

# Deterministic Realization of Classical Dissipation on Quantum Computers

Muhammad Idrees Khan<sup>a,b\*</sup> Sauro Succi<sup>c,d</sup> Hua-Dong Yao<sup>a</sup>

<sup>a</sup>Department of Mechanical Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden

<sup>b</sup>Department of Enterprise Engineering “Mario Lucertini”, University of Rome “Tor Vergata”, Via del Politecnico 1, 00133 Rome, Italy

<sup>c</sup>Italian Institute of Technology, Viale Regina Elena, 291, 00161 Roma, Italy

<sup>d</sup>Department of Physics, Harvard University, 33 Oxford Street, Cambridge, 02138, Massachusetts, USA

## Abstract

Lattice Boltzmann (LB) on quantum devices must reconcile unitary gate evolution with the dissipative *collision* step. In the multiple-relaxation-time (MRT) class, we work in the common setting of *mode-wise diagonal* moment relaxation,  $\delta m'_r = \lambda_r \delta m_r$  with  $\lambda_r \in [-1, 1]$  (overrelaxation if  $\lambda_r < 0$ ). Embedding that contraction in a unitary by block encoding or a linear combination of unitaries (LCU) typically yields subunitary success probability that decays multiplicatively across modes, sites, and time, a key bottleneck for quantum LB. *For the dissipative MRT block alone* we give a *block-encoding-free* construction: a signed *two-rail* population encoding, then a completely positive trace-preserving (CPTP) map (per-rail amplitude damping with survival  $|\lambda_r|$  and, if  $\lambda_r < 0$ , a rail SWAP) so that, after the decode, the map agrees with classical MRT relaxation exactly (expectations of the rail number operators, common encoding–decode scale). Trace preservation gives success probability 1 for that substage. The main result is the dissipative MRT block; construction of the equilibrium moment vector  $m^{\text{eq}} = M f^{\text{eq}}$  (prescribed  $f^{\text{eq}}$ , host moment matrix  $M$ ; notation as in Section 2.1), moment transforms, streaming, and boundaries are composed with it as in a standard host pipeline and lie outside the scope of the formal theorem. Hybrid and fully coherent encodings, adaptive scales, Carleman-based context, and a one-rail no-go in the same nonnegative population framework are in the main text. Audits of the open-channel map on a long LBM collide–stream simulation and on stencil-free inputs both match the target to machine precision.

**Keywords:** lattice Boltzmann method; multiple-relaxation-time; CPTP maps; open quantum systems; two-rail encoding.

## 1 Introduction

Recently, major attention has been directed to the development of quantum algorithms for classical physics problems, most notably nonlinear transport and fluid phenomena [8, 11, 19]. Such algorithms face two major extra-challenges: nonlinearity and non-unitarity (dissipation). Several strategies have been developed in the last few years, a particularly prominent being based on Carleman linearization of the lattice Boltzmann (LB) formulation [3, 4, 5] of fluid dynamics, coupled with linear combination of unitaries (LCU) and/or block encoding (BE) [7, 16]. Despite

---

\*Corresponding author.

numerous attempts, these algorithms are still plagued by a severe bottleneck due to the low success probability of the dissipative update. In this paper we show that such bottleneck can be removed by turning to a deterministic completely positive trace-preserving (CPTP) channel: amplitude damping, which completely dispenses with LCU and BE steps, thereby delivering unit success probability. The strategy is custom-made on a pseudo-spectral version of the LB method, whereby free-streaming occurs in real space, whereas the multiple-relaxation-time (MRT) collision, responsible for loss of unitarity, is performed in the space of kinetic moments. This way, both steps rely upon diagonal representations. This makes it possible to import powerful techniques from quantum information science and the theory of open quantum dissipative systems and to adapt them to the case of nonlinear dissipative classical systems. Choi *et al.* [2] combine homotopy analysis with Lindbladian dynamics for nonlinear partial differential equations (PDEs) in an open-system framework. This work gives a different contribution: Theorem 3.3 provides a deterministic, block-encoding-free CPTP realization of the modewise MRT update  $\delta m'_r = \lambda_r \delta m_r$  on signed two-rail encodings in kinetic-moment space.

In the MRT formulation of the lattice Boltzmann method (LBM) [6, 9, 12], the dissipative part of the collision is modewise diagonal in a chosen moment basis: nonequilibrium moments relax as  $\delta m'_r = \lambda_r \delta m_r$  with  $\lambda_r = 1 - s_r \in [-1, 1]$ . For every dissipative mode this is a *contraction* ( $|\lambda_r| \leq 1$ ), and in the overrelaxation regime  $\lambda_r < 0$  it is also sign-reversing. Contractive, non-unitary maps are not directly implementable on a quantum computer: the standard route is to embed the contraction inside a larger unitary (block encoding) and post-select on an ancilla flag [10, 13]. The per-step success probability of such a block-encoded contraction is bounded above by  $|\lambda_r|^2 / \alpha_r^2$  for subnormalization  $\alpha_r \geq 1$  (equivalently by  $|\lambda_r|^2$  in the tight case  $\alpha_r = 1$ ), and compounds *multiplicatively* over dissipative modes, lattice sites, and timesteps. Multiplicative post-selection or heralding penalties of the same structural form also appear in closely related quantum lattice Boltzmann method (QLBM) routes, notably in LCU-based realizations of linear collision operators [16] and in block-encoded Carleman-linearized lattice Boltzmann constructions [7]; this compounded probability is widely identified in the QLBM literature as the principal obstacle to end-to-end execution at useful Reynolds numbers.

Theorem 3.3 makes the CPTP construction precise: after a signed two-rail population encoding of  $\delta m_r$  (Section 3), amplitude damping with survival  $\alpha_r = |\lambda_r|$  on each rail and a rail SWAP whenever  $\lambda_r < 0$  reproduce  $\delta m_r \mapsto \lambda_r \delta m_r$  exactly at the decoded expectation-value level, with per-step success probability 1 for the dissipative block itself (Corollary 3.4).

The remainder of the paper is organized as follows. Section 2.1 fixes notation for diagonal MRT in moment space. Section 3 contains the encoding, the signed channel, and the main theorem. Section 4 quantifies the block-encoding / success-probability comparison and situates the present result in the QLBM literature. Section 5 sketches how the primitive integrates into a generic block-encoded MRT collision pipeline, comments on its relation to Carleman-linearized QLBM circuits, which share the same structural bottleneck but do not expose a mode-wise diagonal target, and spells out a minimal hybrid Carleman–CPTP pipeline (§5.4) in which Carleman linearization is used only for the nonlinear equilibrium evaluation and the dissipative relaxation is handled by the present primitive. Section 6 reports a machine-precision numerical audit on D3Q19.

The main contributions are the following.

- (C1) The dissipative MRT update  $\delta m'_r = \lambda_r \delta m_r$  is realized as a *deterministic CPTP channel with unit per-step success probability*, in contrast to block-encoded contractions whose success probability decays multiplicatively across modes, sites, and timesteps.
- (C2) Theorem 3.3 proves exact decoded-expectation equivalence on two-rail registers, including the overrelaxation branch  $\lambda_r < 0$ , which is captured by a classically conditioned rail SWAP

composed with amplitude damping at survival  $|\lambda_r|$ .

- (C3) Proposition 3.6 shows that the encoding scale  $S_r(\mathbf{x}, t)$  is classically indexed side information rather than a hidden quantum resource: adaptive scales produce a classically indexed family of encode–channel–decode composites.
- (C4) Within the single-rail nonnegative population-encoding framework of this paper, one rail cannot exactly represent a signed nonequilibrium moment on a two-signed domain (Appendix B), which motivates the two-rail split.
- (C5) Numerical audits: Section 6.1 (D3Q19, decaying Taylor–Green vortex) confirms machine-precision agreement with the target dissipative update when the error is maximized over all dissipative modes and all lattice sites; Section 6.2 independently stress-tests the same per-mode map on stencil-free random  $(\delta m_r, \lambda_r) \in [-X, X] \times [-1, 1]$ , with no particular stencil or flow.

Any QLBM pipeline whose dissipative collision is implemented by block-encoding a diagonal contraction on nonequilibrium moments (that is, any block-encoded MRT scheme in which the dissipative block reduces to  $\Lambda = \text{diag}(\lambda_r)$  on dissipative modes) pays a per-mode, per-site, per-step success-probability cost that decays multiplicatively in  $T$ ,  $|\Omega_h|$ , and  $|\mathcal{D}|$  (Section 4.1). Substituting the primitive proved here for that contraction replaces this compounded cost by 1 for the dissipative block itself, independently of  $\lambda_r$ ,  $S_r$ , and the input state (Corollary 3.4); the overrelaxation branch  $\lambda_r < 0$  is absorbed as a classically conditioned rail SWAP at no probability cost, not as an additional block-encoded conditional. The price paid in exchange, namely a classical nonlinear sign split, two rails per dissipative mode, and classically tracked scales  $S_r(\mathbf{x}, t)$ , is bookkeeping rather than probability loss (Section 4.3). Preparation, extraction, reconstruction, streaming, boundaries, and readout remain separate, host-determined pipeline stages (Section 5) and are not claimed to benefit from Theorem 3.3 in isolation. Block-encoded Carleman-linearized lattice Boltzmann circuits [7] share the same structural bottleneck of a non-unitary collision operator embedded with ancilla post-selection, but their non-unitarity is carried by the full (non-sparse, non-block-diagonal) Carleman collision matrix rather than by a mode-wise diagonal  $\Lambda$ , so the present primitive does not apply to them as a drop-in replacement; the relation is discussed in Section 5.3.

The focus of this article is the dissipative block of MRT collision and its open-channel realization on rails. Coherent equilibrium loading, full one-node collision reconstruction, streaming, lattice boundaries, hardware execution, finite-shot behavior, and asymptotic speedup claims are outside the scope of this work. The dissipative primitive proved here is of standalone theoretical interest as a block-encoding-free CPTP realization of a signed modewise contraction: its validity does not depend on any specific host pipeline, and Section 5 shows that it is compatible with generic block-encoded MRT collision pipelines. Its relation to Carleman-linearized QLBM circuits, which illustrate the same block-encoding bottleneck but do not expose a mode-wise diagonal target, is also discussed there.

Throughout the paper, items are labeled (Letter)(number) so they can be cited in one pair of parentheses. The letter only names which list is meant (it is not a variable): C marks the contribution list; I marks the practical-interpretation checklist; T marks the three tradeoffs of the deterministic CPTP route; P marks the four on-site MRT collision pipeline stages; U marks the three host specializations in the list immediately after (23) (block-encoding, LCU, or Carleman); H marks the six substeps in the hybrid Carleman–CPTP write-up (26); and S marks the five stencil-free synthetic audit cases. The number indexes an entry within that list. Thus (C1)–(C5) are the items above; (I1)–(I7) appear in Section 3.1; (T1)–(T3) in Section 4.3; (P1)–(P4) in Section 5.1; (U1)–(U3) in that list (right after (23)); (H1)–(H6) in (26); and (S1)–(S5) in Section 6.2.

