

VOLUME II

Molecular Wave Mechanics

*Numerical Recipes:
Theorems, Derivations, and Worked Exercises*

A chapter-by-chapter rigorous companion
to Volume I

Abdelkader BOUSABAA

University of Paris-Saclay

Volume II — First Edition — 2026

SEVEN LIONS EDITIONS

Contents

Preface	xv
1 Notation, Prerequisites, and How to Read this Companion	1
I Mathematical Foundations	9
2 Probability and Statistics for the Molecular Sciences	11
2.1 Notation and prerequisites	11
2.2 Theorems and full proofs	12
2.2.1 The Poisson limit of the binomial	12
2.2.2 The negative binomial as a Poisson–gamma mixture	13
2.2.3 Limit theorems	15
2.2.4 Concentration inequalities	16
2.2.5 Maximum-likelihood estimation: consistency, normality, ef- ficiency	18
2.2.6 The Born rule (K1): two derivations	20
2.3 Worked exercises with full solutions	22
2.4 Outlook	27
3 Information Theory and Entropy	29
3.1 Notation and prerequisites	29
3.2 Theorems and full proofs	30
3.2.1 Shannon entropy: definition and bounds	30
3.2.2 Differential entropy and the Gaussian maximum	31
3.2.3 KL divergence and Gibbs’ inequality	32
3.2.4 The cross-entropy identity: the four losses are one	34
3.2.5 Mutual information and the data processing inequality	35
3.2.6 Fidelity, overlap, and the K2 identity	37
3.3 Worked exercises with full solutions	39
3.4 Outlook	44
4 Optimization and Variational Methods	45
4.1 Notation and prerequisites	45
4.2 Theorems and full proofs	46
4.2.1 Convexity and the existence of minimisers	46
4.2.2 Gradient descent on smooth convex objectives	47
4.2.3 Stochastic gradient and the Robbins–Monro condition	49

4.2.4	Constrained optimisation: Lagrangian duality and KKT . .	50
4.2.5	The Rayleigh quotient (K3): Courant–Fischer characterisation	50
4.2.6	From Rayleigh quotient to eigenvalue equation: Lagrangian derivation	53
4.3	Worked exercises with full solutions	54
4.4	Outlook	59
5	Bayesian Inference	61
5.1	Notation and prerequisites	61
5.2	Theorems and full proofs	62
5.2.1	Bayes’ rule and the marginal likelihood	62
5.2.2	Conjugate priors: Beta–Binomial and Gamma–Poisson . . .	63
5.2.3	Markov chain Monte Carlo: detailed balance and ergodicity	64
5.2.4	Variational Bayes and the ELBO decomposition	66
5.2.5	Bernstein–von Mises: posterior asymptotic normality	67
5.2.6	The Bayes-as-amplitude identity (K9)	68
5.3	Worked exercises with full solutions	69
5.4	Outlook: Part I closes, Part II opens	75
II	Machine Learning and Neural Networks	77
6	Statistical Learning Theory: Generalisation and the Variational Certificate	79
6.1	Notation and prerequisites	79
6.2	Theorems and full proofs	80
6.2.1	Uniform convergence for finite hypothesis classes	80
6.2.2	VC dimension and the Sauer–Shelah bound	81
6.2.3	Rademacher complexity	82
6.2.4	The PAC-Bayes bound	83
6.2.5	Double descent for ridgeless random-features regression . .	84
6.2.6	The variational gap as a generalisation certificate	86
6.3	Worked exercises with full solutions	88
6.4	Outlook	93
7	Neural Networks and Backpropagation: The Variational Bridge to Wave Mechanics	95
7.1	Notation and prerequisites	95
7.2	Theorems and full proofs	96
7.2.1	Universal approximation	96
7.2.2	Backpropagation: correctness and complexity	98
7.2.3	Variational Monte Carlo: unbiased gradient estimation . . .	99
7.2.4	The K4 neural-quantum-state identification	100
7.2.5	Depth as renormalisation-group coarse-graining	103
7.3	Worked exercises with full solutions	104
7.4	Outlook	109

