Correlation Boundaries and the First Law: Entanglement Anisotropy, Holographic Screens, and the Thermodynamics of Surfaces

David Neale Goleudy.ai, Rochester, New York

Abstract

We introduce the concept of correlation boundaries — surfaces across which the total pairwise correlation vanishes — and show that these are the only surfaces on which an analog of the first law of thermodynamics holds generically. For surfaces with nonzero cross-boundary correlation $P(\Sigma)$, we derive a thermodynamic deficit $\Delta = P \cdot A$ proportional to the product of the cross-boundary entanglement entropy and the entanglement anisotropy. This result resolves the finding of Wang and Braunstein [Nat. Commun. 9, 2977 (2018)] that holographic screens away from horizons are not thermodynamic: horizons are correlation boundaries ($P = 0$), stretched horizons have $P \approx 0$, and spherically symmetric screens have vanishing anisotropy ($A = 0$) despite nonzero P. We identify the cross-boundary correlation $P(\Sigma)$ with the entanglement entropy S_{EE} across Σ, and show that the Ryu-Takayanagi formula emerges as the minimization of P over surfaces homologous to a boundary region. The framework requires only pairwise correlations with a monogamy (finite capacity) constraint, making it independent of specific microscopic dynamics. We discuss implications for the emergent gravity program, the quantum extremal surface prescription, and the relationship between entanglement structure and gravitational thermodynamics.

I. INTRODUCTION

The deep connection between gravity and thermodynamics, first glimpsed in the laws of black hole mechanics [1] and the Bekenstein-Hawking entropy formula [2,3], has been developed into a powerful program for understanding spacetime as emergent. Jacobson's 1995 derivation of the Einstein equations from thermodynamic relations at local Rindler horizons [4] showed that gravitational dynamics follow from assuming that $\delta Q = TdS$ holds at causal horizons, with entropy proportional to horizon area.

Verlinde [5,6] extended this program by proposing that gravity is an entropic force arising from thermodynamic properties of holographic screens — surfaces of constant Newtonian potential that encode information about the enclosed region. This entropic gravity program makes strong claims: it aims to derive Newton's laws, reproduce general relativity, and explain dark matter effects from the competition between area-law and volume-law entanglement entropy at cosmological scales [6].

A central assumption of Verlinde's program is that holographic screens — ordinary surfaces away from horizons — are intrinsically thermodynamic, possessing a well-defined temperature $T = \kappa/(2\pi)$ and an entropy proportional to area. Wang and Braunstein [7] tested this assumption rigorously. For static asymptotically flat spacetimes, they proved that the first law holds on horizons, holds approximately on stretched horizons, but fails on ordinary surfaces away from horizons. The sole exception is spherically symmetric configurations, where ordinary surfaces do satisfy the first law.

This result has been widely interpreted as a serious challenge to the emergent gravity program. We argue that it is not a challenge but a prediction — one that follows from a simple structural principle: surfaces are thermodynamic if and only if they are correlation boundaries, defined as surfaces across which the total pairwise correlation vanishes. We develop this principle, derive the thermodynamic deficit for general surfaces, connect it to entanglement entropy, and show that the Ryu-Takayanagi formula [8] for holographic entanglement entropy emerges naturally.

The paper is organized as follows. Section II introduces correlation fields with monogamy constraints and defines correlation boundaries. Section III derives the thermodynamic deficit theorem. Section IV establishes the identification between cross-boundary correlation and entanglement entropy. Section V derives the Ryu-Takayanagi formula from this framework. Section VI presents the resolution of the Wang-Braunstein result. Section VII discusses implications and open questions.

II. CORRELATION FIELDS AND CORRELATION BOUNDARIES

A. Correlation fields with monogamy

Consider a system of N degrees of freedom (which we call features) labeled $i = 1, \ldots, N$. Each pair (i, j) carries a correlation strength $\lambda_{ij} \in [0, \Lambda]$, where Λ is the maximum correlation capacity. We impose the monogamy constraint: each feature has finite total correlation capacity,

$$\sum_{j \neq i} \lambda_{ij} \leq \Lambda \quad \forall\, i. \tag{1}$$

This constraint is not an assumption about a specific physical system — it is a structural feature shared by quantum entanglement [9], mutual information in classical systems [10], and correlation measures in tensor networks [11]. In quantum mechanics, monogamy of entanglement ensures that if subsystem A is maximally entangled with B, it cannot be significantly entangled with C [12]. The constraint (1) is the general form of this principle.

In the continuum limit (large N, with features densely distributed), the discrete correlation field λ_{ij} becomes a two-point function $\lambda(x, y)$ on a manifold, with the monogamy constraint becoming

$$\int \lambda(x, y)\, d\mu(y) \leq \Lambda \quad \forall\, x. \tag{2}$$

The continuum limit assumes that the feature density n is sufficiently large and uniform that discrete sums transition to integrals without topological defects. This requires $n \cdot \xi^d \gg 1$, where ξ is the correlation length and d is the spatial dimension — each correlation volume contains many features. The monogamy constraint scales cleanly under this limit: the per-feature budget Λ is fixed while the number of partners within correlation range grows as $n \cdot \xi^d$, so individual correlations scale as $\lambda \sim \Lambda/(n \cdot \xi^d)$, ensuring the weak-correlation regime required by the first-order expansion (Sec. III.D, Step 4). No specific geometric embedding is assumed beyond the existence of a smooth density n(x); the formalism applies to any manifold structure. Throughout, we work with both discrete and continuum descriptions as appropriate.