## 2 MRT-LBM in moment space

### 2.1 Generic moment-space collision

Consider a lattice Boltzmann scheme with  $q$  discrete velocities  $\{\mathbf{c}_i\}_{i=0}^{q-1}$  and population vector

$$\mathbf{f} = (f_0, \dots, f_{q-1})^\top \in \mathbb{R}^q. \quad (1)$$

Let  $M \in \mathbb{R}^{q \times q}$  be any fixed *invertible* moment matrix and define  $\mathbf{m} = M\mathbf{f}$ . Given a prescribed equilibrium  $\mathbf{f}^{\text{eq}}(\rho, \mathbf{u})$ , where  $(\rho, \mathbf{u})$  are the hydrodynamic fields extracted by the baseline MRT model (the usual mass and momentum densities constructed from  $\mathbf{f}$ ), set  $\mathbf{m}^{\text{eq}} = M\mathbf{f}^{\text{eq}}$ . The MRT collision step is

$$\delta \mathbf{m} = \mathbf{m} - \mathbf{m}^{\text{eq}}, \quad (2)$$

$$\delta m'_r = (1 - s_r) \delta m_r, \quad m^+ = m^{\text{eq}} + \delta m', \quad \mathbf{f}^* = M^{-1} \mathbf{m}^+. \quad (3)$$

Post-collision populations stream by

$$f_i(\mathbf{x} + \mathbf{c}_i, t + 1) = f_i^*(\mathbf{x}, t). \quad (4)$$

Let  $\mathcal{C}$  denote the conserved moment index set and  $\mathcal{D}$  its complement (dissipative modes), with  $s_r = 0$  for  $r \in \mathcal{C}$ . Define the modewise relaxation multiplier

$$\lambda_r \stackrel{\text{def}}{=} 1 - s_r. \quad (5)$$

Under the usual MRT admissible range  $0 \leq s_r \leq 2$  for  $r \in \mathcal{D}$ , one has

$$\lambda_r \in [-1, 1] \quad \text{for every } r \in \mathcal{D}. \quad (6)$$

The regime  $\lambda_r < 0$  is *overrelaxation*. Below we only require  $M$  to be invertible; orthogonality is not assumed. The collision operator in (3) is *modewise diagonal* in the moment coordinates defined by  $M$ : each nonequilibrium moment  $\delta m_r$  relaxes with its own scalar factor. Block-coupled or collision models that are nondiagonal in this basis lie outside the scope of the theorem below. The open-channel construction in Section 3 applies independently to each  $r \in \mathcal{D}$  whenever (3) holds in that basis; particular stencils (D2Q9, D3Q19, D3Q27, ...) specify  $q$ , the  $\mathbf{c}_i$ , weights, and the numerical values of  $s_r$ .

## 3 Signed two-rail population encoding and open-channel realization

Fix the generic MRT setting of Section 2.1. The central result is Theorem 3.3, which addresses dissipative modes only; Appendix C records optional algebraic consequences when the open dissipative update is embedded in the same moment reconstruction as a classical baseline. D3Q19 (Appendix A) appears only in the numerical consistency check of Section 6.

For each dissipative mode  $r \in \mathcal{D}$ , let  $S_r = S_r(\mathbf{x}, t) > 0$  denote the **encoding scale**, which may be chosen globally, per mode, or adaptively per node and time, provided the normalized rail populations defined below remain in  $[0, 1]$ . Define

$$p_r^+ = \frac{\max(\delta m_r, 0)}{S_r}, \quad p_r^- = \frac{\max(-\delta m_r, 0)}{S_r}. \quad (7)$$

We require  $p_r^\pm \in [0, 1]$ ; the scale  $S_r$  carries the physical magnitude and is restored by the decoder. Assign one qubit to each rail and define diagonal one-qubit states

$$\rho_r^+ = (1 - p_r^+) |0\rangle\langle 0| + p_r^+ |1\rangle\langle 1|, \quad \rho_r^- = (1 - p_r^-) |0\rangle\langle 0| + p_r^- |1\rangle\langle 1|. \quad (8)$$

The signed two-rail population encoding is  $\rho_r = \rho_r^+ \otimes \rho_r^-$ . Introduce rail number operators  $n_+ = |1\rangle\langle 1| \otimes I$  and  $n_- = I \otimes |1\rangle\langle 1|$ . The decoder is

$$\mathcal{D}_r(\rho_r) \stackrel{\text{def}}{=} S_r (\text{Tr}[n_+ \rho_r] - \text{Tr}[n_- \rho_r]) = S_r (p_r^+ - p_r^-) = \delta m_r. \quad (9)$$

**Remark 3.1** (Choosing the encoding scale  $S_r$ ). *At each use, feasibility  $p_r^\pm \in [0, 1]$  in (7) is equivalent to  $S_r \geq |\delta m_r|$  when  $\delta m_r \neq 0$  (if  $\delta m_r = 0$ , any  $S_r > 0$  suffices). An automatic per-step choice at fixed  $(\mathbf{x}, t)$  is  $S_r = \max(|\delta m_r|, \varepsilon)$  with fixed  $\varepsilon > 0$  for numerical stability, using the same  $S_r$  in encoding and decoding (9). A manual or campaign-wide choice is any fixed  $S_r$  (or per-mode bound) with  $S_r \geq \sup |\delta m_r|$  over the nodes, times, and runs of interest; larger  $S_r$  remains valid but dilutes rail populations. The theorem does not fix  $S_r$  from Reynolds number or mesh spacing alone without such state information or a chosen bound.*

**Remark 3.2** (Nonlinear encoding map). *For fixed  $S_r$ , the map  $\delta m_r \mapsto (p_r^+, p_r^-)$  defined by (7) is nonlinear because of the positive/negative split. Accordingly, the result is an exact decoded CPTP realization on encoded rail registers (nonlinear encoding, CPTP channels on the rails, then decoding), not a single linear CPTP map acting directly on the unencoded classical moment variable  $\delta m_r$ .*

The positive/negative split in (7) is not a technical convenience: within the present nonnegative population-encoding framework, a single rail cannot exactly represent a signed scalar on any domain containing both signs. We defer the precise statement and proof to Appendix B, since the construction below does not rely on it; it is recorded there to motivate the two-rail split rather than a single-rail encoding.

Let  $\mathcal{A}_\alpha$  denote the one-qubit amplitude-damping channel with survival factor  $\alpha \in [0, 1]$ , which maps population  $p \mapsto \alpha p$  [10, 13]. Set  $\alpha_r = |\lambda_r|$ . Define the signed channel on the two-rail register by

$$\mathcal{E}_r = \begin{cases} \mathcal{A}_{\alpha_r} \otimes \mathcal{A}_{\alpha_r}, & \lambda_r \geq 0, \\ \text{SWAP} \circ (\mathcal{A}_{\alpha_r} \otimes \mathcal{A}_{\alpha_r}), & \lambda_r < 0. \end{cases} \quad (10)$$

Both branches are CPTP. The full construction is summarized in Figure 1.

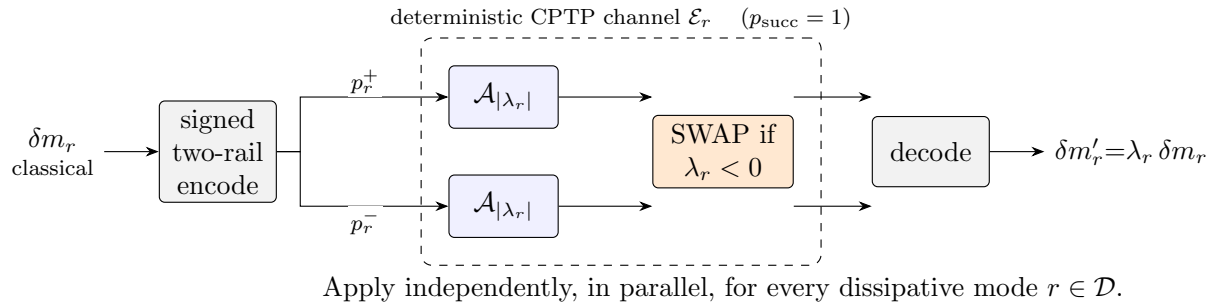


Figure 1: Block-encoding-free CPTP realization of the diagonal MRT dissipative update. The classical nonequilibrium moment  $\delta m_r$  is mapped onto two rails  $(p_r^+, p_r^-)$  by the signed population encoding (7). Each rail is damped by an amplitude-damping channel with survival  $|\lambda_r|$ , and a rail SWAP is applied when  $\lambda_r < 0$ ; the decoder (9) reads off  $\delta m'_r = \lambda_r \delta m_r$  exactly (Theorem 3.3). The boxed region is trace-preserving by construction, so the per-application success probability is unity: no heralding, post-selection, or amplitude amplification is required, in contrast to block-encoded realizations of the same contraction (Section 4). The SWAP branch is selected by the classical sign of  $\lambda_r$  and carries no probability cost.

**Theorem 3.3** (Modewise CPTP realization of MRT dissipative relaxation). *For each dissipative mode  $r \in \mathcal{D}$ , let  $S_r = S_r(\mathbf{x}, t)$  be any positive scale such that the encoded populations  $p_r^\pm$  in (7) lie in  $[0, 1]$ , and use the same  $S_r$  consistently in encoding and decoding. The map  $\delta m_r \mapsto (p_r^+, p_r^-)$*

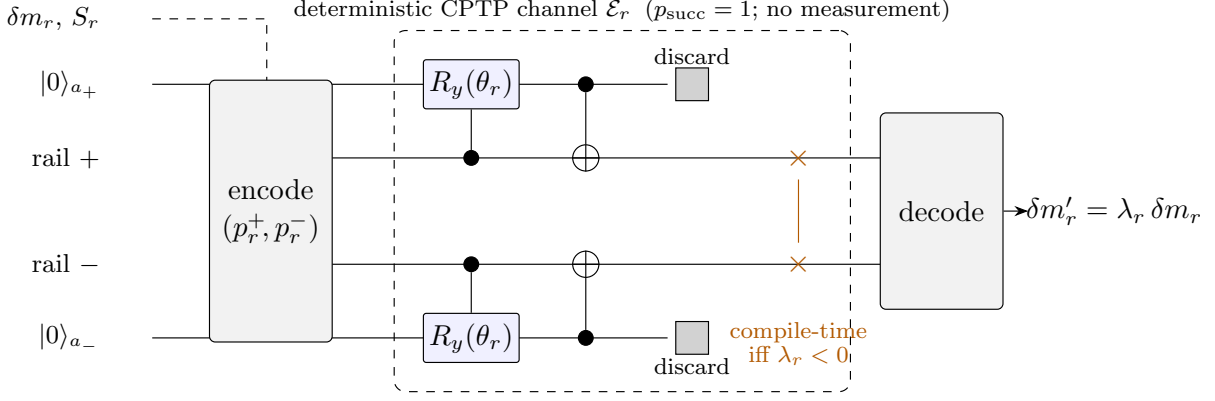


Figure 2: Gate-level realization of the CPTP channel  $\mathcal{E}_r$  of Figure 1. Each amplitude-damping box is opened into its Stinespring dilation: a fresh ancilla  $|0\rangle_{a_{\pm}}$ , a rail-controlled rotation  $R_y(\theta_r)$  with  $\theta_r = 2 \arccos\sqrt{|\lambda_r|}$  (equivalently  $2 \arcsin\sqrt{1-|\lambda_r|}$ ), a controlled NOT (CNOT) gate from the ancilla back to the rail, and a final *discard* (trace-out) of the ancilla. The rail SWAP is a *compile-time* branch selected by  $\text{sign}(\lambda_r)$  per (10), either present (if  $\lambda_r < 0$ ) or absent, so no classical control wire or runtime conditional is needed (no coherent conditional on  $\lambda_r$  is required; Section 5.1). Crucially, no ancilla is ever measured: the dashed CPTP frame contains only coherent unitaries followed by an ancilla trace-out, both of which are deterministic, so  $p_{\text{succ}} = 1$  per application (Corollary 3.4). This is the concrete gate-level contrast with measurement-based realizations of the same contraction, where ancilla post-selection introduces a  $\|\cdot\|$ -dependent success-probability penalty (Section 4).

is generally nonlinear (Remark 3.2); the CPTP map  $\mathcal{E}_r$  acts on the two-qubit rail state after this encoding. Under amplitude damping with  $\alpha_r = |\lambda_r|$ , followed by SWAP when  $\lambda_r < 0$ , the decoded output satisfies

$$\delta m'_r = \lambda_r \delta m_r \quad (11)$$

exactly at the decoded expectation-value level. Therefore the dissipative part of the classical MRT collision operator is reproduced exactly by the encode–channel–decode construction on the rails (for any fixed valid scale choice); when the scale is chosen adaptively, Proposition 3.6 describes the resulting indexed family of composites.

