This supplement expands the numerical definition and reports the complete case comparisons underlying the main text. Fixed ANUGA fields are independent reference inputs; their finite resolution and source closure remain part of the interpretation.

S1 Discrete face and retained-state operators

This section specifies the robust HLL interior operator used in the comparisons. Physical and masked boundary closures are detailed in S15 and S26.1. The description defines the implemented arithmetic; it does not establish omission safety for arbitrary terrain.

S1.1 Thresholds, reconstruction and source flow

Table S1: Shared terminology, numerical constants and discrete branch rules.

Item	Discrete rule or value
Stored domain / N	Allocated grid cells, including dry surroundings; distinct from wet cells and selected transport support
Physical wet region	Region containing water in the continuum problem; not directly observed by a numerical threshold
Diagnostic wet set	Cells with strict $h > \tau$ in saved cell-average fields, usually $\tau = 1$ or 10 mm; used for reported wet-set discrepancy
Selected support / A	Cumulative cell-equivalent transport selections across stages (block counts times 256); a work proxy, not total floating-point instructions or memory traffic
Initial support / S_0, W_0	Initial selected cell-equivalent support and initial wet-cell count, respectively; held fixed in the static models
Steps / K; T1/T2	Actual physical steps; both T1/T2 use two-stage SSP RK2 transport. T1/T2 have asymmetric/symmetric additive-source splitting, respectively; coupling orders are formal smooth-flow statements, not measured whole-algorithm convergence (main Sect. 2.2)
F/C/R	Filtered blocks / compacted blocks / restored full-block work in C's preparation context; R is an intervention, not the RAW clock
MONO/RAW	Adjusted CLOCK_MONOTONIC / CLOCK_MONOTONIC_RAW host durations, over the same synchronized advancement interval
Gravity	$g = 9.81$ m s^{-2}
Flux	HLL/HLLE
Slope limiter	`minmod(a,b)` returns zero for opposite signs and otherwise the smaller absolute slope with the common sign; water surface and bed are reconstructed at faces, velocity in the robust path
Near-dry fallback	$h_{\mathrm{near}} = 10^{-2}$ m; inspect raw four-cell depths before adding bed elevations; if any is at/below this value, use first-order bed-difference hydrostatic reconstruction (HR). Nonnegative face/reconstructed depths at/below it also trigger fallback
Transport threshold	$h_{\mathrm{tr}} = \mathrm{float32}(10^{-3})$ m at reconstructed faces; when both depths are below it, mass and transverse-momentum fluxes are zero and normal pressure is $g(h_L^2 + h_R^2)/4$. Near-dry HR face depth never exceeds its raw cell depth
Dry-state threshold	$h_{\mathrm{dry}} = \mathrm{float32}(10^{-3})$ m; block wet flags require $h \geq h_{\mathrm{dry}}$, while momentum is cleared for $h < h_{\mathrm{dry}}$; cell-index seeds retain all $h > 0$ cells; source/boundary activity can still select a cell or block
Positivity epsilon	$\epsilon_h = 10^{-7}$ m; post-update depths below this value are set to zero with zero momentum
Flux positivity factor	$\theta = 0$ if $h \leq \epsilon_h$; $\theta = 1$ if outgoing depth change is at most 10^{-12} m; otherwise $\theta = \min(1, \max(h - \epsilon_h, 0)/\Delta h_{\mathrm{out}})$; shared face fluxes use the donor's factor, except mass flux magnitudes at most 10^{-12} m^2 s^{-1} use unity
Froude cap	$\mathrm{Fr}_{\max} = 10$ in post-transport clipping; if $\|\boldsymbol{u}\| > 10\sqrt{gh}$, both momentum components are scaled by $10\sqrt{gh}/\|\boldsymbol{u}\|$

Table S1 continued.

Item	Discrete rule or value
Source exhaustion	Exact constant-rate net depth flow; after exhaustion available supply is assigned to negative scalar source, infiltration, then drainage (formula below); unmet demand is recorded
Rainfall conversion	$r_{\text{depth}} = R_{\text{mm/h}}\Delta t/(3.6 \times 10^6)$

Face routines reconstruct from original cell depths and local bed differences. First-order hydrostatic depth is $h^* = \max(0, \text{fl}(h + \text{fl}(b - b_{\max})))$, with $b_{\max} = \max(b_L, b_R)$; explicit rounded CUDA operations prevent fast-math reassociation into $(h+b)-b_{\max}$. Thus $0 \le h^* \le h$, including at the transport threshold, and a constant bed offset cannot change this branch. MUSCL tests the raw four-cell depths first and falls back before elevation arithmetic if any is at/below 1 cm. Fully wet reconstruction uses water surface and bed relative to the left cell's bed datum; open-boundary reconstruction uses the edge cell's datum and passes ghost depth separately. Invariance under a constant elevation offset requires the supplied float32 bed differences to survive the shift; quantizing new terrain inputs can change those differences.

For positive face directions east and north, the outgoing depth increment is $\Delta h_{\text{out}} = \Delta t\{[\max(F_{h,E}, 0) + \max(-F_{h,W}, 0)]/\Delta x + [\max(G_{h,N}, 0) + \max(-G_{h,S}, 0)]/\Delta y\}$. Limiting scales all three components of a shared flux; hydrostatic bed corrections retain their separate balance terms. These thresholds and fallback rules are fixed numerical operations in every execution-path comparison.

For a source substep of length τ, put $a = r + \max(q_s, 0) + j$, $b = \max(-q_s, 0) + i + d$ and $v = a - b$. The new depth is $h^+ = \max(0, h + \tau v)$. If $v < 0$, the wet duration is $t_w = \min(\tau, \max(h, 0)/(-v))$; otherwise $t_w = \tau$. Over $t_d = \tau - t_w$, distribute the supply rate in the fixed order $\rho_q = \min(\max(-q_s, 0), a)$, $\rho_i = \min(i, a - \rho_q)$ and $\rho_d = \min(d, a - \rho_q - \rho_i)$. The applied removal of type k is $b_k t_w + \rho_k t_d$; its requested amount is $b_k \tau$. Injection adds $\tau j(u_j, v_j)$ to momentum; mass-only sinks do not remove proportional momentum. Then depths below ϵ_h are zeroed and momentum is cleared below h_{dry}.

For sparse injection, conversion from volumetric discharge to a depth rate requires cell area greater than 10^{-9} m^2. In the following step description, U denotes the public conserved fields and R their retained residuals.

```
Apply the first Manning half-flow to (U, R).
For T2, apply the constrained additive half-flow.
For both integrators, save the transport initial state (U0, R0).
For each of the two Euler transport stages:
    Form selection from that stage's state plus active sources/boundaries.
    Expand the required stencil; prepare indexed output and positivity storage.
    Compute donor positivity factors, then transport with matching face fluxes.
    Preserve omitted public/residual state; form the next stage's wet flags.
On every valid cell, average (U0, R0) with the second stage (U2, R2).
Apply the remaining additive flow (a full step for T1, a half step for T2).
Apply the final Manning half-flow, positivity/dry/Froude clipping and residuals.
Rebuild wet flags and reduce wave speed for the next physical step.
```

S1.2 Face fluxes and retained state

At an x face, the four raw depths are checked before elevation arithmetic. If any is at/below 0.01 m, first-order reconstruction uses $h_s^* = \max(0, \text{fl}_{32}(h_s + \text{fl}_{32}(b_s - \max(b_L, b_R))))$. The reconstructed momentum is $h_s^* q_s / h_s$ for $h_s \ge h_{\text{dry}}$ and zero otherwise. Fully wet MUSCL first subtracts the left-cell bed datum, then reconstructs water surface and bed with minmod slopes. Write $w_{L,f} = w_L + \sigma_{w,L}/2$, $w_{R,f} = w_R - \sigma_{w,R}/2$ and analogously $b_{L,f}, b_{R,f}$. The pre-HR face depth is $w_{s,f} - b_{s,f}$ and $h_s^* = \max(0, w_{s,f} - \max(b_{L,f}, b_{R,f}))$. A reconstructed/pre-HR depth

at/below 0.01 m also triggers first-order fallback. Wet velocities are minmod-reconstructed from $v(q,h) = 2hq/[h^2 + \max(h^2, 10^{-7}) + 10^{-30}]$. Only the robust production branch is described here; the unused nonrobust reconstruction has different surface limiting.

For side $s \in \{L, R\}$ in normal coordinates, define $U_s^* = (h_s^*, q_{n,s}^*, q_{t,s}^*)^T$ from the hydrostatic reconstruction above. The normal velocity used by the flux is $u_s = q_{n,s}^*/h_s^*$ for $h_s^* \geq h_{\mathrm{dry}}$ and zero below it. Its side flux is explicitly

$$F_s = \begin{pmatrix} h_s^* u_s \\ q_{n,s}^* u_s + \frac{1}{2} g (h_s^*)^2 \\ q_{t,s}^* u_s \end{pmatrix}. \tag{S1}$$

Hydrostatic truncation can give $0 < h_s^* < h_{\mathrm{dry}}$ while retaining nonzero $q_{n,s}^*$ or $q_{t,s}^*$ from an originally wet cell. In that case the side's mass and advective momentum contributions are zero because $u_s = 0$, while its pressure remains. The reconstructed momenta are retained in U_s^* for the HLL jump term. Thus F_s is the thresholded side flux defined here, not an unrestricted substitution into the continuous $F(U)$ of the main text; the combined interface mass flux can still be nonzero when the other side is wet.

Let $c_s = \sqrt{g h_s^*}$, $\widetilde{u} = (\sqrt{h_L^*} u_L + \sqrt{h_R^*} u_R)/(\sqrt{h_L^*} + \sqrt{h_R^*} + 10^{-30})$ and $\widetilde{c} = \sqrt{g(h_L^* + h_R^*)/2}$. The HLL speeds are $s_L = \min(u_L - c_L, \widetilde{u} - \widetilde{c})$ and $s_R = \max(u_R + c_R, \widetilde{u} + \widetilde{c})$. If both face depths are below the transport threshold, the routine returns $(0, g[(h_L^*)^2 + (h_R^*)^2]/4, 0)$ directly. Otherwise the interface flux is F_L for $s_L \geq 0$, F_R for $s_R \leq 0$, and $[s_R F_L - s_L F_R + s_L s_R (U_R^* - U_L^*)]/(s_R - s_L)$ between the waves, using the retained reconstructed conserved vectors. The y-face operator swaps normal and tangential components.

For each side the MUSCL pressure correction is $K_s = (h_s^*)^2 - (w_{s,f} - b_s)^2$; the latter pressure uses the cell bed, which includes the centred bed-source contribution. First-order correction is $K_s = (h_s^*)^2 - h_s^2$. The x momentum source is $g(K_E - K_W)/(2\Delta x)$, and the y expression is analogous. The transport increment uses the donor-limited common face fluxes from Table S1, while these pressure corrections remain separate. Their sign convention is the one used by the shared numerical face and transport operators.

For each conserved scalar, retain a float32 pair (u, r) representing float64$(u) + r$. Flux/reconstruction uses public u; the Euler update adds its increment to that pair in double precision. In the source-free transport stages used here, the depth main value is rounded to float32, its residual is the float32 difference from the double update, and depth/momentum are clipped using Table S1. Zero depth clears its residual. Momentum is rounded and its residual stored after dry/Froude clipping. In the finalizer, the double expression is evaluated in the recorded order $0.5(u_0 + u_2 + r_0 + r_2)$ for each component. The remaining constrained source flow and final Manning half-flow precede the final dry/Froude clipping and pair storage. In T1 the source placement is the full remaining flow; T2 uses a half flow. Stage and source-flow roundings must not be replaced by a public float32 average or by importing a snapshot, which clears hidden residuals.

```
Stage input: three public fields U, three residuals R, active sources and boundaries.
Build stage seeds; expand block neighbours or cell stencil; form selected lists.
For indexed paths, copy all six state fields and clear positivity storage.
For a filtered block that skips transport, preserve all six fields and bookkeeping.
Compute matching reconstructed face fluxes and donor positivity factors.
Update selected pairs in double intermediates; apply the specified clipping/storage.
Repeat selection from the actual second-stage input, not the first-stage mask.
On every valid cell, combine both stage pairs, finish sources/friction, and
↪  clip/store.
Recompute wet flags and speed from the actual final state; reduce the global step
↪  bound.
```

This makes explicit why a transport-stage mask is insufficient for averaging: an omitted cell may still contain a nonzero stage-zero contribution or residual. An identity claim must preserve

the full pair, pressure/source corrections, clipping, flags and speed contribution as well as mass transport.

S2 Time control and boundary scope

The CFL speed is $a_i = \sqrt{gh_i} + \sqrt{(q_{x,i}/h_i)^2 + (q_{y,i}/h_i)^2}$ for valid $h_i \geq h_{\mathrm{dry}}$, and zero below it. The global bound is the maximum over these cells and prescribed incoming boundary wave bounds. The implementation uses $\delta = \mathrm{CFL}\min(\Delta x, \Delta y)/(a_{\max} + 10^{-6})$; it uses a 1 s first-step proposal when $a_{\max} \leq 10^{-3}$ m s^{-1}, and caps later low-speed proposals at 1 s. Subsequent growth is limited to ten times the previous step. Otherwise the case-specific $\Delta t_{\max}$ caps the proposal. A CFL proposal below $\Delta t_{\min}$ is raised to that floor and increments a stagnation counter; 100 consecutive such steps terminate the run. An event or output tail shorter than the floor is allowed and resets that counter. Finally the step is clipped to the next event/output time. This is a scalar speed bound with the smaller cell spacing, not a directional wave-speed sum. Case-specific CFL values and caps are in main Sect. 3. The floor is a controlled termination policy, not a proof of stability when it is active. With local bed-difference near-dry reconstruction, two subthreshold raw cells cannot generate an above-threshold mass-transport face by a change of elevation datum. Their zero-speed classification is therefore consistent with this mass gate. The analytical checks constrain specific cases; they do not establish a universal face-based CFL theorem for nonlinear reconstruction, sources or clipping.

S3 Numerical criteria and analytical cases

Numerical criteria distinguish errors against reference solutions from agreement between execution paths. Analytical and channel tests also check boundary response and volume balance; these criteria define the tested numerical scope. Ritter and Thacker belong to the analytical families compiled by Delestre et al. (2013); the channel cases follow MacDonald et al. (1997), both cited in the main bibliography. The depth, discharge and wet-set limits are study acceptance tolerances, not published accuracy standards. They distinguish field agreement, front disagreement and conservation at different scales; passing a depth or volume tolerance cannot override a failed wet-set check. The archived runners record these criteria before their recorded executions, but do not establish that every threshold preceded all exploratory results. We therefore make no preregistration claim. Reported errors, including failed front checks, remain available independently of the pass/fail labels.

Table S2: Fixed numerical diagnostics. They define the tested scope, not application-calibrated flood-prediction accuracy.

Check	Criterion
Analytical/channel depth	$E_1(h) \leq 0.05H$, with $H = \max(\max h^{ref}, \max h^0, 10^{-6}$ m)
Reflective static lake	All three fields have $E_\infty \leq 5 \times 10^{-6}$ in their corresponding units
Subcritical channel discharge	$E_1(q_x) < 0.1$ m^2 s^{-1}; the two smooth 64-cell cases use the stricter < 0.02 m^2 s^{-1} criterion
Circular independent reference	Depth $E_1 \leq 0.025$ m, each momentum $E_1 \leq 0.08$ m^2 s^{-1} and $D_\eta \leq 25\%$
Paired-source independent reference	Depth $E_1 \leq 0.005$ m, each momentum $E_1 \leq 0.01$ m^2 s^{-1} and $D_\eta \leq 25\%$
Reference sensitivity	Depth $E_1 \leq 0.003$ m, momentum $E_1 \leq 0.01$ m^2 s^{-1} and $D_\eta \leq 5\%$; an empirical sensitivity test, not an exact-error bound
Paired-source volume	Inter-solver volume difference $\leq$0.5 % of gross supplied volume; separate manual-step requested/applied ledger residual $\leq 5 \times 10^{-5}$

Table S2 continued.

Check	Criterion
Source-free/urban performance runs	Relative volume residual $\leq 2 \times 10^{-6}$ where available; the strict release interface does not expose the urban ledger
Execution paths	Bitwise public fields and identical steps, with finite/nonnegative states

Analytical/channel runs also require relative volume residual at most 5×10^{-5}. Channel states remain subcritical, and prescribed inflow is checked to relative tolerance 10^{-5}. Circular volume residual is normalised by initial volume and limited to 10^{-5}. Finite, nonnegative public states and zero values in inactive cells are checked separately.

Table S3: Analytical and channel field errors.

Case	Depth E_1 (m)	q_x E_1 (m^2 s^{-1})	q_y E_1 (m^2 s^{-1})
Ritter, 64 / T1	0.01434979	0.06123397	0
Ritter, 64 / T2	0.01434979	0.06123397	0
Ritter, 128 / T1	0.008204481	0.03712224	0
Ritter, 128 / T2	0.008204481	0.03712224	0
Ritter, 256 / T1	0.005059751	0.02488825	0
Ritter, 256 / T2	0.005059751	0.02488825	0
Thacker, 128 / T1	0.00077261	0.001042989	0.001286902
Thacker, 128 / T2	0.00077261	0.001042989	0.001286902
Static lake, 64 / T2	1.488092e-08	4.817984e-08	5.138792e-08
Steady channel, 64 / T1	0.0002918022	0.00123588	0
Steady channel, 64 / T2	0.0002918115	0.001235846	0
Perturbed channel, 64 / T1	0.000291872	0.001235867	0
Perturbed channel, 64 / T2	0.0002918692	0.001235835	0
Steady channel, 128 / T1	7.179962e-05	0.0003037769	0
Steady channel, 128 / T2	7.181824e-05	0.0003037611	0

S4 Complete time comparisons

Intervals are conditional descriptive 95 % bootstrap intervals over four processes and paired repeats. They are not adjusted for multiple comparisons, and cases share numerical protocols and the same GPU. Process repeats do not add independent spatial realizations. Ratios above one favour the denominator path.

S4.1 All 36 main cases

Table S4: Complete-time ratios for all 36 main cases.

Case	Dense/filtered	Dense/compacted	Dense/cell
Continuous, 1 % / T1	1.944 [1.933, 1.954]	1.711 [1.703, 1.723]	1.483 [1.480, 1.505]
Continuous, 1 % / T2	1.960 [1.932, 1.972]	1.717 [1.692, 1.750]	1.489 [1.479, 1.496]
Continuous, 5 % / T1	1.792 [1.781, 1.800]	1.547 [1.537, 1.555]	1.410 [1.400, 1.423]
Continuous, 5 % / T2	1.894 [1.886, 1.903]	1.640 [1.636, 1.647]	1.463 [1.457, 1.467]
Continuous, 10 % / T1	1.662 [1.657, 1.671]	1.448 [1.417, 1.455]	1.270 [1.262, 1.274]
Continuous, 10 % / T2	1.792 [1.783, 1.797]	1.514 [1.484, 1.523]	1.331 [1.309, 1.337]
Continuous, 25 % / T1	1.394 [1.390, 1.396]	1.214 [1.206, 1.224]	1.079 [1.075, 1.080]
Continuous, 25 % / T2	1.533 [1.528, 1.554]	1.288 [1.284, 1.301]	1.117 [1.114, 1.123]
Continuous, 50 % / T1	1.173 [1.171, 1.175]	1.055 [1.051, 1.059]	0.932 [0.931, 0.935]
Continuous, 50 % / T2	1.264 [1.259, 1.269]	1.080 [1.067, 1.088]	0.909 [0.903, 0.911]

Table S4 continued.

Case	Dense/filtered	Dense/compacted	Dense/cell
Continuous, 100 % / T1	0.994 [0.993, 0.996]	0.915 [0.914, 0.916]	0.825 [0.823, 0.827]
Continuous, 100 % / T2	0.996 [0.992, 0.996]	0.865 [0.855, 0.867]	0.732 [0.729, 0.734]
Fragmented, 1 % / T1	1.262 [1.257, 1.268]	1.114 [1.104, 1.120]	1.353 [1.338, 1.361]
Fragmented, 1 % / T2	1.301 [1.296, 1.306]	1.124 [1.117, 1.128]	1.382 [1.369, 1.388]
Fragmented, 5 % / T1	1.003 [1.001, 1.006]	0.899 [0.895, 0.901]	1.006 [1.004, 1.009]
Fragmented, 5 % / T2	1.005 [1.000, 1.008]	0.868 [0.867, 0.873]	1.021 [1.017, 1.023]
Fragmented, 10 % / T1	0.994 [0.990, 0.996]	0.902 [0.895, 0.905]	0.891 [0.885, 0.896]
Fragmented, 10 % / T2	0.992 [0.987, 0.995]	0.866 [0.864, 0.870]	0.868 [0.865, 0.869]
Fragmented, 25 % / T1	0.994 [0.986, 0.997]	0.915 [0.913, 0.915]	0.811 [0.810, 0.813]
Fragmented, 25 % / T2	0.994 [0.992, 1.000]	0.874 [0.872, 0.877]	0.742 [0.740, 0.745]
Fragmented, 50 % / T1	0.997 [0.995, 0.998]	0.915 [0.913, 0.917]	0.783 [0.781, 0.784]
Fragmented, 50 % / T2	0.995 [0.992, 0.997]	0.873 [0.872, 0.876]	0.696 [0.696, 0.699]
Fragmented, 100 % / T1	0.995 [0.994, 1.006]	0.916 [0.915, 0.920]	0.825 [0.821, 0.829]
Fragmented, 100 % / T2	0.995 [0.992, 0.996]	0.868 [0.864, 0.869]	0.733 [0.730, 0.734]
Bands, 1 % / T1	1.356 [1.352, 1.359]	1.215 [1.214, 1.219]	1.292 [1.289, 1.296]
Bands, 1 % / T2	1.448 [1.437, 1.453]	1.265 [1.258, 1.270]	1.374 [1.371, 1.377]
Bands, 5 % / T1	1.356 [1.353, 1.358]	1.217 [1.215, 1.218]	1.219 [1.217, 1.222]
Bands, 5 % / T2	1.447 [1.443, 1.450]	1.268 [1.259, 1.271]	1.282 [1.278, 1.283]
Bands, 10 % / T1	1.268 [1.266, 1.272]	1.158 [1.155, 1.162]	1.233 [1.231, 1.236]
Bands, 10 % / T2	1.345 [1.340, 1.346]	1.204 [1.197, 1.207]	1.244 [1.220, 1.248]
Bands, 25 % / T1	1.163 [1.160, 1.165]	1.075 [1.068, 1.078]	1.037 [1.034, 1.041]
Bands, 25 % / T2	1.230 [1.225, 1.236]	1.103 [1.098, 1.106]	1.038 [1.036, 1.040]
Bands, 50 % / T1	1.062 [1.061, 1.064]	0.969 [0.968, 0.972]	0.883 [0.881, 0.886]
Bands, 50 % / T2	1.094 [1.090, 1.096]	0.956 [0.954, 0.963]	0.831 [0.829, 0.834]
Bands, 100 % / T1	0.995 [0.993, 0.996]	0.915 [0.914, 0.916]	0.826 [0.824, 0.827]
Bands, 100 % / T2	0.994 [0.991, 0.996]	0.862 [0.861, 0.868]	0.732 [0.727, 0.734]

S4.2 Sources and urban grids

Table S5: Complete-time ratios for source and urban cases.

