Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 13 Issue 4, April - 2026

GPU-Accelerated Lattice Boltzmann Thermal Simulation Of Sand Energy Storage With AI-Based Volumetric Prediction

Antonis P. Papadakis[1,2]*, Sofia Nikolaidou[1], Varnavas Mikrommatis[1]
[1]KYAMOS LTD, 37 Polyneikis Street, Strovolos, 2047, Nicosia, Cyprus
[2]Frederick University, 7. Y. Frederickou, Pallouriotisas, 1036, Nicosia, Cyprus

Abstract—This paper presents a GPU-accelerated Lattice Boltzmann thermal solver with a multiple-relaxation-time collision operator (LBM-MRT) for sand-based thermal energy storage, coupled with a three-dimensional volumetric AI surrogate for rapid thermal prediction. The D3Q19 MRT solver is implemented in CUDA C++ with CUDA-aware MPI and assessed against four benchmark problems. A campaign of 43 scenarios was executed at 96^3 resolution (884 736 cells) on a Tesla V100 InfiniBand cluster. A direct comparison between a single central heater pipe and a 3×3 array of nine distributed heaters at the same total power reduces the peak internal temperature rise from approximately 712 K to 188 K, more than 70 %, while producing a more uniform heated volume; we note that the two configurations also differ in total heater volume and surface area, so the reduction combines spatial distribution with a lower per-pipe volumetric source density. The three-dimensional volumetric surrogate TUNet-AI achieves 0.661 K MAE in a single forward pass at < 5 ms latency, and 0.636 K MAE with eight-way test-time augmentation at approximately 40 ms, on an offline benchmark using per-scenario min/max Normalization from the full scenario trajectory. Results are obtained under a pure-conduction model with fixed material properties and should be read as comparative design trends for the modelled configuration.

Keywords—Lattice Boltzmann Method; sand thermal energy storage; GPU computing; CUDA-aware MPI; U-Net; TUNet-AI; distributed heating; thermal prediction

I. INTRODUCTION

Sand-based thermal energy storage is attractive because it combines high thermal stability (above 1700 °C), long cycle life, and low material degradation, together with abundance and the absence of rare or toxic elements, as demonstrated in industrial deployments by Polar Night Energy [1]. Laboratory-scale studies have further characterized desert sand as a candidate storage medium [2] and investigated the enhancement of effective thermal conductivity through mixed-material bed designs [3]. Design optimisation of sand-storage vessels requires a quantitative understanding of how heat propagates from embedded resistance heaters, how temperature distributions evolve across charging and discharging, and how heater placement affects storage uniformity. Addressing these issues requires high-fidelity, multi-material thermal simulation, for which the packed-bed TES literature has developed a range of numerical approaches [4], [5].

The Lattice Boltzmann Method [6], [7] offers an attractive algorithmic path for such simulations: its explicit, local update rule maps naturally onto GPU architectures and enables massive parallelism with modest communication cost [8]. Combined with AI-based surrogate models for the thermal field, which have recently been explored for conduction and multi-physics problems using U-Net, neural-operator, and physics-informed architectures [9], [10], [11], the approach delivers both the physics fidelity needed for design and the millisecond inference latency needed for real-time monitoring and for integration into closed-loop dispatch controllers.

In this work, we present a GPU-accelerated LBM-MRT thermal solver for sand storage, assess it against a standard battery of benchmarks, execute a three-dimensional scenario campaign, including a distributed-heater optimisation study, and demonstrate an AI surrogate stack that achieves sub-Kelvin mean-absolute-error with a four-orders-of-magnitude speedup over the direct solver. Throughout, results are presented as comparative design trends within an explicit pure-conduction modelling scope; the inclusion of radiative transport, temperature-dependent properties, and a reformulation of the enthalpy-conservative solver are identified as future work. The paper is structured as follows: Section II describes the LBM methodology; Section III presents the sand-storage domain, materials and scenario matrix; Section IV defines the solver validation; Section V presents the three-dimensional simulation results, including skin-surface and distributed-heater behaviour; Section VI describes the AI surrogate; and Section VII includes the conclusions.

II. LATTICE BOLTZMANN METHODOLOGY

A. D3Q19 MRT formulation and solver context

The underlying continuum model is the transient heat equation $\rho C_p \partial T/\partial t = \nabla\cdot(k\nabla T) + q_{vol}$, where q_{vol} is the volumetric Joule source in heater cells and zero elsewhere. The solver implements this model through a D3Q19 Lattice Boltzmann scheme with a Multiple Relaxation Time (MRT) collision operator [7]. MRT transforms populations to moment space, relaxes non-conserved moments independently, and transforms back, providing improved numerical stability over the single-relaxation BGK operator at the high conductivity contrasts present in the sand–steel–insulation–heater composite (ratio 10 000:1 between heater copper-alloy and insulation). For the thermal population set, the equilibrium distribution takes the standard scalar form

g_k^eq = w_k T with D3Q19 weights w_k, and thermal diffusivity is controlled through the single scalar moment relaxation corresponding to the energy flux; the full moment matrix and equilibrium construction follow Lallemand and Luo [7]. The LBM framework used here is the general-purpose KYAMOS thermal–flow code: the formulation is retained because its explicit, strictly local update rule has excellent GPU locality, because the same framework is reused across flow, conduction, and coupled problems, and because the MRT collision operator gives stable behaviour at the large inter-material conductivity ratios of the sand-storage composite.

B. Thermal solver, volumetric source and per-material relaxation

The thermal field uses a separate D3Q19 population set. For sand storage every cell of the domain is tagged as solid, which disables the Navier–Stokes flow branch and reduces the coupled problem to a pure-conduction transport with internal heat sources. The lattice diffusivity follows $\alpha_{LB} = c_s^2(\tau_T - 1/2)$ and is mapped to the physical diffusivity $\alpha = k/(\rho C_p)$ per material region via $dt_{phys} = \alpha_{LB} \cdot dx^2 / \alpha$. The solver uses a single global physical time step dt_phys, set by targeting $\tau_T = 0.8$ in the sand region at dx = 5.2 mm, giving dt_phys = 11.75 s. Each 96^3 scenario at 5,000 steps therefore simulates approximately 16.3 hours of physical time. Because dt_phys is fixed globally by the sand target, the other materials operate at their own (larger) lattice diffusivities and relaxation times, summarized in Table I. In particular, the heater pipes operate at $\tau_T \approx 151$. No per-region time stepping or special stabilization treatment is implemented in the present solver; empirical stability was observed for all reported benchmark and scenario cases.