B. Surfaces and cross-boundary correlation

A surface Σ is a codimension-1 partition of the features into two sets: Inside (I) and Outside (O). For any such surface, define the cross-boundary correlation:

$$P(\Sigma) = \sum_{i \in I} \sum_{j \in O} \lambda_{ij}. \tag{3}$$

This non-negative scalar measures the total correlation threading through Σ. In the continuum:

$$P(\Sigma) = \int_\Sigma \int_O \lambda(x, y)\, d\mu_\Sigma(x)\, d\mu_O(y). \tag{4}$$

C. Correlation boundaries

Definition 1. A surface Σ is a *correlation boundary* if $P(\Sigma) = 0$.

At a correlation boundary, every feature inside is uncorrelated with every feature outside. The surface completely isolates the interior from the exterior in terms of pairwise correlations.

Definition 2. A surface Σ is an *ε-approximate correlation boundary* (or *stretched correlation boundary*) if $P(\Sigma)/P_{max}(\Sigma) < \varepsilon$, where P_{max} is the maximum cross-boundary correlation the surface could support.

Definition 3. A surface Σ is *ordinary* if $P(\Sigma)/P_{max}(\Sigma) \sim O(1)$.

The monogamy constraint (1) ensures that correlation boundaries exist. A feature's finite correlation budget Λ cannot extend to arbitrarily many partners. At some point, the budget is exhausted and correlations reach zero — creating a boundary.

D. Connection to horizons

In a gravitational context, black hole horizons are correlation boundaries: an observer outside the horizon has zero quantum correlation with degrees of freedom behind it (up to Hawking radiation, which represents the small but nonzero P of an approximate correlation boundary). The Rindler horizon of an accelerating observer is also a correlation boundary — by construction, it is the limit of the observer's causal (and therefore correlational) reach.

This identification is not new. The connection between causal horizons and entanglement structure has been developed extensively in the context of holographic entanglement entropy [8,13,14] and the ER=EPR conjecture [15]. What we add is a precise criterion — $P(\Sigma) = 0$ — that can be tested for any surface, not just horizons, and that makes quantitative predictions about the degree to which a surface fails to be thermodynamic.

III. THE THERMODYNAMIC DEFICIT

A. Energy partition across a surface

For a system partitioned by a surface Σ into I and O, the total energy decomposes as

$$E_{total} = E_I + E_O + E_{cross}(\{\lambda_{ij}\}_{i\in I, j\in O}), \tag{5}$$

where E_cross is the energy associated with cross-boundary correlations. Similarly, for any extensive potential that is additive under statistical independence and monotonic in the number of accessible configurations. In gravitational thermodynamics, this potential plays the role of the generalized free energy of the correlation field — it reduces to the Helmholtz free energy $F = E - TS$ for equilibrium systems and to the Bekenstein-Hawking entropy (times temperature) for horizon thermodynamics. The proof of Theorem 1 requires only three structural properties of the potential: additivity under independence (P1), smoothness (P2), and monotonicity in accessible configurations (P3). These are satisfied by a broad class of thermodynamic potentials, including the von Neumann entropy, free energy, and information-theoretic functionals. The results are independent of the specific choice:

$$\Phi_{total} = \Phi_I + \Phi_O + \Phi_{cross}. \tag{6}$$

When $P(\Sigma) = 0$, there are no cross-boundary correlations, so E_cross = 0 and Φ_cross = 0. The partition is informationally clean: Φ_total = Φ_I + Φ_O exactly.

When $P(\Sigma) > 0$, the cross term is nonzero. To leading order in the cross-boundary correlations:

$$\Phi_{cross} = \sum_{i\in I, j\in O} \lambda_{ij} \cdot \varphi_{ij} + O(\lambda^2), \tag{7}$$

where φ_ij is the contribution to the potential per unit correlation between features i and j.

B. Perturbations and the first law

Consider a perturbation that displaces the surface Σ by δε in the outward normal direction n̂(x) at each point x ∈ Σ. This perturbation changes three quantities:

(i) **Surface flux δQ:** the energy crossing the surface due to the displacement.

(ii) **Entropy change TdS:** the change in entropy of the interior, related to internal reconfiguration.

(iii) **Cross-boundary correlation response δΦ_cross:** the change in the cross term due to the perturbation.

The perturbed energy balance is:

$$\delta Q = T\,dS + \delta\Phi_{cross}. \tag{8}$$

The thermodynamic deficit is defined as

$$\Delta_{FL}(\Sigma) \equiv \delta Q - T\,dS = \delta\Phi_{cross}. \tag{9}$$

The first law δQ = TdS holds if and only if δΦ_cross = 0.

C. The correlation anisotropy

When the surface Σ is displaced by δε at point x, features near the surface are reassigned between I and O. The change in cross-boundary correlation depends on how correlations are distributed directionally at x.

