

Capacity, Dissipation, and the Onset of Motility-Induced Phase Separation: An Empirical Test of RSVP Field Dynamics

Flyxion
Independent Researcher

July 2026

Abstract

A construct earns empirical standing when it can be reconstructed from data, compared against genuine alternatives, shown stable across nuisance parameters, and shown to improve prediction beyond an established baseline. This paper applies that standard to a capacity–transport–dissipation decomposition inspired by the RSVP framework, tested against motility-induced phase separation (MIPS) in active Brownian particle systems—a system chosen because strong conventional baselines already exist, so a new decomposition has no explanatory gap to fill by default. Three competing capacity operationalizations and two dissipation estimators were reconstructed from simulated trajectories and evaluated, via regression and classification with held-out and permutation-based testing, for predictive value beyond conventional density observables. Across replications spanning ten runs at a primary parameter point, five at a higher activity level, and twenty across three packing fractions, no capacity or dissipation feature showed incremental predictive value beyond density that survived replication. A separate causal analysis was then built: a counterfactual intervention framework using common-random-numbers paired simulation with a localized activity perturbation, validated and applied in an initial four-region-pair pilot testing whether pre-intervention capacity predicts the magnitude of a region’s response to perturbation. The pilot is explicitly underpowered and settles nothing either way. The paper’s conclusion is correspondingly narrow: this operationalization of capacity is observationally redundant with density in the regimes tested, and whether it carries causal content beyond density remains open, pending a statistically powered matched-region study whose design is specified here.

1 Introduction

The companion essay *Prediction Before Interpretation* [13] argues that theoretical field constructs earn empirical standing only after reconstruction, comparison against alternatives, stability analysis, and successful prediction beyond established baselines. This paper evaluates one proposed capacity–transport–dissipation decomposition against those criteria.

The test case is a capacity–transport–dissipation decomposition of active Brownian particle (ABP) dynamics, evaluated against motility-induced phase separation (MIPS). MIPS was chosen for being a hard test rather than a hospitable one: existing kinetic and hydrodynamic theories already give a reasonably complete account of the phenomenon without any capacity-like vocabulary, so a new decomposition has no explanatory vacuum to fill by default. If it adds nothing here, that is informative. If it adds something, the addition is real.

The paper reports two lines of evidence, kept deliberately separate because they answer different questions. The first is observational: do capacity and dissipation fields, reconstructed from simulated trajectories, carry information about the near-future evolution of density-domain structure that conventional density observables do not already carry? The second is causal: does a region’s pre-intervention capacity reading predict how strongly that region responds to an identical, experimentally administered perturbation? Observational prediction, even if strong, could only justify describing a construct as an early-warning variable. Only a surviving intervention result would justify describing it as a driver.

2 Methods

2.1 Simulation engine and physical validation

The reference system is a two-dimensional overdamped active Brownian particle model [1, 2], of the general class reviewed in [5, 4]: N self-propelled disks of diameter σ , interacting via a Weeks–Chandler–Andersen (WCA) repulsive potential, integrated with Euler–Maruyama (appropriate for this first-order overdamped system; Verlet integration, appropriate for inertial dynamics, was deliberately not used). Motility-induced phase separation, the phenomenon of interest, is reviewed in [3]. The simulation engine was built to be theory-neutral: it knows nothing about capacity or dissipation and emits only particle trajectories and forces, so that RSVP-style quantities are reconstructed afterward rather than baked into the data-generating process.

Before any field reconstruction, the engine was validated against six standard ABP/MIPS benchmarks (Figures 1–2): orientation autocorrelation against the analytic prediction $\langle u(0) \cdot u(t) \rangle = e^{-D_r t}$, mean-squared displacement showing the expected ballistic-to-diffusive crossover, the radial distribution function’s expected exclusion zone and contact peak, and a coarse phase scan over packing fraction ϕ and Péclet number showing the expected qualitative trend of increasing clustering with density. An ideal-gas sanity check ($v_0 = 0$, $\epsilon = 0$, expecting $g(r) \approx 1$ and $S(q) \approx 1$) also passed.

A recurring numerical finding, encountered and re-confirmed several times across the project, is that explicit Euler–Maruyama integration of the WCA potential is only condi-

tionally stable: the safe timestep depends on both the propulsion speed v_0 and the thermal diffusion strength D_t , not only on ϵ , σ , and mobility, and instabilities can appear only after an extended stable-looking run rather than immediately (full numerical detail in Appendix A). A permanent stability diagnostic was added that raises an explicit error above a force-magnitude threshold, so that any future run at an unvalidated parameter combination terminates explicitly rather than producing potentially invalid output.

