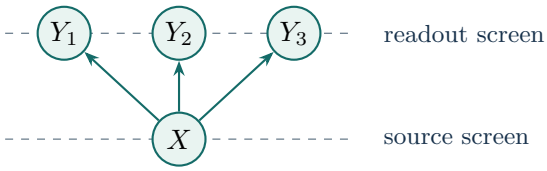
RESEARCH MANUSCRIPT

Replication and Persistence in Spacetime

Higher topos semantics, causal computation, entropy, and free energy



Prepared for Higher Categorical Semantics
Anonymous manuscript · November 2026

Abstract

A record persists when successive observations admit a coherent continuation. It replicates when several designated carriers acquire individually decodable instances of a shared label. We give these conditions a semantics in sheaves of spaces on a spacetime site, and give their implementations a separate semantics in finite causal networks. A homotopy fiber describes possible reconstructions with a fixed boundary observation. For local marginal laws on a tree cover, a separator construction produces the unique entropy-maximizing joint law, with an exact relative-entropy deficit for every other coupling. Cyclic covers can obstruct even the existence of a joint law. A causal cut controls how much of a source can be recovered beyond it, with a conditional Fano bound that accounts for preexisting side information.

For r local decoders with error probabilities ε_i, we prove a lower bound on conditional total correlation of the carriers by $(r-1)\mathsf{H}(X \mid W) - \sum_i f_m(\varepsilon_i)$, where X has m labels and W is declared side information. The exact identity underlying this bound separates ambiguity about the source from correlation already present within its conditional ensembles. At product-energy endpoints the same correlation contributes $k_{\mathrm{B}}T\mathsf{TC}$ to nonequilibrium free energy. In a specified isothermal protocol class the work ledger has a telescoping relative-entropy loss and yields a resource-sensitive replication inequality. With flat, uncorrelated thermal targets and symmetric binary readout error e, the work infimum is $rk_{\mathrm{B}}T(\log 2 - h_{\mathrm{b}}(e))$ in a class with controlled conditional equilibration. Pure blanks can instead fund exact reversible fan-out. A refresh theorem gives the exact restoration cost, in the same sense of infimum, for a noisy binary memory with a retained reference.

We also give finite and compact persistence criteria, an example of incompatible coherent lifts over arbitrarily long finite horizons, and a causal comparison of exact copies, erasure-coded fragments, and actively refreshed memories. Classical record copying and finite-dimensional quantum broadcasting obey distinct criteria. Throughout, the spatial support, coupling, observables, and bath model appear as explicit data.

Keywords. Higher topoi, causal networks, descent, persistence, replication, total correlation, Gibbs-preserving dynamics, nonequilibrium free energy.

Authorship and assistance. This manuscript uses an anonymous byline. Generative AI was used to draft, typeset, and check selected finite examples. The proofs supply the mathematical evidence under the stated hypotheses.

Contents

Abstract i

1 The problem of a continuing record 1

2 Spacetime and observation sites 5

3 Reconstruction by higher descent 9

4 Causal computation and cut bounds 13

5 Logical and effective semantics 17

6 Persistence over increasing horizons 20

7 Entropy of jointly decodable records 23

8 Marginal descent and entropy of gluing 27

9 Transport of couplings and symmetry 32

10 Finite isothermal protocols 36

11 The work of creating replicas 39

12 Optimal preparation at a prescribed readout accuracy 44

13 The work of maintaining a record 49

14 Causal designs for storing a label 53

15 Classical labels at a quantum interface 56

16 Fibrations of histories and continuations 57

17 Information across spacetime screens 61

18 Continuous-time storage and finite activity 65

19 Local gates and inaccessible correlations 69

20 Relative boundary conditions and topological obstructions 72

21 Noisy broadcast along a tree 77

22 Lineage survival and growing populations 79

23 Identifying the label shared by several carriers 81

24 Interactive cuts and zero-error interfaces 84

25 Refinement, coarse observation, and hidden state 87

26 Combining the reconstruction and resource tests 91

A Finite information inequalities 94

B Finite designs with explicit probabilities 96

C Higher diagrams of reconstruction 99

D Operational certificates for finite models 109

References 114

CHAPTER 1

The problem of a continuing record

Consider a source that records one label, a memory that moves along a world tube, and two further memories that receive signals in disjoint regions. A final observer can ask whether each memory yields the source label, whether the moving memory at an intermediate time and the memory at a later time belong to one history, and how much energy had to be supplied to the device. These questions have different invariances. A change of device name leaves readout accuracy unchanged. A change in the admissible causal paths can change that accuracy. A change in the energy assigned to a blank can change the work, even if the entire classical input–output map stays fixed.

We develop a theory in which these distinctions survive composition. The region-indexed object of histories is a sheaf of spaces. Its restriction maps take observations of a history to smaller regions. A causal process generates histories by transmitting signals from inputs to later outputs. An ensemble on a finite realization of this process adds the laws from which entropy and work can be calculated. On an overlapping cover the relevant question is joint realizability. On a directed circuit the relevant question is causal reachability with a sufficiently informative cut. On an isothermal protocol the relevant question is the difference between full free energies and the reference distributions used in each step.

1.1 Three tests on one device

Take a fair bit X and two later registers Y_1, Y_2. The requirement $Y_1 = Y_2 = X$ can be checked at each terminal with its own decoding procedure. Yet the two terminal one-bit marginals alone cannot test the requirement. The independent fair pair has the same marginals as (X, X). Even the joint condition $Y_1 = Y_2$ leaves open whether the common value was inherited from X. It might instead be a fair bit prepared at a third site. The joint law of the source and terminal observables, or an intervention on the source, resolves the ambiguity.

The entropy of the diagonal pair is $\log 2$ in nats. The independent fair pair has entropy $2 \log 2$. Their marginal energies can coincide in a device with a flat spectrum. If the two memories have independent thermal equilibrium distributions, the diagonal law then carries $k_{\mathrm{B}} T \log 2$ more free energy than the independent law. This excess refers to an ensemble and a specified reference. It gives no bound on the work of an operation until the initial state, ancillary degrees of freedom, and retained registers enter the calculation. A controlled

exclusive-or on a source and a pure blank is a permutation. On its prepared input it creates a diagonal pair while leaving joint entropy and flat energy fixed.

An observer who reads at progressively later times faces a different problem. A memory can exist at every finite horizon and still lack a single history compatible with the choices already made. In a finitely branching realization, compactness remedies this when every finite horizon has a possible continuation. In a homotopy-coherent realization, paths witnessing agreement may themselves require coherent choices. The reconstruction is a space of fillers. Even the existence of a filler and uniqueness up to contractible choice require different statements.

1.2 Statement of results

The hypotheses in the left column identify the data used by the corresponding conclusions.

Data	Consequence
History sheaf and prescribed observation	Reconstruction is a homotopy fiber. Its formation respects descent and gluing along matching interfaces (theorem 3.1).
Finite causal network and fixed past cut	Every downstream decoder factors through the cut, with declared side information. Conditional entropy at the source bounds the amount that must cross the cut (theorems 4.1 and 4.2).
Tower of compact reconstruction spaces	Finite extendibility implies an infinite continuation under stated compactness and compatibility assumptions (theorems 6.1 and 6.2).
Declared local decoders and a joint law	Total correlation bounds simultaneous recoverability and has a residual-correlation equality term (theorem 7.1).
Compatible marginals on a tree cover	The separator law uniquely maximizes entropy and exposes the entropy deficit of every other coupling (theorems 8.1 and 8.2).
Isothermal finite protocol with retained records	A stepwise relative-entropy ledger gives a replication work inequality and, under controlled equilibration, a sharp binary refresh bound (theorems 10.1, 11.1 and 13.1).

The causal and thermodynamic results apply to finite classical readouts. The higher semantics governs the witnesses, identifications, and restrictions for which those readouts are defined. A sheaf of spaces can record many distinctions that a decoder ignores. Passing from a homotopy type of implementations to a law of finite readouts therefore requires an observation map and an ensemble.

1.3 Choice of primitives

Fix a time-oriented spacetime M, an observation site $\mathcal{C}_M$, and a sheaf $\mathcal{H}$ on it with values in spaces. A *history* is a section, and its observation is the image of that section in another sheaf. A *carrier* is a designated physical or logical interface, with a specified spacetime support and a finite observable when used in the quantitative theory. A *label* is an element of a finite set L. Equality of labels is part of the readout specification. Equality of two histories is an identity in $\mathcal{H}$, so paths and higher paths can witness equivalences of realizations.

A causal implementation adds a finite directed acyclic graph whose arrow supports are future-directed. A process at a vertex acts on an explicitly declared interface. A readout is a finite function or channel of the variables available to an observer. A replication contract then specifies a reference label, source boundary, readout boundaries, and separate decoders at the targets. It can additionally specify whether the source is retained. A persistence contract specifies a tower of horizons, its observation constraints, and which past choices must be fixed. These data resolve an ambiguity in a putative copy at a later time where two terminals could be two observations of the same degree of freedom.

For thermodynamics we choose a finite microstate set S, an energy function E, and a bath at temperature $T > 0$. A protocol on S alternates Hamiltonian changes at fixed law with Gibbs-stationary stochastic steps at fixed Hamiltonian. Those steps have a relative-entropy accounting identity. A reservoir interpretation requires the corresponding physical hypotheses, such as detailed balance or an explicit driven implementation. The finite energy on S is a model input. A globally conserved energy on the ambient spacetime would be additional structure.

1.4 A separation of directions

A causal arrow $a \to b$ transports an input at a into data at b. A restriction $\mathcal{H}(U) \to \mathcal{H}(V)$ for $V \subset U$ forgets information about the already-described region. The direction of restriction is contravariant even when the experimental dynamics is future-directed. For a finite wire with input A and output B, the history relation can be presented as a span $A \leftarrow R \to B$. Its pullback along another history relation composes histories with matching middle boundaries. Supplying a stochastic kernel on the span additionally prescribes how those histories are weighted.

The higher structure concerns coherence of those history comparisons. A code trace may admit distinct valid refinements, and a region may support a gauge action on its realizations. A groupoid-valued history theory already

retains automorphisms of realizations. A space-valued theory retains all higher coherences of those automorphisms and their comparisons. Taking discrete fibers is a hypothesis about the observed model, and keeps the finite formulas exact.

1.5 A model comparison

The following four examples distinguish the scope of the word replication.

1. A deterministic register with retained source and a blank receiver implements $(x, 0) \mapsto (x, x)$. Each carrier has an individual decoder. The state-preparation cost of the blank belongs to the boundary accounting.
2. A two-of-three erasure code can preserve a message when a carrier is lost. Each surviving pair has a decoder. A single share need not determine the message, so the pair is the readout interface for that contract.
3. A population of descendants may grow while its source label becomes unidentifiable along any one branch. The population size, existence of a surviving branch, and recovery of source information are distinct random variables.
4. An independently prepared fair bit can agree with a source on half of a sample of trials. If their common randomness is declared as side information, the residual causal demand is $\mathsf{H}(X \mid W)$. An intervention can test that the receiving process depends on its input beyond that preparation.

A special case of the categorical model is a deterministic Boolean circuit embedded in a causal network. A special case of the stochastic model is a memoryless binary channel. A special case of the energy model is a degenerate bit array with a common bath. We work these cases through the same set of definitions so that the translations between the abstract results and their operational content are explicit.

1.6 Relation to prior work

The use of sheaves to express compatibility of observations follows standard topos theory [22, 21, 27]. Interval sheaves and wiring diagrams give established semantics of behavior and composition [28, 29]. This manuscript uses spatially overlapping regions as well as time horizons. Contextuality provides a precedent for local probabilistic data with no global joint law [1]. We use a classical finite obstruction and a tree gluing result to locate what information the regional marginals determine.

The causal graph semantics draws on distributed event order, event structures, causal models, and the category of Markov kernels [17, 33, 25, 13]. The entropy identities build on Shannon information and the probabilistic graphical model literature [31, 8, 19]. The thermodynamic ledger uses finite versions of known entropy production and nonequilibrium free-energy relations [18, 4, 30, 11, 24]. England's kinetic bound concerns the forward and reverse rates of a coarse self-replication event [10]. Our primary work inequality concerns formation of jointly decodable records and tracks initial correlations and ancillary free energy. We give a precise comparison in chapter 10.

The mathematical contribution is the explicit combination of the reconstruction fiber, causal cut, replica-correlation identity, and resource ledger. Each theorem states the layer of structure on which it depends. The later examples compare settings with the same label multiplicity and different causal or thermodynamic resources.

CHAPTER 2

Spacetime and observation sites

Let M be a smooth, time-oriented Lorentzian manifold. Write $p \ll q$ when a future-directed timelike curve connects p to q, and $p \preceq q$ when $p = q$ or a future-directed causal curve connects them. If M has no closed causal curves, $\preceq$ is a partial order. Its reflexivity and transitivity give a thin category. This category records reachability. There may be many distinct causal trajectories with the same endpoints. To distinguish them, choose a graph of admissible world tubes or use a category of paths with a specified equivalence relation and concatenation. The results about computation below take the finite graph as input and use the ambient relation to test admissibility.

2.1 Two bases with different variance

The observation site $\mathcal{C}_M$ may be the poset of open subsets of M, with inclusions and ordinary open covers. If M is strongly causal, chronological diamonds

$$D(p, q) = I^+(p) \cap I^-(q), \qquad p \ll q,$$

form a basis for the manifold topology. A basis site with the induced covers gives the same sheaf semantics. Intersections of diamonds need only be unions of basic opens. Cover descent consequently uses the topology generated by the diamonds, while the arrows of a causal network are tested with $\preceq$. These two uses of causal geometry have different purposes. A receiver on a null ray need

not be strictly timelike to the source, even though the topology is described using timelike diamonds.

For globally hyperbolic M, a smooth temporal function t with spacelike Cauchy levels exists [5]. We can choose finitely many levels as reference screens, and then select a finite set of events for any protocol within the modeled slab. All graph arrows point to strictly later levels of t except where a protocol has been represented by events ordered within one nominal round. In the latter case it is enough that the individual event placements increase t. Given an acyclic graph with a placement satisfying the causal arrow constraints, any topological sort is a valid algebraic evaluation order. A sort need not arise from an assigned foliation of M.

The region U containing a measurement can be open even if a physical device is supported on a compact world tube or an idealized hypersurface. A distinguished set of variables is attached to such a boundary by an observation morphism or span. Boundary evaluation is additional data. When modeling sharp boundaries with cells, the incidence site below provides one concrete representation.

Definition 2.1 (Causal placement)**.** A finite directed graph of signal dependencies has a causal placement if its vertices have event positions in M and every arrow is equipped with a future-directed causal path from its origin to its endpoint. Logical subcircuits can be represented by a group of vertices and paths in a region. Simultaneous readout carriers may be chosen with pairwise spacelike separation.

Pairwise spacelike endpoints can share a cause. A source can affect several such endpoints if all are in its causal future. For instance, in $1+1$ Minkowski space with $c = 1$, put the source at $(0, 0)$, and receivers at $(2, -1)$ and $(2, 1)$ in (t, x)-coordinates. Each receiver is timelike to the source and they are spacelike to each other. From a source at $(0, 0)$ to $(1, 2)$ there is no causal path. A model with independent receiving noise and no side information cannot implement fresh exact copying to the latter event. The proof depends on the causal Markov factorization of the chosen implementation, and is given in chapter 4.

2.2 A finite incidence site

A finite local model consists of a set of variable symbols V, finite domains A_v, and a finite family of constraint cells e, with scopes $S_e \subseteq V$ and relations $R_e \subseteq \prod_{v \in S_e} A_v$. Form a poset D whose elements are variables and cells, with $v \leq e$ for $v \in S_e$. The incidence relation is independent of the directed communication

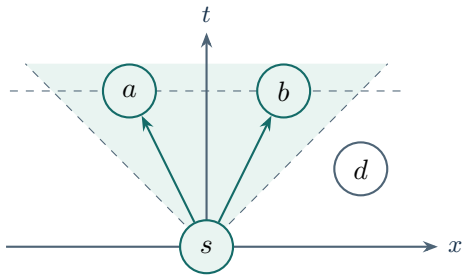


Figure 2.1: Two spacelike-separated receiving events lie in the future cone of s. Event d is outside that cone. The shaded region illustrates admissible signal placement.

order. Equip the lattice of downsets $U \subseteq D$ with union covers. Define

$$\mathsf{Hist}_D(U) = \left\{ a \in \prod_{v \in V \cap U} A_v \;\middle|\; (a_v)_{v \in S_e} \in R_e \text{ for every cell } e \in U \right\}. \tag{2.1}$$

Each scope of a cell in U is contained in U. Restrictions forget variables and constraints outside a smaller downset. One can take R_e to be the graph of a deterministic update, or the positive support of a kernel on its inputs and outputs.

Proposition 2.2 (Finite history sheaf)**.** *The assignment in* (2.1) *is a sheaf of sets, hence also a* 0*-truncated sheaf of spaces. For a binary cover* $U = U_1 \cup U_2$*, it satisfies*

$$\mathsf{Hist}_D(U) \cong \mathsf{Hist}_D(U_1) \times_{\mathsf{Hist}_D(U_1 \cap U_2)} \mathsf{Hist}_D(U_2).$$

Proof. Take a matching family over a cover. Each variable in U appears in a member, and the matching property makes its assigned value independent of the member. A cell e in U lies in some member, which also contains its entire scope. Its relation holds there and therefore holds for the assembled assignment. The assembled assignment is unique since the cover contains every variable. For a union of no members, U is empty and the empty assignment is the unique section. □

When only the variable symbols are included, the sheaf is the product of coordinate choices on that region. If a constraint spans two pieces and its cell lies in neither piece, those pieces do not cover the full constrained region. This is how the site distinguishes agreeing on values from satisfying a dynamical rule whose entire scope has not been included. Similarly, a spacetime cover must account for transition data carried through an interface. Forgetting that data changes the local model.

Example 2.3 (One crossed wire)**.** Put one symbol a upstream and one b downstream, with $A_a = A_b = \{0, 1\}$, and a cell imposing $b = a$. The downset $\{a, b\}$ has four histories. The downset $\{a, b, e\}$ has two. Both have the same two possible terminal marginals as sets. A boundary observer that reads both values distinguishes the joint admissible sets. The inclusion of the unconstrained downset into the constrained one induces the inclusion of diagonal histories into the four possible observations.

2.3 Relabeling and transport

An isomorphism of placed networks consists of a directed graph isomorphism, bijections of the alphabets, and corresponding reindexings of observations and factors. In a Lorentzian model, a time-orientation-preserving isometry carries an admissible placement to another. Under a bijective reindexing preserving the kernels, every finite trajectory probability is carried to its counterpart. Relative entropy, the entropy of fixed readouts, and the energy expectation with the transported energy function are preserved. These claims follow from changing variables in the defining sums. They express covariance of the model and leave room for inequivalent physical clocks or reservoir choices.

On opens, a homeomorphism $f : M \to M'$ induces an equivalence of sites and of their sheaf topoi. Under this equivalence, the fiber of a history map over a fixed observation is transported to the corresponding fiber. If f also preserves the time orientation and causal relation, it transports admissible causal placements. For an inclusion of a region, the restricted model retains observations available inside that region. It can have more possible boundary states than the image of actual global histories. A local law on a restricted region may arise by marginalizing over states from outside it, in which case that image must be specified.

2.4 Geometry used by the energetic model

The proofs below compare energies of explicit modeled states. A laboratory may use a rest frame and a bath with fixed temperature over a chosen interval of its proper time. A collection of spacelike terminals gives a boundary for the protocol only after the state and the energetic reference have been defined. Global hyperbolicity supplies Cauchy slices and order. A timelike Killing field, a conserved charge, a stationary equilibrium or a rule for comparing baths at separated sites would be further data. For a nonstationary background, the results apply to a specified finite thermodynamic realization with a common T, or

separately to individual isothermal modules when their interfaces are accounted for.

The functorial perspective of locally covariant field theory also separates background transport from dynamical properties [6]. In that setting observables form algebras and spacelike locality is formulated using commuting algebras. In the classical finite setting here, locality is established by conditional factorization of a specified graph with explicit noise sources. The mathematical role of the ambient spacetime is to constrain realizable placements and observation regions. The finite model then supplies its own operational and thermodynamic rules.

CHAPTER 3

Reconstruction by higher descent

Fix a small observation site $\mathcal{C}$ and let $\mathcal{E} = \mathrm{Sh}_\infty(\mathcal{C})$. We use sheaves of spaces for ordinary cover descent. Where hypercovers are used, we either work in the hypercomplete topos or assume that the particular objects satisfy hyperdescent. Let U stand also for the sheafification of its representable presheaf. For $A \in \mathcal{E}$, $A(U) = \mathrm{Map}_{\mathcal{E}}(U, A)$. A section is a point of this mapping space. For a map $o : \mathcal{H} \to B$ and $b : U \to B$, define

$$\mathscr{R}_o(U, b) = \mathrm{Map}_{\mathcal{E}_{/B}}\big((U, b), (\mathcal{H}, o)\big) \simeq \mathrm{hofib}_b\big(\mathcal{H}(U) \longrightarrow B(U)\big). \tag{3.1}$$

A point consists of a history and an identification of its observation with b. The specified identification matters for automorphisms and for gluing. This is also the section space of $U \times_B \mathcal{H} \to U$. Nonemptiness gives existence of reconstruction. A singleton of π_0 gives uniqueness up to equivalence. Contractibility additionally makes all higher comparisons unique up to higher equivalence.

3.1 The relative descent theorem

Let $U_\bullet \to U$ be the Čech nerve of a covering family, using the sheafified representables and their coproduct and intersections in $\mathcal{E}$. Write b_n for the pullback of b to U_n.

Theorem 3.1 (Reconstruction commutes with descent)**.** *There is a natural equivalence*

$$\mathscr{R}_o(U, b) \simeq \lim_{[n]\in\Delta} \mathscr{R}_o(U_n, b_n). \tag{3.2}$$

The same conclusion holds for any augmented diagram over U that presents U as its colimit in $\mathcal{E}$. In particular, it holds for hypercovers under hyperdescent.

Proof. Covers in the topos give effective epimorphisms. Their Čech nerves have colimit the covered object. The forgetful functor $\mathcal{E}_{/B} \to \mathcal{E}$ creates colimits, with structural map the one induced by the cocone. The diagram in the slice given by (U_n, b_n) therefore presents (U, b) as its colimit. Mapping out of a colimit into $(\mathcal{H}, o)$ yields the limit of the mapping spaces. Equivalently, the two sheaf descent equivalences for $\mathcal{H}$ and B can be inserted in the homotopy-fiber square. Small limits commute with pullbacks in spaces. □

Corollary 3.2 (Two overlapping regions)**.** *For a cover by $U_1, U_2 \subseteq U$, the space of reconstructions is*

$$\mathscr{R}_o(U_1, b_1) \times^h_{\mathscr{R}_o(U_1 \cap U_2, b_{12})} \mathscr{R}_o(U_2, b_2).$$

If both local spaces and the intersection space are contractible, the global space is contractible.

Proof. In an ∞-topos the union of two subobjects is their pushout along the intersection. Apply theorem 3.1 to the pushout diagram. The pullback of three contractible spaces is contractible because the diagram is equivalent to the terminal diagram. □

Contractible choices on each member of a general cover suffice when the corresponding reconstruction spaces on all finite intersections are also contractible. In that case the entire descent diagram is pointwise equivalent to the terminal diagram. Nonempty local choices alone give a different property. For sets, a matching family must agree on each pairwise overlap. For groupoids, chosen overlap isomorphisms must satisfy the cocycle equation. For spaces, comparison on a triple intersection is an additional homotopy between composites of paths, and the totalization imposes all coherence conditions.

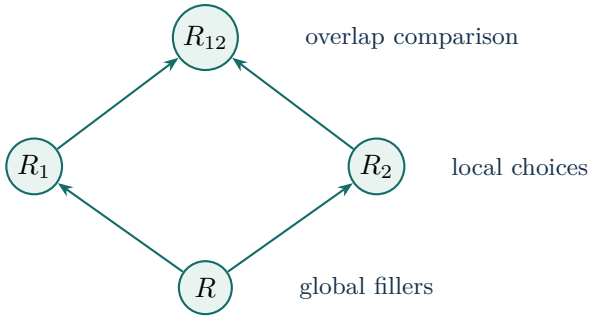


Figure 3.1: The binary reconstruction square is a homotopy pullback. A filler includes a comparison of the two restrictions at the top vertex.

3.2 A twist visible in a groupoid of records

On a circle, consider the two-sheeted connected cover $p : S^1 \to S^1$, $p(z) = z^2$. It defines a locally constant sheaf of two choices on the base. Over any small arc one can choose a sheet. A global section would lift a degree-one loop to a closed loop and induce an integer k with $2k = 1$ on fundamental groups, which is impossible. Thus local witnesses for a record can all be present while the global continuation around a spatial cycle fails. The obstruction is a permutation of choices when transported once around the circle.

There is a related stack-valued example. Let G be a discrete group and let BG denote its classifying space. The groupoid of locally constant G-torsors on U classifies locally constant objects with fiber a free transitive G-set. On S^1 the groupoid of such torsors is equivalent to the action groupoid of G acting on itself by conjugation. A torsor has a monodromy g once a point in the fiber is chosen. A new choice conjugates g. Its automorphism group is the centralizer $C_G(g)$. A cover by arcs therefore permits every local torsor to be trivial while distinguishing global histories labeled by conjugacy classes, each with its own group of automorphisms. These examples can be extended trivially in the time direction by pulling the cover and the torsor back along projection to the spatial circle.

Example 3.3 (Higher identifications on a spatial sphere)**.** Take $M = \mathbb{R} \times (S^2 \times \mathbb{R})$ with product metric $-dt^2 + g_{S^2} + dz^2$, and an open slab U homotopy equivalent to S^2. Let $\mathcal{H}$ be the locally constant sheaf of spaces with fiber the Eilenberg–Mac Lane space $K(A, 2)$, for an abelian group A. On this manifold, its sections identify with $\mathrm{Map}(\mathrm{Sing}\, U, K(A, 2))$. This is the standard description of sections of a constant local system on a locally contractible manifold, or follows by a good cover and the homotopy colimit description of its nerve. Since $K(A, 2)$ admits an abelian group space model, evaluation at a chosen point of U splits off the factor $K(A, 2)$, and the based mapping space is $\Omega^2 K(A, 2) \simeq A_{\mathrm{disc}}$. Thus

$$\mathcal{H}(U) \simeq A_{\mathrm{disc}} \times K(A, 2).$$

If $A \neq 0$, each component has nonzero π_2. Prescribing evaluation at a point gives a discrete fiber with components A. A model that retains only global connected components misses the change from $A_{\mathrm{disc}} \times K(A, 2)$ to A_{disc} produced by the boundary observation.

3.3 A replication object over a reference base

Let B classify reference situations and let $\lambda : B \to L$ give the reference label, with L 0-truncated. For $i = 1, \dots, r$, fix a carrier object $Y_i \to B$ and a decoder $d_i : Y_i \to L$. Write $P = \prod_B Y_i$, and let $\boldsymbol{d} : P \to L^r$ collect the decoders. The replication object is

$$\mathrm{Rep}_B(Y_1, \dots, Y_r; \lambda) = P \times_{B \times L^r} B, \qquad B \xrightarrow{(\mathrm{id}, \lambda, \dots, \lambda)} B \times L^r. \tag{3.3}$$

Here $P \to B \times L^r$ combines the structural map and $\boldsymbol{d}$. Thus a point over b contains carriers over b whose labels all equal $\lambda(b)$. This relation makes sense even when no natural operation can choose such points.

Let $\mathcal{H} \to B$ be a history object with observation maps to the carriers. Pulling it back along the subobject in (3.3) gives the subobject of histories satisfying the copy condition. For a specified $b : U \to B$ its sections are reconstruction witnesses for that copy contract. The formation of this subobject commutes with pullback along $B' \to B$. A section in a family supplies choices that transform coherently across parameter changes. A local section exists in the topos sense if the corresponding image subobject contains the observation. A global section may still be obstructed by a twist such as the double cover.

If the carriers are indistinguishable up to permutation, the labeled list and its quotient represent different observations. For r identical carrier types with the same decoder, Rep_B has an S_r-action. Its homotopy quotient retains stabilizers. For example, a point fixed by all permutations has an orbit with symmetry group S_r. Taking the ordinary orbit set forgets those automorphisms. Physical multiplicity still refers to separate carrier supports or register factors before the quotient. Grouping carriers as an unordered collection alone cannot certify spatial separation.

3.4 Composition along an interface

Let a history from A to B be presented by a span $A \leftarrow H \to B$. A second history from B to C is $B \leftarrow K \to C$. The composite is

$$A \longleftarrow H \times_B K \longrightarrow C.$$

This is the space of compatible pairs, with the full homotopy of the middle identification. For three spans both iterated composites have the universal property of the limit of the same zigzag, yielding the canonical associator and its coherence. The diagonal identity spans are units. Spans therefore present a composition rule intrinsic to the topos and independent of representatives. The symmetric monoidal product of spans is induced by the cartesian product.

On finite stochastic kernels, ordinary composition sums conditional path weights. Passing to supports gives an ordinary relational span if we identify repeated witnesses with the corresponding pair. In that finite case,

$$\operatorname{supp}(KL)(a,c) \iff \exists b\ \big(L(a,b) > 0 \ \wedge\ K(b,c) > 0\big).$$

The proof is the positivity of the finite matrix sum. Thus support commutes with relational composition and products. For a pullback of higher history spaces there can be several matching histories over one visible boundary pair. The sum of probabilities assigned to those histories depends on the supplied ensemble. Neither the span nor its support relation selects those probabilities.

3.5 What uniqueness can mean

A contractible fiber permits a well-defined result up to coherent equivalence. A global section of a map supplies a choice, even when other choices exist. A (-1)-truncated image reports whether a witness exists locally, and has no information about its multiplicity. In a probabilistic implementation, even a nonempty space of successful histories may be given probability zero. Conversely a probability-one class of histories can contain different physical trajectories. The purpose of retaining the space-valued fiber is to separate these questions before specializing to ensembles.

For a finite set H with observation map $o : H \to B$, $\mathscr{R}_o(U, b)$ specializes to the set of histories compatible with the observation at U. An observation value with exactly one fiber element has an unambiguous lift. An observation value with multiple elements has a space of alternatives whose entropy requires a conditional distribution. Even in this elementary case, cardinality gives only an upper bound on that entropy. This is the discrete interface between the descent theorem and the information theory.

CHAPTER 4

Causal computation and cut bounds

Let $G = (V, E)$ be a finite directed acyclic graph with a causal placement. For a vertex v, $\mathrm{pa}(v)$ denotes its parents. Let Z_v take values in a finite alphabet A_v. A kernel $K_v(a_v \mid a_{\mathrm{pa}(v)})$ specifies the distribution of the output at v for each parent assignment. If a gate emits several outgoing messages, its output is a tuple with a specified projection for each wire. One can give every wire variable its own vertex and a deterministic projection kernel. We use that expanded convention when counting cut alphabets.

Fix the law of root inputs and conditionally independent fresh randomness for other gates. For a single joint root tuple Z_{root} this gives

$$p(z) = p_{\mathrm{root}}(z_{\mathrm{root}}) \prod_{v \notin V_{\mathrm{root}}} K_v(z_v \mid z_{\mathrm{pa}(v)}). \tag{4.1}$$

Summing in reverse topological order shows that this is normalized. Evaluating in any topological order gives this joint law, as can be checked by multiplication of the conditional factors. Its support is the set of assignments that satisfy each positivity constraint and the root support. On a site of incidence cells, restrictions of admissible assignments define the history sheaf of proposition 2.2. The weights in (4.1) also prescribe how to pass to observable marginals.

4.1 A causal cut with side information

Write $V_- \subseteq V$ for an order ideal of the reachability poset. Every parent of a vertex in V_- also lies there. Let C be the tuple of transmitted variables on wires leaving V_-. For the theorem, include any messages the receiver can read from the past in C. Let X be a function of variables in V_-, W be declared initial side information, and Y be any tuple of observations in the complement made using W. The outside root inputs have a joint law conditional on W that is independent of the past variables conditional on W. The outside transition kernels may depend on W as an additional root input. No dependence on the past except through the listed cut wires is allowed. This conditional independence may be realized by independent fresh random seeds given W. In particular, if W and the exogenous inputs are omitted from the diagram, causal reachability alone does not justify the factorization.

Theorem 4.1 (Factorization across a causal cut)**.** *There is a kernel Q such that, on positive-probability conditioning events,*

$$p(y \mid x, c, w) = Q(y \mid c, w).$$

Consequently $\mathsf{I}(X; Y \mid W) \le \mathsf{I}(X; C \mid W) \le \mathsf{H}(C \mid W)$*. For an intervention fixing $X = x$ at a root with other initial laws held fixed conditional on W, the same downstream kernel Q applies.*

Proof. Hold the cut tuple and side information fixed and evaluate the outside gates in a topological order. The first outside gates depend on W, outside root randomness, or cut variables. Inductively every later gate uses these and previously sampled outside outputs. The assumption on outside roots lets this construction be performed with the same outside randomness law for every past assignment with the given (c, w). This constructs Q, also at cut values not used in a particular input law. The factorization implies $\mathsf{I}(X; Y \mid C, W) = 0$. Expanding $\mathsf{I}(X; C, Y \mid W)$ in the two orders yields the first inequality. The second follows from $\mathsf{I}(X; C \mid W) = \mathsf{H}(C \mid W) - \mathsf{H}(C \mid X, W)$. Replacing a root's input does not change any outside factors or noise laws by the stated intervention rule. □

For an intervention at a nonroot vertex, the intervention semantics must instead specify replacement of its kernel and which exogenous data are held fixed. Conditional probability and intervention can disagree when a hidden ancestor is shared. A simple example takes independent fair U, V, sets $X = U$, and $Y = U$ with no signal from X to Y. Observationally $X = Y$ with probability one. Under an intervention replacing X by V, the two are independent. In the theorem one may include U in W; then $\mathsf{H}(X \mid W) = 0$ for the observational law.

Write $m = |L| \ge 2$ for the number of

possible labels. For $e \in [0,1]$ put

$$f_m(e) = h_b(e) + e \log(m-1),$$
$$h_b(e) = -e \log e - (1-e)\log(1-e).$$

As usual $0 \log 0 = 0$. This form applies to the actual error. When given only an upper bound e on error, the maximum of f_m over $[0, e]$ is $f_m(\min(e, 1 - 1/m))$.

