a change in time . . . 256
16.8 Two hypotheses the steady state cannot separate . . . 257
16.9 Events and cascades . . . 258
16.10Worked example: two service reservoirs . . . 259
16.11In practice . . . 261
16.12What is missing . . . 262
Notes and further reading . . . 262
Summary . . . 263
Exercises . . . 263
Solutions to selected exercises . . . 264

17 Hierarchy as physical decomposition **265**
17.1 One conservation law used twice . . . 265
17.2 Nested messages . . . 266
17.3 What the parent needs to know . . . 267
17.4 Worked example: a supply zone in three levels . . . 268
17.5 The cut, three times . . . 272
17.6 Worked example: a metered zone under a junction . . . 273
17.7 Hierarchy in time . . . 275
17.8 In practice . . . 277
17.9 What is missing . . . 277
Notes and further reading . . . 278
Summary . . . 278
Exercises . . . 278
Solutions to selected exercises . . . 279

18 Partitioning at scale **281**
18.1 The cost of a cut . . . 281
18.2 Elimination width of physical networks . . . 282
18.3 Choosing the cut . . . 283
18.4 Coupling the regions: messages, propagation, or iteration . . . 285
18.5 Enumeration by region . . . 289
18.6 A second domain, on purpose . . . 292
18.7 In practice . . . 293
18.8 What is missing . . . 293
Notes and further reading . . . 293
Summary . . . 294
Exercises . . . 294
Solutions to selected exercises . . . 295

VI Case studies **296**

19 Case studies **297**
19.1 What a case study is here . . . 297
19.2 Each chapter removes an advantage . . . 298
19.3 What is not here . . . 298
19.4 Reading a case study . . . 298

20 Three questions on one set of rows **300**
20.1 The question . . . 300
20.2 The system . . . 301
20.2.1 Why a closed form exists . . . 301
20.2.2 One regularization, and what it costs . . . 302
20.3 The model as a factor graph . . . 302
20.4 Method . . . 303
20.5 Results . . . 304
20.5.1 The state, four ways . . . 304

20.5.2 How big, and whether the interval is honest . . . 305
20.5.3 Which junction . . . 306
20.5.4 Where two gauges stop seeing . . . 306
20.5.5 Cost . . . 307
20.6 What surprised . . . 308
20.7 Reproduction . . . 308
Notes . . . 309

21 Half the pipes have meters **310**
21.1 The question . . . 310
21.2 The system . . . 311
21.3 The model as a factor graph . . . 311
21.4 Method . . . 311
21.5 Results . . . 313
21.5.1 Is the error bar honest? . . . 313
21.5.2 It is weighted least squares . . . 314
21.5.3 What the cut separates . . . 314
21.5.4 Where are the breaks? . . . 316
21.5.5 Two hypotheses that converge slowly . . . 317
21.5.6 Cost . . . 318
21.6 What surprised . . . 318
21.7 Reproduction . . . 319
Notes . . . 319

22 Screen, then confirm **320**
22.1 The question . . . 320
22.2 The two networks . . . 320
22.3 The model: one unknown, many snapshots . . . 321
22.4 Method . . . 322
22.5 Results . . . 323
22.5.1 Which junction leaks . . . 323
22.5.2 What a window buys . . . 323
22.5.3 What the leak competes with . . . 324
22.5.4 A wrong roughness becomes a leak . . . 325
22.5.5 Modena: screen, then confirm . . . 326
22.5.6 Cost . . . 327
22.6 What surprised . . . 327
22.7 Reproduction . . . 328
Notes . . . 328

23 Held to what a utility knows **330**
23.1 The question . . . 330
23.2 The system . . . 331
23.3 The two layers . . . 331
23.4 Method . . . 331
23.5 Results . . . 332
23.5.1 What calibration buys, and what it does not . . . 332

23.5.2 The year, read forward . . . 333
23.5.3 The official score . . . 334
23.5.4 Are the stated probabilities honest? . . . 334
23.5.5 The second evaluation, which is worse . . . 335
23.5.6 Cost . . . 336
23.6 What surprised . . . 336
23.7 Reproduction . . . 336
Notes . . . 337

A Notation **338**
A.1 Conventions . . . 338
A.2 Symbols . . . 339
A.3 Letters with more than one meaning . . . 342

B Gaussian identities **344**
B.1 The two forms of a Gaussian . . . 344
B.2 Block matrices . . . 345
B.3 Marginals and conditionals . . . 346
B.4 Products and linear-Gaussian models . . . 347
B.5 Rank-one changes . . . 348
B.6 Integrals, quadratic forms and divergences . . . 348
B.7 The Laplace approximation . . . 349
B.8 Where the book uses each identity . . . 350

C From the book to the engine **351**
C.1 What TERRA is . . . 351
C.2 A model in a dozen lines . . . 352
C.3 Where each construction lives . . . 353
C.4 The case studies . . . 353

D A catalogue of closures **356**
D.1 How to read an entry . . . 356
D.2 The four patterns in the engine . . . 358
D.3 Thermal . . . 358
D.4 Fluid . . . 360
D.5 Moist air . . . 362
D.6 Electrical . . . 362
D.7 Storage and electrochemistry . . . 364
D.8 Plant equipment . . . 365

Chapter 0

Introduction

In the small hours, when almost nobody is drawing water, a district metered area tells the truth about itself. The meter at its inlet reads what enters; the customers are asleep; what the district still takes is, nearly all of it, what its pipes are losing. When the night flow climbs, the control room has a short list of explanations and a short list of instruments with which to choose between them. A main has burst somewhere along a few dozen kilometres of pipe. Or the inlet meter has drifted. Or a boundary valve that the records say is shut is passing water in from the neighboring district, and the district is not losing more water at all, only being metered wrongly. Or a large customer is filling a tank. The instruments are the inlet meter, a handful of pressure loggers on hydrants, and a hydraulic model last calibrated against a fire-flow test some years ago.