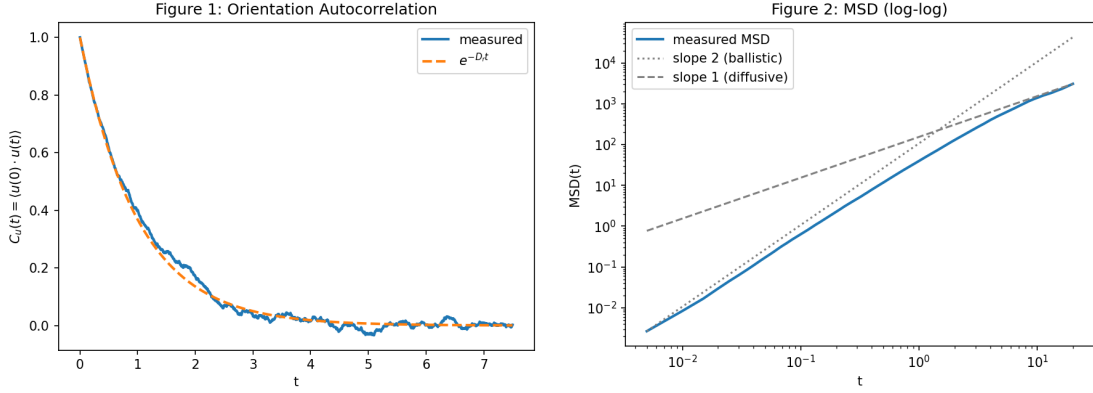


Figure 1: Left: measured orientation autocorrelation against the analytic prediction $e^{-D_r t}$. Right: mean-squared displacement on log-log axes, showing the expected ballistic (slope 2) to diffusive (slope 1) crossover.

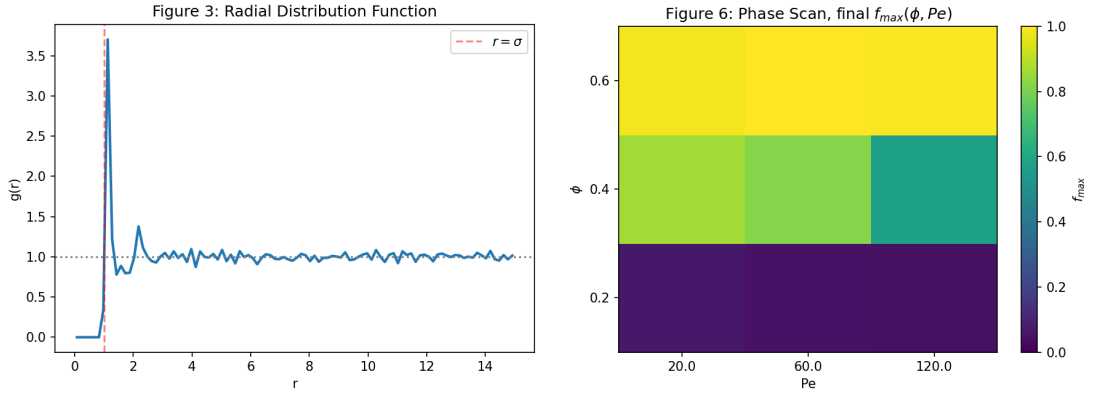


Figure 2: Left: radial distribution function showing the expected exclusion zone below σ and contact peak just above it. Right: an early coarse phase scan over packing fraction and Péclet number; this specific figure was later flagged provisional once a measurement artifact affecting the underlying order parameter was discovered (Section 2.3).

2.2 Field reconstruction

Density ρ_h , current J_h , and velocity v_h were reconstructed from particle trajectories via a normalized periodic Gaussian kernel [8], purely kinematically—this stage makes no reference to capacity or dissipation. Three competing operationalizations of capacity were then defined and reconstructed side by side, rather than one being chosen in advance: Φ_{input} (the imposed propulsion speed, trivially uniform for this constant-speed model and retained as a null control); $\Phi_{\parallel}$ (signed realized alignment between propulsion direction and actual velocity, $\langle u_i \cdot v_i \rangle_h$, which can go negative when collisions push a particle backward); and Φ_+ (the non-negative part of the same quantity). Each is available in a total-velocity and a deterministic-velocity-only variant, since the total-velocity variant was found to be visibly contaminated by thermal noise relative to the deterministic variant on a real clustering system; the deterministic variant was adopted as the default for downstream identification work.

Two dissipation estimators were similarly defined, drawing on the stochastic-thermodynamics framework of [6, 7]: $\dot{S}_{\text{work}}$, a direct force-times-velocity estimator, and $\dot{S}_{\text{current}}$, built from the already-reconstructed density and current fields via the standard Fokker–Planck local entropy-production formula rather than from any per-particle force. It was proved analytically, and confirmed numerically to machine precision, that $\dot{S}_{\text{work}} = (v_0/D_t) \Phi_{\parallel}$ exactly for this model—the two are collinear, not independent variables, which matters for any regression that might otherwise treat their joint selection as two pieces of evidence rather than one. $\dot{S}_{\text{current}}$ was confirmed to be a genuinely independent estimator, with correlation on the order of 0.1 against $\dot{S}_{\text{work}}$ on a real clustering system and a visually distinct spatial pattern (Figure 3).