Case	Dense/filtered	Dense/compacted	Dense/cell
Paired sources, 512^2 / T2	1.254 [1.232, 1.264]	0.995 [0.956, 1.030]	0.936 [0.899, 0.981]
Uniform rainfall, 256^2 / T1	1.033 [1.016, 1.062]	0.706 [0.679, 0.721]	0.797 [0.789, 0.812]
Uniform rainfall, 512^2 / T2	1.041 [1.036, 1.051]	0.869 [0.859, 0.883]	0.857 [0.848, 0.861]
Single source, 256^2 / T1	1.208 [1.162, 1.251]	0.749 [0.711, 0.793]	0.726 [0.707, 0.757]
Merewether coarse	0.949 [0.947, 0.952]	0.634 [0.627, 0.642]	0.768 [0.765, 0.771]
Merewether fine	1.004 [1.002, 1.007]	0.770 [0.768, 0.770]	0.922 [0.920, 0.925]

S4.3 Copy/fused intervention

Table S6: Matched copy/fused time ratios.

Case	Mode	Copy/fused time
Continuous, 1 % / T1	Compacted	1.030 [1.019, 1.045]
Continuous, 1 % / T1	Cell indexed	1.057 [1.046, 1.069]
Continuous, 1 % / T2	Compacted	1.039 [1.029, 1.059]
Continuous, 1 % / T2	Cell indexed	1.058 [1.048, 1.063]
Continuous, 100 % / T1	Compacted	1.017 [1.015, 1.020]
Continuous, 100 % / T1	Cell indexed	1.010 [1.009, 1.012]
Continuous, 100 % / T2	Compacted	1.026 [1.024, 1.032]

Table S6 continued.

Case	Mode	Copy/fused time
Continuous, 100 % / T2	Cell indexed	1.014 [1.011, 1.018]

S4.4 Direct paired comparisons at 1 % wet area

Table S7: Direct paired ratios for the same T2, 1 % inputs. Paired ratios need not equal ratios of the separately displayed median times.

Geometry	Filtered/compacted	Filtered/cell	Compacted/cell
Continuous	0.880 [0.868, 0.890]	0.760 [0.758, 0.765]	0.867 [0.855, 0.875]
Fragmented	0.864 [0.859, 0.867]	1.062 [1.052, 1.066]	1.228 [1.223, 1.234]
Bands	0.874 [0.871, 0.878]	0.950 [0.946, 0.954]	1.087 [1.083, 1.091]

S4.5 Instrumented cost decomposition

Table S8: Separate instrumented runs for T2, 1 % continuous input. Means are CUDA-event milliseconds, not host-clock milliseconds or fractions of an ordinary advancement budget. Embedded preservation is not zero cost or an equivalent transport accounting category; see the clock audit in S23.

Path	List	Preparation	Transport	Finalization	Profile/ordinary wall time
Dense	—	Embedded	5.689	1.629	1.067
Filtered	—	Embedded	1.299	1.651	1.138
Compacted	0.676	1.200	0.530	1.623	1.191
Cell indexed	1.517	1.180	0.631	1.535	1.147

S5 Translation and evolving support

The comparison undoes a seven-cell shift over the common 512 by 2041-cell region. Only exported public conserved fields at saved times are compared; hidden residuals and intervening stages are not exported. The bands agree through 2 s. The later fragmented differences coincide with different wall interactions, so their timing cannot isolate alignment alone.

Table S9: Conserved-field agreement after undoing translation.

Input	Time (s)	Bitwise	Depth max (m)	qx max (m^2 s^{-1})	qy max (m^2 s^{-1})	Wet edge rows
Fragmented	0.2	Yes	0	0	0	0/0
Fragmented	1.0	No	0.0030926047	0.0012711436	0.0002109427	0/20
Fragmented	2.0	No	0.038485499	0.016104686	0.02955899	36/72
Bands	0.2	Yes	0	0	0	0/0
Bands	1.0	Yes	0	0	0	0/0
Bands	2.0	Yes	0	0	0	0/0

Stage-weighted support uses selected updates divided by twice the physical step count and domain size. This is not physical-time weighting. Initial support and window support must be kept separate.

S6 Independent-reference and source-budget diagnostics

Table S10: Independent-reference criteria by family.

Family	Reference	Comparisons	Depth failures	Momentum failures	Front failures	Undefined wet sets
circular	finest	32	0	0	0	0
source	standard	160	0	0	28	12
source	Half-step	160	0	0	32	12
source	closure	160	0	0	32	12

Counts above refer to combinations of grid, integrator, output time and threshold for the dense result. Equal execution paths are not counted as independent replicates. Undefined wet-set ratios are included in front failures. Small mean field errors do not remove those failures.

Table S11: Reference/reference sensitivity: each ANUGA timestep or source-closure variant is compared with the standard ANUGA reference after restriction.

Comparison grid	Reference control	Max depth E_1 (m)	Max momentum E1 (m^2 s^{-1})	Max finite D (%)	Front criteria exceeded
64	Half-step	8.479698e-06	8.128117e-06	4.545	0
64	closure	5.633834e-05	5.586508e-05	11.765	2
256	Half-step	1.368053e-05	1.26423e-05	0.950	0
256	closure	8.835293e-05	8.553852e-05	17.857	7

These reference/reference values compare conservatively restricted states at the same target grid. They do not quantify unresolved reference spatial error or provide an exact-solution bound.

Table S12: Requested and applied source-budget residuals.

Grid	Integrator	Max requested residual	Max applied residual
64	1	6.7897187e-07	7.5273293e-08
64	2	9.0010897e-07	8.2775623e-08
256	1	8.9879829e-05	2.1435927e-05
256	2	0.00012559868	7.573881e-06

The budget rows use the separate manually selected float32 step sequence and gross supplied volume for normalisation. The original requested criterion remains exceeded on the fine grid; the alternative applied ledger does not replace it. The separate diagnostic-precision control in Sect. S11 identifies accumulation error while preserving these original outcomes. Release trajectories and their reference-volume comparisons are stored separately.

S7 Observational interpretation

Merewether point locations are fixed independently of wetting. Fine-grid sites 2 and 3 remain dry; site 2 has bed elevation above the observed mark. Conditional MAEs over wet sites use different subsets on coarse and fine grids, so they are not directly a convergence measure. Both fixed and containing-cell records, point distances, peak depth, arrival and velocity diagnostics remain in the raw tables. The supplied TUFLOW values are model comparisons, not field observations.

Numerical states, timing samples and figure-generation scripts are included in the research package described in the code and data availability statement.

S8 Grouped dry-domain reproduction

The development groups are a 64×64 square, sixteen 4×16 patches (seed 2026091401), sixty-four 8×8 patches (2026091402), and a 128×128 square. The held-out groups are a 32×128 rectangle, sixteen 16×4 patches (2026092401), thirty-two 8×16 patches (2026092402), and sixty-four 16×16 patches (2026092403). Patches occupy distinct aligned sites in a 192^2 region within the 256^2 local array. Sizes, offsets, windows and repeats stay in the same split. The 36 trajectory configurations per split arise from four groups, three domain sizes and three offsets; each has four execution paths and three continuing windows.

The 6120 trajectories comprise 288 release references, 72 development controls, 1152 warmups, 3456 ordinary samples and 1152 separate profiles. All 18360 saved windows satisfy state/step/exterior checks. Seventy-two full-state archives retain canonical references. The ordinary target is synchronized advancement; file output, setup, downloads and Python diagnostics are excluded. All host seconds and coefficients here are on the archived CLOCK_MONOTONIC scale; separate stage profiles use CUDA-event milliseconds.

Models are fitted separately per path using nonnegative relative least squares. Design and response are divided by observed seconds, columns scaled by L2 norm, and coefficients restored to physical units after NNLS. No constant independent of step count is added. Dense drops the exactly redundant domain feature. The complete coefficients, ranks, conditioning, group errors, all predictions and pairwise rankings are provided in `dry_extension/frozen_models.json` and `dry_extension/evaluation.json` under the new delivery campaign.

The Linux model/protocol freeze is recorded on 28 September 2026 in `linux4090_20260928/depth/dry_extension/`. The model was fixed before the Linux holdout run. The geometry families were already observed in earlier research; this is a platform reproduction, not a newly unseen scientific test. Frozen source and model hashes, rather than historical editorial commit dates, define this campaign.

The support+domain model has 7.35/6.74/13.30/15.27 % mean/median/P90/maximum trajectory APE and 0/36 non-best choices. Original no-domain baselines have mean APE 40.27 % for initial fraction and 28.15 % for evolving support. The stronger domain/steps comparator is 7.67 %, leaving a 0.32 percentage-point mean improvement and no additional best-path selection benefit. Engineering targets in the archived protocol are internal thresholds, not external model-validity standards. Complete model comparisons appear in Tables S13/S24.

S9 Isolated cell-selection intervention

The original square-neighbourhood expansion is replaced by two separable passes.

The new horizontal pass leaves intermediate positions unmasked; the vertical pass applies the valid output mask. Existing buffers are reused and CUB stable selection consumes the final mask. Numerical updates, two stage rebuilds, count readback, six-field preparation and RK finalization remain unchanged. The candidate adds a launch and intermediate traffic, included in ordinary time.

Thirty-two combinations cover dry/full-wet states, corners, irregular masks, subthreshold positive water, withdrawal/restart, rainfall and open x boundaries, with T1/T2 and space orders 1/2. Twelve fixed steps capture stage zero, both Euler stage inputs/outputs and final RK fields/residuals. Independent square dilation verifies 1536 lists and 2304 frames; four release trajectories check paired-source and Merewether coarse continuation. Timed release contains no probe exports. The 522 performance trajectories comprise 18 references, 72 warmups, 360

ordinary paired samples and 72 profiles. The limited large-domain large-domain gains and small/full-wet penalties are retained in Table S42.

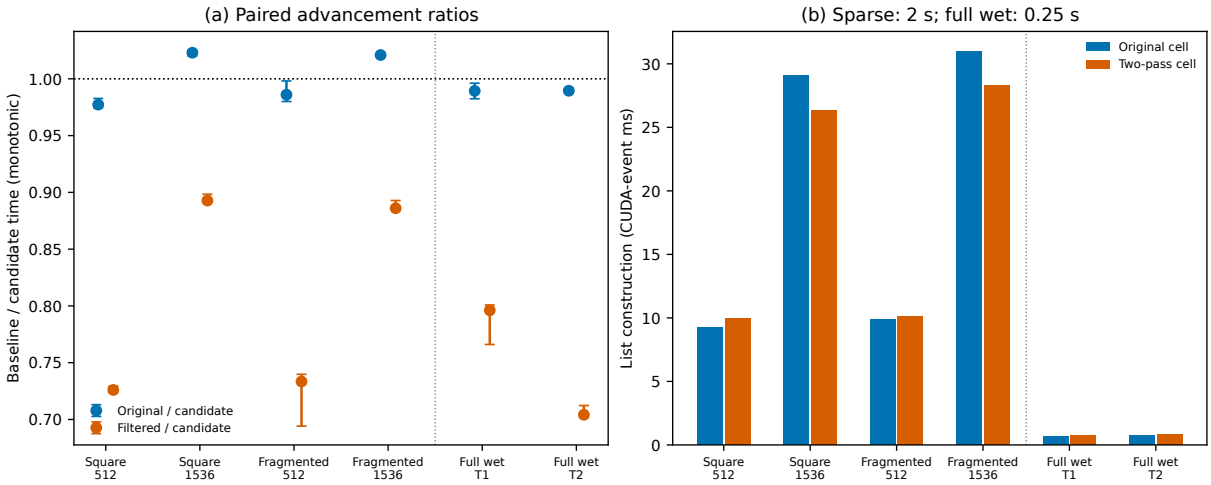


Figure 11: Cell-selection intervention: paired original/candidate and filtered/candidate advancement ratios, plus separately profiled CUDA-event list-build milliseconds. Sparse cases advance once to 2 s; fully wet cases advance to 0.25 s, so their absolute spans are not matched across case families. Three processes use five ordinary repeats and separate warmups/profiles. Host ratios use monotonic time; event spans are not fractions of that budget.

S10 Numerical and reporting cutoff sensitivity

Reporting cutoffs are 0.1/0.5/1/2/5/10/20/50 mm, applied to fixed conserved cell averages. The 2112 rows reproduce 512 original comparisons. Empty unions have undefined wet-union error; empty reference with nonempty numerical wet set has undefined relative discrepancy. Both-empty zero discrepancy is a convention, not improved front prediction.

Numerical cutoffs are 0.5/1/2 mm through the single active shared definition, covering transport, block wetness, CFL, reconstruction, momentum cleanup and friction. Cell seeds retain all positive water and positivity epsilon remains 10^{-7} m. Ritter uses 64/128/256 and paired withdrawal 64/256, T1/T2 and all paths, at fixed 1/10 mm reporting thresholds. Six cost inputs use three processes and three paired repeats (Table S30). The 57-file native inventory also includes legacy DRY_TOL files that are absent from build.sh's 13 translation units; their values do not govern these builds. The common policy also changes source momentum cleanup and the hazard gate; new hazard accuracy is not claimed. The 2 mm source policy has no matched ANUGA closure reference. Original reference arrays are reused with fixed hashes.

The planned 1296 trajectories comprise 162 validation runs and 1134 performance runs (54 references, 216 warmups, 648 ordinary samples, 216 profiles). All within-cutoff public state and step comparisons pass. Source front failures are 52/52/46 for 0.5 mm, 28/32/32 for 1 mm, and 28/30/32 for 2 mm against standard/half-step/closure references. Each reference contains 160 correlated comparisons per numerical cutoff. The 0.5 mm independent-volume failure counts are 32/32/28, duplicated across reporting cutoffs; field L1 checks still pass. No alternate cutoff is selected by a pooled failure count.

At 256^2/T2 and 4.5 s, the 0.5/1/2 mm policies give strict 1 mm discrepancy 203.57/75/50 % against the standard reference, while wet-union depth error is 0.003057/0.004524/0.005393 m. Missed area stays 0.683594 m^2; excess area falls from 2.099609 to 0.341797 to zero. Smaller discrepancy here does not imply smaller depth error. Ritter strict 1 mm discrepancy is unchanged across these policies: 4.08/6.19/6.70 % at 64/128/256 cells.

S11 Source diagnostic accumulation precision

Only four counters widen: positive and negative applied qsrc volume, total blocked source volume, and positivity-zeroed volume. Each original per-cell float32 contribution is explicitly preserved before conversion to double. Device/host storage, zeroing and readback sizes are changed consistently in an isolated development module. State precision and numerical arithmetic do not change. Double atomic accumulation is not claimed deterministic generally.

Across 64/256, T1/T2, four paths and three repeats, 96 trajectories have identical public field bytes and manual time/dt/phase/wave columns. At 256^2, float32 requested residual maxima reach 8.03145e-5 (T1) and 1.15518e-4 (T2); double accumulation reduces them to 6.08527e-7 and 6.21199e-7. Applied maxima decrease from 1.57553e-5/5.78153e-6 to 2.36094e-8/2.35850e-8. All requested checks then pass at the unchanged 5e-5 tolerance. These controls explain a diagnostic accumulation contribution without changing front error, reference-volume agreement, source physics or original failed records. No performance comparison or production adoption is claimed.

S12 Terrain and fixed-location input audit

An independent four-corner interpolation reproduces both complete coarse/fine ground arrays bitwise and checks building/road masks, roughness and fixed gauges. At original site 2, DEM is 23.580493 m against a 23.36 m mark; the fine containing cell adds 0.009528 m. Site 3 original DEM is 23.073623 m against a 23.14 m mark, so its dry result has a separate unresolved forcing/propagation explanation. Neither site lies inside buildings or receives interpolation weight from edge NoData. WGS84/UTM horizontal metadata do not establish a vertical datum. No survey correction is invented and no observation is moved to a wet cell.

S13 Reproduction provenance and scope

The two Linux campaigns use freshly compiled modules for sm_89 and sm_120, respectively. The full primary campaign and bounded secondary campaign have different scopes, stated in main Table 5. Source snapshots, binary hashes, raw timings and original failed orchestration records remain archived. `bindings.json` in each promoted campaign maps exact bytes to their original measurement-host locators. Earlier campaigns remain immutable but their performance values and fitted coefficients are not evidence for this revision. The four fixed ANUGA reference fields are reused with their original identities. The primary grouped campaign used CLOCK_MONOTONIC; later endpoint comparisons also measured CLOCK_MONOTONIC_RAW. No common raw-clock boundary is imputed to historical single-clock measurements within the Linux suite.

S14 Domain/step ablation and two new geometry groups

The review exposed a missing (K, NK) comparator. Its coefficients were fitted only to the original delivery development windows, using the same nonnegative relative least-squares procedure. The original full-model coefficients were retained without refitting. Original holdout results are explicitly post hoc. Both coefficient sets and the protocol were frozen before the Linux replay of the additional groups. These groups were known from earlier research, so the replay is not new independent geometry discovery. The campaign is `experiments/results/linux4090_20260928/domain/`.

The new 48×96 rectangle and 48 nonoverlapping 8×16 patches (seed 2026091391) are distinct complete inputs, not repetitions or translations of old groups. Each retains the three sizes, offsets and continuing windows of Sect. S8. Three processes each use one warmup, three

ordinary repeats and one separate profile per variant. There are 1152 trajectories: 54 references, 18 development interface checks, 216 warmups, 648 ordinary runs and 216 profiles. All 3456 saved-window checks and 18 canonical state archives pass the original common state, inverse-translation, boundary, dry-exterior, step and support guards.

Table S13: Complete-trajectory model ablation. Original hold-out rows are post hoc; new-group rows use coefficients and groups frozen before execution. All models require actual steps; the full model also requires actual support.

Data / model	Mean APE (%)	Median APE (%)	P90 APE (%)	Maximum APE (%)	Non-best / cases
Original / domain + steps (post hoc)	7.67	8.14	12.65	16.95	0 / 36
Original / support + domain	7.35	6.74	13.30	15.27	0 / 36
New / domain + steps	8.60	8.55	13.80	19.46	0 / 18
New / support + domain	8.48	9.01	13.27	14.96	0 / 18

The mean APE reduction is 0.12 percentage points on the subsequent groups. Both models select every point-estimate fastest path, with zero selection extra time. The groups remain source-free and sparse; these descriptive errors do not estimate uncertainty over a broad geometry population.

Two inherited protocol fields require clarification without rewriting the frozen file. The `uncertainty` prose says four processes, but the explicit `processes=3`, run records and actual conditional bootstrap all use three. The retained original `engineering_targets` field is not applied to this ablation: its explicit evaluation rule and pre-run protocol impose no superiority threshold. Original engineering targets retain their original scope. No failures or unfavorable comparisons were removed.

S15 Flux components and boundary and splitting details

For $h > 0$, the fluxes and forcing fields in Sect. 2.1 are

$$F(U) = \begin{pmatrix} q_x \\ q_x^2/h + gh^2/2 \\ q_x q_y/h \end{pmatrix}, \qquad G(U) = \begin{pmatrix} q_y \\ q_x q_y/h \\ q_y^2/h + gh^2/2 \end{pmatrix}, \tag{S2}$$

$$S_b = \begin{pmatrix} 0 \\ -gh\,\partial_x b \\ -gh\,\partial_y b \end{pmatrix}, \qquad S_c = \begin{pmatrix} r - i - d + q_s + j \\ j_u \\ j_v \end{pmatrix}. \tag{S3}$$

Here b is bed elevation, r uniform rainfall, i and d nonnegative infiltration and drainage rates, and q_s the signed scalar source. Sparse injection has depth rate $j = Q_{\mathrm{inj}}/(\Delta x \Delta y)$ and momentum rates $j_u = j u_j$ and $j_v = j v_j$ for specified injection velocity. Hydrostatic reconstruction applies bed forcing through face corrections. Fully wet reconstruction also uses a local bed datum;

invariance to a constant elevation shift requires the input float32 bed differences to survive that shift.

At an open boundary normal to x, prescribed discharge or stage is combined with the outgoing shallow-water characteristic. Second-order closure uses three interior cells; short, masked or near-dry stencils revert to first order. Incoming tangential momentum uses the prescribed exterior transverse velocity, zero in the channel tests; outgoing momentum uses the interior value. Other analytical/channel edges reflect. Urban outflow follows Sect. S26.1.

For smooth wet solutions, T1's local splitting error contains $\Delta t^2(C'L - L'C)/2$, where C is the additive-source vector field and L the transport/bed operator. This term vanishes in the uniform flat-bed source tests because transport vanishes throughout the source flow; it need not vanish under spatially varying forcing. T1 generally has first-order transport–source coupling, whereas the symmetric T2 composition has formally second-order coupling in the smooth regime. This is a splitting statement, not measured convergence of the complete implementation. Clipping, exhaustion and moving wet/dry interfaces fall outside that analysis. No temporal convergence order is measured here. The complete CFL and minimum-step rules remain in Sect. S2.

S16 Evidence scope and falsifiable cost interpretation

The tested hypothesis, subject to the audit coverage in Table S14, is that fixing checked local evolution and absolute selected transport support does not fix the relative cost of the execution paths as stored size changes. The control varies dry extent, checks saved common-domain fields, steps and support, and compares synchronized complete advancement. The measurements support that dependence: the filtered/compacted ratio moves substantially with stored extent on both tested platforms while the checked quantities are held fixed.

The stronger form of the hypothesis, that this dependence is large enough to reverse the ranking within the tested sizes, is not supported. Filtering is faster at every sampled size on both devices, so the campaign reports a narrowing advantage rather than a crossover. A failure of common-domain checks would invalidate the controlled interpretation. The absence of a crossover does not exclude one for other implementations, workloads, devices or sizes beyond 1536^2.

Morales-Hernández et al. (2020, author manuscript Sect. 5.1–5.3, pp. 14–24) explicitly compare wet/all execution, precision and output, including complete CPU/GPU/MPI/I/O time. Their size comparison changes 30 m to 10 m resolution. Chowdhury et al. (2023, Sect. 2.2.3 and 3.2) use CUB compaction after adaptive tree traversal and compare runtimes across refinement/tolerance choices. Chowdhury and Kesserwani (2025, Sect. 2.3, Table 1) report evolving cell counts, steps, adaptation/update effort and total cumulative wall-clock cost. These full texts establish that sparse updates, evolving-work metrics and management costs are precedents. Fixed discrete local fields and absolute selected work across dry-storage changes are not explicitly reported in those comparisons. Our control checks these quantities before interpreting the measured cost difference and then restores omitted execution. This does not claim those earlier controls were absent or make unique per-kernel attribution from an aggregate domain term. Sætra's available author draft (2013, Sect. 3–4, pp. 3–7, Fig. 5) explicitly compares full-domain execution, early exit, CUDPP-compacted sparse compute with unchanged underlying storage, and sparse memory with block-linear storage. Its refinement test holds a 100 m square physical domain fixed; its 4096^2 trajectory shows sparse memory crossing below early-exit performance at about 35 % wet cells. These are direct precedents for sparse scheduling and changing path preference. Exact synchronized timing boundaries and fixed local-state/ absolute-support controls across dry extensions are not explicitly reported. The consulted seven-page version says "This is a draft"; identity with the published chapter is not assumed. Locators and its hash are recorded in `LITERATURE_COMPARISON_20260914.md`.