The engineer on shift already has the physics of that district: a balance at every junction, a head-loss relation along every pipe, a fixed head at every reservoir and every boundary. Given the demands, a solver turns those equations into every head and every flow. What the solver cannot do is run the other way: take three pressure readings and one flow reading and say which of the four explanations they support, how strongly, along which stretch of main a crew should start listening, and whether one more logger would settle it. Probabilistic graphical models supply that machinery. They are usually taught apart from conservation laws, head-loss curves and network solvers, and the engineer who wants both has had to learn two subjects and build the bridge alone. This book is the bridge. Its claim is that the bridge is short, because the graph the probabilistic machinery needs is already there in the network equations the engineer wrote, and that once the graph is recognized, every question on the shift's list is a readout of one object. This chapter says why that claim is worth a book, what the book inherits from four older traditions, what it adds, and how to read it.

0.1 The questions a water engineer asks

A model of a physical system, as engineering teaches it, is a set of equations and a solver. The equations are the balances, the closures and the boundary conditions; the solver turns known inputs into a computed state. The pipeline is

$$\text{inputs} \longrightarrow \text{solver} \longrightarrow \text{outputs}, \tag{0.1}$$

and it answers one question well: given these inputs, what is the state? Most textbooks on physical modeling stop there, and for design that is often enough.

Operation asks different questions. A water network is instrumented, and its instruments are sparse: meters at the district boundaries and the pumping stations, pressure loggers on a few chosen hydrants, almost nothing in between. They are noisy, and occasionally wrong. The pipes' roughness coefficients were assigned when the pipes were laid and have been changing since. The demands are forecasts. The valves are open or shut on a drawing that may be out of date. The questions an operator or a network engineer actually asks are:

- What are the heads and flows across the district now, given a few pressure readings and a model I only partly trust?
- Which pipe's roughness is off, and by how much, and is it the model that is wrong or the network?
- What is the pressure at the critical node nobody logs, the high point where customers complain first?
- Is the night flow a burst, a drifting meter, a passing boundary valve or legitimate use, and if it is a burst, on which stretch of main?
- How uncertain is each of those answers?
- Where should the next pressure logger go, and how much would it narrow the search?
- What would happen if I shut that valve, slowed that pump, or lowered that pressure-reducing valve's setpoint?
- Which uncertain demand or roughness drives the risk of low pressure at peak, and is the number I am about to report a floor, a ceiling or a guess?

None of these is a question the pipeline (0.1) can answer. Each asks the model to run in a direction other than inputs to outputs, or to weigh a reading against a law, or to compare explanations, or to say how far its own answer can be trusted. The pipeline has no place to put a reading, no notion of an explanation, and no measure of trust.

The book's starting observation is that the equations do not have to be rewritten to answer these questions. They have to be read differently. A balance says that certain flows sum to zero at a node; a closure says that a flow and two potentials satisfy a relation; a boundary condition says that a quantity has a value. Each is a statement about which combinations of the variables are admissible. Attach to each a statement of how far from admissible the system may be, attach to each uncertain input a statement of what is known about it, attach to each sensor a statement of what it read and how well, and the collection is a probability distribution over the entire state of the system, built from nothing but the physics and the instruments. Every question in the list is then a question about that distribution, and Parts II to IV of this book are about answering them.

0.2 Where the graph comes from

A textbook on probabilistic graphical models begins with random variables and their joint distributions, defines directed and undirected graphs and the factorizations they encode, develops conditional independence, and then teaches inference on the graph. The examples are alarms and burglaries, diseases and symptoms, letters and their noisy transmissions. It is a coherent subject and the machinery is powerful, but an engineer arriving from the physical side asks a question the textbook does not answer: where did the graph come from? For an alarm and a burglary the graph is a modeling decision, drawn from a story about what causes what. For a pumping main, a district metered area or a city's trunk network it is not a decision at all.

For a physical model the graph is the sparsity pattern of the equations. Each balance reads

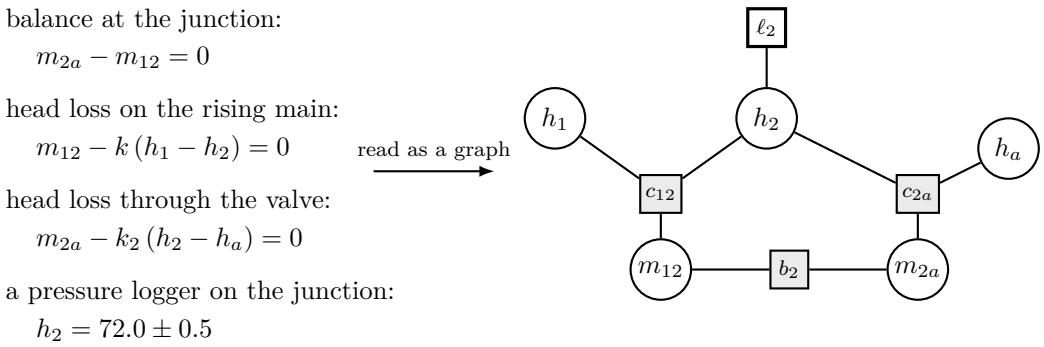


Figure 0.1: The graph comes from the equations. Left: three physics equations of the pumping chain of Example 1.1 and one pressure-logger reading, as an engineer writes them. Right: the same four statements as a factor graph. Each equation is a square joined to exactly the variables it reads; the sensor is a square of degree one. Nothing was decided in drawing it.