*Proof.* Amplitude damping scales  $p_r^{\pm} \mapsto \alpha_r p_r^{\pm}$ . If  $\lambda_r \geq 0$ :  $\alpha_r = \lambda_r$ , no swap. Then  $\mathcal{D}_r(\mathcal{E}_r(\rho_r)) = S_r(\alpha_r p_r^+ - \alpha_r p_r^-) = \lambda_r S_r(p_r^+ - p_r^-) = \lambda_r \delta m_r$ . If  $\lambda_r < 0$ :  $\alpha_r = -\lambda_r$ , and the swap exchanges the rails after damping, so  $\mathcal{D}_r(\mathcal{E}_r(\rho_r)) = S_r(\alpha_r p_r^- - \alpha_r p_r^+) = -\alpha_r S_r(p_r^+ - p_r^-) = \lambda_r \delta m_r$ .  $\square$

**Corollary 3.4** (Channel stage is deterministic). *For every dissipative mode  $r \in \mathcal{D}$  and every admissible two-rail input state  $\rho_r$ , the channel  $\mathcal{E}_r$  of (10) satisfies*

$$\text{Tr}[\mathcal{E}_r(\rho_r)] = \text{Tr}[\rho_r] = 1, \quad p_{\text{succ}}^{(r)} = 1, \quad (12)$$

independently of  $\lambda_r$ ,  $S_r$ , and the input state. Composed independently across  $r \in \mathcal{D}$ , lattice sites  $\mathbf{x} \in \Omega_h$ , and timesteps  $t = 1, \dots, T$ , the product dissipative channel

$$\bigotimes_{t=1}^T \bigotimes_{\mathbf{x} \in \Omega_h} \bigotimes_{r \in \mathcal{D}} \mathcal{E}_r(\mathbf{x}, t) \quad (13)$$

is CPTP with per-step and end-to-end success probability equal to 1. In particular, the dissipative block itself does not accumulate any multiplicative success-probability penalty in  $|\mathcal{D}|$ ,  $|\Omega_h|$ , or  $T$ , in contrast to block-encoded realizations of the same contraction (Section 4.1).

*Proof.* Amplitude damping  $\mathcal{A}_{\alpha_r}$  with  $\alpha_r \in [0, 1]$  is CPTP [10, 13]. Tensor products of CPTP maps are CPTP, and composition with the unitary SWAP (a CPTP map) preserves CPTP. Both branches of (10) are therefore CPTP, giving (12) by trace preservation. The product channel (13) is a tensor product of CPTP maps and is itself CPTP, hence trace-preserving, which gives unit end-to-end success probability.  $\square$

**Remark 3.5** (Scope of Corollary 3.4). *The statement concerns only the dissipative CPTP block  $\otimes_r \mathcal{E}_r$ . Preparation of the rails from classical side information or from a coherent population register, equilibrium loading, nonequilibrium extraction, reconstruction  $f^* = M^{-1}m^+$ , streaming, boundary treatment, and finite-shot readout are separate pipeline stages (see Section 5); their probabilistic costs, if any, are not covered by Corollary 3.4.*

Because admissible scale choices  $S_r(\mathbf{x}, t)$  may vary across nodes, times, and modes, it is useful to state explicitly what remains fixed and what is classically indexed in the construction.

**Proposition 3.6** (Adaptive scales: classically indexed encode–channel–decode family). *Suppose  $S_r = S_r(\mathbf{x}, t)$  may vary with node and time (subject to  $p_r^\pm \in [0, 1]$ ). For fixed relaxation physics, the middle map  $\mathcal{E}_r$  in (10) depends only on  $\lambda_r$  (hence on the chosen  $s_r$ ), not on  $S_r$ . What adapts with  $S_r(\mathbf{x}, t)$  is the composite update  $\mathcal{D}_r \circ \mathcal{E}_r \circ (\text{encode})$  on dissipative mode  $r$ . Thus, as  $(\mathbf{x}, t)$  and  $r$  vary, one obtains a nodewise, classically indexed family of encode–channel–decode maps; equivalently, classical side information  $(\mathbf{x}, t)$  (and mode label  $r$ ) indexes which composite acts. It is not, in general, a single fixed CPTP map acting directly on the raw (unencoded) nonequilibrium moments  $\delta m_r$  (cf. Remark 3.2).*

*Proof.* Fix  $(\mathbf{x}, t)$  and  $r \in \mathcal{D}$ . Encoding with the scale  $S_r(\mathbf{x}, t)$  produces a product state  $\rho_r$  on the rails;  $\mathcal{E}_r$  in (10) is CPTP and independent of  $S_r$ . Letting  $(\mathbf{x}, t)$  vary replaces one such triple (encode,  $\mathcal{E}_r$ , decode) by another only through the classically chosen  $S_r$ . The composite map on unencoded  $\delta m_r$  alone incorporates this  $S_r$ -dependent encoding and need not coincide with any single CPTP channel on the moment variable (Remark 3.2).  $\square$

**Remark 3.7** (Node-level composite dissipative update). *With disjoint two-qubit registers for distinct  $r \in \mathcal{D}$ , applying  $\mathcal{E}_r$  mode-wise and decoding with the same  $S_r$  used in encoding yields  $\delta m'_r = \lambda_r \delta m_r$  simultaneously for all  $r \in \mathcal{D}$ , i.e. the full dissipative moment vector update matches the classical MRT operator on the dissipative subspace. This follows from Theorem 3.3 and independence of registers.*

**Remark 3.8** (Stencil and modewise structure). *Theorem 3.3 assumes only modewise diagonal relaxation (3) with  $\lambda_r \in [-1, 1]$  and invertible  $M$ ; block-coupled collisions in the same moment basis are out of scope.*

**Remark 3.9** (Encoding and decoded observables). *Only the normalized rail populations  $p_r^\pm$  must lie in  $[0, 1]$ ; raw  $\delta m_r$  need not. The exactness claim is for the diagonal two-rail encoding and decoded expectations of rail number operators.*

### 3.1 Practical interpretation

At an operational level, Theorem 3.3 and its surrounding lemmas can be read as a short list of circuit-level facts, independent of the specific lattice or moment basis in use.

- (I1) **One dissipative mode = two rails.** Each  $r \in \mathcal{D}$  is carried by its own two-qubit register  $(p_r^+, p_r^-)$ ; no coupling between distinct modes is required.
- (I2) **Dissipation is amplitude damping with survival  $|\lambda_r|$ .** The contractive factor  $|\lambda_r| \leq 1$  is realized physically as the survival probability of the  $|1\rangle$  population on each rail under  $\mathcal{A}_{|\lambda_r|}$ , not as a block-encoded amplitude rescaling requiring post-selection.
- (I3) **Sign reversal is a rail SWAP, not a negative amplitude.** The overrelaxation branch  $\lambda_r < 0$ , which is *not* a free byproduct of any unitary acting on nonnegative populations, is

captured by a classically conditioned SWAP of the two rails after damping. There is no “negative amplitude,” no coherent conditional on the sign of  $\lambda_r$ , and no additional heralding wiring.

- (I4) **The channel stage is deterministic.** The map  $\mathcal{E}_r$  in (10) is CPTP by construction, so its per-application success probability is 1 and the dissipative sweep does not accumulate any multiplicative success-probability penalty across modes, sites, or timesteps (Corollary 3.4).
- (I5) **Modes are mutually independent.** Since registers for distinct  $r \in \mathcal{D}$  are disjoint, the full dissipative block is a tensor product of per-mode channels and can be executed in parallel over  $\mathcal{D}$  without cross-talk (Remark 3.7).
- (I6) **The encoding scale  $S_r$  is classical side information.**  $S_r(\mathbf{x}, t)$  is a deterministic, classically tracked quantity used consistently at encoding and decoding, not a hidden quantum resource (Proposition 3.6).
- (I7) **Stencil- and basis-agnostic.** The construction assumes only modewise diagonal relaxation (3) in some invertible moment basis  $M$  with  $\lambda_r \in [-1, 1]$ ; it does not depend on the particular velocity stencil (D2Q9, D3Q19, ...) or on orthogonality of  $M$ .

## 4 Block-encoding vs CPTP dissipation: the practical advantage

The purpose of this section is to make explicit what the construction of Section 3 delivers relative to the textbook route for realizing the contraction  $\delta m_r \mapsto \lambda_r \delta m_r$  on a quantum computer. The essential comparison is not how the contraction is computed, but what price is paid per application in *success probability*.

### 4.1 The block-encoding route and its bottleneck

A standard way to implement a contractive, non-unitary linear map  $A$  with  $\|A\| \leq 1$  on a quantum computer is to block-encode it inside a larger unitary  $U$  acting on system and ancilla registers, so that a projection onto the  $|0\rangle$ -ancilla subspace yields  $A$  (up to subnormalization  $\alpha \geq \|A\|$ ) on the system [10, 13]. For a modewise scalar contraction  $\lambda_r$  with  $|\lambda_r| \leq 1$ , the per-application success probability on a generic input state is bounded above by

$$p_{\text{succ}}^{(r)} \leq \frac{|\lambda_r|^2}{\alpha_r^2} \leq |\lambda_r|^2, \quad (14)$$

with equality achievable only in the tight block-encoding case  $\alpha_r = 1$ . Composed independently across dissipative modes  $r \in \mathcal{D}$ , lattice sites  $\mathbf{x} \in \Omega_h$ , and timesteps  $t = 1, \dots, T$ , the end-to-end success probability for a block-encoded MRT dissipative sweep is bounded by

$$p_{\text{succ}}^{\text{end}} \leq \prod_{t=1}^T \prod_{\mathbf{x} \in \Omega_h} \prod_{r \in \mathcal{D}} |\lambda_r(\mathbf{x}, t)|^2, \quad (15)$$

which tends to zero as either  $|\Omega_h| \rightarrow \infty$  or  $T \rightarrow \infty$  whenever some  $|\lambda_r| < 1$  (the generic case away from conservation and strict involutive limits). Amplitude amplification sharpens the prefactor but does not alter the asymptotic statement that (15) decays multiplicatively in  $T$  and  $|\Omega_h|$ . More sophisticated constructions, including global block encodings that treat the diagonal operator  $\Lambda$  over modes and sites jointly with a single subnormalization  $\alpha \geq \max_r |\lambda_r|$ , likewise improve the per-application prefactor by replacing the naive product bound with a state-dependent quantity  $\|\Lambda \otimes I |\psi\rangle\|^2 / \alpha^2$ ; they do not, however, remove the multiplicative decay over  $T$  under repeated application when at least one dissipative mode is strictly contractive. This compounded probability is widely identified in the QLBM literature as the principal obstacle to end-to-end

execution. Analogous multiplicative post-selection penalties, of the same structural form though not expressed in the mode-wise factors  $|\lambda_r|$ , arise in block-encoded Carleman-linearized LB pipelines [7], whose depth/shot budgets are further driven by the non-sparse, non-block-diagonal structure of the block-encoded Carleman collision matrix [7].