Table S14: Scope of execution checks. A dash means not exported or not compared by that experiment, rather than a failed check.

Experiment	Saved public fields	Hidden residual/stage frames	Individual dt	Stage lists	Cumulative support
Grouped dry extension / new groups	Common crop, inverse shift, dry exterior; hashes and step counts	—	—; window step counts only	—	Across sizes and paths as specified
Earlier translation control	Nonwrapping public overlap	—	—	—	Window totals
Cell probes	Full compared frames	Six field/residual frames per fixed step	Fixed steps	Lists at both stages	Stage checks
Manual source ledger precision	Public fields	—	Recorded manual time/dt columns	—	Not the dry-extension support test
Same-device follow-up timing	Same-size path field hashes and window steps	—	—	—	Recorded execution counters
Core trajectory audit (S20)	Every step and saved window; full exterior checked	Six public/residual frames per adaptive step	Every actual dt	Both block sets and compacted lists, independent oracle	Both stage counts and window counters
Default preparation audit (S27)	All 12 F/C/Cf endpoints, matched to release windows	Six public/residual captures per adaptive step; hashes and zero exterior	Every actual dt	Both stages and compacted coordinates, independent dilation	Stage counts and release window counters
Thin-momentum flat endpoints (S31)	12 F/C/Cf trajectories	Six public/residual captures per step	Every actual dt	Both-stage lists/support	Stage checks
Stable-scan flat endpoints (S31)	12 matched trajectories	Six public/residual captures per step	Every actual dt	Both-stage lists/support	Stage checks
Original candidate Thacker 256^2 (S30/S33)	F/C fail against dense; cell agrees	First mismatch localized in six-field stage replay	Equal until first mismatch	Omitted block confirmed at first mismatch	Historical failure retained
Positive-depth selection, Thacker 256^2 (S33)	Original dense and corrected D/F/C release outputs agree	All 17,496 six-field captures per corrected path	Every actual dt	Isolated seed change	No performance claim
Default non-flat extension (S34)	15 D/F/C/Cf/R trajectories; release binding	17,496 six-field captures each; zero exterior	Every actual dt across sizes/paths	Independent dilation and compacted lists	645,007 cumulative block-stage updates, invariant across sizes

Core dry-domain results establish equality at the saved times and of the reported window steps/support. They do not establish equality of every hidden residual or every internal timestep. Cell-intervention stage checks do not extend that scope to the grouped dry-domain experiment. Snapshots are not complete restart checkpoints. The additional S20 audit establishes stronger checks for its eight specified trajectories; S27 adds 12 default preparation combinations. S31 adds separate 12-trajectory audits for stable-scan selection and thin-momentum flat endpoints. None of S20, S27 or the flat S31 audits supplies a matched dense stage comparison. They neither pass the dense omission gate nor supply missing stages for the full geometry matrix. S33 separately locates the original non-flat candidate failure and verifies its selection-only repair; S34 audits the default non-flat dry-extension application.

S17 Targeted Ritter front diagnostics

The unchanged T2 dense solver was run on 64/128/256/512/1024 cells with CFL 0.25 and caps 0.05/0.025/0.0125 s to t=3 s. All fifteen fields are retained. The finer grids were already CFL limited, so an explicitly subsequent, pre-run-committed addition used caps 0.003125/0.0015625 s at 256/512/1024 (six runs). These produce 960/1920 steps; the original 512/1024 runs had 464/940 steps. All strict 1 mm wet-set discrepancies remain unchanged.

Table S15: Ritter at cap 0.05 s. Edge errors compare the rightmost strict 1 mm wet-cell boundary with the exact cell-average reference wet set.

Cells	Depth L1 (m)	Discrepancy (%)	Edge error (m)	Edge error (cells)
64	0.014350	4.08	-3.125	-2
128	0.008204	6.19	-4.6875	-6
256	0.005060	6.70	-5.078125	-13
512	0.003698	7.22	-5.46875	-28
1024	0.003059	7.35	-5.56640625	-57

The continuous analytical threshold crossing is $x_\eta = 50{+}3(2\sqrt{9.81 \times 2}{-}3\sqrt{9.81\eta})$ m. At 1024 cells the error relative to this crossing is -5.5681 m, close to the cell-average comparison. Thus cell quantization alone cannot explain the observed lag. Time refinement changes depth norms but not this wet-set metric. The study does not isolate a single responsible spatial operator or prove nonconvergence under joint refinement of numerical thresholds.

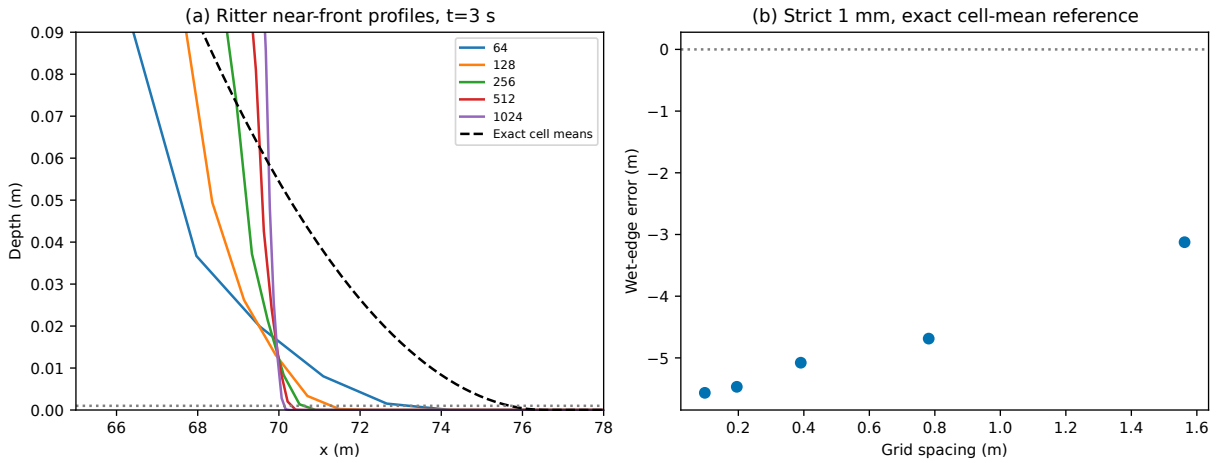


Figure 12: Local near-front depth profiles and wet-edge displacement at t=3 s for the baseline cap. The exact reference uses analytical cell integrals. The 1 mm reporting threshold is indicated; the error panel uses the exact cell-average wet set. This analytical reference has no external reference-grid uncertainty.

For paired withdrawal, an additional ANUGA 3.3.10 DE1_7 reference uses 256 base-grid squares per direction with four triangles each, CFL 0.5, cap 0.02 s and the original 1 mm source momentum closure. Protocol `034440c` precedes execution. Both this reference and the original 512 reference are projected as conserved fields before restriction to 64/256 and strict 1/10 mm reporting. All 160 time/grid/integrator/reporting comparisons are retained in `anuga_spatial/comparison.json` under the new campaign.

At 256^2, t=4.5 s and 1 mm, reference wet areas are 1.171875 and 1.3671875 m^2; their symmetric difference is 0.1953125 m^2 (14.29 % relative to the fine reference). Their full-domain depth L1 difference is 1.746e-6 m, compared with 0.001788 m over their wet union. RapidFlow T2 discrepancy changes from 70.83 % against the coarse reference to 75 % against the fine reference. Thus reference resolution sensitivity is measurable without explaining away the front error. The original 512 reference remains primary. Two reference resolutions do not establish spatial convergence or a certified error bound; source-closure and small-area sensitivity remain relevant.

S18 Separate-session cost location and telemetry

Two separately launched sessions on the primary platform each contain 84 trajectories: two development geometries, three sizes, two paths, one warmup, five randomized paired ordinary repeats and one separate profile. These are two short same-device sessions, not independent devices, dates or geometry populations. Sessions are not pooled. The same-size public-field hashes and window step counts agree across paths; hidden residuals are not exported.

Raw telemetry includes timestamp, temperature, P-state, SM/memory clocks, power, utilization and memory use every 500 ms. The initial nvidia-smi process snapshot and actual session start/end times are retained. No fixed-clock, controlled-temperature or guaranteed exclusive-host condition is claimed. The two sessions span 55–58 degrees Celsius with median SM clock 2520 MHz and a 2745 MHz maximum. Observed power spans approximately 23–115 W. Samples include initialization and downloads as well as advances; these ranges are not per-kernel thermal or power measurements. Ranges and medians are in `timing_summary.json`; ordinary paired intervals, list building/readback, preparation, transport, finalization, other event ranges and signed wall-minus-event-sum differences are in `timing_summary.md` under `experiments/results/linux4090_20260928/review/`.

All twelve session/geometry/size point ratios favour filtering. The four large-domain values are 0.948–0.962, and the four small-domain values 0.765–0.820, consistent with the endpoint and scan campaigns; within-session intervals use five paired repeats and do not measure long-term variability. The profiles retain path-dependent disturbance. Profile wall time exceeds the sum of reported event spans by 2.03–2.64 ms in these Linux sessions. This positive remainder does not make the spans undisturbed causal components. In particular, filtered transport includes preservation work while compacted preparation is separately recorded.

These instrumented spans indicate where compaction spends additional time. At 1536^2 the compacted path records a separate preparation span of 41.8 and 43.6 ms in session 0 for the continuous and fragmented inputs, plus 6.5 and 7.1 ms of list construction with readback, while its transport span falls to 5.1 and 8.6 ms against filtered transport of 45.9 and 48.3 ms. Filtered preservation is embedded in that transport span rather than recorded separately, so the two columns are not equivalent accounting categories. The pattern is consistent with the ordinary-time result that selection and preparation cost more than the dry-block transport they avoid, but event spans locate candidate costs without uniquely apportioning the ordinary-time difference. Section S21 supplies the event-envelope audit, mutually exclusive host partitions and an ordinary-time intervention.

S19 Isolated Ritter Froude-cap intervention

For Ritter's exact fan, $\mathrm{Fr} = 2(\sqrt{h_0/h} - 1)$. With $h_0 = 2$ m, Fr=10 occurs at 0.05556 m depth, at approximately 69.93 m when t=3 s. The continuous 1 mm location has Fr approximately 87.4. The default velocity cap therefore changes part of the fan well above the reporting threshold. This correspondence motivates an intervention; spatial coincidence alone is not treated as causal evidence.

The isolated builds replace only the cap at both Euler updates and RK finalization, with values 10, 100 and disabled. They retain all other numerical rules and record every clipping event: actual step/dt, stage, cell index, preclip depth, Froude number and signed removed momenta. Instrumentation is optional and its on/off final fields and dt sequences agree exactly. The cap-10 fields and step counts also agree with the original release module. All fifteen configurations finish with finite nonnegative depth and no stagnation. This is a bounded stability observation, not a general stability proof or a recommendation to disable limiting in applications.

Table S16: Ritter cap intervention, t=3 s, CFL 0.25. Cap100 and disabled are identical for all compared fields and dt sequences and have no clipping events. The last two rows use identical 1920-step sequences across all caps.

n	dt cap (s)	Fr cap	Depth L1 (m)	D1mm (%)	Edge error (m)
128	0.05	10	0.008204	6.19	-4.6875
128	0.05	100 / off	0.007872	4.12	-3.1250
256	0.05	10	0.005060	6.70	-5.0781
256	0.05	100 / off	0.004218	5.15	-3.9063
512	0.05	10	0.003698	7.22	-5.4688
512	0.05	100 / off	0.002391	5.15	-3.9063
1024	0.05	10	0.003059	7.35	-5.5664
1024	0.05	100 / off	0.001578	5.03	-3.8086
1024	0.0015625	10	0.003071	7.35	-5.5664
1024	0.0015625	100 / off	0.001648	5.16	-3.9063

At 1024 cells and the original cap, Fr10 produces 540800 cell-stage clipping events over 940 steps; raised/disabled variants take 969 steps. The diagnostic sum of removed hu contributions is 416.162 in stored discharge units before cell-area weighting. This sum counts both Euler

updates and finalization and is not a physical momentum loss integral. The smaller-step cap10 run records 1075328 events. Raw events retain the spatial and stage distribution and 1/10 mm comparisons; counts are not independent experimental samples.

Removing the cap reduces the 1024-cell edge lag by 1.76 m under adaptive CFL, and by 1.66 m with the same 1920 timesteps. Depth L1 approximately halves, yet roughly 3.8–3.9 m of lag remains. The cap is therefore a causal contributor to this error, but does not explain it completely. Reconstruction fallback, dry momentum closure and the flux discretization remain possible contributors to the residual. No new convergence or revised-default performance claim is made. Results are in `experiments/results/linux4090_20260928/mechanisms/`.

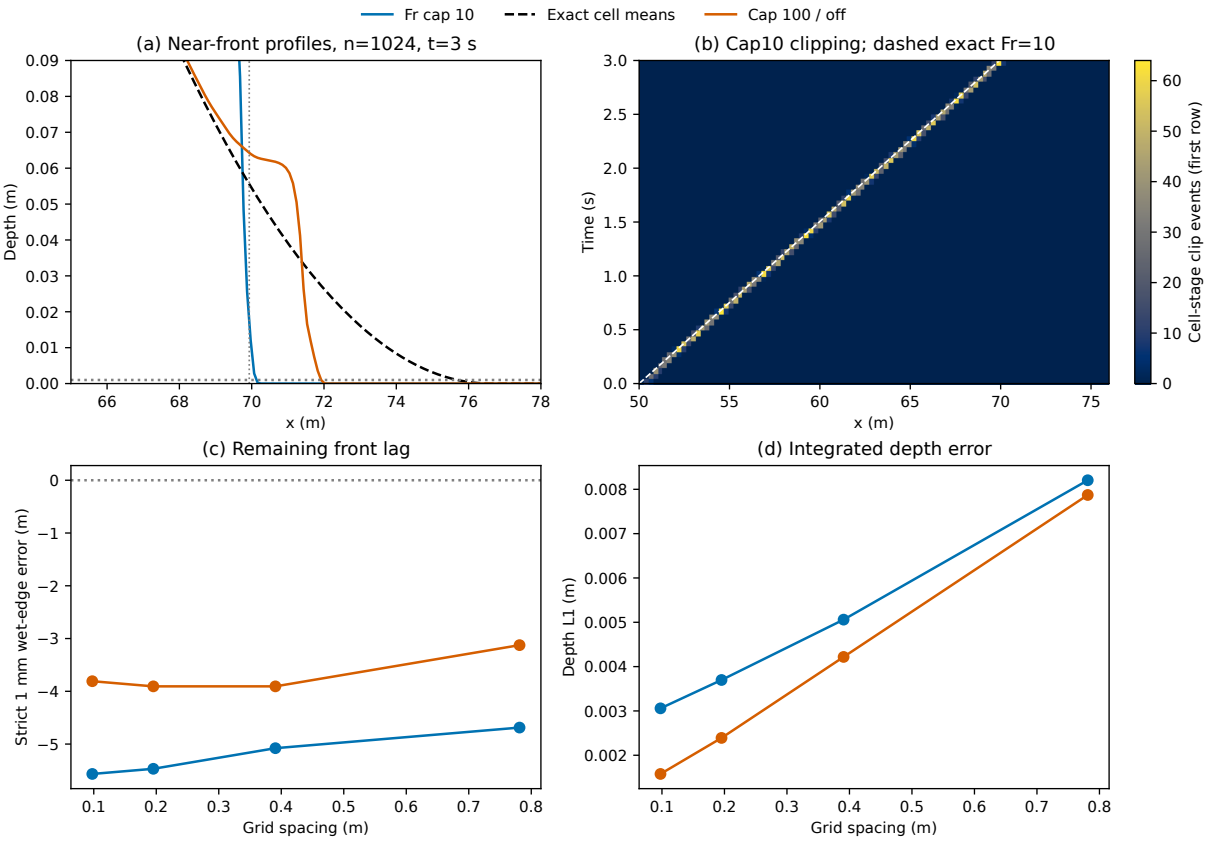


Figure 13: Isolated cap intervention: near-front profiles at n=1024, locations of actual cap10 clipping over the trajectory, and remaining front/depth errors across resolutions. The event histogram shows the first grid row, with both Euler stages and finalization counted; the dashed line is the exact Fr=10 locus. Final-step cap10 clipping at n=1024 occurs at cell centres 69.971–70.068 m, with preclip depths 2.69–14.24 mm. These are depths in the already modified solution, not the exact fan's 55.56 mm. Raised and disabled caps have identical states and do not clip in these runs.

S20 Core dry-domain complete trajectory audit

The audit uses dev_square_64 and dev_patch_8x8, offset0, n=512/1536, filtered and compacted paths, and the original continuing targets 0.25/1/2 s. The eight adaptive trajectories comprise 1524 steps, 9144 six-field frames and 3048 stage-support checks. Every step records stage zero, both Euler inputs and outputs and final RK state. Hashes cover all three public float32 fields and all three retained float32 residuals over the common 512-square region; all stored values outside it are explicitly checked for zero. Raw first-step and each window's

final-step frames are also retained.

Across paths and domain sizes, all corresponding frames, dt values and stage supports agree exactly. Continuous trajectories take 185 steps and fragmented trajectories 196. Each input's launch flags agree with an independently computed one-block dilation of inclusive float32 1 mm wet flags, and compacted coordinate sets agree with those flags. Nonzero residuals are present, so this does not reduce to a comparison of all-zero hidden fields. Compacted cumulative support equals the sum of recorded stage counts and the uninstrumented window counters. Saved fields and window steps agree with the original release runs.

Only host capture and bindings change; the complete multiset of 27 device function occurrences retains the original instruction words. These F/C-only checks do not compare either path with dense and therefore do not pass the complete-operator omission gate. They strengthen the representative cross-path and cross-size comparison while leaving the scope of the full grouped study as stated in Table S14. They do not establish equivalence for arbitrary source, boundary or geometry configurations.

S21 Timing envelope audit and restored dry-work intervention

The StageTimings collector uses consecutive events on the same compute stream, resets the event index for each step and accumulates each adjacent interval once. Continuing-window analysis differences cumulative counters. There is no explicit nested range or repeated summation of windows. An isolated audit adds the first-to-last event elapsed time and enclosing host duration to each step. All 36 event profiles match their sum to the event envelope within 0.000001 ms over the trajectory. On these platforms the envelope also lies inside its enclosing host step duration, by 0.82 to 1.07 ms. The clock checks for these Linux measurements are reported in S23. Event durations are still not converted into ordinary-time shares or treated as calibrated hardware timelines.

A second isolated build replaces CUDA events with stream synchronization and host-clock boundaries at each labelled stage. Those intervals are mutually exclusive on one clock. Their sums equal the first-to-last host partition within 1e-6 ms; complete advance exceeds the partition by 0.130–0.200 ms across all 36 profiles. Synchronizing each range changes launch overlap and cost, so this is an instrumented location check, not an undisturbed budget. Full per-process components, envelopes and ordinary/profile ratios are in `experiments/results/linux4090_20260928/mechanisms/summary.json`.

Table S17: Mutually exclusive synchronized host-stage durations (CLOCK_MONOTONIC ms), medians of three instrumented trajectories. Transport includes filtered preservation; other includes positivity, friction, scheduling and reductions. These are perturbed profiles, not ordinary-time shares. The raw records include all three sizes; endpoint medians show the growth in work.

Input / n / path	List/readback	Prepare	Transport	Finalize	Other
Continuous / 512 / F	0	0	11.29	5.61	17.88
Continuous / 512 / C	6.04	8.85	6.27	5.94	17.25
Continuous / 1536 / F	0	0	47.17	38.15	35.36
Continuous / 1536 / C	6.08	42.71	6.30	38.53	31.26
Fragmented / 512 / F	0	0	12.91	6.25	20.16
Fragmented / 512 / C	6.40	9.46	9.08	6.25	18.70
Fragmented / 1536 / F	0	0	49.53	40.33	36.89
Fragmented / 1536 / C	6.61	44.32	9.58	39.89	34.28

At 1536^2 the compacted preparation partition is 42.7 and 44.3 ms for the two inputs, against filtered transport partitions of 47.2 and 49.5 ms. Compacted transport drops to 6.3 and 9.6 ms,

and list construction with readback adds about 6.1 and 6.6 ms. The preparation and selection cost therefore accounts for most of what compaction saves in transport, which is consistent with the ordinary-time result that compaction stays slower. These partitions are perturbed by their own synchronization and are not an undisturbed budget.

The primary mechanism test instead changes ordinary work. Let F be filtered, C compacted, and R compacted with the original full interior/boundary coordinate arrays and counts restored after list construction and preparation. R retains selection, host count readback, six state/residual copies, theta zeroing and wet-flag zeroing. Its positivity and transport launch the original filtered kernels across all blocks, restoring dry-block theta reset, six-field copying, wet reduction and zero speed outputs. Finalization is unchanged. The selected list counters still report the compacted support; they are not R's actual full-grid launch count. There is no hidden state/list probe in the timing module. All 27 device-function occurrences retain the original instruction words.

Three serial processes each measure two reused development geometries at three sizes, one warmup and seven randomized paired ordinary repeats of F/C/R, followed by separate event and host profiles of F/C. The 504 trajectories all preserve same-size fields and steps; C/R selected support agrees. The ordinary contrast R-C measures restored work in the presence of compacted preparation, including redundant copies and potential cache interactions. R-F measures the added management/preparation context. Each sample obeys F-C=(R-C)-(R-F); medians of the contrasts need not be exactly additive.

Table S18: Paired ordinary-time intervention differences in CLOCK_MONOTONIC ms, with conditional 95 % hierarchical bootstrap intervals over three processes and seven repeats. Negative F-C means filtering is faster. Ordinary path times are process-median point estimates.

Input	n	F ms	C ms	R ms	F/C [95 %]	F-C	R-C	R-F
Continuous	512	28.28	34.59	40.50	0.811 [0.779, 0.866]	-6.6 [-7.6, -4.3]	6.8 [5.9, 7.7]	13.6 [11.9, 13.9]
Continuous	1024	46.60	52.04	62.97	0.880 [0.833, 0.891]	-6.5 [-8.9, -6.0]	9.3 [8.3, 11.2]	15.7 [14.9, 16.3]
Continuous	1536	109.74	116.81	152.65	0.955 [0.943, 0.960]	-5.3 [-6.6, -4.7]	36.8 [34.6, 37.5]	42.6 [41.2, 43.8]
Fragmented	512	30.42	40.30	44.69	0.774 [0.759, 0.804]	-9.0 [-10.0, -7.2]	5.4 [1.8, 5.8]	14.4 [12.2, 14.9]
Fragmented	1024	50.00	57.75	66.74	0.866 [0.862, 0.874]	-7.7 [-8.1, -7.1]	9.2 [8.9, 9.8]	16.8 [16.3, 17.8]
Fragmented	1536	117.51	122.92	162.73	0.951 [0.945, 0.957]	-6.0 [-6.8, -5.3]	38.7 [37.6, 41.2]	44.5 [43.8, 47.8]

Main Figure 5 displays the dual-clock endpoint measurements of the same construction on both platforms; this table gives the six-size mechanism session. The two agree on every direction and magnitude to within session variation. The causal interpretation depends on R's controlled construction, not the algebraic identity alone. R-F includes the added management/preparation context and its interactions, rather than measuring pure management overhead.