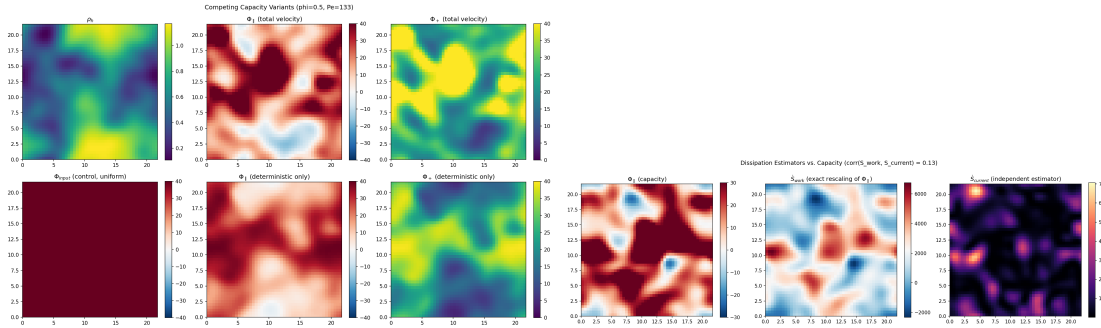


Figure 3: Left: the three capacity variants and their total-velocity/deterministic-velocity forms, on a dense clustering system—the deterministic variants are visibly smoother, consistent with the total-velocity variant carrying thermal-noise contamination. Right: $\dot{S}_{\text{work}}$ and $\Phi_{\parallel}$ share an identical spatial pattern up to rescaling (the proved collinearity), while $\dot{S}_{\text{current}}$ shows a visually distinct, more localized pattern.

2.3 Target validity and a measurement artifact

Before building any identification pipeline, a narrow validity scan was run specifically to check whether the natural candidate order parameter—particle-contact cluster fraction

$f_{\max}$, computed with a fixed contact-distance cutoff—was actually a valid measure of phase separation across the packing fractions to be studied. It was not. At moderate-to-high packing fraction ($\phi = 0.5$, $N = 300$), $f_{\max}$ read exactly 1.000 at $t = 0$, before any dynamics had occurred, for both grid-based and random-sequential-adsorption initial configurations. The cause is geometric rather than dynamical: at sufficiently high packing fraction, the typical inter-particle spacing implied by packing alone falls below the fixed contact cutoff, so any configuration—phase-separated or perfectly homogeneous—trivially satisfies the criterion. This was confirmed independently: density bimodality, which should track genuine dense/dilute coexistence, showed near-zero or negative correlation with $f_{\max}$ in this regime (Figure 4).

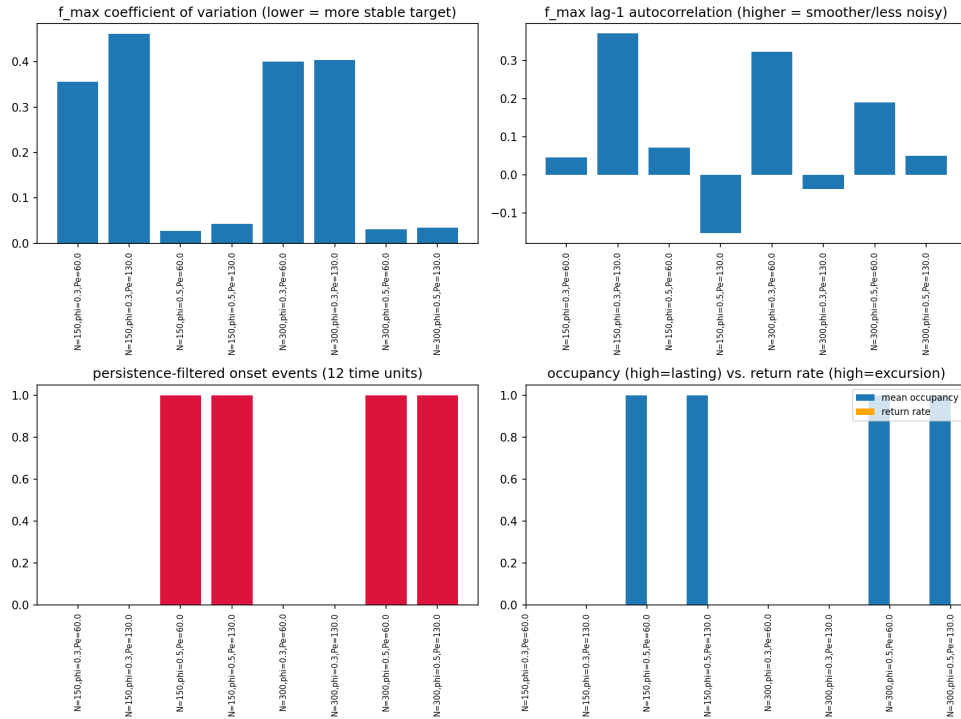


Figure 4: The scan that surfaced the $f_{\max}$ artifact: coefficient of variation, lag-1 autocorrelation, persistence-filtered onset event counts, and occupancy/return-rate diagnostics across a small grid of packing fraction and Péclet number, run before any exporter or identification infrastructure was built.

The fix was to replace particle-contact cluster fraction with a composite of density-domain indicators built from the coarse-grained density field: $f_{\text{dense,max}}$ (the largest connected region of ρ_h exceeding a threshold defined relative to that snapshot’s own mean density, via a periodic-boundary connected-components procedure), f_{void} (the analogous dilute-region fraction), the low- q structure factor of the density field, and the density bimodality coefficient. Because the dense/void thresholds are defined relative to each snapshot’s own mean density rather than an absolute level, high packing fraction alone cannot trigger a false “phase separated” reading; this was verified directly by confirming $f_{\text{dense,max}} =$

0.000 at the exact configuration where the old metric had read 1.000. Particle-contact cluster fraction was retained afterward only as a diagnostic, never again as the primary target. One further consequence: high packing fraction ($\phi > 0.5$) could not be included in the parameter sweeps reported below, because random-sequential-adsorption initialization—adopted as the default preparation protocol precisely because it avoids the lattice-regularity artifacts of grid initialization—cannot itself reach packing fractions near the two-dimensional hard-disk jamming limit of approximately 0.55.