Table 0.1: What each object of a physical model becomes when probability is attached. The graph is unchanged; only the meaning of the squares is.

In the equations	In the factor graph	With probability attached
balance at a node	a factor on the node's flows	conservation, to within a small width
constitutive relation	a factor on a flow and two potentials	closure, to within its model error
boundary condition	a factor of degree one	a prior on the boundary value
uncertain parameter	a variable in every closure reading it	a prior on the parameter
sensor	(absent)	a likelihood of degree one on its variable
component in or out of service	(absent)	a discrete variable in its closure's scope

the flows at one node. Each closure reads one flow and the potentials at its ends. Each boundary condition reads one variable. Draw a node for every variable and a node for every equation, join each equation to the variables it reads, and the result is a factor graph, the bipartite graph on which the probabilistic machinery operates (Figure 0.1). No modeling decision was made in drawing it, because every edge was already an entry of the Jacobian that the solver assembles. The graph is the model's structure, seen.

Attaching probability then changes what the squares mean and not where they are. A balance equation becomes a factor that says the balance holds to within a small tolerance. A closure becomes a factor that says the constitutive law holds to within its model error. An uncertain parameter or boundary condition gets a prior, a square of degree one on its variable. A sensor becomes a likelihood, another square of degree one, on the variable it reads. A component that may be in service or out becomes a discrete variable in the scope of the closure it governs. Table 0.1 lists the correspondence, which Chapter 3 develops. The bridge from the engineer's equations to the graphical model is this table, and it is short.

0.3 One subject, not four

Little in this construction is new in isolation, and it would be wrong to present it as if it were. Four traditions have each built part of it, and the book is written in the conviction that they are one subject that has been taught as four.

Port-based physical modeling. Bond graphs, port-Hamiltonian systems and network thermodynamics supply the domain-independent view of a physical system as conserved quantities at nodes, flows driven by potentials along edges, and conjugate flow-and-potential pairs at ports, the same structure whether the domain is hydraulic, thermal, electrical or mechanical [39, 66, 69, 95]. This tradition stops at the equations: it says how to write $\mathbf{F}(\mathbf{z}) = \mathbf{0}$ for any physical system, and then solves.

Equation-based modeling. Acausal modeling languages and the structural analysis of differential-algebraic systems establish that a physical law is a relation among variables, not an assignment of an output from inputs, and they give the bipartite equation-variable structure and its matching theory [5, 20, 63, 68]. This tradition, too, stops at the solve, and it does not attach probability.

Probabilistic graphical models and sparse elimination. Factor graphs, Markov properties, separators, elimination, fill and treewidth come from the graphical-model and sparse-matrix literatures [16, 45, 47, 49, 78]. This tradition starts from a factorized distribution and asks how to compute with it; it does not say where the factorization comes from, and its graphs are drawn, not derived.

Bayesian inverse problems and state estimation. Priors on uncertain physical inputs, sensors as likelihoods, weighted least-squares estimation under conservation constraints, and the diagnosis of bad data are established in inverse problems, in power-system state estimation, and in process data reconciliation [1, 37, 57, 64, 88], and in water distribution they are the calibration of network models and the localization of leaks from pressure residuals [71, 79]. This tradition does attach probability to physics, but usually around the solver: the physics is a black-box map from inputs to observables, and the posterior is over the inputs alone.

Put the four end to end and they make one chain,

$$\boxed{\text{physical topology} \;\rightarrow\; \text{local equations} \;\rightarrow\; \text{factorization} \;\rightarrow\; \text{probability} \;\rightarrow\; \text{inference} \;\rightarrow\; \text{engineering decisions}} \tag{0.2}$$

of which each tradition owns one or two links and none carries the whole. Figure 0.2 draws it, with the chapters of this book placed along it. The book's purpose is to make the chain feel like one subject: to start where the engineer starts, at the topology and the equations, and to arrive where the operator needs to be, at a decision with its uncertainty stated, without changing the object along the way.

0.4 Probability in the structure, not around the solver

There is an obvious way to combine a physical model with probability, and most of the fourth tradition uses it: leave the solver as it is, treat it as a function from uncertain inputs to observable outputs, draw inputs from their priors, run the solver, and reason about the outputs. The physics is a black box inside a probabilistic wrapper. Figure 0.3 draws it on the left. It is correct, it is general, and it is how a great deal of uncertainty quantification is done.

The book takes the other route, drawn on the right: the equations themselves are the factors, and probability lives inside the structure, on every variable the physics has. The thesis of the book, in one sentence, is

$$\boxed{\text{Do not put probability around the solver. Put it into the structure of the physical model.}} \tag{0.3}$$