F-C is negative at all three sizes, so filtering is faster throughout, while R-C grows from 5.4–6.8 ms at 512^2 to 36.8–38.7 ms at 1536^2. Restoring the omitted dry work therefore recovers a rapidly growing amount of execution that the compacted implementation does not convert into a net advantage in this range.

The archived telemetry covers initialization and downloads as well as advances. No competing compute process is listed at the session snapshots. Clocks and cooling are not controlled, the container does not guarantee exclusive use of the physical machine, and the sessions remain

on one day and one device. The mechanism is supported for these configurations; it does not establish portability or a hardware bandwidth bottleneck.

S22 Remaining front bias and numerical-policy connection

After disabling Froude clipping, three single-factor interventions separate post-update momentum clearing, flux dry gates and near-dry reconstruction fallback. The common Ritter control uses n256/1024, CFL 0.25, T2 dense, cap 0.0015625 s and the unchanged exact cell-average reference. Reporting remains strict 1/10 mm. All cases complete with finite nonnegative depth. At n1024, the cap-off control reproduces the earlier fields and all 1920 timesteps.

The momentum intervention retains hu/hv below 1 mm at both Euler updates and RK finalization but keeps the 1e-7 m positivity zeroing branch. The transport intervention lowers only gates in flux_ops.cuh from 1 mm to 1e-7 m, including hydrostatic momentum reconstruction, face velocities and HLL dry classification. It leaves block wet flags, CFL gates, friction and post-update clearing fixed. The reconstruction intervention lowers only the 1 cm fallback gate to 1e-7 m; genuinely dry stencils still fall back. None changes the reporting threshold.

Table S19: Isolated cap-off factors at n1024, t=3 s. Rows use the same 1920-step sequence. The combined row was triggered by the disclosed residual error criterion after observing the single factors.

Factor	Depth L1 (m)	Strict 1 mm edge error (m)
Cap off, other defaults	0.00164814	-3.90625
Retain thin-layer momentum	0.00098643	-1.26953
Lower transport gates	0.00164793	-3.90625
Lower reconstruction fallback gate	0.00167015	-4.00391
Momentum plus lower transport gates	0.00098506	-1.36719

The combined policy has edge errors -1.9531/-1.7578/-1.3672 m at n256/512/1024. At n1024, halving the cap again produces 3840 steps without changing the momentum-only or combined wet-edge positions; their depth L1 values become 0.00098555 and 0.00098430 m. Post-update momentum clearing is therefore a substantial cause of the cap-off residual in this experiment. Lower transport gates alone do not fix it, and lowering reconstruction fallback is not an improvement. The combination does not improve edge placement over the momentum-only control. Residual bias remains, so this is not a joint space/time/threshold convergence result or a validated new numerical default.

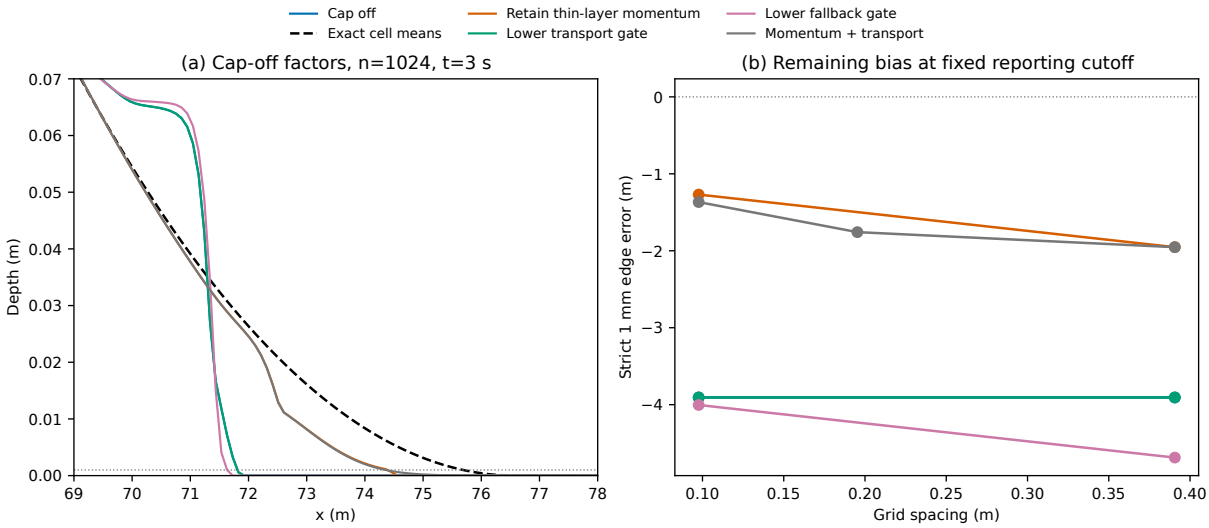


Figure 14: Cap-off numerical factors with unchanged reporting thresholds. The left panel shows n1024 profiles; the right shows remaining edge errors at the sampled resolutions. Thin-layer momentum retention explains much of the residual, while transport and reconstruction factors are retained with their neutral or adverse outcomes. Lines connect tested resolutions and do not establish an asymptotic convergence rate.

The core timing inputs are not exempt from clipping: default Fr10 records 264 cell-stage events in the continuous input and 2377 in the fragmented input through 2 s at n512. Recording on/off fields and dt sequences agree with each other and saved fields/window steps agree with the original release. Absolute hu/hv removals sum to 0.01059/0.01059 for the continuous case and 0.16558/0.16243 for the fragmented case before area weighting; signed sums can cancel. These are operator diagnostics, not physical momentum budgets. The observation triggers the Fr100 endpoint comparison. A separate follow-up uses the effective single-factor momentum-preserving cap-off policy; it was selected after numerical diagnosis and frozen before its timing runs.

Each endpoint comparison retains the original two development groups, 512/1536 domains, filtered/compacted paths and continuing 0.25/1/2 s targets. Three serial processes each use one warmup and five randomized paired repeats of default and candidate. No probe/profile code is included in candidate release modules. Every policy's common fields, window steps and cumulative stage support agree across domains and paths as applicable; exterior public state remains zero. Different policies are allowed to have different fields, steps and support. The earlier full hidden-state audit is not repeated here.

Table S20: Policy connection, paired filtered/compacted ordinary-time ratios on CLOCK_MONOTONIC_RAW. Intervals are conditional 95 % hierarchical bootstrap intervals over three processes and five repeats. Default is sampled separately within each phase; phases are not pooled.

Policy	Input	n512 ratio [95 %]	n1536 ratio [95 %]	Steps	Cumulative block support
Default (Fr100 session)	Continuous	0.812 [0.743, 0.820]	0.956 [0.947, 0.977]	185	40108
Default (Fr100 session)	Fragmented	0.820 [0.775, 0.830]	0.952 [0.947, 0.955]	196	115255

Table S20 continued.

Policy	Input	n512 ratio [95 %]	n1536 ratio [95 %]	Steps	Cumulative block support
Fr100	Continuous	0.811 [0.763, 0.817]	0.958 [0.944, 0.963]	185	40108
Fr100	Fragmented	0.813 [0.775, 0.818]	0.950 [0.946, 0.955]	196	115255
Default (momentum session)	Continuous	0.801 [0.700, 0.866]	0.956 [0.942, 0.962]	185	40108
Default (momentum session)	Fragmented	0.807 [0.796, 0.821]	0.952 [0.949, 0.956]	196	115255
Cap off + retain momentum	Continuous	0.777 [0.722, 0.832]	0.959 [0.938, 0.969]	184	39828
Cap off + retain momentum	Fragmented	0.793 [0.733, 0.801]	0.955 [0.944, 0.960]	196	115352

Steps and support in Table S20 agree across sizes within each row. The default continuous input takes 185 steps and 40108 block updates; default fragmented input takes 196 steps and 115255 updates. The momentum-retention policy changes the continuous case to 184 steps and 39828 updates, so the improved-front policy does alter the evolution relevant to scheduling. All sixteen rows nevertheless favour filtering at both sizes, and the compaction deficit narrows with stored extent in every row, so the measured execution behaviour does not depend on the default clipping policy. This is a limited connection between numerical and performance evidence, not full source or urban validation of either candidate. All 576 timing trajectories and 13 Ritter factor/interaction runs remain archived under `experiments/results/linux4090_20260928/policy/`.

S23 Independent clock metrology on both platforms

A standalone CUDA program includes no RapidFlow source. It brackets three consecutive events on one stream with host steady_clock and CLOCK_MONOTONIC_RAW, explicitly synchronizing before measurement and after the final event. Events are either created before each measurement or reused only after synchronization; both default and explicit nonblocking streams are tested. Allocations, event creation and result downloads are outside the measured interval. Two dependent FMA kernels use 0/4096/65536/1048576/16777216 iterations, or two groups of 16-MiB device copies use 1/16/128 repetitions. Each condition has one warmup and seven retained repeats, giving 224 samples per platform. Device globaltimer intervals are also recorded for the arithmetic kernels. Millisecond and nanosecond conversions, output checksums, source/binary hashes and raw rows are retained. The probes ran on both platforms before the timed campaigns.

Both Linux platforms show consistent host clocks. For the longest arithmetic workload, the median event/steady ratio is 1.0003 on the primary platform and 0.9999 on the second, and the raw/steady ratio is 1.0005 and 0.99999 respectively. Short measurements still show ratios well below one because the host interval contains launch and synchronization work that the event envelope excludes; the relative gap generally decreases as the arithmetic workload grows. At the largest arithmetic workload, median event/steady ratios are within about 0.04 % of one. An independent CPU-only program reads adjtimex with modes=0, without changing any time setting, and compares monotonic, raw, boottime, realtime and C++ clocks around 1/10/100/1000 ms sleeps. Both platforms report `tick=10000` microseconds, the standard value. Raw over monotonic duration ratios are 1.00049 median on the primary platform and 0.999992 median on the second, so a nominal 1000 ms monotonic interval corresponds to 1000.5 and

1000.0 ms raw respectively.

Neither host clock is traceably calibrated to an external standard. Consistency between clocks is a diagnostic, not an external calibration.

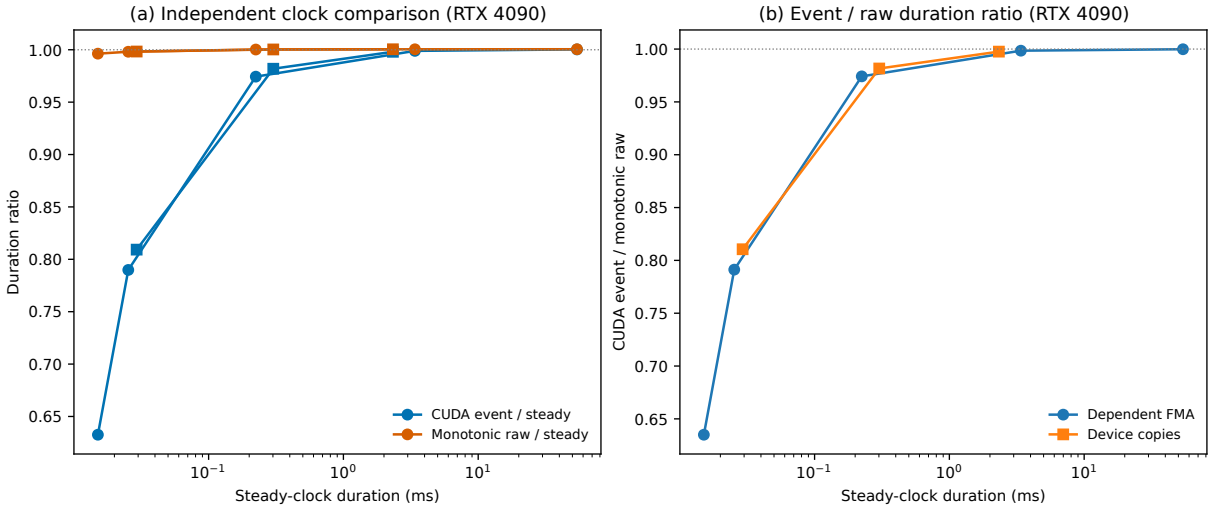


Figure 15: RTX 4090 standalone metrology for fresh events on an explicit stream; the archive also includes reused events and the default stream. CUDA-event, raw-host and steady-clock durations agree closely once the arithmetic workload is large. Short measurements retain event/host launch and synchronization overhead, which is why their ratios fall below one. Read-only time-adjustment metadata reports the standard tick; the second-platform metrology is archived separately.

The endpoint, scan and policy experiments record `perf_counter` and the monotonic raw clock around the same synchronized advance. Their paired ratios agree to within 0.00011 for F/C/R endpoints and 0.0015 for policy comparisons. The baseline/grouped campaigns retain single-clock monotonic timings; they must not be described as having the later dual-clock protocol retroactively. A constant common clock scale cancels from a paired ratio, and agreement between two clocks on one host does not turn conditional intervals into long-term robustness estimates.

Across the policy comparisons, raw/monotonic whole-trajectory duration ratios span 0.99806 to 1.00044, and the largest difference between the two clocks' individual paired F/C ratios is 0.00147. Both clocks give the same ranking at every size; their raw paired intervals are reported in Table S20. Read-only clock metadata before and after the timed processes report `tick=10000` on both platforms.

Instrumented event channels remain separate from host durations. They include Supplement Table S8, the event list-build panel in Figure 11 (asset fig09), grouped execution_costs event summaries, and the S18 and S21 event profiles. Their units are CUDA-event milliseconds, and they are not mixed with host durations as calibrated fractions or converted into bandwidth estimates. The synchronized host partitions in Table S17 are internally exclusive but instrumented, so they perturb the cost they locate.

S24 Dual-clock default F/C/R endpoints on both platforms

The protocol and runner were hash-bound before sampling on each platform. Three serial processes use the continuous and fragmented development inputs at $512^2/1024^2/1536^2$, offset zero, default numerics, and continuing 0.25/1/2 s targets. Every geometry has one warmup

and seven ordinary repeats of all nine size/F/C/R variants in random order. The baseline and instruction-audited R module are used without numerical changes; no profiles are included. All 432 trajectories and 1296 saved windows per platform pass common-field/window-step, C/R support, finite/nonnegative, mass, zero-exterior and boundary checks. Representative common states are archived. Hidden-state auditing remains the separate S20 experiment. Both clocks enclose the same synchronized native advance, with setup, download and metric reads outside.

Table S21: Same-sample dual-clock contrasts on the primary platform; entries are conditional 95 % bootstrap intervals over three processes and seven repeats. RAW is CLOCK_MONOTONIC_RAW; M is CLOCK_MONOTONIC. Difference units follow the named clock. Values below one favour filtering.

Input / n	Clock	F/C [95 %]	F-C ms [95 %]	R-C ms [95 %]	R-F ms [95 %]
Continuous / 512	M	0.811 [0.756,0.821]	-6.1 [-9.1,-5.7]	6.5 [6.3,7.4]	12.4 [12.2,16.2]
Continuous / 512	RAW	0.811 [0.756,0.821]	-6.1 [-9.0,-5.7]	6.5 [6.3,7.4]	12.4 [12.2,16.2]
Continuous / 1024	M	0.877 [0.867,0.880]	-6.2 [-7.1,-6.0]	8.8 [8.7,9.8]	15.0 [14.7,16.8]
Continuous / 1024	RAW	0.877 [0.867,0.880]	-6.2 [-7.1,-6.0]	8.8 [8.7,9.8]	15.0 [14.7,16.8]
Continuous / 1536	M	0.961 [0.956,0.963]	-4.4 [-5.2,-4.2]	36.9 [36.6,37.2]	41.5 [40.8,42.4]
Continuous / 1536	RAW	0.961 [0.956,0.963]	-4.4 [-5.2,-4.2]	36.9 [36.6,37.2]	41.5 [40.8,42.4]
Fragmented / 512	M	0.805 [0.737,0.815]	-7.3 [-11.4,-6.8]	5.0 [4.9,5.2]	12.2 [11.9,16.4]
Fragmented / 512	RAW	0.805 [0.737,0.815]	-7.3 [-11.4,-6.8]	5.0 [4.9,5.2]	12.2 [11.9,16.4]
Fragmented / 1024	M	0.874 [0.854,0.878]	-7.1 [-8.7,-6.9]	9.4 [9.3,9.5]	16.4 [16.2,18.3]
Fragmented / 1024	RAW	0.874 [0.854,0.878]	-7.1 [-8.7,-6.9]	9.4 [9.3,9.5]	16.4 [16.2,18.3]
Fragmented / 1536	M	0.954 [0.952,0.960]	-5.6 [-6.0,-5.0]	38.7 [38.1,38.8]	44.3 [43.5,44.8]
Fragmented / 1536	RAW	0.954 [0.952,0.960]	-5.6 [-6.0,-5.0]	38.7 [38.1,38.8]	44.3 [43.5,44.8]

The two clocks agree to the displayed precision in every ratio. Individual RAW/M duration ratios span 0.999758 to 0.999962, and the maximum paired F/C ratio difference between clocks is 0.000109. Read-only `adjtimex` checks before and after every process report `tick=10000`. GPU telemetry is sampled every 500 ms; no fixed clocks or exclusive-host guarantee applies.

F/C is below one at all three sizes, so filtering is faster throughout, and the ratio rises with stored extent. R-C and R-F both grow strongly with size, from 5.0–6.5 ms and 12.2–12.4 ms at 512^2 to 36.9–38.7 ms and 41.5–44.3 ms at 1536^2. F-C remains negative at every size, between -7.3 and -4.4 ms. Restoring the omitted dry work therefore recovers an increasingly large amount of execution, without the compacted path becoming faster overall in this range.

Table S22: The same protocol on the second platform, reported on its raw clock. Table 8 additionally reports F-C differences recovered from the frozen raw paired durations. The CPU reanalysis in `experiments/analysis/review_endpoint_differences.py` uses 10,000 process/repeat resamples, seed 2026092801, and retains input digests in `experiments/results/mock_review_20260928/endpoint_differences.json`. These within-platform differences do not identify an architecture effect between hosts.

Input / n	F/C [95 %]	R-C ms [95 %]	R-F ms [95 %]
Continuous / 512	0.710 [0.699,0.715]	4.4 [3.9,4.7]	13.8 [13.7,14.1]
Continuous / 1024	0.784 [0.781,0.786]	7.9 [7.5,8.3]	17.8 [17.7,18.5]
Continuous / 1536	0.962 [0.958,0.967]	22.6 [22.3,22.9]	25.6 [25.4,25.8]
Fragmented / 512	0.740 [0.736,0.743]	4.8 [4.7,5.8]	14.1 [14.0,14.4]
Fragmented / 1024	0.779 [0.774,0.782]	7.5 [7.4,7.9]	19.2 [19.1,19.4]
Fragmented / 1536	0.951 [0.949,0.953]	23.3 [22.8,23.5]	27.4 [27.3,27.8]

On the second platform, RAW/monotonic ratios span 0.999704 to 0.999976. Every F/C interval again lies below one and rises with stored extent, and both restoration contrasts again grow with size. The second platform's own summary records that no primary-platform coefficients were transferred or refitted.

The frozen cost model is also evaluated without refitting or rescaling on the primary-platform F/C trajectories. Mean and maximum APE are 9.17 % and 16.27 % against monotonic seconds, and 9.17 % and 16.28 % against raw seconds. This is a bounded session-transfer diagnostic on already-used geometries, not an independent geometry test and not a cross-platform validation.

S25 Static initial geometry as a cost-model baseline

This sensitivity was proposed after observing both original and subsequent evaluation groups. All new coefficients use only the original four development groups. Each group's sizes, offsets, windows and repeats remain together; no evaluation group is used for fitting, tuning or exclusions. Initial geometry is reconstructed deterministically and matched to the archived initial support before fitting. S_0 is fixed at t=0 for every window, rather than refreshed at window starts. The equation for every path is $\widehat{t} = aK + cNK + bX$, where X is KW_0, $2KS_0$ or actual A. For dense $A = 2NK$ and $S_0 = N$; the equivalent domain coefficient below absorbs the redundant support term. Domain/steps uses $b = 0$.

Table S23: Full per-path empirical coefficients on CLOCK_MONOTONIC. a is seconds per step; c is seconds per stored-cell step; b is seconds per selected-cell update (or per wet-cell step for KW_0). Displayed values are rounded; exact coefficients, rank, conditioning and errors are archived. These terms do not isolate kernel costs.

Model / X	Path	a	c	b
Domain/steps / none	Dense	8.833549575e-05	3.852418177e-10	0.000000000e+00
Domain/steps / none	Filtered	9.444292840e-05	1.810666893e-10	0.000000000e+00
Domain/steps / none	Compacted	1.334427500e-04	1.806992897e-10	0.000000000e+00
Domain/steps / none	Cell	1.370542667e-04	2.124846788e-10	0.000000000e+00
Initial wet / KW_0	Dense	8.833549575e-05	3.852418177e-10	0.000000000e+00

Table S23 continued.

Model / X	Path	a	c	b
Initial wet / KW_0	Filtered	9.417527926e-05	1.810657028e-10	4.202595244e-11
Initial wet / KW_0	Compacted	1.334427500e-04	1.806992897e-10	0.000000000e+00
Initial wet / KW_0	Cell	1.370542667e-04	2.124846788e-10	0.000000000e+00
Initial support / $2KS_0$	Dense	8.833549575e-05	3.852418177e-10	0.000000000e+00
Initial support / $2KS_0$	Filtered	8.438552073e-05	1.809274840e-10	1.775502821e-10
Initial support / $2KS_0$	Compacted	1.202742077e-04	1.805522914e-10	2.322849557e-10
Initial support / $2KS_0$	Cell	1.364315366e-04	2.124791741e-10	3.829146376e-11
Evolving support / A	Dense	8.833549575e-05	3.852418177e-10	0.000000000e+00
Evolving support / A	Filtered	7.996059572e-05	1.808705638e-10	1.508628875e-10
Evolving support / A	Compacted	1.161794384e-04	1.805137319e-10	1.792166794e-10
Evolving support / A	Cell	1.263289701e-04	2.123293598e-10	2.015157965e-10

Table S24: Static-support sensitivity error distributions on observed evaluation sets. Mean/median/P90/max refer to complete trajectory APE (%). Each row contains four paths; original/subsequent sets have four/two geometries.