2.4 Identification pipeline

The identification target is the vector $M(t) = (f_{\text{dense,max}}, f_{\text{void}}, S_\rho(q_{\text{low}}), B_\rho)(t)$, with two representations used for evaluation: $\Delta_\tau M(t) = M(t+\tau) - M(t)$ for regression, and $Y_\tau(t) = \mathbb{K}\{\text{onset starts within } (t, t+\tau]\}$ for classification, where onset is defined by a persistence criterion (sustained crossing of a high threshold following an extended dwell above a lower threshold) rather than a raw threshold crossing, since raw density-domain order parameters were found to be highly volatile at the system sizes studied. The candidate feature library Θ consists of spatially-aggregated scalar functionals of each reconstructed field—mean, variance, skewness, upper quantile, gradient-squared and Laplacian-squared means, low- q structure factor, and a Fourier-based correlation length, plus cross-field covariance, mean-product, and gradient-alignment terms—rather than pointwise spatial-PDE terms; the target is prospective time-series identification over field functionals, not spatial PDE discovery, and a genuinely spatial PDE for capacity’s own dynamics is treated as a separate, shelved auxiliary project. Three competing capacity variants were evaluated as separate branches ($\Phi_{\text{input}}, \Phi_{\parallel}, \Phi_{+}$) rather than pooled into one regression, with nested model comparison at each: M_0 (density-only baseline), M_1 (M_0 plus that branch’s capacity features), M_2 (M_1 plus dissipation and branch-filtered interaction terms).

All simulated trajectories, reconstructed fields, extracted features, targets, and onset labels were written to a canonical per-run data file recording full preparation and reconstruction provenance, with an explicit completion flag set only after a run finished without numerical failure (schema and provenance fields detailed in Appendix B). This proved practically important: several long export runs were interrupted mid-write by execution-time limits during this project, and in every case the resulting file was either correctly unreadable or correctly flagged incomplete, rather than entering the analysis as valid data.

Regression evaluation used LassoCV [9] with a chronological (not random) train/test split [10], since random splitting would let temporally autocorrelated future frames leak into the training set; the inner cross-validation used to select the regularization strength likewise used a time-series-respecting split. Classification evaluation pooled leave-one-run-out held-out predictions before computing ROC-AUC, PR-AUC, and Brier score, since per-run AUC is unstable with only a handful of positive onset windows per run. Significance testing throughout the observational program used run-level permutation tests [11] rather than parametric assumptions, given the small and unevenly sized replicate sets involved.

2.5 Intervention protocol

A counterfactual intervention framework was built to test causal rather than purely observational claims. A localized activity pulse—elevated propulsion speed for particles within a specified radius of a specified center, during a specified time window—is applied to a treatment branch cloned from a common starting state; a control branch, cloned from the same state, receives no perturbation. Both branches are advanced with independently instantiated random-number generators seeded identically, so that at zero perturbation amplitude the two branches are guaranteed bit-identical, and any divergence at nonzero amplitude is attributable to the perturbation rather than to independent sampling noise (common random numbers, the standard variance-reduction design for causal comparison in stochastic simulation [12]). A matched-region extension characterizes candidate regions of a real simulated system by local density, local capacity, local polar order, dense-domain membership, and local density-gradient magnitude, then pairs regions matched on density but contrasting in capacity, applying the identical pulse to each member of a pair.

3 Results

3.1 Observational identification: no incremental value beyond density

The primary regression result, replicated across ten independent runs at one parameter point ($N = 200$, $\phi = 0.5$, Péclet = 130), found that a density-only baseline model achieved $R^2 = 0.348$ for predicting $\Delta_\tau f_{\text{dense,max}}$, and that adding any capacity variant, with or without dissipation and interaction terms, was statistically indistinguishable from the baseline (range 0.348–0.349). Sparse selection under cross-validated regularization retained almost exclusively density features.

Because a coefficient hitting exactly zero under L_1 regularization is a weaker claim than a directly tested null, two further analyses were run. A residualized correlation test regressed each capacity and dissipation feature against the density baseline via ridge regression, then tested whether the residual still correlated with the target, using an exact (or, at larger replicate counts, sampled) permutation test at the level of entire runs rather than individual frames—frames within a run are strongly autocorrelated, so the effective independent sample size is the number of runs, not the number of frames. A second test block-permuted which run’s capacity features were paired with which run’s density baseline and target, refitting leave-one-run-out and comparing the true ΔR^2 to the permutation-derived null distribution. At the primary parameter point and replicate count, all thirty-two tested feature columns showed $p > 0.05$ under residualized-correlation testing, and three of four block-permutation tests were clean nulls; one (the combined capacity-plus-dissipation-plus-interactions model) reached nominal significance ($p = 0.017$) but did not survive correction for the four tests run, and, critically, did not replicate when the replicate count was doubled to ten runs ($p = 0.990$ at $n = 10$)—direct evidence that the marginal result at $n = 5$ was a small-sample artifact rather than a real effect.