Define the correlation current density at x ∈ Σ:

$$\vec{J}(x) = \sum_j \lambda_{xj}\,\hat{r}_{xj}\,\delta(x \in \Sigma), \tag{10}$$

where r̂_xj is the unit vector from x toward feature j. The correlation anisotropy at x is the projection of the net current onto the surface normal:

$$\mathcal{A}(x) = \hat{n}(x) \cdot \left(\vec{J}^{\,out}(x) - \vec{J}^{\,in}(x)\right), \tag{11}$$

where the superscripts denote contributions from outside and inside features respectively.

The anisotropy measures the directional asymmetry of correlations at the surface. If correlations are distributed isotropically, J^out and J^in are balanced and A = 0. If there is a preferred direction, A ≠ 0.

D. The thermodynamic deficit theorem

Theorem 1. For a surface Σ in a correlation field satisfying the monogamy constraint (1), the thermodynamic deficit under a normal displacement δε(x) is

$$\Delta_{FL}(\Sigma) = \int_{\Sigma} P(x) \cdot \mathcal{A}(x) \cdot h[\delta\varepsilon(x)]\, d\mu_{\Sigma} + O(P^2), \tag{12}$$

where P(x) is the local cross-boundary correlation density, A(x) is the local correlation anisotropy, and h is a function of the perturbation profile.

Proof sketch. The argument proceeds in four steps; the complete proof with error bounds is given in Appendix A.

Step 1: Potential decomposition. The additivity property (P1) implies that the total potential decomposes as Φ_total = Φ_I + Φ_O + Φ_cross, where Φ_cross depends only on the cross-boundary correlations {λ_ij : i ∈ I, j ∈ O}. The first law holds if and only if δΦ_cross = 0 under the surface perturbation.

Step 2: Feature reassignment. Displacing Σ by ε · δε(x) in the outward normal transfers features in a thin shell near the surface from O to I. Each reassigned feature at position x changes the cross-boundary correlation set: correlations from α to partners already in I become internal (reducing P), while correlations from α to partners remaining in O become cross-boundary (increasing P). The net change at x is

$$\delta P(x) = C^{out}(x) - C^{in}(x), \tag{12a}$$

where C^out and C^in are the total correlation from x to outside and inside partners respectively.

Step 3: The P · A factorisation. The change in Φ_cross from reassigning features in the shell is, to first order in ε,

$$\delta\Phi_{cross} = \varepsilon\bar{\varphi} \int_{\Sigma} n(x)\, \delta\varepsilon(x) \cdot \left[C^{out}(x) - C^{in}(x)\right] d\mu_{\Sigma} + O(\varepsilon^2, 1/N), \tag{12b}$$

where φ̄ is the mean correlation susceptibility and n(x) is the feature density (Appendix A, Step 3). The factor C^out − C^in is precisely the correlation anisotropy A(x). It vanishes in two cases: (a) when P(x) = 0 (no correlations cross the surface at x, so there is nothing to be asymmetric), or (b) when correlations are isotropic (C^out = C^in, so the directional asymmetry vanishes despite nonzero P). The product structure P · A follows: the deficit requires *both* nonzero cross-boundary correlation *and* directional asymmetry of that correlation.

Step 4: Error control. The first-order expansion is valid when individual cross-boundary correlations are weak ($\lambda_{ij} \ll \Lambda$), which the monogamy constraint ensures at large N since each feature's budget Λ is shared among O(N) partners. A subleading geometric effect — the change in susceptibilities due to the partition shift — contributes at $O(\varepsilon/N)$ relative to the leading term. Both corrections are controlled in the thermodynamic regime ($N \gg 1$, smooth surfaces, small perturbations). ■

For uniform perturbations, Eq. (12) simplifies to

$$\Delta_{FL} \propto \langle P \cdot \mathcal{A} \rangle_{\Sigma}. \tag{13}$$

E. The anisotropy tensor and multipole decomposition

Promoting the scalar anisotropy to a tensor:

$$\mathcal{A}_{ab}(x) = \sum_j \lambda_{xj} \left(\hat{r}_{xj,a}\, \hat{r}_{xj,b} - \frac{1}{d}\delta_{ab} \right), \tag{14}$$

the scalar anisotropy is $A(x) = n^a n^b A_{ab}(x)$. Expanding in spherical harmonics on a surface of approximate spherical topology:

$$\mathcal{A}(x) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mathcal{A}_{lm}\, Y_{lm}(\hat{x}). \tag{15}$$

This decomposition is central to understanding the spherical symmetry exception (Sec. VI).

IV. IDENTIFICATION WITH ENTANGLEMENT ENTROPY

A. Entanglement entropy and cross-boundary correlations

The entanglement entropy of a bipartite quantum system in pure state $|\Psi\rangle$, partitioned into A and B, is

$$S_{EE}(A) = -\mathrm{Tr}(\rho_A \ln \rho_A), \tag{16}$$

where $\rho_A = \mathrm{Tr}_B |\Psi\rangle\langle\Psi|$. For a system described by pairwise correlations λ_{ij}, the entanglement entropy decomposes as

$$S_{EE} = \sum_{i \in I, j \in O} f(\lambda_{ij}) + \text{multi-body corrections}, \quad (17)$$

where $f(\lambda)$ is a monotonically increasing function with $f(0) = 0$ and $f(\lambda) \to \ln d$ as $\lambda \to \Lambda$, where d is the local Hilbert space dimension.