Set / model	Mean	Median	P90	Maximum	Non-best / cases
Original / domain-steps	7.67	8.14	12.65	16.95	0/36
Original / initial wet	7.67	8.17	12.61	16.91	0/36
Original / initial support	7.59	7.39	12.90	14.24	0/36
Original / evolving support	7.35	6.74	13.30	15.27	0/36
Subsequent / domain-steps	8.60	8.55	13.80	19.46	0/18
Subsequent / initial wet	8.61	8.55	13.77	19.46	0/18
Subsequent / initial support	8.61	9.26	14.10	16.02	0/18
Subsequent / evolving support	8.48	9.01	13.27	14.96	0/18

Actual K is required by all models, so even initial-geometry models are retrospective. Against initial support, evolving support changes mean APE from 7.59 to 7.35 % on original groups and from 8.61 to 8.48 % on subsequent groups. These modest mean gains do not establish uniformly smaller errors or improved path selection. All coefficients and individual predictions belong to the Linux primary campaign; the second device did not replay these models.

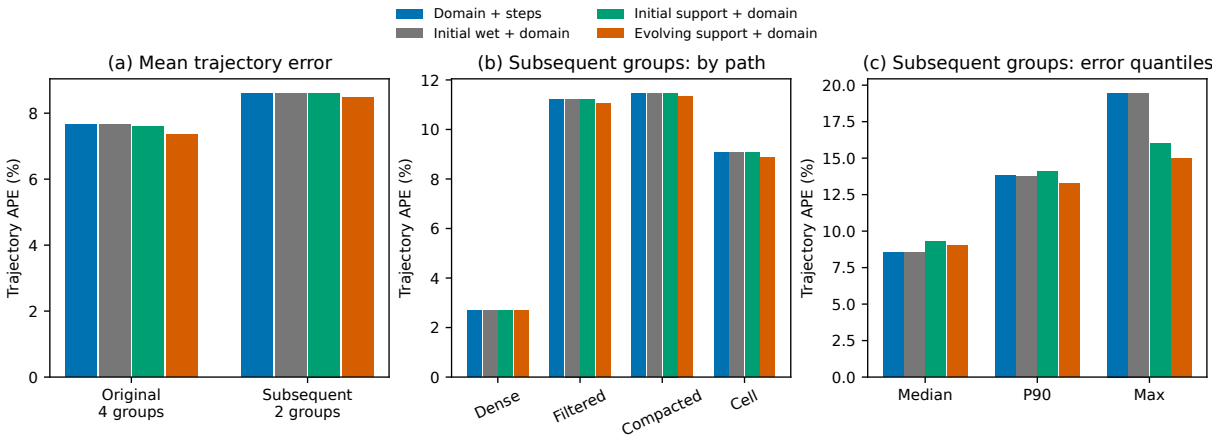


Figure 16: Post-hoc initial-geometry sensitivity: domain/steps, initial wet count, initial selected support and evolving support models share a domain term and development-only fitting. Panels show pooled trajectory mean APE, subsequent-group path means and subsequent-group error quantiles. The 18 configurations derive from two geometries. No extra best-path choices follow from the evolving support. The earlier domain/steps versus full-model comparison was frozen before these two groups; the added static comparisons were proposed afterward. Bars and quantiles are descriptive, not intervals.

S26 Ancillary inputs and execution evidence

Detailed protocols and supporting results are consolidated here so that the main text follows the controlled dry-domain evidence. All historical host absolute times use CLOCK_MONOTONIC; event spans use CUDA events.

S26.1 Analytical, source and urban inputs

The numerical problems probe complementary aspects of wetting, drying and boundary response (Table S25).

Table S25: Numerical problems and the aspect of the solution they probe.

Problem	Comparison	Purpose
Ritter dam break	Analytical cell averages, 64–1024 cells	Separate field convergence from wet-region discrepancy
Thacker oscillation	Analytical quarter-period state	Moving shoreline over non-flat topography
Lake and smooth channels	Steady analytical states	Balance and prescribed-boundary response
Circular expansion and paired withdrawal	Fixed independent triangular-mesh solutions	Two-dimensional wetting, retreat and reference sensitivity
Merewether	Five prescribed flood marks	Restricted real-event spatial comparison

Analytical tests include the Ritter dam break at 64, 128 and 256 cells in x using the SWASHES solution (Delestre et al., 2013), Thacker (1981) oscillations at a quarter-period with 128 cells, a reflective static lake, and smooth subcritical channels with steady and perturbed initial conditions (MacDonald et al., 1997). Both integrators are used except in the static-lake test. Finite-volume reference values are compared with the conserved cell averages. These checks do

not establish long-time Thacker convergence, hydraulic-jump pointwise accuracy or supercritical boundary validation. The theoretical splitting discussion in main Sect. 2 is not a measured temporal-convergence study in this campaign.

Circular expansion and paired-source withdrawal/restart are evaluated on 64^2 and 256^2 grids, both integrators and all four execution paths, with development controls. The circle uses a 40×40 m flat domain, reflecting walls, radius 5 m, 2 m inner depth, dry exterior, zero momentum and zero friction. Exact geometric integrals initialize Cartesian cell averages. CFL is 0.25, the step cap is 0.05 s and outputs occur every 0.25 s until 1 s.

The paired-source domain is also flat, reflective and initially dry. Aligned rectangular patches supply 3.125 m^3 during 0–2 s, stop during 2–4 s, request 112.5 m^3 of unequal withdrawals during 4–6 s, stop during 6–8 s and restart supply during 8–10 s. No state is reset. CFL is 0.25, the step cap is 0.02 s and outputs occur every 0.5 s. Mass-only sinks retain momentum while wet and clear it below their stated dry threshold.

Source coordinates are metres in $[0, 40]^2$. Paired supply rectangles are $[15, 17.5) \times [18.75, 21.25)$ and $[22.5, 25) \times [18.75, 21.25)$, each at 0.125 m s^{-1}. The single-patch control uses $[17.5, 22.5) \times [18.75, 21.25)$ at the same rate. During withdrawal, $[10, 20) \times [12.5, 27.5)$ requests 0.125 m s^{-1} and $[20, 30) \times [12.5, 27.5)$ requests 0.25 m s^{-1}. These are requested rates before exhaustion constraints; other cells have zero prescribed mass rate. The phase/output protocol is also recorded in `experiments/protocols/sources.json`.

Independent finite-resolution solutions use ANUGA 3.3.10 with the DE1_7 algorithm (ANUGA developers, n.d.). The same fixed reference fields are used for every RapidFlow execution path. The circular reference uses 768 squares per direction with four triangles per square and CFL 0.25. Three paired-source references use 512 squares per direction: a standard CFL 0.5 run, a CFL 0.25 run with halved time step cap, and a source momentum-clearance control at 0.5 mm instead of 1 mm. The independent constrained mass operator implements the same prescribed source model. ANUGA's native velocity-preserving extraction would be a different model. These finite references retain resolution and closure uncertainty; no exact front-error bound is claimed.

A separate development diagnostic uses host-selected float32 time steps to retain requested, blocked, applied and clipped source volumes. It starts from the same physical source case but has its own step sequence. Its budget failures must not be attributed to the release controller without a separate comparison.

The Merewether workload uses public terrain from the Water Research Laboratory (WRL), 57 building polygons, roads and five field flood marks distributed with the ANUGA benchmark (Geoscience Australia, n.d.) from the June 2007 Pasha Bulker flood benchmark. The 321×416 m domain is represented by 321×416 and 161×208 cells. Seventy-three edge NoData raster cells are filled from nearest valid terrain before bilinear sampling at model cell centres. Buildings add 3 m, and road/non-road Manning coefficients are 0.02/0.04. A constant 19.7 m^3/s enters a radius-10 m disk for 1000 s from dry initial conditions, using cell overlaps with a 1024-sided disk approximation and normalisation. South/west walls reflect, and north/east mask interfaces allow outward flow. The inlet centre is (382265, 6354280) m in the supplied coordinates, or (15,15) m from the southwest corner. At an outflow mask interface the first-order exterior state copies local depth, momentum and bed; an inward mass flux sets all three flux components to zero. Outward flux retains donor positivity limiting. This closure differs from the prescribed characteristic channel boundary in Sect. S15. This is a Cartesian peak-flow adaptation, not a reconstruction of the complete event hydrograph or the original ANUGA inlet/outlet implementation.

Observation cells are chosen before simulation as the nearest valid nonbuilding cell, without searching for a favourable wet cell afterward. Both containing-cell and fixed-cell results are retained. TUFLOW values supplied with the benchmark are a separate model comparison. Five marks without uncertainty estimates or time series cannot validate velocities, arrival histories or full-event dynamics.

S26.2 Matrix and ancillary timing protocols

The main matrix crosses six initial wet fractions (1, 5, 10, 25, 50 and 100 %), three fixed geometries (centred rectangle, eight bands and random aligned 8×8 patches) and T1/T2 on 2048×512 cells. Initial depth is 1 m, momentum is zero, the bed is flat, walls reflect and Manning is 0.03. Spacing is 0.125 m, CFL is 0.4, the step cap is 0.05 s and the horizon is 0.2 s. Geometry seed 20260911 is fixed; each requested fraction contains round(Np) wet cells. The fully wet labels share their physical input. Four fresh sequential processes randomise case/mode order, with one warmup, three paired ordinary repeats and one separate profile per mode/process. Repeated processes do not add spatial realizations or GPUs.

Four source timing cases use single-patch 256^2/T1, paired-patch 512^2/T2, and uniform rain/withdrawal/restart at 256^2/T1 and 512^2/T2. The horizon is 10 s, CFL 0.25 and cap 0.02 s. Source-upload plus synchronized advancement defines complete source cost; downloads and metric reads are reported separately. The two urban grids use horizon 1000 s, CFL 0.25 and cap 1 s, with two ordinary repeats per mode/process plus warmup and separate profiling.

The alignment experiment shifts each 1 % wet input by seven cells without wrapping. Bands and a fragmented case with a dry frame continue through 0.2, 1 and 2 s; window timings are 0–0.2, 0.2–1 and 1–2 s. All position/path variants are interleaved within each repetition. Saved conserved fields are compared after undoing the translation on the common domain. Later fragmented wall interactions limit causal interpretation. Initial and stage-weighted selected support are reported separately.

The copy/fused intervention uses continuous 1 % and fully wet inputs, both integrators and the two indexed paths. It combines state preparation into one kernel while holding the numerical update and arithmetic fixed within each integrator. Dense and filtered controls remain present. It does not isolate the friction-coefficient difference between T1 and T2 or establish a general optimal cell-indexed implementation.

S26.3 Expansion, withdrawal and reference sensitivity

Figure 17 shows the time-resolved front diagnostics underlying main Sect. 5.3. Reference/reference timestep and closure comparisons are defined in S6; the separate reference-grid comparison is in S17. These finite-reference checks provide sensitivity information, not exact-solution error bounds.

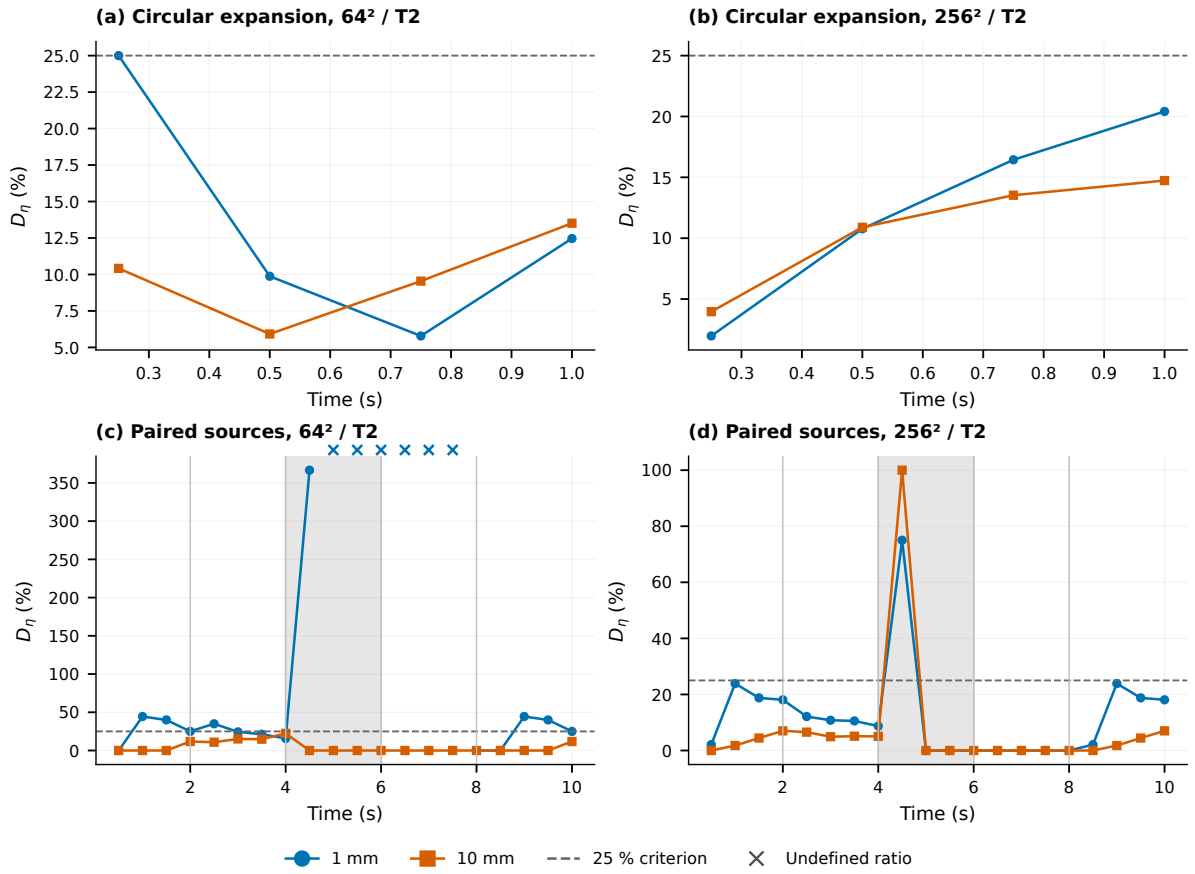


Figure 17: Wet-set discrepancy against the independent ANUGA references for T2: (a, b) circular expansion and (c, d) paired-source evolution on 64^2 and 256^2 grids. Squares and circles distinguish the strict depth thresholds. Dashed lines mark the 25 % diagnostic criterion. Shading marks withdrawal during 4–6 s; vertical lines mark source-phase changes. Crosses above the axes denote undefined ratios where the numerical wet set is nonempty and the reference wet set is empty; they are not finite discrepancy values.

The failed manual-step source budgets and unchanged-state double-accumulation control are retained in S6/S11; release and manual trajectories remain distinct.

S26.4 Merewether observations and terrain

Merewether is a real-terrain execution workload with restricted observational comparison. Its 1000 s constant peak inflow, five fixed marks and undocumented vertical datum do not validate arrival histories or full-event skill. Table S26 reports wet-site coverage with conditional stage mean absolute error (MAE). Since the wet subsets differ, their MAEs are not a direct grid-convergence comparison. Figures 18 and 19 show the spatial flood indicators and observational diagnostics. Table S27 distinguishes simulated wet stages from dry terrain/observation diagnostics.

Table S26: Merewether fixed-site coverage and conditional peak-stage error. The wet subsets differ, so the MAEs are not a grid-convergence comparison.

Grid	Wet fixed sites / all marks	Conditional MAE (m)
Coarse	4 / 5	0.174
Fine	3 / 5	0.152

Table S27: Five fixed fine-grid locations. A dry model cell has no simulated flood-mark water level.

ID	Bed (m)	Observed mark (m)	Peak depth (m)	Ever wet	Interpretation
0	19.471	19.980	0.686	Yes	Wet stage residual +0.177 m
1	17.690	18.380	0.763	Yes	Wet stage residual +0.073 m
2	23.590	23.360	0.000	No	Bed above observed mark
3	23.090	23.140	0.000	No	Observed positive depth; model dry
4	22.575	23.010	0.228	Yes	Wet stage residual -0.207 m

At site 2, the model bed is about 0.230 m above the observed mark. Nonnegative depth cannot reproduce that lower elevation even if the cell becomes wet. Terrain representation, point mapping or vertical datum may contribute, so this mismatch is not solely evidence of failed propagation. Site 3 has a physically representable positive observed depth but remains dry. Neither site is moved after inspecting the result.

The maximum relative volume residual is 1.704×10^{-9}, considerably tighter than the spatial agreement with observations. Volume conservation, wetting extent and observed water level thus provide complementary evidence. The prescribed peak-flow experiment does not validate the full event hydrograph, velocity field or observed arrival times. Its source and urban settings are useful execution workloads within these physical limitations.

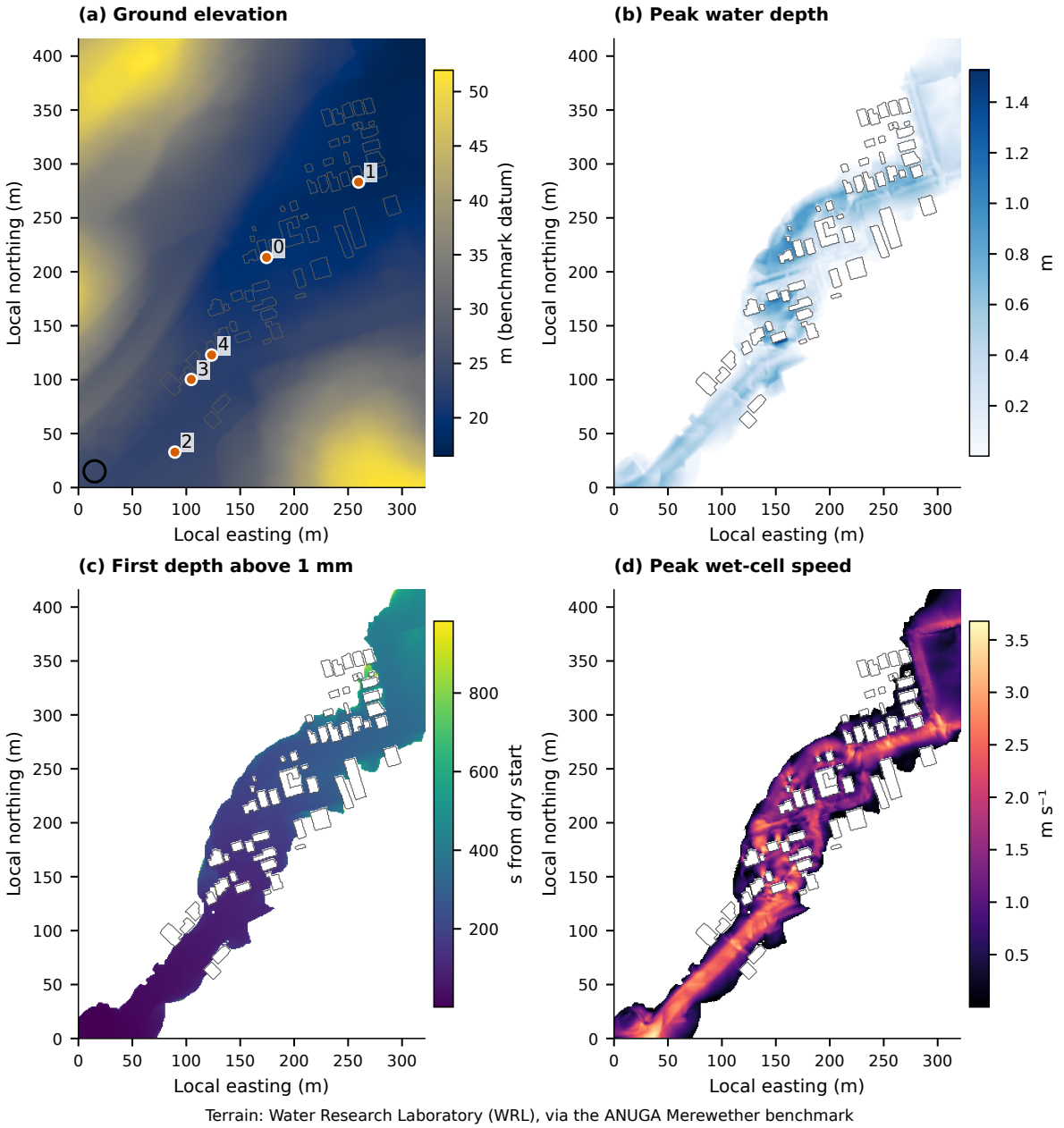


Figure 18: Merewether on the fine grid: (a) ground elevation and building outlines, (b) peak water depth, (c) first crossing of the 1 mm depth threshold and (d) peak speed in wet cells. Numbered points identify the five fixed observation sites; the circle marks the prescribed inflow disk. White areas in flood panels have no qualifying inundation. Time is measured from the prescribed dry start. Terrain supplied by the Water Research Laboratory, Manly, New South Wales, via the ANUGA benchmark (Geoscience Australia, n.d.).

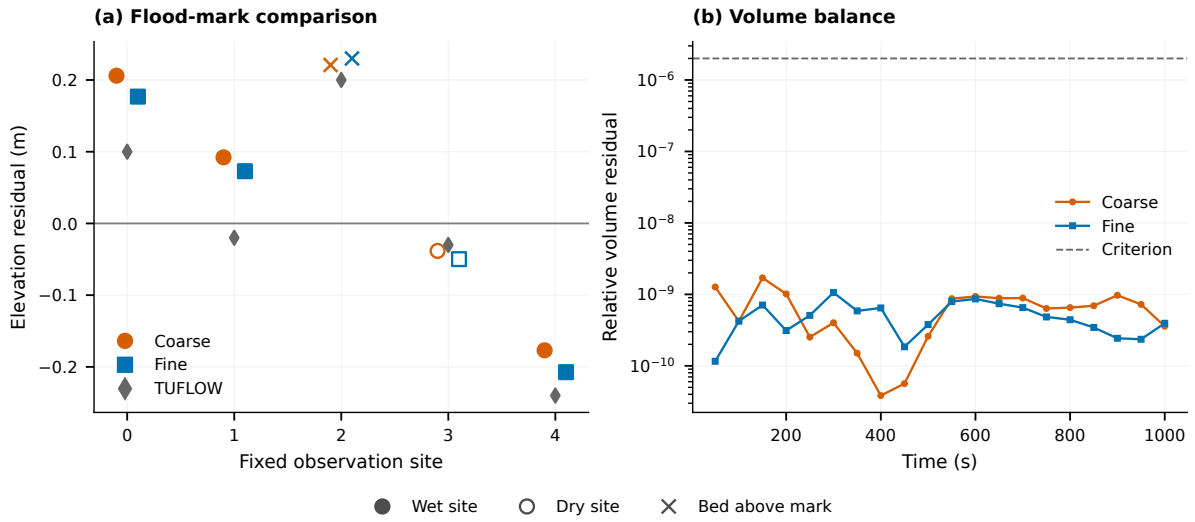


Figure 19: Merewether (a) fixed-site elevation residuals and (b) relative volume residuals. Filled coarse/fine markers show simulated wet-site peak stages minus observed marks; hollow markers show dry terrain-plus-depth diagnostics, not simulated flood levels. Crosses identify bed elevations above the observed mark. TUFLOW values are a supplied model comparison. The dashed budget line denotes the diagnostic criterion of 2×10^{-6}.

An independent reconstruction verifies the terrain and fixed-site mapping (Sect. S12). Site 2's bed/mark mismatch is already present in the original DEM; site 3 instead has unresolved propagation/forcing error. The vertical datum remains undocumented. No datum correction or movement of observation points is introduced.