Theorem 4.2 (Cut capacity for recoverability)**.** *If Y, W admit a decoder of $X \in L$ of error probability e, then*

$$\begin{aligned} \mathsf{H}(X \mid W) - f_m(e) &\le \mathsf{I}(X; C \mid W) \\ &\le \mathsf{H}(C \mid W) \\ &\le \sum_{a \in \mathcal{A}} \log |A_a|. \end{aligned} \tag{4.2}$$

where $\mathcal{A}$ indexes the transmitted cut symbols, with finite alphabets A_a. For zero error one has $\mathsf{H}(X \mid C, W) = 0$.

Proof. For deterministic decoding, append the error indicator $B = 1_{\{X \neq d(Y,W)\}}$. It is a function of X, Y, W. Then

$$\begin{aligned} \mathsf{H}(X \mid Y, W) &= \mathsf{H}(B \mid Y, W) \\ &\quad + \mathsf{H}(X \mid B, Y, W) \\ &\le h_b(e) + e \log(m-1). \end{aligned}$$

The second term is zero on success, and on failure at most $m - 1$ labels remain. If the decoder uses independent extra randomness, include the random seed in Y, preserving $\mathsf{I}(X; Y \mid W)$ and the factorization. Now apply theorem 4.1 and the bound by the logarithm of the cardinality of the cut alphabet. At zero error the first bound reaches $\mathsf{H}(X \mid W)$, forcing equality in the conditional mutual information identity. □

4.2 Reachability, multiplicity, and cost of a wire

If X is the only source input and there is no directed path from it to an observer Y, take V_- to contain X and its ancestors, but evaluate only the ancestors of Y outside the ideal. None of the wires leaving the ideal and entering the ancestors of Y can carry any dependence on X through a directed path. For an exogenous X independent of other roots conditional on W, an equivalent proof is to restrict the graph to ancestors of Y; this removes X and keeps a law independent of it given W. In either case, optimal success based on Y, W is at most $\sum_w p(w) \max_x p(x \mid w)$. If a receiver is outside the causal future in the placement, transitivity of $\preceq$ rules out a directed path from X.

Conversely, suppose every terminal is reachable from the source, every chosen edge can carry the full alphabet L, and deterministic local copying is available. Select for every vertex in the reachable subgraph except the source a predecessor on a path from the source. The selected edges form an arborescence rooted at the source because their predecessor chains must terminate and reach the source. Retain the finite subtree spanning the required terminals. Forwarding the label on each selected edge gives exact local decoding for every input. A shared upstream edge can serve arbitrarily many downstream labels so long as its one alphabet carries the source label. Thus a bound requiring $r\mathsf{H}(X)$ to cross a common upstream cut would be false for r recipients.

If the cut has d symbols of at most q values each, an exact, uniform m-label problem without side information requires $d \log q \ge \log m$. In a series of channels for which each stage is a cut, this inequality holds for every stage. In a graph with multiple receivers, each receiver has its own family of cuts. A favorable count on one cut does not suffice in the presence of noisy channels or a restriction on allowable coding. The full network coding theorem addresses a different asymptotic throughput problem with coding over blocks [2].

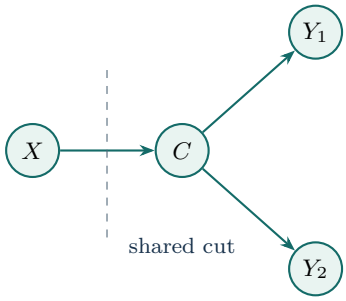


Figure 4.1: One transmitted symbol can carry the full source entropy and subsequently feed two separate decoders. The joint readout carries redundancy across its carriers.

4.3 From circuit traces to composition

A deterministic gate $g : A \to B$ presents the span $A \leftarrow A \to B$. A pipeline g, h has trace $A \leftarrow A \times_B B \to C$, with the left projection an isomorphism and the right map $h \circ g$. For nondeterministic support, the pullback retains every compatible intermediate witness. For a stochastic computation, summing those witnesses with product weights gives composition in **FinStoch**, the category of finite sets and kernels. Identity kernels are point masses and tensor products are independent parallel composition. Conditional independence is a property of the law or of its factorization; a correlated joint input can yield correlated outputs even for a tensor of independent kernels.

Independent finite circuits can be evaluated in any interleaving. More generally, for a process on a fixed product state, specify read and write sets R_v, W_v at each gate, with $W_v \subseteq R_v$ if a gate may consult the previous value of its own target, and otherwise allow $R_v \cup W_v$ as its scope. Two gates commute whenever their write sets are disjoint and neither reads the other's writes. Their independent conditional randomness then yields the same composite kernel in either order. This condition permits the gates to read a common immutable label. If these conditions hold for every incomparable pair of events, the kernel obtained by composing the gates is independent of the linear extension. The proof uses the fact that any two linear extensions of a finite partial order are related by swaps of adjacent incomparable elements. To see this, move the first element of the desired order leftward past every element preceding it in the current order. None of those elements can precede it in the poset, and all swaps preserve a linear extension. Repeat on the remainder.

4.4 Recoverability is a contract on readout interfaces

A family of possibly overlapping observation tuples can be used for decoding. An erasure code specifies a family of authorized subsets S, and for each a decoder from $(Y_i)_{i \in S}$. This is different from a family in which each of r disjoint carriers has a decoder. For example, let X and R be independent fair bits. Give R to one carrier and $X \oplus R$ to another. Either carrier alone is independent of X, and the pair determines it. The terminal pair is independent marginally, so its total correlation is zero. The theorem in chapter 7 applies to individually decoding readouts. When applied to two overlapping authorized subsets, common coordinates can themselves be the source of correlation; thermodynamic additivity across those subsets then requires a separate boundary model.

If X is retained at the source, it can be counted as one of the final decodable carriers. This changes the replica multiplicity in the entropy calculation, but need not add a wire or a receiver. For the energetic result the full final state may include X even if the specification names only the new receivers.

We distinguish these conventions by giving separate expressions for r receivers and $r+1$ individually decoding final registers.

CHAPTER 5

Logical and effective semantics

The same contract can be read in the internal language of a topos and executed on a finite graph. The relation between these readings is clearest when the observable label object is 0-truncated, the realizations may have higher identifications, and the verifier reads an explicitly finite part of a realization.

5.1 A proposition, a chosen witness, and a uniform section

Let $p : R \to B$ be the object of allowed histories over reference data, incorporating an observation and a success predicate. In a slice over B, local existence of a witness is the (-1)-truncation of R in that slice, equivalently its image subobject. A specified global lift over $b : U \to B$ is a point of $\mathrm{Map}_{\mathcal{E}_{/B}}(U, R)$. A choice for every reference situation, coherent in parameters, is a section $B \to R$ in that slice. An effective epimorphism $R \to B$ asserts local inhabitance everywhere. It can be nonsplit. For a covering with monodromy, a section would produce a compatible choice of sheet along every loop. The double cover from chapter 3 demonstrates the gap between local existence and uniform choice.

For finite labels, the copy predicate on chosen readouts is a finite conjunction of equalities. Equality of two label maps has a subobject as equalizer. An internal statement that a readout is correct can be pulled back along arbitrary parameter change. If each local model restricts correctly, the equalities glue because maps into a sheaf satisfy descent. If one has a locally existing readout with no globally chosen point, the existence predicate may hold globally in the internal sense while the space of global witnesses is empty. Evaluating the external space of witnesses distinguishes the two statements.

Existence in a finite circuit with a finite search space is decidable by exhaustive search. In constructive logic this uses decidable finite alphabets and decidable membership in every local constraint. The global existence question for a history sheaf on an arbitrary site has no such general algorithm. In particular, restricting to a finite truncation of the observation graph does not make an infinite persistence quantifier decidable.

5.2 Classical copying and the tensor of kernels

For finite sets, the diagonal $\Delta_A : A \to A \times A$ and deletion $A \to \mathbf{1}$ form a commutative comonoid. Both are natural for deterministic maps. In **FinStoch**, deletion stays natural because each conditional output law sums to one. The diagonal is a distinguished copying map on each classical alphabet, but a general stochastic transformation is not a homomorphism for it.

Proposition 5.1 (Exactly the deterministic kernels preserve copying)**.** *A stochastic kernel $K : A \rightsquigarrow B$ satisfies*

$$\Delta_B \circ K = (K \otimes K) \circ \Delta_A$$

if and only if $K(\cdot \mid a)$ is a point mass for every a.

Proof. At input a, the left law assigns zero probability to every off-diagonal pair in B^2. The right law assigns $K(b \mid a)K(b' \mid a)$ to (b, b'). Equality requires each pair of distinct

entries to have zero product. Every row is normalized, so exactly one entry is one. For point mass rows the two compositions give the same diagonal point mass. □

Running a randomized computation twice and copying one result therefore express different histories. They can give the same output marginals. If K ignores its input and returns a fair bit, the two output entropies are $\log 2$ for copy-after-sampling and $2 \log 2$ for independent sampling-after-copy. The order of these operations is visible in joint observations even when hidden from local readouts. Markov categories formalize this interaction of copying, deletion, and probabilistic transformations [13].

5.3 Reversible realization of finite deterministic computation

A finite function $g : A \to B$ need not be injective. A reversible realization keeps a register A, represents B in a finite abelian group G of sufficient cardinality, and chooses the permutation

$$(a, z) \longmapsto (a, z + g(a)). \tag{5.1}$$

Its inverse subtracts the same value. On a prepared zero target it produces $g(a)$ while retaining a. For a circuit with gates g_j, place one fresh register per intermediate result and apply controlled addition at each gate. Copy the required final result to a designated output. Then reverse the gate sequence to clear the scratch registers. The reverse is valid because the necessary control registers remain until their own reversals are applied. This realizes the output relation without logically erasing the inputs or leaving scratch data, using initialized registers and additional time [4].

The abstract permutation for a many-input gate can be causally placed only where those inputs can be jointly accessed. The uncomputation sequence requires later communication to return intermediate results to their registers. A purely formal inverse of a channel is not a future-directed trajectory reversing the ambient causal order. To implement the inverse logic, build a new set of future-directed events that reverse the order of the gate functions after the output has been copied. If the original hardware cannot route the required information, reversible realization in that hardware can fail even though (5.1) is a permutation on the declared alphabets.

Corollary 5.2 (Reversible routing of an exact label)**.** *Suppose each internal edge of a finite fan-out tree can transport a group register and every target has a zero blank. With an input label in the group register at the root and sufficient causal access for the controlled additions, exact fan-out is a permutation on the modeled array restricted to the prepared boundary. Scratch cleanup is possible whenever the reversed logical circuit is also causally placeable.*

For a degenerate energy function every permutation preserves Gibbs equilibrium and all finite-state entropies. This yields a zero-drive-work idealization of the computation. Treating initially pure blanks as ordinary equilibrium blanks would change the initial law. The proof of the work inequality later uses the actual ensemble.

5.4 Decision procedures for finite realizations

There are separate finite decision problems. Reachability of an event from a source is graph search. Given a deterministic circuit and a set of allowed inputs, whether it meets an exact decoding contract can

be tested by enumerating the input and comparing observed labels. Given a rational stochastic circuit, exact probabilities of output events are rational and computable by enumerating gate assignments or eliminating latent variables. A cut bound is an efficiently checkable necessary test if one knows the alphabets and source cardinality, but it need not decide the probabilistic decoding problem.

Suppose that a horizon n gives a decidable finite set R_n and computable restrictions $R_{n+1} \to R_n$. By the compactness theorem below, the existence of an infinite continuation is equivalent to nonemptiness of R_n for every n. Thus on this effective presentation it is a Π^0_1 predicate. One can search for an empty finite level as a certificate of failure. In the general class of such presentations there is no decision procedure.

Theorem 5.3 (Undecidability of unbounded finite-horizon witnesses)**.** *For towers of finite decidable sets presented by primitive-recursive level generators and restriction maps, asking whether the inverse limit is nonempty is Π^0_1-complete, even when each level has at most one element. For general computable presentations the same classification applies with totality of the presentation as a promise.*

Proof. The membership statement follows from finite compactness and the universal quantification over the computably decidable levels. For hardness, given a program P and its input, define $R_n = \{*\}$ if P has not halted within n steps, and $R_n = \varnothing$ otherwise. These tests are computable. If the later set is nonempty the earlier set is also nonempty, so the restriction is always well-defined, either unique between singletons or with empty domain. A continuation exists exactly when P never halts. The map from a program to this tower presentation is effective. □

This theorem concerns a uniformly indexed presentation, which can perform more simulation as n increases. A single fixed finite transition graph with an initial vertex and an allowed-state predicate has a decidable infinite-path problem. An infinite allowed path eventually repeats a vertex, and a reachable allowed cycle produces an infinite path. Thus one must specify whether the local mechanism itself is bounded-state or only each finite observation has a finite witness set.

5.5 Finite existence can differ from effective choice

There are computable finitely branching trees with paths and no computable infinite path. One direct construction starts from an effective enumeration (φ_e) of partial computable integer functions. A binary string s of length n is allowed when for every $e < n$, if the computation $\varphi_e(e)$ halts within n steps with a bit as output, then s_e differs from that bit. This predicate is decidable. It is closed under prefixes because a constraint observed by time $m < n$ is still observed by time n. Every finite level is nonempty because each relevant coordinate forbids at most one bit. An infinite path must disagree with the diagonal value at every coordinate where that value is a bit. A computable infinite path would be the total binary function of some φ_e, contradicting the required disagreement at e.

The presentation yields finite reconstruction sets with a coherent path classically by compactness, but the proof does not furnish an algorithm to produce a path. In a realizability interpretation such as the effective topos [16], a uniform realizer for a binary sequence is additional structure. The present result identifies an obstruction to such a realizer. It uses classical compactness for existence in the external set model.

5.6 An interface test for logical refinement

A refinement replaces a gate by a finite subnetwork with the same visible boundary kernel. Let the surrounding graph see the refinement only through those boundary variables, and let the refinement randomness be independent of other gate seeds conditional on its inputs. Then summing over its hidden internal states yields the original factor. Hence every outer observable law is preserved, along with entropies depending only on those readouts. The state spaces of histories before and after refinement can differ, and energetic summaries can differ if the refinement adds baths, ancillary states or different energies. This separation lets one prove functional correctness through a change of implementation and still account for its physical resource use.

CHAPTER 6

Persistence over increasing horizons

A persistence question specifies which earlier observations are fixed and how horizons are related. Let $U_0 \subseteq U_1 \subseteq \cdots$ be increasing regions and let b_n be coherent observations on U_n. A history at U_n may have to satisfy success predicates at several earlier screens. Assume these predicates are chosen so that a successful history restricts to a successful history. This gives a tower of spaces R_n, which can be the reconstruction spaces in (3.1) or their indicated subspaces. Its inverse limit in spaces is the space of persistent histories.

When $U = \bigcup_n U_n$, a sheaf evaluates on U as the limit over the increasing regions. One way to see this is that the sheafified representable for U is the union of the corresponding subobjects. Each finite intersection of U_n is the smallest region in it, and the cover descent diagram is equivalent for purposes of the limit to the tower of restrictions. If the observations glue to b on U, the reconstruction at b is the limit of the finite-horizon reconstructions. Additional conditions defining R_n simply specify which histories in that limit are allowed.

6.1 Compactness for extensions

Theorem 6.1 (Compact finite-horizon principle)**.** *Let R_n be nonempty compact Hausdorff spaces with continuous restriction maps $\rho_n : R_{n+1} \to R_n$. Their ordinary inverse limit is nonempty. In particular the result applies to nonempty finite sets. The same proof gives an extension of a fixed earlier witness if its descendant sets at every later horizon are nonempty.*

Proof. The product of the R_n is compact. Each equation $r_n = \rho_n(r_{n+1})$ cuts out a closed subset since the diagonal of R_n is closed. Any finite collection of these equations can be solved by choosing a point at a sufficiently late level and restricting it down to the earlier levels. Other coordinates can be arbitrary. The family of closed sets has the finite intersection property, so their intersection is nonempty. For a fixed earlier witness impose the additional closed equation and take the late choice in its descendant set. □

For finite sets this argument is also König's infinity lemma. At level n choose a witness with descendants at arbitrarily late levels, which is possible because the level is finite and the later levels are nonempty. Among the finitely many immediate extensions at least one also has arbitrarily late descendants. Repeating yields a compatible sequence. A time-indexed construction can have its early choices destroyed by later

conditions. The compactness proof selects choices with unbounded extension and does not rely on the surjectivity of the bonding maps.

If the allowed root witness r_0 is fixed, compactness of the descendant fibers is automatic under the stated hypotheses: the fibers of the composite $\rho_0 \cdots \rho_{n-1}$ over r_0 are closed. If the initial condition is a compact subset of R_0, the same argument allows the root to be selected in it. Compactness matters even for a single countable tower. The sets $R_n = \{m \in \mathbb{N} : m \geq n\}$ with inclusion restrictions are all nonempty and have empty inverse limit. The integer m can encode a finite lifetime chosen at creation. Existence of some device for every requested lifetime differs from a compatible choice of one device for all lifetimes.

6.2 A higher version and a component obstruction

Theorem 6.2 (Nonempty homotopy limit from compact components)**.** *Let $R_\bullet$ be a countable tower of spaces. Suppose each $\pi_0 R_n$ admits a nonempty compact Hausdorff topology for which the induced maps are continuous. Then* $\operatorname{holim}_n R_n$ *is nonempty. More generally, for a tower of spaces a compatible sequence of components suffices for nonemptiness.*

Proof. By theorem 6.1, the first hypothesis gives compatible components. Choose one point r_n in each component. Its compatibility supplies a path in R_n from r_n to $\rho_n(r_{n+1})$. A homotopy limit model for a tower has points given by precisely these pairs of sequences of points and paths. For example, it is the homotopy equalizer of identity and shift on $\prod_n R_n$, the shift taking (r_n) to $(\rho_n(r_{n+1}))$. The model follows by replacing arrows with path-space fibrations, or from the fact that the indexing category is freely generated by the consecutive arrows. It gives a point of the homotopy limit. □

The compact component criterion does not give uniqueness. An explicit example takes $R_n = K(\mathbb{Z}, 1)$ and each bonding map induced by multiplication by two. There is a compatible sequence of basepoints and every $\pi_0 R_n$ is a singleton. Use the homotopy equalizer model and choose every r_n to be the basepoint. The paths can then be loops, with classes $b_n \in \mathbb{Z}$. Every component of the homotopy limit has a representative of this form because each R_n is connected. Moving between two such representatives gives a sequence of loops with classes a_n, one at each coordinate. The square formed by the two comparison paths and the coordinate loops commutes up to homotopy exactly when the difference of the comparison classes equals $a_n - 2a_{n+1}$, up to the overall sign chosen for the path orientation. Conversely this equality gives a path between the representatives by filling those squares. Thus the component set is the quotient of $\prod_n \mathbb{Z}$ by the image of

$$(a_n)_{n\geq 0} \longmapsto (a_n - 2a_{n+1})_{n\geq 0}.$$

This is the usual abelian derived-limit description for this tower.

Take $b_{2j} = 1$ and $b_{2j+1} = 0$. If $a_n - 2a_{n+1} = b_n$ for all n, then iteration to $n = 2k$ would give

$$a_0 = 4^k a_{2k} + \sum_{j=0}^{k-1} 4^j = 4^k a_{2k} + \frac{4^k - 1}{3}.$$

Hence $3a_0+1$ would be divisible by every 4^k, which forces $3a_0 + 1 = 0$ and is impossible over $\mathbb{Z}$. The nonzero class becomes trivial on each finite initial part of the tower: prescribe the last $a_n = 0$ and solve backwards. Every finite comparison can thus be made with the base history, but there is no single compatible infinite comparison.

A separate failure of existence comes from taking $R_n = \mathbb{Z}_{\mathrm{disc}}$ with affine restrictions $\rho_n(a) = 2a + b_n$. The same computation shows that no compatible points exist even though every finite truncation has a solution. The groupoid example has a persistent history and more than one component. The affine example has no persistent history. Together they exhibit the roles of components and coherent identifications.

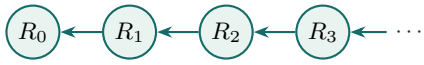


Figure 6.1: Persistence uses restrictions toward the past. Choosing physical transitions forward and finding a compatible limit of observations are different operations.

6.3 Persistence as a probability

Let a single probability law on trajectories be given. Write A_n for the event that the contract has held through horizon n, including every earlier requirement. Then $A_{n+1} \subseteq A_n$. Countable additivity gives

$$\mathbb{P}\left(\bigcap_{n\geq 0} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

This follows by applying continuity from below to the increasing complements. In a countable finite-state model, normalized compatible finite-horizon laws determine a path law on the product by the usual cylinder construction. An implementation with a specified initial law and transition kernels gives such compatible laws. Arbitrary local marginals on overlapping regions can fail to do so, as discussed in chapter 8.

When $\mathbb{P}(A_0) > 0$ and the previous survival probabilities are positive, let $q_j = \mathbb{P}(A_{j+1}^c \mid A_j)$. The chain rule gives

$$\mathbb{P}(A_n) = \mathbb{P}(A_0) \prod_{j=0}^{n-1} (1 - q_j).$$

If any $q_j = 1$, all later probabilities vanish. If $0 \leq q_j < 1$ for every j, survival forever has positive probability exactly when $\sum_j q_j < \infty$. Indeed, $-\log(1-q) \geq q$, and for $q \leq 1/2$, $-\log(1-q) \leq 2q$. A finite sum implies only finitely many terms exceed $1/2$. The corresponding product convergence follows. This statement uses conditional hazards and requires no independence across time.

An endpoint decoder at horizon n asks whether its current estimate equals the source. A device can fail at one screen and be corrected at a later screen. Then the endpoint events need not be nested. A fixed bound $\mathbb{P}(\text{endpoint error at } n) \leq e$ is compatible with eventual failure somewhere with probability one. If independent per-round errors each have probability $e > 0$, the probability of no error through n independent rounds is $(1 - e)^n$. A union bound gives a sufficient condition for persistent accuracy when there is a summable sequence of per-round error bounds: the probability of any error is at most their sum.

6.4 A killed Markov model

Let P be the transition matrix of a finite-state memory with allowed set A. Treat the first exit from A as failure. The submatrix $Q = (P_{ab})_{a,b\in A}$ is substochastic. For initial row law p supported in A, survival through n transitions is

$$s_n = pQ^n\mathbf{1}.$$

This follows by summing the probabilities of paths that stay in A and using matrix multiplication.

If for some integer k and $\delta > 0$ every state of A has probability at least δ of exit in at most k steps, then $Q^k\mathbf{1} \le (1-\delta)\mathbf{1}$. Iteration gives $s_n \le (1-\delta)^{\lfloor n/k \rfloor}$. In a finite model the exit hypothesis holds if each state can reach the complement along a positive-probability path. Choose finitely many such paths and take a maximum length and a minimum positive path probability.

If q is a probability row vector with $qQ = \alpha q$ and $0 < \alpha < 1$, then $s_n = \alpha^n$ for initial law q, and the conditional law given survival remains q. Such a quasi-stationary state has a time-independent conditional law and finite mean remaining lifetime $1/(1-\alpha)$ when counted in steps to exit. It illustrates the difference between stability of a conditional ensemble and persistence of the event being conditioned on [9].

At a deterministic boundary, the corresponding support may have a cycle entirely in A. Thus a possible infinitely persistent trajectory can exist with probability zero even when every starting state almost surely eventually exits. For one state with a self-loop of probability $1-\delta$ and an exit of probability δ, the all-loop path is a global witness but has measure zero. The distinction survives the passage from finite relations to higher reconstruction spaces.

CHAPTER 7

Entropy of jointly decodable records

All variables in this chapter are finite. Logarithms are natural. For a law p, write $\mathsf{H}(p) = -\sum_z p(z)\log p(z)$, and write $\mathsf{H}(Z \mid W) = \sum_w p(w)\mathsf{H}(p_{Z|w})$. The relative entropy is

$$\mathsf{D}(p \,\|\, q) = \sum_{z:p(z)>0} p(z)\log\frac{p(z)}{q(z)}$$

with value $+\infty$ if any positive $p(z)$ has $q(z) = 0$. Conditional quantities ignore zero-probability values of the conditioning variable. Mutual information is $\mathsf{I}(X;Y \mid W) = \mathsf{H}(X \mid W) - \mathsf{H}(X \mid Y,W)$. This equals the conditional average divergence of the joint law from the product of its marginals.

7.1 Relative entropy and stochastic transformations

For $p \ll q$ and a finite stochastic kernel K, consider joint laws $p(x)K(y \mid x)$ and $q(x)K(y \mid x)$. The chain rule for divergence gives

$$\mathsf{D}(p \,\|\, q) - \mathsf{D}(pK \,\|\, qK) = \sum_y (pK)(y)\mathsf{D}(p(\cdot \mid y) \,\|\, q(\cdot \mid y))\,. \tag{7.1}$$

The forward conditionals cancel in the expansion in x first. In the expansion in y first they give the displayed difference. Each divergence on the right is nonnegative by the log-sum inequality. Equality holds if and only if the posterior laws agree for every y with $(pK)(y) > 0$. If q has full support, its Bayesian reverse kernel at any y with $(qK)(y) > 0$ therefore recovers p from pK when equality holds. It always recovers q from qK.

The same computation shows that the information about a source decreases when a readout is passed through a channel with no access to the source conditional on the readout and the declared context. More explicitly, if $p(x, z \mid y, w) = p(x \mid y, w)p(z \mid y, w)$, then

$$\mathsf{I}(X; Y \mid W) - \mathsf{I}(X; Z \mid W) = \mathsf{I}(X; Y \mid Z, W).$$

This difference can be positive even if the channel preserves the marginal entropy of Y. For a fair input bit and a binary symmetric channel with crossover e, the marginal output stays fair and the information drops by $h_{\mathrm{b}}(e)$.

7.2 The replica identity

For $Y = (Y_1, \dots, Y_r)$ define its conditional total correlation by

$$\begin{aligned} \mathsf{TC}(Y \mid W) &= \sum_{i=1}^{r} \mathsf{H}(Y_i \mid W) - \mathsf{H}(Y \mid W) \\ &= \sum_{w} p(w) \mathsf{D}\left(p_{Y \mid w} \,\middle\|\, \prod_i p_{Y_i \mid w} \right). \end{aligned} \tag{7.2}$$

The second formula follows by expanding the logarithm of the product and marginalizing. Total correlation is nonnegative. It is zero exactly when the carriers are conditionally independent given W. It equals the sum of $\mathsf{I}(Y_i; Y_{<i} \mid W)$ in any order of the carriers. This description is useful for adding one more carrier, even though the total does not depend on the ordering.

Theorem 7.1 (Correlation required for local decoding)**.** *Let X take values in an alphabet of size $m \geq 2$. For any joint law of $X, W, Y_1, \dots, Y_r$,*

$$\begin{aligned} \mathsf{TC}(Y \mid W) = (r-1)\mathsf{H}(X \mid W) - \sum_i \mathsf{H}(X \mid Y_i, W) \\ + \mathsf{H}(X \mid Y, W) + \mathsf{TC}(Y \mid X, W). \end{aligned} \tag{7.3}$$

If for each i the pair Y_i, W admits a decoder with error e_i, then

$$\mathsf{TC}(Y \mid W) \geq \max\left\{ 0, (r-1)\mathsf{H}(X \mid W) - \sum_i f_m(e_i) \right\}. \tag{7.4}$$

If all decoders are exact, then

$$\mathsf{TC}(Y \mid W) = (r-1)\mathsf{H}(X \mid W) + \mathsf{TC}(Y \mid X, W). \tag{7.5}$$

For $r = 1$, the identity holds with both total correlations zero.

Proof. Expand each $\mathsf{H}(Y_i \mid W)$ into $\mathsf{H}(Y_i \mid X, W) + \mathsf{I}(X; Y_i \mid W)$, and expand $\mathsf{H}(Y \mid W)$ in the same way. Their difference is $\mathsf{TC}(Y \mid X, W) + \sum_i \mathsf{I}(X; Y_i \mid W) - \mathsf{I}(X; Y \mid W)$. Insert the entropy formulas for the mutual informations. For approximate decoding use conditional Fano from theorem 4.2 and nonnegativity of the remaining terms. The Fano proof itself requires only the joint law, independent of any cut. Exact decoding makes all conditional entropies of X given the relevant readouts zero. □

This identity assigns a precise role to ancillary correlations between carriers. Suppose $Y_i = (X, R_i)$ and the R_i are conditionally independent given X, W. The exact lower bound is attained. If all R_i equal a further fair bit R independent of X, W, the excess is $(r-1)\log 2$. There are the same individual decoders in both cases. The excess counts common information carried by the joint histories that does not contribute to deciding X.

If a separate bound B on $\mathsf{I}(X; Y \mid W)$ is available, then

$$\mathsf{TC}(Y \mid W) \geq \sum_i \big(\mathsf{H}(X \mid W) - f_m(e_i)\big)_+ - B, \tag{7.6}$$

where $u_+ = \max(u, 0)$. For a common causal cut, $B = \mathsf{I}(X; C \mid W)$ is valid. Formula (7.6) can use zero for a receiver with a loose error guarantee, and uses a measured upper bound on joint source information when one is known. In particular, many receivers whose individual information is close to the cut information must have substantial total correlation. The cut limits unique source information, while the correlation term measures redundancy among carrier laws.

7.3 Including the source in the final array

If X is retained, let $Y^+ = (X, Y_1, \ldots, Y_r)$. A direct expansion yields

$$\mathsf{TC}(Y^+ \mid W) = r\mathsf{H}(X \mid W) - \sum_i \mathsf{H}(X \mid Y_i, W) + \mathsf{TC}(Y \mid X, W). \tag{7.7}$$

Here the tuple Y^+ knows X exactly even when every target decoder has error. In the formula r denotes the number of *new* receivers. If conditional independence holds given X, W, the retained total correlation is exactly $\sum_i \mathsf{I}(X; Y_i \mid W)$. The entropy of the final array satisfies $\mathsf{H}(Y^+ \mid W) = \mathsf{H}(X \mid W) + \mathsf{H}(Y \mid X, W)$.

These formulas distinguish the total entropy of the records from the sum of their separate entropies.

For independent symmetric errors $Y_i = X \oplus N_i$ with fair X, and independent N_i of crossover e_i, each Y_i is marginally fair. Thus the retained total correlation is $\sum_i(\log 2 - h_\mathrm{b}(e_i))$. Dropping X removes the amount $\mathsf{I}(X;Y)$, which need not be small even with many receivers. For $r = 2$ and equal e, the difference bit $Y_1 \oplus Y_2$ is independent of Y_1 and has crossover $2e(1-e)$. The unretained total correlation is

$$\mathsf{TC}(Y_1, Y_2) = \log 2 - h_\mathrm{b}(2e(1-e)).$$

Comparing this with the retained value gives a numerical distinction between the two accounting boundaries.

7.4 Successful decoding as a coupling

The decoded tuple $\widehat{X} = (d_1(Y_1, W), \ldots, d_r(Y_r, W))$ has an actual coupling with the diagonal tuple $(X, \ldots, X)$ on the same trial. The probability that the tuples differ is at most $\sum_i e_i$. For any event in L^r, the difference of the two probabilities is at most this disagreement probability. Therefore

$$\left\| p_{\widehat{X}} - p_{(X,\ldots,X)} \right\|_{\mathrm{TV}} \leq \min\{1, \textstyle\sum_i e_i\}.$$

There is also a conditional statement at each w, using the conditional errors. A correlation inequality follows by applying data processing to the deterministic local decoders conditional on $W = w$: the joint law and its product of marginals are mapped to the decoded joint law and its product of marginals. Consequently $\mathsf{TC}(\widehat{X} \mid W) \leq \mathsf{TC}(Y \mid W)$. Equality may hold when each decoder loses only noise independent across carriers conditional on the decoded labels, with the corresponding conditional law preserved by the reverse channel for the product reference.

Pairwise independence is weaker than mutual independence. If U, V are independent fair bits and $Y = (U, V, U \oplus V)$, all pairs are independent while the total correlation is $\log 2$. No nonconstant label is exactly decodable from all three individual carriers: if it is a function of U and also of V, independence forces it to be constant. Thus positive total correlation alone is insufficient to certify replication of a specified label. The contract and the source coupling are essential.

7.5 Erasure and conditional information

For an erasure channel with erasure indicator E independent of X, a receiver obtains X on success and a distinct erasure symbol on failure. If $\mathbb{P}(E = 1) = e$, the posterior uncertainty is $e\mathsf{H}(X)$ and the mutual information is $(1 - e)\mathsf{H}(X)$. If one uses an error indicator for a forced guess rather than allowing erasure, the error and the Fano bound change. The work accounting of an erasure-flag carrier also uses the actual alphabet and its energy model, which can differ from those for an ordinary binary bit.

If W contains a perfect copy of X, all decoders can output W without inspecting their carriers. Then $\mathsf{H}(X \mid W) = 0$ and there is no required new conditional redundancy. If W is a noisy estimate, $\mathsf{H}(X \mid W)$ specifies the residual information. The effect on free energy depends on whether the physical registers constituting W are retained, used catalytically, or treated as part of an external experiment description.

CHAPTER 8

Marginal descent and entropy of gluing

A probability law on a set of histories induces laws on every observation. The converse is an extension problem. Let V be a finite set of variables, with finite alphabets A_v, and let $\mathcal{U} = (U_i)_{i \in I}$ be a cover of V by scopes. Write p_i for a law on the variables in U_i. Call the data compatible when p_i and p_j have the same marginal on $U_i \cap U_j$ for every i, j. A coupling is a law p on the union whose scope marginals are all the p_i. If the underlying history set is constrained, one also requires its support condition. The extension problem on the full product already has obstructions.

Disjoint scopes give an elementary failure of uniqueness. Two fair binary marginals have a product coupling, a diagonal coupling, and an antidiagonal coupling. They agree on the empty overlap. Thus the presheaf sending a scope to the set of all probability laws on its variables, with marginal restriction, need not be a sheaf. The support history sheaf still satisfies descent. Its sections have a different type from distributions on those sections. A choice of stochastic gluing can impose conditional independence at interfaces.

8.1 Gluing along one separator

Let A, S, B be disjoint tuples, and let p_{AS} and p_{SB} agree on S, with separator marginal p_S. Define

$$q(a, s, b) = p_S(s)p_{A|S}(a \mid s)p_{B|S}(b \mid s)$$

on $p_S(s) > 0$, and zero otherwise. The output has the two required marginals by summing out B or A. This is the unique coupling with A and B conditionally independent given S. If p is any coupling with the same marginals, then $p \ll q$: positive mass at (a, s, b) gives positive mass to each projected event. Moreover,

$$\mathsf{D}(p \,\|\, q) = \mathsf{I}_p(A; B \mid S) = \mathsf{H}(p_{AS}) + \mathsf{H}(p_{SB}) - \mathsf{H}(p_S) - \mathsf{H}(p).$$

The first equality follows directly from the conditional product, the second from the entropy chain rule. It proves that q uniquely maximizes entropy. This is an exact deficit, including when some separator values are never observed.

