

A concurrent global–local numerical method for multiscale parabolic equations

YULEI LIAO

Department of Mathematics, Faculty of Science, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Singapore

YANG LIU

State Key Laboratory of Mathematical Sciences, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Beijing 100190, China

School of Mathematical Sciences, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China

AND

PINGBING MING*

State Key Laboratory of Mathematical Sciences, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Beijing 100190, China

School of Mathematical Sciences, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China

*Corresponding author: mpb@lsec.cc.ac.cn

[Received on 19 June 2025; revised on 14 December 2025]

This paper presents a concurrent global–local numerical method for solving multiscale parabolic equations in divergence form. The proposed method employs hybrid coefficient to provide accurate macroscopic information while preserving essential microscopic details within specified local defects. Both the macroscopic and microscopic errors have been improved compared with existing results, eliminating the factor of $\Delta t^{-1/2}$ when the diffusion coefficient is time-independent. Numerical experiments demonstrate that the proposed method effectively captures both global and local solution behaviours.

Keywords: concurrent global–local method; heterogeneous multiscale method; multiscale parabolic equation; local error estimate.

1. Introduction

Multiscale problems are pervasive in various scientific and engineering disciplines, characterized by phenomena occurring at multiple spatial and temporal scales (Fish, 2009). Within this broad field, multiscale parabolic equations are fundamental in modelling diffusion processes in heterogeneous media. These equations often involve diffusion coefficients that vary significantly across different regions, posing additional difficulties for numerical simulations. In the present work, we focus on a multiscale parabolic equation in divergence form with Dirichlet boundary condition:

$$\begin{cases} \partial_t u^\varepsilon(x, t) - \nabla \cdot (a^\varepsilon(x) \nabla u^\varepsilon(x, t)) &= f(x, t), & (x, t) \in D \times (0, T) =: D_T, \\ u^\varepsilon(x, 0) &= u_0(x), & (x, t) \in D \times \{0\}, \\ u^\varepsilon(x, t) &= 0, & (x, t) \in \partial D \times (0, T), \end{cases} \quad (1.1)$$

where $D \subset \mathbb{R}^n$ is a convex bounded domain and $\varepsilon > 0$ is a small parameter that signifies the multiscale nature of the problem. The source term $f \in L^2(0, T; H^{-1}(D))$ and initial value $u_0 \in H_0^1(D)$. The time-independent rough diffusion coefficient $a^\varepsilon(x)$ lies in $\mathcal{M}(\lambda, \Lambda; D)$ for some $\lambda, \Lambda > 0$, where

$$\mathcal{M}(\lambda, \Lambda; D) := \left\{ a \in [L^\infty(D)]^{n \times n} \mid \xi \cdot a(x)\xi \geq \lambda|\xi|^2, \xi \cdot a(x)\xi \leq \frac{1}{\Lambda}|a(x)\xi|^2 \right. \\ \left. \text{for any } \xi \in \mathbb{R}^n \text{ and a.e. } x \in D \right\}.$$

On the analytic level, the homogenization theory, particularly H-convergence, provides a powerful framework for approximating the behaviour of such multiscale systems by averaging the effects of small-scale variations. For a sequence $\{a^\varepsilon(x)\}_{\varepsilon>0} \in \mathcal{M}(\lambda, \Lambda; D)$, there exists a subsequence (still denoted by $\{a^\varepsilon(x)\}_{\varepsilon>0}$ for convenience) and $A(x) \in \mathcal{M}(\lambda, \Lambda; D)$ such that $a^\varepsilon(x)$ H-converges to $A(x)$ (Tartar, 2010). Set $u(x, t)$ the solution to the corresponding homogenization problem

$$\begin{cases} \partial_t u(x, t) - \nabla \cdot (A(x) \nabla u(x, t)) = f(x, t), & (x, t) \in D_T, \\ u(x, 0) = u_0(x), & (x, t) \in D \times \{0\}, \\ u(x, t) = 0, & (x, t) \in \partial D \times (0, T). \end{cases} \quad (1.2)$$