S26.5 Reporting and numerical thresholds

S10 defines the reporting-only scan and joint numerical-policy experiment, including all 512 reproduced comparisons and 150/92/90 pooled source-front failures. These correlated counts cannot select a preferred numerical default. Figure 23 connects fixed-reporting-cutoff accuracy to conditional execution cost; the 0.5 mm policy retains volume failures and the 2 mm policy has no matched ANUGA source closure. Ritter refinement and front interventions are analysed in S17/S19/S22.

S26.6 Equal wet area does not imply equal cost

Across the 36 main cases, filtered-block dense/variant ratios range from 0.992 to 1.960; compacted-block and cell-indexed ranges are 0.862–1.717 and 0.696–1.489. None is uniformly faster. At a nominal 1 % wet area, geometry changes stage-weighted block support from about 2.7 % to 46.6 % of the domain. Cell support stays smaller, yet the cell-indexed path is slower than filtering for the continuous case and becomes competitive on the fragmented case (Table S28; direct paired ratios are reported in Sect. S4). Spatial work count (Fig. 20) is therefore an incomplete predictor of complete advancement cost.

Table S28: T2 at 1 % nominal initial wet area. Support is averaged over actual stage updates. Times are medians of process medians on CLOCK_MONOTONIC; initialization and downloads are excluded.

Geometry	Block support (%)	Cell support (%)	Dense (ms)	Filtered (ms)	Compacted (ms)	Cell (ms)
Continuous	2.704	1.453	9.294	4.738	5.406	6.238
Bands	25.000	7.805	9.586	6.625	7.579	6.974
Fragmented	46.621	7.946	9.414	7.234	8.363	6.811

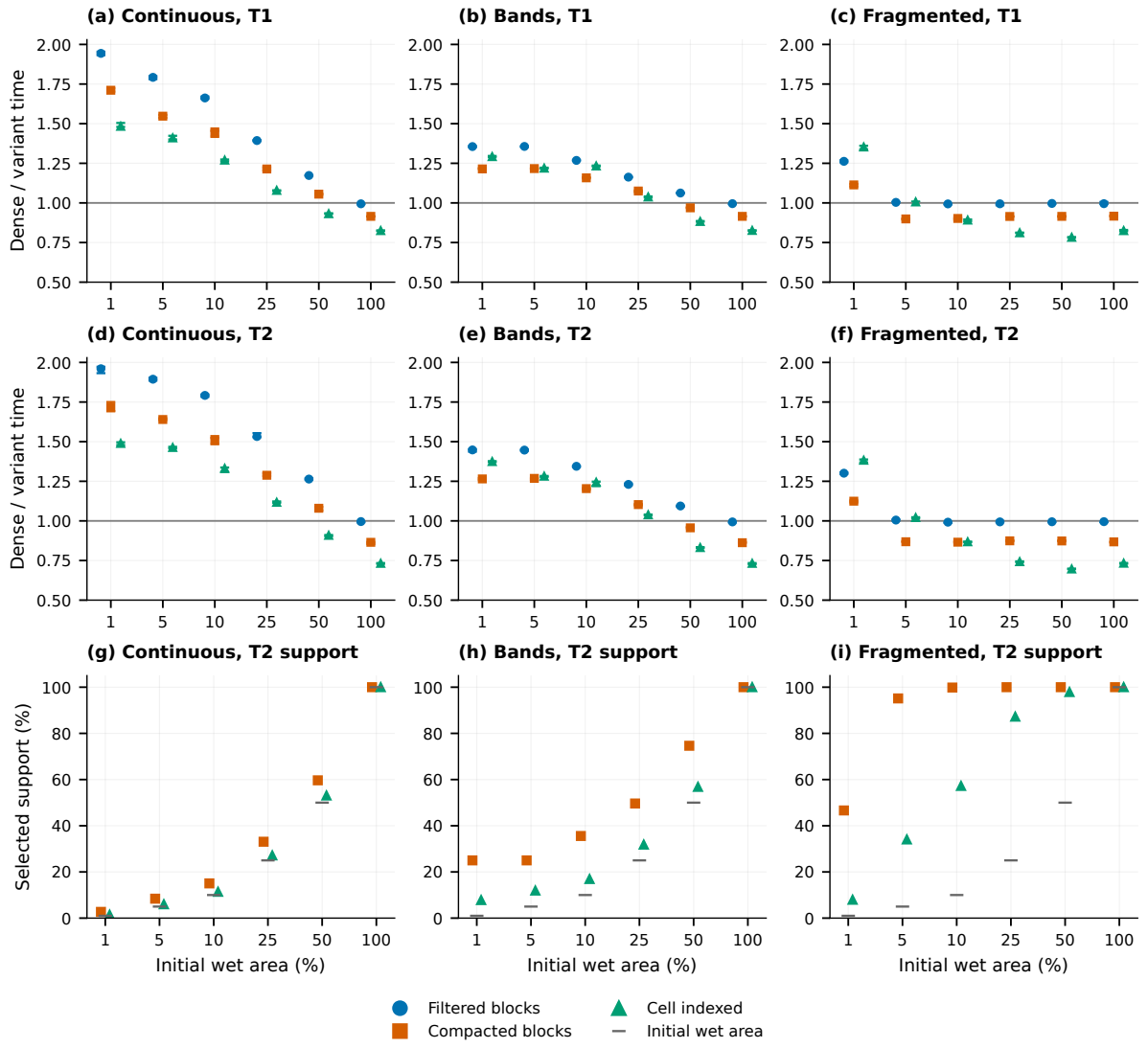


Figure 20: Execution time and selected support across the geometry matrix. Panels (a–c) and (d–f) show dense/variant complete-time ratios for T1 and T2; bars are conditional 95 % bootstrap intervals from four processes. Ratios above one favour the variant. Panels (g–i) show T2 stage-weighted block and cell support, with initial wet area for comparison. The filtered and compacted paths share the block-selection rule. Wet fractions are discrete categories; repeated fully wet labels have identical physical inputs.

The continuous wet region lets block filtering avoid most transport work while preserving a

regular tiled stencil. Fragmentation expands its selected block neighborhoods, reducing that advantage. Cell indexing avoids part of this geometric expansion but also changes list construction, memory access and state preparation. The ordering of complete costs consequently depends on their combination, not only on how many cells enter the transport kernel.

S26.7 Initial alignment advantage changes during evolution

Table S29 and Fig. 21 show the window support and timing contrasts for the translation checks defined in S5. Bands retain identical saved fields after undoing the shift; later fragmented fields differ and do not isolate alignment.

Table S29: Translation controls. The last column is unshifted/shifted filtered time; support uses compacted-block counts with the same selection rule.

Input	Window (s)	Unshifted support (%)	Shifted support (%)	Relative change (%)	Paired time ratio
Bands	0–0.2	25.000	23.068	-7.726	1.032 [1.029, 1.036]
Bands	0.2–1	33.183	33.492	+0.930	0.995 [0.992, 0.997]
Bands	1–2	44.834	45.371	+1.198	0.994 [0.993, 0.997]
Fragmented	0–0.2	46.707	47.122	+0.889	0.997 [0.994, 1.004]
Fragmented	0.2–1	67.107	67.063	-0.066	0.994 [0.991, 0.997]
Fragmented	1–2	80.453	80.108	-0.429	1.000 [0.999, 1.002]

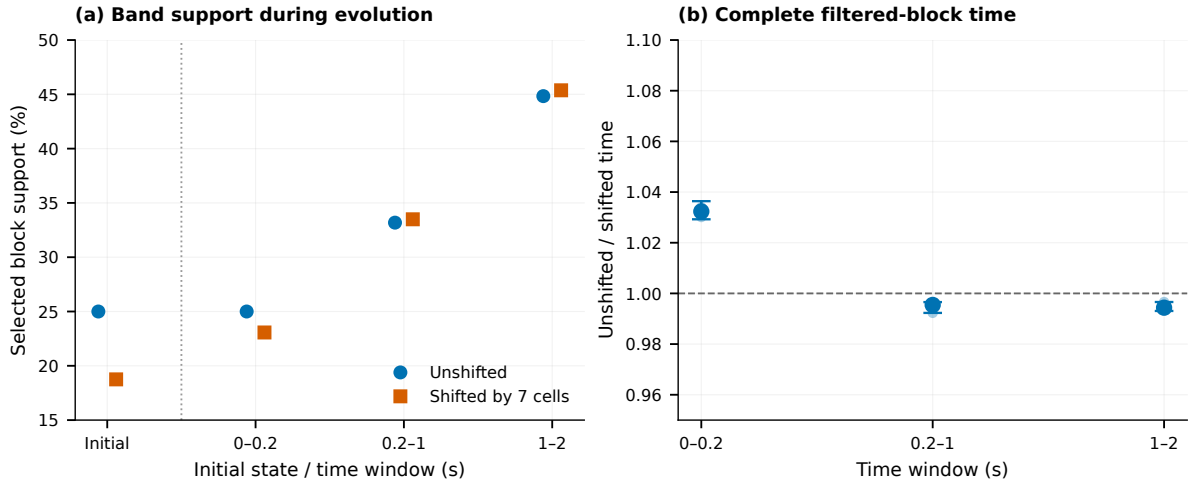


Figure 21: Effect of a seven-cell translation on the 1 % band input (T2). (a) Initial and stage-weighted selected block support. The initial state is separated from window averages, which weight transport stages rather than physical time. (b) Paired complete-time ratios for filtered blocks; small points show process medians and bars show conditional 95 % bootstrap intervals. Ratios above one favour the shifted input. Saved conserved fields agree after undoing the translation through 2 s.

S26.8 Source events and urban extent change the balance

Source and urban timings are summarized in Fig. 22. Localized sources retain opportunities for filtering, whereas domain-wide rainfall reduces them. The paired 512^2 source case gives a filtered ratio of 1.254 [1.232, 1.264]; the corresponding indexed comparisons are close to one.

On the smaller rain and single-patch grids, indexed maintenance costs are much larger relative to the available transport savings. These contrasts show that an implementation efficient for a large sparse field need not be efficient for a smaller problem with frequently updated sources.

Merewether shows a related size dependence: the filtered ratio is 0.949 [0.947, 0.952] on the coarse grid and 1.004 [1.002, 1.007] on the fine grid. The coarse case favours dense execution; the fine case has a small resolved filtering benefit within this session. Both indexed variants are slower on these two grids. All paths retain the same saved physical states and hazards at a given grid, so these differences concern the execution strategy rather than a more favourable flood solution.

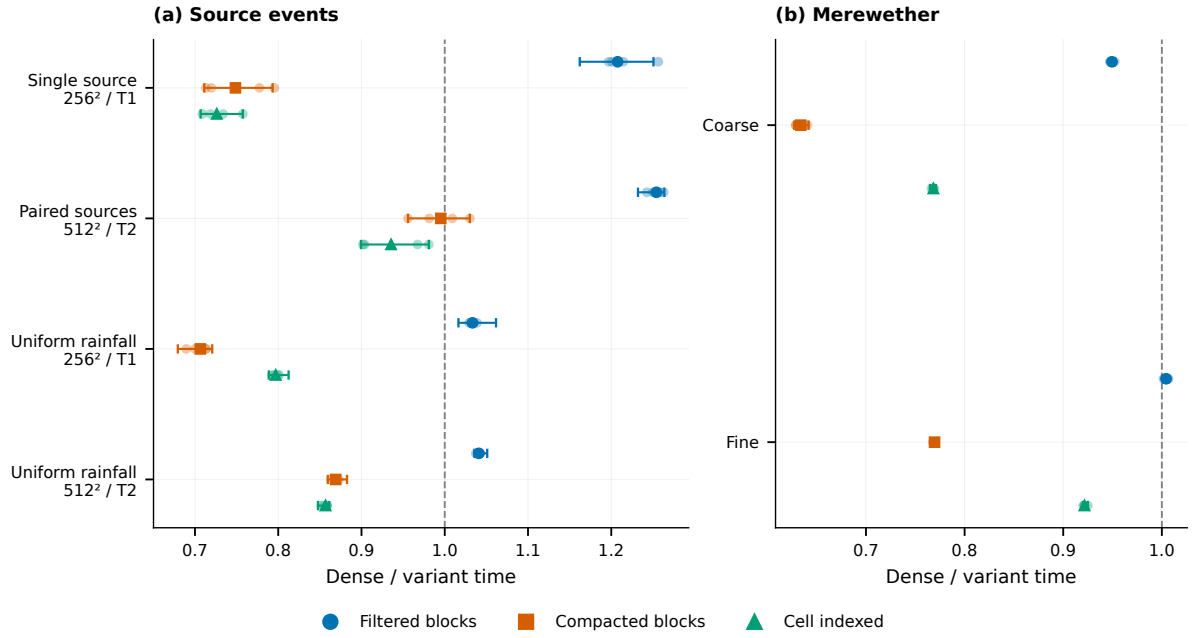


Figure 22: Paired complete-time ratios for (a) source-event cases and (b) the two Merewether grids. Symbols distinguish execution paths; small points are the four process medians, and bars are conditional 95 % bootstrap intervals. Ratios above one favour the variant. The source interval includes source uploads and synchronized advancement; urban intervals include synchronized advancement.

S26.9 Preserving state is part of the update cost

The indexed preparation inventory (main Table 3) and fused/copy comparison (S4, Table S6) show modest resolved fusion benefits on the base matrix. The separate fixed-support endpoint study (S27) instead resolves a large-size fusion penalty on the primary platform. Benefits do not transfer automatically between workloads and sizes. This does not establish fully optimized cell indexing.

Separate Linux CUDA-event stage values are reported in Supplement Table S8 with the metrology qualification in Sect. S23. The ordinary restored-work intervention and synchronized host partitions provide the current mechanism evidence (Fig. 5 and Sect. S21). Untimed free-memory deltas of 126 MiB for dense/filtering/compacted paths and 132 MiB for cell indexing in this case are allocation snapshots, not peak-memory measurements. Allocation granularity can conceal small differences. Retaining a float32 main value and residual uses eight bytes per scalar; no memory or runtime advantage over a fully float64 solver is established without that separate implementation.

S26.10 Numerical-cutoff cost and accuracy panel

Figure 23 brings together the numerical-policy results defined in S10 and Table S30. Cell-list and accumulator interventions have their full execution and validation inventories in S9 and S11, respectively.

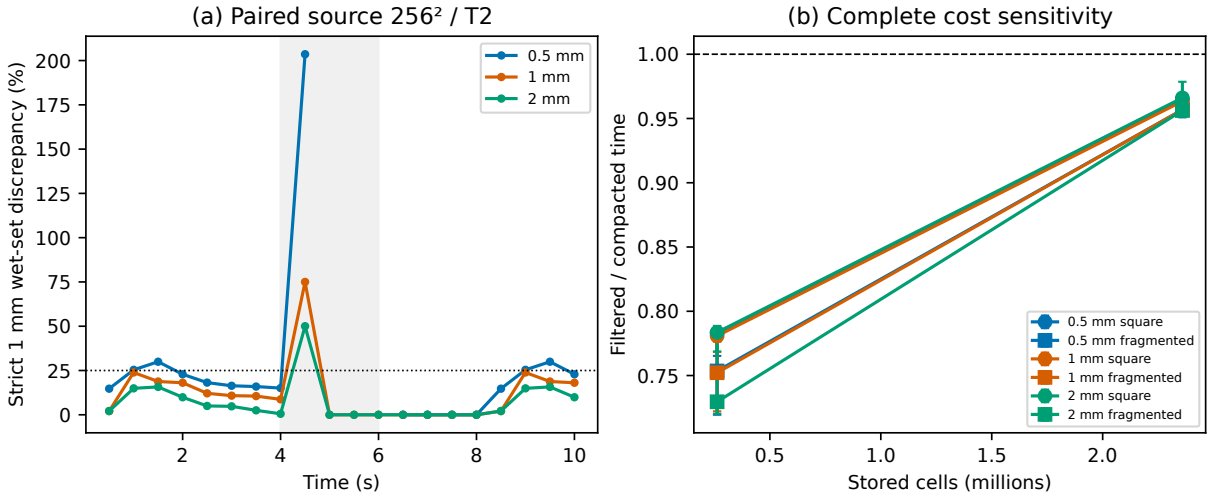


Figure 23: Numerical cutoffs of 0.5, 1 and 2 mm: (a) paired-source 256^2/T2 discrepancy at the fixed strict 1 mm reporting cutoff and (b) paired filtered/compacted complete-time ratios for square and fragmented dry-domain inputs. Shading denotes withdrawal. The two panels answer different questions: physical-front sensitivity and conditional scheduling cost. The 2 mm source comparison has no matched ANUGA closure reference.

S26.11 Auxiliary protocol inventory

Table S30: Consolidated core, auxiliary and numerical protocols. MONO/RAW denote adjusted/raw host clocks; E is separate CUDA-event profiling. Repeats exclude one warmup per variant. Numerical and stage checks are untimed; all paths means dense, filtered, compacted and cell indexed. Source coordinates, rates and terrain/boundary closures remain in S26.1. Core rows use flat bed, zero sources and 0.125 m spacing. Table 8 in the main text reports dual-clock endpoint effects; the preparation sensitivity is specified in S27.

Experiment / geometry	Grid / spacing	Integration and targets	Policy / paths	Processes and repeats / clock	Control and check scope
Dry extension / 8 local groups	512^2, 1024^2, 1536^2; offsets 0/7/15	T2; 0.25, 1, 2 s	Default / all	4 by 3; MONO, separate E	Common state, shift, exterior, steps, cumulative support
Restored work / 2 development groups	Same sizes; offset 0	T2; 0.25, 1, 2 s	Default / F,C,R	3 by 7; new MONO/RAW, historical MONO	Saved common state/steps, C/R support, mass; earlier stage audit S20

Table S30 continued.

Experiment / geometry	Grid / spacing	Integration and targets	Policy / paths	Processes and repeats / clock	Control and check scope
Front-policy connection / same 2 groups	512^2, 1536^2; offset 0	T2; 0.25, 1, 2 s	Each candidate plus default / F,C	3 by 5; MONO/RAW	Within-policy common state, steps, support; no cross-policy equality
Subsequent / 2 groups	512^2, 1024^2, 1536^2; 0.125 m; offsets 0/7/15	T2; 0.25, 1, 2 s	Default / all	3 by 3; MONO, separate E	Common fields, shifts, exterior, steps, support; models frozen first
Cell-list intervention / square, patches, fully wet	512^2/1536^2; 0.125 m	T2 sparse: once to 2 s; T1/T2 full wet: once to 0.25 s	Default / old,new cell; dense,F controls	3 by 5; MONO, separate E	Full final fields/steps; separate stage/list probes
Cutoff sensitivity / same cost inputs	512^2/1536^2; 0.125 m	T2 sparse: 2 s; T1/T2 full wet: 0.25 s	0.5/1/2 mm / all	3 by 3; MONO, separate E	Within-policy state/steps; separate Ritter/source accuracy
Ritter / analytical flat dry front	64–1024 x cells; 100/n m	T2; 3 s; CFL 0.25, capped/fixed dt	Default and factors / dense; earlier T1/path controls	Untimed numerical tests	Exact cell means, mass, wet edge, dt and clipping
Source / flat withdrawal-restart	64^2/256^2 accuracy; up to 512^2 timing; 40/n m	T1/T2; accuracy every 0.5 s; timing every 2 s to 10 s; cap 0.02 s	Default / all	4 by 3 timing; MONO, separate E	Saved fields, finite references, budgets; uploads included
Merewether / terrain, constant inlet	161 by 208 / 321 by 416; about 2 / 1 m	T2; every 50 s to 1000 s; cap 1 s	Default / all	4 by 2; MONO, separate E	Saved fields, hazards, steps; five fixed marks

S27 Matched preparation endpoints on two platforms

The protocol fixes the two core geometries, offset zero, 512/1536 sizes, default T2, and continuing targets 0.25/1/2 s before sampling. F is filtered/copy, C compacted/copy, and Cf compacted/fused, using the existing release module. F is paired with both C variants in the same repeat. Three serial processes use one warmup and five ordinary repeats; the six size/path combinations are randomized within each geometry/repeat. No numerical default, timing exclusion or adaptive repeat is introduced. The campaign was run three times: twice on the primary platform, one hour apart on the same instance without a reboot, and once on the second platform. Evidence is in `experiments/results/linux4090_20260928/preparation_session1/`, `preparation_session2/` and `experiments/results/linux5090_20260928/preparation/`. These are same-day implementation tests, not independent-date evidence.

The existing audited trajectory module checks all 12 combinations before ordinary timing on each platform. At every step, six captures hash the common public and residual fields and require zero exterior. Stage-one and stage-two launch support is checked against independent block dilation; compacted coordinates must match. All normalized hashes, supports and actual dt agree across sizes/preparations. The 13,716 captures and 4572 stage-support checks pass; all 216 release trajectories per session match their audited saved fields, window steps and C/Cf

support, and pass mass <2e-6, finite/nonnegative, boundary and completion guards. The second primary-platform session reuses that unchanged-binary stage audit rather than repeating it, and contributes fresh ordinary samples and warmups only; its `run.json` records that reuse explicitly. These checks apply to the default operator, not the momentum-retention candidate.

Table S31: Preparation sensitivity on the primary platform, session 1: paired RAW trajectory ratios and differences. Intervals are conditional 95 % hierarchical process/repeat bootstrap intervals, with 10000 draws and the median of process medians as estimate. Positive a-b and ratios a/b above one favour b. Savings are relative to a and obtained as 100(1-1/r), with bounds transformed from the ratio interval. Process medians for both clocks, all samples and audit bindings are in the archived summary.

Input / width	a/b	Ratio [95 %]	a-b (ms) [95 %]	Saving (%) [95 %]
Continuous / 512	F/C	0.813 [0.742, 0.841]	-5.9 [-10.6, -5.0]	-23.0 [-34.8, -18.9]
Continuous / 512	F/Cf	0.973 [0.928, 0.994]	-0.7 [-2.1, -0.2]	-2.8 [-7.8, -0.6]
Continuous / 512	C/Cf	1.189 [1.179, 1.341]	5.0 [4.9, 10.4]	15.9 [15.2, 25.4]
Continuous / 1536	F/C	0.965 [0.952, 0.981]	-3.9 [-5.5, -2.1]	-3.6 [-5.0, -1.9]
Continuous / 1536	F/Cf	0.925 [0.912, 0.929]	-8.8 [-10.5, -8.6]	-8.1 [-9.6, -7.7]
Continuous / 1536	C/Cf	0.958 [0.957, 0.971]	-4.9 [-5.1, -3.4]	-4.4 [-4.5, -3.0]
Fragmented / 512	F/C	0.814 [0.810, 0.816]	-6.8 [-7.0, -6.7]	-22.9 [-23.4, -22.5]
Fragmented / 512	F/Cf	0.924 [0.920, 0.927]	-2.5 [-2.6, -2.4]	-8.2 [-8.7, -7.9]
Fragmented / 512	C/Cf	1.132 [1.131, 1.138]	4.3 [4.2, 4.4]	11.6 [11.6, 12.1]
Fragmented / 1536	F/C	0.958 [0.956, 0.958]	-5.1 [-5.3, -5.0]	-4.4 [-4.6, -4.3]
Fragmented / 1536	F/Cf	0.916 [0.913, 0.918]	-10.7 [-11.0, -10.4]	-9.2 [-9.5, -8.9]
Fragmented / 1536	C/Cf	0.957 [0.955, 0.958]	-5.5 [-5.7, -5.3]	-4.5 [-4.7, -4.4]

Table S32: The same campaign repeated in primary-platform session 2.