For example, let S be a fair bit, with $A = S \oplus N_a$ and $B = S \oplus N_b$, and prescribe both crossover probabilities as e. If the noises are independently drawn given S, the joint entropy is $\log 2 + 2h_{\mathrm{b}}(e)$. If the same noise is used for both, the joint entropy is $\log 2 + h_{\mathrm{b}}(e)$. Their prescribed marginals agree. The loss of $h_{\mathrm{b}}(e)$ in the second coupling is the conditional dependence of A and B given S. If S is physically unavailable as a conditioning register, conditional independence used in this mathematical gluing rule still describes the law, but does not supply a mechanism to sample it.

8.2 The tree extension theorem

A tree T with nodes indexed by the scopes U_i has the *running intersection* property if the nodes whose scope contains any fixed variable form a connected subtree. For an edge ij, let $S_{ij} = U_i \cap U_j$. If the edge is deleted, the two unions of scopes on its sides intersect precisely in S_{ij}. Indeed, a variable occurring on both sides belongs to every scope along the path between the two occurrences.

Theorem 8.1 (Separator gluing)**.** *Suppose the scope tree has running intersection and the laws agree on each edge separator. Choose a root i_0, orient the edges away from it, and let $j(i)$ be the parent of $i \neq i_0$. There is a coupling q given by*

$$q(x) = p_{i_0}(x_{U_{i_0}}) \prod_{i \neq i_0} p_i(x_{U_i \setminus S_{i,j(i)}} \mid x_{S_{i,j(i)}}). \tag{8.1}$$

A separator value of zero prescribed mass may be given any normalized conditional row in this formula. The resulting joint law is independent of those choices and

of the root. It is the only coupling that is conditionally independent across the two sides of each edge given the separator.

Proof. Sample the root and then visit each child after its parent. When visiting a child, its overlap with the variables already sampled is exactly its separator. Inclusion of the separator follows because the parent has been sampled. A variable in the child and any already sampled scope lies in the parent by running intersection. Sample the remaining variables using the conditional row in (8.1). The separator has its prescribed law by induction. Thus each child has law p_i, and zero-probability separator rows are never used. Repeating gives a normalized joint law with all required marginals.

On assignments where all separator masses are positive, the formula is $\prod_i p_i(x_{U_i}) / \prod_{ij \in T} p_{S_{ij}}(x_{S_{ij}})$. If a separator value has zero mass, the resulting joint value is zero because that separator has the prescribed marginal. This proves root independence. Root the tree on either side of a chosen edge. The generative description shows that the child side, conditional on the parent side, is sampled using only the separator and fresh randomness. For uniqueness, prune a leaf. Conditional independence across the leaf edge forces its conditional law given the rest to be the prescribed leaf conditional on the separator. The marginal on the remaining tree still has the other edge independence properties, since each such property is stable under discarding variables on either side. Induct over the remaining tree. □



Figure 8.1: In this scope tree each variable occurs on a connected set of nodes. Each edge contributes one separator law.

Theorem 8.2 (Exact junction entropy deficit)**.** *For the coupling q of theorem 8.1 and every coupling p with the specified marginals,*

$$\mathsf{H}(q) = \sum_i \mathsf{H}(p_i) - \sum_{ij \in T} \mathsf{H}(p_{S_{ij}}), \tag{8.2}$$

$$\mathsf{H}(q) - \mathsf{H}(p) = \mathsf{D}(p \,\|\, q)\,. \tag{8.3}$$

Consequently q uniquely maximizes entropy on the set of couplings. If an energy is a sum of functions supported on the scopes, then q uniquely minimizes nonequilibrium free energy over the same set at $T > 0$, and the free-energy excess of p is $k_{\mathrm{B}} T \mathsf{D}(p \,\|\, q)$.

Proof. For a positive joint mass of p, all clique and separator masses at the projected values are positive, so $p \ll q$. On this support, expand the logarithm of the product divided by the separator masses. The expectation of $\log q$ is the same under p and q because every term is fixed by a scope or separator marginal. With q as the law, the expansion gives (8.2); using p gives (8.3). The energy expectation of a sum of local functions is also fixed by these marginals. Inserting $\mathcal{F}_E(p) = \mathbb{E}_p E - k_{\mathrm{B}} T \mathsf{H}(p)$ gives the last claim. Nonnegativity and the equality condition of divergence give uniqueness. □

If physical admissibility imposes constraints whose entire scopes are among the U_i, and each p_i gives probability one to its constraint, then q is admissible. A missing global constraint can exclude the separator coupling. For instance, consider the disjoint scopes of three bits whose marginals are fair, with a global even parity rule. The product law gives parity zero with probability $1/2$ and is inadmissible. On the allowed four assignments the uniform law is entropy maximizing, with entropy $2 \log 2$. The local marginals and the explicitly imposed parity rule together specify the optimization problem.

8.3 Cycles and a finite obstruction

For three fair bits A, B, C, prescribe $A = B$ almost surely on AB, $B = C$ on BC, and $A \neq C$ on AC. The three pair laws are compatible on their singleton intersections. Their conjunction is impossible, so there is no global coupling. The obstruction persists when the support relations themselves all become full. If each pair of fair bits is required to disagree with probability 0.9, every pair outcome has positive probability. For any joint assignment, the number of disagreements among the three pairs is either zero or two. Its expected value would be 2.7, a contradiction.

For a more precise calculation use signs $a, b, c \in \{-1, 1\}$. With fair marginals the local pair laws are specified by correlations $\rho_1 = \mathbb{E}[AB]$, $\rho_2 = \mathbb{E}[BC]$, and $\rho_3 = \mathbb{E}[CA]$. The possible triples of pair products are

$$(1, 1, 1), \qquad (1, -1, -1), \qquad (-1, 1, -1), \qquad (-1, -1, 1).$$

The four weights of their distribution are fixed by the three means and normalization. For each such triple u, its weight is $(1 + \rho \cdot u)/4$: this follows by inversion of the four-by-four sign matrix, or by taking expectations of $(1 + u_1 AB + u_2 BC + u_3 CA)/4$, the indicator of the triple u. A joint law exists if and only if these four weights are nonnegative. In that case choose the product pattern with those weights, then choose uniformly between the two sign

assignments realizing it. This has fair marginals and the required correlations. It is the unique maximum-entropy law since the entropy inside each two-point fiber is at most $\log 2$, attained only by uniform choice at positive-weight fibers.

The obstruction with disagreement 0.9 has $\rho_i = -0.8$ and gives negative weight to $(1, 1, 1)$. By contrast, for disagreement $2/3$ on every pair, the all-equal weight is zero and the remaining three weights are $1/3$. The unique entropy maximizer is uniform on six assignments. This boundary example shows that a compatible solution may have zeros even with all prescribed pair outcomes individually positive.

8.4 A dual certificate for arbitrary finite covers

Let Ω be a nonempty finite set of globally allowed assignments. Send each assignment ω to the tuple of indicator vectors of its restrictions $(\omega|_{U_i})_i$. The set of joint marginal lists is the convex hull of these vectors. In particular a marginal list (p_i) is realizable if and only if, for every choice of real functions f_i on the local alphabets,

$$\sum_i \mathbb{E}_{p_i} f_i \le \max_{\omega \in \Omega} \sum_i f_i(\omega|_{U_i}). \tag{8.4}$$

Every realizable list satisfies the inequality by averaging. If a list lies outside the hull, a separating linear functional exists since the hull is compact and convex. Its coefficients are precisely functions f_i on the individual scope assignments, giving a violation. In the triangle obstruction take f_i to be disagreement indicators. The bound is two on the global maximum. Equation (8.4) is also useful for constraints on causal history support, when Ω is the set of histories permitted by the gate relations. A law on that support may still fail to have the desired independent gate randomness, a stronger implementation condition.

8.5 Descent and statistical choice

The separator formula is a choice of one joint law on a special class of covers. It can be iterated along a tree using conditional distributions. A general descent limit of distribution sets describes compatible marginal data and can include lists without a joint law. A sheafification of that marginal presheaf changes the mathematical object and can identify global laws with the same local data. In the thermodynamic ledger this identification can discard a measurable free-energy difference. The coupling must therefore be part of the finite realization unless one explicitly chooses a gluing convention and proves it applies.

CHAPTER 9

Transport of couplings and symmetry

The passage from compatible histories to laws on those histories is a finite probabilistic operation. This chapter specifies it as a natural comparison, describes its fibers, and gives two changes of observation for which the entropy effects can be calculated exactly. One change coarsens a separator. The other forgets the positions of carriers related by a finite group.

9.1 The distribution comparison

For a finite set A, let $\mathsf{P}(A)$ be its simplex of probability laws. A map $g : A \to B$ gives a pushforward by summing the weights in each fiber. Functoriality follows by grouping the summands by the composite map. A kernel $K : A \rightsquigarrow B$ can equivalently be written $A \to \mathsf{P}(B)$. Composing such maps averages a conditional law against the preceding law. This is the Kleisli composition of the finite-support distribution monad on sets. Its unit is the point-mass assignment and its multiplication is barycentric averaging. On a finite kernel, associativity follows by interchanging two finite sums.

Let $H(U)$ be a finite history set and let a cover identify it with the limit of a diagram of local history sets. The cone maps induce a natural comparison

$$\mathsf{P}(H(U)) \longrightarrow \lim_i \mathsf{P}(H(U_i)),$$

where the limit on the right includes compatibility over overlaps. This comparison exists by functoriality of pushforward. The result of chapter 8 shows both ways it can fail to be an isomorphism. Disjoint scope marginals can have many couplings, and cyclic scope marginals can lack all couplings. Even when compatible sections form the stated limit, the probability assignment preserves the comparison and need not preserve the limit itself.

$$\begin{array}{ccc} \mathsf{P}(H(U)) & \longrightarrow & \mathsf{P}(H(U_1)) \\ \downarrow & & \downarrow \\ \mathsf{P}(H(U_2)) & \longrightarrow & \mathsf{P}(H(U_{12})) \end{array}$$

Figure 9.1: The probability square commutes by marginalization. Whether a matching family of marginal points lifts, and how many lifts it has, are separate questions from descent of the histories.

For finite sets, regard the probability spaces and the matching space of marginals as convex subsets of Euclidean spaces. The comparison is affine.

Its ordinary fiber at a matching family consists of the solutions of linear equalities with nonnegative coordinates and total mass one. This fiber is a compact convex polytope, possibly empty. If it is nonempty it is contractible in its Euclidean topology by linear interpolation to any selected law. That contractibility is compatible with having many different couplings, with different entropies and different work costs. Homotopy type of the law space under continuous deformation and uniqueness of a probability law are distinct. For a tree of coordinate scopes, theorem 8.2 provides a canonical point of this polytope by a strictly concave optimization.

9.2 Coarsening an interface before gluing

Let A, S, B have the separator law q so that $A \perp B \mid S$. Apply local deterministic readouts $A' = f(A)$, $B' = g(B)$, and $S' = h(S)$. Denote the resulting joint law by $\widetilde{q}$. The two new scope laws on $A'S'$ and $S'B'$ agree on S'. Let q' be their new separator coupling. Pushing forward the old coupling and gluing after coarsening need not agree.

Theorem 9.1 (Cost of a coarsened separator)**.** *In the stated setting,*

$$\mathsf{H}(q') - \mathsf{H}(\widetilde{q}) = \mathsf{D}(\widetilde{q} \| q') = \mathsf{I}_{\widetilde{q}}(A'; B' \mid S'), \tag{9.1}$$

$$\mathsf{I}_{\widetilde{q}}(A'; B' \mid S') \le \min\{\mathsf{I}(A'; S \mid S'), \mathsf{I}(B'; S \mid S')\} \le \mathsf{H}(S \mid S'). \tag{9.2}$$

The two procedures agree exactly when the new readouts are conditionally independent given the new separator.

Proof. The first equality is the two-scope instance of theorem 8.2. The second follows by expanding divergence from the conditional product given S'. Given S, S', the two readouts are independent because S' is a function of S and the old law has separator S. Thus $\mathsf{I}(A'; S, B' \mid S') = \mathsf{I}(A'; S \mid S')$. Expanding this mutual information in the other order and using nonnegativity gives $\mathsf{I}(A'; B' \mid S') \le \mathsf{I}(A'; S \mid S')$. The symmetric bound is identical with the readouts exchanged. Either term is at most $\mathsf{H}(S \mid S')$. Equality of the couplings is the zero condition for their divergence. □

For instance let S be fair and $A = B = S$. Conditional on S, the old outputs are deterministic and independent as a product of point masses. If S' is a singleton and the outputs are retained, the pushed law is diagonal while the glued law is independent fair. Their entropy difference is $\log 2 = \mathsf{H}(S \mid S')$, reaching the bound. At flat energies the corresponding free-energy difference is $k_{\mathrm{B}}T \log 2$. A physical observer who loses access to S can still see the output

correlation in a joint experiment. Replacing it by the new separator law would alter that experiment.

If instead $A' \perp S \mid S'$, the first term in the minimum of (9.2) is zero and the two operations do agree. For example an old separator can include irrelevant coordinates that neither readout uses once S' is specified. The condition gives a sufficient criterion for simplifying the interface without changing the separator law. It has a causal interpretation when S' is the only message made available to two downstream devices. Their conditional independence must come from their independent randomness and from all other shared causes having been declared. A statistical pushforward by itself does not guarantee it.

On a larger scope tree, simultaneous variablewise coarsening preserves the running intersection of named variables, and the coarsened clique marginals remain compatible. The new separator law then gives a global entropy upper bound for the pushed joint law with the same exact KL deficit. Applying coarsening consistently to an overlap variable is essential. If a stochastic coarsening is used, the overlap is sampled once and its result supplied to every adjacent scope. Sampling independent coarse values for the same overlap changes the underlying variable set and hence the comparison.

9.3 Symmetrizing a family of physical carriers

Let G be a finite group acting by permutations on a finite microstate set S. Assume $E(gs) = E(s)$. For a law p, write $p_g = g_* p$ and $\bar{p} = |G|^{-1} \sum_g p_g$. The energy expectation of every p_g is the same. If a family of linear marginal or success constraints is invariant under the group action and p is feasible, then $\bar{p}$ is feasible by convexity. The entropy gain has an exact expression:

$$\mathsf{H}(\bar{p}) - \mathsf{H}(p) = \frac{1}{|G|} \sum_g \mathsf{D}(p_g \| \bar{p}). \tag{9.3}$$

Indeed each $p_g \ll \bar{p}$, each has entropy $\mathsf{H}(p)$, and their expectations of $\log \bar{p}$ average to its expectation under $\bar{p}$. Expanding gives (9.3). Thus $\mathcal{F}_E(\bar{p}) \leq \mathcal{F}_E(p)$, with equality if and only if p is G-invariant. This supplies a reduction for an optimization over permutation-invariant copy contracts. It does not say that the corresponding symmetrized law is achievable when allowed causal protocols themselves are restricted or when random permutation needs extra resources.

Let $O = S/G$ be the orbit set, and let π be the Gibbs law of E. Its pushed law on O is Gibbs for

$$\bar{E}(o) = E(s_o) - k_{\mathrm{B}} T \log |o|,$$

where $s_o \in o$. This is (25.1). Within each orbit the conditional Gibbs law is uniform. For any ensemble p, the hidden part of its free energy under observation of the orbit is

$$k_{\mathrm{B}}T \sum_{o} p_O(o)\big[\log|o| - \mathsf{H}(p(\cdot \mid o))\big]. \tag{9.4}$$

The orbit–stabilizer identity gives $|o| = |G|/|G_{s_o}|$. Thus the stabilizers appearing in the homotopy quotient of the observation theory also enter the degeneracy count when the individual orbit elements are physically distinct microstates. If the action identifies different descriptions of the same physical state rather than different states, the physical counting measure and energy need to be specified on the actual state space. Counting redundant descriptions as thermodynamic multiplicity would change the physical model.

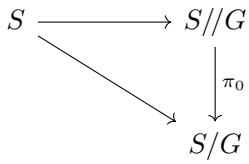


Figure 9.2: The action groupoid retains the stabilizer of every orbit. Its components give the observed orbit set, while the map from the physical state set has finite fibers of size $|G|/|G_s|$.

9.4 An unordered binary array

Consider r flat target bits, with the permutation group acting on their positions. An orbit is determined by the Hamming weight k. Its size is $\binom{r}{k}$, so its coarse energy relative to the original flat zero is $-k_{\mathrm{B}}T\log\binom{r}{k}$. The induced equilibrium law on k is binomial with success $1/2$. A majority decoder depends only on k. If the conditional ensemble given the reference is exchangeable, it is uniform within each weight orbit of positive probability. Then (9.4) is zero within each reference block. The coarse weight law with the coarse energy gives the same endpoint free energy as the full exchangeable array.

If a device always sends a different error pattern to particular locations, the majority outcome can have the same distribution while the conditional distribution within an orbit is nonuniform. For example with two targets, restrict the final conditional error pattern to exactly one wrong bit. Choosing the wrong position uniformly yields conditional entropy $\log 2$, and always choosing the first position yields conditional entropy zero. Both have the same weight of the error pattern. At flat energy, the latter law carries $k_{\mathrm{B}}T\log 2$ more free energy in that block. The orbit observer has no access to the distinction. An individual site,

however, can have a different decoding reliability, so a permutation-invariant majority contract and a site-specific replica contract use different observations.

These results connect change of observation to the statistical and energetic layers without assigning a canonical measure to an arbitrary higher groupoid. The symmetry calculation uses a finite action, an explicitly counted physical microstate set, and a Gibbs reference. The coupling calculation uses a declared finite law. The categorical maps specify the fibers over which those finite calculations are performed.

CHAPTER 10

Finite isothermal protocols

Let S be a finite microstate space, $T > 0$ a fixed temperature, and $\beta = (k_{\mathrm{B}}T)^{-1}$. An energy is a function $E : S \to \mathbb{R}$. Its Gibbs law and equilibrium Helmholtz free energy are

$$\pi_E(s) = Z_E^{-1} \exp(-\beta E(s)), \quad Z_E = \sum_s \exp(-\beta E(s)), \quad F_{\mathrm{eq}}(E) = -\beta^{-1} \log Z_E.$$

For an ensemble p, the nonequilibrium free energy is

$$\mathcal{F}_E(p) = \sum_s p(s)E(s) - \beta^{-1}\mathsf{H}(p) = F_{\mathrm{eq}}(E) + \beta^{-1}\mathsf{D}(p \,\|\, \pi_E)\,. \tag{10.1}$$

The equality follows from the logarithm of the Gibbs weights. It holds for laws with zeros because the energy is finite and π_E has full support.

We take work and heat into the modeled system to be positive. A quench changes E with the ensemble held fixed. It performs mean work given by the expected change in energy. A fixed-Hamiltonian step is a kernel K with $\pi_E K = \pi_E$. In the protocol accounting this step performs no drive work, and the change in its expected energy is assigned to heat. This is a physical idealization of the specified class. A microscopic interpretation as heat exchange with one equilibrium bath can be made for a step satisfying detailed balance. If a hidden controller or driving supply is used to implement a kernel, its energetic cost belongs to an enlarged boundary. Mere stationarity gives the mathematical inequality below, and the actual thermodynamics of a driven device can contain additional terms.

10.1 Exact accounting by relative entropy

Start with (p_0, E_0). At stage j quench to E_j, then apply a fixed-Hamiltonian Gibbs-stationary kernel K_j, yielding $p_j = p_{j-1}K_j$. Identity kernels allow a final

quench with no following evolution. Permutations preserving E_j are allowed kernels and can represent reversible flat-energy logical gates. For the j-th stage write

$$W_j = \mathbb{E}_{p_{j-1}}(E_j - E_{j-1}), \qquad Q_j = \mathbb{E}_{p_j} E_j - \mathbb{E}_{p_{j-1}} E_j.$$

The energy change over that stage is $W_j + Q_j$. In particular, the net heat delivered to the bath in a protocol with zero net internal-energy change equals the input work under this accounting.

Theorem 10.1 (Isothermal work ledger)**.** *For the protocol just specified,*

$$W_{\mathrm{in}} - \left[\mathcal{F}_{E_N}(p_N) - \mathcal{F}_{E_0}(p_0)\right] = \beta^{-1} \sum_{j=1}^{N} \sigma_j, \tag{10.2}$$

$$\sigma_j = \mathsf{D}\big(p_{j-1} \,\big\|\, \pi_{E_j}\big) - \mathsf{D}\big(p_j \,\big\|\, \pi_{E_j}\big) \geq 0. \tag{10.3}$$

The expression for each σ_j is also $\mathsf{H}(p_j) - \mathsf{H}(p_{j-1}) - \beta Q_j$.

Proof. At a quench the entropy is unchanged, so the free energy increases by W_j. At the subsequent kernel step the equilibrium free energy stays fixed and (10.1) gives the change as $-\beta^{-1}\sigma_j$. Nonnegativity follows from (7.1) with $q = \pi_{E_j}$ and $qK_j = q$. Summing telescopes. The formula involving Q_j follows by expanding the two divergences and using the expression for $\log \pi_{E_j}$. □

Each loss vanishes exactly when the Gibbs reverse of the step recovers the actual incoming law from the outgoing one, in the sense of (7.1). A permutation preserving E has zero loss for all input laws. For the full equilibration kernel with all rows equal to π_E, the loss is precisely $\mathsf{D}(p \,\|\, \pi_E)$, and it vanishes only for the equilibrium input.

If an independent bath acts on a component A of a larger state A, B, with fixed additive energy, the loss decomposes into a marginal term and a correlation term. For $K_A \otimes \mathrm{id}_B$,

$$\begin{aligned} &\mathsf{D}(p_{AB} \,\|\, \pi_A \pi_B) - \mathsf{D}(p'_{AB} \,\|\, \pi_A \pi_B) \\ &\quad = \mathsf{D}(p_A \,\|\, \pi_A) - \mathsf{D}(p'_A \,\|\, \pi_A) + \mathsf{I}_p(A;B) - \mathsf{I}_{p'}(A;B). \end{aligned} \tag{10.4}$$

This follows by expanding both divergences against the product reference. The marginal law on B is unchanged. If A is already thermal marginally, its independent complete equilibration changes the correlation to zero without changing any marginal energy. The total free-energy loss is then $k_{\mathrm{B}} T \mathsf{I}_p(A;B)$, though the drive work and mean heat in the step can both be zero. The entropy of the joint state increases.

10.2 Trajectory reversal

For a fixed-Hamiltonian step with $\pi = \pi_E$, its Gibbs adjoint has entries

$$K^\dagger(x \mid y) = \frac{\pi(x)K(y \mid x)}{\pi(y)}. \tag{10.5}$$

It is a normalized kernel by stationarity. Compare the forward joint law $P(x, y) = p(x)K(y \mid x)$ with the reversed-order joint law $Q(x, y) = p'(y)K^\dagger(x \mid y)$. On a forward-positive pair,

$$\log \frac{P(x, y)}{Q(x, y)} = \log \frac{p(x)}{p'(y)} - \beta\big(E(y) - E(x)\big).$$

Averaging gives exactly σ, and $K^\dagger = K$ as a transition rule when detailed balance holds. This comparison uses time reversal of the statistical protocol, including the law at its starting boundary. For variables with physical time-reversal parity, the corresponding parity action must be included in the physical reverse. The finite model here uses time-even state labels.

For the full forward trajectory, start the reverse law at p_N, apply the adjoint kernels in reverse order and reverse each quench after its adjoint step. Both laws can be compared on the same ordered state paths. The sum of the individual likelihood ratios telescopes in the p_j factors, giving

$$\mathsf{D}(P_{\text{path}} \| Q_{\text{path}}) = \sum_j \sigma_j.$$

Consequently if $\mathfrak{s}(\omega) = \log(P_{\text{path}}(\omega)/Q_{\text{path}}(\omega))$ on forward support, then $\mathbb{E}_{P_{\text{path}}} e^{-\mathfrak{s}} = Q_{\text{path}}(\operatorname{supp} P_{\text{path}}) \leq 1$. Equality holds when the reverse law gives no mass to paths outside the forward support. With full initial support and finite energies, support coincidence follows: adjoints reverse every positive transition and all intermediate marginals are full by stationarity of a full-support reference. Prepared pure states need not satisfy this support hypothesis.

10.3 Coarse events and self-replication rates

Let A be an event on paths with $P(A) > 0$ and $Q(A) > 0$, where Q is the reverse law written on forward-oriented paths. Restricting the likelihood ratio to A gives the exact relation

$$\mathbb{E}_P[\mathfrak{s} \mid A] = \log \frac{P(A)}{Q(A)} + \mathsf{D}\big(P(\cdot \mid A) \| Q(\cdot \mid A)\big) \geq \log \frac{P(A)}{Q(A)}.$$

The binary event also gives $\mathsf{D}(P\|Q) \geq \mathsf{D}(\mathrm{Bern}(P(A)) \| \mathrm{Bern}(Q(A)))$ by data processing. When A is the formation of a successful record and $Q(A)$ is small, the successful trajectories can be thermodynamically asymmetric. If $P \ll Q$,

$P(A) > 0$ already forces $Q(A) > 0$. In a physical fluctuation theorem, the exact reverse experiment and boundary conditioning must be specified to interpret $Q(A)$ as a reverse transition probability.

England's treatment of biological self-replication uses coarse forward and reverse transitions with a thermal bath and estimates a dissipation constraint [10]. The path comparison above is a finite algebraic analogue under the Gibbs-adjoint protocol. The endpoint copy inequality in the next chapter needs only (10.2) and a replica-decoding requirement. It contains no kinetic rate or growth law. Conversely, a kinetic coarse bound can distinguish two protocols with the same endpoint joint law. The respective statistics are the full joint endpoint ensemble and the ratio of coarse path probabilities.

10.4 Boundary choices and work in physical units

At one fixed energy, $\beta^{-1}\mathsf{D}(p\|\pi_E)$ is the free energy of p above equilibrium. It has units of energy. Entropies and divergences in this manuscript are dimensionless numbers in nats, and their physical entropy equivalents are multiplied by k_{B}. At a flat energy on m states, a pure state has free-energy excess $k_{\mathrm{B}}T\log m$. If there are multiple baths, the work accounting can keep the heat for each bath and use a sum of $-Q_j/T_j$ in the entropy production. The single-temperature free-energy differences in (10.2) rely on the common T.

A register used as a work or information resource belongs to the modeled system unless it is an external supply whose contribution is separately recorded. Apparent work savings can occur if a blank is initially out of equilibrium, if a controller's memory is consumed, or if the machine receives preexisting correlations. Ancillary systems that finish with their original marginals can still carry new correlations. A genuinely returned auxiliary state with the same full joint law and energy makes its free-energy change zero. The next chapter gives a factorized endpoint formula that records these contributions.

CHAPTER 11

The work of creating replicas

Fix an endpoint boundary of individually readable carriers $Y = (Y_1, \dots, Y_r)$. Assume their full energetic state at the endpoints is their joint finite state, with a product Hamiltonian in the sense $E(y) = \sum_i E_i(y_i)$. The notation permits different initial and final energies, each additive at the indicated endpoints. One may include the retained source as a carrier or place it in an explicit side register.

The endpoint requirement does not constrain the Hamiltonian at intermediate times. Interactions during copying can be essential to the attainable work.

11.1 Additive energies and the value of correlations

For the product reference $\pi = \prod_i \pi_i$, direct expansion gives

$$\mathsf{D}\left(p_Y \,\middle\|\, \prod_i \pi_i\right) = \mathsf{TC}(Y) + \sum_i \mathsf{D}(p_{Y_i} \,\|\, \pi_i)\,, \tag{11.1}$$

$$\mathcal{F}_E(p_Y) = \sum_i \mathcal{F}_{E_i}(p_{Y_i}) + k_{\mathrm{B}}T\mathsf{TC}(Y). \tag{11.2}$$

The first identity subtracts and adds $\log \prod_i p_{Y_i}(y_i)$ in the sum, with zero-mass states omitted. The second follows from the entropy formula or from (10.1) and additivity of F_{eq}. It holds for any ensemble, even when the marginals are each thermal. For an interaction term E_{int} at an endpoint the same entropy expansion is valid, with the additional term $\mathbb{E}_p E_{\mathrm{int}}$ on the right. A model with interacting equilibrium can already favor the correlated target, which affects the work comparison.

When W is a retained register with p_W fixed, define conditional endpoint energies by $E(w, y) = E_W(w) + \sum_i E_i(y_i, w)$, and allow E_W and the E_i to change with the protocol. Then

$$\mathcal{F}_E(p_{W,Y}) = \mathcal{F}_{E_W}(p_W) + \sum_i \mathbb{E}_{w\sim p_W} \mathcal{F}_{E_i(\cdot,w)}(p_{Y_i|w}) + k_{\mathrm{B}}T\mathsf{TC}(Y \mid W). \tag{11.3}$$

This is a single application of the conditional entropy chain rule. If E_W and p_W are fixed, the first term cancels in endpoint differences. When there is no conditioning, use W a singleton. If W includes the source label, the information required beyond W is zero. In that convention the change in conditional marginal free energies of the receivers still measures their preparation. Alternatively include the source as one of the Y_i and use the retained correlation identity (7.7). These conventions produce the same free energy when applied to the same full boundary.

Theorem 11.1 (A resource-sensitive work inequality)**.** *Suppose an isothermal protocol in theorem 10.1 ends with r carriers Y and retains a declared W. Use the endpoint form in (11.3), with the same side law and side energy. Let ΔF_i be the change of the averaged conditional free energy of carrier i, and let TC_0 be the initial conditional total correlation of the carriers. If each final Y_i, W decodes*

the label X with error e_i, then

$$W_{\mathrm{in}} \geq \sum_i \Delta F_i - k_{\mathrm{B}}T\mathsf{TC}_0 + k_{\mathrm{B}}T \max\left\{0, (r-1)\mathsf{H}(X \mid W) - \sum_i f_m(e_i)\right\}. \tag{11.4}$$

The exact identity replaces the maximum in (11.4) by the final $\mathsf{TC}(Y \mid W)$ and adds $k_{\mathrm{B}}T\sum_j \sigma_j$ on the right.

Proof. Use (11.3) at each endpoint of (10.2). The common $\mathcal{F}_{E_W}(p_W)$ cancels. This gives the identity. For the lower bound apply theorem 7.1 to the final joint readout law and use $\sigma_j \geq 0$. □

The full-system boundary is needed for this version. If Y consists only of coarse observations of physical registers, the ledger applies to the physical registers. A local coarse decoder cannot increase total correlation of the physical registers, by the data-processing argument in chapter 7, so the final lower bound still applies to their correlation under product endpoint energies. The initial term and marginal free energies must then be calculated on the physical states. Likewise, if additional ancillary variables are present, include them in the free-energy difference. They contribute their marginal change and their changing correlations with the carriers, or cancel if the full ancillary factor is returned independent at the endpoints.

11.2 A retained binary reference and flat targets

Let X be fair, retain its register, and let r new target bits begin independently fair and independent of X. All endpoints have flat energy. For a final bit Y_i whose error under its named binary decoder is e_i, conditional Fano gives $\mathsf{H}(Y_i \mid X) \leq h_{\mathrm{b}}(e_i)$ when the decoder is the identity: the error is $N_i = X \oplus Y_i$, so $\mathsf{H}(Y_i \mid X) = \mathsf{H}(N_i \mid X) \leq h_{\mathrm{b}}(e_i)$. The same bound on the entropy of a binary target with an informative deterministic decoder follows by relabeling if the decoder is a bijection. For a constant decoder the error is $1/2$ for fair X, and the entropy bound is then trivial. For these bit targets,

$$W_{\mathrm{in}} \geq k_{\mathrm{B}}T\,[r\log 2 - \mathsf{H}(Y \mid X)] \geq k_{\mathrm{B}}T\sum_{i=1}^{r}\big(\log 2 - h_{\mathrm{b}}(e_i)\big). \tag{11.5}$$

The first inequality uses the actual full initial and final entropies. The second uses conditional subadditivity. Equality of the endpoint bound holds for independently drawn symmetric errors conditional on X. If a target carries extra internal states, this binary equality need not apply to its physical preparation.

With exact decoding, the free-energy increase is $rk_{\mathrm{B}}T\log 2$. With pure zero blanks instead of fair targets, initial entropy is $\mathsf{H}(X) = \log 2$. The exact final array has the same entropy. The r controlled XORs form a reversible permutation and can have zero work in the flat-energy idealization. In the decomposition (11.2), the sum of the marginal free energies drops by $rk_{\mathrm{B}}T\log 2$ as each target becomes marginally fair, while total correlation increases by $r\log 2$. This balance explicitly uses the initial purity. Returning a consumed blank to its pure state in an independent cycle is a separate protocol.

For an arbitrary finite X with alphabet of size m and r initially uniform independent targets of that alphabet, an exact final retained array has joint entropy $\mathsf{H}(X)$. The initial entropy is $\mathsf{H}(X) + r\log m$, so its full free-energy increase is $rk_{\mathrm{B}}T\log m$, even when X is biased. The final correlation component is $rk_{\mathrm{B}}T\mathsf{H}(X)$; the remainder is the cost of changing the initially uniform marginals to the biased source law. If the initial targets instead each have the same biased law as X and are independent, the entire free-energy increase is $rk_{\mathrm{B}}T\mathsf{H}(X)$. A thermal state with a biased finite alphabet requires a corresponding nonflat energy, or a higher-dimensional microstate model. The formula for E must agree with the intended interpretation of the blanks.

11.3 Achieving the bound with controlled equilibration

We give an explicit protocol class in which the endpoint lower bound is a work infimum. Let the state be (X, Z), where X is retained, Z is finite, and the initial and target conditional laws $p(z \mid x), q(z \mid x)$ have full support for all x of positive $p_X(x)$. At unused x choose arbitrary full-support conditional laws. Permit arbitrary finite energy changes on (x, z) and a complete conditional reset of Z given x to the Gibbs conditional for that energy. This is a Gibbs-stationary kernel on the full state: it fixes x and equilibrates inside each x-block. It also satisfies detailed balance since within a block its joint Gibbs flux is symmetric.

Theorem 11.2 (Controlled work infimum)**.** *In the protocol class above, for any finite initial and final energies on (X, Z), the infimum of input work among protocols realizing the target law exactly is*

$$\mathcal{F}_{E_{\mathrm{out}}}(p_X q) - \mathcal{F}_{E_{\mathrm{in}}}(p_X p).$$

For a target with zero conditional entries, the corresponding endpoint lower bound is approached by a sequence of full-support target approximations converging to it.