Table I. Per-material lattice Boltzmann parameters at dx = 5.21 mm and global dt_phys = 11.75 s.

Material	α_physical (m²/s)	α_LB	τ_T	Regime
Sand	2.31×10^{-7}	0.100	0.800	Target regime
Steel shell	1.32×10^{-5}	5.70	17.6	Large τ, usable
Insulation	6.67×10^{-7}	0.289	1.37	Normal
Heater pipe	1.16×10^{-4}	50.2	151.2	Very large τ, no special treatment

Joule heating is applied as a cell-by-cell volumetric source $\Delta T_{src} = (q_{vol} / \rho C_p) \times dt_{phys}$, distributed across the thermal populations as $g_k \mathrel{+}= w_k \times \Delta T_{src}$ in heater cells only; the source discretisation follows a standard local-forcing philosophy [12]. The present storage model is intentionally a pure-conduction representation with fixed material properties. Temperature-dependent thermal conductivity and heat capacity, inter-particle and bed-to-boundary radiative transfer, contact resistance, and gas convection in the pore space are not included. The results should therefore be interpreted as comparative trends for the modelled geometry and materials rather than as a complete high-temperature system model.

C. GPU implementation and MPI

The solver is implemented in CUDA C++ with one thread per cell. CUDA-aware MPI enables direct GPU-to-GPU halo exchange over InfiniBand. MPI consistency is verified by direct comparison of multi-rank against single-rank execution on identical problems, yielding bitwise equality on deterministic test cases. Per-scenario wall-clock cost at 96^3 is approximately 33 s.

III. SAND-STORAGE MODEL

A. Domain, materials and geometry

The storage cell is a 0.5 m cube resolved on a 96^3 Cartesian grid (884 736 cells). Four material regions are defined analytically, stacked radially from the outside in: the insulation layer is the outermost two-cell band, the steel shell is the next four cells inward (approximately 21 mm thick), and the sand fill occupies the remaining interior volume. One or more cylindrical vertical heater pipes are embedded within the sand volume. Material properties and per-material LBM parameters appear in Tables I and II. Figure 1 overlays the vessel geometry, the heater pipe and the thermal field in a single cutaway view, making explicit how the material layout organises the temperature distribution: the heater supplies the central source, the sand conducts outward, and the shell–insulation system moderates the external boundary.

Table II. Material properties of the 3D sand-storage model.

Region	Material	k (W/m·K)	ρC_p (MJ/m³·K)	Geometry
Insulation	Insulation (outermost)	0.04	0.06	Outer 2-cell band
Steel shell	Steel (beneath insulation)	50	3.80	Next 4-cell band inward
Sand	Sand (effective)	0.30	1.30	Interior bulk fill
Heater pipe	Heater	400	3.45	Cylindrical, vertical, in sand

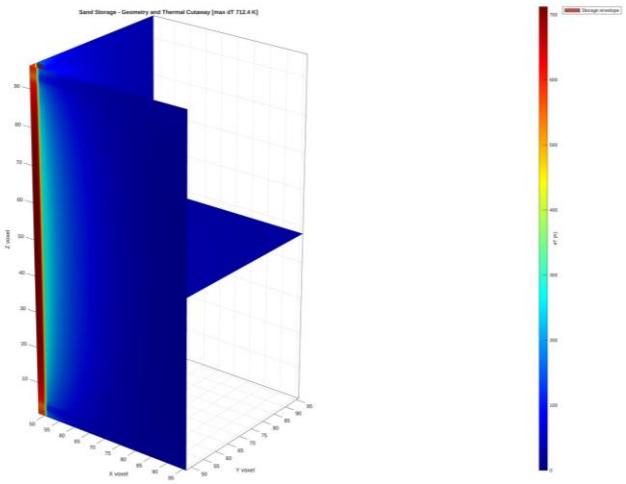

Fig. 1. Geometry and thermal-field overlay for the nominal charging case T01. The material layout organises the temperature distribution: the heater supplies the central source, the sand conducts outward, and the shell–insulation system moderates the external boundary.

B. Heater configurations

Three heater arrangements are compared at the same total power of 353.4 W: a single central pipe (T01, pipe radius 15 mm), a 2×2 four-pipe array (T41, pipe radius 15 mm each), and a 3×3 nine-pipe array (T42, pipe radius 10 mm each). Table III reports the geometric parameters that distinguish these configurations. The total heater cross-sectional area and the per-pipe volumetric source density q_vol differ substantially between configurations: at fixed total power, the 3×3 array has four times the total heater cross-sectional area of the single-pipe case, which reduces the per-pipe q_vol by a factor of four. The peak-temperature reduction reported in Section V-G therefore combines the effect of spatial distribution with the effect of reduced per-pipe source density; disentangling the two effects would require a separate study at matched heater volume, which is identified as future work. A further scenario T43 repeats the 3×3 geometry at twice nominal total power.

Table III. Heater-configuration geometry at fixed total power of 353.4 W.

Scenario	Pipes	Radius (mm)	Total cross-section (mm^2)	q_vol (W/m^3)	Power per pipe (W)
T01(1pipe)	1	15	707	1.00×10^6	353.4
T41 (2×2)	4	15	2 827	2.50×10^5	88.4
T42 (3×3)	9	10	2 827	2.50×10^5	39.3

C. Scenario Matrix

Table IV defines the nine scenarios used throughout Sections V and VI. The matrix includes four single-pipe cases (nominal charging T01, amplified charging T03, ambient-cooled charging T04, and the charge-to-relaxation transition T40), two Gaussian-discharge cases (T08 at 600 K peak, T36 at 450 K peak), and three distributed-heater cases (T41 four-pipe, T42 and T43 nine-pipe at one and two times nominal power).

Because T40 has adiabatic boundary conditions and no external heat sink, the "relaxation" phase after source switch-off is internal redistribution of stored heat rather than net discharge to ambient; the total enthalpy remains constant during that phase. All standard scenarios run for 5 000 LBM steps (16.3 h physical time at 96^3); T40 runs for 10 000 steps (32.6 h) with the heater source switched off after step 4999.