B. The pairwise-dominated regime

In the regime where most cross-boundary correlations are weak, the function f is approximately linear. We now show that the monogamy constraint ensures this regime at large N, and bound the multi-body corrections explicitly.

Pairwise dominance. Each feature has total correlation budget Λ shared among O(N) partners. At large N, individual correlations scale as $\lambda_{ij} \sim \Lambda/(N-1)$. The function f, expanded around zero, gives $f(\lambda) \approx c_1\lambda + c_2\lambda^2 + \ldots$, so the pairwise contribution to S_EE is

$$S_{EE}^{(2)} = c_1 \sum_{i \in I, j \in O} \lambda_{ij} + O(\lambda^2) = c_1 \cdot P(\Sigma) + O(\Lambda^2/N). \quad (18)$$

The $O(\Lambda^2/N)$ correction arises from the $c_2\lambda^2$ terms: there are O(nN) cross-boundary pairs, each contributing $c_2\lambda^2 \sim c_2\Lambda^2/N^2$, giving total $c_2\Lambda^2 n/N \sim O(\Lambda^2/N)$ for a balanced partition.

Multi-body corrections. The leading multi-body term in S_EE involves genuine three-party entanglement — correlations among triples (i, j, k) that cannot be decomposed into pairwise contributions. Each triple contributes at order $\lambda^3 \sim \Lambda^3/N^3$. The number of cross-boundary triples (with at least one member on each side) is $O(n^2(N-n)) \sim O(N^3)$ for a balanced partition, giving a total three-body contribution of $O(\Lambda^3/N^3) \cdot O(N^3) = O(\Lambda^3)$. This is O(1) — comparable to the pairwise term at leading order in N.

However, the three-body contribution factors differently from the pairwise contribution under surface perturbations. For the thermodynamic deficit, what matters is not S_EE itself but its response to surface displacement (Eq. 12). The three-body response involves the change in triples straddling the surface when the surface is displaced, which introduces an additional geometric factor that suppresses the contribution by 1/N relative to the pairwise response (Appendix A, Step 4). Therefore:

$$S_{EE}(\Sigma) = c_1 \cdot P(\Sigma) + O(1/N). \quad (19)$$

This identifies the cross-boundary correlation $P(\Sigma)$ with the entanglement entropy S_EE, up to normalization and corrections that vanish at large N.

C. Where the identification is exact

The identification $P \propto S_{EE}$ is exact in three important cases:

(i) Tensor networks. In discrete tensor network models of holography (e.g., the HaPPY code [16]), each bond carries a fixed entanglement ln d. The entanglement entropy of a boundary region equals the number of bonds cut by the minimal separating surface, times ln d. This is precisely $P(\Sigma)$ with uniform bond weights.

(ii) Free (Gaussian) field theories. For Gaussian states, the entanglement entropy is entirely determined by the two-point correlation function [17], which maps directly to the pairwise λ_{ij}. Multi-body corrections vanish identically.

(iii) Holographic CFTs at leading order in $1/G_N$. The Ryu-Takayanagi entropy is a geometric (area-based) quantity that captures the leading-order contribution to entanglement entropy. This leading-order contribution is pairwise-dominated [18].

D. Multi-body corrections as irreducible content

The discrepancy between $P(\Sigma)$ and S_{EE} at finite N measures genuine multi-partite entanglement — correlations that cannot be decomposed into pairwise contributions. These corrections, while important for precision, do not affect the qualitative structure of the thermodynamic deficit theorem: the first law still fails as $\Delta \propto S_{EE} \cdot A$ whether S_{EE} is computed from P alone or includes multi-body terms.

E. The rewritten deficit

With the identification (19), the thermodynamic deficit theorem becomes:

$$\Delta_{FL}(\Sigma) \propto S_{EE}(\Sigma) \cdot \langle \mathcal{A} \rangle_\Sigma. \tag{20}$$

The first law fails in proportion to the entanglement entropy across the surface, weighted by the entanglement anisotropy. This is the central result of this paper.

V. DERIVATION OF THE RYU-TAKAYANAGI FORMULA

A. Minimal P surfaces

Given a boundary region A in a holographic setting, consider all bulk surfaces Σ homologous to A (i.e., sharing the boundary ∂A and continuously deformable to A). The surface that correctly computes the entanglement entropy of A is the one that minimizes the cross-boundary correlation:

$$S_{EE}(A) = c_1 \cdot \min_{\Sigma \sim A} P(\Sigma). \tag{21}$$

The minimization follows from strong subadditivity of entanglement entropy: the surface that captures the actual entanglement content is the minimum cut through the correlation network. Any other surface overcounts by including correlations internal to one side.

B. Area law in the vacuum

For the vacuum state of a conformal field theory, the correlation function falls off as a power law:

$$\lambda(x, y) \sim \frac{\Lambda_0}{|x-y|^{2\Delta}}, \tag{22}$$

where Δ is the scaling dimension. The cross-boundary correlation integral is dominated by short-distance pairs near the surface, giving

$$P(\Sigma) \sim \text{Area}(\Sigma) \cdot \Lambda_{UV}^{d-2}, \tag{23}$$

where Λ_UV is the UV cutoff. In the vacuum, the correlation density is uniform, so minimizing P(Σ) is equivalent to minimizing Area(Σ):

$$S_{EE}(A) = \frac{1}{4G_N} \min_{\Sigma \sim A} \text{Area}(\Sigma). \tag{24}$$

This is the Ryu-Takayanagi formula [8], with the identification $c_1\Lambda_0$Λ_UV^{d−2} → 1/(4G_N).