The null was then tested for stability across four further axes, since the essay’s own criterion requires stability across nuisance parameters as a precondition for any claim, positive or negative (Table 1). It held cleanly across coarse-graining and replicate count—

strengthening, in fact, at the higher replicate count, since the one borderline permutation result at five runs vanished at ten ($p = 0.990$)—and across the higher activity level. Packing fraction produced the one genuinely mixed result: the block-permutation test remained completely flat at $\phi = 0.35$, but the residualized correlation test flagged six of thirty-two columns, with an internally consistent pattern across the mean and skewness of $\Phi_{\parallel}$, Φ_{+} , and $\dot{S}_{\text{work}}$ (the latter matching the former exactly, as required by the proved collinearity—itsself a useful internal consistency check). This was investigated rather than left unresolved: neither neighboring packing fraction reproduced the same signature, arguing that the $\phi = 0.35$ pattern is an isolated, non-systematic anomaly rather than genuine density-dependent information content, though this is not certain given the asymmetric replicate counts involved.

Table 1: Stability of the observational null across four axes. “Residualized” reports the fraction of 32 tested feature columns significant at $p < 0.05$; “Permutation” reports the block-permutation ΔR^2 test outcome across the four branch/model combinations tested.

Axis	Setting	Replicates	Residualized	Permutation
Primary regime	$\phi = 0.5$, $\text{Pe} = 130$, $h = 1.2$	5	1/32	3/4 null, 1 marginal ($p = 0.017$)
Replicate count	(as above)	10	1/32	4/4 null (marginal result vanished)
Coarse-graining	$h = 0.8$	3	0/8	1/1 null
Coarse-graining	$h = 1.8$	2	0/8	1/1 null (low power)
Activity level	$\text{Pe} = 200$	5	0/32	4/4 null
Packing fraction	$\phi = 0.35$	10	6/32 (mixed)	4/4 null
Packing fraction	$\phi = 0.2$	5	no replication of $\phi = 0.35$ pattern	
Packing fraction	$\phi = 0.45$	5	no replication of $\phi = 0.35$ pattern	

Classification, revisited with the full ten-run primary-regime dataset after an earlier five-run pilot had been explicitly inconclusive, converged on the same qualitative story: pooled leave-one-run-out ROC-AUC was 0.53–0.56 across every branch and model level—barely above chance, but real, since PR-AUC (≈ 0.32) sat modestly above the base positive rate (0.28)—and the density-only baseline matched or slightly exceeded every model that added capacity or dissipation features.

3.2 Intervention validation and a caught measurement error

Before any causal claim could be evaluated, the intervention mechanism itself required validation. The zero-amplitude identity property—that treatment and control branches must be bit-identical when no perturbation is applied—was confirmed directly rather than assumed. A substantive implementation error was caught during this process: an early version of the intervention timing logic evaluated the perturbation window against the wrong time reference, with the consequence that the intervention did not activate under realistic post-equilibration starting conditions (full account in Appendix C). The error was caught only by directly verifying, on a demonstration run, that the treatment and control trajectories had actually diverged; they had not. This is reported because it illustrates a general point about computational causal inference: an inactive intervention and a genuinely null causal effect are numerically indistinguishable from the output alone, and only a positive-

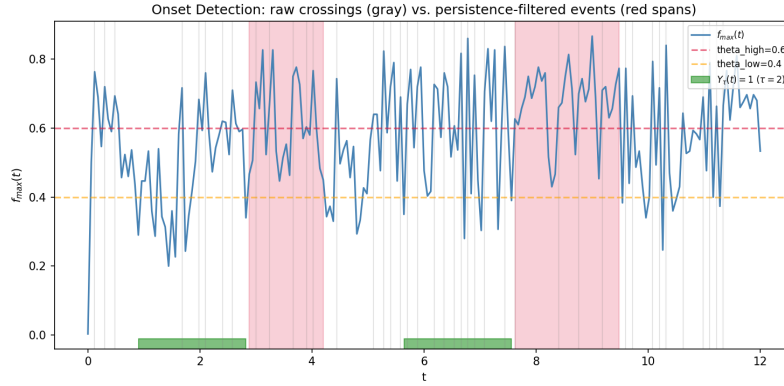


Figure 5: Persistence-filtered onset detection on a real trajectory: raw threshold crossings (gray) versus events that clear both a high threshold and a minimum-dwell-time requirement (red spans), illustrating why a raw crossing criterion was rejected as the onset definition.

control check—confirming the intervention can produce a nonzero effect at all—can distinguish them.

With the error corrected, a demonstration intervention (a threefold local activity increase, radius three particle diameters, two-time-unit duration) produced a substantial, temporally structured, physically interpretable response: local particle count within the perturbed region fell from a control-matched baseline of approximately twenty to a treatment minimum of two during the pulse (consistent with elevated local propulsion increasing persistence length and driving particles out of the region faster than they arrive), then overshoot the control trajectory afterward (twenty-eight versus sixteen, three time units after the pulse ended), with a corresponding difference in the whole-system density-domain indicator persisting to the end of the observation window (Figure 6). This establishes that the intervention variable is causally consequential and that the counterfactual framework can resolve both the immediate and delayed components of that response; it does not by itself establish anything about capacity.