D. Grid-resolution study

Two grid resolutions were evaluated before the full campaign: 64^3 and 96^3. Table V reports the comparison at step 5000 for the nominal charging configuration. Because the physical timestep is set by the lattice diffusion constraint and scales with dx^2, the 64^3 run at step 5000 corresponds to 36.7 h of physical time while the 96^3 run corresponds to 16.3 h; the comparison is therefore at matched simulation steps rather than at matched physical time. The factor-of-two peak-ΔT difference between grids reflects both the longer physical integration time at 64^3 and a numerical concentration of the voxelised heater source on the coarser grid, and should not be read as a grid-convergence error in the usual sense. At 64^3 the cylindrical heater and near-heater gradients are only weakly resolved and the conductive front appears artificially broadened by grid diffusion (Fig. 2a). At 96^3 the heater cross-section is reproduced cleanly, the radial gradient is sharp, and the thermal field retains the expected cylindrical symmetry (Fig. 2b). The 96^3 configuration was selected as the coarsest resolution that preserves the essential charging physics at acceptable computational cost for the present comparative study. A matched-physical-time convergence study against a 128^3 reference is identified as future work.

Table IV. Scenario definitions used in Sections V and VI.

ID	Heater configuration	Total power (W)	Initial / BC	Description
T01	1 pipe, r = 15 mm, central	353.4	300 K uniform; adiabatic walls	Nominal charging
T03	1 pipe, r = 15 mm	1 767.1	300 K uniform; adiabatic walls	Amplified (5×) charging
T04	1 pipe, r = 15 mm	353.4	300 K uniform; 300 K Dirichlet walls	Charging with ambient-cooled outer boundary
T08	No heater	0	Gaussian peak 600 K, σ = 0.08 m; 300 K Dirichlet walls	Discharge from pre-charged Gaussian
T36	No heater	0	Gaussian peak 450 K, σ = 0.08 m; 300 K Dirichlet walls	Moderate-temperature discharge
T40	1 pipe, active steps 0–4999 then off	706.9 → 0	300 K uniform; adiabatic walls; 10 000 total steps	Charge-to-relaxation transition (adiabatic; source off at step 5000)
T41	4 pipes 2×2, r = 15 mm each	353.4	300 K uniform; adiabatic walls	4-pipe distributed, nominal total power
T42	9 pipes 3×3, r = 10 mm each	353.4	300 K uniform; adiabatic walls	9-pipe distributed, nominal total power
T43	9 pipes 3×3, r = 10 mm each	706.9	300 K uniform; adiabatic walls	9-pipe distributed, 2× nominal total power

Table V. Grid-comparison data at step 5000 for the nominal charging configuration (T01). Physical times differ between grids because dt_phys ∝ dx².

Grid	dx (mm)	Peak ΔT (K)	Stored enthalpy (MJ)	Runtime (s)	Mem (MB)
64³	7.81	1225.6	15.68	11.2	395
96³	5.21	712.4	8.30	32.6	503

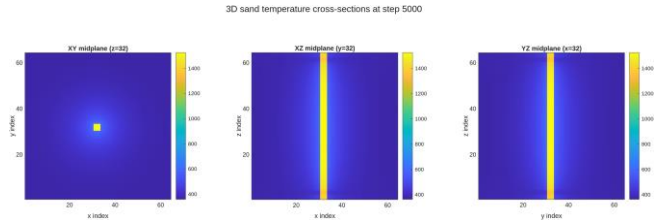

Fig. 2a. Cross-sections of the nominal charging case at 64³ resolution. Interfaces are broadened by grid diffusion.

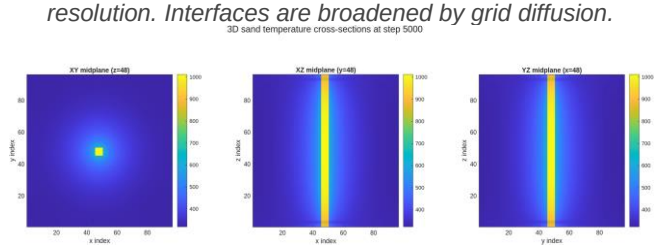

Fig. 2b. Cross-sections at 96³ resolution. The heater cross-section is reproduced cleanly, the radial gradient is sharp, and the expected cylindrical symmetry is preserved.

IV. SOLVER VALIDATION

Four benchmarks were executed, summarised in Table VI: a Poiseuille channel (flow branch, unused in the sand-storage configuration but reported for completeness), a one-dimensional thermal source problem (diffusion + source term + MPI halo exchange), a heated cavity at Ra = 0 (pure conduction, Nu = 1), and a heated cavity at Ra = 10^4 (coupled buoyancy [13]). The thermal-source 1D benchmark is particularly informative because it exercises the volumetric source term, diffusive transport, Dirichlet boundary conditions, and MPI halo exchange simultaneously across rank boundaries. Because the benchmarks probe different parts of the solver, the reported metrics are heterogeneous and should be interpreted per benchmark rather than collapsed into a single percentage claim.

Table VI. Benchmark validation summary.

Benchmark	Validates	Metric	Result
Poiseuille channel	Flow profile + wall BC (unused in sand model)	NRMSE	3.4 %
Thermal source 1D	Diffusion + source + MPI halo	L2	1.73×10^{-4}
Cavity Ra = 0	Pure conduction (Nu = 1)	Nu error	0.008 %
Cavity Ra = 10^4	Coupled buoyancy [13]	Nu error	0.4 %

Note on total-enthalpy balance

The present solver evolves temperature using the volumetric source expression q_vol/(ρC_p) applied cell-wise, rather than an enthalpy-conservative formulation. A previously reported 0.01 % residual corresponds to a temperature-sum / source-consistency check comparing Σ(q_vol/ρC_p) · dt against Σ ΔT, rather than a full enthalpy balance of the form ∫ ρC_p · ΔT dV against ∫₀ᵗ ∫ q_vol dV · dt'. A direct full-domain enthalpy evaluation for the adiabatic nominal charging case T01 at step 5000 yields a stored enthalpy of 8.30 MJ (full domain, including sand, steel shell, insulation, and heater regions) against an injected enthalpy of 19.93 MJ (gridded volumetric source integrated in time), for a ratio E_stored / E_injected ≈ 0.416 at 96³. The corresponding ratio on the 64³ grid is 0.432. The sand region alone accounts for 5.96 MJ at 96³. The specific numerical origin of this ratio is under investigation and is likely associated with the interaction between the temperature-based source expression and material-dependent ρC_p across sharp heterogeneous interfaces; resolving it is the top priority of the identified enthalpy-conservative reformulation. The scenario-level thermal fields reported in Section V and the qualitative comparative findings remain valid as comparative design trends within the stated pure-conduction modelling scope, but absolute stored-energy magnitudes should be read with the above ratio in mind.