C. Non-vacuum generalization

In non-vacuum states, the correlation density is non-uniform. The cross-boundary correlation integral receives a bulk contribution from correlations between features deep inside I and deep inside O that happen to thread through Σ. The correct expression is:

$$S_{EE}(A) = \min_{\Sigma \sim A} \left[\frac{\text{Area}(\Sigma)}{4G_N} + S_{bulk}(\Sigma) \right], \tag{25}$$

where S_bulk captures the non-geometric contribution from bulk entanglement. This is the quantum extremal surface prescription of Engelhardt and Wall [19], derived here from the structure of P(Σ) in a non-uniform correlation field.

The derivation makes clear why the quantum correction S_bulk appears: it represents correlations whose spatial distribution does not follow the vacuum area law but instead reflects the specific entanglement pattern of the excited state.

D. The RT surface as approximate correlation boundary

The Ryu-Takayanagi surface γ_A is the surface that minimizes P, not the surface where P vanishes. It is the best approximation to a correlation boundary for the given boundary region:

$$\gamma_A = \arg\min_{\Sigma\sim A} P(\Sigma). \tag{26}$$

This perspective places the RT surface in a hierarchy between true correlation boundaries (horizons, where P = 0) and ordinary surfaces (where P ~ O(1)):

Surface type	**P(Σ)**	**Thermodynamic?**
Horizon	0	Exactly
RT surface (symmetric region)	min, A ≈ 0	Approximately
RT surface (generic region)	min, A ≠ 0	With corrections
Stretched horizon	ε ≈ 0	Approximately
Ordinary (spherical)	O(1), A = 0	Degenerate
Ordinary (generic)	O(1), A ≠ 0	No

A new prediction follows: the thermodynamic properties of RT surfaces depend on the symmetry of the boundary region, not just the bulk geometry. Highly symmetric boundary regions (e.g., half-spaces) produce RT surfaces with small anisotropy and good thermodynamic behavior. Irregular boundary regions produce RT surfaces with larger anisotropy and greater thermodynamic deficit.

VI. RESOLUTION OF WANG-BRAUNSTEIN

A. Summary of their result

Wang and Braunstein [7] consider static asymptotically flat spacetimes with Killing vector ξ. For a surface Σ of constant Newtonian potential φ, they test whether perturbations satisfy the first law in the form

$$\delta M = \frac{\kappa}{8\pi}\delta A + \text{matter terms}, \tag{27}$$

where κ is the surface gravity and A the area. They find:

(a) On horizons, the first law holds exactly.

(b) On stretched horizons, it holds to excellent approximation.

(c) On ordinary surfaces, additional terms appear that depend on metric perturbations k_1, which cannot be eliminated by any choice of surface perturbation.

(d) Exception: for spherically symmetric configurations, the first law holds on any surface of constant φ.

B. Resolution from the correlation boundary formalism

Each of these results follows from Theorem 1 (Eq. 12):

Result (a): Horizons. A horizon is a correlation boundary: P = 0. By Eq. (12), Δ_FL = 0 regardless of A. The first law holds exactly.

Result (b): Stretched horizons. A stretched horizon has P(Σ) = ε ≪ 1. The deficit is Δ_FL ~ ε · |A| · h, which is small. The first law holds approximately, with corrections of order ε.

Result (c): Ordinary surfaces. For generic surfaces, P ~ O(1) and A ~ O(1). The deficit Δ_FL is order unity. The first law fails. The "unwanted terms" in Wang-Braunstein's Eq. (3) correspond to the δΦ_cross contribution from cross-boundary correlations responding to the perturbation.

Result (d): Spherical symmetry. This is the most interesting case and requires detailed analysis.

C. The spherical symmetry degeneracy

For a spherically symmetric mass distribution, every surface of constant φ (i.e., every concentric sphere) has the following property: the correlations threading through it are isotropic.

At any point x on the sphere, the correlations from x to outside features are distributed uniformly over solid angle (by spherical symmetry). The correlation current J has no preferred angular direction; it is purely radial. Therefore the anisotropy A(x) — the directional asymmetry of correlations — decomposes into spherical harmonics with:

$$\mathcal{A}_{lm} = 0 \quad \text{for all } l \geq 1. \tag{28}$$

The l = 0 (monopole) component may be nonzero, but it has the same angular structure as the standard thermodynamic terms δQ and TdS. It can therefore be absorbed into redefined thermodynamic quantities T' and S' (see Supporting Information Sec. S6 for the complete absorption argument):

$$\delta Q = T'\, dS', \tag{29}$$

where T' and S' incorporate the monopole contribution from δΦ_cross. The first law holds with renormalized thermodynamic quantities.