3.3 Matched-region pilot: an open causal question

An initial matched-region experiment applied the validated intervention to four density-matched, capacity-contrasting region pairs drawn from a single simulated system at one random seed. The result is reported as a pilot rather than a finding, for the same reason several observational results earlier in the project turned out not to replicate at higher sample sizes: local response magnitude (peak post-pulse overshoot, integrated treatment-control difference) showed no consistent direction across the four pairs, with two pairs showing a larger response in the higher-capacity region and two showing the opposite (Figure 7). The delayed whole-system effect was, suggestively, more negative for the higher-capacity region in all four pairs, but four pairs at one seed is exactly the sample size at which earlier apparent effects in this project’s observational program (a marginal permutation result, a packing-fraction anomaly) did not survive replication; this pattern is recorded as a candi-

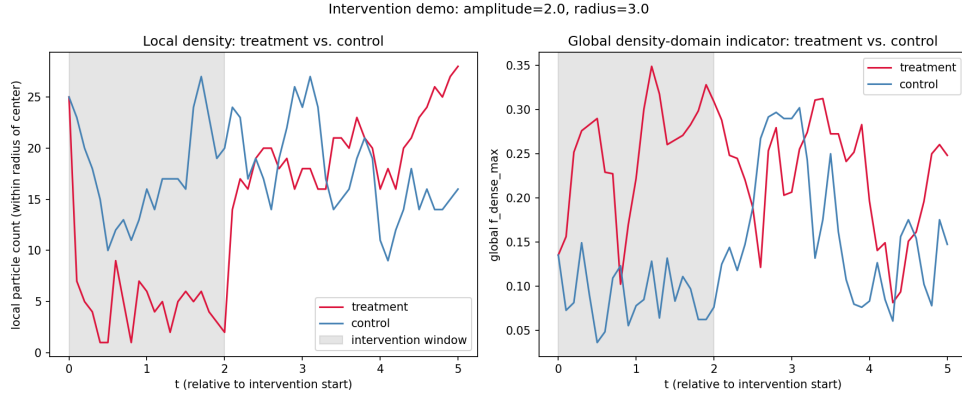


Figure 6: Local particle count and global density-domain indicator, treatment versus control, for a single demonstration intervention: an immediate depletion during the pulse, a later overshoot past the control trajectory, and a persistent difference in the global indicator.

date primary outcome for a properly powered follow-up, not as a result. A further finding from the pilot bears directly on how that follow-up must be designed: local density and local capacity were consistently anti-correlated across the candidate regions sampled, meaning that naive density-matching risks comparing capacity extremes only at the edges of the available density range, where the resulting comparison would extrapolate rather than genuinely compare. A statistically powered version accordingly requires common-support diagnostics, the region-pair (not the individual region) as the unit of analysis, a single pre-registered primary outcome, and the remaining recorded covariates (polarization, dense-domain membership, density gradient) incorporated into the matching procedure rather than only recorded alongside it.

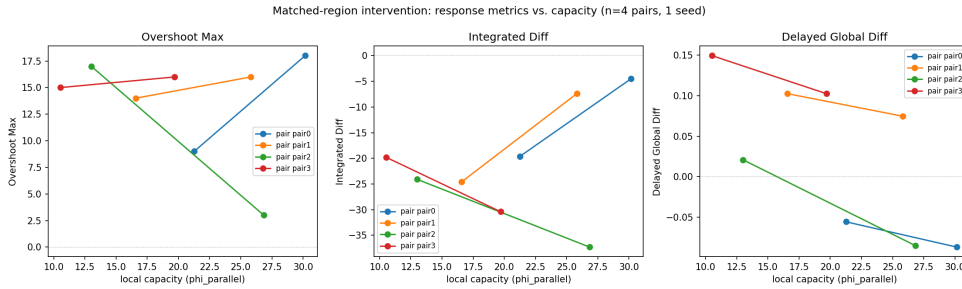


Figure 7: Response metrics against local capacity for the four pilot matched pairs. The absence of a consistent slope direction across pairs in the local-response metrics is the central honest finding of this pilot: it does not resolve the causal question.

4 Discussion

4.1 Interpretation

Read against the interpretation scheme proposed in the companion essay, the observational result is close to, but short of, the essay’s own “clean failure” row. It clears three of the criterion’s stability clauses cleanly—coarse-graining, replicate count, and activity level—and the fourth (packing fraction) after investigation of an initial anomaly that did not replicate. Within the regimes actually tested, the finding is considerably better supported than a single Lasso fit could establish on its own: capacity and dissipation features, however constructed from this model’s dynamics, behave as downstream expressions of density-driven clustering rather than as sources of independent predictive information, at least for near-term prediction of density-domain structure.

This is not evidence that the constructs are meaningless, nor that a different active-matter system or a different capacity operationalization would show the same pattern; the essay’s own logic is explicit that a negative result here is a scoped, technical finding about one attempt, not a verdict on the underlying idea. Nor does it resolve the causal question at all. A construct can be observationally redundant with density while still identifying real differences in a region’s causal susceptibility to perturbation—exactly the distinction the matched-region pilot was designed to test, and exactly what the pilot’s small sample size prevents it from deciding.

The paper’s honest current position is therefore two claims of different strength, kept deliberately apart. First, with substantial supporting evidence across four stability axes: this capacity–dissipation decomposition does not improve near-term prediction of MIPS-relevant structure beyond density, in the parameter regimes tested. Second, with essentially no supporting evidence either way: whether capacity nonetheless identifies heterogeneous causal susceptibility to intervention remains an open, well-posed, and now fully instrumented question.

More broadly, the study illustrates that a principled empirical workflow can yield informative negative results: reconstruction, competing operationalizations, robustness testing, and intervention design together narrow the space of viable theoretical claims even when no positive predictive effect is found.