8	CNNs, RNNs, and Transformers: Architectures as Wave-Function Ansätze	111
8.1	Notation and prerequisites	111
8.2	Theorems and full proofs	112
8.2.1	Convolutional networks: translation equivariance	112
8.2.2	Recurrent networks: state-space universal approximation	114
8.2.3	Self-attention: the Q-K-V mechanism	114
8.2.4	The K5 attention-pair-correlation identification	115
8.2.5	Direct coupling analysis and the precision-matrix correspondence	118
8.2.6	Architectural inductive biases as wave-function structures	119
8.3	Worked exercises with full solutions	120
8.4	Outlook	125
9	Generative Models: VAEs, GANs, Diffusion	127
9.1	Notation and prerequisites	127
9.2	Theorems and full proofs	128
9.2.1	The variational autoencoder: ELBO and reparameterisation	128
9.2.2	Generative adversarial networks: optimal-discriminator analysis	129
9.2.3	Anderson's reverse-time stochastic differential equation	130
9.2.4	Score matching: Hyvärinen's identity	131
9.2.5	The K6 identification: score as amplitude gradient	132
9.2.6	Wick rotation: diffusion as imaginary-time Schrödinger evolution	134
9.3	Worked exercises with full solutions	135
9.4	Outlook	140
10	Foundation Models and Self-Supervised Learning	143
10.1	Notation and prerequisites	143
10.2	Theorems and full proofs	144
10.2.1	Self-supervised objectives as Kullback–Leibler minimisations	144
10.2.2	The K7 structural decomposition: pre-training loss as variational gap	145
10.2.3	The K7 rate statement: asymptotic scaling	148
10.2.4	Transfer learning: data-efficiency bounds	149
10.2.5	The K7 conjectural extension and the K10 connection	150
10.3	Worked exercises with full solutions	151
10.4	Outlook	155
11	Machine Learning for Biological Sequences and Structures	157
11.1	Notation and prerequisites	157
11.2	Theorems and full proofs	158
11.2.1	Sequence embeddings as coordinate charts on the wave-function support	158
11.2.2	Geometric deep learning: equivariance and quotient spaces	159
11.2.3	Multi-modal models as entangled joint wave functions	160

11.2.4	Transfer learning as projection	161
11.2.5	The K8 master correspondence	162
11.2.6	The Part II dictionary, formally	164
11.3	Worked exercises with full solutions	165
11.4	Outlook: Part II closes, Part III opens	170
III De Broglie Waves in Molecular Biology		171
12 De Broglie Waves for Biomolecules		173
12.1	Notation and prerequisites	173
12.2	Theorems and full proofs	174
12.2.1	The thermal de Broglie wavelength	174
12.2.2	Migration of wave nature into internal coordinates	175
12.2.3	Joos–Zeh decoherence: two-regime structure of the master equation	177
12.2.4	Coherence length: kinematic bound from coherence time	180
12.2.5	The three regimes: rigorous classification	181
12.3	Worked exercises with full solutions	181
12.4	Outlook: Part III begins with the equation of motion	185
13 A Biological Schrödinger Equation		187
13.1	Notation and prerequisites	187
13.2	Theorems and full proofs	188
13.2.1	Self-adjointness of the biological Hamiltonian	188
13.2.2	Spectral decomposition and discreteness of the bound-state spectrum	189
13.2.3	Time evolution: the Trotter–Kato product formula	190
13.2.4	Madelung decomposition: amplitude and phase	191
13.2.5	Wick rotation: rigorous duality with the Smoluchowski equation	193
13.3	Worked exercises with full solutions	195
13.4	Outlook: from dynamics to dissipation	199
14 Tunnelling, Decoherence, and Enzymatic Catalysis		201
14.1	Notation and prerequisites	201
14.2	Theorems and full proofs	202
14.2.1	The Lindblad / GKSL master equation	202
14.2.2	WKB transmission probability	203
14.2.3	Joos–Zeh as a Lindblad realisation	206
14.2.4	Kinetic isotope effect: tunnelling enhancement	207
14.2.5	Marcus theory of electron transfer	208
14.3	Worked exercises with full solutions	210
14.4	Outlook: from single-particle tunnelling to cellular and multi-residue wave mechanics	213
15 Cellular Wave Function and Stochastic Gene Expression		215