D. Breaking spherical symmetry: multipole corrections

For a nearly spherical mass distribution with multipole moments Q_l, the anisotropy acquires nonzero l ⩾ 1 components:

$$\mathcal{A}_{lm} \propto Q_l \cdot r^{-(l+1)}. \tag{30}$$

These components have angular dependence that cannot be absorbed into scalar thermodynamic quantities. The thermodynamic deficit becomes:

$$\langle |\Delta_{FL}|^2 \rangle_\Sigma \propto \sum_{l \geq 2} \frac{Q_l^2}{r^{2(l+1)}} \cdot P(r), \tag{31}$$

where the l = 1 (dipole) term vanishes for surfaces centered on the center of mass.

The dominant correction is the quadrupole (l = 2):

$$\langle |\Delta_{FL}|^2 \rangle \sim \frac{Q_2^2}{r^6} \cdot P(r). \tag{32}$$

This is a testable prediction: the thermodynamic deficit on holographic screens around oblate or prolate mass distributions should scale with the quadrupole moment squared (see Supporting Information Sec. S7 for the detailed multipole analysis).

E. Why this resolves the tension

Wang and Braunstein's result was interpreted as evidence against emergent gravity. The correlation boundary formalism shows that it is instead evidence for a more refined version of emergent gravity. The original programs of Jacobson [4] and Verlinde [5] both involve thermodynamic properties of surfaces, but they differ in which surfaces:

Jacobson applies δQ = TdS at local Rindler horizons — surfaces that are correlation boundaries by construction. His derivation is unaffected by Wang-Braunstein.

Verlinde extends thermodynamic properties to arbitrary holographic screens — surfaces that are not correlation boundaries. Wang-Braunstein correctly identifies that this extension fails.

The resolution is that gravitational thermodynamics is a property of correlation boundaries specifically, not of surfaces in general. The thermodynamic structure exists everywhere in the correlation field (as a geometric relation), but the thermodynamic interpretation — temperature, entropy, first law — applies only at surfaces where the partition into inside and outside is informationally clean (P = 0) or degenerate by symmetry (A = 0).

VII. DISCUSSION

A. Relation to Jacobson's entanglement equilibrium

Jacobson [20] argued that the Einstein equations follow from the condition that entanglement entropy across local Rindler horizons is in equilibrium. In our language, this is the statement that cross-boundary correlation P is stationary at correlation boundaries:

$$\delta P(\Sigma_H) = 0 \quad \text{for all perturbations of the horizon } \Sigma_H. \tag{33}$$

Since P = 0 at the horizon, stationarity means that P remains zero under small perturbations — the horizon is a stable correlation boundary. This condition, imposed at every point, constrains the geometry and reproduces the Einstein equations.

The connection between P = S_EE and Jacobson's entanglement equilibrium makes this precise: the equilibrium condition (33) is the entanglement first law [21], applied at the special surface where entanglement across the partition vanishes.

B. Implications for the emergent gravity program

The results of this paper suggest that the emergent gravity program can be salvaged by restricting thermodynamic assumptions to correlation boundaries:

Strong version (Jacobson). Apply δQ = TdS only at local Rindler horizons. This remains valid and leads to the Einstein equations.

Modified Verlinde. The geometric structure underlying Verlinde's derivation — the relationship between energy, entropy, and surface properties — exists everywhere. But the thermodynamic interpretation should be restricted to surfaces with high thermodynamic index (Sec. III). This preserves Verlinde's mathematics where it works while correctly predicting where it fails.

Verlinde's dark matter proposal. The cosmological horizon is a true correlation boundary (it defines the limit of causal correlation in de Sitter space). The thermodynamic properties Verlinde assigns to it [6] are legitimate. Whether the quantitative predictions for galaxy rotation curves are correct is a separate question, but the cosmological horizon is not affected by Wang-Braunstein's objection.

C. The non-thermodynamic contribution to gravity

The cross-boundary correlation response δΦ_cross represents a contribution to gravitational dynamics that does not have a thermodynamic interpretation. It is the energy associated with correlations threading through surfaces, responding to perturbations through channels that bypass the surface flux.

For an observer who assumes all gravitational dynamics are thermodynamic (i.e., who assumes $\delta Q = TdS$ holds on all surfaces), this non-thermodynamic contribution would appear as unaccounted energy — dynamics that cannot be attributed to visible matter through standard gravitational accounting. This suggests a natural research direction connecting $\delta\Phi\_cross$ to modified gravity phenomenology, though developing this connection quantitatively requires further work.

D. Computational testability

The thermodynamic deficit theorem makes predictions that can be tested in computational systems:

(i) In tensor networks, define surfaces and compute $P(\Sigma)$ and $A(x)$. Verify that thermodynamic behavior (well-defined temperature, first-law compliance) correlates with low P and low A.

(ii) In lattice gauge theories, compute the anisotropy of entanglement entropy distribution across surfaces of different geometry. The deficit should scale as $S\_EE \cdot A$.

(iii) In holographic models, test the prediction that RT surface thermodynamic properties depend on boundary region symmetry.

E. Open questions

Several directions for future work emerge:

(i) Quantitative mapping. The derivation establishes $\Delta \propto P \cdot A$ but does not fix the proportionality constant. This requires specifying the relationship between the correlation field λ_ij and the gravitational variables (metric, curvature).

(ii) Volume-law entropy. Verlinde's 2016 proposal distinguishes area-law entropy (horizons) from volume-law entropy (cosmological de Sitter space). The correlation boundary formalism should accommodate this distinction, potentially through the relationship between P and the bulk correlation density.