Proof. The lower bound is theorem 10.1. For an integer N, define $p_j(z \mid x) = (1-$

$j/N)p(z \mid x) + (j/N)q(z \mid x)$. Quench to the energy $E^{(0)}(x, z) = -\beta^{-1} \log p_0(z \mid x)$. Its Gibbs conditional is p_0, so one may use the identity kernel. For $j = 1, \ldots, N$, quench to $E^{(j)} = -\beta^{-1} \log p_j(z \mid x)$ and apply a conditional reset. The law becomes $p_X p_j$. The full Gibbs law of this chosen energy has uniform x because each conditional partition function is one. The x-marginal is fixed throughout, so the full divergence loss in this step is

$$\sigma_j = \sum_x p_X(x) \mathsf{D}(p_{j-1}(\cdot \mid x) \,\|\, p_j(\cdot \mid x)) \,.$$

For positive v and probability vector u, $\mathsf{D}(u\|v) \leq \sum_z (u(z) - v(z))^2 / v(z)$, from $\log t \leq t - 1$. All entries of p_j on used blocks have a common positive lower bound. Differences of consecutive rows are $O(N^{-1})$, so the sum of the N losses is $O(N^{-1})$. Quench to the desired endpoint energy with no further stochastic evolution. This realizes the exact q for each N and approaches the lower bound. If q has zeros, replace it by $(1 - \delta)q + \delta u$ with u uniform on Z. The full endpoint free energy converges by continuity of entropy on the finite simplex. Choose N after δ to make the total loss tend to zero. □

For fair X and r symmetric independent targets with crossover $e \in (0, 1/2]$, this proves the infimum $rk_{\mathrm{B}}T(\log 2 - h_{\mathrm{b}}(e))$ in (11.5). One can realize the conditional reset independently on each target when the conditional law is a product, using controlled energies additive over Z_i for fixed x. An implementation needs causal access to x at the place where each such coupling is performed. The theorem allows those couplings as operations; the cut results govern their placement. The limit $e \to 0$ establishes the stated asymptotic cost for increasingly accurate binary targets. For fixed finite energies, a Gibbs-stationary kernel sends a full-support law to a full-support law, so a finite protocol starting from full support in this same state space cannot reach a target with zero entries. Indeed, each target state has a positive incoming transition from some state because the stationary law has full support. A pure blank, an ideal infinite gap, or another resource changes the attainable class.

11.4 Consuming correlation instead of blank purity

Initially correlated blank registers can also fund replication. Let X, R be independent fair bits and let the initial flat-energy array be (X, R, R). The permutation $(x, y, z) \mapsto (x, y, z \oplus y \oplus x)$ sends it to (X, R, X). All marginal laws and the total entropy are preserved. The initial correlation was between the two blanks; the final correlation is between the retained label and the new target. If R must be left independent and in its initial marginal, the permutation does that on the

retained R-register. If the two initial blank registers were required to be returned as a correlated ancillary factor, this same operation would not meet the contract. The same endpoint resource cannot both be returned and be consumed as the new correlation.

11.5 Equality conditions and possible extra resources

The lower bound in (11.4) can be strict for three logically distinct reasons. The Fano bounds can be loose for the actual error distribution. The carriers can retain correlation conditional on the source, or their joint readout can leave residual source uncertainty. Finally the implementation can incur nonzero relative-entropy losses. If an implementation is chosen before the final coupling is chosen, the constraint set can prohibit the conditional-independent coupling needed for equality. If the final coupling is prescribed and the operations allow arbitrary conditional equilibration, theorem 11.2 supplies the asymptotic equality at full support. The scope of an equality claim therefore depends on both the coupling problem and the permitted protocol.

CHAPTER 12

Optimal preparation at a prescribed readout accuracy

The replica inequality treats the endpoint marginals and error probabilities as data. In some designs the energy of each target is fixed, while the joint output law can be chosen subject to decoding requirements. This gives a constrained minimization of nonequilibrium free energy. We solve it for retained classical reference labels and independent thermal targets, allowing nonuniform energies and decoders that group several microstates into one label. The proof also gives a sharp m-ary work formula at flat energy.

12.1 A single physical target with a fixed decoder

Let $X \in L$ have law μ, and keep its register unchanged. Let A be the finite microstate space of a receiver, E its endpoint energy, and π its Gibbs law. At the initial boundary A is independent of X and has law π. Write $d : A \to L$ for the fixed decoder. For label x, let $S_x = d^{-1}(x)$ and put $s_x = \pi(S_x)$. Assume every x of positive μ-mass has $0 < s_x < 1$. If a label has no correct microstate, its conditional accuracy is necessarily zero. If the correct set is the whole receiver

alphabet, the accuracy is necessarily one in that block. Those blocks can be accounted for by fixed contributions before optimizing the others.

A final conditional law q_x on A has mean success $a(q) = \sum_x \mu(x) q_x(S_x)$. The energy of X is fixed, and the endpoint energy is additive between X and A. The free-energy increase from the independent thermal target to the desired q is

$$\Delta\mathcal{F} = k_{\mathrm{B}}T \sum_x \mu(x)\mathsf{D}(q_x\|\pi). \tag{12.1}$$

This follows from conditional free energy, or the chain rule against the reference $\mu(x)\pi(a)$. The distribution μ need not be an equilibrium distribution of the source. Its contribution cancels because its law and energy are retained. A control interaction may use x at intermediate times, provided the source is accessible to that interaction.

For each x, the coarse correct/incorrect partition induces a Bernoulli law of mean $u_x = q_x(S_x)$. Expanding the divergence over the two cells gives

$$\begin{aligned}\mathsf{D}(q_x\|\pi) = {}& \mathsf{D}(\mathrm{Bern}(u_x)\|\,\mathrm{Bern}(s_x)) \\ &+ u_x\,\mathsf{D}(q_x(\cdot \mid S_x)\|\pi(\cdot \mid S_x)) \\ &+ (1-u_x)\,\mathsf{D}(q_x(\cdot \mid A\setminus S_x)\|\pi(\cdot \mid A\setminus S_x)).\end{aligned} \tag{12.2}$$

If a cell has zero q_x-mass its term is zero. Thus for a fixed accuracy in each block the unique least-cost microstate law uses the Gibbs conditionals inside its correct and incorrect cells. The cost of a given correct probability is then a binary divergence from the equilibrium correct probability. A physical decoder with multiple allowed microstates for each label can use those microstates according to their Gibbs weights.

12.2 One multiplier for average accuracy

Let a_* be the required average success, with equilibrium success $a_0 = \sum_x \mu(x) s_x$. Define

$$\psi(\lambda) = \sum_x \mu(x)\log\bigl(1 + s_x(e^\lambda - 1)\bigr).$$

For $\lambda \geq 0$ define a blockwise tilted law

$$q_x^\lambda(z) = \frac{\pi(z)e^{\lambda 1_{S_x}(z)}}{1 + s_x(e^\lambda - 1)}. \tag{12.3}$$

A direct calculation gives $\psi'(\lambda) = a(q^\lambda)$ and $\psi''(\lambda) = \sum_x \mu(x) u_x^\lambda(1-u_x^\lambda) > 0$, with $u_x^\lambda = q_x^\lambda(S_x)$. Thus ψ' increases continuously from a_0 to one.

Theorem 12.1 (Minimum free energy at a decoding accuracy)**.** *In the setting above, for $a_* \in [0,1]$ the minimum dimensionless endpoint free energy over*

conditional laws with $a(q) \geq a_*$ is

$$J(a_*) = \sup_{\lambda \geq 0} \{\lambda a_* - \psi(\lambda)\}. \tag{12.4}$$

For $a_* \leq a_0$ it is zero, with $q_x = \pi$ on the used blocks. For $a_0 < a_* < 1$, the unique optimizer is $q_x^{\lambda_*}$, where $\psi'(\lambda_*) = a_*$. At $a_* = 1$ it is $\sum_x \mu(x)(-\log s_x)$, with $q_x = \pi(\cdot \mid S_x)$. When controlled equilibration with retained X is allowed, $k_{\mathrm{B}}TJ(a_*)$ is the work infimum for $a_* < 1$. At exact accuracy it gives the lower bound and the limit of the infima as $a_* \uparrow 1$.

Proof. For any q, substitute (12.3) in the divergence to get

$$\sum_x \mu(x)\mathsf{D}(q_x \| \pi) = \sum_x \mu(x)\mathsf{D}(q_x \| q_x^\lambda) + \lambda a(q) - \psi(\lambda).$$

For feasible q and $\lambda \geq 0$ this is at least $\lambda a_* - \psi(\lambda)$. If $a_* \leq a_0$, the original Gibbs conditionals are feasible and give zero, matching $\lambda = 0$. If $a_0 < a_* < 1$, the strict increase of ψ' gives a unique $\lambda_* > 0$. Its tilted law is feasible with equality of the accuracy bound and zero divergence to itself, so it attains the dual value. At $a_* = 1$, every used block must have success one. By (12.2), the minimum is the stated mean of $-\log s_x$; the dual expression tends to it as $\lambda \to \infty$. Uniqueness on positive-mass blocks follows from the divergence equality condition. For a finite multiplier all target conditionals have full support. Apply theorem 11.2 to attain the endpoint infimum in work. For exact accuracy the tilted laws converge to the indicated conditional Gibbs laws, so the finite-state endpoint values converge. □

The multiplier λ_* is dimensionless. Energetically, the tilted equilibrium on each x-block is obtained by subtracting $k_{\mathrm{B}}T\lambda_*$ from the energies of the correct microstates, then restoring the endpoint Hamiltonian after the desired law has been prepared. A single quench to that tilt followed by equilibration can be dissipative. The work infimum uses a sequence of stages that matches the conditional law along the path, as in the proof of theorem 11.2. The theorem concerns the best mean work given available control and the finite energetic state, rather than a prescribed duration.

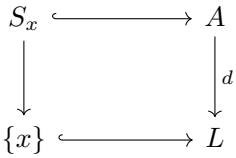


Figure 12.1: The correct microstate cell is the decoder fiber. Its Gibbs mass s_x sets the cost of concentrating the receiver in that cell. The square is a pullback of finite sets.

12.3 Many targets and the cost of an additional coupling

Take receiver energies E_i and Gibbs laws π_i, with the full initial receiver law $\prod_i \pi_i$ independent of X. Fix one decoder per carrier and a separate average success threshold a_i. For each final joint conditional law $q(\cdot \mid x)$, write $q_i(\cdot \mid x)$ for its marginal. Expanding relative entropy against the product Gibbs law conditional on X gives

$$\frac{\Delta \mathcal{F}}{k_{\mathrm{B}} T} = \sum_i \sum_x \mu(x) \mathsf{D}(q_i(\cdot \mid x) \| \pi_i) + \mathsf{TC}_q(Y \mid X). \tag{12.5}$$

The constraints concern only the q_i. Therefore the minimal endpoint cost is $k_{\mathrm{B}} T \sum_i J_i(a_i)$. It is achieved by taking the conditional product of the single-target optimizers. Any residual dependence conditional on X adds exactly the last term of (12.5) for the same chosen marginals. If a causal design requires a common noise, the optimal product coupling may be unavailable. In that case this is still a lower bound, and the residual term computes the excess due to that particular correlation.

If all thresholds are strictly below one and every target has causal access to the control, independent controlled thermalizations attain the infimum. These operations can occur at separate spacetime sites after the control is brought to each. Any retained registers used to transport the control belong to the full energetic account. The controlled-target formula applies to the part of that protocol with its reference fixed. A device with access to an imperfect reference is constrained by its source posterior; adding an energetic tilt conditioned on that reference cannot remove a causal information loss.

For m labels and exactly m flat microstates per receiver, let each decoder be a bijection. Then $s_x = 1/m$ for every x, and the optimal conditional law at error $e \le 1 - 1/m$ places mass $1 - e$ on the correct label and $e/(m-1)$ on each wrong label. Thus for separate thresholds e_i,

$$W_{\mathrm{inf}} = k_{\mathrm{B}} T \sum_i \big[\log m - h_{\mathrm{b}}(e_i) - e_i \log(m-1)\big]. \tag{12.6}$$

For positive errors in that range, each term follows either from theorem 12.1 or by computing the entropy of the symmetric conditional law. At $e_i = 0$ it is the exact endpoint bound and a limit from positive errors. A biased label law μ does not change this formula with a retained source and flat thermal targets. The cost of the target is relative to uniform thermal preparation, and the same conditional output entropy is attainable in every source block.

12.4 Choosing the readout cells at exact accuracy

Now suppose the decoder can be chosen in advance. Let the target have N flat microstates and let L have $m \le N$ labels of positive prior mass. Choose integer cell sizes $n_x \ge 1$ with $\sum_x n_x = N$. By theorem 12.1 the dimensionless exact endpoint cost is

$$\sum_x \mu(x) \log(N/n_x) = \mathsf{H}(\mu) + \mathsf{D}(\mu\|s), \qquad s_x = n_x/N.$$

The equality follows by subtracting and adding $\log \mu(x)$ termwise. Hence it is at least $\mathsf{H}(\mu)$, with equality exactly when $n_x/N = \mu(x)$ for every label. For uniform labels and N divisible by m, equally sized cells give cost $\log m$, independent of the extra within-cell degeneracy. For a biased source, designing larger cells for likely labels reduces the marginal preparation cost while keeping a deterministic decoder.

The optimal integer choice can be computed by a finite allocation rule. Start with all $n_x = 1$ and assign the remaining slots successively to a label with largest current marginal gain $\mu(x) \log((n_x+1)/n_x)$. The objective is to maximize $\sum_x \mu(x) \log n_x$; each label has a decreasing sequence of gains, and choosing any slot for that label requires all its earlier gains. Selecting the largest currently available gain chooses one of the greatest remaining gains overall because all inaccessible gains for that label are smaller. Induction shows that the rule maximizes the sum for a given number of slots. This optimizes exact endpoint free energy, with the same support caveat for finite-time writing from full support.

12.5 An asymmetric numerical example

Take a two-state receiver with equilibrium probabilities $9/10, 1/10$ and decode the two microstates as labels $0, 1$. With a fair reference the unprepared target has mean success $1/2$. To reach mean success $3/4$, write $z = e^{\lambda_*}$. The correctness probabilities in the two reference blocks are $9z/(1+9z)$ and $z/(9+z)$. Their mean is $3/4$ exactly when $9z^2 - 82z - 27 = 0$, with positive root

$$z = \frac{82 + \sqrt{7696}}{18}.$$

The more favorable equilibrium label retains a higher final correctness probability. Requiring equal errors in both blocks would generally impose a larger cost than this average-accuracy contract. For exact accuracy the minimum dimensionless endpoint free energy is $\frac{1}{2}\log(10/9) + \frac{1}{2}\log 10 = \log(10/3)$. This exceeds $\log 2$, the cost of exact writing into a two-state flat target, because the thermal prior of the physical receiver is biased relative to the fair source.

For the same receiver law with a biased source matching its equilibrium probabilities, the exact endpoint cost is $\mathsf{H}(9/10, 1/10)$. The final marginal of the target is then its original Gibbs law. All cost above the initial independent state is the correlation with the retained reference. The example connects the minimum-work formula with the product-energy decomposition without assuming equal marginal energies or equiprobable labels.

12.6 Restrictions on optimizing a physical design

An ensemble optimizer does not supply a design with sufficient spacetime access. To realize the tilted laws, a control must access the source label or another variable that yields the desired conditional law at each target. A distant observer may have access only to a noisy causal cut. The strongest attainable average accuracy can then be below the optimizer's threshold. Also the work infimum relies on a protocol class whose energy landscape may depend on the control. A device restricted to blind independent target operations cannot in general create the joint correlation with X, even if its scalar work budget is large. In the mathematical model, feasibility is checked first for causal controls and the entropy optimization gives the minimal endpoint cost among the resulting constraints.

CHAPTER 13

The work of maintaining a record

Replication creates a new decoding relation. Maintenance restores an existing relation after a specified disturbance. Both can be realized as maps of finite ensembles on the same retained state space. Their initial conditions differ, and that difference appears in the free-energy ledger.

13.1 A general restoration inequality

Let p^* be the desired joint state of reference and memory, and let a passive fixed-energy Gibbs-stationary channel K produce $p^- = p^*K$. Both desired and degraded states use energy E. Suppose a repair protocol begins at (p^-, E) and ends at (p^*, E). Then

$$W_{\mathrm{repair}} \geq \mathcal{F}_E(p^*) - \mathcal{F}_E(p^-) = k_{\mathrm{B}}T\big[\mathsf{D}(p^* \,\|\, \pi_E) - \mathsf{D}\big(p^- \,\big\|\, \pi_E\big)\big]. \tag{13.1}$$

This follows directly from theorem 10.1. The right side is nonnegative by data processing. The full cycle begins and ends with the same state and energy, so its work is the heat delivered to the bath under the stated convention. The

passive step has already dissipated the indicated relative entropy, regardless of how efficiently repair is performed. A reversible repair infimum can restore the state without incurring further dissipation in that idealization.

Consider an additive energy on X, Y and a passive channel acting only on Y. If both marginals are unchanged, (10.4) reduces the right side of (13.1) to

$$k_{\mathrm{B}}T\big[\mathsf{I}_{p^*}(X;Y) - \mathsf{I}_{p^-}(X;Y)\big].$$

The information here is correlation with the retained reference, rather than the marginal entropy of the memory. A fair memory bit can have unchanged marginal entropy while losing all information about its source.

13.2 A sharp binary refresh cycle

Let X be a retained fair bit and let Y be a bit with $Y = X \oplus N$, where N is independent of X with probability e_0 of one. Take $0 < e_0 < 1/2$. Keep both energies flat at the endpoints. Suppose the passive channel flips Y with probability $\eta \in [0, 1/2]$ independently of the incoming pair. Then the crossover after the passive step is

$$e_- = (1-\eta)e_0 + \eta(1-e_0) = e_0 + (1-2e_0)\eta. \tag{13.2}$$

The pair entropy is $\log 2 + h_{\mathrm{b}}(e)$ at crossover e.

Theorem 13.1 (Binary refresh work)**.** *Any isothermal restoration of the pair to its initial crossover law with the reference retained requires*

$$W_{\mathrm{repair}} \geq k_{\mathrm{B}}T[h_{\mathrm{b}}(e_-) - h_{\mathrm{b}}(e_0)]. \tag{13.3}$$

In the class of controlled equilibration of theorem 11.2, this value is the work infimum. For r conditionally independent binary targets with independent passive flips and product restored conditional laws, the corresponding costs add.

Proof. Flat energy makes the increase in free energy from degraded to restored equal to $k_{\mathrm{B}}T$ times the decrease in joint entropy. Formula (13.2) computes that difference. Both conditional laws have full support. The allowed controlled protocol therefore attains the infimum by theorem 11.2. For multiple targets with the indicated conditional independence, the joint entropy is $\log 2 + \sum_i h_{\mathrm{b}}(e_i)$ and controlled resets may act independently on each target given the reference. □

There is no need for the individual target to reveal X to the controller by a measurement with an extra recorded output. A controlled Hamiltonian can depend on the immutable physical reference state directly. Its implementation nevertheless requires a causal channel or local interaction permitting access to

that state. If the allowed repair is a channel on Y alone and never sees a reference or other relevant side information, data processing precludes increasing $\mathsf{I}(X;Y)$. Since h_{b} is strictly increasing on $[0, 1/2]$, such a repair cannot restore the lower crossover whenever $e_- > e_0$. A decoder using other noisy copies can help only to the extent that the accessible tuple still carries information about X.

For a pure target $e_0 = 0$ with a positive passive noise rate, (13.3) gives the limiting free-energy cost $k_{\mathrm{B}}T h_{\mathrm{b}}(e_-)$. Exact restoration in the finite-energy class cannot reach the required zero probabilities from full support. The work interpretation at this endpoint is a bound and a limit over increasingly accurate target laws. With $e_0 = 1/2$ the desired record carries no source information and no refresh is required.

13.3 Continuous time between refreshes

A passive symmetric jump process on a bit, with flip rate $\gamma > 0$ in each direction, produces the crossover

$$e(t) = \tfrac{1}{2}\big[1 - (1 - 2e_0)e^{-2\gamma t}\big].$$

Indeed, the error probability obeys $\dot{e} = \gamma(1 - 2e)$ with the given initial value. If refresh operations are allowed at the end of identical intervals of length $\tau > 0$, the ideal infimum of work per interval is

$$w(\tau) = k_{\mathrm{B}}T[h_{\mathrm{b}}(e(\tau)) - h_{\mathrm{b}}(e_0)].$$

Taking the repair duration to be negligible is a convention for the scheduling calculation. Theorem 11.2 minimizes work without a bound on the physical time needed for equilibration. A finite-time device would require a rate model that accounts for the repair time. The mathematical schedule below uses a separate passive-time coordinate.

For $0 < e_0 < 1/2$, w is increasing and strictly concave. For $t > 0$,

$$w''(t) = k_{\mathrm{B}}T\big[h_{\mathrm{b}}''(e(t))\,\dot{e}(t)^2 + h_{\mathrm{b}}'(e(t))\,\ddot{e}(t)\big] < 0$$

because $h_{\mathrm{b}}''(e) = -1/[e(1-e)]$, $h_{\mathrm{b}}'(e) = \log((1-e)/e) > 0$, and $\ddot{e} < 0$. Also $w(0) = 0$, hence $w(t)/t$ is decreasing. This follows from concavity by comparing $t_1 = \theta t_2$ to the convex combination of 0 and t_2. The limit at zero is

$$\lim_{\tau \downarrow 0} \frac{w(\tau)}{\tau} = k_{\mathrm{B}}T\gamma(1 - 2e_0)\log\frac{1 - e_0}{e_0}.$$

The limiting slope diverges as $e_0 \downarrow 0$, while the cost of a fixed refresh interval stays finite. The divergence concerns maintaining an arbitrarily accurate state under arbitrarily frequent correction in this model.

Proposition 13.2 (Optimal passive interval under a uniform error cap)**.** *Assume open-loop intervals, identical restored crossover e_0, and a maximum permitted instantaneous crossover $e_* \in (e_0, 1/2)$. Allow each repair the work infimum above and neglect its duration in the passive-time metric. Then the least long-run work per unit passive time is $w(\tau_*)/\tau_*$, where*

$$\tau_* = \frac{1}{2\gamma} \log \frac{1 - 2e_0}{1 - 2e_*}.$$

Proof. Since $e(t)$ increases, every interval must have length at most τ_*. Concavity and $w(0) = 0$ give $w(t) \geq t\, w(\tau_*)/\tau_*$ on that interval. Summing gives the lower bound on completed intervals. A bounded last partial interval does not affect the long-run average. Periodic intervals of length τ_* and repairs arbitrarily close to their infima approach the bound. □

The scheduling result maintains the one-time joint error distribution at each modeled time. It does not give positive probability of success at all checks forever. Repeated checks may have a strictly positive conditional hazard, in which case the path survival probability vanishes as in chapter 6. A specification of maintenance should state whether it bounds endpoint error, a worst-case instantaneous law, or a pathwise survival event.

13.4 Noisy repair with an imperfect reference

Suppose the controller has a side register C produced by a binary symmetric channel from X, with crossover $u \in [0, 1/2]$. A fresh output $Y = C \oplus N$ with independent crossover e has error $u + (1 - 2u)e$ relative to X. If C is the only information available at the repair site, no controller can beat the Bayes error u at that site. For each c, the most likely x is c with probability $1 - u$; extra randomized decisions cannot improve this. The conditional work infimum for writing Y from a thermal bit with reference C retained is $k_{\mathrm{B}}T[\log 2 - h_{\mathrm{b}}(e)]$. Thus a stringent error requirement relative to X can be causally unattainable at a site, even with arbitrarily large work for operations conditioned on C.

If the goal is merely to refresh a noisy target relative to C, the earlier repair theorem uses its crossover with C directly. If the full energetic model includes X, C, Y and the initial and final distributions satisfy $X \perp Y \mid C$, the difference in full free energy due to changing Y is the same conditional entropy difference: the law on X, C is fixed. This is an example of declaring the correct immutable reference instead of inferring one from a target marginal.

13.5 Persistent invariants and passive storage

Let a passive kernel preserve a label $d : S \to L$ on every positive transition. Then an initial state supported on $d(s) = x$ remains in that set and has zero decoding error at every time. A nontrivial such label implies the kernel is reducible on the corresponding communicating classes. When the energy is fixed and no drive is used, maintaining that label in this model requires no work after preparation. A trivial instance is an ideal isolated register. A less trivial instance is a random permutation of several carriers preserving their joint parity. An observer of the entire array reads the parity indefinitely while a designated one-bit carrier can lose all record of its original value.

For a finite irreducible aperiodic Gibbs-stationary kernel, the law tends to its unique Gibbs law from any initial state. If the immutable reference is held separately and the same irreducible memory kernel acts for every reference value, the conditional memory laws tend to a common equilibrium. Then $\mathsf{I}(X; Y_t) \to 0$ by continuity on the finite simplex. Passive storage can be metastable for long times if mixing is slow. Permanent retention of a nonconstant label requires an invariant, active repair, or a change of the observation interface. The theory supplies a ledger for particular noise and repair models, rather than a universal cost per unit of persistence.

CHAPTER 14

Causal designs for storing a label

We now evaluate three different realizations of a finite record. The examples use the same contract language and make the difference between regional compatibility, causal access, and energetic cost explicit.

14.1 A fork with a common noisy trunk

Let X be fair. A trunk bit is $C = X \oplus M$, with M independent of X and crossover $u \in [0, 1/2]$. At r receiving events, put

$$Y_i = C \oplus N_i,$$

where N_i are independent of each other and of X, M, with crossover $e \in [0, 1/2]$. The graph is a fork through C, and it can be placed by taking a common relay in the causal future of the source and the past of every receiver. The errors relative to X are correlated through M. The error probability at each receiver is $s = u + (1-2u)e$. The cut C carries $\log 2 - h_{\mathrm{b}}(u)$ nats about X. Every collection of receiver readouts has at most that amount of information about X.

Conditioning on C, receiver entropy is $r h_{\mathrm{b}}(e)$. Conditioning on X instead gives

$$\mathsf{H}(Y \mid X) = h_{\mathrm{b}}(u) + r h_{\mathrm{b}}(e) - \mathsf{H}(C \mid X, Y). \tag{14.1}$$

To derive it, expand the joint conditional entropy of C, Y in the two possible orders. The extra $h_{\mathrm{b}}(u)$ enters because the common trunk noise is unobserved when only X is given. In particular

$$\mathsf{TC}(Y \mid X) = r h_{\mathrm{b}}(s) - h_{\mathrm{b}}(u) - r h_{\mathrm{b}}(e) + \mathsf{H}(C \mid X, Y).$$

This residual vanishes at $u = 0$ and can grow with the number of receivers when the trunk has its own noise. If $e = 0$ and $u > 0$, the receivers are all exactly identical to the trunk and their common residual conditional on X is $(r-1)h_{\mathrm{b}}(u)$.

The best observer with access only to Y can infer C increasingly accurately as r grows for $e < 1/2$. Indeed, majority decoding of C from independent errors has error tending to zero by the law of large

numbers. If $u < 1/2$, the optimal error for X is bounded below by u by the Markov relation X, C, Y. A majority estimate of C gives an estimate of X with error tending to u. The redundancy is useful against independent errors at the receivers, while a common upstream error cannot be removed by those extra receivers.

Suppose the trunk C is retained during preparation of receiver bits. Start the Y_i as fair independent targets and hold the X, C law fixed. At flat endpoint energies the full entropy drop is $r(\log 2 - h_{\mathrm{b}}(e))$. It can be approached with separate controlled equilibrations given C, in the class of theorem 11.2. The apparatus that prepares C from X has its own initial resources and ledger. If one sums the two pieces, the work bound uses a full boundary containing the trunk wherever it is retained or charges its later reset. If one ignores C entirely, (14.1) computes the correlation change visible in the coarse X, Y observation. It generally gives a different free-energy difference.

14.2 An erasure archive on a field

Let $\mathbb{F}_q$ be a finite field, choose distinct evaluation points $a_1, \dots, a_n$, and let $1 \le k \le n \le q$. Take the label $X = (X_0, \dots, X_{k-1})$ uniform on $\mathbb{F}_q^k$. The i-th carrier receives

$$Y_i = \sum_{j=0}^{k-1} X_j a_i^j.$$

Any k distinct evaluations determine the polynomial of degree less than k: a nonzero difference polynomial has at most $k-1$ roots. The corresponding linear map has k-by-k full rank. For $t < k$, any t evaluation rows are linearly independent, since the first t monomial columns form an invertible Vandermonde matrix on those points. Thus a tuple of t surviving carriers is uniform on $\mathbb{F}_q^t$. For $t \ge k$ it is uniform on a k-dimensional image.

It follows that for each subset S of size $t \ge 1$,

$$\begin{aligned} \mathsf{H}(Y_S) &= \min(t,k) \log q, \\ \mathsf{H}(X \mid Y_S) &= (k - \min(t,k)) \log q, \\ \mathsf{TC}(Y_S) &= (t - \min(t,k)) \log q. \end{aligned}$$

Each single share has $\log q$ nats of entropy, and an authorized k-set has the full $k \log q$ nats of source information. If $k > 1$ the best exact guess based on one share succeeds with probability $q^{-(k-1)}$, since its posterior is uniform over a fiber of size q^{k-1}. This archive preserves the source against loss of up to $n - k$ carriers. It does not require individual recoverability from each surviving carrier.

Suppose carrier survival is independent with probability s. The chance of enough surviving data is

$$\sum_{j=k}^{n} \binom{n}{j} s^j (1-s)^{n-j}.$$

If failures are correlated, the distribution of the surviving set replaces the binomial law. A common upstream physical hazard can defeat all carriers at once. In addition, reconstruction at a repair event requires at least k usable shares accessible in its causal past, or other side information that fills the missing rank. For a set of observed linear combinations with matrix A, an exact decoder of the entire uniform label exists precisely when A has full column rank. This rank test is an instance of the cut condition where entropy equals the rank times $\log q$.

If X is retained and the n new q-state targets are initially uniform independent with flat energies, the final exact codeword is deterministic given X. The full entropy drop is $n \log q$. It costs the same endpoint free energy to write any deterministic string of n symbols from these thermal targets while retaining X, even one carrying no source

information. In the erasure archive the use of those symbols is specified by the reconstruction contract. Repeating the entire k-symbol label at each carrier would use nk target symbols and, for the same initial condition, cost $nk\,k_{\mathrm{B}}T\log q$. The larger boundary provides a stronger individual-decoding contract. Prepared pure targets instead permit a controlled addition as in (5.1), storing the codeword reversibly with the source.

14.3 Noisy recursive regeneration

For another design, let a binary output gate read three observations of a source. Given the source, suppose the three observations have independent error probability e. The majority is wrong if exactly two or all three are wrong, giving $g(e) = 3e^2 - 2e^3$. Suppose the majority output itself is flipped independently with probability $q \in [0, 1/2]$. The new error is

$$T_q(e) = q + (1 - 2q)(3e^2 - 2e^3).$$

A depth-n formula with three disjoint subformulas at each internal gate can realize this recurrence if the leaf noises and all gate noises are conditionally independent given the reference. A causal placement must allow the three inputs to converge at each majority gate, and the initial leaf observations must be generated with the stated independence. The recurrence does not hold if the three inputs are obtained by copying one noisy predecessor.

Let $m = 1 - 2e$ be the input bias toward correctness. In these coordinates the update is

$$m \longmapsto (1 - 2q)\frac{3m - m^3}{2}.$$

For $q < 1/6$, the positive fixed point is $m_* = \sqrt{3 - 2/(1 - 2q)}$. For $q = 0$ this is 1. The map is increasing on $[0, 1]$, and the difference between output and input is $m((1 - 2q)(3 - m^2)/2 - 1)$. Thus an initial bias $m > 0$ increases toward m_* from below or decreases toward it from above. For $q \geq 1/6$ the difference is negative for every $m > 0$, and iteration converges to zero. This establishes a threshold for this particular formula family. Reliable computation under general gate noise has additional constraints on circuit size and depth [26].

Preparing more independent noisy inputs or refreshing them repeatedly uses physical resources. Even when the formula improves endpoint error, the entropy and work of its intermediate observations are part of its ledger. A majority function realized as an irreversible overwrite can lose correlations among its inputs. It can instead be implemented reversibly with initialized scratch and output registers. The protocol choice determines whether those correlations are recovered or dissipated.

14.4 Comparing the contracts

The fork, the erasure archive, and the majority formula make three different uses of additional carriers. In the fork each new carrier approximates the whole label. The common trunk constrains all of them. In the archive no one share is sufficient for $k > 1$, and persistence is measured by the survival of an authorized set with accessible rank. In the formula the intermediate carriers are used jointly to reduce independent errors, with conditional independence as an essential premise. A copy count alone cannot identify which of these mechanisms is present. The history site specifies the local constraints, the causal graph specifies availability, and the full ensemble and energy supply the accounting.

CHAPTER 15

Classical labels at a quantum interface

A finite classical label can be represented by orthogonal pointer states, or measured from a quantum system and stored in a classical register. Copying the record of a measurement need not copy the premeasurement state. The distinction can be expressed operationally using quantum channels and quantitatively using the entropy of the full output state.

15.1 Copying an orthogonal label

Let $\mathcal{H}_A$ have a pointer basis $\{|x\rangle : x \in L\}$. Give each target a copy of that Hilbert space and an initialized basis state $|0\rangle$. After identifying the basis with the elements of a cyclic group, a unitary controlled addition sends $|x\rangle|0\rangle$ to $|x\rangle|x\rangle$. On an ensemble diagonal in that basis it implements exact copying of the classical label. The entropy of the full state is preserved by the unitary. A thermodynamic realization needs prepared targets and an energy model, as in the classical discussion.

On a coherent superposition $\alpha|0\rangle + \beta|1\rangle$, the same gate gives $\alpha|00\rangle + \beta|11\rangle$. Each marginal is diagonal in the pointer basis, and in general the joint state is entangled. It is not the tensor square of the original pure state. For a fair bit, the diagonal correlated mixture and the pure equal-amplitude state have the same computational-basis joint observation. They have different von Neumann entropies. The mixture has entropy $\log 2$, while the pure state has zero entropy. Their total correlations across the two carriers are $\log 2$ and $2 \log 2$, respectively. A classical observation therefore need not account for all available physical free energy.

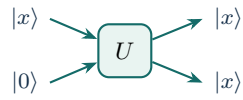


Figure 15.1: Controlled addition copies a pointer-basis label into a prepared target. Its action on a superposition is determined by linearity.