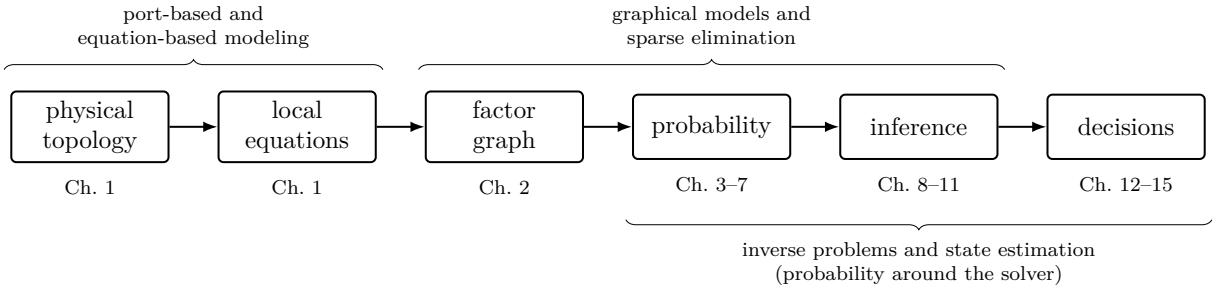
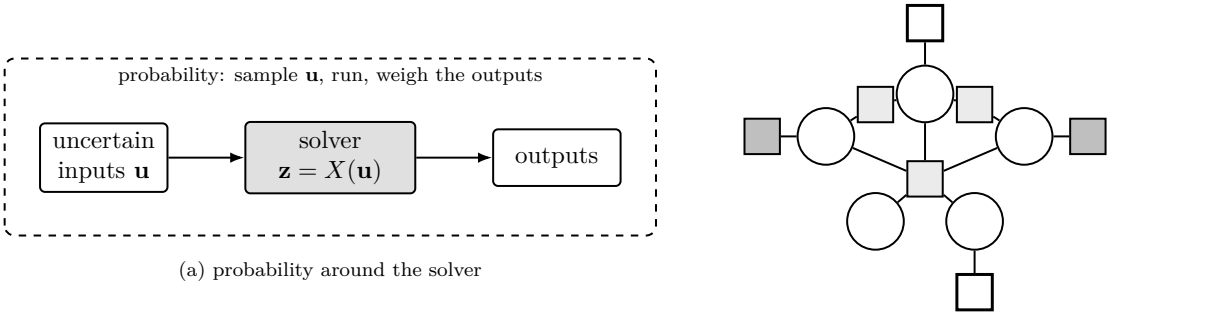


Figure 0.2: The chain of ideas the book follows, with the chapters that cover each link and the traditions that own parts of it. No one tradition carries the object from the topology to the decision.



(a) probability around the solver

(b) probability in the structure:
priors (dark) and sensors (open) added to the physics graph

Figure 0.3: Two ways to combine physics and probability. (a) The solver is a black box; probability lives outside it, over its inputs and outputs. (b) The physics equations are the factors; priors and sensors are squares added to the same graph. The book takes the second route and says why.

Three consequences follow, and they are what the later chapters are about.

There is one object. The deterministic solve is the special case of the probabilistic model in which the priors and sensors are absent and the physics is exact; the two share every row and every Jacobian (Chapter 6 proves this). Adding a sensor adds a row. Adding a question adds nothing: the estimate, the diagnosis, the counterfactual and the risk are readouts of one posterior (Chapter 12).

Structure is preserved, and structure is what computation costs. The wrapper's map from inputs to outputs is dense: every output depends on every input. The physics graph is sparse: every equation reads a few variables. The posterior built in the structure has the sparsity of the physics, and every computation in Part III is affordable because of it. The wrapper's route discards the sparsity the moment it closes the box.

Violation can be seen. A wrapper enforces the physics exactly, because the solver does. When the sensors report something no state of the physics can produce, the wrapper has no answer. The structural model gives each law a width, lets it be violated at a cost, and reports the violation as a diagnosis: the burst, the drifting meter, the valve believed shut (Chapters 4 and 12).

0.5 The boundary

Every system worth modeling is a piece cut out of a larger one. A pressure zone trades water with the zones above and below it, a district metered area trades water with the trunk main that feeds it, a pumping station trades water with its aquifer and head with the network. The piece has an inside that the modeler describes and a boundary across which water passes to an outside the modeler does not describe. Chapter 1 calls such a piece a *boundary-exchanging system*, and the book returns to the boundary at every level for a reason that is easiest to state in the language of the third tradition.

In a graphical-model text, a *separator* is a set of variables whose knowledge makes two parts of the graph independent, and finding one is a graph-theoretic exercise whose reward is a cheaper computation. Water engineers drew their separators before any theory asked them to. A district metered area is defined by its boundary: every main that crosses it is either metered or valved shut, so that what enters is known and a night-flow balance can be struck. The flows and heads at those metered inlets, a flow and its conjugate potential at each port, are exactly the variables that separate the district's interior from the rest of the network (Chapter 5 proves it, and proves that nothing smaller does). So the instruction "find a separator" becomes "cut the network where the meters are", which is where the utility has already cut it, and the mathematics then explains why that cut is also the cheap place to divide the computation. It explains, too, why a boundary valve believed shut and in fact passing water does the damage it does: it is a hole in the separator, and Chapter 8 measures what it costs.

This gives one decomposition three readings at once. Physically, a subsystem is summarized to its neighbors by what crosses its ports, exactly, because conservation says the cut flows equal the net source inside. Probabilistically, the ports are a separator, so a subsystem's message to the rest of the system is a function of the ports alone. Computationally, eliminating the interior onto the ports is the Schur complement of sparse elimination, the Kron reduction of network theory, and the region message of a junction tree, which are one operation (Chapter 8). A district can be an equivalent inlet with error bars, computed once and handed to its neighbors; a large network can be a tree of districts (Chapter 17 works one, a district metered area under a distribution junction); and a partition can be judged by its ports (Chapter 18). The three-way connection is the part of the book for which the author has found the least direct precedent, and it is the reason the boundary is in the title.