Input / width	a/b	Ratio [95 %]	a-b (ms) [95 %]	Saving (%) [95 %]
Continuous / 512	F/C	0.813 [0.792, 0.825]	-6.0 [-7.3, -5.5]	-23.1 [-26.3, -21.2]
Continuous / 512	F/Cf	0.971 [0.966, 0.974]	-0.8 [-0.9, -0.7]	-3.0 [-3.5, -2.7]
Continuous / 512	C/Cf	1.197 [1.178, 1.226]	5.2 [4.7, 6.7]	16.5 [15.1, 18.5]
Continuous / 1536	F/C	0.959 [0.957, 0.975]	-4.7 [-4.9, -2.8]	-4.3 [-4.5, -2.5]
Continuous / 1536	F/Cf	0.927 [0.924, 0.934]	-8.6 [-8.9, -7.8]	-7.9 [-8.2, -7.0]
Continuous / 1536	C/Cf	0.966 [0.962, 0.968]	-4.0 [-4.4, -3.9]	-3.5 [-3.9, -3.4]
Fragmented / 512	F/C	0.811 [0.807, 0.815]	-7.0 [-7.1, -6.8]	-23.4 [-23.9, -22.6]
Fragmented / 512	F/Cf	0.923 [0.916, 0.925]	-2.5 [-2.7, -2.4]	-8.4 [-9.1, -8.1]
Fragmented / 512	C/Cf	1.137 [1.131, 1.142]	4.4 [4.3, 4.6]	12.1 [11.6, 12.4]
Fragmented / 1536	F/C	0.953 [0.952, 0.954]	-5.7 [-5.8, -5.6]	-4.9 [-5.0, -4.8]
Fragmented / 1536	F/Cf	0.919 [0.918, 0.921]	-10.3 [-10.4, -10.0]	-8.9 [-8.9, -8.6]
Fragmented / 1536	C/Cf	0.964 [0.963, 0.966]	-4.5 [-4.7, -4.2]	-3.7 [-3.9, -3.5]

Table S33: The same campaign on the second platform. The large-size C/Cf intervals include one, so fusion's direction at that size is unresolved on this device. They are not reported as equivalence.

Input / width	a/b	Ratio [95 %]	a-b (ms) [95 %]	Saving (%) [95 %]
Continuous / 512	F/C	0.707 [0.673, 0.709]	-9.5 [-11.1, -9.4]	-41.5 [-48.5, -41.0]
Continuous / 512	F/Cf	0.826 [0.810, 0.829]	-4.8 [-5.4, -4.7]	-21.0 [-23.4, -20.6]
Continuous / 512	C/Cf	1.171 [1.144, 1.227]	4.7 [4.1, 6.3]	14.6 [12.6, 18.5]
Continuous / 1536	F/C	0.964 [0.959, 0.970]	-2.8 [-3.2, -2.3]	-3.7 [-4.2, -3.0]

Table S33 continued.

Input / width	a/b	Ratio [95 %]	a-b (ms) [95 %]	Saving (%) [95 %]
Continuous / 1536	F/Cf	0.951 [0.945, 0.968]	-3.9 [-4.4, -2.5]	-5.1 [-5.9, -3.3]
Continuous / 1536	C/Cf	0.984 [0.982, 1.004]	-1.3 [-1.4, 0.3]	-1.7 [-1.9, 0.4]
Fragmented / 512	F/C	0.736 [0.733, 0.740]	-9.5 [-9.7, -9.3]	-35.8 [-36.5, -35.1]
Fragmented / 512	F/Cf	0.799 [0.783, 0.818]	-6.7 [-7.3, -6.1]	-25.2 [-27.7, -22.3]
Fragmented / 512	C/Cf	1.086 [1.065, 1.090]	2.8 [2.2, 3.0]	7.9 [6.1, 8.2]
Fragmented / 1536	F/C	0.960 [0.952, 0.961]	-3.4 [-4.1, -3.3]	-4.2 [-5.1, -4.1]
Fragmented / 1536	F/Cf	0.945 [0.945, 0.959]	-4.8 [-4.8, -3.5]	-5.8 [-5.9, -4.3]
Fragmented / 1536	C/Cf	0.985 [0.981, 1.007]	-1.3 [-1.6, 0.6]	-1.5 [-1.9, 0.7]

RAW/monotonic ratios remain within 0.9997 to 1.0000 on both platforms, and the maximum paired ratio difference between clocks is below 0.0002; no `tick=9000` condition appears on either Linux host. Neither clock is externally calibrated.

Small-domain C/Cf exceeds one in all six geometry/session combinations, so fusing the preparation copy and clear helps at 512^2 on both devices. Large-domain C/Cf is about 0.958 to 0.966 on the primary platform, wholly below one in both sessions and geometries, so fusion is resolved as slower there despite removing launches. On the second platform the large-domain point estimates are 0.984 and 0.985 with intervals crossing one, so the direction is undetermined. Fewer preparation launches therefore do not imply shorter complete advancement, and the large-size evidence resolves a penalty on the primary platform and leaves the direction unresolved on the second. This difference in significance does not itself test a between-platform effect difference.

F/Cf stays below one in all twelve platform/geometry/size rows, so fused preparation never makes the compacted path faster than filtering. This preparation contrast compares two variants of the compacted path and is separate from the filtered/compacted ranking of S29. No timing data from other sessions are pooled with these experiments.

The initial analysis rejected integer-valued floating-point support counters from the native API. All audits and worker processes had completed successfully. The failed orchestration receipt and original frozen source bytes are retained; only analysis was repaired to accept exact integral counts and reject fractional, nonfinite or boolean counters. A fresh CPU reanalysis verifies the frozen source snapshots, all raw reports and audit hashes. `summary.json` distinguishes complete measurement/analysis from the retained original analysis failure; no simulation or timing sample was replaced. The source/record contract and an intentional mismatch example are described in `docs/ENDPOINT_PROTOCOL.md`.

S28 Subgroup cost-model errors

Table S34: Subsequent-set mean trajectory APE (%), recomputed from the same archived trajectory predictions as Table S24. Path rows pool geometries; geometry rows pool paths. These are marginal summaries, not path-by-geometry interactions. All four models still select the same paths; subgroup errors need not improve uniformly.

Subgroup	Initial support	Evolving support
block	11.232	11.061
block_compact	11.471	11.332
cell	9.064	8.860
dense	2.676	2.676
new_patch_8x16_48	8.668	8.318
new_rectangle_48x96	8.553	8.647

`check_scientific_values.py` recomputes these means from individual trajectory predictions, checks both this table and the main-text values, and binds the input hashes. All rows refer to the Linux primary campaign.

S29 Eight-size scan on two platforms

The frozen plans, binary/source bindings, three worker reports per session, GPU telemetry and summaries are in `experiments/results/linux4090_20260928/scan_session1/`, `scan_session2/` and `experiments/results/linux5090_20260928/scan/`. Each plan fixes the same eight sizes, two geometries, one warmup and five ordinary repeats in each of three serial processes, with randomized size/path ordering within each geometry/repeat. No timing-based exclusions are made. Synchronized advances and saved-window guards reuse the endpoint runner. Each session contains 576 trajectories including 96 warmups, each with three continuing windows, giving 1728 trajectories in total. All pass; common hashes and steps match across F/C and sizes within each process, and compacted support agrees across sizes. These campaigns add no per-stage audit. Raw records retain both clocks and each summary retains individual process medians.

The two primary-platform sessions ran on the same rented instance on the same UTC date, with the same binaries and no host reboot; the second began about one hour after the first. They are reported separately throughout and never pooled, because two sessions one hour apart do not constitute independent dates or independent hardware. The second-platform session ran on a different rented instance whose host processor and driver also differ, so between-platform differences belong to complete platforms rather than to GPU architecture alone.

Table S35: Paired RAW F/C ratios [conditional 95 % interval] for the primary platform, session 1. Values below one favour filtering. Intervals resample processes and paired repeats with 10,000 draws; they do not supply independent hardware sessions.

Size	Continuous	Fragmented
512	0.816 [0.721, 0.856]	0.789 [0.717, 0.817]
640	0.855 [0.671, 0.864]	0.826 [0.783, 0.854]
768	0.853 [0.780, 0.877]	0.843 [0.796, 0.865]
896	0.876 [0.838, 0.898]	0.857 [0.827, 0.890]
1024	0.880 [0.843, 0.900]	0.875 [0.834, 0.937]
1152	0.888 [0.875, 0.898]	0.901 [0.870, 0.941]
1280	0.961 [0.948, 0.979]	0.942 [0.906, 0.962]
1536	0.962 [0.953, 0.973]	0.949 [0.941, 0.956]

Table S36: The same scan repeated in primary-platform session 2, about one hour later on the same instance without a reboot. Every direction agrees with Table S35; point estimates differ by at most about 0.04.

Size	Continuous	Fragmented
512	0.852 [0.808, 0.880]	0.806 [0.792, 0.852]
640	0.861 [0.820, 0.917]	0.827 [0.825, 0.856]
768	0.853 [0.840, 0.872]	0.851 [0.827, 0.870]
896	0.882 [0.877, 0.912]	0.871 [0.817, 0.912]
1024	0.880 [0.878, 0.883]	0.877 [0.874, 0.896]
1152	0.885 [0.834, 0.901]	0.888 [0.887, 0.910]
1280	0.954 [0.952, 0.993]	0.952 [0.929, 0.959]

Table S36 continued.

Size	Continuous	Fragmented
1536	0.957 [0.943, 0.977]	0.955 [0.949, 0.962]

Table S37: The same protocol on the second platform. Ratios are lower at small and intermediate sizes, meaning filtering wins by a wider margin there, and converge to nearly the primary-platform value at 1536^2.

Size	Continuous	Fragmented
512	0.703 [0.675, 0.718]	0.741 [0.731, 0.743]
640	0.741 [0.711, 0.755]	0.770 [0.751, 0.774]
768	0.794 [0.763, 0.805]	0.793 [0.769, 0.794]
896	0.757 [0.755, 0.774]	0.752 [0.737, 0.771]
1024	0.783 [0.774, 0.789]	0.776 [0.771, 0.785]
1152	0.802 [0.792, 0.816]	0.797 [0.790, 0.800]
1280	0.838 [0.836, 0.842]	0.829 [0.825, 0.853]
1536	0.964 [0.958, 0.977]	0.952 [0.949, 0.958]

All 48 intervals in Tables S35 to S37 lie wholly below one. There is no sampled size, geometry, session or device at which the compacted path becomes faster, so the scan reports no crossover rather than locating one.

The ratio generally rises with stored size, but not at every consecutive step. On the second platform the continuous input falls from 0.794 at 768^2 to 0.757 at 896^2, and the fragmented input from 0.793 to 0.752 across the same pair; the primary platform shows smaller reversals of the same kind, such as the continuous input at 640^2 and 768^2 in session 1. We therefore describe a narrowing trend with sampled exceptions and do not fit a slope law, a change point or an asymptote to these points. Because no crossover occurs, the earlier question of which size first resolves in favour of compaction does not arise in these data.

The CUDA runtime query `cudaDeviceGetAttribute(cudaDevAttrL2CacheSize)` returns 75,497,472 bytes (72 MiB) on the primary device and 100,663,296 bytes (96 MiB) on the second. Equating each capacity to the minimum 48 bytes of state per stored cell places candidate boundaries near 1254 and 1448 cells per side respectively. On the primary platform, 1280^2 and 1536^2 both exceed the 1254-cell boundary and still favour filtering in both geometries and both sessions. On the second platform, 1536^2 exceeds the 1448-cell boundary and still favours filtering in both geometries. The measurements therefore contradict this capacity-equality rule on these two devices. This heuristic was devised for exploration here; it is not a theoretical cache model or a prediction taken from prior literature.

That rejection concerns one specific rule. Allocation does not specify reuse distance or the live working set, extra arrays and access order affect residency, and distinct smooth per-step and per-domain cost growth can also shape these curves. The failure of the equality rule therefore does not exclude other cache effects, and none is established here.

Nsight Compute probes requested DRAM read/write bytes, L2 sector hit rate and achieved occupancy on both platforms. Neither rented container granted counter access: the profiler reported insufficient permissions for GPU performance counters, and no counter value was obtained on either device. The failure records are retained with the campaign evidence. The proposed F/C/R endpoint profiling matrix was consequently not run on either platform; a failed probe is not a hardware measurement and is excluded from ordinary timings. Every statement about cache or bandwidth behaviour in this paper is therefore a hypothesis consistent with timing, not a measured mechanism.

S30 Bounded numerical candidate validation

The pre-run protocol is `experiments/REVISION_VALIDATION_20260928.md`. The fixed candidate disables the Froude cap and retains thin-layer momentum; it preserves transport gates and positivity zeroing. It is compared with the default on all four execution paths. This is a validation of a previously selected candidate, not an independent search for a favourable policy.

The initial matrix contains 88 trajectories. All configurations reach their targets and retain finite nonnegative depths, with matching step counts. Eighty-six satisfy saved-output bitwise agreement with their own policy's dense path; the two exceptions are the candidate filtered and compacted Thacker paths at 256^2. These two paths agree with each other, while the cell path agrees with dense. The maximum depth difference from dense over eight outputs is 0.0008104 m (momentum maxima 0.0002015 and 0.0007156 m^2/s). The discrepancy first appears after the initial quarter period. The subsequent stage-resolved diagnosis in S33 identifies a subthreshold nonidentity update omitted by the original block selector; the historical failure remains unchanged. The failure is retained and blocks default promotion; existing default-operator consistency results are unaffected. Closed-domain relative volume error stays below 3e-9. The masked lake-at-rest check meets the unchanged 5e-6 equilibrium tolerance. These guards do not certify accurate moving fronts or general stability.

Table S38: Ritter at 3 s, CFL 0.25 and maximum step 0.05 s. Negative edge error means lag. Depth norms are full-domain cell-average errors; discrepancy uses a strict 1 mm threshold. Default and candidate keep the same spatial operator except for the two explicitly changed clipping rules.

Grid	Policy	Depth L1 (m)	Wet discrepancy (%)	Edge error (m)
128	Default	0.008204	6.19	-4.6875
128	Candidate	0.007747	2.06	-1.5625
256	Default	0.005060	6.70	-5.0781
256	Candidate	0.003976	2.58	-1.9531
512	Default	0.003698	7.22	-5.4688
512	Candidate	0.001959	2.32	-1.7578
1024	Default	0.003059	7.35	-5.5664
1024	Candidate	0.000978	1.68	-1.2695

The candidate reduces depth L1 approximately by half on successive grid refinements, but its wet-edge error is not monotonic over all four grids. At n=1024 the remaining 1.2695 m lag equals 13 cells and fails the retained two-cell accuracy criterion. Maximum steps 0.0015625 and 0.00078125 leave that edge unchanged. This rules out adopting the candidate as a validated default on the strength of the improved Ritter result alone.

The Thacker case uses the existing analytical translating free surface over a parabolic bed, cell-average reference quadrature and reflective boundaries. Eight quarter-period outputs cover two full wetting/retreat cycles. This adds moving-shoreline validation on a non-flat two-dimensional bed to the earlier quarter-period check. It does not introduce a new field-event validation.

Table S39: Thacker after two full periods. Errors are compared with the analytical cell averages; discrepancy uses the strict 1 mm threshold.

Grid	Policy	Depth L1 (m)	Wet discrepancy (%)	Relative volume error
64	Default	0.006350	29.62	1.734e-09
64	Candidate	0.006346	27.73	1.979e-09
128	Default	0.002219	15.25	8.258e-10
128	Candidate	0.002229	14.18	1.892e-10
256	Default	0.000783	4.91	2.480e-10
256	Candidate	0.000800	4.75	1.642e-10

Both policies' depth errors and endpoint wet-set discrepancies decrease with refinement. The candidate is not uniformly more accurate: at 128 and 256 its depth L1 is slightly higher while its wet-set discrepancy is lower. A first n128 maximum-step reduction from 0.05 to 0.025 s leaves actual steps and fields unchanged because the CFL limit is already tighter; it is retained as a null intervention, not temporal-refinement evidence.

An explicit pre-run amendment adds 16 n128 trajectories at maximum steps 0.003125 and 0.0015625 s. Actual total steps increase from 9080 to 18160 for both policies, and all four paths agree within each policy. Endpoint depth L1 changes from 0.002224 to 0.002224 m for the default and from 0.002229 to 0.002229 m for the candidate at the displayed precision. The strict 1 mm discrepancy is 14.91 % for the default at both caps and 14.06/14.03 % for the candidate. Endpoint depth differences between caps reach 0.000655/0.000950 m in Linf for default/candidate. This is a measured timestep sensitivity, not a demonstration of asymptotic temporal order.

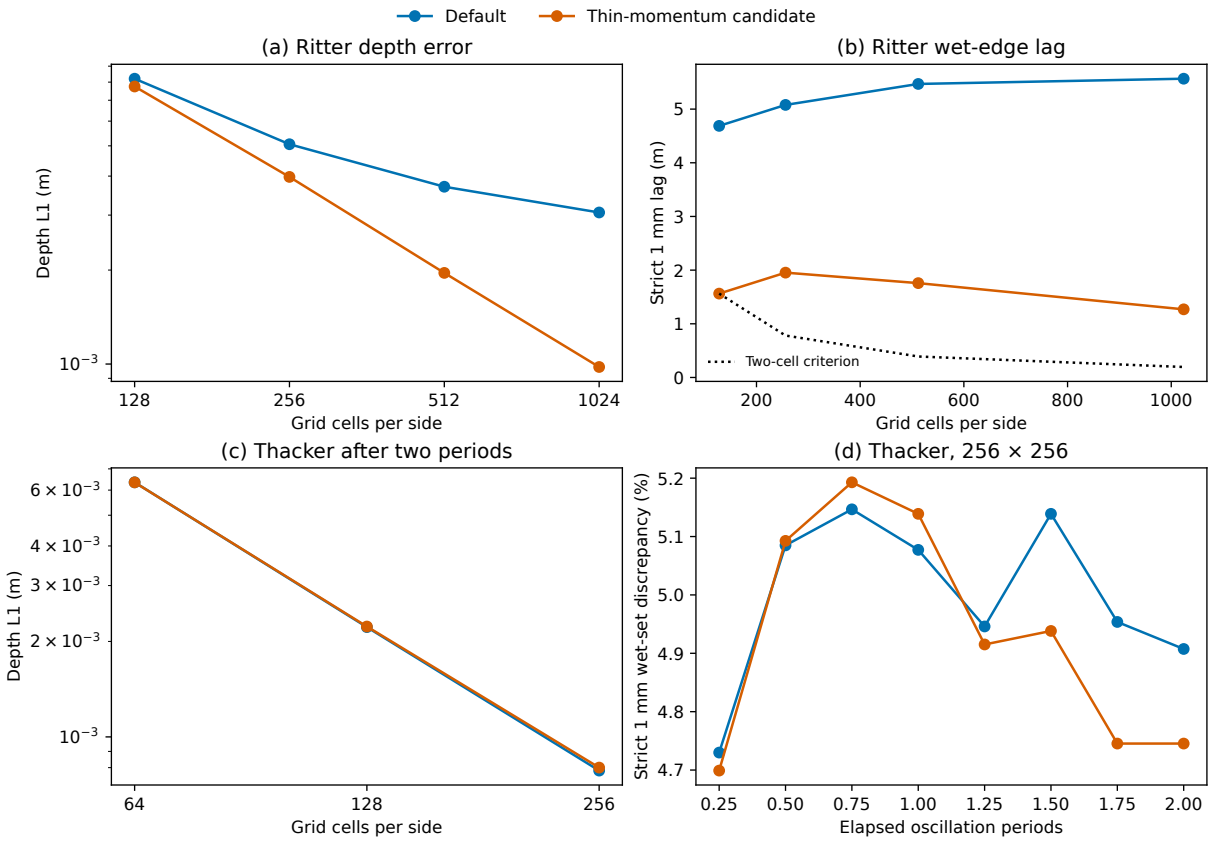


Figure 24: Default and thin-momentum candidate: Ritter depth refinement and strict 1 mm edge lag, followed by Thacker endpoint depth error and the 256-grid wet-set discrepancy at each quarter period. The dotted two-cell line is a development check recorded before this follow-up run, not a fitted target. Saved-output agreement across paths is distinct from error against the analytical reference.

S31 Stable block-scan compaction control

The original scheduler performs two CUB DeviceSelect operations per transport stage, followed by a shared count readback. The isolated alternative uses two 256-thread CTAs, one per interior/boundary list, and stable prefix scans over successive chunks. Coordinate order, host count readback/synchronization, copy or fused preparation, numerical kernels and finalization are retained. The original allocation remains, including unused selection scratch; this is an execution intervention, not a memory-capacity optimization. All 25 device functions shared with the baseline retain identical instruction words.

Before timing, 12 trajectories on two geometries and $512^2/1536^2$ pass all six public/residual stage captures, actual dt and both-stage list/support checks against the matching release build. A separate 12-trajectory audit passes for the thin-momentum candidate on these flat extension inputs; this does not override its non-flat Thacker consistency failures in S30.

The restarted-instance session uses three serial processes, one warmup and five randomized paired repeats for F, original C/Cf and alternative S/Sf, for 360 trajectories. All saved common states, steps and selected support agree. No earlier session is pooled with it. The host reports an uptime of about 453 days: the user's instance restart is not evidence of a host reboot, and all sessions still occur on one date.

Table S40: Stable block-scan control, raw-clock paired ratios with conditional 95 % process/repeat intervals. S uses copy preparation and Sf uses fused preparation. Ratios above one favour the denominator. The last two columns compare filtering with the new implementations, rather than with baseline C.

Input / size	C/S	Cf/Sf	F/S	F/Sf
Continuous / 512	1.038 [1.015, 1.055]	1.009 [1.001, 1.023]	0.807 [0.803, 0.818]	0.938 [0.936, 0.985]
Continuous / 1536	0.954 [0.948, 0.969]	0.956 [0.948, 0.970]	0.912 [0.910, 0.925]	0.882 [0.879, 0.895]
Fragmented / 512	1.019 [1.016, 1.047]	1.014 [0.984, 1.018]	0.785 [0.771, 0.787]	0.922 [0.921, 0.928]
Fragmented / 1536	0.953 [0.950, 0.954]	0.954 [0.952, 0.957]	0.905 [0.900, 0.906]	0.875 [0.874, 0.876]

At 512^2 the copy alternative is faster than C by about 2–4 % in ratio terms; the fused comparison is small and unresolved for the fragmented case. At 1536^2 both forms of the alternative are slower, with original/alternative ratios near 0.954–0.956. Filtering remains faster than either alternative at both endpoints. Replacing two general-purpose selections by a stable small-list scan therefore does not supply a generally faster compaction path. This narrows one implementation concern while leaving other scheduler designs, count-readback removal and state-maintenance optimizations untested. It does not establish that compaction as a technique is inferior or that C is optimal.