V. THREE-DIMENSIONAL SIMULATION RESULTS

The full 3D campaign comprises 43 scenarios at 96³ resolution, executed on a Tesla V100 cluster with two MPI ranks. The matrix includes charging, discharge, mixed-mode, and distributed-heater configurations. The campaign completes in approximately 12 minutes of wall-clock time (33 s per scenario; 503 MB of GPU memory per rank).

A. Nominal charging — orthogonal midplanes

Figure 3 shows the three orthogonal midplanes of the nominal charging case T01 at the end of the charging window. The XY plane reveals the expected radial symmetry of the temperature field around the cylindrical pipe. The XZ and YZ planes show the heater as a hot vertical column extending through the full height of the domain, with radial heat propagation visible on either side. The field is consistent with a vertically extended line source in a conduction-dominated medium: the steepest gradients sit immediately adjacent to the pipe, and the temperature field decays outward with approximate cylindrical symmetry.

B. Multi-isosurface view of the charging field

Figure 4 displays multiple temperature isosurfaces simultaneously for the T01 case, producing a nested set of shells around the heater. The shells read as a radial thermal hierarchy: the innermost shell isolates the strongly heated core in the immediate neighbourhood of the pipe, while successive outer shells capture the conductive spread of energy through the sand toward the steel shell and insulation. The finite radial reach of the hottest contour is a direct measure of how far the strongest thermal penetration has progressed during the simulated charging window.

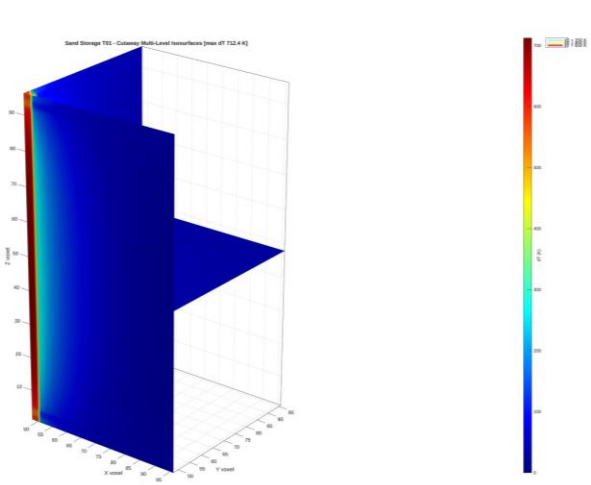

Figure 4: Nominal charging T01 — nested temperature isosurfaces at several levels. The concentric thermal hierarchy reads from the heater core outward, showing the radial layering of stored energy.

C. Fixed-scale jet cutaway

Figure 5 presents a fixed-scale jet cutaway of the final T01 state. Unlike the locally rescaled views of Figures 3 and 4, the fixed colour scale preserves the true relationship between the global maximum and the rest of the field. Under this view the heater neighbourhood reaches very high temperatures while most of the vessel volume remains substantially cooler, making explicit that the simulated charging interval captures the early-to-intermediate conductive front rather than full-vessel thermal saturation.

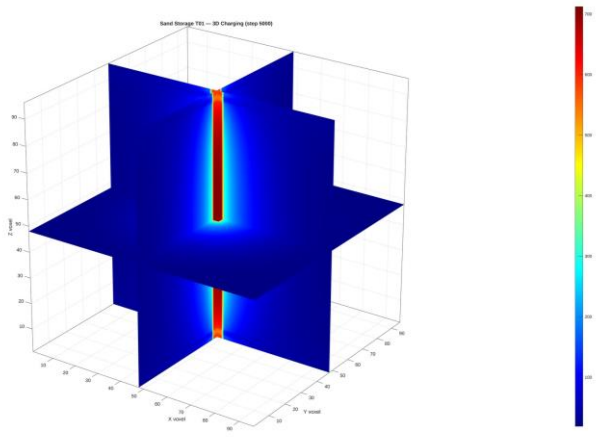

Fig. 3. Nominal charging T01 at step 5000 — orthogonal XY, XZ and YZ midplanes. The cylindrical heater pipe is visible as a hot vertical column; heat spreads radially into the sand with approximate cylindrical symmetry.

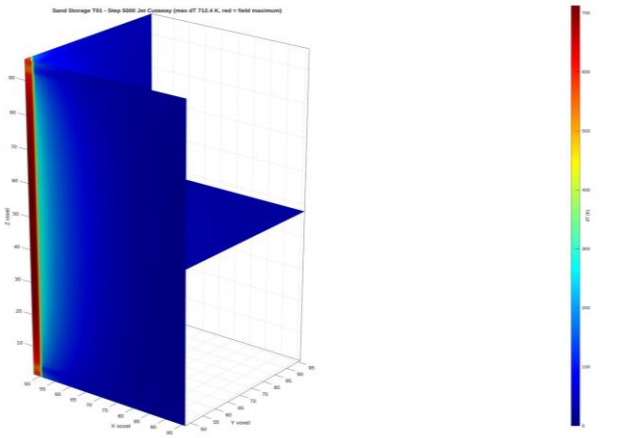

Figure 5: Nominal charging T01 — fixed-scale jet cutaway of the final state. The preserved global colour scale shows that intense heating is localized near the heater while most of the vessel volume remains at substantially lower temperatures.

D. Amplified charging — temporal evolution

Figure 6 shows the temporal evolution of charging on the XY midplane (scenario T03, 5× nominal power). The thermal front expands monotonically with time and retains a nested concentric structure throughout, demonstrating diffusion-dominated storage rather than advective transport, the expected result for an all-solid sand domain. The stronger source drives the local temperature up faster and pushes the conductive front outward more rapidly than at nominal power, but the cylindrical symmetry of the field is preserved.