0.6 From "what is the state" to "what can I know"

The pipeline (0.1) has a direction: inputs in, outputs out. A physical law does not. The relation

$$m - k_2\,(h_2 - h_a) \;=\; 0 \tag{0.4}$$

says that four quantities satisfy one constraint: the flow through the pumping chain's throttling valve, the heads either side of it, and the valve's conductance. It does not say that m is computed from the heads; given m and h_2 it determines the trunk main's head h_a, and given all three it determines k_2, which is how a valve that is throttled harder than the records say is found. Which variable is unknown is a property of the question, not of the physics (Chapter 2 makes this precise, and shows that for a looped system no fixed direction exists at all). The book keeps the relations as relations, which is what the second tradition insists on and what the first tradition's bond graphs draw, and it is why the factor graph rather than the directed graph is the book's language.

Once uncertainty and observations enter, the absence of direction becomes the point. A reading of one head is evidence about everything the physics ties to it. On the pumping chain that runs through the book, a single pressure logger on the junction cuts the uncertainty in the pumping station's unmeasured delivery by a factor of three and a half, and predicts the unmeasured discharge head to within about a metre and a half, before any second sensor is read (Chapter 3). That prediction is a *virtual sensor*, and its error bar is part of the answer. The question the model answers has changed from "what is the state, given the inputs" to "what can I know, given what I have", and the second question is the operator's.

0.7 What is inherited and what is the book's

A book that combines old ingredients owes its reader a clear account of which is which. Each chapter closes with a section of notes that does this for its own contents. The account for the whole is short.

Inherited: the port-based view of physics; the acausal reading of its equations; the factor graph, its Markov properties, elimination and treewidth; the Gaussian machinery of least squares, sparse Cholesky and the Laplace approximation; Bayesian inverse problems; state estimation and data reconciliation with their bad-data statistics; the causal calculus of interventions; variance-based attribution; reliability evaluation and contingency analysis. None of these is the book's, and the notes say whose each is.

The book's: the synthesis. Taking an equation-based, conservation-structured physical model and turning its residual structure, unchanged, into the factorization on which inference, diagnosis, uncertainty propagation, attribution and decision queries are performed; the identification of physical ports with probabilistic separators and computational cuts; the intervention semantics for acausal equations; and the insistence, carried through every chapter, that the deterministic model is a corner of the probabilistic one.

The author has looked for this treatment and not found it. Pieces of it exist in several communities that have discovered the connection independently. Multibody dynamics has been written as a factor graph whose sparse structure serves both simulation and estimation, with forward and inverse dynamics as different queries on one graph [10], which is the "relations, not assignments" reading in a single domain. Bond graphs have been converted into Bayesian networks for fault diagnosis [54], which attaches probability to a port-based physical model but does so by assigning the causal directions this book declines to assign. Power systems have been estimated by message passing on a factor graph built from the power-flow equations [14], which is Chapters 6 to 12 in one domain and one query. Table 0.2 sets these beside the four traditions and beside this book. What the table shows is not that any column is empty but that no prior row fills them all, and that several communities have arrived at parts of one construction without a text that says they are parts of one construction. That is the gap the book is written to fill, and "not found" is the strongest claim the author is willing to make for it.

0.8 How to read this book

Who it is for. The reader is an engineer who knows conservation laws and linear algebra and has never met a probabilistic graphical model. Nothing about probability beyond the meaning of a density is assumed; Chapter 3 supplies what is needed in a page and Appendix B the Gaussian identities. A reader who knows graphical models and wants the physics will find Chapters 1, 2

Table 0.2: The closest prior work, by what it covers. A check marks a substantial treatment; a word marks a partial or specialized one.

	physics graph	equations as factors	probability	general domains	ports as separators	broad inference
bond graphs, port-Hamiltonian [69, 95]	✓	—	—	✓	✓	—
multibody factor graphs [10]	✓	✓	limited	multibody	—	solve, estimate
bond-graph Bayesian networks [54]	✓	indirect	✓	general idea	—	fault diagnosis
factor-graph state estimation [14]	power grid	specialized	✓	power only	regional	state estimation
graphical-model texts [45, 49]	—	—	✓	abstract	graph-theoretic	✓
inverse-problem texts [37, 88]	equations	not generally	✓	✓	—	inverse only
this book	✓	✓	✓	✓	✓	✓

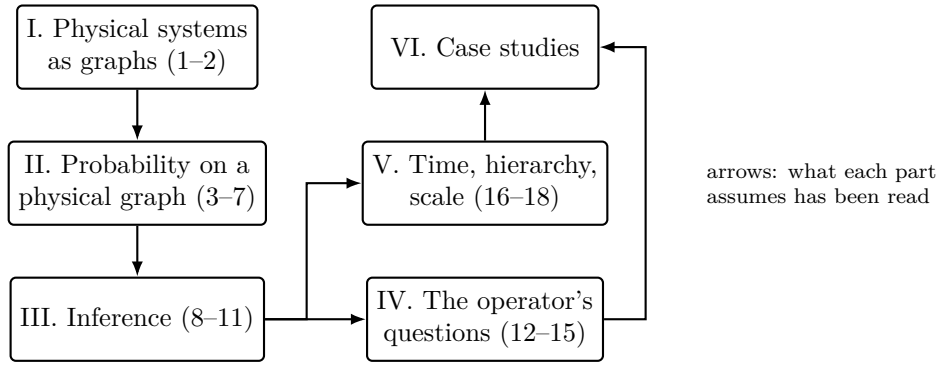


Figure 0.4: The parts of the book and their dependencies. Parts IV and V each need Part III and not each other; Part VI needs both.

and 5 the ones to read first, since they say what the graph is and why the ports separate. The examples are water examples; the constructions are not, and a reader who works on heat, charge, power or any other conserved quantity will find nothing in Parts I to V that needs changing but the names.