15.1	Notation and prerequisites	215
15.2	Theorems and full proofs	216
15.2.1	The Doi–Peliti Fock-space formulation	216
15.2.2	The telegraph model: rigorous derivation of the negative- binomial steady state	218
15.2.3	The Gillespie algorithm: exact sampling from the CME	221
15.2.4	Spectral decomposition of the temporal correlation function	222
15.2.5	Wick-rotation duality for the cellular wave function	223
15.3	Worked exercises with full solutions	224
15.4	Outlook: from cells to multi-residue entanglement	228
16	Entanglement and Long-Range Biomolecular Correlations	229
16.1	Notation and prerequisites	229
16.2	Theorems and full proofs	230
16.2.1	The Schmidt decomposition	230
16.2.2	Entanglement entropy as von Neumann entropy of the re- duced density matrix	232
16.2.3	Potts maximum-entropy model and mean-field DCA inversion	234
16.2.4	Average product correction	235
16.2.5	Lieb–Robinson bound and the biomolecular allosteric velocity	237
16.2.6	Attention as entanglement detector (K5 revisited)	238
16.3	Worked exercises with full solutions	239
16.4	Outlook: Part III complete, Parts IV–VI ahead	242
IV	Sequence: Genomics and Transcriptomics	245
17	NGS Pipelines and the Wave Function over Sequence Space	247
17.1	Notation and prerequisites	247
17.2	Theorems and full proofs	248
17.2.1	The FM-index and alignment cost	248
17.2.2	Lander–Waterman coverage statistics	250
17.2.3	Phred quality score as Bayesian probability	251
17.2.4	Pile-up genotype maximum-likelihood estimator	252
17.2.5	Sequence-space wave function as projective measurements	253
17.3	Worked exercises with full solutions	255
17.4	Outlook: from pile-ups to variants	258
18	Variant Calling: From Likelihoods to Quantum Measurements	259
18.1	Notation and prerequisites	259
18.2	Theorems and full proofs	260
18.2.1	Bayesian genotype-posterior formulation	260
18.2.2	Pair-HMM haplotype likelihood (GATK HaplotypeCaller)	261
18.2.3	Deep-learning variant calling as posterior approximation	263
18.2.4	Coverage requirements for somatic variant detection	264
18.2.5	Variant call as projective measurement	265
18.3	Worked exercises with full solutions	266

18.4 Outlook: from variants to transcripts	270
19 RNA-seq Differential Expression and Wave Packets in Transcript Space	271
19.1 Notation and prerequisites	271
19.2 Theorems and full proofs	272
19.2.1 Negative-binomial variance formula	272
19.2.2 Empirical-Bayes dispersion shrinkage	273
19.2.3 Benjamini–Hochberg false-discovery-rate control	275
19.2.4 Welch t -test power and the five-replicate guideline	276
19.2.5 Transcript-space wave packet	277
19.3 Worked exercises with full solutions	278
19.4 Outlook: from bulk to single-cell	281
20 Single-Cell Trajectories: Cell States in an Expression Hilbert Space	283
20.1 Notation and prerequisites	283
20.2 Theorems and full proofs	284
20.2.1 Zero-inflated negative binomial for dropout	284
20.2.2 PCA on the centred cell-gene matrix	285
20.2.3 UMAP local-neighbourhood preservation	286
20.2.4 Leiden community detection	287
20.2.5 RNA velocity as kinetic generator	288
20.2.6 K8 master correspondence for single-cell foundation models	290
20.3 Worked exercises with full solutions	291
20.4 Outlook: Part IV complete, Part V opens	294
V Structure: Proteins and Macromolecules	297
21 AlphaFold, ESMFold, and the Energy Landscape as Effective Hamiltonian	299
21.1 Notation and prerequisites	299
21.2 Theorems and full proofs	300
21.2.1 Funnelled landscape and Levinthal resolution	300
21.2.2 Triangle-inequality enforcement in the Evoformer pair representation	302
21.2.3 Invariant Point Attention: $SE(3)$ -equivariance	303
21.2.4 ESMFold as K7 foundation-model specialisation	304
21.2.5 pLDDT as estimator of the eigenvalue gap	305
21.2.6 AlphaFold as variational ground-state minimisation	307
21.3 Worked exercises with full solutions	308
21.4 Outlook: from structure to function	311
22 Docking and Binding Affinity: Tunnelling at Transition States	313
22.1 Notation and prerequisites	313
22.2 Theorems and full proofs	314