15.2 An obstruction to universal copying

Proposition 15.1 (Exact copying of pure states)**.** *If one channel sends each of two pure input states to the tensor square of that same pure state, then the states are identical or orthogonal [34].*

Proof. Use an isometric dilation of the channel. For input $|\psi\rangle$, the output on the two carriers is pure, so its purification in the dilation has the form $|\psi\rangle|\psi\rangle|e_\psi\rangle$ with a unit environment vector. The same holds for $|\phi\rangle$. An isometry preserves inner products. Writing $s = |\langle\psi|\phi\rangle|$, we obtain

$$s = s^2 |\langle e_\psi | e_\phi \rangle| \leq s^2.$$

For $0 \leq s \leq 1$ this permits only zero and one. These cases correspond to orthogonal states and the same pure state. □

Broadcasting asks for equal reduced states on two carriers, permitting correlations. Pure marginals force a product, so broadcasting pure states is already subject to the same argument. For a finite family of commuting density matrices, choose a common eigenbasis and copy its index into two outputs, discarding the input coherences if necessary. The resulting output for a diagonal input with weights p_i is $\sum_i p_i |ii\rangle\langle ii|$, with both marginals equal to the input. The finite-dimensional no-broadcasting theorem states the converse: a family of density matrices broadcastable by one channel must commute [3]. The

classical replication object of chapter 3 is appropriate for a record after its label has been specified. A coherent quantum input requires a different state object and a different class of allowed copying operations.

15.3 Local readouts and the energetic boundary

For finite-dimensional systems with additive endpoint Hamiltonian $H = \sum_i H_i$, define

$$S(\rho) = -\operatorname{Tr} \rho \log \rho,$$
$$F_H(\rho) = \operatorname{Tr}(\rho H) - k_{\mathrm{B}} T S(\rho).$$

The quantum Gibbs law is $\gamma_H = e^{-\beta H}/Z_H$. Expanding quantum relative entropy gives $F_H(\rho) = F_{\mathrm{eq}}(H) + k_{\mathrm{B}} T D(\rho \| \gamma_H)$. Also

$$F_H(\rho) = \sum_i F_{H_i}(\rho_i) + k_{\mathrm{B}} T T_q(\rho),$$
$$T_q(\rho) = \sum_i S(\rho_i) - S(\rho) = D(\rho \| \bigotimes_i \rho_i).$$

The support needed in the last expression holds since a positive state has zero mass on the orthogonal complement of the support of any marginal, hence lies in the product of marginal supports. A product of local measurement channels takes the reference $\bigotimes_i \rho_i$ to the product of the outcome marginals. Monotonicity of quantum relative entropy then bounds the classical total correlation of those outcomes by $T_q(\rho)$ [20]. The classical decoding theorem applied to the outcomes lower bounds the quantum correlation term when the source is a specified classical label in a joint classical–quantum ensemble.

The same stage calculation as in theorem 10.1 works for finite quantum systems if quenches preserve ρ and each stochastic step is replaced by a channel preserving its Gibbs density matrix. The entropy contraction used in the proof is quantum data processing. This gives a free-energy lower bound for a declared quantum protocol class. Work costs of a particular measurement device still require its ancillas, operations, and reset boundary. A mathematical completely positive map can represent many devices with different costs.

15.4 Spacelike observations and shared preparations

On a bipartite state ρ_{AB}, a trace-preserving channel on A leaves the reduced state on B unchanged. For each observable O_B, the new expectation is the old expectation of $\Phi_A^*(I_A) \otimes O_B$, which equals $I_A \otimes O_B$ because the adjoint is unital. If a source chooses between local channels at A, a B-only measurement cannot infer that choice through these channels. A selection conditioned on a measurement outcome at A can change the conditional state at B; using that conditioned information requires access to the outcome. This is the quantum analogue of keeping shared preparation and declared communication distinct.

The manuscript uses this finite setting to identify a boundary of the classical record model. Full relativistic quantum field theory would replace finite tensor factors and local density matrices with the appropriate algebras, states, and locality conditions. Under those hypotheses even defining energy and entropy of sharply bounded regions takes further work [6]. The classical finite causal and thermodynamic results above are statements about the chosen readouts and the finite energetic state on which the protocol acts.

CHAPTER 16

Fibrations of histories and continuations

A family of history spaces can be represented as a fibration over contexts. This

keeps the data of a history together with the region on which it is a history, and distinguishes restrictions from time evolution at the level of arrows.

16.1 The right fibration of observations

Let $\mathcal{C}$ be a site, regarded as an ∞-category, and let $\mathcal{H} : \mathcal{C}^{\mathrm{op}} \to \mathcal{S}$ be the history sheaf. By the straightening correspondence [21, 27], it classifies a right fibration

$$p : \int_{\mathcal{C}} \mathcal{H} \longrightarrow \mathcal{C}.$$

The fiber at U is $\mathcal{H}(U)$. An object over V is a history $h \in \mathcal{H}(V)$, and its cartesian lift over an inclusion $f : U \to V$ starts at f^*h. More explicitly, the space of arrows over f from (U, a) to (V, h) is

$$\mathrm{Map}_{\mathcal{H}(U)}(a, f^*h).$$

Here the right side is a path space in $\mathcal{H}(U)$. A specific arrow over f therefore includes a chosen comparison. The cartesian arrow corresponds to the identity path on f^*h.

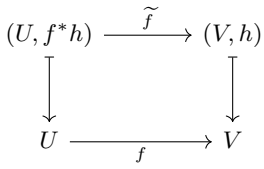


Figure 16.1: A cartesian lift of a region inclusion transports a section by restriction. The arrow in the total category has its domain over the smaller region.

For a third region T and $g : T \to U$, the comparison between g^*f^*h and $(fg)^*h$ belongs to the functoriality data. It determines the 2-simplex comparing the composite of the two lifts with a cartesian lift over fg. In a strict presheaf of sets, the comparison is the equality of two computed restrictions. In a stack of field configurations it can be a nontrivial equivalence between representatives. Composable triples require the higher comparison, supplied by functoriality of the sheaf as a diagram of spaces.

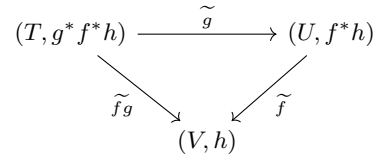


Figure 16.2: The filled triangle compares the composite lift with the lift of fg. Its top-left object uses the canonical equivalence with $(T, (fg)^*h)$.

16.2 The relative fibration over observed histories

An observation map $o : \mathcal{H} \to B$ gives a map of right fibrations over $\mathcal{C}$. Fixing a point (U, b) in the total space of B, the homotopy fiber over that point is $\mathscr{R}_o(U, b)$. As (U, b) varies, these fibers form a presheaf on $\int_{\mathcal{C}} B$. For an arrow given by $f : U \to V$ and a comparison $\alpha : b \simeq f^*c$, restrict a history over c along f and compose its observed-boundary path with α^{-1}. The chosen identification is used in the construction. Two different paths α can act differently on the reconstruction fiber.

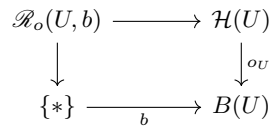


Figure 16.3: At a fixed region and observed value, reconstruction is the pullback of the history map along the point b. A point in the top-left corner includes the path represented by this square.

For instance, at a point take $o : * \to BG$.

The homotopy fiber at the unique basepoint is $\Omega BG \simeq G_{\mathrm{disc}}$. An automorphism g of the observation acts by translating the loop label, with a left or right action according to the chosen path convention. The fiber has $|G|$ elements for finite G, though the domain of o has only one point. The multiplicity here is multiplicity of compatible identifications, not multiplicity of distinguishable physical carriers. Quotienting the observation groupoid before taking the fiber would change the result.

16.3 Forward systems over causal paths

A deterministic system on a path category $\mathcal{P}$ can be represented by a covariant functor $F : \mathcal{P} \to \mathbf{Set}$, sending a path to a state transition. Its category of elements over $\mathcal{P}$ has cocartesian arrows $(a, s) \to (b, F(f)s)$ over a path $f : a \to b$. The arrow is now a forward extension: its domain is supplied and its codomain is determined by the transition. A covariant functor to spaces similarly defines a left fibration. With arbitrary categories of states and interventions one can use a cocartesian fibration with those fibers.

An observation fibration and a forward fibration are related by a space of transition histories, represented by a span between boundary objects. This relation need not turn a conditional law into a section of the observation fibration. A random transition gives a law on extensions, not a chosen target. One representation takes the finite space of input, seed, and output triples satisfying the gate rule. The projection to input forgets seed and output. A distribution of seeds conditional on input equips that projection with a stochastic lift.

16.4 A triangle for overlap coherence

For a cover U_0, U_1, U_2, take local sections h_i with identifications α_{ij} between their restrictions on pairwise intersections. On U_{012} every one of the three sections restricts again. One needs a 2-simplex whose edges are the restricted comparisons, oriented so that the direct comparison and the two-step comparison have the same endpoints.

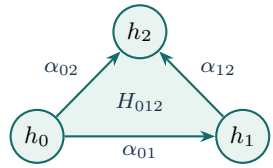


Figure 16.4: A triple-overlap 2-simplex witnesses $\alpha_{12}\alpha_{01} \simeq \alpha_{02}$. All objects and arrows in the picture have been restricted to U_{012}.

If the fiber is a groupoid, H_{012} is the equality of composites. If it is a general space, there may be distinct fillers. On quadruple overlaps, those fillers must themselves satisfy a compatibility represented by a 3-simplex. Equivalence of two families is again coherent across all overlaps. This is the content of totalization in (3.2). It can be calculated by replacing the cosimplicial diagram with a fibrant model and solving successive matching equations, as described in the diagram appendix.

16.5 Changing coordinates in a persistence cone

For a tower $R_{n+1} \to R_n$, a limit point consists of compatible choices of r_n and paths $h_n : r_n \to \rho_n(r_{n+1})$. A path between two limit points consists of paths $a_n : r_n \to r'_n$ and higher homotopies making the squares with h_n, h'_n commute.

If R_n is a groupoid, those higher homotopies are equations between arrows. In the tower of $K(\mathbb{Z}, 1)$ with doubling maps in chapter 6, the labels of the squares satisfy $b'_n - b_n = \pm(a_n - 2a_{n+1})$, depending on orientations. This makes the derived-limit component calculation a compatibility condition on comparisons rather than on the levelwise objects.

$$\begin{array}{ccc} r_n & \xrightarrow{h_n} & \rho_n r_{n+1} \\ {\scriptstyle a_n}\downarrow & & \downarrow{\scriptstyle \rho_n a_{n+1}} \\ r'_n & \xrightarrow[h'_n]{} & \rho_n r'_{n+1} \end{array}$$

Figure 16.5: A path between coherent continuations includes a filler of this square at each level. Merely choosing the vertical paths does not supply the filler.

16.6 Naturality as a comparison of restrictions

Let $F, G : \mathcal{C}^{\mathrm{op}} \to \mathcal{S}$ be two history theories, and let $\eta : F \to G$ be a map. It induces a functor over $\mathcal{C}$ preserving cartesian arrows. For an inclusion $u : U \to V$, the naturality datum is the equivalence between first restricting in F and then applying η_U, or first applying η_V and then restricting in G.

$$\begin{array}{ccc} F(V) & \xrightarrow{\eta_V} & G(V) \\ {\scriptstyle u^*}\downarrow & & \downarrow{\scriptstyle u^*} \\ F(U) & \xrightarrow[\eta_U]{} & G(U) \end{array}$$

Figure 16.6: Naturality supplies the homotopy between the two composites. Under unstraightening it compares the images of cartesian lifts.

For a chosen $h \in F(V)$, this homotopy produces an arrow over u whose domain is equivalent to $u^*\eta_V(h)$. The mapping-space property in chapter C then makes that arrow cartesian. A collection of maps on separate fibers with no such naturality need not preserve compatible continuations. For example, a map on each fiber of a two-sheeted local system that selects one sheet and discards the other could fail to commute with monodromy. It would not define a map of the local systems, even though the set map at each point exists. The higher comparisons are what make observation and restriction commute on all contexts.

CHAPTER 17

Information across spacetime screens

The cut theorem bounds one interface of a causal computation. A sequence of screens gives a related conservation calculation: the unique source information at an outward-moving cut can only decrease as fresh conditional noise is introduced, while local replicas and redundancy may grow. We formulate the cut-to-cut process and connect it with the time at which two spatial observations can first be read jointly.

17.1 A cut as the state of wires in transit

Use the finite directed acyclic model of chapter 4, with each transmitted symbol attached to a particular edge. A vertex reads symbols on its incoming wires and generates its outgoing wire tuple using a kernel. An edge may retain its symbol for several intermediate screens. The full joint law follows by sampling each vertex once after its parents. Fix a past ideal containing all source inputs other than declared conditional-independent outside data. Let X be the source label and W a side context. For each larger past ideal I, define C_I as the tuple of variables on edges starting in I and ending outside I. Include a wire terminating at a later observer whenever its message is still available to that observer. If an observer has a retained copy from an earlier round, that memory is a wire or a state input in this model.

Take ideals $I \subseteq J$ above the source. Some wires persist on both cuts. Other wires from I terminate in $J \setminus I$ and are read by gates there. All inputs to gates in $J \setminus I$ are already on the I cut, generated earlier inside $J \setminus I$, or supplied from independent exogenous data conditional on W. No edge can enter $J \setminus I$ from the complement of J, since J is an ideal. This makes it possible to evaluate just the layer between the cuts.

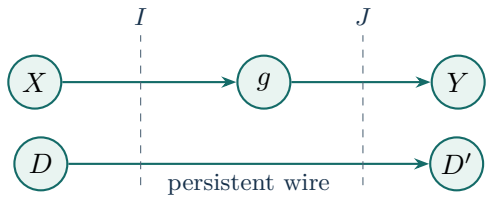


Figure 17.1: A cut-to-cut transition carries through the lower wire and updates the upper wire. An old carrier still accessible in the future must remain represented on the outgoing cut.

Theorem 17.1 (Source information across nested screens)**.** *For $I \subseteq J$ as above, the joint law of X, C_I, C_J, W satisfies $X \perp C_J \mid C_I, W$. Therefore*

$$\mathsf{I}(X; C_I \mid W) - \mathsf{I}(X; C_J \mid W) = \mathsf{I}(X; C_I \mid C_J, W). \tag{17.1}$$

For a chain of nested screens the total decrease is the sum of these conditional losses at each step. Equality of the information at two screens holds exactly when the posterior of X on the first cut agrees with the posterior on the second, almost surely on joint cut observations.

Proof. Evaluate the between-screen layer conditional on (C_I, W) and retain the values of shared wires. The independence assumption on the additional exogenous data and the Markov factorization give a kernel from (C_I, W) to C_J independent of X. Expand the conditional information of X with the two cuts in the two orders, as in the cut theorem. This gives (17.1), whose sum telescopes. By the factorization, $p(X \mid C_I, C_J, W) = p(X \mid C_I, W)$. The conditional mutual information on the right of (17.1) is the expectation of the divergence from that posterior to $p(X \mid C_J, W)$. Its zero condition is equality on the joint support. □

A gate that advances the screen by one event has incoming tuple A, outgoing tuple B, and an unchanged shared part S of the cut. Then its information loss is $\mathsf{I}(X; A \mid B, S, W)$. Indeed $C_I = (A, S)$ and $C_J = (B, S)$, omitting irrelevant constant symbols. This expression may be zero for a deterministic many-to-one gate if the discarded distinctions do not change the source posterior once B, S, W are known. A source-preserving reversible rearrangement with all outputs retained always has zero loss. If the source-independent noise of a gate is discarded after use, the loss reflects what can still be inferred from the surviving wire tuple.

The sum of losses depends only on the bounding cuts and their joint evaluation, whereas an individual term can depend on which other wires share the cut at that step. For instance consider two exact copies of a fair bit on the same initial cut. If both wires are independently randomized, the first randomization leaves source information $\log 2$ on the other wire. The second randomization eliminates it. The same two operations can also be evaluated in the opposite order. Conditional loss is then assigned to the opposite gate. This order dependence concerns decomposition of a fixed information difference. Under the work commutation hypotheses of chapter 19, total work is also schedule-independent.

17.2 Transport through Cauchy levels

Take a smooth Cauchy temporal function t in a globally hyperbolic spacetime and represent each wire by a future-directed causal path between two placed events. For a regular level $t = \tau$ avoiding vertex times, let I_τ contain the vertices with $t(v) < \tau$. The strict increase of t along each nonconstant causal path implies that I_τ is an ideal. A path crosses the level exactly once when its two endpoint times straddle it, by continuity and strict monotonicity. Thus the finite crossing tuple is an observation on the screen. A model may subdivide a wire if its symbol actually evolves during transit. In that case the coordinate at a screen represents the state after the last event in the subdivided trace rather than a fixed label along the whole wire.

If $\tau < \tau'$ and both cuts are above the source, theorem 17.1 applies. The resulting finite inequality uses only the nested ideals and the permitted wires. A second temporal function can define a different family of cuts. When two levels for the two functions induce the same ideal, their crossing tuple labels are the same up to reindexing, given the convention of a fixed wire symbol. When the ideals are nested, the same monotonicity applies. For incomparable ideals no direction of inequality follows. One ideal can have received an intact signal on a protected wire while another has passed a noisy copy and lost the original. The comparison concerns causal exposure of registers rather than a scalar field of information defined on the ambient manifold.

If a future region must decode X with error e, every preceding screen across which its usable input flows has $\mathsf{I}(X; C_I \mid W) \geq \mathsf{H}(X \mid W) - f_m(e)$. Consider a family of devices with more screens and uniform required error below the guessing error of a uniform source. If the remaining cut information tends to zero along those screens, this inequality forbids a uniform decoder that always reads only their future. Thus recovery can fail even when a compatible support history of exact copying persists through all screens.

17.3 When two separated records become jointly usable

In Minkowski space with $c = 1$, place two readout events at (t_1, x_1), (t_2, x_2), with $x_i \in \mathbb{R}^d$. Assume signals can be sent along straight null paths and arbitrary spatial gathering locations are allowed. At a location x, a future event receiving both data can occur no earlier than

$$t(x) = \max\{t_1 + \|x - x_1\|,\ t_2 + \|x - x_2\|\}.$$

This follows from the two future-cone conditions, and the bound is attained in the ideal zero-latency receiver model by waiting for both signals. Set $d_{12} = \|x_1 - x_2\|$.

Proposition 17.2 (Earliest joint readout)**.** *The earliest time over gathering locations is*

$$\min_x t(x) = \max\left\{t_1, t_2, \tfrac{1}{2}(t_1 + t_2 + d_{12})\right\}.$$

If the readout events are spacelike separated, it is achieved on the line segment between x_1 and x_2, where the two earliest arrivals coincide.

Proof. The two endpoint times are lower bounds. The triangle inequality also gives $2t(x) \geq t_1 + t_2 + d_{12}$. If $t_2 \geq t_1 + d_{12}$, the location x_2 gives time t_2, attaining the maximum; the reverse case is symmetric. Otherwise $d_{12} > |t_2 - t_1|$. Choose the point on the segment at distance $(d_{12} + t_2 - t_1)/2$ from x_1. Its distances to the two endpoints sum to d_{12}, and the two arrival times at it are equal to the asserted lower bound. □

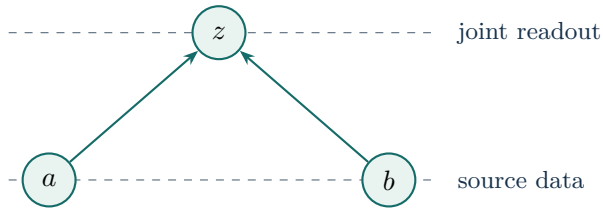


Figure 17.2: For simultaneous spacelike readings, the earliest event in both future cones is at their midpoint after half the light-travel time between them.

A threshold code can therefore have enough surviving shares and still miss a deadline. For example, let the message be $X \in \{0, 1\}$, split as R and $X \oplus R$ for independent fair R, and place the two shares at $t_1 = t_2 = 0$ and positions separated by distance $d_{12} > 0$. Assume no other data about X are available beyond those sites and signals derived from them. Before time $d_{12}/2$, an observer can be in the future of at most one share. That share is marginally independent of X, so its best success is $1/2$ by the cut theorem. At the earliest joint gathering event the two messages decode exactly. No additional processing at a single earlier site changes the information bound. A local exact copy available at either site would meet a strictly earlier local deadline. This is a physical distinction between complete replication and distributed recoverability.

17.4 Information loss and heat depend on the boundary

Equation (17.1) is a statement about a source posterior. A local channel can lose information about X while costing no work at fixed flat energy, for example by

independently randomizing a fair target already correlated with X. If X remains as a physical spectator, the step increases joint entropy by the lost correlation and the full nonequilibrium free energy falls. When both marginals are unchanged and the spectator is exactly X, this drop equals $k_{\mathrm{B}}T$ times the information loss. If the memory marginal also changes, the marginal free-energy term must be added. If the spectator is inaccessible and is later needed, the local gate cannot restore the correlation it has lost using its output alone. These are particular energetic realizations of the screen identity, with the precise assumptions supplied by (10.4) and theorem 19.1.

In an energetic model a cut may be used to choose which degrees of freedom are retained at an interface. Boundary laws at different screens need not live on the same product state space. To use the work ledger directly, embed the gates in a common energetic state, including prepared incoming registers and unused outgoing registers or an explicit account of changing system boundaries. The screen information formula itself does not require this embedding. The two formulations allow a causal statement on a small interface and a separate full-system accounting of the protocol implementing it.

CHAPTER 18

Continuous-time storage and finite activity

The discrete ledger gives a work infimum over protocols with no duration constraint. To relate it to a physical timescale, one can add bounded transition activity. We use a finite continuous-time jump process with local detailed balance. This section supplies both a mixing result for passive memory and a lower bound on dissipation during a finite-time repair.

18.1 Currents and entropy production

Let S be finite and let $k_{ij}(t) \geq 0$ for $i \neq j$ denote the transition rates from i to j. Assume the law $p(t)$ is positive on the modeled state space throughout the calculation, and that the rates and energy are regular enough for the differential equations and integrals below, for example bounded piecewise continuous rates and a piecewise C^1 finite energy with separate quenches at jumps. Instantaneous detailed balance means

$$\pi_i(t)k_{ij}(t) = \pi_j(t)k_{ji}(t), \qquad \pi_i(t) = Z(t)^{-1}e^{-\beta E_i(t)}.$$

For an unordered pair $i < j$, put $a_{ij} = p_i k_{ij}$, $b_{ij} = p_j k_{ji}$, $J_{ij} = a_{ij} - b_{ij}$. Rate-zero pairs contribute zero. By detailed balance, an active pair has both fluxes

positive under the support assumption. The master equation expresses $\dot{p}$ as the sum over pair currents, with $-J_{ij}$ at i and $+J_{ij}$ at j.

Let

$$\sigma(t) = \sum_{i<j} (a_{ij} - b_{ij}) \log \frac{a_{ij}}{b_{ij}} \geq 0. \tag{18.1}$$

The nonnegativity holds termwise. The drive work rate is $\dot{W} = \sum_i p_i \dot{E}_i$. The energy flow due to the jumps is $\dot{Q} = \sum_i \dot{p}_i E_i$. Differentiating entropy using $\sum_i \dot{p}_i = 0$ and pairing the current contributions gives

$$\begin{aligned} \frac{d}{dt}\mathsf{H}(p(t)) - \beta\dot{Q} &= \sum_{i<j} J_{ij}\big(\log(p_i/p_j) - \beta(E_j - E_i)\big) \\ &= \sigma(t). \end{aligned}$$

The last equality uses $\log(k_{ij}/k_{ji}) = -\beta(E_j - E_i)$. Therefore

$$\frac{d}{dt}\mathcal{F}_{E(t)}(p(t)) = \dot{W} - k_{\mathrm{B}}T\,\sigma(t).$$

Integrated with the quench contributions, this is the continuous-time counterpart of (10.2). For fixed energy and no drive, it says $-\frac{d}{dt}\mathsf{D}(p(t)\|\pi) = \sigma(t)$.

18.2 A finite-activity repair bound

Put $A(t) = \sum_{i<j}(a_{ij} + b_{ij})$. This equals the expected jump intensity, since it counts each directed rate contribution once. The integral $\mathcal{A} = \int_0^\tau A(t)\,dt$ is the expected number of jumps during the protocol. Quenches alone leave p fixed and give no activity. We allow them at finitely many instants.

Theorem 18.1 (Work above the endpoint cost under bounded activity)**.** *Let* $\Sigma = \int_0^\tau \sigma(t)\,dt$. *Then*

$$\|p(\tau) - p(0)\|_1^2 \leq 2\Sigma\mathcal{A}. \tag{18.2}$$

If $\mathcal{A} > 0$, *the protocol work satisfies*

$$W_{\mathrm{in}} \;\geq\; \Delta\mathcal{F} + \frac{k_{\mathrm{B}}T\,\|p(\tau) - p(0)\|_1^2}{2\mathcal{A}}.$$

If $A(t) \leq A_{\max}$ *with* $A_{\max} > 0$, *the excess is at least* $k_{\mathrm{B}}T\|p(\tau)-p(0)\|_1^2/(2A_{\max}\tau)$.

Proof. For $a, b > 0$, $(a - b)\log(a/b) \geq 2(a - b)^2/(a + b)$. To prove it set $z = (a - b)/(a + b)$. The logarithm is $2\,\mathrm{artanh}(z)$, and $z\,\mathrm{artanh}(z) \geq z^2$ by the integral of $1/(1 - u^2)$ or its power series. Summing and using Cauchy–Schwarz yields

$$\sum_{i<j} |J_{ij}| \leq \sqrt{A(t)\sigma(t)/2}.$$

The master equation bounds $\|\dot{p}(t)\|_1 \le 2\sum_{i<j}|J_{ij}|$. Integrate and apply Cauchy–Schwarz to $\int \sqrt{A(t)\sigma(t)}\,dt$. This proves (18.2). The work bound follows from the continuous ledger and the bound on $\mathcal{A}$. □

This is a finite-state speed limit of the form developed for stochastic processes in [32]. It constrains the work infimum when restoration must occur at finite activity. For the binary refresh law of theorem 13.1, the two pair laws at crossovers e_- and e_0 have 1-distance $2|e_- - e_0|$. Thus any such restoration incurs work at least the entropy drop plus

$$\frac{2k_{\mathrm{B}}T(e_- - e_0)^2}{\mathcal{A}}.$$

This term vanishes in an unbounded-activity limit. For fixed activity and a nontrivial restoration it is strictly positive. In particular, a proof of the quasistatic work infimum gives no finite-time attainment.

18.3 Passive relaxation and loss of a source

Now take a time-homogeneous reversible memory generator with positive equilibrium π and spectral gap $\lambda > 0$. For a law p, let $u_i = p_i/\pi_i - 1$, which has mean zero under π. Define

$$\chi^2(p,\pi) = \sum_i \pi_i u_i^2, \qquad \mathcal{D}_\pi(u) = \frac{1}{2}\sum_{i,j}\pi_i k_{ij}(u_j - u_i)^2.$$

Detailed balance implies the density u evolves under the same generator acting on functions. Differentiating its squared norm and pairing symmetric terms gives

$$\frac{d}{dt}\chi^2(p(t),\pi) = -2\mathcal{D}_\pi(u(t)).$$

The gap is a constant with $\mathcal{D}_\pi(u) \ge \lambda\sum_i \pi_i u_i^2$ for mean-zero u. Hence $\chi^2(p(t),\pi) \le e^{-2\lambda t}\chi^2(p(0),\pi)$. The bound $\mathsf{D}(p\|\pi) \le \chi^2(p,\pi)$ follows from $\log s \le s - 1$.

Suppose X is retained in another register and the same memory generator acts for every $X = x$, with no direct access to it. For each conditional initial memory law $p^x(0)$ we get the same decay. The decomposition

$$\sum_x p_X(x)\mathsf{D}(p^x(t)\|\pi) = \mathsf{I}(X;Y_t) + \mathsf{D}(p_{Y_t}\|\pi)$$

then proves

$$\mathsf{I}(X;Y_t) \le e^{-2\lambda t}\sum_x p_X(x)\chi^2(p^x(0),\pi). \tag{18.3}$$

Combined with conditional Fano, this limits the time for which a decoding contract can be met without repair. For a fair m-ary source and a target error

cap $e < 1 - 1/m$, the required information is $\log m - f_m(e) > 0$. Let B be the constant multiplying the exponential in (18.3). If B is at least the required information, a necessary condition for the cap is

$$t \le \frac{1}{2\lambda} \log \frac{B}{\log m - f_m(e)}.$$

If B is smaller, this bound rules out meeting the cap at any time, including time zero. The estimate can be coarse when the initial ensemble has mass on rare equilibrium states. It nevertheless separates memory mixing from pure reachability.

18.4 An energetic example with a biased bit

Let $E(0) = 0$, $E(1) = \Delta > 0$ and choose rates $k_{01} = a$, $k_{10} = b$, with $a/b = e^{-\beta\Delta}$. Then $\pi(1) = a/(a+b)$. A conditional probability $p^x(t) = \mathbb{P}(Y_t = 1 \mid X = x)$ satisfies

$$p^x(t) = \pi(1) + (p^x(0) - \pi(1))e^{-(a+b)t}.$$

For two reference labels the difference in their conditional probabilities decays at rate $a+b$, regardless of the equilibrium bias. A binary decoder using only Y_t cannot preserve source information at equilibrium. If the source input law is fair and the conditional probabilities are ordered, its optimal error is $\frac{1}{2}[1 - |p^1(t) - p^0(t)|]$; this follows by maximizing the two likelihoods at each of the two outcomes. In a repair ledger with these nonflat energies, the mean energy of the target can change as well as its conditional entropy. The loss decomposes into marginal equilibration and correlation destruction as in (10.4). Keeping the energy term is necessary even when the readout is binary.

18.5 Rate assumptions and possible mechanisms

A positive spectral gap describes mixing in the chosen finite memory model. Energy barriers can make the gap small and increase passive retention time. A conserved sector instead makes the global gap vanish when the equilibrium measure includes multiple sectors. A rate bound for repair concerns expected activity during the repair and can be imposed by engineering restrictions on transitions. Neither assumption follows from the spacetime site. The site orders potential causal access, while the rates, energies, and bath coupling give the physical constraints used by the speed and mixing results.

CHAPTER 19

Local gates and inaccessible correlations

A processor can act on a region without having access to all degrees of freedom correlated with its inputs. As a result, even a protocol that approaches the work infimum for the visible marginal can dissipate a definite amount in the full system. This distinction matters for spatially separated replicas and for memory repairs that can access some, but only some, of the surviving records.

19.1 A gate with a controlled interface

Let A be the finite state a module can update, B its immutable control state, and C an inaccessible, unchanged spectator. At every step assume the energy has the form

$$E(a, b, c) = E_{AB}(a, b) + E_{BC}(b, c),$$

with the second term fixed. A quench is allowed to change E_{AB}, and each local stochastic step acts only on A with a kernel controlled by b. Its conditional stationary law is Gibbs for $E_{AB}(\cdot, b)$. On the full state this also preserves the Gibbs law of E, because the conditional equilibrium on A given B, C depends only on B. The gate can have many stages with this property. The law of B, C stays fixed.

Writing $\mathcal{F}_{AB}$ for the free energy with energy E_{AB} evaluated on the AB marginal, the chain rule gives

$$\mathcal{F}_E(p_{ABC}) = \mathcal{F}_{AB}(p_{AB}) + \mathbb{E}_p E_{BC} - k_{\mathrm{B}}T\mathsf{H}(C \mid B) + k_{\mathrm{B}}T\mathsf{I}(A; C \mid B). \quad (19.1)$$

Indeed, $\mathsf{H}(ABC) = \mathsf{H}(AB) + \mathsf{H}(C \mid AB)$ and $\mathsf{H}(C \mid AB) = \mathsf{H}(C \mid B) - \mathsf{I}(A; C \mid B)$. This decomposition allows interactions within AB and BC. If a direct A, C interaction at an endpoint is present, its energetic contribution would have to be added, and the same local Gibbs stationarity would generally fail.

Theorem 19.1 (Loss of inaccessible correlations)**.** *Let A_0, A_1 be the initial and final A of a module above, with boundary laws viewed on the same trajectory. Then*

$$W_{\mathrm{in}} - \Delta\mathcal{F}_E = k_{\mathrm{B}}T\Sigma_{AB} + k_{\mathrm{B}}T\mathsf{I}(C; A_0 \mid A_1, B), \quad (19.2)$$

where $\Sigma_{AB} \geq 0$ is the sum of the relative-entropy losses of the marginal AB protocol. The last term equals the decrease of $\mathsf{I}(A; C \mid B)$.

Proof. The work is the same for the full and marginal protocols, because all quenched terms depend only on A, B. The marginal AB process is a Gibbs-

stationary protocol at each fixed energy, since the conditional Gibbs distributions are stationary and B is held fixed. Its work minus $\Delta\mathcal{F}_{AB}$ is $k_{\mathrm{B}}T\Sigma_{AB}$ by theorem 10.1. Subtract the endpoint difference of (19.1). The two BC-dependent terms cancel because their law and energy are fixed. Finally C and A_1 are conditionally independent given A_0, B, as the whole module composes to a channel on A controlled by B. Expanding $\mathsf{I}(C; A_0, A_1 \mid B)$ in the two orders proves the last equality. □

This is a finite version of the thermodynamic cost of modular information processing [7]. Even if the gate loses no work relative to its own endpoint marginal, the full accounting has the correlation term. The lost information could be used only by a protocol with access to the relevant degrees of freedom, or to a sufficient summary of them. This access is a causal restriction.

19.2 Blind erasure and controlled erasure

Take an initially flat bit $A = C$ with a fair spectator C, and a singleton B. Reset A to a fixed zero. The full endpoint entropy is $\log 2$ before and after, so the full free-energy difference is zero. The marginal entropy of A falls by $\log 2$, so a blind reset requires work at least $k_{\mathrm{B}}T \log 2$, approached by full-support approximations to the pure target. The full entropy production includes exactly the lost initial mutual information $\log 2$ at the reversible marginal infimum.

If C is accessible as the control, the same initial pair admits the reversible flat-energy gate $A \mapsto A \oplus C$. This sends $A = C$ to $A = 0$ with the control preserved, as a permutation on all pairs. It uses zero work in the permutation model and no conditional law has to be brought from full support to a point mass: the prepared input is already supported on two states. The difference between the protocols is precisely access to the correlated register. More generally, controlled approximate erasure from a positive conditional law to a nearly pure conditional output has work approaching $k_{\mathrm{B}}T\mathsf{H}(A \mid B)$ at flat energies, with B retained. Blind approximate erasure has infimum approaching $k_{\mathrm{B}}T\mathsf{H}(A)$. The resource difference is $k_{\mathrm{B}}T\mathsf{I}(A; B)$.

If both schemes use a noisy local summary B of a more distant C, the controlled scheme can consume only the correlation visible to that summary. The residual inaccessible information is conditional on B, exactly as in (19.2). When A is replicated at multiple sites, the control summary need not coincide with the original source. Its availability should be tested in the placed causal graph before applying a controlled-work formula.