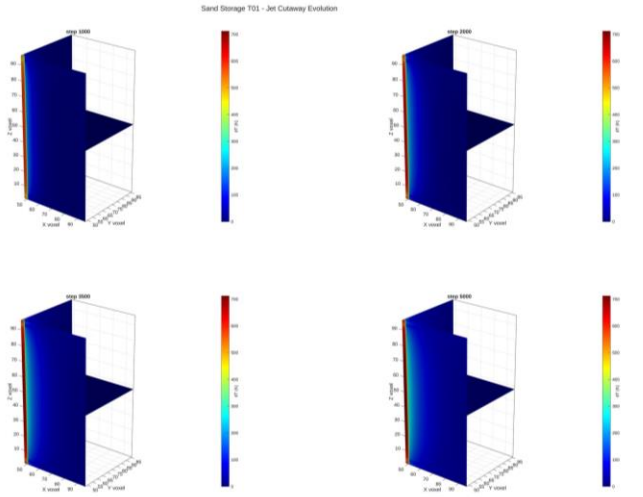

Figure 6: Temporal evolution of charging on the XY midplane. The thermal front expands monotonically and retains a nested concentric structure throughout, demonstrating diffusion-dominated transport.

E. Discharge behavior and outside-in cooling

Figure 7 shows the discharge case T08, initialized from a smooth radially symmetric Gaussian charged state with a 600 K peak (σ = 0.08 m) above a 300 K baseline and evolving without active heating. The outer boundary is a 300 K fixed-temperature Dirichlet, so the vessel discharges by passive conductive loss through the sand and structural layers to the ambient environment. The three orthogonal

midplanes are shown at three timepoints. The field follows the expected outside-in cooling pattern: the outer region relaxes toward ambient first, while the center remains warmer for longer because stored heat must diffuse through the surrounding sand and the structural layers before it can leave the system. The persistence of the warm core is an indicator of residual internal thermal energy during the discharge window.

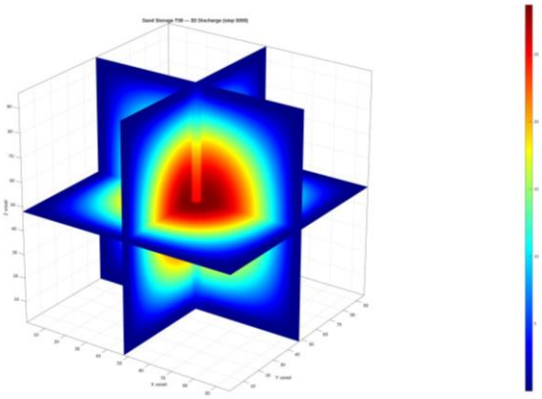

Figure 7: Discharge T08 — orthogonal XY, XZ, YZ midplanes at representative discharge times. Heat is lost outside-in; the warm core persists longest at the center of the domain.

F. External skin-surface response — observability decoupling

Figure 8 presents the insulation-skin temperature fields across the charging (T01), discharge (T08), and nine-pipe distributed-heating (T42) cases on matched cube views. Three observations emerge directly from the figure. In the charging case the external skin carries a localised rise that is attenuated relative to the interior heater region (peak $\Delta T \approx 375$ K at the pipe footprint). In the discharge case the exterior skin has returned essentially to ambient (peak $\Delta T \approx 0.5$ K) while the interior still holds stored energy. In the nine-pipe distributed case the skin carries a much smaller peak rise ($\Delta T \approx 115$ K) and a flatter external footprint than the single-pipe case.

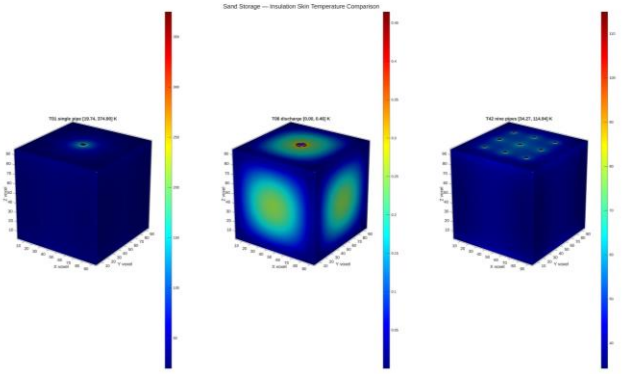

Figure 8: External insulation-skin temperature fields for nominal charging T01 (left, ΔT up to ≈375 K at the pipe footprint), discharge T08 (center, ΔT up to ≈0.5 K — interior still holds energy), and nine-pipe distributed charging T42 (right, ΔT up to ≈115 K, flatter footprint).

The observability-decoupling result has direct instrumentation consequences. External thermography alone is insufficient to estimate residual stored energy in these configurations and would benefit from either an inverse-model reconstruction informed by the internal geometry, or direct internal calibration with a small number of in-bed temperature probes. The flattening effect of distributed heating on the external footprint is additionally relevant to material-stress budgets and to ambient heat-loss control and reinforces the case for distributed-heater designs beyond the internal-uniformity argument of the next subsection.

G. Distributed-heater study — one pipe vs nine pipes

Figure 9 places the single-pipe (T01) and nine-pipe (T42) cases on the same cutaway view and the same colour scale, so the difference in transport physics can be read directly. The single heater produces a sharply concentrated peak temperature rise of approximately 712 K and a steep radial gradient, while the nine-pipe arrangement produces a peak ΔT of approximately 188 K — more than 70 % reduction — and a broader, more spatially uniform heated volume. As noted in Section III-B (Table III), the two configurations share total power but differ in total heater cross-sectional area (707 vs 2 827 mm²) and per-pipe volumetric source density (1.00×10^6 vs 2.50×10^5 W/m³). The observed peak-ΔT reduction therefore combines the effect of spatial distribution with the effect of reduced per-pipe source density; isolating the spatial-distribution contribution alone would require a further comparison at matched total heater volume, which is identified as future work.