The parts. Figure 0.4 shows the six parts and what each depends on. Part I is deterministic and builds the objects. Part II attaches probability and states, rather than skips, the places where that attachment needs care. Part III organizes the computation. Part IV asks the operator's questions. Part V handles time, hierarchy and scale. Part VI is a set of complete studies, each reproducible from the material in the earlier parts.

The running examples. Two small systems appear in almost every chapter: the pumping chain of Example 1.1, a borehole station lifting into a trunk main through two sections of rising main and a throttling valve, and the water loop of Example 1.2, three junctions in a triangle fed from a trunk main, with the head loss linearized so that every number can be checked by hand; Chapter 6 restores the Hazen–Williams exponent and measures what the linearization cost. Both are introduced in Chapter 1, both are small enough to solve by hand, and both acquire priors, loggers, a failed pipe, a second district, an intervention and a failure set as the chapters go on. The reader who follows them through will have seen every construction of the book on a system whose every number can be checked. A third, the district cooling loop of Example 1.3, carries mass and energy on one graph and is kept to show that nothing in the method assumes a

single conserved quantity. Larger systems appear where the small ones cannot show the point: a two-district network joined by one transfer main for decomposition and failure, a booster station with four pumps for common cause, a grid of streets for cost at scale, and the case studies of Part VI.

The numbers. Every number in every worked example is produced by a short script that accompanies the book, one per chapter, and the text says which. The scripts are a few hundred lines each and use nothing beyond a numerical library; they are meant to be read and rerun, and several exercises ask for exactly that. The case studies of Part VI read their numbers from committed result files of the studies they describe. No number in the book was typed by hand.

Conventions. Each chapter ends with a section on what is in practice, notes on what is inherited and from where, a summary, five exercises, and worked solutions to two of them. Load-bearing conclusions are set apart as numbered principles; there are about three per chapter, and a reader who remembers only those will have the book's argument. Propositions are proved where they are stated. Notation is collected in Appendix A, and the reference implementation that the author used to run every script is mapped to the book's constructions in Appendix C; nothing in the body depends on it.

0.9 What this book does not do

It does not teach physical modeling; it assumes the reader can write a balance and a closure, and it uses the simplest ones. It does not teach probability or graphical models in generality; it teaches the parts a physical model needs, and points to the texts that teach the rest. It does not develop machine learning: no factor in this book is learned from data, though Chapter 12 estimates parameters and Chapter 14 fits a surrogate, and the notes say where the learned versions of its constructions live. It does not treat water quality: no chlorine decay, no age, no contaminant tracking, though each would ride on the same graph. It does not treat transients: there is no water hammer here, and time enters in Chapter 16 only as service reservoirs filling and draining. It does not treat continuous space: every system here is a graph of lumped elements, and a field discretized on a mesh appears only as a large graph whose treewidth is a warning. And it does not claim that the constructions it teaches are new; it claims that they are one subject, and it tries to teach them as one.

Notes and further reading

The five gaps that §0.1 to §0.6 describe were articulated in a review of the manuscript, and the positioning of §0.7 follows a literature search made in response to it. The closest works found, Blanco-Claraco, Leanza and Reina's factor-graph multibody dynamics [10], Lo, Wong and Rad's bond-graph Bayesian networks [54], and Chavali and Nehorai's factor-graph state estimation [14], are each discussed in the chapter whose subject they touch. The four traditions of §0.3 are cited there and again in the notes of the chapters that inherit from them; the reader who wants one text per tradition is directed to Karnopp and Rosenberg [39], Benveniste, Caillaud and Malandain [5], Koller and Friedman [45], and Kaipio and Somersalo [37].

Part I

Physical systems as graphs

Chapter 1

Conservation, closure, and the boundary

Every physical system a modeler cares about is a piece cut out of a larger one. A building trades heat with the weather. A transmission network trades power with the networks around it. A district metered area trades water with the zone around it. The piece has an inside, which the modeler describes in detail, and a boundary, across which conserved quantities pass to an outside the modeler does not describe at all.

This chapter builds the deterministic model of such a piece. It has three ingredients and no probability. The three ingredients are conservation at places, closure along paths, and exchange at the boundary. The reason to build the deterministic model first, and to build it carefully, is that the probabilistic model of Part II inherits its structure entirely: the places, paths and boundary of this chapter become the nodes, factors and separators of the graphical model. Nothing is added to the structure later. Only uncertainty is.

1.1 Three questions about any system

Take any conserved quantity: energy, mass, electric charge, a chemical species. Three questions describe how it lives in a system.

1. Where can it accumulate?
2. How does it move from one place to another?
3. What crosses the boundary?

The first question identifies the *places*. Each place has an amount of the quantity, which may change in time. The amount is an *extensive* state: the energy in a wall, the mass of water in a tank, the charge on a capacitor. Doubling the place doubles the state.

The second question identifies the *paths*. Along a path the quantity flows at some rate: watts, kilograms per second, amperes. What drives the flow is a difference in an *intensive* quantity between the two ends of the path: temperature, pressure, voltage. Doubling the place does not double the intensive quantity; it is a property of the state at a point, not of the amount.

The third question identifies the *boundary*. Some paths lead to places the model does not describe. The quantity that crosses those paths enters or leaves the system. The places on the far side are held at imposed conditions, either an imposed intensive value (the room temperature, the grid voltage) or an imposed flow (a heat load, a scheduled injection).