**The same bottleneck under LCU.** The linear-combination-of-unitaries (LCU) realization of a contraction  $A = \sum_k c_k U_k$  with  $c_k \geq 0$  and  $\sum_k c_k = \|c\|_1$  prepares a “select” ancilla in  $|c\rangle \propto \sum_k \sqrt{c_k} |k\rangle$ , applies the multiplexed unitary  $\sum_k |k\rangle\langle k| \otimes U_k$ , and uncomputes the ancilla; projection of the ancilla on  $|c\rangle$  extracts  $A|\psi\rangle/\|c\|_1$ , so the per-application success probability on a pure input is bounded by  $\|A|\psi\rangle\|^2/\|c\|_1^2$ , of the same form as (14) with  $\alpha_r$  replaced by  $\|c\|_1$  [10, 13]. The linear-equilibrium QLBM of Liu, John, and Emerson [16] is a concrete realization of this route: they cast the Bhatnagar–Gross–Krook (BGK) collision (with bounce-back absorbed) as a non-unitary matrix  $C$  in the velocity basis and implement it through singular value decomposition (SVD) followed by LCU of the resulting diagonal singular-value factor. The same bound therefore applies to their construction, and, a fortiori, to any mode-wise MRT extension in which  $\Lambda = \text{diag}(\lambda_r)$  is realized by LCU of its nonnegative-eigenvalue SVD factor: the product structure in (15) is inherited over  $(T, |\Omega_h|, |\mathcal{D}|)$ . The common cause is structural: both block-encoding and LCU target a *coherent linear-operator* realization, in which the output is required to be the pure state  $A|\psi\rangle$  up to normalization, and the lost norm necessarily reappears as failure probability on a measured ancilla.

## 4.2 The deterministic CPTP route

The channel  $\mathcal{E}_r$  defined in (10) is completely positive and trace-preserving by construction: amplitude damping is CPTP [10, 13], tensor products of CPTP channels are CPTP, and composition with a unitary SWAP preserves CPTP. Consequently

$$\text{Tr}[\mathcal{E}_r(\rho_r)] = \text{Tr}[\rho_r] = 1 \quad \text{for every valid two-rail input state } \rho_r, \quad (16)$$

so the per-application success probability is  $p_{\text{succ}}^{(r)} = 1$ , independent of  $\lambda_r$ . The contraction on decoded populations arises from coherent probability transfer to the damping ancilla, which is subsequently discarded (or reset) rather than post-selected. The end-to-end success probability of a full MRT dissipative sweep is therefore

$$p_{\text{succ}}^{\text{end}} = \prod_{t=1}^T \prod_{\mathbf{x} \in \Omega_h} \prod_{r \in \mathcal{D}} 1 = 1, \quad (17)$$

for every  $T$  and every lattice size. The overrelaxation branch  $\lambda_r < 0$ , which is sign-reversing and thus *not* a free byproduct of block-encoded amplitude control, is absorbed into a *classically* conditioned SWAP on the two rails at no probability cost.

**Lindblad/GKSL viewpoint.** At the density-matrix level, Theorem 3.3 implements the dissipative collision as a completely positive trace-preserving channel built from amplitude damping, i.e. the standard open-system mechanism of Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) theory, rather than as closed-system unitary (Schrödinger) evolution [10, 13, 14]. The continuous-time GKSL generator

$$\dot{\rho} = -i[H, \rho] + \sum_k \gamma_k \left( L_k \rho L_k^\dagger - \frac{1}{2} \{ L_k^\dagger L_k, \rho \} \right), \quad \gamma_k \geq 0, \quad (18)$$

is the most general generator of a Markovian semigroup of CPTP maps on a finite-dimensional system [14]. GKSL is normally written in continuous time ( $\dot{\rho} = \mathcal{L}[\rho]$ ); here each LBM collision

step applies one discrete CPTP map  $\mathcal{E}_r$ , without identifying that step with a specific  $e^{\Delta t \mathcal{L}}$  unless a lattice–time embedding is fixed separately.

In structural terms, the moment-basis transform  $m = Mf$  plays the role of the diagonalization that decouples the dissipative GKSL generator into independent single-channel jump terms. The full dissipative collision factorizes over modes as  $\mathcal{E}^{\text{diss}} = \bigotimes_{r \in \mathcal{D}} \mathcal{E}_r$  (Remark 3.7), with each  $\mathcal{E}_r$  a single-channel amplitude-damping map on its two-rail encoding (Section 3). Under this analogy: the dissipative MRT index  $r \in \mathcal{D}$  corresponds to the GKSL jump index  $k$  in (18); the MRT rate  $s_r$  corresponds to the dissipation rate  $\gamma_k$ ; the rail-lowering Kraus operator inside  $\mathcal{E}_r$  corresponds to the jump operator  $L_k$ ; and the conserved moments  $(\rho, \mathbf{j})$ , which are exact fixed points of  $\mathcal{E}^{\text{diss}}$  because  $\lambda_r = 1$  on conserved  $r$ , correspond to the fixed-point manifold of the dissipative generator. If, in addition, a lattice–time embedding is fixed by identifying one MRT step with a continuous time interval  $\Delta t$ , then for  $\lambda_r = 1 - s_r > 0$  the per-step contraction  $\alpha_r = |1 - s_r|$  is the time- $\Delta t$  sample

$$\alpha_r = e^{-\gamma_r \Delta t}, \quad \gamma_r = -\frac{1}{\Delta t} \ln |1 - s_r|, \quad (19)$$

of an exponential single-jump GKSL semigroup acting on the residual encoding. The overrelaxation branch  $\lambda_r < 0$  has no continuous-time semigroup analogue at the same single rate alone: it is realized in Theorem 3.3 as that same amplitude-damping step composed with a unitary rail SWAP (Section 3), so it is a unitary dressing of a one-jump GKSL semigroup rather than an element of one.

In the same spirit as the standard Lindbladian jump terms in (18), which dissipate coherences and drive  $\rho$  toward an associated fixed manifold, MRT collision dissipates nonequilibrium kinetic information by relaxing moments toward  $m^{\text{eq}}$  (diagonally,  $\delta m'_r = \lambda_r \delta m_r$  in (3)), or populations toward  $f^{\text{eq}}$  in a BGK-type picture  $\Omega_{ij}(f_j - f_j^{\text{eq}})$  after a change of basis; the analogy is at the level of *relaxation to local equilibrium* and of a diagonalization-into-independent-channels structure, not an identity of linear algebraic form between superoperators on  $\rho$  and the classical collision matrix on  $(f, m)$ .

**Stinespring dilation, not coherent linear-operator realization.** The structural reason the bound (14) does not apply to the primitive proved here is that Section 3 realizes a completely positive trace-preserving *channel* on a rail density matrix, not a coherent linear operator on a pure rail state. By Stinespring’s theorem [10, 13], every CPTP map admits the form

$$\mathcal{E}_r(\rho_r) = \text{Tr}_E[U_r(\rho_r \otimes |0\rangle\langle 0|_E)U_r^\dagger], \quad (20)$$

for some unitary  $U_r$  on system plus damping environment  $E$ , in which the environment is *discarded* (or reset) rather than measured and post-selected. Trace preservation is therefore automatic, and the per-application success probability is 1 by construction (Corollary 3.4). The bound that governs block-encoding (14) and LCU is a theorem about the coherent recovery of  $A|\psi\rangle$  from a pure input via a projectively measured ancilla; it does not apply to the Stinespring/discard realization solved here, in which the decoded output (9) is a linear functional of the rail density matrix on diagonal observables. The contraction  $p_r^\pm \mapsto \alpha_r p_r^\pm$  induced by amplitude damping on rail populations is, at the level of coherent amplitudes, a non-unitary map whose lost norm is coherently *transferred* to the damping ancilla and subsequently reset; what is saved relative to block-encoding and LCU is probability loss, not the entropy/Landauer cost of the reset itself, which is charged to the environment in the standard open-system way.

### 4.3 What is traded in exchange

The deterministic CPTP realization is not free: it trades post-selection for a different, classically accountable, set of costs.

- (T1) **A nonlinear classical encoding step.** The map  $\delta m_r \mapsto (p_r^+, p_r^-)$  in (7) is nonlinear because of the positive/negative split (Remark 3.2). In a hybrid loop where  $\delta m_r$  is classically known at each use, this split is free; in a fully coherent pipeline, a reversible sign-split gadget is required and must be accounted for separately.
- (T2) **Two rails per dissipative mode.** One extra qubit per mode relative to a hypothetical single-rail encoding. This is a *qubit* overhead, not a probability overhead.
- (T3) **Classically indexed scale.** The encoding scale  $S_r(\mathbf{x}, t)$  is classical side information used consistently in encoding and decoding (Proposition 3.6); it does not introduce coherent ancilla cost, but it must be tracked.

In every QLBM pipeline known to the authors, converting per-step probability loss (15) into the classical bookkeeping (T1)–(T3) is the strictly dominant trade.

#### 4.4 Summary comparison

Table 1: Methodological comparison of the two possible routes for realizing the diagonal MRT dissipative contraction  $\delta m_r \mapsto \lambda_r \delta m_r$  on a quantum computer: the standard block-encoding route with ancilla post-selection (Section 4.1), and the deterministic two-rail CPTP route introduced here (Theorem 3.3, Section 4.2). The comparison is between the two *approaches*, not between specific published works; entries on the right-hand column refer to the CPTP primitive of this paper. Quantities are per dissipative mode unless stated otherwise.

Quantity	Block-encoding route (post-selected)	Two-rail CPTP route (this paper)
Per-step success probability	$\leq  \lambda_r ^2 / \alpha_r^2$	1
End-to-end over $(T,  \Omega_h ,  \mathcal{D} )$	decays multiplicatively	1
Overrelaxation $\lambda_r < 0$	extra conditional structure	classically conditioned SWAP
Ancilla per mode	$\geq 1$ flag, post-selected	1 damping ancilla, reset
Post-selection / heralding	required	none
Amplitude amplification	typically required	not applicable
Qubits carrying $\delta m_r$	1 rail	2 rails (sign split)
Encoding of $\delta m_r$	linear	nonlinear (sign split; Rem. 3.2)
Scale $S_r$ handling	implicit in normalization	explicit classical side info

#### 4.5 Relation to recent QLBM work

Representative algorithmic lines in the recent literature include nonlinear fluid QLBM [11], linearized advection–diffusion lattice Boltzmann schemes with dynamic circuits and related linear-collision QLBM formulations [17, 20, 21], linear-equilibrium constructions that treat collision and streaming linear-algebraically and decompose the collision via SVD–LCU (singular value decomposition and linear combination of unitaries) [16], multi-circuit resource reduction [18], one-step device-oriented lattice Boltzmann algorithms [1], lattice Boltzmann–Carleman quantum algorithms and circuits for moderate Reynolds number flows [7], learned surrogate circuits for dissipative BGK collisions [15], and a recent synthesis of realizability issues [19]. With the exception of learned surrogates, all of the above face, to varying degrees, multiplicative post-selection or heralding penalties in their dissipative/contractive block: the mode-wise form (15) applies directly to block-encoded MRT schemes, and structurally analogous decays arise in LCU-based realizations of linear collision operators [16] and in block-encoded Carleman-linearized collision matrices [7] (with the caveat, in the Carleman case, that the non-unitarity is carried by a non-diagonal operator in velocity/Carleman-variable space rather than by a moment-basis diagonal  $\Lambda$ ). Theorem 3.3 provides a block-encoding-free alternative for the mode-diagonal case: it isolates the dissipative MRT block as a deterministic CPTP primitive, leaving equilibrium loading, streaming, and boundaries to be composed on top, as discussed in Section 5.