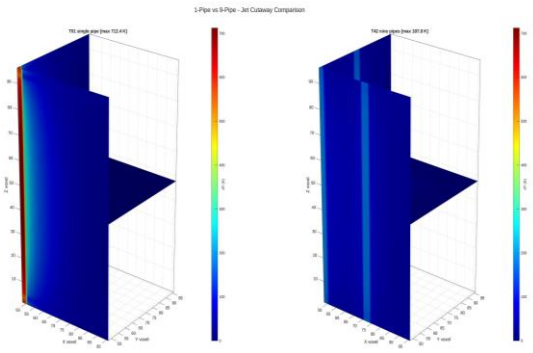

Figure 9: Direct comparison between the single-pipe case T01 (left, peak ΔT ≈ 712 K) and the nine-pipe case T42 (right, peak ΔT ≈ 188 K) at the same total heater power, on a common cutaway and a common colour scale. The nine-pipe arrangement reduces peak internal temperature rise by more than 70 % while broadening the heated volume; the two configurations also differ in total heater cross-sectional area (4×) and per-pipe source density (¼), so the reduction combines spatial distribution with lower per-pipe source density.

Within the present modelling scope, these results suggest that distributing the heat input into a larger number of lower-power pipes may reduce peak internal temperatures and improve spatial uniformity of the heated volume. As noted in Section II-B, inclusion of radiative transport and temperature-dependent thermal conductivity is expected to reduce peak temperatures in the single-pipe case

specifically, and may therefore further modify the quantitative magnitude of the 70 % figure; the qualitative design direction is expected to be preserved. These implications should be verified through coupled system-level charging and discharging studies. Figure 10 shows the 9-pipe field on orthogonal midplanes: the 3×3 pipe pattern is clearly visible in the XY plane, and the individual plumes merge into a broader, more uniform heated region across the vessel cross-section.

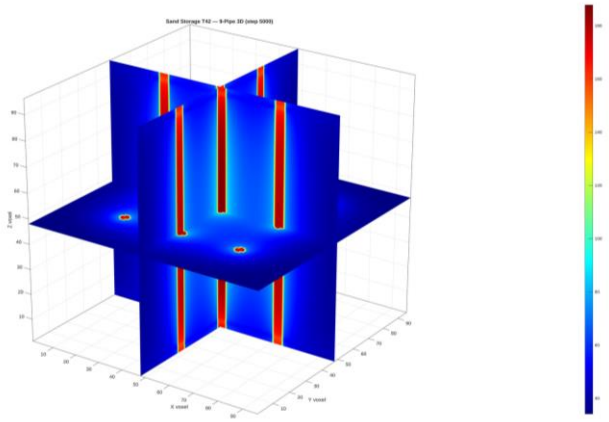

Figure 10: Nine-pipe distributed charging T42 — orthogonal midplanes. The 3×3 pipe pattern is visible in the XY plane; the individual plumes merge into a broader, more uniform heated region.

VI. AI-BASED THERMAL PREDICTION

A. Dataset construction and splits

Two linked datasets were produced by the LBM solver of Section II. The 2D dataset comprises 40 slab-geometry scenarios at 256 × 256 resolution, split 28 / 5 / 7 into train / validation / test partitions. The 3D dataset comprises 43 volumetric scenarios at 96^3 resolution, split 30 / 6 / 7 into train / validation / test partitions. The 3D test partition contains T01, T04, T08, T36, T40, T41, and T42; the validation partition contains T02, T11, T18, T25, T32, and T43. Splits are performed at the scenario level, not at the snapshot level, in order to prevent temporal leakage between neighbouring time steps.

Normalization of the 3D benchmark was performed per-scenario using minimum and maximum ΔT computed over the full available trajectory of each scenario, including for held-out test scenarios. This constitutes offline benchmark leakage: the Normalization statistics for a test scenario are computed from frames beyond what would be observable in a deployment setting. The 0.636 K TUNet-AI test MAE reported below should therefore be interpreted as an offline benchmark figure rather than a deployment-ready evaluation; a retraining run with per-scenario normalization restricted to the network's own input-window frames (or to global training-set statistics) is identified as future work.

B. 2D U-Net

A U-Net encoder–decoder [14] was trained on the 2D dataset. The input tensor combines four consecutive ΔT frames with the static q_vol map, for a total of five channels. The model predicts the temperature residual relative to an analytic baseline rather than the absolute temperature, which is important for stable training across the 60–500 K dynamic range between mild and amplified charging scenarios. On the held-out test set, the 2D U-Net achieves a mean absolute error of 0.23 K, a thermal-front intersection-over-union of 0.998 with the front defined as the set of cells satisfying $\Delta T > 0.5 \cdot \Delta T_{max}$ where ΔT_{max} denotes the spatial maximum ΔT in that snapshot, and a maximum pointwise error of 14.0 K at 5 ms inference latency per sample.

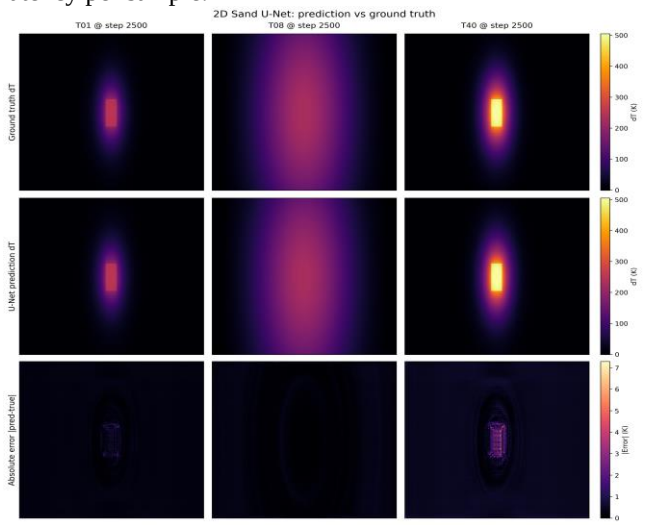

Figure 11: 2D U-Net surrogate — ground truth (top row), prediction (middle row), and absolute error (bottom row) for three representative scenarios: T01 charging, T08 discharge, and T40 charge-to-discharge transition.