Table 1.1: Four conservation domains and their three quantities.

Domain	Extensive state	Flow	Intensive driver
Mass (water network)	mass M [kg]	mass rate $\dot{m}$ [kg/s]	head h [m]
Energy (thermal)	internal energy E [J]	heat rate Q [W]	temperature T [K]
Charge	charge q [C]	current I [A]	voltage V [V]
Power (DC network)	—	line flow F [W]	angle θ [rad]

Table 1.1 lists the four conservation domains that recur in this book with their extensive state, their flow, and the intensive quantity that drives it. The rows are physically different; the columns are the same three questions. That parallel is the foundation of everything that follows. A method that works on the columns works on every row.

Two rows are worth a comment. The first is the one this book works in. For a pressurized water network the driver is written as *head*, metres of water column, rather than as pressure in pascals: the two differ by the constant ρg, and head is what a gauge on a main is calibrated in and what a pump curve is quoted in. Every statement in this book about a potential applies to it unchanged.

The last row is there to make the generality visible. In the DC approximation of a power network the conserved quantity is real power, the flow on a line is the power it carries, and the driver is the voltage angle difference between its ends. There is no extensive state because the approximation is steady: buses do not store power. The row still fits the three questions; the first answer is simply "nowhere". Nothing in this book is specialized to water or to heat, and the occasional appearance of another domain in a table is the reminder.

1.2 Nodes conserve

The three questions define a graph. The places are nodes; the paths are edges. Write $\mathcal{N}$ for the set of nodes, $\mathcal{E}$ for the set of edges, and $\mathcal{D}$ for the set of conservation domains the model tracks. A model with heat and water flowing through the same pipework has $\mathcal{D} = \{\text{energy}, \text{mass}\}$, which is a district heating network and is where Part VI ends; a cold water distribution network has $\mathcal{D} = \{\text{mass}\}$ alone.

1.2.1 The balance law

At each node n and for each domain d in which n can accumulate, the rate of change of the extensive state equals what flows in, minus what flows out, plus what is generated, minus what is lost:

$$\boxed{\frac{\mathrm{d}X_{n,d}}{\mathrm{d}t} \;=\; \sum_{e \in \mathcal{E}(n)} A_{n,e}\, P_{d,e} \;+\; G_{n,d} \;-\; L_{n,d}} \tag{1.1}$$

Here $X_{n,d}$ is the extensive state of domain d at node n, $\mathcal{E}(n)$ is the set of edges incident on n, $P_{d,e}$ is the flow of d along edge e, $G_{n,d}$ is a source at the node and $L_{n,d}$ a loss. The coefficient $A_{n,e}$ is the sign that says whether edge e delivers to or draws from node n; it is defined in §1.2.3.

Equation (1.1) is the whole of physics that a node contributes. It is linear in the flows, always and exactly. It has no parameters. Whether the domain is energy or mass or charge, whether the node is a wall or a tank or a bus, the balance row looks the same. This is not a modeling

convenience. It is the reason the graph is the right object: the graph is the conservation skeleton, and every domain that the model tracks uses that same skeleton.

1.2.2 Node roles

Not every node accumulates. Three roles cover the cases.

Balanced. The node has an extensive state that can change. Equation (1.1) holds with its left side live. A wall, a water tank, a battery.

Zero-storage. The node is a junction or a bus. It has no extensive state, so equation (1.1) holds with $\mathrm{d}X_{n,d}/\mathrm{d}t \equiv 0$: what comes in goes out. A pipe tee, an electrical bus, a mixing point.

Reservoir. The node is on the far side of the boundary. Its intensive state is imposed, not solved. It contributes no balance row, because the model does not claim to know its balance; it appears only as the far end of the edges that cross the boundary. The atmosphere, the utility grid, a feed line.

The first two roles are the inside of the system. The third is the outside. Which nodes are reservoirs is the modeler's first and most consequential decision, and §1.4 returns to it.

A node also declares which domains it is balanced in. A pipe carrying water may need a mass balance at each tee but no energy balance if the pipe is adiabatic; a bus needs a power balance and nothing else. The balance rows exist only where the modeler asks for them. The roles are declared per domain as well: a return manifold whose pressure is held by a pump is a reservoir for mass and yet a balanced node for energy, because the temperature of the water leaving it is the model's to determine. Example 1.3 has such a node.

1.2.3 Incidence

Give every edge a direction, from a tail node to a head node. The direction is a bookkeeping convention, not a physical claim: a flow that runs against the edge's direction is simply negative. The *incidence matrix* $\mathbf{A}$ has one row per non-reservoir node and one column per edge, with entries

$$A_{n,e} = \begin{cases} +1 & \text{edge } e \text{ enters } n \text{ (n is its head)}, \\ -1 & \text{edge } e \text{ leaves } n \text{ (n is its tail)}, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

Every column of the full incidence matrix (with reservoir rows included) has one $+1$ and one -1, so the columns sum to zero. Dropping the reservoir rows breaks that property for the columns of boundary edges, and that is exactly right: the flow on a boundary edge is not balanced by any row of the model. It is what the system exchanges.

With $\mathbf{A}$ in hand, the sum in equation (1.1) is one row of a matrix product. Stack the flows of domain d into a vector $\mathbf{P}_d$ and the balances into a vector; the balance law for all nodes at once is $\mathrm{d}\mathbf{X}_d/\mathrm{d}t = \mathbf{A}\,\mathbf{P}_d + \mathbf{G}_d - \mathbf{L}_d$.