## 5 Integration in an MRT collision pipeline

Theorem 3.3 establishes a block-encoding-free CPTP realization of the dissipative MRT step. For that result to be useful, the primitive must compose inside a recognizable circuit pipeline. This section describes, at a schematic level, how the primitive slots into a *generic block-encoded MRT collision pipeline* at the site level, and comments on its relation to Carleman-linearized QLBM circuits [7], which share the same block-encoding bottleneck but do not expose a mode-wise diagonal target that the primitive could substitute for directly. The description is deliberately generic: the host pipeline is any quantum-compatible realization of the standard MRT collision stages on a site, and the primitive replaces only the contractive dissipative block within that host. Full coherent-state preparation, streaming, boundaries, and readout are not treated here.

### 5.1 A generic MRT collision pipeline

At the site level, any quantum-compatible realization of one MRT collision decomposes into four logically distinct stages, composed per timestep:

- (P1) **Equilibrium preparation.** Compute or encode the equilibrium  $m^{\text{eq}}(\rho, \mathbf{u})$  in an arithmetic register. The specific realization of this stage, including whether division is performed explicitly or postponed by a numerator-register representation, is part of the host pipeline and is not fixed here.
- (P2) **Nonequilibrium extraction.** Form  $\delta m_r$  for each dissipative mode  $r \in \mathcal{D}$  on the arithmetic register of the host pipeline.
- (P3) **Dissipative CPTP block (Theorem 3.3).** For each  $r \in \mathcal{D}$ , prepare the signed two-rail register  $\rho_r = \rho_r^+ \otimes \rho_r^-$  via (7), apply  $\mathcal{E}_r$  in (10), and decode (9) to obtain  $\lambda_r \delta m_r$  on the arithmetic register. The classical sign of  $\lambda_r$  selects the SWAP branch at compile time; no coherent conditional on  $\lambda_r$  is required.
- (P4) **Reconstruction.** Combine the relaxed  $\delta m'$  with  $m^{\text{eq}}$  and apply  $M^{-1}$  to obtain post-collision populations on the host's output register.

In compact notation, stages (P2)–(P3) realize the vector-form dissipative update

$$\delta m' = \Lambda \delta m, \quad \Lambda = \text{diag}\left((1)_{r \in \mathcal{C}}, (\lambda_r)_{r \in \mathcal{D}}\right), \quad (21)$$

on the moment register, where  $\Lambda$  is modewise diagonal by construction of MRT collision in the moment basis  $M$  and the conserved block of  $\Lambda$  is the identity. At each site  $\mathbf{x} \in \Omega_h$ , stage (P3) acts on the dissipative rails as the node-level composite channel

$$\mathcal{E}_{\mathbf{x}}^{\text{diss}} \stackrel{\text{def}}{=} \bigotimes_{r \in \mathcal{D}} \mathcal{E}_{r, \mathbf{x}}, \quad (22)$$

with each  $\mathcal{E}_{r, \mathbf{x}}$  given by (10);  $\mathcal{E}_{\mathbf{x}}^{\text{diss}}$  is CPTP as a tensor product of CPTP maps (Corollary 3.4) and reproduces the diagonal action of  $\Lambda$  restricted to  $\mathcal{D}$  on the decoded moments (Theorem 3.3, Remark 3.7). Because registers for different dissipative modes are disjoint, stage (P3) is parallel across  $r \in \mathcal{D}$ . Because (P3) is deterministic, *no success-probability discount accumulates across modes, sites, or timesteps*: the probabilistic cost of the whole dissipative sweep is  $p_{\text{succ}} = 1$ , matching Section 4.2.

### 5.2 Resource sketch per dissipative mode

At the level of abstract circuit primitives, the dissipative CPTP block (P3) for one mode  $r$  requires

- two rail qubits (the signed two-rail register);
- one damping ancilla per rail (reset after use), realizing  $\mathcal{A}_{\alpha_r}$  with survival  $\alpha_r = |\lambda_r|$  via the standard two-qubit damping gadget (one controlled rotation of angle  $2 \arccos \sqrt{\alpha_r}$  from rail to ancilla, one CNOT, one reset) [10, 13];
- one rail SWAP applied iff  $\lambda_r < 0$  (classically conditioned, compile-time);
- no heralding flag and no post-selection control wiring.

The angle  $2 \arccos \sqrt{\alpha_r}$  is classical, fixed by the lattice scheme, and can be computed once per mode. Across all  $|\mathcal{D}|$  dissipative modes and  $|\Omega_h|$  sites, the per-timestep dissipative-block cost is linear in  $|\mathcal{D}| |\Omega_h|$ , with no probability discount. A detailed gate-count breakdown for stages (P1), (P2), and (P4) depends on the specific host pipeline and is outside the scope of this paper.

### 5.3 Scope of drop-in replacement and relation to other QLBM pipelines

The dissipative CPTP block (P3) acts only on the signed two-rail register at a fixed site and mode, given a classical target relaxation multiplier  $\lambda_r$ . It is therefore pipeline-agnostic *within the MRT family*: any host circuit whose dissipative collision is implemented as a block-encoded mode-wise diagonal contraction  $\Lambda = \text{diag}(\lambda_r)$  on nonequilibrium moments (equivalently, any block-encoded MRT scheme that carries  $\delta m_r$  as an accessible register after a moment-basis transform) can replace its block-encoded dissipative block with Theorem 3.3, upgrading the contribution of that block to the end-to-end success probability from (15) to 1. The cost paid in exchange is (T1)–(T3) of Section 4.3, dominated by the classical nonlinear sign split; this split is free whenever  $\text{sign}(\delta m_r)$  is classically available from the host’s moment register.

Constructions in which the non-unitarity is *not* of mode-wise diagonal form in a moment basis lie outside this drop-in statement. The Carleman-linearized Lattice Boltzmann circuit of Sanavio and Succi [7] is representative: it linearizes the BGK collision by Carleman tensor-embedding rather than by MRT mode decomposition, and its non-unitary core is the full (non-sparse, non-block-diagonal) Carleman collision matrix, not a diagonal  $\Lambda$  on nonequilibrium moments. The present primitive does not substitute for that non-unitarity as such, and the Carleman truncation does not, on its own, expose MRT nonequilibrium moments  $\delta m_r$  with classical sign structure; making the primitive applicable in that setting would require, additionally, an explicit moment-basis transform of the Carleman state. It is nevertheless relevant to Carleman-LB as a *target for reformulation*: any Carleman-LB variant that exposes an explicit MRT mode decomposition of its relaxation block would gain the same replacement, at no probability cost for the dissipative block. A minimal such reformulation is sketched in Section 5.4 below.

### 5.4 A hybrid Carleman–CPTP pipeline

The observation that motivates this subsection is that the utility of Carleman linearization in QLBM is *the coherent evaluation of a nonlinear equilibrium*  $f^{\text{eq}}(\rho, \mathbf{u})$ , which is polynomial (typically quadratic) in  $f$ ; the use of Carleman for the dissipative BGK relaxation itself is a structural consequence of coupling equilibrium and relaxation in a single non-unitary operator, not an intrinsic requirement. Separating the two responsibilities, with Carleman used only for equilibrium and the MRT-diagonal open channel used for relaxation, yields a pipeline in which the present primitive applies without modification.

**Host-agnostic composite-map decomposition.** The combination of open-channel CPTP dissipation with the mainstream unitary machinery of QLBM is *not specific to Carleman linearization*: the same schema covers Carleman linearization, linear combination of unitaries (LCU), and block-encoding (BE) interchangeably, since all three are unitary (or block-encoded

unitary) realizations of *linear* operators and can therefore occupy the same pre/post slots. Abstracting away this choice, a single MRT collision step at a site can be written at the *composite-map* level as a sandwich of three logically distinct stages,

$$\Phi_{\text{step}} = \underbrace{\mathcal{U}_{\text{post}}}_{\substack{\text{unitary mainstream} \\ \text{(reconstruction,} \\ M^{-1}, \text{ streaming, } \dots)}} \circ \underbrace{\mathcal{E}^{\text{diss}}}_{\substack{\text{open-channel CPTP} \\ \text{(this paper,} \\ p_{\text{succ}}=1)}} \circ \underbrace{\mathcal{U}_{\text{pre}}}_{\substack{\text{unitary mainstream} \\ \text{(Carleman / LCU /} \\ \text{block-encoding,} \\ \text{moment extract, } \dots)}} \quad (23)$$

where  $\mathcal{U}_{\text{pre}}$  is any coherent (up to fixed subnormalizations absorbed in the host encoding) preparation that delivers the nonequilibrium moments  $\delta m$  on an accessible register,  $\mathcal{E}^{\text{diss}} = \bigotimes_{r \in \mathcal{D}} \mathcal{E}_r$  is the deterministic CPTP dissipation of (10) (site-level; cf. (22)), and  $\mathcal{U}_{\text{post}}$  is any coherent recombination and inverse moment transform that returns post-collision populations. Equation (23) is the general combination schema: *the mainstream unitary machinery is confined to  $\mathcal{U}_{\text{pre}}$  and  $\mathcal{U}_{\text{post}}$ , and the only intrinsically non-unitary stage is  $\mathcal{E}^{\text{diss}}$ , which is realized deterministically with unit per-step success probability.* Any dissipation-dependent multiplicative success-probability decay of the form (15) is therefore removed at the level of the collision step itself, regardless of which unitary mainstream method is chosen for  $\mathcal{U}_{\text{pre}}$  and  $\mathcal{U}_{\text{post}}$ . Concretely, (23) admits at least three interchangeable specializations, labeled (U1)–(U3) in the list below:

- (U1) **Block-encoding (BE) host.**  $\mathcal{U}_{\text{pre}}, \mathcal{U}_{\text{post}}$  are block-encoded realizations of the linear moment transform  $M$ , equilibrium subtraction, and inverse moment transform  $M^{-1}$ ; the present primitive  $\mathcal{E}^{\text{diss}}$  replaces the block-encoded, post-selected realization of the diagonal contraction  $\Lambda = \text{diag}(\lambda_r)$  on dissipative modes (Sections 5.1 and 5.3).
- (U2) **LCU host.**  $\mathcal{U}_{\text{pre}}, \mathcal{U}_{\text{post}}$  implement the linear stages through a linear combination of unitaries (LCU), with the non-unitary contractive factor handled separately. In extant LCU-based QLBM constructions the contractive factor is itself realized by LCU of a non-unitary collision matrix, either directly or via SVD of the collision: Liu, John, and Emerson [16] apply SVD followed by LCU to the non-unitary BGK collision in the velocity basis; the present primitive  $\mathcal{E}^{\text{diss}}$  replaces that LCU-of-a-contraction step, eliminating the associated penalty  $p_{\text{succ}} \leq \|A|\psi\rangle\|^2 / \|c\|_1^2$  (notation of Section 4.1) on the dissipative stage.
- (U3) **Carleman host (hybrid Carleman–CPTP).**  $\mathcal{U}_{\text{pre}}$  additionally contains the Carleman lift and coherent evaluation of the nonlinear equilibrium,  $f^{\text{eq}} = E_K \tilde{f}$ , detailed in (26) below;  $\mathcal{E}^{\text{diss}}$  replaces the dissipative part of the fused Carleman-BGK block [7].