19.3 When correlation loss vanishes

By the equality condition of conditional data processing, the second term in (19.2) vanishes if and only if

$$p(c \mid a_0, a_1, b) = p(c \mid a_1, b)$$

on the support. Equivalently the new accessible state A_1, B is sufficient for the information about C formerly available at A_0, B. A permutation of A controlled by B preserves this information for every input ensemble. A genuinely many-to-one gate can also preserve it for a particular ensemble, if all inputs merged by the gate have the same conditional law of C given B. Thus a logical loss of distinctions can be harmless for one spectator and dissipative for another.

For example, let $A = (D, N)$ with $D = C$ a fair bit and N independent fair noise. A gate may rethermalize N independently without losing $\mathsf{I}(A; C)$, since D remains. A gate that keeps N but resets D loses the whole correlation. If C is a future observer whose record will be used in a repair, the two gates therefore leave different usable resources, even though both act on one coordinate of a two-bit system.

19.4 Spacelike scheduling of independent modules

Let two gates have scopes A, B and D, B, where A, D are disjoint targets and B is immutable. Assume each gate's rate laws and quenches depend only on its target and B, and its transition seeds are independent of the other gate conditional on those states. Their kernels commute: at given b, both orders have joint transition law $K_A(a' \mid a, b)K_D(d' \mid d, b)$. This remains true if A, D are initially correlated. The expected work assigned to the A-gate is unaffected by scheduling the other gate first, since its complete A, B trajectory law is unaffected. The same holds for the D-gate.

Proposition 19.2 (Schedule-independent full ledger)**.** *In a finite causal protocol, suppose every adjacent swap of incomparable operations is between modules of this form, with separately additive quench changes on the stated scopes. Then the full input work, final law, and total relative-entropy loss are independent of the chosen linear extension.*

Proof. The preceding calculation proves invariance of the work and final law under one swap. The final energy is also the same because the quench changes add. The initial energy and initial law are fixed, so the total loss is determined

by the full ledger. Any two linear extensions are connected by the swaps proved in chapter 4. □

The assignment of correlation loss to an individual operation can still depend on the order. In a flat pair of identical fair bits, independently reset both to fair. The first reset destroys the single correlation $\log 2$, while the second changes nothing. The full loss is $\log 2$ in either order. The operations have distinct spacelike supports if their common correlation was prepared in the past, and no communication is needed to erase it by independent noise. This illustrates an order-independent endpoint result whose decomposition into marginal increments can vary.

19.5 A constraint on compositional thermodynamics

At the level of visible boundary channels, two realizations with the same stochastic kernel are interchangeable for computing downstream laws. At the level of work, their interchange requires agreement of initial and final auxiliary resources and of the ledger, or at least a bound valid for both. In particular a hidden spectator can carry correlations between two visible modules. Discarding it when composing estimates can double-count or omit a resource.

One can attach a boundary free energy and a nonnegative loss to every thermodynamic implementation with fixed initial law. When two implementations are composed on a shared full state with matching law and energy, the free-energy terms telescope and the losses add. For independent parallel implementations with product input, work and losses add by additivity of entropy and energy. For correlated parallel inputs, work can still add when scopes and quench changes are disjoint, while the total loss additionally records the correlation change. A symmetric monoidal rule for scalar losses therefore needs a product-input hypothesis or a correction term. This is the energetic counterpart of the distinction between tensoring channels and declaring a joint state.

CHAPTER 20

Relative boundary conditions and topological obstructions

A regional history is often constrained by data on an incoming boundary, an interface, or a world tube that is already observed. A general history map handles the boundary. In examples with a local system of higher choices, ordinary

relative cohomology can compute the space of possible extensions. This supplies obstructions beyond a comparison of the set of local readouts.

20.1 Base change and sections

Let $p : H \to B$ be a history map in the observation topos and $f : B' \to B$ a change of reference parameter. Set $H' = B' \times_B H$. If $b' : U \to B'$, then

$$U \times_{B'} H' \simeq U \times_B H, \qquad b = fb'.$$

The equivalence follows from pasting pullback squares. Taking section spaces over U gives the canonical equivalence of reconstruction spaces for the two boundary descriptions.

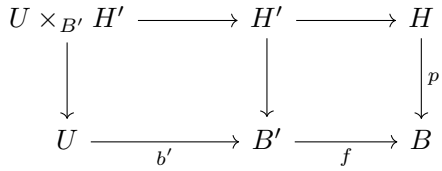


Figure 20.1: Both squares are homotopy pullbacks. Their outer rectangle identifies reconstruction after changing the reference base.

More generally, in an ∞-topos each pullback between slices has dependent sum and product adjoints. Dependent sum along $g : E \to B$ takes an object over E to its composite over B. Its fiber at b is the space of a choice of e over b and an element over e. Dependent product has fiber the space of compatible choices for every e over b. These operations implement different contracts. If e is a possible failure pattern, existence of a repair for one pattern uses a dependent sum and its image. A policy that supplies coherent repairs for all admissible patterns is a point of a dependent product. A causal implementation of such a policy also has to observe enough of e before acting.

For a pullback square of bases, dependent product satisfies the usual base-change equivalence. In the following diagram let g' and u be the pullbacks of g and v, respectively, and let $A \to E$ be an object.

To prove the displayed equivalence, test it against any $T \to B'$. The right side has mapping space $\mathrm{Map}_{E'}(T \times_{B'} E', u^*A)$, writing slice maps with a subscript. This identifies with $\mathrm{Map}_E(T \times_B E, A)$ by the pullback universal property. By adjunction this equals $\mathrm{Map}_B(T, \Pi_g A)$, where T is regarded over B. This is the mapping space into the left side. Naturality in T gives the equivalence. The result explains why changes of parameter commute with reconstruction conditions quantified over families of overlaps or contexts.

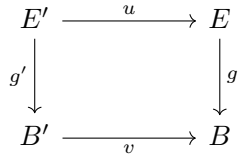


Figure 20.2: A pullback of parameter bases gives $v^{*}\Pi_{g}A \simeq \Pi_{g'}u^{*}A$. The two sides describe the same coherent family after changing parameters.

20.2 Mapping-space fibration for a pair

For a CW pair $A \subset U$, let $K = K(G, n)$ with G abelian and $n \geq 1$. An observation is a map $b : A \to K$. We use topological maps up to homotopy, equivalently maps of their corresponding homotopy types. The extension space is

$$\mathscr{R}(U, A; b) = \mathrm{hofib}_{b}\big(\mathrm{Map}(U, K) \longrightarrow \mathrm{Map}(A, K)\big).$$

Because the inclusion is a cofibration, a restriction model with the homotopy extension property makes strict extensions a model for that homotopy fiber. In a locally constant history theory, maps of this form arise as section spaces or their restrictions along boundary subspaces. The boundary can be modeled as its own observation object, so no assertion that a closed subspace is an open region is needed.

Theorem 20.1 (Relative cohomology of reconstructions)**.** *A reconstruction exists if and only if the connecting homomorphism sends $[b] \in H^{n}(A; G)$ to zero in $H^{n+1}(U, A; G)$. If one reconstruction exists, choosing it as an origin identifies the reconstruction space up to equivalence with $\mathrm{Map}_{*}(U/A, K)$. In particular, at that origin the homotopy groups in degree $i \geq 1$ are $H^{n-i}(U, A; G)$, taken to be zero in negative degrees. Its components form a torsor for $H^{n}(U, A; G)$.*

Proof. Homotopy classes of maps to $K(G, n)$ represent H^{n} [15]. The exact sequence of the pair identifies the image of the restriction map with the kernel of the connecting homomorphism. This is equivalent to nonemptiness of the homotopy fiber. Model $K(G, n)$ as an abelian group space. Translation by a chosen extension takes the fiber over b to the fiber over the constant observation, with any boundary comparison incorporated by the homotopy extension property. Mapping out of the cofiber U/A identifies that fiber with the based mapping space. Taking its i-fold loop space gives based homotopy classes of maps from the i-fold suspension of U/A into K. The suspension isomorphism in reduced cohomology gives the stated groups. The translation action on components is free and transitive. □

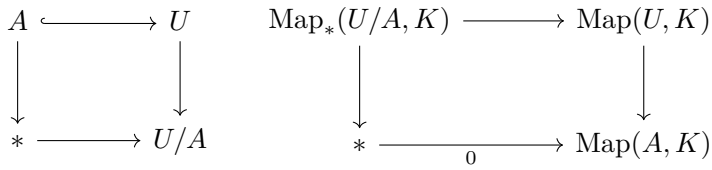


Figure 20.3: The cofiber square of the pair becomes a homotopy fiber square after mapping into K. The right-hand fiber computes the freedom of a reconstruction once one extension has been chosen.

20.3 Three boundary calculations

Take $U = D^2$ and $A = S^1$, and let $K = K(\mathbb{Z}, 1)$. A boundary record is its winding number. Only winding number zero can extend. Indeed, the restriction image of $H^1(D^2; \mathbb{Z})$ is zero, and the connecting map to $H^2(D^2, S^1; \mathbb{Z}) \cong \mathbb{Z}$ is an isomorphism. Every small arc of the boundary admits local compatible extensions into a collar. Those local facts cannot remove the global winding obstruction. At zero winding, the reconstruction fiber is contractible because $D^2/S^1 \simeq S^2$ and $\Omega^2 K(\mathbb{Z}, 1)$ is contractible.

For the same pair and $K = K(G, 2)$, every boundary observation extends, since $H^2(S^1; G) = 0$. The extension components form a torsor for G, and each component is contractible, because $\Omega^2 K(G, 2) \simeq G_{\mathrm{disc}}$. Thus identical boundary observations can permit a discrete family of inequivalent interiors. No choice among them is supplied by the boundary alone.

For $K = K(G, 3)$ the extension space at the trivial boundary has type $K(G, 1)$. Its unique component has automorphisms G, interpreted as paths from an extension to itself up to homotopy. In this example taking components collapses all possibilities to a singleton but loses the symmetries of the chosen lift. Each calculation can be localized to a spacetime chart with a designated spatial disk and boundary, or extended through a time interval by projection onto that disk.

20.4 A twisted example

A local system over a spatial circle can act nontrivially on its coefficient group G by transport $\theta : G \to G$. The cellular cochain complex has one copy of G in degree zero and one in degree one, with differential $\theta - \mathrm{id}$ up to choice of orientation. Thus global 0-cochains are invariant elements and the H^1 group is $G/(\theta - \mathrm{id})G$. For $G = \mathbb{Z}$ and $\theta = -1$, the invariant group is zero and $H^1 \cong \mathbb{Z}/2$. One can see the differential by cutting the circle into an interval, choosing a value

at its initial end, and comparing its transported value with the end value upon gluing. A twisted local history space can therefore have two global extension classes detected by a comparison around the cycle even if there is no nonzero globally fixed local value.

This topological information is independent of a probability law on the extension space. For example, a finite observer may record the parity class of a twisted lift, in which case there are two labels to which a finite ensemble can assign probabilities. A different observer may see only a local region where those classes agree. The entropy of the global parity readout depends on the ensemble and may have any value between zero and $\log 2$. The cohomological calculation determines the set of available classes and its behavior under observation. It supplies no canonical law on that set.

20.5 The first nontrivial lift of a Postnikov tower

A convenient general pattern starts with a classifying map $k : B \to K(G, n+1)$. Let H be its homotopy fiber over the basepoint. For a map $b : U \to B$, a lift to H is precisely a nullhomotopy of kb. If kb is nullhomotopic, the space of nullhomotopies is a torsor, after choosing one, for the loop space of $\mathrm{Map}(U, K(G, n+1))$ at the constant map. This loop space is $\mathrm{Map}(U, K(G, n))$. The proof is the path-space model for a homotopy fiber and the fact that mapping into a space preserves limits. A nonzero class in $H^{n+1}(U; G)$ is an obstruction, and even at a vanishing class there is generally a higher space of choices. Twisted Postnikov systems require their corresponding local coefficients. This simple untwisted form already gives an obstruction and a witness space for a higher reconstruction contract.

20.6 Locating the remaining freedom on an overlap

Suppose $U = U_1 \cup U_2$ is an open cover, $V = U_1 \cap U_2$, and local lift classes are modeled by cohomology classes in degree n with constant coefficients. The Mayer–Vietoris sequence [15] contains

$$H^{n-1}(V; G) \longrightarrow H^n(U; G) \longrightarrow H^n(U_1; G) \oplus H^n(U_2; G) \longrightarrow H^n(V; G).$$

The last arrow is the difference of restrictions. If the two classes agree on V, exactness supplies a global class. The set of such classes is a coset of the image of the left arrow. This identifies the remaining freedom in global classes with possible changes of the overlap comparison.

CHAPTER 21

Noisy broadcast along a tree

A source can remain physically represented by many descendants while the information obtainable from them tends to zero. In this chapter we analyze a process in which the number of readouts grows with the horizon. Each edge transmits a noisy copy, and the observer sees the entire generation. The calculation gives a criterion for collective reconstruction that cannot be inferred from a single branch.

21.1 Model and single-branch information

Take a rooted b-ary tree with $b \geq 2$ and an unbiased sign $X \in \{-1, 1\}$ at its root. For each edge, multiply the parent's sign by an independent sign of mean $\theta \in [0, 1]$. Thus the crossover probability is $(1 - \theta)/2$. Denote the level-n signs by L_n, and let $I_n = \mathsf{I}(X; L_n)$. Every finite truncation has a causal placement, for example with generations separated by time intervals longer than all wire lengths for that generation. A graph can branch within a bounded region if there is no additional packing constraint. The finite calculation assumes distinct carrier registers and independent edge noise.

For a single leaf, the conditional mean given X is $\theta^n X$, by multiplication of the independent edge signs. Its mutual information with X is

$$g(\theta^n), \qquad g(m) = \log 2 - h_{\mathrm{b}}((1 - m)/2).$$

For $0 \leq \theta < 1$ it tends to zero. In contrast, $|L_n| = b^n$. The issue is whether the new observations are numerous enough to recover the root despite their dependence through earlier shared branches.

21.2 Posterior biases and a contraction

If Y is any finite readout of a fair sign X, put $m(Y) = \mathbb{E}[X \mid Y]$. Then

$$\mathsf{I}(X; Y) = \mathbb{E}[g(m(Y))].$$

The optimal success probability is $(1 + \mathbb{E}|m(Y)|)/2$, since the most probable sign at each outcome is its posterior mode. The function g is even and has the expansion

$$g(m) = \sum_{k \geq 1} \frac{m^{2k}}{(2k)(2k - 1)} \quad (-1 \leq m \leq 1). \tag{21.1}$$

Indeed $g'(m) = \operatorname{artanh}(m)$ in the interior, so integrating the power series from zero gives the formula, and monotone convergence gives its endpoint values. The coefficient sum is $\log 2$. Thus

$$\tfrac{1}{2} m^2 \leq g(m) \leq (\log 2) m^2.$$

In particular, vanishing I_n implies $\mathbb{E}|m(L_n)| \to 0$, by Cauchy–Schwarz and the lower bound.

If X' is a child of X, it is also fair. If Y is downstream of that child alone, the Markov property gives $\mathbb{E}[X \mid Y] = \theta \mathbb{E}[X' \mid Y]$. Using (21.1) term by term proves

$$\mathsf{I}(X; Y) \leq \theta^2 \mathsf{I}(X'; Y).$$

The quartic term gives a strict improvement:

$$\mathsf{I}(X; Y) \leq \theta^2 \mathsf{I}(X'; Y) - \frac{\theta^2 - \theta^4}{12} \mathbb{E}\big[\mathbb{E}(X' \mid Y)^4\big]. \tag{21.2}$$

The coefficient is positive when $0 < \theta < 1$. By Jensen's inequality and $g(m) \leq (\log 2) m^2$, the expectation in the last term is at least $(\mathsf{I}(X'; Y)/\log 2)^2$.

21.3 The collective threshold

Theorem 21.1 (Binary tree reconstruction)**.** *In the model above the Bayes success probability from L_n tends to $1/2$ when $b\theta^2 \le 1$. When $b\theta^2 > 1$, it is bounded away from $1/2$ uniformly in n.*

Proof. Partition the level-n observations into the b subtrees descending from the root's children. Their observation tuples are conditionally independent given X. Hence the joint conditional entropy given X is the sum of their conditional entropies. Subadditivity of unconditional entropy gives

$$I_n \le \sum_{j=1}^{b} \mathsf{I}(X; L_n^{(j)}).$$

Each child has the same fair prior as the root. Applying (21.2) with child-level information I_{n-1} yields

$$I_n \le b\theta^2 I_{n-1} - cI_{n-1}^2, \qquad c = \frac{b(\theta^2 - \theta^4)}{12(\log 2)^2}.$$

Below the threshold the first term alone forces geometric decay. At equality $0 < \theta < 1$ and $c > 0$. A positive limiting value of the nonincreasing sequence would force a fixed positive drop at each step, which is impossible. Therefore $I_n \to 0$, giving the first assertion.

For $b\theta^2 > 1$, let S_n be the sum of the signs at level n, and set $Z_n = S_n/(b\theta)^n$. Given all signs through level n, each child sign has mean θ times its parent's sign and variance $1-\theta^2$. The next generation has conditionally independent edge noises. It follows that Z_n is a martingale with $Z_0 = X$, and the conditional variance of its increment at level $n+1$ is $(1-\theta^2)/(b\theta^2)^{n+1}$. Thus $\mathbb{E}Z_n^2 \le M$, where

$$M = 1 + \frac{1-\theta^2}{b\theta^2 - 1}.$$

Since $\mathbb{E}[XZ_n] = 1$ and Z_n is a function of L_n, conditioning on L_n and applying Cauchy–Schwarz gives $\mathbb{E}[\mathbb{E}(X \mid L_n)^2] \ge 1/M$. For $|m| \le 1$, $|m| \ge m^2$. Therefore the Bayes success probability is at least $(1 + 1/M)/2$ for every n. □

The critical parameter here is for binary symmetric independent edges and a regular tree [12]. The proof of the upper bound uses the independence of disjoint subtrees conditional on the root, and the lower bound uses the variance of independent child noises. A common random fault at a generation can violate both. If all children of a vertex receive the same bit flipped by one shared edge error, the new carrier count alone does not give the variance improvement. Causal access to the aggregate L_n also requires a readout event after the relevant carriers; a local observer at a leaf only has the single-branch information above.

21.4 The total correlation of a generation

Every leaf is marginally fair. Hence

$$\mathsf{TC}(L_n) = b^n \log 2 - \mathsf{H}(L_n).$$

This can be large even when root information tends to zero. For example, take any $0 < \theta < 1$. Two siblings have pair product with mean θ^2 since the parent cancels. Their mutual information is $g(\theta^2) > 0$. Group the leaves into disjoint sibling pairs within each family, with an unmatched leaf when b is odd. By entropy subadditivity across these groups and the fairness of each leaf,

$$\mathsf{TC}(L_n) \ge b^{n-1} \lfloor b/2 \rfloor \, g(\theta^2).$$

Thus total correlation can grow exponentially while the specified root label becomes unrecoverable below the threshold. The common recent ancestors generate redundancy of a different label. The replication theorem lower bounds correlation under a

decoding contract for a named source. A converse using total correlation alone would give the wrong conclusion here.

21.5 Reading a finite genealogy

A full realized lineage includes every parent and every child sign. Conditional on the root, the independent edge signs can be recovered from that complete history by multiplying the signs on each edge. For a depth-n tree with $E_n = \sum_{j=1}^{n} b^j$ edges, the entropy of the complete history is therefore

$$\log 2 + E_n \, h_{\mathrm{b}}((1-\theta)/2).$$

The entropy of the leaf projection is different because it sums over internal signs. With retained parent registers and freshly thermal fair targets, controlled writing of each child at crossover $(1-\theta)/2$ has an endpoint work infimum $k_{\mathrm{B}}Tg(\theta)$ per child in the stated energetic class. Preparing every child conditional on its parent preserves the joint factorization. Summing over a fixed finite depth gives total work infimum $E_n k_{\mathrm{B}}Tg(\theta)$ for this explicit realization with all those registers retained. The same total follows from the full entropy drop, so the independent conditional writing construction is asymptotically optimal for that boundary. A different procedure that discards intermediates must account for them or state a different input–output resource model.

A high work investment in the individual writes does not necessarily preserve root information when $b\theta^2 \leq 1$. The resulting histories have a well-defined stochastic coupling and nonzero local correlations. Their root-reconstruction behavior depends on how many independent observations cross the relevant causal generations.

CHAPTER 22

Lineage survival and growing populations

The tree in chapter 21 has deterministic branching. A replication process can also fail to produce descendants. A lineage semantics must then keep track of survival and label fidelity separately. We give a branching model in which the distinction can be proved and account for work conditional on a finite realized genealogy.

22.1 A random genealogical presheaf

Fix a finite maximum offspring count B, and let a rooted ordered tree have at most B children per vertex. A finite-depth observation contains the tree through that depth and any observable labels on its vertices. Restriction prunes the last generation. The tower of such finite observations has an inverse limit representing locally finite genealogies and their labels, with the usual finite cylinder topology. A stochastic realization can generate the next generation from the present one with independent offspring counts and transition kernels. The finite laws are consistent under pruning because the offspring and child-label kernels each sum to one. This gives a law on the infinite tower by the countable cylinder construction.

The genealogy itself is a record of possible transmissions. A surviving infinite genealogy has an infinite branch by finite branching and König's lemma. This makes it a witness to an unbounded lineage. A faithful copy of the initial label along that branch is a separate condition on the edge label assignments. The chance of having any faithful branch can be smaller than the chance of having an infinite genealogy.

22.2 Extinction and the first moment

Let the offspring count be D with generating polynomial $f(s) = \mathbb{E}s^D$. Start from one individual. Write Z_n for the population size at generation n, and $q_n = \mathbb{P}(Z_n = 0)$. By independence of the child subtrees,

$$q_0 = 0, \qquad q_{n+1} = f(q_n).$$

Since zero is absorbing, q_n increases to the probability q of eventual extinction. Continuity yields $q = f(q)$. If $u \in [0,1]$ is any fixed point, then $q_0 \le u$ and monotonicity gives $q_n \le u$, so q is the smallest fixed point. The mean population obeys $\mathbb{E}Z_{n+1} = f'(1)\mathbb{E}Z_n$, since conditional on Z_n it is the sum of Z_n iid offspring counts. Hence $\mathbb{E}Z_n = \mu^n$ where $\mu = \mathbb{E}D$.

If $\mu \le 1$ and D is not identically one, then $q = 1$. For $s < 1$, convexity gives $f(s) \ge 1 + \mu(s-1) \ge s$. When $\mu < 1$ the inequality is strict. At $\mu = 1$, a nonconstant law with mean one has some mass at $D \ge 2$; strict convexity on $(0,1)$ gives strict inequality there and $f(0) > 0$. If $\mu > 1$, then for $s < 1$ sufficiently close to one, $f(s) < s$. Since $f(0) \ge 0$, continuity gives a fixed point less than one. This includes the case $f(0) = 0$, when $q = 0$. Thus supercritical growth means positive survival probability, with deterministic one-child persistence as a separate critical case.

For instance, if $D = 2$ with probability p and $D = 0$ otherwise, the fixed-point equation is $q = 1 - p + pq^2$. For $p > 1/2$ the smaller root is $(1-p)/p$, so survival has probability $(2p-1)/p$. A population can have exponentially growing expected size and still have substantial probability of extinction. At $p = 3/4$ survival has probability $2/3$ and the mean size is $(3/2)^n$.

22.3 Faithful sublineages

Suppose each offspring transmission preserves its parent's label with probability a, independently across edges and independently of offspring counts. For the moment impose the stricter fidelity condition that every edge on the path preserves the label. Starting from an intact source, the subtree of intact paths is a branching process whose offspring count is binomial with parameter a conditional on D. Its generating function is

$$f_{\text{intact}}(s) = f(1 - a + as).$$

The conditional mean is aD, so the mean number of intact children is $a\mu$. Under the nondegeneracy just discussed, positive survival of an intact branch holds exactly when $a\mu > 1$. The proof is the previous generating-function argument applied to this thinned offspring law.

This intact criterion treats an edge error as a failure even if a later edge restores the original label. Endpoint recovery permits cancellation or coding-based correction. For binary symmetric edge errors an observer of many labels can exploit statistical dependence, as theorem 21.1 shows on a regular tree. Thus $a\mu > 1$ and $b\theta^2 > 1$ apply to different observables and different success events. An intact-path count can be large even when the path is not identifiable from unlabeled internal traces. Conversely a successful collective decoder need not find a path with no flips.

22.4 An entropy and work calculation at fixed genealogy

Condition on a finite realized genealogy τ, independent of the initial fair binary label and the edge noise. Let $N(\tau)$ count all nonroot vertices in it. Assume every nonroot

carrier initially has an independent fair bit, the root is retained, and all endpoint energies are flat. Require each child bit to be the parent plus an independent crossover e. Then the final conditional entropy of all carriers given the root and τ is $N(\tau)h_{\mathrm{b}}(e)$. The proof gives a bijection between child assignments and edge error assignments once the root and tree are fixed. Before writing, that entropy is $N(\tau)\log 2$.

In a protocol whose full modeled boundary for this conditional branch consists of the root and these targets, the input work is at least

$$N(\tau)k_{\mathrm{B}}T(\log 2 - h_{\mathrm{b}}(e)).$$

With causal access to each parent and independently controlled Gibbs resets, the conditional work infimum equals this value for $0 < e \leq 1/2$. At $e = 0$ it is approached through approximate writes. Each subprotocol acts with the parent as an immutable control. The full initial and final conditional laws match the factorized construction, so the sum of conditional write costs meets the endpoint bound. Any temporary communication resources are either returned or included separately.

If the tree is selected independently in advance, revealed to the protocol, and retained as a side register with fixed law and energy, the conditional cost can be averaged. At depth n the expected number of new carriers is $\sum_{j=1}^{n} \mu^{j}$. One can realize an equivalent fixed finite state space by preallocating all slots in the maximum B-ary tree, leaving unused slots with their same independent thermal law conditional on τ, and acting only on selected slots. Those unused factors cancel at the endpoints. If instead the device generates the tree using energy-consuming transitions or a mutable controller, its resource ledger includes that process. Averaging a conditional cost alone would not account for preparing, measuring, or erasing the genealogy.

22.5 A local trace and a population-level specification

At finite horizon, label equality on each edge is a local cell predicate. Existence of at least one intact root-to-leaf path is a disjunction over paths, and so differs from a predicate on the entire set of local transitions. The path can be recorded as a witness, or omitted after existential quantification. If a lineage is judged by the survival of any compatible branch, the resulting reconstruction object can have many witnesses at the same horizon, with different extension prospects. The compactness theorem applies to the tower of intact path witnesses because each level is finite. It does not assert that a random sample from the genealogy has positive probability of admitting such a branch. The branching calculation provides that probability under independent offspring and edge noise.

In an application with limited energy or bandwidth, offspring rates can depend on remaining resources or on a shared environment. Then the simple product recurrence $q_{n+1} = f(q_n)$ generally fails. One can still include those environmental variables in a causal graph and evaluate survival conditional on them. The history model and work ledger continue to state their requirements. The generating-function result specifically uses independent, identically distributed reproduction conditional on each live individual.

CHAPTER 23

Identifying the label shared by several carriers

So far the label X has been named in the replication contract. Given only a terminal

joint law, one can instead ask which random variables are exactly identifiable from every individual carrier. The answer is determined by the connected components of a finite incidence construction, a form of the Gács–Körner common variable [14]. It provides a precise converse for shared exact decodability, complementary to the entropy lower bound.

23.1 An incidence colimit for common information

Let $Y = (Y_1, \dots, Y_r)$ have finite joint support Ω. Keep only values of each Y_i with positive marginal probability. Form a disjoint union $D = \bigsqcup_i A_i$ of their value sets. For every $\omega \in \Omega$, identify the r tagged values (i, ω_i) with one another. Let J be the quotient set for the equivalence relation they generate. Equivalently, form an undirected bipartite incidence graph with a vertex for each ω and each tagged value, joining ω to its tagged coordinates. The set J is the set of connected components. The quotient maps $j_i : A_i \to J$ satisfy $j_1(Y_1) = \cdots = j_r(Y_r)$ almost surely. Write this common value as J_Y.

Theorem 23.1 (Maximal shared exact label)**.** *If Z is any random variable that is a deterministic function of each Y_i almost surely, then Z is a deterministic function of J_Y. Conversely every function of J_Y is exactly decodable from all the carriers. In particular the greatest entropy of an exactly common label is $\mathsf{H}(J_Y)$.*

Proof. Choose functions d_i with $Z = d_i(Y_i)$ almost surely. Every ω has positive probability, so the values $d_i(\omega_i)$ must all agree. This means the map from the disjoint union taking (i, y) to $d_i(y)$ is constant on each generating identification. It factors through J, and the converse follows from the quotient maps j_i. Entropy cannot increase under a deterministic function, and J_Y itself is allowed. □

The construction is a colimit in sets: take the coequalizer enforcing equality of all maps out of a support tuple into its coordinate values. Its universal property is precisely the common-label property. A homotopy colimit of the incidence graph retains loops as well as components, while exact classical labels are discrete, so a map into their label set factors through π_0. Hence the criterion can be expressed either by the set colimit or by the components of a higher incidence object.

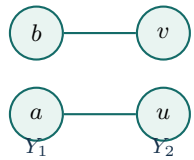


Figure 23.1: For two carriers the support pairs are edges. This disconnected support gives two possible shared exact label values. A positive edge a–v would join the components.

For a pair of fair bits with strictly positive crossover and a full-support binary symmetric joint law, the incidence graph is connected, so its maximal exact shared label is constant. Its mutual information is positive if the crossover differs from $1/2$. For perfectly equal fair bits the graph has two components, recovering one full bit of exact common information. Thus the maximal exact label can change discontinuously when a zero entry becomes positive. The entropy formulas of chapter 7 change continuously with the law on the fixed finite alphabet. Exact and approximate contracts have genuinely different stability properties.

23.2 A named source and a family of side states

To test whether a prescribed X is exactly copied into all Y_i, include X as an additional carrier in the support graph, and test whether the component quotient distinguishes all positive-probability values of X. If it does, X factors through the common component label and each Y_i decodes it. If two distinct values of X are joined, their identification is forced by one or more positive support assignments, and no decoder on every individual carrier can distinguish them. This criterion uses the actual joint support. Identifying a common terminal label alone does not identify its origin.

With side information W, form the incidence graph separately for each w of positive mass. A family of conditional exact decoders is possible if each conditional support has the required distinctions. For a single joint object record the pair (w, J_w), with J_w the conditional component. The entropy of the maximal shared component beyond W is $\sum_w p(w)\mathsf{H}(J_w \mid w)$. A temporal restriction that forgets W can merge components. For example X and W may be independent fair bits and $Y = X \oplus W$. Given W the support pairs are a permutation graph and Y decodes X exactly. After forgetting W, the support of X, Y is full and the common component is trivial.

23.3 Interfaces determined by a posterior

Suppose an observer sees Y, W and needs to choose an action a with loss $\ell(X, a, w)$. Its expected loss conditional on (y, w) is $\sum_x p(x \mid y, w)\ell(x, a, w)$. Thus the posterior vector $P_Y = p(\cdot \mid Y, W)$, together with W, is sufficient for every finite decision task about X. The optimal Bayes success for guessing X is $\mathbb{E}\max_x P_Y(x)$. Two observations with identical posterior vectors have the same conditional value for every loss. Conversely if posterior vectors differ, some coordinate differs, and one can choose a two-action loss with a threshold between the corresponding coordinate probabilities. The two observations then prefer different actions. This makes the posterior a minimal statistical distinction if all decision tasks about X are retained.

For a causal cut C, the posterior given C, W determines the maximal conditional success obtainable from the cut. A channel from C to a downstream readout replaces each posterior by an average of cut posteriors with that readout. The maximum-coordinate function is convex. Its expected value cannot increase under this averaging, which proves the decision version of the cut bound without entropy. If an output value merges cut states with different unique optimal guesses, the success can drop strictly. If it merges cut states whose optimal guess agrees, guessing can be unaffected even when other tasks and mutual information degrade.

23.4 Loss of common labels under restriction

A local map may merge tagged values and hence join components of the common-label graph. Independently applied local channels cannot improve an individual carrier's optimal inference about the old source. A new shared random input can create a common label, which then has a different causal origin. The component graph decides exact decodability, while the ensemble determines approximate recovery.

CHAPTER 24

Interactive cuts and zero-error interfaces

A stationary cut in a directed acyclic graph isolates a past ideal from its future. Some storage protocols permit messages in both directions between a persistent pair of parties. Every individual message still travels forward in time. A cut across the parties can be crossed repeatedly, so a bound must account for the transcript and the receiver's private information. We give an interactive extension of the information bound and a finite criterion for exact one-message interfaces.

24.1 A finite interactive execution

Let a sender receive X, and let W be shared prior data. The receiver has a private random seed R independent of X conditional on W. Additional sender randomness is permitted. Conditioned on all private seeds and the inputs, the protocol is deterministic. Messages are exchanged in a finite causal order. The sender-to-receiver messages are $C_1, \dots, C_n$, with declared alphabets A_i. The receiver may reply using W, R and its current transcript. At the end its observation Y is a function of $W, R, C_1, \dots, C_n$. If availability of a message, its timing, or a choice of channel conveys information, include the corresponding symbols in the transcript. This prevents an uncounted signaling side channel.

Theorem 24.1 (Interactive transcript bound)**.** *If the receiver outputs a label with error probability e for an m-ary X, then*

$$\begin{aligned} \mathsf{H}(X \mid W) - f_m(e) &\le \sum_{i=1}^{n} \mathsf{I}(X; C_i \mid C_{<i}, W, R) \\ &\le \sum_{i=1}^{n} \log |A_i|. \end{aligned} \tag{24.1}$$

The first inequality holds regardless of the number of reverse messages.

Proof. The final observation is a function of $W, R, C_{\le n}$, so conditional Fano gives $\mathsf{H}(X \mid W, R, C_{\le n}) \le f_m(e)$. Since $\mathsf{I}(X; R \mid W) = 0$, the conditional information $\mathsf{I}(X; R, C_{\le n} \mid W)$ equals $\mathsf{I}(X; C_{\le n} \mid W, R)$. Apply the mutual-information chain rule in the forward message order. Each summand is bounded by the conditional entropy of C_i, in turn bounded by its alphabet logarithm. □

The partial sums on the right of the first inequality are the reductions of posterior uncertainty at the receiver with its seed included. This identity concerns one source value chosen at the start. Directed information is a broader treatment

of time-indexed inputs with feedback [23]. The point here is that an ordinary statistic of the forward transcript can understate the information available to the receiver. For example, let X, R be independent fair bits, let the receiver first transmit R, and let the sender then transmit $C = X \oplus R$. The forward bit is marginally independent of X, but $\mathsf{I}(X; C \mid R) = \log 2$ and the receiver decodes exactly. The two messages must be assigned causally ordered events. In spacetime they can represent a query and a later response from a relay.