S32 Auxiliary cost explanation and controls

S32.1 Initial geometry and evolving support

Four domain-aware models explain window time per execution path as $\widehat{t} = aK + cNK + bX$: domain/steps ($b = 0$), initial wet cells ($X = KW_0$), initial selected support ($X = 2KS_0$), and evolving support ($X = A$). N is stored cells, K actual steps, W_0 the initial wet-cell count, and S_0 initial selected cell-equivalent support, fixed across windows. A sums selected cell-equivalent updates over both transport stages; block counts are multiplied by 256. Initial support includes halo and alignment. Dense has $S_0 = N$ and $A = 2NK$, so redundant columns are dropped. Coefficients are nonnegative relative least-squares fits on four development geometries. All sizes, offsets, windows and repeats of each geometry stay together; evaluation uses four original and two subsequent complete groups.

The evolving-support coefficients were frozen before original held-out execution (S8); the added domain/steps comparator was post hoc there. Both were frozen before the subsequent rectangle and patch groups ran (S14). The two initial- geometry baselines were proposed after both evaluation sets were observed, but fit only the original development groups (S25). No evaluation data were used to refit or tune these coefficients. Actual steps and cumulative support make the comparison retrospective time explanation, not online prediction.

Table S41: Retrospective trajectory-time explanation on complete geometry groups. APE is absolute percentage error in complete trajectory time. Models use actual steps and the specified initial or cumulative support. Static geometry rows are post hoc on both evaluation sets; Supplement Fig. 16 shows the distributions. The subsequent-set domain/steps and evolving-support coefficients were frozen before those two groups ran; original-group domain/steps evaluation was post hoc.

Model	Mean APE (%)	Median APE (%)	P90 APE (%)	Wrong best / cases
Original / domain + steps	7.67	8.14	12.65	0 / 36

Table S41 continued.

Model	Mean APE (%)	Median APE (%)	P90 APE (%)	Wrong best / cases
Original / initial wet + domain	7.67	8.17	12.61	0 / 36
Original / initial support + domain	7.59	7.39	12.90	0 / 36
Original / evolving support + domain	7.35	6.74	13.30	0 / 36
Subsequent / domain + steps	8.60	8.55	13.80	0 / 18
Subsequent / initial wet + domain	8.61	8.55	13.77	0 / 18
Subsequent / initial support + domain	8.61	9.26	14.10	0 / 18
Subsequent / evolving support + domain	8.48	9.01	13.27	0 / 18

Domain size and actual steps give 7.67 % and 8.60 % mean error on the two evaluation sets; adding evolving support reduces this to 7.35 % and 8.48 %, gains of 0.32 and 0.12 percentage points, without improving every quantile: original-group P90 increases (13.30 % versus 12.65 %). All models choose the same point-estimate winner in all 36/18 cases. The support features therefore establish no additional selection value. These are retrospective primary-platform diagnostics: actual steps and evolving support are known after execution. Coefficients, feature definitions and full error distributions are in S13, S23–S25; no online or cross-device selector is validated.

S32.2 A cell-selection implementation control

The original cell selector expands each active seed over a square neighbourhood in one pass. The candidate computes the same mask with horizontal then vertical expansion, retaining the final valid-cell mask and stable selection (S9). It changes selection work while retaining count readback, state preparation and numerical updates.

Table S42: Paired complete-time ratio of original cell indexing to the isolated two-pass expansion candidate. Intervals are conditional 95 % bootstrap intervals. Values above one favour the candidate.

Input	Original / candidate [95 %]	Filtered / candidate [95 %]
Square 512^2	0.977 [0.974, 0.983]	0.725 [0.720, 0.728]
Square 1536^2	1.023 [1.020, 1.025]	0.905 [0.903, 0.907]
Fragmented 512^2	0.986 [0.980, 0.998]	0.737 [0.734, 0.745]
Fragmented 1536^2	1.021 [1.020, 1.022]	0.897 [0.895, 0.898]
Full wet 512^2 / T1	0.989 [0.982, 0.996]	0.816 [0.810, 0.821]
Full wet 512^2 / T2	0.989 [0.988, 0.992]	0.762 [0.760, 0.765]

Both larger dry-domain cases favour the intervention, by about 2 %. The four smaller and fully wet cases favour the original implementation by 1–2 %. The benefit of this selection change is therefore itself size dependent, in the same direction as the other measurements: it helps where stored extent is large. All 27 baseline device-function occurrences retain their instruction words; only the two expansion kernels are added. The 1536 independent stage-list checks and 2304

field/residual-frame comparisons pass. Candidate cell indexing remains slower than filtered execution in every case (Supplement Fig. 11). One bounded improvement therefore strengthens the baseline comparison without establishing fully optimized indexing.

S32.3 Supporting execution controls

Joint numerical cutoffs of 0.5–2 mm do not change the path ranking or its size dependence: across the twelve dry-domain cases, large-domain filtered/compacted ratios are 0.956–0.966 and small-domain ratios 0.730–0.784, with every conditional interval below one. The six fully wet control cases also favour filtering, at 0.752–0.816. Support changes by less than 1 % and steps remain 183/195 for the square/fragmented single-target runs. S10 and Supplement Fig. 23 report the policy and accuracy limits; numerical defaults are unchanged.

Equal wet area does not fix stage support: at nominal 1 % wet area the matrix's block support spans about 2.7-46.6 %. A seven-cell band translation changes initial support by -25 %, but its first-window support change is only -7.73 % and later changes reverse sign. Saved translated band fields agree; later fragmented fields do not, limiting that comparison. Source/urban workloads and a fused preparation intervention likewise show no universally fastest path. The complete matrices, process points and negative outcomes are in S4-S5 and S26; they support the assessment method but do not isolate dry extent as cleanly as Sect. 4.

S33 First-stage omission failure and isolated selection repair

The protocol is `experiments/NONFLAT_PROTOCOL_20260928.md`. This diagnosis replays the previously failed 256^2 candidate Thacker trajectory from its original initial state, preserving retained residuals. Eight continuing quarter-period outputs and CFL 0.25 are unchanged. It compares all six stage captures, all three public fields, all three residuals and actual dt; saved public snapshots are not used as exact restart states.

The first unequal capture is index 4036: zero-based step 672 (step 673 when counted from one), second Euler output. Its step starts at 6.534717753 s with dt=0.009737942368 s. All preceding captures agree bitwise. The 19 affected cells lie in block row 5, column 12, and their input depths range from 0.0000886231 to 0.000998636 m. Actual stage-input flags agree with an independent 1 mm wet-block dilation oracle and omit that block. F and C retain the identical stage input; dense changes momentum. The first maximum differences are 3.0396e-7 and 3.0291e-7 m^2/s in public momentum, and 1.9211e-11 and 8.5282e-11 m^2/s in its residuals. Depth and its residual still agree at this capture.

The dry/dry branch of `hll_flux_interface_2d` sets mass flux to zero but retains momentum pressure; hydrostatic pressure corrections also enter the update. The candidate suppresses the default subthreshold momentum clearing, so this remaining update need not be an identity. Block omission nevertheless copies the old state. This explains why two sparse implementations agree while both fail against dense; it is not evidence of differing time controllers or lists between F and C.

An isolated control seeds blocks using every positive public depth, including initial/imported flags, skipped-block flags and transport/finalization flags. The 3×3 dilation, transport threshold, CFL speed calculation and all numerical arithmetic remain unchanged. The expanded-selector dense, F and C each agree with the original candidate dense at all 17,496 public/residual captures through two periods, with identical actual dt. Original dense release and all three expanded-selector release paths also match the audited outputs. This establishes a bounded omission repair; it neither promotes the candidate as a production policy nor establishes correctness for arbitrary source or boundary problems.

A separately frozen 32^2 one-step fixture has zero momentum, all depths below 1 mm, a varying thin layer and a parabolic bed. The original candidate launches no F/C blocks; dense

generates nonzero momentum, with maximum stage differences 3.9221e-8 and 4.0182e-8 m^2/s. Positive-depth selection launches all four blocks and restores bitwise stage agreement. The default policy also agrees across paths on this fixture. Thus the mechanism does not require pre-existing thin-layer momentum. This is a numerical counterexample, not a physical accuracy benchmark or a restart of the Thacker trajectory.

Pilot postprocessing failures and their corrected runners are retained in the reproduction records. The final report verifies stage hashes and release binding before marking completion; failed numerical records are preserved.

S34 Complete protocol on a non-flat moving shoreline

This application uses the default translating Thacker basin with an analytical reference, not the failed thin-momentum policy. Radius is 10 m, displacement amplitude 2 m and depth scale 1 m. The cell-centre parabolic bed extends across all stored cells; boundaries reflect and Manning friction is zero. The fixed 40 m common domain has 256^2 cells at dx=dy=0.15625 m. Centred dry extensions increase stored widths to 512 and 768 without changing the common input, spacing or alignment. All cells remain valid. Eight continuing quarter-period outputs span two full periods (28.370067596 s at the final float32 target). T2, second-order space, CFL 0.25, dt_max=0.05 s and dt_min=1e-6 s are fixed.

The choice of resolution and acceptance bounds uses already observed refinement results from S30. This is a new controlled performance application, not a blind accuracy discovery. Workload-specific bounds are depth L1 <=0.005 m, each momentum L1 <=0.01 m^2/s, strict 1/10 mm wet discrepancy <=10%, finite nonnegative fields and relative volume error <=5e-5. Norms are always evaluated on the same 40 m common crop; added dry storage cannot dilute the errors. These bounds do not replace the retained Ritter or external-reference failures, nor define a universal flood-prediction tolerance.

References use 16×16 Gauss–Legendre cell quadrature at the actual float32 output times. Increasing to 32×32 changes no 1 mm or 10 mm wet classification at any output; maximum cell-depth difference is 7.7293e-6 m. The finite reference quadrature is therefore checked separately from solver refinement.

Before timing, dense, F, C, Cf and R are compared over all three stored widths. Six public/residual captures per adaptive step, actual dt, common crop and zero exterior are checked. F/C/Cf/R flags are checked against independent wet-block dilation; compacted coordinates are checked against those flags. F/C/Cf retain the same selected support across sizes; R deliberately restores full launch arrays after compacted preparation. Release outputs and steps are bound to the corresponding audit outputs. Instrumentation is absent from ordinary timings.

Three serial process invocations each contain one warmup and five paired repeats of F/C/Cf/R at each width, with randomized size/path order within repeat and telemetry every 500 ms. Complete synchronized advancement includes selection, readback, preparation, transport and finalization, excluding construction and downloads. Both monotonic and raw clocks are retained. Intervals condition on this session and use 10,000 process/repeat bootstrap resamples; they do not add cross-date evidence or isolate GPU architecture. Analysis and raw paired values are reproduced by `experiments/analysis/analyze_nonflat.py`.

All 15 trajectories pass 17,496 captures each (262,440 capture positions). They use 2916 adaptive steps per trajectory. Cumulative selected support is 645,007 block-stage updates (repeated blocks count at each stage) across the 5832 transport stages, invariant under dry extension. All 216 timed/warmup trajectories pass the saved-state, step, support and accuracy guards. This is a single additional same-day session.

Maximum common-crop depth/momentum L1 is 0.000783/0.001847 in the corresponding units; maximum strict 1/10 mm discrepancy is 5.15/1.40 %. Maximum relative volume error is 3.8441e-10. The moving shoreline and the two reporting cutoffs are shown in Fig. 2; no error is

normalized by the enlarged stored domain.

Table S43: Complete RAW-clock Thacker times and paired effects through two periods. Absolute columns show median-of-process-medians milliseconds. Contrasts are paired before summarization and include conditional 95 % intervals; differences of independently summarized absolute medians need not equal paired differences.

Width	F ms	C ms	Cf ms	R ms	F-C ms [95 %]	R-C ms [95 %]	R-F ms [95 %]
256	290.2	467.7	374.4	495.4	-181.2 [-187.7, -140.9]	25.8 [24.1, 30.1]	209.4 [165.9, 214.3]
512	425.1	541.3	447.0	637.8	-116.4 [-125.0, -89.1]	95.9 [91.2, 113.8]	212.5 [201.5, 216.4]
768	537.2	630.9	571.2	748.8	-94.4 [-97.3, -75.9]	118.3 [114.3, 123.1]	213.0 [194.3, 220.0]

F/C intervals are below one at every stored size. C/Cf intervals are above one, while F/Cf remains below one: fusion helps compaction but does not reverse the filtered-path choice. R-C grows from 25.8 to 118.3 ms; R-F remains larger at each endpoint. The latter includes redundant preservation and execution interactions, so the contrasts do not supply independently additive kernel costs.

The S33 delivery includes six compressed, byte-preserving stage/dt/public/residual hash sequences. CPU replay compares corrected D/F/C against original dense at every frame, checks both original sparse prefixes and directly rechecks the first-difference arrays. Digests bind compressed and original bytes.

S35 Absolute costs for the eight-size scan

The nine original process records are included alongside the promoted scan summaries. `render_platform_figures.py` verifies each record against the hash in its frozen summary, reconstructs whole-trajectory paired ratios, and checks those estimates and intervals against the published scan summaries before computing absolute costs and signed differences. Warmups are excluded. All three continuing windows stay together. Each estimate is a median of process medians; paired differences are formed before summarization. Conditional 95 % intervals resample three processes and five paired repeats with 10,000 draws. The replay exports all process difference medians and N, K, cumulative A and 256A/(2KN) in `figures/scan_costs.json` and draws them in the main scan figure.

Table S44: Eight-size absolute complete times and paired F-C on CLOCK_MONOTONIC_RAW, in milliseconds. Absolute columns are summarized independently, so their subtraction need not equal the paired contrast. Sessions remain separate.

Session / input / width	F ms	C ms	Paired F-C ms [95 %]
4090 session 1 / Continuous / 512	26.9	32.5	-5.8 [-11.6, -4.7]
4090 session 1 / Continuous / 640	31.7	36.7	-5.0 [-14.9, -4.8]
4090 session 1 / Continuous / 768	35.2	41.4	-6.1 [-10.5, -4.6]
4090 session 1 / Continuous / 896	40.9	46.3	-5.5 [-7.5, -4.4]
4090 session 1 / Continuous / 1024	46.0	53.0	-6.1 [-8.2, -5.3]

Table S44 continued.

Session / input / width	F ms	C ms	Paired F-C ms [95 %]
4090 session 1 / Continuous / 1152	56.8	64.2	-7.2 [-7.7, -6.1]
4090 session 1 / Continuous / 1280	75.3	78.1	-3.0 [-4.1, -1.6]
4090 session 1 / Continuous / 1536	110.6	115.6	-4.3 [-5.4, -3.2]
4090 session 1 / Fragmented / 512	32.1	40.5	-8.6 [-12.8, -6.8]
4090 session 1 / Fragmented / 640	34.7	42.9	-7.1 [-10.1, -6.3]
4090 session 1 / Fragmented / 768	38.8	47.2	-7.2 [-10.7, -6.5]
4090 session 1 / Fragmented / 896	44.1	52.0	-7.7 [-9.9, -5.9]
4090 session 1 / Fragmented / 1024	50.4	57.7	-7.2 [-10.6, -3.8]
4090 session 1 / Fragmented / 1152	62.7	68.5	-6.6 [-9.2, -3.9]
4090 session 1 / Fragmented / 1280	80.5	84.8	-4.9 [-8.6, -3.1]
4090 session 1 / Fragmented / 1536	118.1	124.7	-6.4 [-7.4, -5.3]
4090 session 2 / Continuous / 512	27.8	32.0	-4.9 [-6.2, -3.8]
4090 session 2 / Continuous / 640	31.7	35.6	-5.1 [-6.7, -2.9]
4090 session 2 / Continuous / 768	32.9	39.0	-6.1 [-6.3, -4.8]
4090 session 2 / Continuous / 896	41.0	46.5	-5.5 [-5.5, -3.9]
4090 session 2 / Continuous / 1024	44.2	51.4	-6.2 [-6.4, -6.0]
4090 session 2 / Continuous / 1152	53.7	64.1	-7.4 [-10.7, -5.9]
4090 session 2 / Continuous / 1280	75.2	78.9	-3.6 [-3.8, -0.5]
4090 session 2 / Continuous / 1536	109.7	116.0	-5.1 [-6.6, -2.6]
4090 session 2 / Fragmented / 512	31.1	37.1	-7.2 [-8.0, -5.5]
4090 session 2 / Fragmented / 640	34.1	41.3	-7.1 [-7.2, -6.0]
4090 session 2 / Fragmented / 768	38.2	44.9	-6.8 [-8.0, -6.0]
4090 session 2 / Fragmented / 896	43.3	51.3	-6.5 [-9.7, -4.4]
4090 session 2 / Fragmented / 1024	49.7	56.7	-7.0 [-7.1, -6.1]
4090 session 2 / Fragmented / 1152	59.7	67.0	-7.5 [-7.8, -6.0]
4090 session 2 / Fragmented / 1280	79.8	83.3	-4.0 [-6.3, -3.4]
4090 session 2 / Fragmented / 1536	117.2	122.2	-5.5 [-6.2, -4.7]
5090 / Continuous / 512	23.2	32.9	-9.8 [-11.7, -9.1]
5090 / Continuous / 640	25.5	34.4	-8.9 [-10.8, -8.2]
5090 / Continuous / 768	29.7	37.2	-7.7 [-9.6, -7.2]
5090 / Continuous / 896	33.3	44.1	-10.7 [-11.0, -9.9]
5090 / Continuous / 1024	36.7	46.9	-10.1 [-11.1, -9.9]
5090 / Continuous / 1152	42.6	53.0	-10.5 [-11.1, -9.6]

Table S44 continued.

Session / input / width	F ms	C ms	Paired F-C ms [95 %]
5090 / Continuous / 1280	48.2	57.4	-9.3 [-9.6, -9.0]
5090 / Continuous / 1536	75.4	78.3	-2.8 [-3.3, -1.8]
5090 / Fragmented / 512	26.8	36.1	-9.4 [-10.7, -9.3]
5090 / Fragmented / 640	29.7	38.4	-8.8 [-10.5, -8.7]
5090 / Fragmented / 768	33.4	42.2	-8.7 [-10.3, -8.7]
5090 / Fragmented / 896	36.7	49.6	-12.1 [-13.0, -11.4]
5090 / Fragmented / 1024	41.1	53.0	-11.9 [-12.2, -11.5]
5090 / Fragmented / 1152	46.9	58.7	-12.2 [-12.3, -11.7]
5090 / Fragmented / 1280	53.3	64.2	-11.0 [-11.2, -9.4]
5090 / Fragmented / 1536	81.7	85.8	-4.1 [-4.4, -3.6]

S36 Spatial source-policy sensitivity

The main paper retains the default-policy view; this figure preserves all three numerical cutoffs against the same strict diagnostic threshold and fixed external reference. Selecting the best-looking cutoff after comparison does not certify its numerical accuracy.

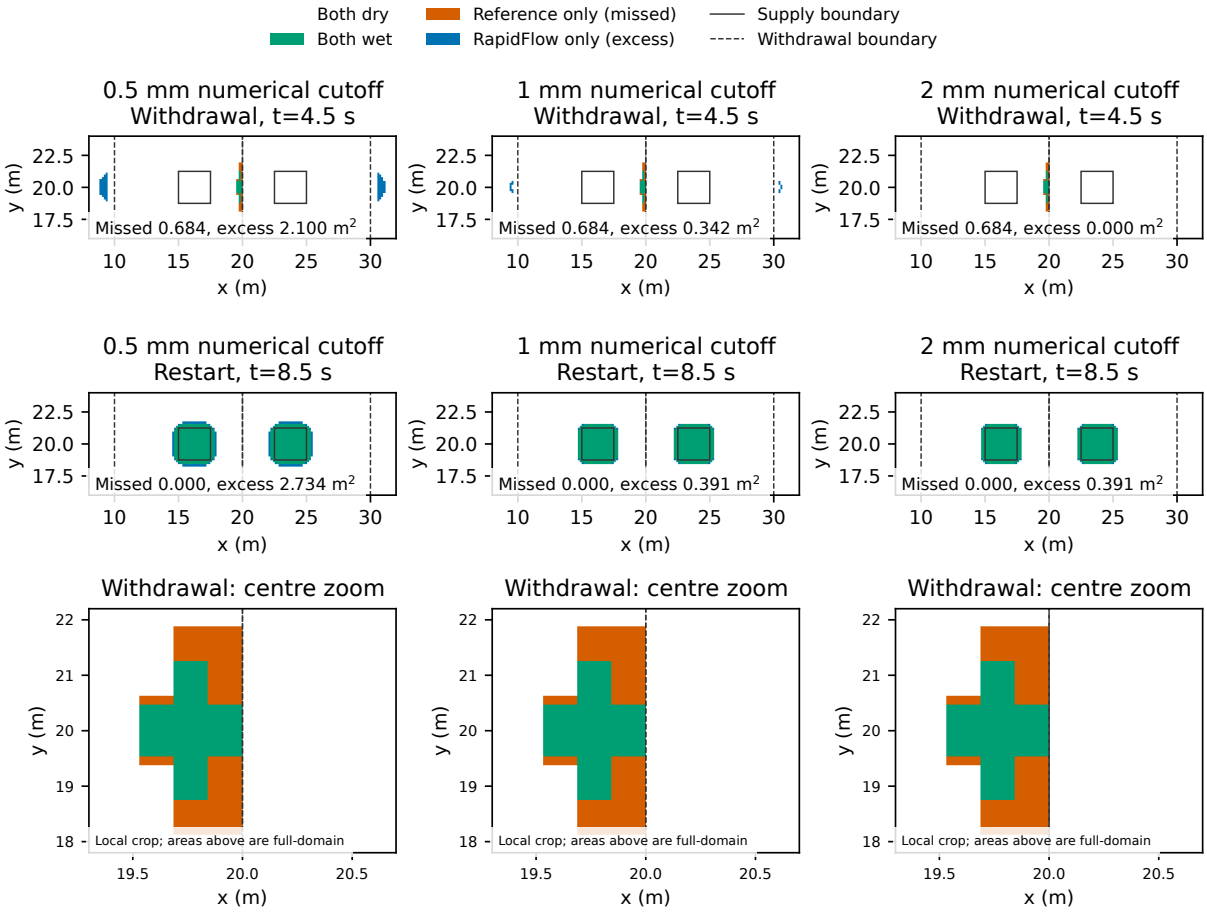


Figure 25: Paired-source wet masks at 4.5 and 8.5 s on 256^2/T2, comparing each numerical cutoff with the same finite ANUGA reference (ANUGA developers, n.d.) at the strict 1 mm reporting threshold. Colours separate agreement, missed wet area and excess wet area; dry surroundings are cropped. Solid/dashed lines mark supply/ withdrawal boundaries (parts of the latter lie outside the crop). Rows denote withdrawal/restart; the third row enlarges the central withdrawal front with independently scaled axes. Area labels remain full-domain values. The 2 mm policy has no matched ANUGA source-closure control.