Here BE stands for *block-encoding*, the standard embedding of a non-unitary map inside a larger unitary with ancilla post-selection (Section 4.1), and LCU for *linear combination of unitaries* (Section 4.1, paragraph on LCU); both are the mainstream coherent techniques for the linear stages of QLBM. The quantum-algebraic realization of  $\mathcal{E}^{\text{diss}}$  itself (amplitude-damping gadget, Stinespring dilation, rail SWAP under  $\lambda_r < 0$ ) is the subject of Section 3 and is not repeated here; (23) is stated at the composite-map level on purpose, so that the combination claim reads the same whatever the host’s unitary stack happens to be.

**Carleman equilibrium lift.** The remainder of this subsection works out, as a concrete example, the Carleman-host specialization (U3) of the general schema (23): we fix  $\mathcal{U}_{\text{pre}}$  and  $\mathcal{U}_{\text{post}}$  to the Carleman lift plus linear moment stages detailed in (26) below, keep  $\mathcal{E}^{\text{diss}}$  as the open-channel CPTP block of Section 3, and trace through the resulting composite map and probability budget. The parallel specializations (U1) and (U2) follow the same template with the Carleman lift dropped (no nonlinear equilibrium to evaluate). Specialization (U1) (BE host) coincides with the generic pipeline of Sections 5.1 and 5.3; specialization (U2) (LCU host) inherits the same pre/post schema, with the LCU success-probability bound of Section 4.1 (paragraph

“The same bottleneck under LCU”) replaced by  $p_{\text{succ}} = 1$  on the dissipative stage. Both are algebraically simpler than the Carleman case and are therefore not written out explicitly here.

Let  $K \geq 2$  denote the Carleman truncation order. For each site  $\mathbf{x}$ , introduce the truncated Carleman tower

$$\tilde{f} = f \oplus (f \otimes f) \oplus \cdots \oplus f^{\otimes K}, \quad (24)$$

amplitude-encoded on a register  $|\tilde{f}\rangle$  of the host pipeline. Because  $f^{\text{eq}}(\rho, \mathbf{u})$  is polynomial of degree  $\leq K$  in  $f$ , there exists a linear operator  $E_K$  on the Carleman tower such that

$$f^{\text{eq}} = E_K \tilde{f} \quad (25)$$

exactly up to the standard Carleman truncation error [7]. Realizing  $E_K$  as a block-encoded (or LCU) unitary is the conventional Carleman-LB task and is host-specific; the present subsection does not fix that realization.

**Pipeline composition.** Fixing an invertible moment basis  $M$  as in Section 2.1, with conserved/dissipative index split  $(\mathcal{C}, \mathcal{D})$  and modewise multipliers  $\lambda_r$  as in (5), the hybrid site-level update decomposes into the following stages, replacing the single fused Carleman-BGK block with a composition of linear stages and a mode-wise CPTP block:

$$\begin{aligned} \text{(H1)} \quad & |f\rangle \longrightarrow |\tilde{f}\rangle && \text{(Carleman lift to order } K), \\ \text{(H2)} \quad & m = Mf, \quad m^{\text{eq}} = ME_K \tilde{f} && \text{(linear; moment-basis transform),} \\ \text{(H3)} \quad & \delta m = m - m^{\text{eq}} && \text{(linear; on the moment register),} \\ \text{(H4)} \quad & \delta m'_r = \lambda_r \delta m_r && \text{(Theorem 3.3; CPTP open channel, } p_{\text{succ}} = 1), \\ \text{(H5)} \quad & m^+ = m^{\text{eq}} + \delta m' && \text{(linear; reconstruction),} \\ \text{(H6)} \quad & f^* = M^{-1} m^+ && \text{(linear; inverse moment transform).} \end{aligned} \quad (26)$$

Equivalently, at the composite-map level,

$$f^* = M^{-1}(ME_K \tilde{f} + \Lambda(Mf - ME_K \tilde{f})), \quad \Lambda = \text{diag}((1)_{r \in \mathcal{C}}, (\lambda_r)_{r \in \mathcal{D}}), \quad (27)$$

in which the action of  $\Lambda$  on the dissipative modes is realized by the CPTP channel  $\bigotimes_{r \in \mathcal{D}} \mathcal{E}_r$  of (10) on signed two-rail registers (Theorem 3.3, Remark 3.7). The resulting register-level pipeline, with (H1)–(H6) slotted into the host-agnostic decomposition (23), is summarized in Figure 3.

**Probability budget.** Under the hybrid decomposition (26):

- Stages (H2), (H3), (H5), (H6) are *linear* and unitary up to fixed subnormalizations absorbed in the host encoding of  $M$ ,  $M^{-1}$ , and arithmetic registers; they carry no *dissipation-dependent* success-probability loss.
- Stage (H4) is the only intrinsically non-unitary step in the dissipative collision. By Corollary 3.4 it is *deterministic*, so

$$p_{\text{succ}}^{\text{diss, end}} = \prod_{t=1}^T \prod_{\mathbf{x} \in \Omega_h} \prod_{r \in \mathcal{D}} 1 = 1, \quad (28)$$

removing the mode-wise penalty (15) that the fused Carleman-BGK block incurs.

- Stage (H1) is the standard Carleman lift. Its realization cost (depth, subnormalization, truncation order) is host-determined and shared with any Carleman-based pipeline [7]; it is *not* affected by the choice of dissipation primitive.

The hybrid pipeline therefore converts the success-probability penalty of the Carleman-BGK collision into the classical bookkeeping (T1)–(T3) of Section 4.3 on the relaxation block, while leaving the Carleman machinery strictly as a coherent equilibrium evaluator.

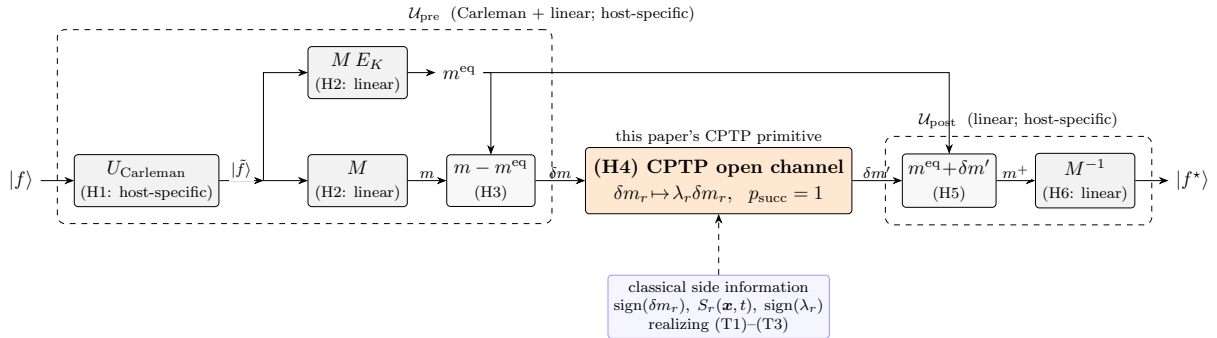


Figure 3: Register-level block diagram of the hybrid Carleman–CPTP collision step (26), realizing specialization (U3) of the host-agnostic schema (23). The outer dashed frames mark the induced grouping  $\mathcal{U}_{\text{pre}} = (\text{H1}) \circ (\text{H2}) \circ (\text{H3})$  and  $\mathcal{U}_{\text{post}} = (\text{H5}) \circ (\text{H6})$ . The orange block (H4) is the CPTP open channel of Theorem 3.3, whose signed two-rail internal structure is detailed in Figures 1–2 (block and gate level). The dashed arrow entering (H4) from below carries the classical side information  $(\text{sign}(\delta m_r), S_r(\mathbf{x}, t), \text{sign}(\lambda_r))$  realizing the bookkeeping items (T1)–(T3) of Section 4.3 (encoding-scale choice: Remark 3.1; adaptive scaling: Proposition 3.6).

**Scope of the claim.** The present paper proves (H4). Stages (H1)–(H3) and (H5)–(H6) are linear and standard in the LBM and Carleman-LB literature; a full specification and resource accounting of (H1) (Carleman truncation order, block-encoding of  $E_K$ , memory scaling in the tower (24)) and (H2), (H6) (coherent moment transforms) are host-design choices and are left to future work. The claim here is architectural: the hybrid pipeline (26) is a concrete, minimal example of a Carleman-based QLBM construction in which the dissipative relaxation is handled by the present CPTP primitive with unit per-step success probability, illustrating the “target for reformulation” statement at the end of Section 5.3.

## 5.5 Scope of this section

The claim of this section is that the dissipative block (P3) is host-agnostic and does not contribute a success-probability penalty to the broader pipeline. Quantitative resource estimates for (P1), (P2), and (P4), coherent preparation of the rails directly from a population register, multi-step composition with streaming and boundaries, and finite-shot readout are properties of the surrounding host, not of the primitive itself, and are therefore determined by the specific architecture in which the primitive is embedded.

## 6 Numerical consistency check

Because Theorem 3.3 is algebraic, we verify the dissipative open-channel construction numerically in two complementary ways. The D3Q19 audit in Section 6.1 uses nonequilibrium moment fields from a *full* collide–stream evolution (decaying Taylor–Green vortex) on the D3Q19 stencil, so the residuals in Table 2 and Figure 4 are probed on *realistic, flow-generated*  $\delta m_r$  data rather than on hand-chosen or purely random values alone. The stencil-free sweeps in Section 6.2 apply the *same* encoder (7), channel (10), and decoder (9) to inputs  $(\delta m_r, \lambda_r) \in [-X, X] \times [-1, 1]$  drawn with no connection to a lattice, thereby stress-testing the per-mode map independently of a particular velocity set or flow. Together, these two checks cover both a complete classical LBM time history in a standard 3D setting and a purely algebraic, stencil-agnostic stress test.

## 6.1 D3Q19 lattice audit (Taylor–Green vortex)

We check the open-channel construction on the D3Q19 stencil (Appendix A) by evaluating the maximum over dissipative modes and lattice sites of  $|\delta m'_r - \lambda_r \delta m_r|$  along the open-channel path. All tests use the *same* decaying Taylor–Green vortex run on the periodic box  $[0, L]^3$  with  $L = 2\pi$ ,  $N^3 = 64^3$ ,  $\tau = 0.5035$ ,  $u_0 = 0.1$ , and adaptive per-node rail scales  $S_r(\mathbf{x}, t) = \max(|\delta m_r|, \varepsilon)$ ; Table 2 summarizes five audits over the short window  $t \in [0, 19.6]$ , while Figure 4 shows the same error series continued over  $t \in [0, 981.7]$ . The reported time is the advective LBM time  $t = i(2\pi/N)u_0$  after  $i$  collide–stream iterations (per-iteration increment  $\approx 9.82 \times 10^{-3}$ ), equivalently  $t^* \equiv t/L$  box-crossings at  $u_0$ .