22.2.1	The Zwanzig FEP identity	314
22.2.2	Thermodynamic integration	315
22.2.3	Bennett acceptance ratio: minimum-variance estimator . . .	316
22.2.4	Eyring-Kramers binding rate with WKB correction	317
22.2.5	Deuterium kinetic isotope effect for binding residence time	319
22.2.6	Binding as joint protein-ligand wave function	320
22.3	Worked exercises with full solutions	321
22.4	Outlook: from docking to design	324
23	Protein Design: Variational Principle for the Folded State	325
23.1	Notation and prerequisites	325
23.2	Theorems and full proofs	326
23.2.1	The design variational principle	326
23.2.2	RFdiffusion as score-matching generative model	327
23.2.3	ProteinMPNN autoregressive sequence decoding	328
23.2.4	Sequence-recovery rate as eigenvalue-gap consequence . . .	329
23.2.5	Binder design hit rate	331
23.2.6	Gap-aware design refinement	332
23.3	Worked exercises with full solutions	333
23.4	Outlook: from design to dynamics	336
24	Allostery and Dynamics: Phonons and Coupled Oscillators	337
24.1	Notation and prerequisites	337
24.2	Theorems and full proofs	338
24.2.1	The ENM Hessian and spectral decomposition	338
24.2.2	The GNM Kirchhoff-matrix structure	339
24.2.3	The ANM $3N$ -dimensional analysis	340
24.2.4	Phonon dispersion on the residue graph	341
24.2.5	AlphaFlow as conformational ensemble sampler	342
24.2.6	Cryptic pockets as surface-projected modes	343
24.3	Worked exercises with full solutions	344
24.4	Outlook: Part V closes, Part VI opens with immunotherapy	348
VI	Signal: Cancer and Immunotherapy	349
25	TCR and BCR Repertoires as Wave Packets	351
25.1	Notation and prerequisites	351
25.2	Theorems and full proofs	352
25.2.1	V(D)J combinatorial diversity	352
25.2.2	Hill numbers and diversity profile	353
25.2.3	IGoR/OLGA generation-probability model	354
25.2.4	Antigen-driven expansion as hierarchical-Bayes test	355
25.2.5	Somatic hypermutation as random walk	356
25.2.6	Repertoire wave packet	357
25.3	Worked exercises with full solutions	358
25.4	Outlook: from repertoires to neoantigens	361

26 Neoantigen Prediction as a Superposition of Binding Modes	363
26.1 Notation and prerequisites	363
26.2 Theorems and full proofs	364
26.2.1 NetMHCpan binding model	364
26.2.2 Eluted-ligand training advantage	364
26.2.3 TCR-pMHC docking framework	365
26.2.4 Attrition-funnel composition	365
26.2.5 Bayesian update with empirical-expansion prior	366
26.2.6 Joint peptide-MHC-TCR amplitude with interference	366
26.3 Worked exercises with full solutions	367
26.4 Outlook: from neoantigens to CAR-T	369
27 CAR-T Design as an Engineered Wave Function	371
27.1 Notation and prerequisites	371
27.2 Theorems and full proofs	372
27.2.1 Modular CAR Hamiltonian decomposition	372
27.2.2 Antigen-density-threshold triggering	373
27.2.3 Affinity Goldilocks window	373
27.2.4 Off-target risk as scFv-healthy-tissue overlap	375
27.2.5 Logic-gated CARs as projector compositions	375
27.2.6 CAR design as constrained variational principle	376
27.3 Worked exercises with full solutions	376
27.4 Outlook: from CAR-T to clinical response	379
28 Clinical Response Prediction: The Measurement Problem in the Clinic	381
28.1 Notation and prerequisites	381
28.2 Theorems and full proofs	382
28.2.1 Kaplan–Meier estimator	382
28.2.2 Cox proportional-hazards partial likelihood	383
28.2.3 Hierarchical-Bayes response prediction in small cohorts	384
28.2.4 Digital twin as patient-specific wave function	385
28.2.5 Counterfactual identification under randomisation	385
28.2.6 Clinical response as wave-function collapse	386
28.3 Worked exercises with full solutions	386
28.4 Outlook: Part VI closes, Part VII opens with quantum biology	389
VII Perspectives	391
29 Quantum Biology: Empirical Evidence	393
29.1 Notation and prerequisites	393
29.2 Theorems and full proofs	394
29.2.1 FMO vibronic Hamiltonian	394
29.2.2 Radical-pair singlet-yield formula	395
29.2.3 Compass anisotropy from anisotropic hyperfine	396
29.2.4 Enzymatic KIE beyond Swain–Schaad	397