(iii) Finite-N corrections. At small N, the distinction between correlation boundary and ordinary surface becomes discrete rather than continuous. A companion paper [22] develops the formalism for mesoscopic N, showing that the thermodynamic deficit acquires a leading 1/N correction from the destruction of irreducible three-body correlation structure (triangular loops) that the pairwise cross-boundary correlation P does not capture. This correction provides a concrete lower bound for the applicability of the first law: bilateral thermodynamic structure requires a minimum of $N = 6$ degrees of freedom (at least three on each side of the partition), with the transition from discrete to continuum behavior passing through three qualitative thresholds at $N = 3$, $N = 6$, and $N \sim 30$.

(iv) Derivation from first principles. This paper takes the monogamy constraint (1) as given. A deeper question is whether it can be derived from more primitive principles — for instance, from conditions on what must hold for physical degrees of freedom to be distinguishable from each other. Such a derivation would ground the monogamy constraint in information-theoretic or ontological foundations rather than treating it as an empirical input.

VIII. CONCLUSION

We have introduced the concept of correlation boundaries — surfaces across which the total pairwise correlation vanishes — and shown that these are the only surfaces on which the first law of thermodynamics holds generically. The thermodynamic deficit of any surface is controlled by two quantities: the cross-boundary entanglement entropy $P \propto S_{EE}$ and the entanglement anisotropy A. The product $P \cdot A$ determines the magnitude of the deficit.

This result resolves the tension identified by Wang and Braunstein between the emergent gravity program and the non-thermodynamic behavior of holographic screens. Horizons are thermodynamic because they are correlation boundaries. Stretched horizons are approximately thermodynamic because they are near-correlation boundaries. Spherically symmetric screens are thermodynamic by a degeneracy — isotropic entanglement mimics a correlation boundary. Generic screens fail because they have nonzero entanglement that is anisotropically distributed.

The identification $P = S_{EE}$ connects this formalism to the holographic entanglement entropy program, with the Ryu-Takayanagi formula emerging as the minimization of cross-boundary correlation over surfaces homologous to a boundary region. The quantum extremal surface prescription follows from non-uniformity of the correlation field in excited states.

These results follow from a minimal set of assumptions — pairwise correlations with finite capacity — making them independent of specific microscopic theories and applicable across quantum gravity, condensed matter, and quantum information contexts.

ACKNOWLEDGMENTS

The author thanks Gabriele Carcassi and Christine Aidala for making the Assumptions of Physics project openly available, and Zhi-Wei Wang and Samuel Braunstein for the clarity of the result that motivated this work.

APPENDIX A: PROOF OF THEOREM 1

We prove that the thermodynamic deficit under a normal surface displacement is $\Delta_{FL} = \int_\Sigma P(x) \cdot A(x) \cdot h(x)\, d\mu_\Sigma + O(P^2)$.

A1. Potential decomposition. Consider the family of correlation fields parameterised by $t \in [0,1]$, with $\lambda_{ij}(t) = t \cdot \lambda_{ij}$ for cross-boundary pairs and all internal correlations held fixed. At $t = 0$, there are no cross-boundary correlations, so by additivity (P1): $\Phi_{total}(0) = \Phi_I + \Phi_O$. By smoothness (P2):

$$\Phi_{cross} = \Phi_{total}(1) - \Phi_{total}(0) = \int_0^1 \sum_{i \in I, j \in O} \varphi_{ij}(t) \cdot \lambda_{ij}\, dt, \tag{A1}$$

where $\varphi_{ij}(t) = \partial\Phi/\partial\lambda_{ij}|_t$ is the correlation susceptibility. To first order: $\Phi_{cross} = \Sigma\, \lambda_{ij} \cdot \varphi^0_{ij} + O(\lambda^2)$, where $\varphi^0_{ij} = \varphi_{ij}(0)$.

A2. Surface displacement. Displacing Σ by $\varepsilon \cdot \delta\varepsilon(x)$ in the outward normal reassigns features in a shell of volume $dV = \varepsilon \cdot \delta\varepsilon(x) \cdot d\mu_\Sigma(x)$, containing $n(x) \cdot dV$ features. Each reassigned feature α at position x contributes a change:

$$\delta\Phi^{(\alpha)}_{cross} = \sum_{j \in O'} \lambda_{\alpha j} \cdot \varphi^0_{\alpha j} - \sum_{i \in I} \lambda_{\alpha i} \cdot \varphi^0_{\alpha i}. \tag{A2}$$

A subleading geometric effect (the change in susceptibilities φ_{ij} due to the altered partition) contributes at $O(\varepsilon/N)$ relative to the reassignment effect (A2), because it involves $O(N^2)$ pairs each shifting by $O(\varepsilon/N^2)$. We retain only the dominant reassignment effect.