Five audits are reported. (i) Standard  $\lambda_r \in [-1, 1]$  from the D3Q19  $s_r$  at  $\tau = 0.5035$ , running maximum over the trajectory. (ii–iv) Endpoint regression with  $\lambda_r$  uniformly set to  $-1, 0, +1$  on all dissipative modes, applied to a representative Taylor–Green vortex (TGV) pre-collision snapshot, probing maximal overrelaxation (SWAP with  $\alpha_r = 1$ ), full dissipation ( $\alpha_r = 0$ ), and the identity branch. (v) Adaptive  $S_r$  coincides with (i) on this run. In all cases the decoded output matches  $\lambda_r \delta m_r$  to well within IEEE-754 double-precision roundoff.

Table 2: Maximum error  $\max_{r \in \mathcal{D}, \text{sites}} |\delta m'_r - \lambda_r \delta m_r|$  for the five audits of Theorem 3.3 on D3Q19, in IEEE-754 double precision; run defined in Section 6. Endpoint tests set all dissipative  $\lambda_r$  to the same value and act on a TGV pre-collision snapshot.

Test	Quantity audited	Max error
TGV trajectory, standard $\lambda_r$	$ \delta m'_r - \lambda_r \delta m_r $	$4.44 \times 10^{-16}$
Endpoint $\lambda_r = -1$ (SWAP, $\alpha = 1$ )	$ \delta m'_r + \delta m_r $	$6.16 \times 10^{-33}$
Endpoint $\lambda_r = 0$ ( $\alpha = 0$ )	$ \delta m'_r $	0
Endpoint $\lambda_r = +1$ (identity)	$ \delta m'_r - \delta m_r $	$6.16 \times 10^{-33}$
TGV trajectory, adaptive $S_r(\mathbf{x}, t)$	$ \delta m'_r - \lambda_r \delta m_r $	$4.44 \times 10^{-16}$

*Note.* For  $\lambda_r = \pm 1$  we have  $\alpha_r = 1$  (damping is the identity; a rail SWAP is applied only if  $\lambda_r = -1$ ). The  $6.16 \times 10^{-33}$  entries are *absolute* maxima, not  $O(1)$  relative accuracies: they are set by a few sites with extremely small  $|\delta m_r|$ , for which the floating-point Kraus implementation leaves a tiny residual of order  $\mathcal{O}(\varepsilon_{\text{mach}} |\delta m_r|)$  (with  $S_r = \max(|\delta m_r|, \varepsilon)$ ; Remark 3.1). The  $4.44 \times 10^{-16}$  trajectory rows are at the expected IEEE-754 double-precision scale for a full D3Q19 time history.

Figure 4 resolves the standard- $\lambda_r$  error per iteration over the extended window, confirming that the machine-precision agreement of Table 2 holds along the dynamics rather than at isolated snapshots.

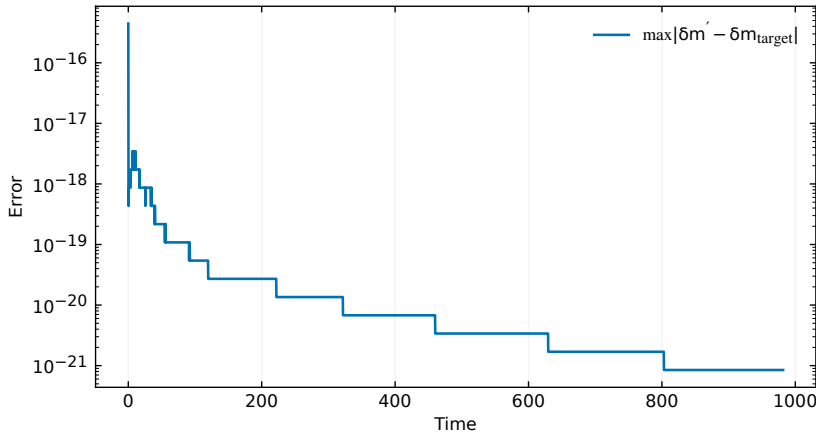


Figure 4: Dissipative-mode audit  $\max_{r \in \mathcal{D}, \text{sites}} |\delta m'_r - \lambda_r \delta m_r|$  versus advective LBM time for the run defined in Section 6.1. The error stays at the IEEE-754 double-precision floor throughout.

## 6.2 Stencil-free synthetic sweep

To confirm that the exactness of Theorem 3.3 is an algebraic property of the construction, not an artifact of a particular stencil or flow, we repeat the audit on inputs that are *not* drawn from any lattice simulation. Random pairs  $(\delta m_r, \lambda_r) \in [-X, X] \times [-1, 1]$  (with  $X = 1$ ) are fed through the same signed two-rail encoder (7), channel (10), and decoder (9) used in Section 6.1, with  $S_r = \max(|\delta m_r|, \varepsilon)$  in the adaptive mode, and the pointwise residual  $|\delta m'_r - \lambda_r \delta m_r|$  is aggregated over samples.

Five sweeps are reported in Table 3:

- (S1) **Dense  $\lambda$  grid.** 101 uniform  $\lambda$  values in  $[-1, 1]$  with 1000 random  $\delta m$  per  $\lambda$  ( $1.01 \times 10^5$  samples), covering the admissible range at resolution  $2 \times 10^{-2}$  in  $\lambda$ .
- (S2) **Uniform random  $(\delta m, \lambda)$ .**  $200 \times 200 = 4 \times 10^4$  samples drawn jointly uniformly, probing generic interior points.
- (S3) **Joint boundary stress.**  $\lambda \in \{-1, -1+\varepsilon_b, -\varepsilon_b, 0, +\varepsilon_b, 1-\varepsilon_b, 1\}$  with  $\varepsilon_b = 10^{-12}$ ; for each such  $\lambda$ ,  $10^4$  random interior  $\delta m \in [-X, X]$  are combined with the deterministic  $\delta m$ -edge vector  $\{-X, -X+\varepsilon_b, -\varepsilon_b, 0, +\varepsilon_b, X-\varepsilon_b, X\}$ , so the joint corners ( $\lambda$ -edge,  $\delta m$ -edge) are probed deterministically rather than only probabilistically. This regime includes exact overrelaxation ( $\lambda = -1$ ,  $\alpha_r = 1$ , SWAP), exact identity ( $\lambda = +1$ ), and exact full dissipation ( $\lambda = 0$ ).
- (S4) **Scale-sensitivity.** 200 random  $(\delta m, \lambda)$  pairs, each evaluated at 50 log-uniform scale values  $S_r \in [\max(|\delta m|, \varepsilon_S), 10^6 \max(|\delta m|, \varepsilon_S)]$  with  $\varepsilon_S = 10^{-12}$ , verifying that the decoded output continues to match  $\lambda \delta m$  to roundoff at every admissible  $S_r$ , consistent with Proposition 3.6.
- (S5) **Deterministic exact-corner audit.** The  $5 \times 5$  grid  $(\delta m, \lambda) \in \{-X, -X/2, 0, X/2, X\} \times \{-1, -1/2, 0, 1/2, 1\}$ , reader-verifiable by hand; this complements the random sweeps (S1)–(S4) with a fully reproducible grid of corner and half-corner points.

Table 3: Stencil-free synthetic audit of Theorem 3.3: maximum residual  $|\delta m'_r - \lambda_r \delta m_r|$  over random and deterministic draws on  $[-X, X] \times [-1, 1]$  with  $X = 1$ , in IEEE-754 double precision.

Sweep	#samples	Max error
(S1) Dense $\lambda$ grid, random $\delta m$	101,000	$1.11 \times 10^{-16}$
(S2) Uniform random $(\delta m, \lambda)$	40,000	$1.11 \times 10^{-16}$
(S3) Joint boundary stress ( $\lambda$ - and $\delta m$ -edges)	70,049	$1.11 \times 10^{-16}$
(S4) Scale sensitivity over $S_r$	10,000	$3.33 \times 10^{-16}$
(S5) Deterministic exact corners $5 \times 5$	25	$1.11 \times 10^{-16}$

All five sweeps saturate at the IEEE-754 double-precision floor (machine  $\varepsilon \approx 2.22 \times 10^{-16}$ ; the  $3.33 \times 10^{-16}$  entry for (S4) corresponds to at most  $1.5\varepsilon$ , consistent with a single division by  $S_r$  at decode). In particular, (S3) exhibits no degradation at the exact boundaries  $\lambda \in \{-1, 0, +1\}$  even when combined with  $|\delta m| = X$ , where the channel reduces to a permutation, a reset, and the identity respectively; (S4) confirms that varying  $S_r$  over six decades above  $|\delta m|$  leaves the decoded output matching  $\lambda \delta m$  at the roundoff floor, as required by Proposition 3.6; and (S5) returns machine-precision agreement at every point of the deterministic  $5 \times 5$  corner grid.

## 7 Conclusion

We proved that the signed diagonal MRT dissipative update

$$\delta m'_r = \lambda_r \delta m_r, \quad \lambda_r \in [-1, 1],$$

admits an exact decoded-expectation realization by a *deterministic* local CPTP channel on signed two-rail registers: amplitude damping with survival  $|\lambda_r|$  on each rail, followed by a rail SWAP in the overrelaxation branch  $\lambda_r < 0$  (Theorem 3.3). Because the channel is trace-preserving by construction, the per-step success probability of the dissipative block is unity and does not compound multiplicatively across modes, sites, or timesteps (Corollary 3.4), in contrast to block-encoded realizations of the same contraction whose end-to-end probability is bounded by (15). Adaptive scale choices  $S_r(\mathbf{x}, t)$  are classically indexed side information rather than hidden quantum resources (Proposition 3.6); within the present nonnegative population-encoding framework a single rail cannot exactly represent a signed moment (Theorem B.1, Appendix B), which motivates the two-rail split. Numerical audits, combining a D3Q19 decaying Taylor–Green vortex trajectory covering the standard  $\lambda_r$  used in that run, the endpoints  $\lambda_r \in \{-1, 0, +1\}$ , and adaptive scales (Table 2, Figure 4) with a stencil-free synthetic sweep over random and deterministic  $(\delta m_r, \lambda_r) \in [-X, X] \times [-1, 1]$ , including dense  $\lambda$  grids, uniform random pairs, joint boundary stress over both  $\lambda$ - and  $\delta m$ -edges, scale-sensitivity tests, and a deterministic  $5 \times 5$  exact-corner audit (Table 3), confirm machine-precision agreement with the target dissipative update across both the physical D3Q19 use case and the general algebraic domain.

This paper isolates the dissipative block of MRT collision as a block-encoding-free CPTP primitive and shows, schematically, how it slots into a generic block-encoded MRT collision pipeline (Section 5). Block-encoded Carleman-linearized LB circuits [7] share the same structural block-encoding bottleneck but do not directly admit the present primitive as a drop-in replacement, since their non-unitarity is not of mode-wise diagonal form in a moment basis; any Carleman-LB variant that exposes an explicit MRT mode decomposition of its relaxation block would, however, inherit the replacement, and a minimal hybrid pipeline of this kind, with Carleman linearization used strictly for the nonlinear equilibrium and open-channel CPTP used for the diagonal dissipative relaxation, is sketched in Section 5.4. Composition of the primitive into a full one-node D3Q19 MRT collision, together with exact multi-step lattice evolution under streaming and standard computational fluid dynamics (CFD) boundaries, is left to future work; Appendix C records only the algebraic consequence of embedding the primitive in a classical MRT baseline that shares the same moment basis and equilibrium.