4.2 Limitations

Several limitations bound how far the results above should be read.

Coverage of parameter space. The packing-fraction sweep is one-sided: high packing fraction ($\phi \gtrsim 0.5$) could not be reached because random-sequential-adsorption initialization, adopted specifically to avoid the lattice-regularity artifacts of grid initialization, cannot itself be driven near the two-dimensional hard-disk jamming limit. The observational null is therefore established for low-to-moderate packing fraction and one activity-level contrast, not for the full phase diagram.

Asymmetric statistical power. Several stability checks were run at unequal replicate counts for practical reasons—ten runs at the primary parameter point, five at the contrasting activity level, ten and five and five across three packing fractions, three and two runs at the two coarse-graining bandwidths. Where a result did not replicate across a resolution of

unequal sample size (the packing-fraction anomaly in particular), the conclusion drawn is appropriately hedged in the text rather than treated as fully decisive.

A single simulator and a single physical system. All results come from one two-dimensional overdamped active Brownian particle implementation. Nothing here speaks to whether the same null would hold in three dimensions, under different interaction potentials, or in other active-matter systems where MIPS or MIPS-like transitions occur.

Classification received a less complete stability battery than regression. The four-axis stability program (coarse-graining, replicate count, activity level, packing fraction) was carried out in full only for the regression target. Classification was re-evaluated once, on the full ten-run primary-regime dataset, and found to match the regression conclusion qualitatively, but it was not independently re-tested across the same four axes.

The causal pilot is a design validation, not a result. Four matched pairs from a single simulated system at one random seed cannot distinguish a real causal effect of capacity from sampling noise, a point emphasized throughout Section 4.3 and repeated here because it is the single most important caveat in the paper: no claim, positive or negative, about capacity’s causal role should be drawn from the pilot data reported here.

5 Future Research Directions

The results above leave a specific, non-speculative set of next steps, several of which were designed in the course of this project but not yet executed.

1. **A statistically powered matched-region study.** The matched-region pilot (Section 4.3) motivates a fully specified follow-up: the region *pair*, not the individual region, as the unit of analysis, with within-pair treatment-effect contrasts pooled across many pairs and multiple random seeds; explicit common-support diagnostics, since local density and local capacity were found to be consistently anti-correlated in the pilot sample, so that naive matching risks comparing capacity extremes only at the edges of the available density range; a single primary outcome variable chosen in advance (the delayed integrated local-density response or the delayed global density-domain effect are the leading candidates, the latter having shown a suggestive but statistically unsupported pattern in the pilot) with all other response metrics treated as secondary; and incorporation of the presently-recorded-but-unused covariates (local polar order, dense-domain membership, local density-gradient magnitude) into the matching procedure itself rather than leaving them as post-hoc descriptive variables.
2. **Extending the packing-fraction sweep to high density.** Testing whether the observational null persists near and above the jamming-adjacent regime requires an initialization protocol capable of reaching those densities, such as compression from a dilute, randomly-seeded configuration under active dynamics rather than one-shot rejection sampling. This is a bounded engineering task rather than an open design question.
3. **A matching stability battery for classification.** The same four-axis program applied to regression—coarse-graining, replicate count, activity level, and packing fraction—

has not yet been applied to the classification target. Given that classification and regression have so far told a consistent story, this is lower priority than the causal study above, but it remains a genuine gap rather than an assumption.

4. **Testing additional or alternative capacity operationalizations.** The three variants tested here (Φ_{input} , $\Phi_{\parallel}$, Φ_{+}) do not exhaust the space of reasonable capacity definitions consistent with the RSVP framework’s informal description. A negative result for these three specific reconstructions constrains, but does not by itself rule out, other operationalizations of the same underlying concept.
5. **Generalizing beyond this system.** The identification pipeline, exporter schema, and intervention framework built for this project are not specific to motility-induced phase separation; they could in principle be applied with modest adaptation to other active-matter transitions, to three-dimensional systems, or to entirely different classes of self-organizing dynamical systems where an analogous capacity–transport–dissipation decomposition has been proposed. Whether the null result reported here is a property of MIPS specifically or a more general feature of this style of decomposition is, at present, unknown.
6. **Revisiting the shelved spatial identification problem.** The present study deliberately treated the identification target as a set of spatially-aggregated scalar functionals rather than attempting to identify a genuine spatial partial differential equation governing capacity’s own evolution (the auxiliary “Option A” problem set aside at the design stage of this project). Should either the observational or the causal program eventually produce a positive result, a natural next step would be to revisit that spatial formulation directly.

6 Conclusion

The methodological essay this paper answers to argued that a construct like capacity earns empirical standing only by surviving reconstruction, comparison, stability, and prediction, in that order, and that the appropriate response to a negative result is neither concealment nor overreach but a precisely scoped statement of what failed and what remains untested. This paper has tried to honor that standard throughout: a measurement artifact was found and corrected before it could contaminate the identification pipeline; a marginal statistical result was flagged, investigated, and shown not to replicate; an intervention bug was caught by a positive-control check rather than trusted on the basis of a clean-looking run; and a causal pilot with a suggestive but underpowered pattern was reported as exactly that. The capacity–dissipation decomposition tested here does not, on current evidence, improve prediction of MIPS-relevant structure beyond density. Whether it nonetheless carries causal content is the question the next, properly powered matched-region study is positioned to answer.