A3. The P · A factorisation. Under the uniform susceptibility condition $\varphi^0_{ij} \approx \bar{\varphi}$ (which holds when individual correlations are weak, i.e., $\lambda_{ij} \ll \Lambda$), the susceptibility factors out:

$$\delta\Phi_{cross} = \varepsilon\bar{\varphi} \int_\Sigma n(x)\, \delta\varepsilon(x) \cdot \left[C^{out}(x) - C^{in}(x)\right] d\mu_\Sigma + O(\varepsilon^2, 1/N), \tag{A3}$$

where $C^{out}(x) = \Sigma_{j \in O}\, \lambda_{xj}$ and $C^{in}(x) = \Sigma_{i \in I}\, \lambda_{xi}$. The factor $C^{out} - C^{in}$ is the correlation anisotropy $A(x)$. It is proportional to $P(x)$ (since $A(x) = 0$ when $P(x) = 0$) and vanishes independently when $C^{out} = C^{in}$ (isotropy). Defining $h(x) = \varepsilon \cdot n(x) \cdot \delta\varepsilon(x) \cdot \bar{\varphi}$:

$$\Delta_{FL}(\Sigma) = \int_\Sigma P(x) \cdot \hat{A}(x) \cdot h(x)\, d\mu_\Sigma + O(\varepsilon^2, 1/N). \tag{A4}$$

A4. Error bounds. Three approximations are used: (i) First-order expansion of Φ_{cross} — error $O(\Lambda^2/N)$ per pair, total $O(\Lambda^2)$ for a balanced partition, controlled when $\lambda_{ij} \ll \Lambda$ (large N). (ii) Local partner classification (partners classified as inside/outside by the local surface normal) — valid when the correlation length $\xi \ll R_\Sigma$ (surface curvature radius). (iii) Neglect of geometric effect — $O(\varepsilon/N)$ relative to leading term.

Combined validity: $N \gg 1$, $\xi \ll R_\Sigma$, $\varepsilon \ll R_\Sigma$. This is the standard thermodynamic regime.

A5. Special cases. (a) Correlation boundary ($P = 0$): $\Phi_{cross} = 0$ exactly by (P1), so $\Delta = 0$ with no approximation required. (b) Approximate boundary ($P \approx \varepsilon_P$): $\Delta = O(\varepsilon \cdot \varepsilon_P)$. (c) Isotropic correlations ($A = 0$): the deficit vanishes at leading order but may acquire $O(\varepsilon^2)$ corrections from the second-order term in (A1). The isotropic case is a degeneracy, not a fundamental property — distinguishing it from the exact result at correlation boundaries. ■

REFERENCES

[1] J. M. Bardeen, B. Carter, and S. W. Hawking, "The four laws of black hole mechanics," Commun. Math. Phys. **31**, 161 (1973).

[2] J. D. Bekenstein, "Black holes and entropy," Phys. Rev. D **7**, 2333 (1973).

[3] S. W. Hawking, "Particle creation by black holes," Commun. Math. Phys. **43**, 199 (1975).

[4] T. Jacobson, "Thermodynamics of Spacetime: The Einstein Equation of State," Phys. Rev. Lett. **75**, 1260 (1995).

[5] E. Verlinde, "On the Origin of Gravity and the Laws of Newton," JHEP **2011**, 29 (2011). arXiv:1001.0785.

[6] E. Verlinde, "Emergent Gravity and the Dark Universe," SciPost Phys. **2**, 016 (2017). arXiv:1611.02269.

[7] Z.-W. Wang and S. L. Braunstein, "Surfaces away from horizons are not thermodynamic," Nat. Commun. **9**, 2977 (2018).

[8] S. Ryu and T. Takayanagi, "Holographic Derivation of Entanglement Entropy from AdS/CFT," Phys. Rev. Lett. **96**, 181602 (2006).

[9] V. Coffman, J. Kundu, and W. K. Wootters, "Distributed entanglement," Phys. Rev. A **61**, 052306 (2000).

[10] T. J. Osborne and F. Verstraete, "General Monogamy Inequality for Bipartite Qubit Entanglement," Phys. Rev. Lett. **96**, 220503 (2006).

[11] G. Vidal, "Entanglement Renormalization," Phys. Rev. Lett. **99**, 220405 (2007).

[12] B. M. Terhal, "Is entanglement monogamous?," IBM J. Res. Dev. **48**, 71 (2004).

[13] A. Lewkowycz and J. Maldacena, "Generalized gravitational entropy," JHEP **2013**, 90 (2013).

[14] M. Van Raamsdonk, "Building up spacetime with quantum entanglement," Gen. Rel. Grav. **42**, 2323 (2010).

[15] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," Fortschr. Phys. **61**, 781 (2013).

[16] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, "Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence," JHEP **2015**, 149 (2015).

[17] I. Peschel, "Calculation of reduced density matrices from correlation functions," J. Phys. A **36**, L205 (2003).

[18] T. Faulkner, A. Lewkowycz, and J. Maldacena, "Quantum corrections to holographic entanglement entropy," JHEP **2013**, 74 (2013).

[19] N. Engelhardt and A. C. Wall, "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime," JHEP **2015**, 73 (2015).

[20] T. Jacobson, "Entanglement Equilibrium and the Einstein Equation," Phys. Rev. Lett. **116**, 201101 (2016).

[21] D. D. Blanco, H. Casini, L.-Y. Hung, and R. C. Myers, "Relative entropy and holography," JHEP **2013**, 60 (2013).

[22] D. Neale, "Distinguishability Boundaries and the Emergence of Thermodynamic Structure." Companion paper. Preprint available at goleudy.ai.