## Appendices

### A D3Q19 stencil for the numerical audit

The numerical consistency check in Section 6 uses standard athermal D3Q19 MRT in the d’Humières/Lallemand–Luo ordering [9, 12] ( $q = 19$ , second-order  $f^{\text{eq}}$ ,  $m^{\text{eq}} = M f^{\text{eq}}$ , collision and streaming (3)–(4)). Conserved indices are  $\mathcal{C} = \{0, 3, 5, 7\}$  [9]; all other modes are dissipative. The tabulated relaxation parameters  $s_r$  are the usual D3Q19 entries used together with this ordering (as in [9, 12] and in the simulation that produces Figure 4); they are not fixed by the theorem.

### B One-rail no-go within the present population-encoding framework

This appendix records the obstruction that motivates the two-rail split in Section 3, *within the population-encoding framework of this paper*: diagonal one-qubit population states and population-expectation decoders of the form  $\rho \mapsto \sigma S \text{Tr}[n\rho]$ . It does not concern other one-qubit encodings (e.g. centered affine decoders on Bloch coordinates), which can realize a bounded signed scalar on one qubit.

**Theorem B.1** (No exact signed one-rail population encoding). *Fix  $X > 0$  and  $S > 0$ . Consider a one-rail population representation of a scalar  $x$  by diagonal one-qubit states*

$$\rho^{(1)}(x) = (1 - p(x)) |0\rangle\langle 0| + p(x) |1\rangle\langle 1|, \quad p(x) \in [0, 1], \quad (29)$$

together with a single-rail population decoder of the form

$$\mathcal{D}_{S,\sigma}^{(1)}(\rho) \stackrel{\text{def}}{=} \sigma S \text{Tr}[n\rho], \quad n = |1\rangle\langle 1|, \quad \sigma \in \{+1, -1\}. \quad (30)$$

Then there is no choice of  $p : [-X, X] \rightarrow [0, 1]$  such that

$$\mathcal{D}_{S,\sigma}^{(1)}(\rho^{(1)}(x)) = x \quad \text{for all } x \in [-X, X]. \quad (31)$$

Consequently, within this one-rail nonnegative population framework, a single rail cannot exactly encode a signed scalar on any domain containing both positive and negative values. In particular, the signed MRT dissipative update  $\delta m'_r = \lambda_r \delta m_r$  on such a domain cannot be realized exactly by a single nonnegative population rail with a population-expectation decoder.

*Proof.* If  $\sigma = +1$ , then for every state of the stated form,

$$\mathcal{D}_{S,+1}^{(1)}(\rho) = S \text{Tr}[n\rho] = S p(x) \in [0, S].$$

Hence the decoded values are nonnegative and cannot equal all  $x \in [-X, X]$ , which include negative values. If  $\sigma = -1$ , then  $\mathcal{D}_{S,-1}^{(1)}(\rho) = -S p(x) \in [-S, 0]$ , so the decoded values are nonpositive and cannot equal all  $x \in [-X, X]$ , which include positive values. Therefore no such one-rail population encoding with the stated decoder can exactly represent a signed scalar on a domain containing both signs.  $\square$

**Remark B.2** (Why sign reversal is the real obstruction). *Within a single nonnegative population rail, decoded values have a fixed sign by construction. The obstruction is therefore not the branch  $\lambda_r < 0$  per se, but the attempt to represent a signed scalar by a single nonnegative population expectation. The two-rail split of Section 3 resolves this by storing positive and negative parts separately and letting the decoder take their difference.*

## C Classical embedding: populations and one-step consequence

This appendix is not part of the main theorem: it records standard algebraic consequences if the dissipative update of Theorem 3.3 is combined with the same moment reconstruction and streaming as the classical MRT baseline.

Theorem 3.3 gives  $\delta m'_r = \lambda_r \delta m_r$  for each  $r \in \mathcal{D}$ . For conserved modes, define  $\delta m'_r = \delta m_r$ . The full nodewise update is therefore

$$\delta m'_r = \begin{cases} \delta m_r, & r \in \mathcal{C}, \\ \lambda_r \delta m_r, & r \in \mathcal{D}. \end{cases} \quad (32)$$

Set

$$m^+ = m^{\text{eq}} + \delta m', \quad f_{\text{open}}^* = M^{-1} m^+. \quad (33)$$

**Assumption C.1** (Shared baseline data). *The open-system construction and the classical MRT baseline use the same moment basis  $M$  and the same equilibrium map  $m^{\text{eq}}(\rho, \mathbf{u})$ .*

**Proposition C.2** (Matching post-collision populations under shared baseline). *Assume Assumption C.1. When the channel of Theorem 3.3 is used as the dissipative module with scales  $S_r = S_r(\mathbf{x}, t)$  chosen so that  $p_r^\pm \in [0, 1]$  at each node and with the same  $S_r$  used in encoding and decoding, the vector  $f_{\text{open}}^*$  defined in (33) equals the classical post-collision vector  $f^*$  in (3).*

*Proof.* For conserved modes, (32) is the identity update. For dissipative modes, Theorem 3.3 gives  $\delta m'_r = \lambda_r \delta m_r = (1 - s_r) \delta m_r$ . Hence  $\delta m'$  coincides with the classical MRT-updated nonequilibrium moment vector. Substituting into (33) and (3) yields the same  $m^+$  and therefore the same  $f^* = M^{-1}m^+$ .  $\square$

**Remark C.3** (One-step map and Chapman–Enskog). *Under Assumption C.1, if decoded post-collision populations agree with classical MRT at each node (Proposition C.2) and streaming is the classical permutation (4), then the deterministic one-step update agrees with the corresponding classical MRT-LBM scheme, and standard Chapman–Enskog analysis for that classical scheme applies unchanged. This observation is included only to connect the dissipative primitive to the classical discrete dynamics.*

## References

- [1] A. Bastida-Zamora *et al.*, “Quantum algorithm for the lattice Boltzmann method with applications on real quantum devices,” *arXiv preprint* arXiv:2603.02127 (2026); <https://doi.org/10.48550/arXiv.2603.02127>.
- [2] E. Choi, J. E. Kim, X. Lu, and Y. Wang, “Lindbladian Homotopy Analysis Method to Solve Nonlinear Partial Differential Equations,” *arXiv preprint* arXiv:2604.18924 (2026); <https://doi.org/10.48550/arXiv.2604.18924>.
- [3] F. J. Higuera and S. Succi, “Simulating the Flow Around a Circular Cylinder with a Lattice Boltzmann Equation,” *Europhysics Letters* **8**(6), 517 (1989); <https://doi.org/10.1209/0295-5075/8/6/005>.
- [4] F. J. Higuera, S. Succi, and R. Benzi, “Lattice Gas Dynamics with Enhanced Collisions,” *Europhysics Letters* **9**(4), 345 (1989); <https://doi.org/10.1209/0295-5075/9/4/008>.
- [5] F. J. Higuera and J. Jiménez, “Boltzmann Approach to Lattice Gas Simulations,” *Europhysics Letters* **9**(7), 663 (1989); <https://doi.org/10.1209/0295-5075/9/7/009>.
- [6] R. Benzi, S. Succi, and M. Vergassola, “The lattice Boltzmann equation: theory and applications,” *Physics Reports* **222**(3), 145–197 (1992); [https://doi.org/10.1016/0370-1573\(92\)90090-M](https://doi.org/10.1016/0370-1573(92)90090-M).
- [7] C. Sanavio and S. Succi, “Lattice Boltzmann–Carleman quantum algorithm and circuit for fluid flows at moderate Reynolds number,” *AVS Quantum Science* **6**(2), 023802 (2024); <https://doi.org/10.1116/5.0195549>.
- [8] S. Succi, W. Itani, K. R. Sreenivasan, and R. Steijl, “Quantum computing for fluids: Where do we stand?” *Europhysics Letters* **144**(1), 10001 (2023); <https://doi.org/10.1209/0295-5075/acfdc7>.
- [9] D. d’Humières, I. Ginzburg, M. Krafczyk, P. Lallemand, and L.-S. Luo, “Multiple-relaxation-time lattice Boltzmann models in three dimensions,” *Philosophical Transactions of the Royal Society A* **360** (2002), 437–451; <https://doi.org/10.1098/rsta.2001.0955>.
- [10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, Cambridge, 2010; ISBN 978-1-139-49548-6; <https://books.google.it/books?id=-s4DEy7o-a0C>.
- [11] B. Wang, Z. Meng, Y. Zhao, and Y. Yang, “Quantum lattice Boltzmann method for simulating nonlinear fluid dynamics,” *npj Quantum Information* **11**(1), 196 (2025); <https://doi.org/10.1038/s41534-025-01142-6>.

- [12] P. Lallemand and L.-S. Luo, “Theory of the lattice Boltzmann method: Acoustic and thermal properties in two and three dimensions,” *Phys. Rev. E* **68**(3), 036706 (2003); <https://doi.org/10.1103/PhysRevE.68.036706>.
- [13] J. Watrous, *The Theory of Quantum Information*, Cambridge University Press, Cambridge, 2018; ISBN 978-1-316-85312-2; <https://books.google.it/books?id=jQFWDwAAQBAJ>.
- [14] D. Manzano, “A short introduction to the Lindblad master equation,” *AIP Advances* **10**(2), 025106 (2020); <https://doi.org/10.1063/1.5115323>.
- [15] M. Lăcătuș and M. Möller, “Surrogate Quantum Circuit Design for the Lattice Boltzmann Collision Operator,” *International Journal for Numerical Methods in Engineering* **127**(4), e70286 (2026); <https://doi.org/10.1002/nme.70286>.
- [16] Z. Liu, B. John, and D. R. Emerson, “Quantum Lattice Boltzmann Method based on linear equilibrium distribution functions,” *Frontiers in Mechanical Engineering*, Volume 11 (2025), Article 1717775 (2026); <https://doi.org/10.3389/fmech.2025.1717775>.
- [17] D. Wawrzyniak, J. Winter, S. Schmidt, T. Indiniger, C. F. Janßen, U. Schramm, and N. A. Adams, “Linearized quantum lattice-Boltzmann method for the advection-diffusion equation using dynamic circuits,” *Computer Physics Communications* **317**, 109856 (2025); <https://doi.org/10.1016/j.cpc.2025.109856>.
- [18] M. Lee, Z. Song, S. Kocherla, A. Adams, A. Alexeev, and S. H. Bryngelson, “A multiple-circuit approach to quantum resource reduction with application to the quantum lattice Boltzmann method,” *Future Gener. Comput. Syst.* **174**, 107975 (2026); <https://doi.org/10.1016/j.future.2025.107975>.
- [19] A. Tiwari, J. Iaconis, J. Jojo, S. Ray, M. Roetteler, C. Hill, and J. Pathak, “Algorithmic Advances Towards a Realizable Quantum Lattice Boltzmann Method,” *arXiv preprint arXiv:2504.10870* (2025); <https://arxiv.org/abs/2504.10870>.
- [20] L. Xu, M. Li, L. Zhang, H. Sun, and J. Yao, “Improved quantum lattice Boltzmann method for advection-diffusion equations with a linear collision model,” *Phys. Rev. E* **111**(4), 045305 (2025); <https://doi.org/10.1103/PhysRevE.111.045305>.
- [21] K.-Y. Zeng, X.-D. Niu, A. Khan, D.-C. Li, and H. Yamaguchi, “A quantum computing-based lattice Boltzmann method with a linearized non-equilibrium collision operator and modular circuit for practical flow simulation,” *Physics of Fluids* **37**(8), 081701 (2025); <https://doi.org/10.1063/5.0278054>.