Appendices

A Numerical stability of the integration scheme

Explicit Euler–Maruyama integration of the Weeks–Chandler–Andersen potential is only conditionally stable, and the instability is easy to miss with a short validation run. In one representative case, a configuration remained numerically well-behaved for approximately 38,000 steps before a close encounter between two particles triggered a force-magnitude runaway that reached 4×10^{10} within a further several hundred steps; a probe run of only the first 5,000 steps at the same parameters would have reported no problem. The safe timestep was found empirically to depend on both the propulsion speed v_0 and the thermal diffusion strength D_t rather than on the interaction parameters alone: representative stable values ranged from $dt = 10^{-3}$ at low activity ($v_0 \approx 8$) down to $dt = 10^{-4}$ at the highest activity levels and packing fractions tested in this study. Because of this sensitivity, every parameter combination used in the reported results was individually probed for stability at its full intended run duration before being used for data collection, rather than assumed safe by similarity to a previously validated configuration. The permanent diagnostic added to the simulation engine raises an explicit error whenever the maximum inter-particle force magnitude exceeds a fixed threshold, so that a run at an unvalidated configuration terminates with a clear error rather than completing and returning a trajectory containing non-physical values.

B Data provenance and export schema

Each simulated run’s reconstructed fields, extracted scalar features, targets, onset labels, and metadata were written to a single per-run data file using a fixed schema: shared reconstructed-field arrays (density, current, velocity, capacity variants, dissipation estimators) stored once rather than duplicated across the three capacity branches; separate feature tables per branch (baseline density features, each capacity variant’s features, dissipation features, cross-field interaction features) so that branch-specific analyses cannot be contaminated by another branch’s features; the four density-domain target indicators stored as independent series rather than collapsed into a single scalar; the deprecated particle-contact cluster fraction retained only in a clearly separated diagnostic location, so that it could not be inadvertently reused as a target; and a metadata block recording initialization method, minimum-distance parameter, burn-in duration, random seed, kernel bandwidth, grid resolution, integration timestep, and software version. Every file additionally carries a boolean completion flag, set to true only after the entire run finishes without a stability failure or other error; files left incomplete by an interrupted process are therefore distinguishable from genuinely finished data without inspecting their contents in detail. This distinction was exercised in practice, not only in principle: across the batch export runs conducted for this study, several were interrupted mid-write by execution-time limits, and in every such case the resulting file was either unreadable or explicitly flagged incomplete, and was excluded from analysis accordingly.

C The intervention timing error

The counterfactual intervention framework represents a perturbation as active during a time window specified relative to the start of the counterfactual experiment (for example, “beginning immediately and lasting two time units”). An early implementation evaluated this window against the simulation state’s absolute internal clock rather than against elapsed time since the experiment began. For a freshly initialized state, whose internal clock starts at zero, this distinction is invisible, because absolute and experiment-relative time coincide. For a realistic post-equilibration starting state, whose internal clock has already advanced through a burn-in period before the counterfactual experiment begins, the two quantities differ by the burn-in duration, and a window nominally beginning “immediately” was in practice being compared against a clock value it could never match, so the perturbation never activated. Every test in the initial validation suite happened to use a freshly initialized state and consequently could not have exposed the discrepancy. The error was identified by directly inspecting particle positions in the treatment and control branches after a demonstration run and finding them identical, which is possible only if the intervention had no effect; a regression test using a state with a nonzero initial clock value was subsequently added to guard against recurrence. The broader methodological point is that an inactive intervention and a genuinely null causal effect produce identical output, so validating that an intervention is capable of producing a nonzero effect at all is a necessary precondition for interpreting any subsequent null result as informative about the system rather than about the implementation.

References

- [1] Y. Fily and M. C. Marchetti, “Athermal Phase Separation of Self-Propelled Particles with No Alignment,” *Physical Review Letters*, 108, 235702 (2012).
- [2] G. S. Redner, M. F. Hagan, and A. Baskaran, “Structure and Dynamics of a Phase-Separating Active Colloidal Fluid,” *Physical Review Letters*, 110, 055701 (2013).
- [3] M. E. Cates and J. Tailleur, “Motility-Induced Phase Separation,” *Annual Review of Condensed Matter Physics*, 6, 219–244 (2015).
- [4] C. Bechinger et al., “Active Particles in Complex and Crowded Environments,” *Reviews of Modern Physics*, 88, 045006 (2016).
- [5] M. C. Marchetti et al., “Hydrodynamics of Soft Active Matter,” *Reviews of Modern Physics*, 85, 1143–1189 (2013).
- [6] U. Seifert, “Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines,” *Reports on Progress in Physics*, 75, 126001 (2012).
- [7] R. E. Spinney and I. J. Ford, “Entropy Production in Full Phase Space for Continuous Stochastic Dynamics,” *Physical Review E*, 85, 051113 (2012).
- [8] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall (1986).

- [9] R. Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society B*, 58, 267–288 (1996).
- [10] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., OTexts (2018).
- [11] P. Good, *Permutation, Parametric and Bootstrap Tests of Hypotheses*, Springer (2005).
- [12] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer (2004).
- [13] Flyxion, “Prediction Before Interpretation: A Criterion for Field-Like Constructs in Complex Systems,” companion manuscript (2026). <https://standardgalactic.github.io/calculus/experiments/prediction-before-interpretation.pdf>