C. 3D TUNet-AI

For the volumetric problem, we introduce TUNet-AI, a three-dimensional surrogate built for rapid volumetric prediction. Architectural adaptations from the 2D baseline include a fully three-dimensional convolutional encoder–decoder, group Normalization tuned for the small per-GPU batch sizes required at 96^3, an attention-augmented bottleneck that pools global information across the volume, and a composite loss function of the form $L = \alpha \cdot L_{MAE} + \beta \cdot L_{grad}$, where L_{MAE} is the per-voxel mean absolute error between prediction and ground truth and L_{grad} is a gradient-matching term that penalizes mismatches in the spatial gradients of the temperature field; the coefficients α and β are tuned so that thermal-front sharpness is explicitly rewarded during training. The total parameter count is 4.1 M, which is compact relative to recent 3D neural-operator and 3D U-Net surrogates for physics fields [15], [16].

Training is distributed across eleven Tesla V100 GPUs using NCCL DistributedDataParallel with automatic mixed precision, for 40 epochs. The mean absolute error falls from 109.1 K at initialization to 0.71 K at convergence. The convergence curve is monotone after the first two epochs, and the same ordering of validation scores is reproduced across independent re-runs with different random seeds. A systematic ablation study over the attention module, the gradient-matching loss term, and TTA, together with further assessment on broader geometries and thermal operating regimes, is identified as future work.

D. Test-time augmentation and per-scenario results

Eight-way test-time augmentation averages predictions across the eight symmetry-preserving reorientations of the cubic domain, exploiting the invariance of the underlying physics under these operations for the storage-cell

geometry. TTA reduces the plain test mean absolute error from 0.661 K (single forward pass) to 0.636 K (8-way average), a 3.7 % relative improvement, and also reduces the root-mean-square and maximum errors. Table VII presents the per-scenario test-set performance.

Table VII. 3D TUNet-AI per-scenario test results (8-way test-time augmentation).

Run	Type	MAE (K)	RMSE (K)	Max error (K)
T01	Nominal charging	0.37	0.64	108.0
T04	Charging + ambient cooling	0.30	0.63	108.7
T08	Discharge, 600 K peak	0.59	1.75	156.0
T36	Discharge, 450 K peak	0.29	0.88	78.0
T40	Charge-to-relaxation	**1.08**	4.72	343.5
T41	4-pipe distributed	0.68	0.81	33.7
T42	9-pipe distributed	0.69	0.88	34.9
Overall	**TTA aggregate**	**0.636**	**2.517**	**343.5**

The per-scenario maximum errors warrant explicit comment. Global MAE values are sub-Kelvin across the test set, but local maximum errors can be large in two identifiable situations. First, sharp material interfaces, in particular the voxel-scale transition between the cylindrical heater surface and the surrounding sand, produce discretisation-sensitive peak errors in cases with steep gradients (e.g. T01 at 108 K, T04 at 108.7 K). These are spatially localised at the pipe boundary and do not propagate. Second, the charge-to-relaxation transition T40 produces the largest maximum error of 343.5 K, concentrated at the temporal discontinuity when the source term is switched off at step 5000 and the field derivative changes sign instantaneously. Mechanistically, this is a temporally-localised phenomenon at a single known instant; the T40 global MAE remains at 1.08 K across the full 10 000-step run. The distributed-heater cases T41 and T42 have the smallest maximum errors of the test set (below 35 K), consistent with the smoother and more spatially uniform thermal fields produced by distributed heating.

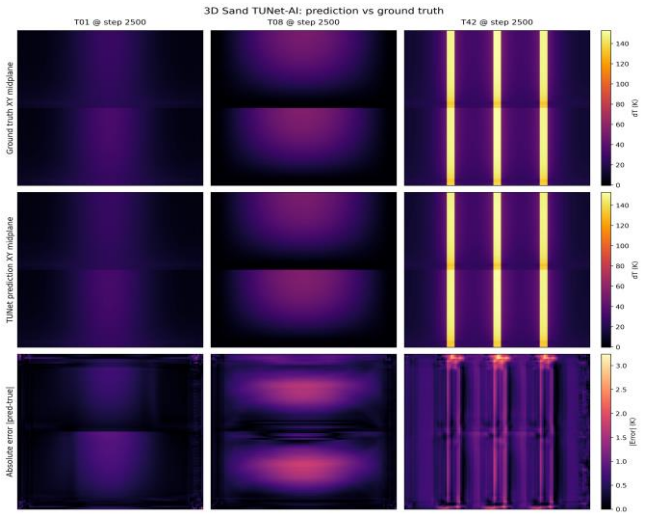

Figure 12: 3D TUNet-AI — representative ground-truth XY midplane (top row), prediction (middle row), and absolute error (bottom row) for three test scenarios: T01 nominal charging, T08 discharge, and T42 nine-pipe distributed heating. Eight-way test-time augmentation applied. The error scale shown here is for representative frames; Table VII reports worst-case maximum errors across full trajectories.

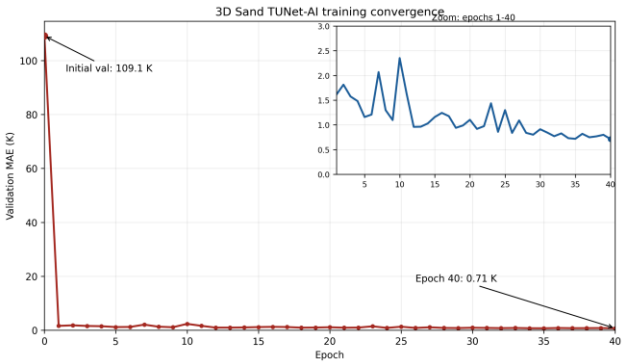

Figure 13: 3D TUNet-AI training convergence on eleven V100 GPUs. Validation MAE decreases from 109.1 K at initialisation to 0.71 K at convergence over 40 epochs. The inset shows the fine-grained progression over the final 30 epochs.

E. Training convergence

Figure 13 shows the training convergence. The rapid initial drop in validation MAE confirms that the architecture captures the dominant thermal physics within the first epoch, with subsequent epochs refining the prediction at material interfaces and in the neighbourhood of sharp thermal fronts.

F. Computational performance

Table VIII summarises the computational performance of the coupled simulation-plus-surrogate stack. The single-pass 3D inference latency below 5 ms corresponds to one forward pass of TUNet-AI on a V100 GPU. The 8-way TTA latency is approximately 40 ms, corresponding to eight forward passes followed by symmetry-averaging, and represents the end-to-end evaluation cost of the 0.636 K MAE offline-benchmark result reported in Table VII. For real-time deployment scenarios where the 3.7 % TTA gain is not required, the single-pass latency applies.