> **Principle 1.1.** *The incidence matrix is shared by every domain. Energy and mass in the same pipework do not have two graphs. They have one graph, and each domain contributes balance rows on the nodes where it is tracked, using the same* $\mathbf{A}$.

An edge that carries one domain but not another, such as a rotating shaft that transmits energy but no mass, or a conduction path through a solid, simply has no flow variable for the absent domain. Its column contributes zero to that domain's balance rows.

Table 1.2: Four closure patterns. Each is written as a residual that vanishes when the relation holds.

Pattern	Residual	Typical use
Carrier flow	$P_e - q_e\,\varepsilon_e$	a quantity carried by another: $P = \dot{m}\,h$ (enthalpy carried by mass), $P = I\,V$ (power carried by current)
Gradient flow	$P_e - g(\phi_i, \phi_j; \theta)$	driven by a difference in intensive state: $Q = U\,A\,(T_h - T_c)$, $F = B\,(\theta_i - \theta_j)$
Property	$y - f(x_1, \ldots, x_k)$	a property table or a device specification: $h = h(p, T)$, $Q = \mathrm{COP} \cdot W$
Piecewise	$y - f_b(x)$	a device that changes regime: a valve that saturates, a limiter, a breaker

1.3 Edges close

The balance law says that the flows at a node sum to the accumulation. It does not say what any flow is. That is the job of the edge.

1.3.1 Closure relations

Along each edge, for each domain it carries, there is a flow variable and a relation that ties that flow to the states at the edge's ends and to some parameters. The relation is called a *closure* because it closes the system of equations: the balance rows alone have more unknowns than rows, and the closures supply the missing rows.

Definition 1.1 (Closure)**.** A closure is a local relation $r(\mathbf{z}_{\mathrm{loc}}; \theta) = 0$ among a small set of variables $\mathbf{z}_{\mathrm{loc}}$, with parameters θ, written in *residual form*: a function that is zero exactly when the relation holds. The closure declares which variables it reads.

Four patterns cover the great majority of physical closures (Table 1.2).

The gradient pattern is the one the three questions anticipated: a flow driven by an intensive difference. Conduction, convection, DC current, DC power flow and laminar pipe flow all take this form, with different functions g and different parameters θ: a conductance, a heat transfer coefficient times an area, a susceptance. The carrier pattern appears whenever one domain rides on another, as enthalpy rides on mass flow. The property pattern is a catch-all for relations among variables at a single place. The piecewise pattern is the first hint of something the deterministic model handles awkwardly and the probabilistic model handles naturally; it returns in Chapter 7. Appendix D catalogues the closures of the domains this book touches. For each it gives the residual and its pattern, the unknowns and parameters, the Jacobian, where the relation kinks, its shape, and the parameter that usually carries the prior.

1.3.2 A closure is a relation, not an assignment

Write the conduction closure as most textbooks do:

$$Q = U\,A\,(T_h - T_c). \tag{1.3}$$

The notation suggests a computation: given the two temperatures, compute the heat rate. But nothing in the physics says which quantities are given. If the heat rate and the hot temperature are known, the same relation determines the cold temperature:

$$T_c = T_h - Q/(UA). \tag{1.4}$$

If all three are measured, the relation determines the conductance UA. The physics asserts that four quantities satisfy one constraint. It does not assert a direction.

The residual form makes this explicit:

$$r(Q, T_h, T_c; UA) \;=\; Q - UA\,(T_h - T_c) \;=\; 0. \tag{1.5}$$

Equation (1.5) has no left-hand side to compute. It is a surface in the space of its variables, and every point on the surface is a physically admissible state of the edge.

> **Principle 1.2.** *A closure relates its variables. It does not compute one of them from the others. Which variable is unknown is a property of the question being asked, not of the physics.*

This principle looks like a remark about notation. It is not. It decides, in Chapter 2, which kind of graphical model a physical system should be written in, and it is the reason the arrows of a Bayesian network are not the natural representation of a closure. Hold on to it. The principle itself is old: it is the acausal view of physical laws that bond-graph modeling and equation-based modeling languages are built on [63, 69], and §1.8 says more.

1.3.3 Parameters

The parameters θ of a closure, a conductance, a susceptance, an efficiency, are ordinarily numbers the modeler supplies. Nothing forces that. A parameter the modeler does not know can be left as an unknown, added to the vector of variables, and determined by the equations if enough other things are known. A conductance inferred from a measured heat rate and two measured temperatures is the simplest case of this. The deterministic model needs one extra equation for each parameter it frees; the probabilistic model of Part II needs only a prior.

1.4 The boundary

The boundary is the set of edges with a reservoir at one end. Everything on the reservoir's side is outside the model. This section says what the model knows about the outside, why that knowledge is exact in one specific sense, and why the boundary is the most important part of the system.

1.4.1 Reservoirs and imposed conditions

A reservoir contributes no balance row. It contributes an imposed condition on the boundary edge, of one of two kinds.

Imposed potential. The reservoir's intensive state is fixed: the room temperature, the grid voltage angle at the point of connection. The boundary edge's closure then determines the flow across it. This is the Dirichlet condition of boundary-value problems.

END OF THE PREVIEW

Probabilistic Modeling of Water Distribution Networks

These are the opening pages. The complete book runs 386 pages.

The full PDF opens with a token. Write to **rmarquez123@gmail.com**, name the book, and I will send you one.

One token opens every book on the site.

Ricardo Marquez, Ph.D.

ricardo-marquez.com