If the private receiver state is correlated with X already, one must either include it in the side information or keep $\mathsf{I}(X; R \mid W)$ on the right. If receiver randomness is repeatedly consumed or measured, it can be represented as a seed sequence and made available to the conditional calculation as a single register. If feedback itself changes the source value, the theorem still applies to a specified exogenous label X; inference about a changing target requires a new definition of the reference variable.

24.2 Deterministic codes with side information

Suppose only the receiver observes W. The sender observes X and sends one message $c(X)$ from a finite alphabet. Build the graph $\mathcal{G}$ on positive-probability labels x, joining distinct x, x' if there is a w with both $p(x, w) > 0$ and $p(x', w) > 0$.

Theorem 24.2 (Exact one-message alphabet)**.** *The least message alphabet size permitting zero-error decoding of X from the message and W is the chromatic number $\chi(\mathcal{G})$. The result also holds if the sender is allowed independent randomization and zero error is required.*

Proof. If two adjacent labels get the same message, the receiver observing the common side value cannot distinguish them. Thus a deterministic code is a proper coloring. Conversely a coloring determines at most one possible label among those compatible with any given w and message, so it defines a zero-error decoder on the support. For a randomized encoder, the supports of the message laws of adjacent labels must be disjoint, for the same ambiguity argument. Selecting one possible message from the support of each label gives a proper coloring with the same alphabet. □

For example, take X uniform on the five vertices of a cycle and, conditional on $X = x$, choose W uniformly among the two cycle edges incident to x. The posterior given W is uniform on its two endpoints, so $\mathsf{H}(X \mid W) = \log 2$. The confusability graph is the odd cycle. A binary alphabet meets the entropy bound

with equality but cannot be a proper coloring. A three-symbol alphabet is both necessary and sufficient. This comparison separates the expected-information condition from a one-shot zero-error feasibility condition. If the sender learns W before transmitting, it can encode the index of X in the ordered pair of endpoints and one bit suffices.

24.3 Fixed pairs and block messages

If independent source labels $X_1, \dots, X_k$ must all be recovered with side contexts $W_1, \dots, W_k$, the joint conditional entropy is the sum when the pairs (X_j, W_j) are independent. Their joint confusability relation holds for two distinct source tuples whenever, at every coordinate, the two labels are equal or adjacent, since one can select a compatible side value at each coordinate. This is the strong product of the single-letter graphs. The one-shot result applies to the product. A coordinatewise coloring gives an alphabet of size at most $\chi(\mathcal{G})^k$, but joint coloring can be smaller.

For the five-cycle example, two source labels can be coded with five messages, rather than the nine obtained by coloring coordinates separately. Label the vertices by residues modulo five with adjacency ± 1 and send $c(x_1, x_2) = x_1 + 2x_2 \pmod 5$. For distinct confusable pairs, their coordinate differences are in $\{-1, 0, 1\}$ and not both zero. The corresponding value $\delta_1 + 2\delta_2$ is nonzero modulo five: its absolute value is at most three and it cannot vanish with such differences unless both are zero. Thus this is a valid coloring. The entropy bound gives four messages as a lower bound, and a simple independent-set argument raises it to five. Indeed in a set of mutually nonconfusable pairs, each fixed first coordinate has at most two second coordinates, and if some row has two such coordinates, they are nonadjacent in the five-cycle. Their closed neighborhoods cover all five possible second coordinates, so both adjacent rows of the first coordinate must be empty. The two remaining rows are adjacent to each other and contain at most two points total: if one contains two, the other is empty, and if neither does, each contains at most one. Hence such a set has size at most four whenever any row has two points. If no row has two, it has size at most five. Every color class has size at most five, so the 25 pairs require at least five colors.

This example gives an explicit causal reason to distinguish coding before a channel from copying at its receivers. The sender may combine input labels to fit an interface whose cardinality is a resource. A later observer possessing the relevant side information can decode both exactly. Without that side information, the same channel is not injective on source tuples. The reconstruction fiber at a message is therefore conditional on the side observation.

24.4 Combining feedback with a work ledger

An interactive transcript can be stored in physical registers. For independent fair X, R in the earlier example, exact reversible communication can write a pure outgoing blank with the bit $C = X \oplus R$ using controlled addition, provided both controls are causally accessible at the write. Sending and later resetting R, C is a physical protocol with its own resources. If both registers are retained, X is a deterministic function of the pair. If an implementation erases C with X, R as controls, the existing correlation can be used by a reversible uncompute. If it erases a fair C blindly, its marginal preparation has work infimum approaching $k_{\mathrm{B}}T \log 2$ at flat energy. The causal protocol determines which of these control sets is available at the erasure event.

More generally, after a receiver computes an answer from a transcript it may keep only the answer and reset a log. The entropy of the log conditional on the retained data gives the controlled free-energy difference for a flat reset. An observer unable to access those retained data can pay more, as in theorem 19.1. Entropic transcript bounds thus constrain inference at one interface, while the thermodynamic ledger constrains preparing and clearing the physical implementation of that interface. The two calculations can share variables without making the inference bound itself a work formula.

CHAPTER 25

Refinement, coarse observation, and hidden state

A model of replication depends on a choice of resolution. One can refine regions, insert intermediate events in a causal trace, or resolve the microstates behind an observed carrier. These changes preserve different quantities. We give precise criteria for preservation in each case and describe what is lost when the criteria fail.

25.1 Refining a site

On a manifold, a basis of opens generates the topology. A sheaf on all opens is determined by its restriction to a basis with the induced topology. Conversely, a sheaf on the basis extends by descent: evaluate an arbitrary open by the limit of sections on basic opens contained in it, with compatible restrictions. The sheaf condition on the basis ensures agreement on basic opens and ensures the arbitrary-open sheaf axiom for the extension. In the ∞-categorical setting this is the comparison equivalence for sheaves of spaces [21]. The analogous

hypercomplete statement uses the hypercomplete reflection on both sites. Hence restriction to a basis preserves reconstruction fibers and relative pullbacks up to equivalence.

A choice of a finite cellular approximation is different from restricting to a basis of the same topology. It can identify distinct worldsheets, omit intersections, or replace a continuum of states by finitely many values. To justify it one needs a comparison from the continuous observation object to the discrete one, or a new independent specification for the finite model. A finite sheaf on an incidence site has exact gluing relative to its cells. It need not recover every observable in the smooth spacetime model. In the proof of the cut theorem the causal graph is supplied, rather than derived from an approximation of every causal curve.

25.2 Coarsening a boundary and the fiber of the fiber

Let $H \xrightarrow{o} B \xrightarrow{f} C$ be composable observation maps in $\mathcal{E}$. Fix $c : U \to C$. Define the spaces of reconstructions over c for fo and for f as in (3.1). Applying o gives a map

$$\mathscr{R}_{fo}(U, c) \longrightarrow \mathscr{R}_f(U, c).$$

A point of the target is a fine observation b and an identification $fb \simeq c$. The fiber over this point is $\mathscr{R}_o(U, b)$. This follows from pasting homotopy pullbacks or by identifying the full fiber as triples consisting of a history, a fine observation, and a comparison in the space over c. The boundary identification carried by b is part of the point in the target, and the comparison to it includes the needed coherence.

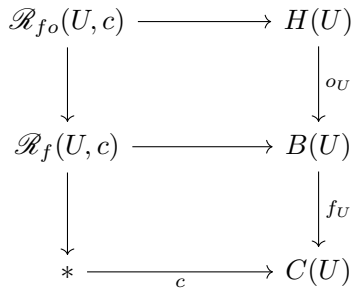


Figure 25.1: Pasting the two homotopy pullback squares gives the coarse reconstruction fiber. A second fiber over a fine observation in the middle row recovers the original reconstruction.

For a finite joint law of fine and coarse values, the chain rule $\mathsf{H}(B) = \mathsf{H}(C) + \mathsf{H}(B \mid C)$ applies when $C = f(B)$. The conditional term quantifies the uncertainty

about a fine observation over its coarse value. It has no value fixed by the homotopy fiber alone. Two ensembles with the same fine and coarse supports can have different conditional entropies. The geometric and statistical constructions therefore agree on which observations are grouped while retaining different information about them.

25.3 Refining a causal wire

Let $K : A \rightsquigarrow B$ be a kernel on a directed edge. Replace the edge by a future-directed path with kernels $L : A \rightsquigarrow D$ and $J : D \rightsquigarrow B$ such that $K = JL$, using composition in **FinStoch**. Conditional on the input, the intermediate randomness is private to this path and only B is exposed to the rest of the graph. Eliminating D by summation gives the original factor in the joint law of all original observed variables. Thus every readout error and entropy at an old interface stays the same. The new intermediate interface can have a larger alphabet and more source information than B, according to data processing.

If two paths transmit jointly prepared intermediate values, each marginal path factorization can be correct while their joint simulation fails. The replacement is valid for the full graph when the factor induced on all exposed outputs, conditional on all exposed inputs, agrees jointly. This is the boundary-kernel criterion. For an example, replace one random gate that outputs a common fair bit twice by two independent fair-bit gates. The replacement gives the two same one-wire laws and changes the joint gate law. Downstream equality tests detect the change. Thermodynamic costs can differ even when joint kernel equivalence does hold, because implementing a new intermediate interface may require nonthermal states.

25.4 Coarse energies from partition functions

Let $g : S \to A$ be a surjective map of finite state sets, and $\bar{p} = g_* p$. For a microstate energy E, define an energy on coarse states by

$$\bar{E}(a) = -\beta^{-1} \log \sum_{s : g(s) = a} e^{-\beta E(s)}. \tag{25.1}$$

Then $\pi_{\bar{E}} = g_* \pi_E$, and both equilibrium partition functions coincide. The log-sum in (25.1) incorporates degeneracy. For a flat $E = 0$, $\bar{E}(a) = -k_{\mathrm{B}} T \log |g^{-1}(a)|$, so a coarse alphabet with different fiber sizes is generally not flat.

Theorem 25.1 (Exact hidden free-energy term)**.** *For every finite microstate*

ensemble,

$$\mathcal{F}_E(p) = \mathcal{F}_{\bar{E}}(\bar{p}) + k_{\mathrm{B}}T \sum_a \bar{p}(a)\mathsf{D}\big(p(\cdot \mid a)\|\pi_E(\cdot \mid a)\big). \tag{25.2}$$

At a fixed coarse law, the unique minimizing microstate ensemble is the one with each conditional microstate law equal to its Gibbs conditional on every positive-mass coarse state.

Proof. Apply the chain rule for relative entropy to the deterministic observation g, with Gibbs reference. The coarse term is $\mathsf{D}(\bar{p}\|\pi_{\bar{E}})$, and the remaining terms are the conditional divergences in (25.2). Insert the free-energy identity and equality of equilibrium partition functions. Nonnegativity and its equality condition prove the optimization claim. □

Suppose a kernel K on S is lumpable through g: for each s with $g(s) = a$, the conditional distribution of the next $g(s')$ is the same $\bar{K}(\cdot \mid a)$. Then $(g_*p)\bar{K} = g_*(pK)$ for every p. If K preserves π_E, $\bar{K}$ preserves $\pi_{\bar{E}}$. This gives a closed marginal dynamics. It does not imply that the hidden term in (25.2) is zero. A separate conditional equilibrium hypothesis at endpoints ensures equality of endpoint free-energy differences. Equality of the full work with the coarse work also follows if each quench satisfies $E'(s) - E(s) = u(g(s))$, since then $\bar{E}'(a) - \bar{E}(a) = u(a)$ and both quenches have the same work. Under these conditions the full and coarse total losses agree by their ledgers.

25.5 Two calculations with the same visible behavior

Let the physical state be $S = \{0,1\} \times \{0,1\}$ with coordinates A, B, observed only through A. Start from the uniform law and leave A unchanged. One device does nothing to B. Another prepares B with small crossover e from a fixed zero value using an isothermal protocol that returns to flat energy. The visible endpoint laws and visible channel on A agree. The first device can have zero work. The second has full endpoint free-energy increase $k_{\mathrm{B}}T[\log 2 - h_{\mathrm{b}}(e)]$, all in the hidden term. As $e \to 0$ the difference tends to $k_{\mathrm{B}}T \log 2$.

Alternatively start with $B = A$ and both bits fair. Let K independently thermalize B, keeping A fixed, at flat energy with no work. Before the step the hidden term of (25.2) is $k_{\mathrm{B}}T \log 2$. Afterwards it is zero. The coarse law has no change and predicts no loss on its own. The full dissipation is $\log 2$ in nats. Thus an omitted term may either increase or decrease across endpoints. A coarse free-energy difference alone need not be a lower bound on the full free-energy difference; one must know the change of the hidden term, or supply a condition such as equilibrium conditional states at both ends.

25.6 Resolution as part of a theorem statement

A geometric refinement preserving the sheaf category transports reconstruction spaces. A causal refinement preserving a joint boundary kernel transports readout laws. A thermodynamic refinement preserving the full resource ledger transports the corresponding work calculation. None of these equivalences implies the next one without its own data. For instance, two stochastic kernels can have the same support and thus define the same finite history sheaf while assigning different probabilities to successes. Two implementations of one kernel can use different blank preparations or dissipate different correlations. The refinement tests in this chapter provide conditions under which a result can be moved between descriptions without changing the quantity it predicts.

CHAPTER 26

Combining the reconstruction and resource tests

The preceding results give a sequence of compatibility tests for a design. To make their interaction explicit, consider a causal protocol with named boundary registers, a full law at each cut, an observation map, and an isothermal energy model. The same contract can be tested at each level without identifying the tests with one another.

26.1 A combined cut and work bound

Let X be a source, W an unchanged side register, and C the messages across a common causal cut. Write $Y = (Y_1, \dots, Y_r)$ for the final receiver tuple. Suppose their joint law satisfies the cut factorization and the i-th receiver meets error e_i. If the final physical array is exactly the carrier state Y together with W, and its energies have the conditional endpoint form of (11.3), the following two bounds apply to a protocol in theorem 10.1: Write $a_i = (\mathsf{H}(X \mid W) - f_m(e_i))_+$, and let $B = \mathsf{I}(X; C \mid W)$. The bounds are

$$B \geq \max_i a_i, \tag{26.1}$$

$$W_{\mathrm{in}} \geq \sum_i \Delta F_i - k_{\mathrm{B}} T \mathsf{TC}_0 + k_{\mathrm{B}} T \max\{0, \textstyle\sum_i a_i - B\}. \tag{26.2}$$

The first follows by using each Y_i as the readout in theorem 4.2. The second follows from (7.6) and the work decomposition. The entropy B appearing here is a joint statistic of the source and cut. If only an upper bound $B' \geq B$ is available, using B' gives a possibly weaker valid inequality. In particular the logarithm of cut alphabet size is an admissible upper bound.

The combination distinguishes throughput from energetic redundancy. For a noiseless shared trunk and exact final receivers without the source in the accounting boundary, $B = \mathsf{H}(X \mid W)$ is sufficient for the cut and $(r-1)\mathsf{H}(X \mid W)$ is still required in the final correlation. If the source register is retained inside the boundary, include it as a separate carrier and use (7.7) or the full entropy difference. It does not have to lie downstream of the common cut. Consequently the causal bound counts only receivers, while the energetic count for the full final array includes the source. The apparent difference in multiplicity comes from the different questions asked of the same circuit.

26.2 What can be composed

Take two finite networks N_1, N_2 whose exposed interface variables agree, and suppose the interior randomness of the second is generated conditionally only from the interface and its declared side data. The history spans compose by a pullback, with the interface identified. The law composes by conditioning on that interface and using the downstream kernel. If the two sides are given instead by marginal laws on their own variables, the separator construction in chapter 8 gives exactly the conditional-independent law when the separator marginals match. This compatibility statement can be iterated over a tree.

If both networks are thermodynamic protocols on the same full state boundary, with equal energy and law at the glued interface in time, the work of their concatenation is the sum of the stage works and their free-energy differences telescope. When only the visible interface is matched, extra degrees of freedom and their correlations can obstruct the corresponding energetic composition. A hidden conditional state that differs across the two protocols may have to be prepared or reset. The hidden free-energy identity (25.2) identifies the needed endpoint term when both models share a specified microstate refinement.

The conditional independence used to glue distributions is an additional modeling condition. Higher descent of support histories does not impose it. Its role can be tested against alternative couplings by (8.3), which gives exactly how much joint entropy is lost relative to the separator law. If the energy expectation is determined by the same local marginals, that loss becomes an energetic difference of $k_{\mathrm{B}}T$ times the divergence. This is one direct bridge between compatibility of local observations and free energy of their global realization.

26.3 Three kinds of obstruction

A topological obstruction prevents a coherent history with the specified boundary, even when every small region admits a local extension. The winding and twisted lift examples in chapter 20 exhibit this failure without any noise or work calculation. A causal obstruction occurs when admissible histories include terminal values equal to the source but the proposed communication cannot produce them uniformly from the source input. A receiver outside the causal future with no preexisting side information illustrates this case. Finally a resource obstruction occurs when the initial physical state, allowed operations, and work budget cannot produce the required full joint ensemble. Thermal blanks and insufficient work give an example even when causal paths exist.

These obstructions have different certificates. A missing homotopy class can be detected by a relative cohomology class. A missing causal degree of freedom can be detected by the factorization through an insufficient cut or by a zero-error confusability graph. An energetic insufficiency can be detected by a free-energy difference or a finite-activity penalty. A model that merges the three tests into a single scalar fails to retain these certificates.

26.4 Evidence and scope

The results are mathematical statements with proofs under the declared hypotheses. The finite illustrations specify their own joint laws and rate parameters. They are examples, and do not constitute measurements of a particular device or a numerical model of biological reproduction. Standard results used in the proofs are cited where they enter, including ∞-topos descent, finite data processing, equilibrium free energy, and quantum relative-entropy monotonicity.

The spacetime geometry constrains observations, causal placements, and comparison of local histories. The quantitative theory uses finite readouts, finite thermodynamic states, and chosen temperature and energy data. Extending it to a general interacting field, multiple nonequilibrium baths with distributed control, or a resource bound involving gravity would require new physical hypotheses. The finite results here identify the interfaces at which such hypotheses would enter. In particular, the existence of a persistent section neither fixes a probability of finding it nor supplies the work of sustaining it.

APPENDIX A

Finite information inequalities

This appendix records reusable finite estimates with the support and error conventions used in the main arguments. All equalities are of finite sums. Logarithms are natural, and conditional statements are imposed only on conditioning events of positive probability.

A.1 Divergence and conditional expansion

For $u, v > 0$, $-\log(v/u) \geq 1 - v/u$. If $p \ll q$, summing this at positive p yields

$$\mathsf{D}(p\|q) \geq 1 - \sum_{z:p(z)>0} q(z) \geq 0.$$

Equality requires all ratios on the positive support to be one and q to give no mass to its complement, so $p = q$. The same conclusion with divergence $+\infty$ is immediate when $p \not\ll q$.

For joint laws P, Q on A, B with $P \ll Q$, writing the logarithm of the ratio as the sum of a marginal log ratio and a conditional log ratio gives

$$\mathsf{D}(P\|Q) - \mathsf{D}(P_A\|Q_A) = \sum_a P_A(a)\, D_a,$$
$$D_a = \mathsf{D}(P_{B|a}\|Q_{B|a}).$$

Every conditional row with positive P_A has positive Q_A. Rows with $P_A = 0$ do not enter. The identity implies monotonicity under deterministic projection. Appending a common kernel to both laws leaves divergence unchanged, since the conditional row divergence is zero. Projection after appending the common kernel gives stochastic data processing. Equality holds exactly when the lost conditionals of the first variable agree at every observed output of positive mass.

For finite p on n values, $\mathsf{D}(p\|u) = \log n - \mathsf{H}(p)$ with u uniform. Thus maximum entropy is $\log n$, attained exactly at the uniform law. Conditional subadditivity follows by expanding divergence relative to a product of conditional marginals. This proves the nonnegativity of the total correlation and its zero condition without requiring a positive joint law on the whole product.

A.2 An entropy bound for list decoding

An observation Y, W might specify a list of possible labels rather than a single guess. Let $L(Y, W) \subseteq \{1, \ldots, m\}$ be its list of size exactly k, with $1 \leq k < m$, and let $\varepsilon = \mathbb{P}(X \notin L(Y, W))$. Let B be the indicator of a miss. The chain rule gives

$$\mathsf{H}(X \mid Y, W) \leq h_{\mathrm{b}}(\varepsilon) + (1-\varepsilon)\log k + \varepsilon \log(m-k). \quad \text{(A.1)}$$

Indeed, condition on the miss flag as in Fano's inequality. On a hit the label is one of k, and on a miss it is one of $m - k$. At $k = 1$ this is the usual finite Fano bound. Randomized lists generated independently conditional on Y, W can be incorporated by giving the decoder the seed. For lists of size at most k, either pad to exactly k and compute its new error, or use the weaker $\log m$ bound on the miss term with the original error.

Combining (A.1) with the cut theorem yields a cut bound for list recoverability. An exact list of size k need only cross enough information to reduce source uncertainty to $\log k$ on average. This is relevant for repairing coded records when an observer can first obtain a finite list and then refine the query using an additional check. It does not imply individual exact recovery from a carrier until the remaining ambiguity has been resolved.

A.3 A total variation estimate

For two finite laws p, q on the same alphabet, define

$$\|p - q\|_{\mathrm{TV}} = \tfrac{1}{2} \sum_z |p(z) - q(z)|.$$

It is the supremum of $|p(A) - q(A)|$ over all events, attained on the set where $p \geq q$. Projection to that event gives a binary divergence lower bound. For $u, v \in (0,1)$, the function $\mathsf{D}(\mathrm{Bern}(u) \| \mathrm{Bern}(v))$ has value and first derivative zero at $u = v$, and its second derivative with respect to u is $1/[u(1-u)] \geq 4$. Integrating twice gives a lower bound $2(u-v)^2$, and limits cover the endpoints. Thus

$$\mathsf{D}(p\|q) \geq 2\|p - q\|_{\mathrm{TV}}^2.$$

For example, a collection of separately measured carrier marginals may be proposed to arise from a product coupling. If the measured joint law differs from that product by variation d, its total correlation is at least $2d^2$. This use requires a joint statistic or a bound on it from a known model.

A.4 Coupling and continuity on a fixed alphabet

A coupling with mismatch probability equal to $\|p - q\|_{\mathrm{TV}} = d$ is obtained by taking common mass $\min(p(z), q(z))$ on each diagonal pair, then coupling the remaining disjoint masses. If Z, Z' have that coupling and the alphabet has size $m \geq 2$, Fano's proof gives both $\mathsf{H}(Z \mid Z') \leq f_m(d)$ and the reverse bound. Since

$$\mathsf{H}(Z) - \mathsf{H}(Z') = \mathsf{H}(Z \mid Z') - \mathsf{H}(Z' \mid Z),$$

we obtain $|\mathsf{H}(p) - \mathsf{H}(q)| \leq f_m(d)$. The bound uses the actual mismatch probability. It proves uniform continuity of entropy on each fixed finite alphabet. For a bounded energy on that alphabet, the difference of mean energies is at most $d(\max E - \min E)$ by the same coupling. Thus a sequence of approximating full-support ensembles has free energies converging to the desired finite-state endpoint value. No uniformity across unbounded alphabet size follows from this estimate.

A.5 Independent binary repetitions

Let $N_1, \ldots, N_r$ be iid binary errors of probability $e < 1/2$, and let $K = \sum_i N_i$. For any $\lambda \geq 0$,

$$\mathbb{P}(K \geq r/2) \leq e^{-\lambda r/2}(1 - e + ee^{\lambda})^r$$

by exponential Markov inequality and independence. For $0 < e < 1/2$, choose $e^{\lambda} = (1-e)/e$, giving

$$\mathbb{P}(K \geq r/2) \leq [2\sqrt{e(1-e)}]^r.$$

At $e = 0$ the probability is zero. The estimate counts ties as failures and therefore also bounds the error with a randomized tie break. It is exponential in r for fixed $e < 1/2$, which justifies the convergence of the aggregate decoder in the fork example when its repeated errors are independent conditional on the trunk.

For odd $r = 2j + 1$, the Bayes decision for a fair input with iid symmetric errors is majority. Its error is exactly $\sum_{k=j+1}^{2j+1} \binom{2j+1}{k} e^k (1-e)^{2j+1-k}$. The likelihood ratio for a given observed vector depends only on its number of ones, and changes sign at the midpoint. For a biased input, the log prior ratio shifts that threshold. For a noisy common trunk, repeated observations can estimate the trunk with this accuracy, but recovering the original source additionally inherits the trunk's posterior uncertainty.

A.6 A formula for the correlation of an independent-noise array

For fair X, let $Y_i = X \oplus N_i$ with iid crossover e. If a vector y has k ones, its probability is

$$p_Y(y) = \tfrac{1}{2}\left[e^k(1-e)^{r-k} + e^{r-k}(1-e)^k\right].$$

Call this number u_k. The full entropy is

$$\mathsf{H}(Y) = -\sum_{k=0}^{r} \binom{r}{k} u_k \log u_k.$$

The marginal entropy is $\log 2$ at every coordinate. Thus the unretained total correlation is $r\log 2 - \mathsf{H}(Y)$, while the retained one is $r[\log 2 - h_{\mathrm{b}}(e)]$. Their difference is $\mathsf{I}(X;Y) = \mathsf{H}(Y) - rh_{\mathrm{b}}(e)$, tending to $\log 2$ when $e < 1/2$ by Fano and the repetition bound. These formulas are useful for finite examples where the exact correlation matters more than the elementary lower bound.

APPENDIX B

Finite designs with explicit probabilities

Here we calculate several small models directly. The purpose is to make the support conditions and accounting boundaries reproducible without estimating them from large simulations.

B.1 A four-point energy and correlation ledger

Use the order $00, 01, 10, 11$ for a source and one receiver. Let X be fair and initially prepare the receiver in a pure zero. The incoming law is

$$p_0 = (1/2, 0, 1/2, 0).$$

A controlled XOR with first bit as control swaps states $10, 11$ and fixes the other two. Its output law is $p_1 = (1/2, 0, 0, 1/2)$. At flat energy all states have equal Gibbs mass. The relative entropy of both laws to that Gibbs law is $\log 2$, and the permutation has zero loss and work. Initially the marginal receiver has free-energy excess $k_{\mathrm{B}}T\log 2$. Finally that marginal has zero excess and the joint correlation term has gained $k_{\mathrm{B}}T\log 2$. The law of the source is unchanged.

Now choose a thermal target so the incoming pair law is uniform. At final crossover e, the output is

$$p_e = ((1-e)/2, e/2, e/2, (1-e)/2).$$

Its divergence to equilibrium is $\log 2 - h_{\mathrm{b}}(e)$. This is the work infimum in units of $k_{\mathrm{B}}T$ for $0 < e \le 1/2$ with a controlled bath as in theorem 11.2. At $e = 1/4$, $p_e = (3/8, 1/8, 1/8, 3/8)$ and the dimensionless work infimum is

$$\tfrac{3}{4}\log 3 - \log 2.$$

To verify the simplification, substitute the probabilities in $\mathsf{D}(p_e\|u)$, yielding $\frac{3}{4}\log(3/2) + \frac{1}{4}\log(1/2)$. This is approximately 0.130812 nats. In the pure-blank experiment with a single reversible permutation the work is zero. In the thermal-blank experiment, the initial ensemble is different even though the source and receiver alphabets are the same.

B.2 A complete refresh calculation

Take the restored law $p_{1/8}$ and a passive independent flip with probability $1/4$. The new error from (13.2) is $1/8 + (3/4)(1/4) = 5/16$. The entropy and loss ledger, in nats, is

Boundary	Crossover	Joint entropy
Restored	1/8	$\log 2 + h_{\mathrm{b}}(1/8)$
Degraded	5/16	$\log 2 + h_{\mathrm{b}}(5/16)$
Restored again	1/8	$\log 2 + h_{\mathrm{b}}(1/8)$

The passive step has no work or heat at flat energy and increases the joint entropy by $h_{\mathrm{b}}(5/16) - h_{\mathrm{b}}(1/8)$. The least ideal work to restore it is $k_{\mathrm{B}}T$ times this difference. Both marginals are uniform throughout. A separately equilibrating controller on the receiver alone cannot restore the correlation. The reference-conditioned reset can restore it if the reference and receiver can interact causally.

If the repair is limited to expected jump activity $a > 0$, the full 1-distance between the two pair laws is $2(5/16 - 1/8) = 3/8$. The extra work lower bound from theorem 18.1 is $(9/128)k_{\mathrm{B}}T/a$. This gives a concrete difference between the finite-activity cost and the unconstrained work infimum.

B.3 Two spacelike targets and four conditional cells

Let X be fair, and use two receiving registers with common error 1/4 in each marginal. On $X = 0$ consider the two following conditional assignments, written in the order $00, 01, 10, 11$:

$$\left(\tfrac{9}{16}, \tfrac{3}{16}, \tfrac{3}{16}, \tfrac{1}{16}\right), \qquad \left(\tfrac{3}{4}, 0, 0, \tfrac{1}{4}\right).$$

For $X = 1$, flip both bits in either row. The first law is generated by independent receiver noise and the second by one shared error bit. Each observed pair X, Y_i has the same law under both realizations. The first has conditional total correlation zero and the second has conditional total correlation $h_{\mathrm{b}}(1/4)$. The latter is exactly the entropy deficit relative to the separator coupling for the scopes $\{X, Y_1\}$, $\{X, Y_2\}$.

For the two output bits alone, the independent-error law has disagreement probability 3/8. The shared-error law has disagreement zero, so its unretained total correlation is $\log 2$. In the shared-error law the two bits always agree with each other, but the common value is wrong with probability 1/4. A local comparison of receiver agreement does not measure fidelity to the source. If the source and targets are kept and their energies are flat, the shared-error endpoint costs $k_{\mathrm{B}}T h_{\mathrm{b}}(1/4)$ more free energy than the independent one for the same initial ensemble. Generating the shared noise can itself require a common cause and an admissible causal placement.

B.4 A one-shot coding calculation over five symbols

Take the field $\mathbb{F}_5$, $k = 2$, and evaluation positions $0, 1, 2, 3$ for a two-symbol uniform label (a, b). The four carrier values are $a, a + b, a + 2b, a + 3b$. Any two values u, v at distinct positions i, j recover

$$b = (v - u)/(j - i), \qquad a = u - ib.$$

For a fixed evaluation u at position i, its fiber consists of the five pairs $(u - ib, b)$. A single observer has success 1/5 for the whole label, and two have success one when their data are jointly available. The entropy of any triple of shares is $2 \log 5$ and its total correlation is $\log 5$. For all four the corresponding correlation is $2 \log 5$. Pairwise independence of the four shares follows from invertibility of the two-row map.

If each carrier survives independently with probability 3/4, the probability of failure to recover is the probability that zero or one survives:

$$(1/4)^4 + 4(3/4)(1/4)^3 = 13/256.$$

Thus the recovery probability is 243/256. If instead all four survive together with probability 3/4 and all fail otherwise, each indi-

vidual marginal is unchanged, but the recovery probability is only $3/4$. The comparison concerns failure coupling, not the code itself. Its thermodynamic cost under the thermal-target writing assumptions is $4k_{\mathrm{B}}T\log 5$ for the four symbol targets with retained source. Storing the entire two-symbol label in each of four carriers would cost $8k_{\mathrm{B}}T\log 5$ at the same energies and blank preparation.

B.5 Temporal checks with independent hazards

Suppose R_n is the finite set of histories in which a bit has survived n independent integrity checks with success probability $1-e$ at each check. At every finite horizon the success path exists for $0 \le e < 1$. The inverse limit contains that success path. Its probability under the independent check law is $\lim_n (1-e)^n = 0$ for $0 < e < 1$.

For a contrasting example choose the j-th conditional hazard $q_j = 1/(j+2)^2$ for $j \ge 0$. The product factors telescope:

$$\prod_{j=0}^{n-1}(1-q_j) = \prod_{k=2}^{n+1}\frac{(k-1)(k+1)}{k^2} = \frac{n+2}{2(n+1)}.$$

The infinite survival probability conditional on the initial success is $1/2$. Here positive survival is obtained because the conditional hazards are summable. The event existence in the support is present in both examples. The path law discriminates them.

B.6 A four-state killed process

Consider two live states a, b and two absorbing exit states. The live submatrix is

$$Q = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}.$$

Give the missing $1/4$ from each live row to its corresponding exit. Both rows have live mass $3/4$, so for any initial live law p, $pQ^n\mathbf{1} = (3/4)^n$. The uniform live law is quasi-stationary. The difference of the live-state probabilities decays with the eigenvalue $1/4$ before conditioning; after dividing by survival it decays as $(1/3)^n$. A history can stay alive forever in the support, and the conditioned ensemble tends toward a stable law on a, b. Its unconditioned survival probability tends to zero. The distinction is visible both in eigenvalues and in the inverse-limit witness.

B.7 A majority contract with identical local errors

Let X be fair and consider three outputs. In one design the three errors relative to X are independent with probability $1/3$. In another design exactly one error is chosen uniformly among the three positions, independent of X. Each output separately has error $1/3$ under either design. The independent law gives majority error

$$3(1/3)^2 - 2(1/3)^3 = 7/27.$$

The exactly-one-error law gives exact majority, since two carriers always agree with the source. Its target conditional entropy is $\log 3$, while the independent law has conditional entropy $3h_{\mathrm{b}}(1/3)$. The decrease is $2\log(3/2)$ nats. If all targets start independently fair, the retained-source endpoint cost at flat energies is therefore greater by $2k_{\mathrm{B}}T\log(3/2)$ for the exact majority law.

Conditional on each x, the error probabilities sum to one. Exact majority requires at most one error, so the count is exactly one almost surely. The local errors then fix the weights of the three patterns. A common selector can realize the coupling when each receiver is causally reachable from it.

APPENDIX C

Higher diagrams of reconstruction

This appendix makes the comparison cells used in the proofs explicit. Each diagram is read in spaces or in the indicated slice of the observation topos. A displayed pullback is a homotopy pullback unless the objects have been made discrete. A comparison between two composites is part of the data of a witness. The shape of a square alone does not specify its comparison.