Table VIII. Computational performance comparison.

Metric	Value
LBM solver wall-clock (96^3, 5000 steps)	33 s per scenario
Full 43-scenario 3D campaign wall-clock	~12 min on 2 nodes
AI inference (3D, single prediction)	< 5 ms
AI inference (3D, 8-way TTA)	≈ 40 ms
End-to-end speedup vs LBM	> 6 600× (using < 5 ms single prediction)
2D AI inference	5.0 ms
3D TUNet-AI parameters	4.1 M
GPU memory at inference	< 500 MB

VII. CONCLUSIONS

A GPU-accelerated LBM-MRT thermal solver for sand-based thermal energy storage has been developed and applied to a three-dimensional campaign comprising 43 scenarios at 96^3 resolution, completing in approximately 12 minutes of wall-clock time. The campaign supports three principal findings. First, the thermal response during charging is conduction-dominated and exhibits approximate cylindrical symmetry about the heater pipe. Second, the temperature field at the external insulation skin is not a reliable proxy for the internal thermal state, indicating that external thermography alone is insufficient without inverse-model reconstruction or internal calibration. Third, at equal total heater power, distributing the heat input across nine 10 mm-radius pipes rather than a single 15 mm-radius pipe reduces the peak internal temperature rise by more than 70 % (from ≈712 K to ≈188 K) while producing a broader and more spatially uniform heated region; the two configurations also differ in total heater cross-sectional area and per-pipe source density, so the reduction combines spatial distribution with lower per-pipe source density.

A two-dimensional U-Net surrogate achieves 0.23 K mean absolute error with 0.998 thermal-front intersection-over-union at 5 ms inference latency. The three-dimensional TUNet-AI surrogate achieves 0.661 K MAE in a single forward pass at sub-five-millisecond latency and 0.636 K MAE with 8-way test-time augmentation at ≈40 ms, on an offline benchmark that uses per-scenario min/max Normalization over the full trajectory of each scenario. Follow-on work is organized into two tiers. The highest priority items directly affect the validity and deployment-readiness of the reported numbers and should be addressed before the method is used for quantitative design: (i) an enthalpy-conservative LBM reformulation to resolve the E_stored / E_injected ≈ 0.416 ratio reported in Section IV; and (ii) a retraining run of the 3D surrogate with deployment-safe Normalization restricted to the network's input-window frames or to global training-set statistics. Additional extensions that broaden the scope of the work but do not affect the validity of the present findings include: (iii) a matched-physical-time grid-convergence study against a 128^3 reference; (iv) extension of the model to include temperature-dependent properties and radiative transport; (v) a matched-heater-volume comparison to isolate the contribution of spatial distribution from that of reduced per-pipe source density; and (vi) a systematic ablation study over the TUNet-AI architectural components.

ACKNOWLEDGMENT

This work is part of the **REALISATION-SAND-AI** Project (Grant number: COM-CONCEPT-ENERGY/0624/0185), which is funded by the EU Recovery and Resilience Facility of the European Union - NextGenerationEU, and the Republic of Cyprus through the Research and Innovation Foundation within the framework of the «RESTART 2016-2020» Programmes for Research, under the Component 6.1 «REPowerEU» of the Cyprus Recovery and Resilience Plan. Computations were performed on the KYAMOS V100 InfiniBand cluster.

REFERENCES

[1] Polar Night Energy, “Sand-based thermal energy storage,” polarnightenergy.fi, 2024.
[2] M. Diago, A. C. Iniesta, T. Delclos, T. Shamim, N. Calvet, “Characterization of desert sand for its feasible use as thermal energy storage medium,” Energy Procedia, vol. 75, pp. 2113–2118, 2015.
[3] S. Tetteh, G. Juul, M. Järvinen, A. Santasalo-Aarnio, “Improved effective thermal conductivity of sand bed in thermal energy storage systems,” J. Energy Storage, vol. 86 (Part B), 111350, 2024.
[4] S. Trevisan, Y. Jemmal, R. Guedez, “Packed bed thermal energy storage: a novel design methodology including quasi-dynamic boundary conditions and techno-economic optimisation,” J. Energy Storage, vol. 36, 102441, 2021.
[5] Á. R. Acosta-Iborra et al., “Comprehensive review of dynamical simulation models of packed-bed systems for thermal energy storage applications in renewable power production,” Heliyon, vol. 11, 2025.
[6] S. Succi, The Lattice Boltzmann Equation for Fluid Dynamics and Beyond. Oxford Univ. Press, 2001.
[7] P. Lallemand and L.-S. Luo, “Theory of the lattice Boltzmann method: dispersion, dissipation, isotropy, Galilean invariance, and stability,” Phys. Rev. E, vol. 61, no. 6, pp. 6546–6562, 2000.
[8] J. Latt et al., “Palabos: parallel lattice Boltzmann solver,” Comput. Math. Appl., vol. 81, pp. 334–350, 2021.
[9] H. Peng et al., “Application of U-Net in 3D steady heat conduction solver,” in Deep Learning-Based Forward Modeling and Inversion Techniques for Computational Physics Problems, CRC Press, 2023.
[10] S. Koric and D. W. Abueidda, “Data-driven and physics-informed deep learning operators for solution of heat conduction equation with parametric heat source,” Int. J. Heat Mass Transf., vol. 203, 123809, 2023.
[11] X. Han, Q. Zhao, J. Kang, J. Wang, “Deep learning based heat transfer simulation of the casting process,” Sci. Rep., vol. 14, 2024.
[12] Z. Guo, C. Zheng, B. Shi, “Discrete lattice effects on the forcing term in the lattice Boltzmann method,” Phys. Rev. E, vol. 65, 046308, 2002.

[13] G. de Vahl Davis, "Natural convection of air in a square cavity: a bench mark numerical solution," Int. J. Numer. Meth. Fluids, vol. 3, pp. 249–264, 1983.
[14] O. Ronneberger, P. Fischer, T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in MICCAI, pp. 234–241, 2015.
[15] Z. Li et al., "Fourier neural operator for parametric partial differential equations," in ICLR, 2021.
[16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in MICCAI, pp. 424–432, 2016.