29.2.5	Inelastic-electron-tunnelling olfactory spectrometer	397
29.2.6	Empirical-evidence scoreboard	398
29.3	Worked exercises with full solutions	398
29.4	Outlook: from quantum biology to quantum computing	401
30	Quantum Machine Learning for Genomics and Structure	403
30.1	Notation and prerequisites	403
30.2	Theorems and full proofs	404
30.2.1	Parameter-shift gradient rule	404
30.2.2	VQE variational convergence	405
30.2.3	Quantum-kernel computation via SWAP test	405
30.2.4	Hamiltonian-simulation complexity	406
30.2.5	Barren-plateau phenomenon	406
30.2.6	Realistic-horizons scoreboard	407
30.3	Worked exercises with full solutions	408
30.4	Outlook: from quantum ML to the K10 conjecture	410
31	Foundation Models as Universal Wave Functions: The K10 Master Conjecture	411
31.1	Notation and prerequisites	411
31.2	Theorems and full proofs	412
31.2.1	The K10 master conjecture	412
31.2.2	The three structural pillars	413
31.2.3	Partial-evidence catalogue	414
31.2.4	Spectral-features and SSL equivalence	415
31.2.5	Open problems: dynamics, multi-scale, causality	415
31.2.6	Closing variational-unification statement	416
31.3	Worked exercises with full solutions	417
31.4	Coda: the wave is launched	419
	Bibliography	421

Preface

This is Volume II of *Molecular Wave Mechanics*. Volume I gave the wave-mechanical reading of biology in its didactic form: analogies of daily life, intuitions, conceptual figures, and the load-bearing conjecture named at the end. The present volume is its rigorous companion. Each chapter here corresponds, one-to-one and in the same order, to a chapter of Volume I. The intuitions of Volume I become theorems with full proofs; the worked examples of Volume I become exercises with complete solutions; the keystone equations of Volume I are stated as formal results and derived from first principles.

The reader is expected to have read, or to be reading in parallel, the corresponding Volume I chapter. Where Volume I says “the variational principle gives the network a principled loss,” Volume II proves the variational principle, derives the loss, and verifies its convergence properties. Where Volume I introduces a wave function on configuration space, Volume II constructs the configuration space, defines the wave function rigorously, and supplies the operator-theoretic prerequisites the construction needs.

A note on the omission of code. Volume I included Python sources for every numerical figure; the present volume contains none. All implementations, datasets, and reproducible figures of both volumes are made available online at quantum.abdelkader.site, the molecular-biology toolkit that has served as the implementation substrate for the author’s work since 2018. The volume’s mathematical content is independent of the implementations; the implementations, when consulted, sit beside the proofs rather than inside them.

A note on the level. The mathematical level of this volume is graduate. The treatment uses, without re-introduction, measure-theoretic probability, the theory of unbounded self-adjoint operators on Hilbert space, functional-analytic methods for partial differential equations, and standard categorical and algebraic constructions at the level of a first-year graduate sequence. The exercises range from routine verifications, suitable for self-testing of the prerequisite material, to research-level open problems flagged as such, often connected to the conjecture that closes the volume.

The author thanks again the readers, students, and collaborators who engaged with Volume I, whose questions are the proximate cause of the formal treatments this companion volume develops. Errata and corrections to either volume are gratefully received through the quantum.abdelkader.site contact channel; both volumes are intended to be read together, and a correction to one almost always concerns the other.

Volume I argued that biological machine learning is wave mechanics. This volume proves it.

Every keystone equation of Volume I is restated as a formal theorem and derived in full: the Born rule from Gleason's theorem, the fidelity–Bhattacharyya identity, the Rayleigh quotient as the Courant–Fischer characterisation, the neural-quantum-state identification, attention as a pair-correlation operator, and the master correspondence that ties them together.

169 theorems. 329 proofs, written out. 150 worked exercises with complete solutions, graded from routine verification to research-level open problem. The thirty chapters mirror Volume I one for one, and each has the same four sections: notation and prerequisites, theorems and full proofs, worked exercises with solutions, and an outlook. Keep the two volumes open side by side — the narrative on the left, the derivation on the right — or read this one alone, as a graduate course in the mathematics of computational biology. The level is graduate throughout: measure-theoretic probability, the spectral theory of unbounded self-adjoint operators and the functional-analytic treatment of parabolic equations are used as working tools. Notebooks and live demos: quantum.abdelkader.site.

Abdelkader BOUSABAA has taught applied mathematics, stochastic methods, scientific computing and artificial intelligence at the University of Paris-Saclay for more than twenty years, and led a biotech company in the orbit of the *Institut Curie* between 2018 and 2023, applying these methods to the design of T-cell immunotherapies against cancer.