C.1 A path model for the overlap

Let $A \xrightarrow{u} B \xleftarrow{v} C$ be spaces and $PB = B^{[0,1]}$ a path-space model. Write $\mathrm{ev}_{0,1}$ for evaluation at the two endpoints. The constant-path map $B \to PB$ is an equivalence and factors the diagonal $B \to B \times B$. Taking the ordinary pullback of the endpoint fibration gives a model

$$A \times^h_B C \simeq \{(a, c, \gamma) \mid \gamma(0) = u(a),\ \gamma(1) = v(c)\}.$$

This formula is invariant under weak equivalence of appropriately fibrant models. In particular, if A, C are points and their images are the same point b, the pullback is the loop space $\Omega_b B$. Replacing B by its set of components before taking the pullback can change the result.

$$\begin{array}{ccc} A \times^h_B C & \xrightarrow{\ \gamma\ } & PB \\ {\scriptstyle (a,c)}\downarrow & & \downarrow{\scriptstyle \mathrm{ev}_{0,1}} \\ A \times C & \xrightarrow[u \times v]{} & B \times B \end{array}$$

Figure C.1: A witness of a homotopy pullback contains the entire overlap path γ. The right-hand fibration evaluates its endpoints.

For a pair of sections h_1, h_2 over two opens, take $A = \mathcal{H}(U_1)$, $C = \mathcal{H}(U_2)$, $B = \mathcal{H}(U_1 \cap U_2)$, with u, v restrictions. A path in the reconstruction space consists of paths in the two section spaces and a homotopy relating their boundary paths. When computing a fiber of a fixed observation, replace each of these section spaces by $\mathscr{R}_o$. The intersection path then also compares the boundary identifications prescribed at that intersection.

If $B = BG$ and $A = C = *$, there is one pair of local objects and a discrete set G of possible identifications up to homotopy. The result depends on the higher homotopy of B. For $B = K(G, 2)$ with G abelian the same computation gives $K(G, 1)$, where the local pair again has just one class of identifications but

that class has automorphisms. This calculation provides a minimal example of a fibration of matchings that cannot be recovered from sets of components.

C.2 Face maps of a cover

Let $U = \bigcup_i U_i$. In the Čech nerve let

$$U^{[n]} = \coprod_{i_0,\dots,i_n} U_{i_0} \cap \cdots \cap U_{i_n}.$$

Each face drops one index and includes the intersection in the larger one. Degeneracies repeat an index. The first three levels and the augmentation are displayed below. Their geometric realization is U in the topos. Applying $\mathrm{Map}(-, \mathcal{H})$ reverses the arrows, takes coproducts to products, and turns that realization into a totalization.

$$U^{[2]} \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow}\; U^{[1]} \;\rightrightarrows\; U^{[0]} \longrightarrow U$$

Figure C.2: The face maps and augmentation of the cover nerve. Reading from right to left after mapping into $\mathcal{H}$ gives restrictions, pair comparisons, and triple coherences.

At degree zero, a potential point of the totalization consists of the local sections h_i. At degree one, it requires paths between their restrictions. At degree two, it requires paths of paths, with endpoints the relevant composites. One must also satisfy degeneracy conditions. For instance the comparison on a repeated index is identified with the identity. An arbitrary list of pairwise isomorphisms can violate the cocycle condition even when all sections are identical as objects. To see this in a groupoid with one object, choose two pair maps with product $g_2 g_1$, and choose the direct map $g_3 \neq g_2 g_1$. There is no 2-simplex for this triple in the nerve of the groupoid.

For a sheaf of sets, the comparison paths are forced when their endpoints are equal. The diagram reduces to the ordinary equalizer of the two restriction maps from the product over opens to the product over their pairwise intersections. This reduction uses a discrete target, in which any two points have either a unique comparison or none. Passing from a sheaf of spaces to its components need not give the same calculation.

C.3 The first comparison above a triple

At a quadruple intersection write the restricted local objects as h_0, h_1, h_2, h_3, their comparison paths as α_{ij}, and the homotopies of compositions as H_{ijk}. The four triangles form a boundary of a tetrahedron. For coherent gluing the triangles must bound a 3-simplex.

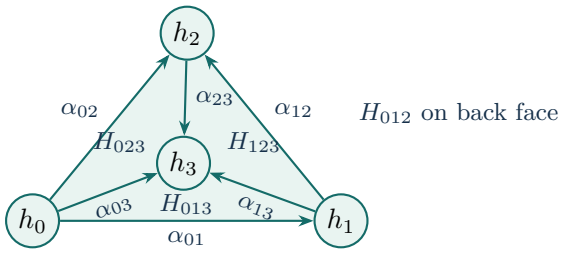


Figure C.3: The four H_{ijk} give the faces of a tetrahedral boundary on U_{0123}. A higher compatibility is a filler of that boundary.

For a space F, the realization of the boundary of a tetrahedron is a sphere S^2, and the tetrahedron is contractible. A specified spherical boundary extends precisely when its map to F is nullhomotopic. When one extension exists, a nullhomotopy transports the boundary to a constant one. The resulting fiber has type $\mathrm{Map}_*(S^3, F) \simeq \Omega^3 F$ at the chosen basepoint, by the cofiber description. This is a calculation in spaces. In a simplicial model the same obstruction is represented using a fibrant replacement and the homotopy fiber of restriction.

For an explicit failure, take $F = K(\mathbb{Z}, 2)$, let all vertices be the basepoint, let every edge be constant, and let three faces be constant. Let the remaining face represent the generator via $\Delta^2/\partial\Delta^2 \simeq S^2$. The resulting tetrahedral boundary has degree one up to sign in $\pi_2 F = \mathbb{Z}$, so it cannot be filled. Specifying all pairwise comparisons and all triple fillers can therefore still fail a higher condition. The same construction can be used in a single intersection space when such choices arise there. It gives an obstruction to a chosen list of witnesses. Other choices of witnesses can have fillers.

C.4 Hypercovers and matching objects

A hypercover augments a simplicial diagram $V_\bullet \to U$ such that $V_0 \to U$ covers and each map from V_n to its n-th matching object is a cover. The matching object consists of lower-dimensional face data with their overlap compatibilities. It can be expressed as the value at n of the coskeleton of the truncated augmented diagram. For the first matching map, the matching object is $V_0 \times_U V_0$. The second matching object describes three edge choices with matching vertices, computed as a limit of the appropriate lower-dimensional diagram. A hypercover can refine that matching data rather than using exactly the Čech intersections.

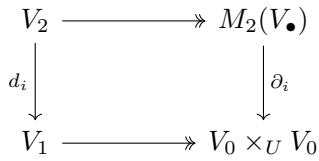


Figure C.4: For a chosen face d_i, the second matching map is compatible with projection to the endpoints of that edge. The horizontal arrows are covers.

The indicated projection from $M_2(V_\bullet)$ factors through the selected edge in V_1. In a hypercomplete topos the augmented diagram presents U after taking its colimit there. An object satisfying hyperdescent has a mapping-space equivalence between its global sections and the totalization over $V_\bullet$. For $\mathscr{R}_o(U, b)$, applying this to history and observation or working in the slice gives the same relative reconstruction formula. An ordinary Čech cover uses the effective-epimorphism argument without the extra hyperdescent hypothesis.

C.5 Cartesian lifting and its test square

Let $P : \mathcal{C}^{\mathrm{op}} \to \mathrm{Cat}_\infty$ be a presheaf of ∞-categories. Under unstraightening it corresponds to a cartesian fibration $p : E \to \mathcal{C}$. The space-valued case of chapter 16 is obtained when each fiber is a groupoid. For $f : U \to V$ and $y \in E_V$, an arrow $e : x \to y$ over f is cartesian if for every object $z \in E$ the following square of mapping spaces is a pullback.

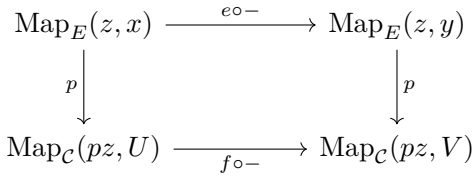


Figure C.5: The mapping-space square of a cartesian lift. Every lift of a composite through y with a specified factor through U factors through x up to contractible choice.

To read the square, fix an arrow $h : pz \to U$ and a map $z \to y$ lying over fh, including its comparison in the base. A point in the fiber over those data is a map $z \to x$ over h with a comparison to the given $z \to y$. Contractibility of the fiber is the lifting property. In particular, two cartesian lifts over the same base arrow with the same endpoint have equivalent domains in the fiber over U, with a contractible space of equivalences respecting the lift data. This proves the coherence of changes in representatives of restrictions. It is a universal property of a comparison. The placement of an actual memory transition is governed by the causal paths in the physical model.

C.6 An observation and its lifts under base change

Suppose $T \to B' \to B$ and $H \to B$ are objects and arrows in the observation topos. Let $H' = B' \times_B H$, $H_T = T \times_B H$. A point of the space of sections over T is a lift of T into H' over B', and by pullback pasting the same space is a lift into H over B. The section triangle can be depicted by including the lift into H_T.

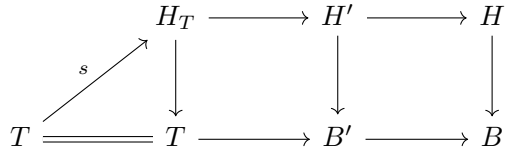


Figure C.6: A chosen lift s is a section over T. Composing horizontally gives the same witness under the two descriptions of the parameter base.

A different lift over the same boundary can change the microscopic realization while preserving the observation. At the level of regional parameter dependence, pullback stability ensures a coherent identification for every restriction $T' \to T$. If a statistical model is also supplied, one can push it forward along observation maps or disintegrate it over a chosen observation when this is defined. Those operations need a law. The equivalence of section spaces by itself identifies the possible witnesses.

C.7 Pasting spans and witnessing associativity

For composable spans with apices H_1, H_2, H_3, the apex of the composite is the limit of a zigzag. A point of the limit consists of points in the H_i and paths in the interface spaces matching consecutive endpoints. The two iterated pullbacks compute that same limit.

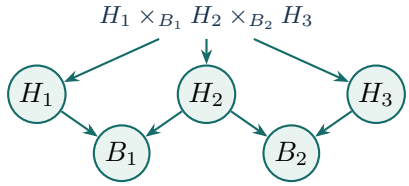


Figure C.7: The composite trace is a limit cone over the zigzag of apices and interfaces. Its comparisons identify the two composites to B_1 and the two to B_2.

For four spans, write ij for the composite between successive apices and use parentheses to indicate successive pullbacks. Let $P_1 = (((12)3)4)$, $P_2 = ((1(23))4)$, $P_3 = (1((23)4))$, $P_4 = (1(2(34)))$, and $P_5 = ((12)(34))$. The standard associator edges connect these five parenthesizations.

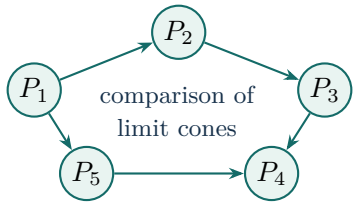


Figure C.8: The associator pentagon for four composable spans. Both paths from P_1 to P_4 compare cones over the same zigzag.

The space of limit cones over a fixed diagram, with the cone structure retained, is contractible when a limit exists. Each edge of the pentagon is the comparison of two cones inducing the identity on the underlying diagram. Its two composites are therefore homotopic as comparisons of limit cones. The same argument gives the higher coherences for longer sequences. It specifies how any particular trace, including the middle identifications, is transported between evaluation orders.

There is also an interchange with parallel composition. Given $H_1 \times_{B_1} H_2$ and $K_1 \times_{C_1} K_2$, their product is equivalent to $(H_1 \times K_1) \times_{B_1 \times C_1} (H_2 \times K_2)$. Indeed, a test map into either side is a pair of test maps into the original two pullback cones. This comparison is natural in all the spans. It justifies the monoidal structural diagram for support histories. A weighted version with product distributions requires independence of the initial parallel systems or a specification of their joint law.

C.8 Computing a limit cone for persistence

For a tower (R_n, ρ_n), put $P = \prod_n R_n$ and $s((r_n)) = (\rho_n(r_{n+1}))$. The two maps $\mathrm{id}, s : P \rightrightarrows P$ have a homotopy equalizer whose path component data were used in chapter 6. A point consists of a sequence r and a path in P from r to $s(r)$. Product path spaces identify that path with the consecutive comparisons.

$$\begin{array}{ccc} \operatorname{holim} R_n & \longrightarrow & P^{[0,1]} \\ \downarrow & & \downarrow{\scriptstyle \mathrm{ev}_{0,1}} \\ P & \xrightarrow[(\mathrm{id},s)]{} & P \times P \end{array}$$

Figure C.9: The homotopy equalizer models the limit of the tower. The path γ in $P^{[0,1]}$ simultaneously witnesses all consecutive restriction comparisons.

To produce a tower of actual fibrations, start with a fibrant model R'_0 of R_0. Given R'_n and an equivalence from R_n, factor the composite $R_{n+1} \to R'_n$ through its mapping-path object. That object has points (x, γ) with $\gamma(0)$ equal to the image of x. Its projection to $\gamma(1)$ is a fibration, and insertion of constant paths

gives an equivalence from R_{n+1}. Iteration produces a levelwise equivalent tower with fibration bonding maps. Its strict limit computes the homotopy limit. This method and the homotopy equalizer both retain the comparisons that ordinary pointwise equality would omit.

Suppose each R_n is itself the homotopy fiber of a history map at b_n and the b_n form a coherent cone. Then one can take the fiber at the limiting observation after taking both limits, or first take every level's fiber and then the limit. The two constructions are equivalent because limits commute. This permits a layered calculation: determine the possible boundary observations at every horizon, choose their coherent restriction data, then take the space of compatible continuations above that choice.

$$\begin{array}{ccc} \lim_n R_n & \longrightarrow & \lim_n \mathcal{H}(U_n) \\ \downarrow & & \downarrow \\ * & \xrightarrow[(b_n)]{} & \lim_n B(U_n) \end{array}$$

Figure C.10: Limiting a diagram of observation fibers gives the fiber of the limiting observation. The lower arrow includes the comparison data of the boundary cone.

C.9 Components do not preserve this pullback

Take the pointed space BG for a nontrivial discrete group. There is a comparison from the set of components of the pullback of its two basepoint maps to the pullback of their sets of components. It has source G and target a singleton.

$$\begin{array}{ccc} \pi_0(* \times^h_{BG} *) & \xrightarrow{\simeq} & G_{\mathrm{disc}} \\ \downarrow & & \downarrow \\ \pi_0(*) \times_{\pi_0(BG)} \pi_0(*) & \xrightarrow[\simeq]{} & * \end{array}$$

Figure C.11: Taking components before gluing forgets the possible identifications around the loop in BG. The downward comparison need not be invertible.

The same phenomenon occurs when one classifies visible labels by taking a discrete quotient of the histories. The quotient can commute with one particular gluing if the overlap fibers have appropriate connectedness. To obtain a valid discrete calculation one can specify an actual 0-truncated sheaf of observations and take its fibers in that layer. To compute lift multiplicity or automorphisms, take fibers in the history layer first. This ordering matters even when the terminal labels are ordinary bits.

C.10 A classifying square for extensions

Take an untwisted classifying map $k : B \to K(G, n+1)$ and its homotopy fiber H. For each observation $b : U \to B$, a lift is a section of $U \times_B H \to U$, which is equivalently a nullhomotopy of kb. The following pasting displays both constructions.

$$\begin{array}{ccccc} U \times_B H & \longrightarrow & H & \longrightarrow & * \\ \downarrow & & \downarrow & & \downarrow{\scriptstyle 0} \\ U & \xrightarrow[b]{} & B & \xrightarrow[k]{} & K(G, n+1) \end{array}$$

Figure C.12: The outer pullback is the family of paths from kb to the basepoint. A section over U gives the nullhomotopy required to lift the observation.

If U has the homotopy type of S^{n+1} and kb represents a nonzero element of G, there is no lift. If kb is constant, the space of lifts is $\mathrm{Map}(U, K(G, n))$. For instance take $U \simeq S^n$; then unbased maps have a factor $K(G, n)$ and based components G. These calculations separate existence of a lift, existence of a unique component of lifts, and contractibility of the whole lift space. They also provide a family whose higher symmetry can be varied through the degree n.

C.11 The long exact comparison at an observation

For a pointed map $o : H \to B$ with fiber F at the basepoint, there is a sequence of spaces

$$\Omega H \longrightarrow \Omega B \longrightarrow F \longrightarrow H \longrightarrow B.$$

The arrow from ΩB to F changes the boundary identification of a fixed lift by a loop. For example, use $H = *$ and $B = BG$. That arrow is an equivalence, reproducing the fact that all the fiber choices are identifications of the single history with its observed value. More generally, exactness on homotopy groups says that a boundary loop induced from a history loop does not change the lift component. A boundary loop not in that image can change it.

$$\begin{array}{ccccc} \Omega B & \longrightarrow & F & \longrightarrow & * \\ \downarrow & & \downarrow & & \downarrow{\scriptstyle b} \\ * & \xrightarrow[h]{} & H & \xrightarrow[o]{} & B \end{array}$$

Figure C.13: The right pullback defines the fiber. Given a chosen history h and a comparison $o(h) \simeq b$, the outer pullback identifies the fiber of $F \to H$ with ΩB.

Indeed, the outer square is equivalent to the pullback of the two basepoint maps after identifying the endpoints by the comparison. This also shows explicitly the dependence on an initial chosen history. Without such a point the reconstruction fiber may be empty and the based exact sequence cannot be initialized there.

C.12 A local image versus a chosen section

For a map $R \to B$, the relative image factors it as an effective epimorphism $R \to I$ followed by a monomorphism $I \hookrightarrow B$. Over a point of I, the statement that R has a lift is local in the internal sense. A global lift of B is a section $B \to R$, which includes consistent choices. The two-sheeted circle example gives a nonsplit effective epimorphism in a sheaf topos: its image is the entire base but it has no global chosen sheet.

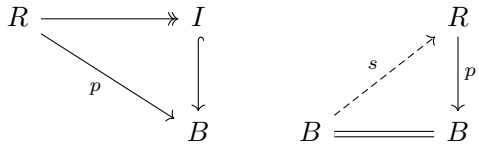


Figure C.14: The image factorization records local existence. The dashed arrow in the second diagram would supply a single coherent selection of witnesses.

In a finite implementation, the corresponding uniform selection can be the program implementing the protocol. A stochastic implementation instead supplies a conditional ensemble on the fiber, potentially with several witnesses at every parameter. Its probabilities depend on seed choices and resources. The category of sections models which realizations are available. The finite kernels and work ledger of the main text specify how a particular selection is realized.

C.13 A groupoid quotient with and without a free action

Let a finite group G act on a finite set A. Its action groupoid has objects $a \in A$ and arrows $a \to a'$ labeled by g with $ga = a'$. Its nerve $A//G$ maps to BG, with fiber A. A point in the fiber over the unique base object includes the identification of its image with that object. Each orbit in A contributes one component, equivalent to BG_a for a representative a, where G_a is its stabilizer. To see this, every object in the orbit is isomorphic to a, and its full one-object subgroupoid has precisely G_a as automorphisms. This inclusion is fully faithful and essentially surjective.

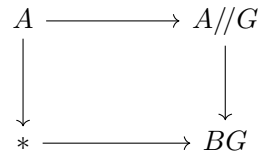


Figure C.15: The homotopy fiber of the action groupoid over the basepoint recovers the original set of carrier configurations. The monodromy on the fiber is the G-action.

For $G = S_r$ permuting identical carrier positions, a configuration with distinct tagged values has trivial stabilizer. A configuration fixed under a permutation has a corresponding loop in the quotient. With all tags equal, the stabilizer is S_r. The ordinary orbit set records each such configuration as one orbit and forgets the difference between a free orbit and one with symmetries. Neither the size of the stabilizer nor the number of orbits determines the number of physical sites unless those sites are part of the original object A. This is the finite version of retaining symmetry data while keeping the energetic register count explicit.

C.14 Transition histories and the two boundary maps

Take a finite relation $R \subseteq A \times B$, viewed as the trace support of a stochastic transition. The projections π_A, π_B give a span. A point of R specifies one admissible input–output pair. If R is replaced by a space of detailed traces, there can be many traces with the same pair. For instance a measurement outcome can be the same at a screen whether a stored bit was directly transported or reconstructed from several shares. The history map to the boundary forgets that choice, while the total trace retains it.

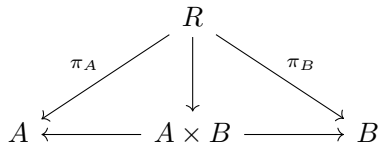


Figure C.16: The pair of boundary maps defines a single observation $R \to A \times B$. Its homotopy fiber at a pair is the space of histories implementing that pair.

A stochastic law over the input can be extended by a conditional distribution on the trace fiber over that input. The marginal on B is the resulting output law. For finite states and an input a with a nonempty finite trace fiber, one may choose any normalized conditional law on it. Distinct choices can induce the same conditional output kernel while differing in intermediate energy or source access. For a general higher fiber, the assignment of a law is additional structure and does not follow from a point of the homotopy limit.

C.15 A small higher descent computation with a loop

Let a good cover of a circle have three arcs U_0, U_1, U_2, with each pair overlapping in a contractible arc and no triple overlap. The nerve of the cover is a three-cycle. Take a constant local system of groupoids with one-object fiber BG. In a local trivialization, choose a comparison g_{ij} on each edge oriented cyclically. There is no triple-face condition because the triple intersection is empty. A change of local trivializations by $a_i \in G$ sends g_{ij} to $a_j g_{ij} a_i^{-1}$. Thus the cyclic composite $g_{20}g_{12}g_{01}$ changes by conjugation by a_0. Choosing two of the trivializations gauges two edge comparisons to the identity, and the remaining comparison represents the cyclic composite. This computes the conjugation groupoid directly from the descent data.

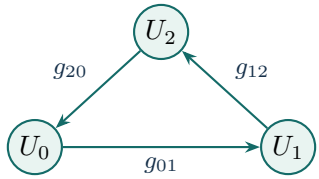


Figure C.17: A cycle in the cover nerve with no triple intersection. Its composite is a holonomy, while the filled triangle in the earlier overlap diagram imposes a relation on composites.

If a triple intersection is added and its restriction fibers have the same local model, a compatible triangle requires the product around its boundary to be the identity. This describes concretely how incidence of intersections changes the global reconstruction. A gauge-invariant finite observer might record a conjugacy class, in which case a stochastic realization supplies a law on those classes. The descent computation itself identifies the classes and their stabilizers.

APPENDIX D

Operational certificates for finite models

For a finite realization one can verify the implications used in the manuscript with finite data. This appendix describes certificates and gives their relation to the categorical construction. In particular, positivity of a certificate of stochastic success depends on the weights as well as the admissible support.

D.1 From constraints to a probability table

A finite model can be described by a list of vertices with alphabet sizes, a list of parent sets in topological order, and a rational table for each kernel. Verify that

every table entry is nonnegative and each conditional row sums to one. Include the root joint law, and specify which root values are treated as exogenous when an intervention is tested. Also include local observation functions and a decoding criterion. A trace probability is a product of these rational numbers. Finite sums give exactly the error probabilities and all marginals used by the decoding and entropy calculations. Logarithms enter the entropy evaluation only after those rational probabilities are formed.

For large products it is often preferable to eliminate variables before enumerating full traces. Suppose a hidden variable v appears in several factors. Multiply the factors containing it, sum over v, and replace them by the resulting factor on the union of their scopes minus v. This leaves the marginal on all other variables unchanged by distributivity of the sum. Repeating eliminates exactly the intended hidden variables. The largest intermediate scope determines the dominant size of an exact table. In a tree of separators, one can instead pass conditional weights or messages between adjacent scopes, maintaining the stated interface. The junction entropy calculation does not require storing the entire joint table when the needed logarithmic terms are fixed by those margins.

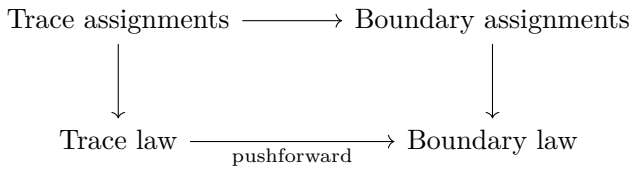


Figure D.1: The upper arrow forgets hidden variables. Applying it to a specified ensemble gives the lower marginal calculation. The vertical steps indicate supplying a law and its induced law, rather than canonical maps from sets to measures.

The support of each rational factor determines a local constraint. Their global conjunction is precisely the support of the joint product when root constraints are included. One may verify successful traces either by enumeration or by working with intersections and restrictions on the finite site. The agreement of those methods is proposition 2.2. In the unweighted setting only the truth of the compatibility predicate is computed. In the weighted setting the rational trace factors determine the probability of the compatible assignments.

D.2 Witnesses and countercertificates

For existence of a finite global history, a single assignment of all variable values certifies that every constraint can be met. The checker evaluates each finite relation on its scope. For nonexistence, enumeration of a finite search space gives a decision, or one may produce a structured contradiction such as the

odd triangle in chapter 8. If the issue is existence of a joint law with specified marginals, a feasible joint table is a positive certificate. An inequality of the form (8.4) with a strict violation is a negative certificate. For rational input data it can be expressed using rational coefficients: the marginal hull is a rational polytope and a strict separating functional can be approximated by rational coefficients while preserving the strict gap.

For a cut bound, give the order ideal, list all transmitted alphabets, and identify the accessible outside seeds. A claimed zero-error decoder of a uniform m-label problem whose cut has $d < m$ total possible states is infeasible in the absence of side information. The counting argument is that distinct source labels require disjoint positive supports on the cut, or else the same downstream kernel would have to give both labels. With side information, the conditional support and the ability of the sender to adapt to it decide whether a counting bound is sufficient. The graph-coloring certificate of theorem 24.2 covers the one-message encoder that lacks the side information.

For a finite-horizon persistence failure, an empty reconstruction level certifies failure of the full infinite contract. A cycle in an allowed-state finite transition graph is a finite positive certificate of persistence when it is reachable from the chosen initial condition. For general computable towers of finite sets, existence of an infinite path need not have a finite decidable certificate. This is consistent with the distinction between a proof of the compactness theorem and a method for making its choices effective.

D.3 A robust bound from approximate marginals

If a law is specified experimentally, computed marginals may differ from their nominal values. In the finite mathematical model one can derive sensitivity bounds from a declared total variation tolerance. If the full law is within distance δ of a nominal law, every marginal is within δ by data processing for variation. This follows because each marginal event pulls back to an event of the full sample space. The probability of a fixed decoding error changes by at most δ. To bound changes in conditional entropy, expand it as a difference of two unconditional entropies. For variables A, B with fixed alphabets of sizes $a, b \geq 2$,

$$|\mathsf{H}_p(A \mid B) - \mathsf{H}_q(A \mid B)| \leq \overline{f}_{ab}(\delta) + \overline{f}_b(\delta), \qquad \overline{f}_m(\delta) = f_m(\min(\delta, 1 - 1/m)).$$

Use the coupling continuity bound of chapter A on the joint and on the marginal. The monotone envelope is necessary when the actual marginal distances are only upper bounded. The same method can bound the change in a finite total-correlation ledger by applying the estimate to each entropy term. A bound on

the separate marginals alone gives no such estimate on joint error or correlation because couplings with the same marginals may differ in full variation.

For a controlled work protocol, if both the initial and final states are specified within a declared tolerance and the energy is bounded on the corresponding finite alphabets, the endpoint free-energy estimate has an explicit uncertainty range. For each endpoint it is at most

$$\delta\,(\max E - \min E) + k_{\mathrm{B}}T\,\overline{f}_{|S|}(\delta).$$

This follows from the energy expectation and entropy continuity bounds. The two endpoint errors add by the triangle inequality. The bound refers to endpoint free energy, while estimating actual work also requires controlling the stage energies and laws in the quench expectation.

D.4 Verifying a work ledger without a microscopic trajectory sample

For a finite isothermal stage, provide the energy table E_j, incoming law p_{j-1}, and a fixed-energy transition table K_j. Its Gibbs distribution can be computed from the energy table. The stationarity equation is a finite matrix equation. If the numbers are represented via rational Gibbs weights and the energies are defined by their negative logarithms plus a constant, stationarity can be verified exactly in rational arithmetic. A particularly simple case uses a flat Hamiltonian, for which stationarity is the assertion that the kernel is doubly stochastic. Another uses conditional block resets with rational Gibbs conditionals and the chosen block weights.

The mean work for the quench is computed from p_{j-1} and the energy difference. The stage loss can be evaluated from the relative entropies before and after the kernel. Providing exact rational laws and specifying each energy as a logarithm gives formulas in elementary constants; decimal displays can be checked numerically to any chosen precision. The identity (10.2) supplies the check that these differences telescope. For a physical device the premises that the quench is the only drive work and that fixed-energy changes are heat come from the device model. A numerical equality in the declared ledger cannot by itself verify those premises.

For a toy exact certificate, choose two states, a flat energy E_0, and a quench to E_1 with values $0, k_{\mathrm{B}}T\log 3$. Then the Gibbs law is $(3/4, 1/4)$. Start at $(1/2, 1/2)$, and after the quench fully reset to the Gibbs law. The input work is $\frac{1}{2}k_{\mathrm{B}}T\log 3$. The reset loss is $\mathsf{D}((1/2,1/2)\|(3/4,1/4)) = \log 2 - \frac{1}{2}\log 3$. Quench back to flat energy at the new law for work $-\frac{1}{4}k_{\mathrm{B}}T\log 3$. The net work is $\frac{1}{4}k_{\mathrm{B}}T\log 3$,

while the endpoint free-energy increase is $k_{\mathrm{B}}T[\frac{3}{4}\log 3 - \log 2]$. Their difference is exactly the loss just computed, multiplied by $k_{\mathrm{B}}T$. This evaluates a finite protocol with a strict gap between actual work and the reversible infimum.

D.5 Separating semantic checks from resource checks

A causal placement certificate can assign each graph edge a permitted future-directed path. For a finite circuit placed in Minkowski coordinates with straight wires, checking $t(v) - t(u) \geq \|\mathbf{x}(v) - \mathbf{x}(u)\|$ and a positive time difference suffices for the corresponding arrow. A field model may require more geometry, but the graph algorithms still operate after that placement has been established. Checking a joint decoding law then uses the available register variables at the observation points. Finally checking the work uses all energetic degrees of freedom at each endpoint, including those deliberately left outside a coarse readout. A useful finite report gives each of those objects its own boundary list.

For a higher reconstruction, a comparable certificate may specify a presentation of the site, the history and observation maps, and local sections with their chosen comparisons. The atlas shows where the 2-simplices, 3-simplices, and path fibers enter such a certificate. In a truncated model the required comparison dimensions can be bounded using the truncation level. In a general homotopy type an arbitrary finite list of pairwise comparisons is not a certificate of all higher conditions. One can instead provide a coherent functor into the relevant total category, or prove that the local comparisons form the image of a genuine global section. Such a semantic certificate does not require a law or energy until quantitative claims are made.

D.6 Reproducing the examples

The accompanying source contains a small independent calculation of finite examples. It enumerates a binary fork, checks the separator entropy deficit, computes the two-of-four field code, evaluates the refresh example, and compares the exact work ledger for the two-state quench. These are reproducibility checks of explicitly stated finite laws. The general claims are established by the proofs, including the compactness and coherence results that do not reduce to those examples.

References

[1] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13:113036, 2011. arXiv:1102.0264.

[2] Rudolf Ahlswede, Ning Cai, Shuo-Yen Robert Li, and Raymond W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000. DOI.

[3] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76:2818–2821, 1996. Full text.

[4] Charles H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17(6):525–532, 1973. Full text.

[5] Antonio N. Bernal and Miguel Sánchez. Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes. *Communications in Mathematical Physics*, 257:43–50, 2005. arXiv:gr-qc/0401112.

[6] Romeo Brunetti, Klaus Fredenhagen, and Rainer Verch. The generally covariant locality principle: A new paradigm for local quantum physics. *Communications in Mathematical Physics*, 237:31–68, 2003. arXiv:math-ph/0112041.

[7] Alexander B. Boyd, Dibyendu Mandal, and James P. Crutchfield. Thermodynamics of modularity: Structural costs beyond the Landauer bound. *Physical Review X*, 8:031036, 2018. Repository record.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*, second edition. Wiley, 2006. Publisher record.

[9] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability*, 2(1):88–100, 1965. Publisher record.

[10] Jeremy L. England. Statistical physics of self-replication. *Journal of Chemical Physics*, 139:121923, 2013. MIT repository.

[11] Massimiliano Esposito and Christian Van den Broeck. Three faces of the second law. I. Master equation formulation. *Physical Review E*, 82:011143, 2010. arXiv:1005.1683.

[12] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the Ising model. *Annals of Applied Probability*, 10(2):410–433, 2000. DOI.

[13] Tobias Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 2020. arXiv:1908.07021.

[14] Péter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):119–162, 1973. Full text.

[15] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002. Author's edition.

[16] J. M. E. Hyland. The effective topos. In *The L. E. J. Brouwer Centenary Symposium*, Studies in Logic and the Foundations of Mathematics, 110:165–216. North-Holland, 1982. Full text.

[17] Leslie Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, 1978. Author's record.

[18] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961. DOI.

[19] Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996. Bibliographic record.

[20] Göran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40:147–151, 1975. Publisher record.

[21] Jacob Lurie. *Higher Topos Theory*. Princeton University Press, 2009. arXiv:math/0608040.

[22] Saunders Mac Lane and Ieke Moerdijk. *Sheaves in Geometry and Logic*. Springer, 1992. Publisher record.

[23] James L. Massey. Causality, feedback and directed information. In *Proceedings of the International Symposium on Information Theory and its Applications*, pp. 303–305, 1990. Full text.

[24] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature Physics*, 11:131–139, 2015. Publisher record.

[25] Judea Pearl. *Causality: Models, Reasoning, and Inference*, second edition. Cambridge University Press, 2009. DOI.

[26] Nicholas Pippenger. Reliable computation by formulas in the presence of noise. *IEEE Transactions on Information Theory*, 34(2):194–197, 1988. Repository record.

[27] Emily Riehl and Dominic Verity. *Elements of ∞-Category Theory*. Cambridge University Press, 2022. Publisher record.

[28] Patrick Schultz and David I. Spivak. *Temporal Type Theory: A Topos-Theoretic Approach to Systems and Behavior*. Birkhäuser, 2019. arXiv:1710.10258.

[29] Patrick Schultz, David I. Spivak, and Christina Vasilakopoulou. Dynamical systems and sheaves. *Applied Categorical Structures*, 28:1–57, 2020. arXiv:1609.08086.

[30] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75:126001, 2012. arXiv:1205.4176.

[31] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, 1948. Full text.

[32] Naoto Shiraishi, Ken Funo, and Keiji Saito. Speed limit for classical stochastic processes. *Physical Review Letters*, 121:070601, 2018. arXiv:1802.06554.

[33] Glynn Winskel. Event structures. In *Advances in Petri Nets 1986*, Lecture Notes in Computer Science, 255:325–392. Springer, 1987. Full text.

[34] William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982. DOI.