According to Brahim-Otsmane *et al.* (1992); Bensoussan *et al.* (2011), the solutions u^ε and u exist in $L^2(0, T; H_0^1(D))$, and satisfy

$$\begin{cases} u^\varepsilon \rightharpoonup u & \text{weakly in } L^2(0, T; H_0^1(D)), \\ a^\varepsilon \nabla u^\varepsilon \rightharpoonup A \nabla u & \text{weakly in } [L^2(0, T; L^2(D))]^d, \\ \partial_t u^\varepsilon \rightharpoonup \partial_t u & \text{weakly in } L^2(0, T; H^{-1}(D)). \end{cases}$$

Numerous numerical methods founded on the idea of homogenization have been suggested and extensively examined in existing studies such as multiscale finite-element method (Hou & Wu, 1997) and heterogeneous multiscale method (HMM) (Abdulle & E, 2003; E & Engquist, 2003; Chen *et al.*, 2005). HMM aims to capture the macroscopic behaviour of the equation without microscopic details. It consists of the macroscopic solver to approximate the homogenization solution u , which rely on the effective matrix A and the microscopic solver to approximate this effective matrix. Review on HMM may be found in E *et al.* (2007) and Abdulle *et al.* (2012).

In this work, we shall employ an HMM-type method to capture the microscopic information in a small defect $K_0 \subset D$ and macroscopic information outside the region. For such scenario, there are two types of numerical methods. The first one is based on domain decomposition (Hu & Gunzburger, 2000; Gervasio *et al.*, 2001; Glowinski *et al.*, 2004; Apoung Kamga & Pironneau, 2007; Lozinski & Pironneau, 2011), particularly, the Arlequin method (Dhia & Rateau, 2005; Albella *et al.*, 2019; Gorynina *et al.*, 2021). These methods apply different operators to different subdomains of the computational domain. The second one is the global–local method (Oden & Vemaganti, 2000, 2001) that compute the homogenized solution u firstly, and then exploit it as the boundary condition to approximate the multiscale solution u^ε on a local region. Following Huang *et al.* (2018); Ming & Song (2022) several concurrent global–local methods have been proposed for multiscale elliptic equations with matching and nonmatching grids, respectively. The key idea is to hybridize the microscopic and macroscopic coefficients by a function as

$b^\varepsilon(x) = \rho(x)a^\varepsilon(x) + (1 - \rho(x))A(x)$, where the transition function $\rho = 1$ in K_0 and vanishes away from K_0 without any smoothness requirement.

In this work, we extend the concurrent global-local method proposed in [Huang et al. \(2018\)](#) to the multiscale parabolic problems. Specifically, we plan to study the microscopic information in a subdomain $K_0 \subset D$, with $\text{diam}K_0 = d$. We slightly extend K_0 , i.e., set $K_0 \subset K$, with $\text{dist}(K_0, \partial K \setminus \partial D) = cd$, where c is a positive universal constant. The hybrid coefficient $b^\varepsilon(x) = \rho(x)a^\varepsilon(x) + (1 - \rho(x))A(x)$, where $\rho(x)$ is a transition function which satisfies

$$\begin{cases} \rho(x) = 1, & x \in K_0, \\ 0 \leq \rho(x) \leq 1, & x \in K, \\ \rho(x) = 0, & x \in D \setminus K. \end{cases}$$

An interesting finding in Theorem 10 suggests that the pollution of the local energy error may be reduced if we take ρ as the indicator function of K .

We use the backward Euler scheme as the temporal discretization with a uniform time step Δt and the Lagrange linear finite element to approximate the solution space with local mesh size h over K and global mesh size H over D . Specifically, let $\mathcal{T}_{h,H}$ be a shape-regular triangulation of D in the sense of [Ciarlet \(2002\)](#), which satisfies

$$h = \max_{T \in \mathcal{T}_{h,H}, T \cap K \neq \emptyset} h_T \quad \text{and} \quad H = \max_{T \in \mathcal{T}_{h,H}} h_T,$$

and the Lagrange linear finite-element space is defined as

$$X_{h,H} := \{v \in H_0^1(D) \mid v|_T \in \mathbb{P}_1(T) \quad \text{for all} \quad T \in \mathcal{T}_{h,H}\}.$$

Generally, we do not have the information of the effective matrix A . In practice, the hybrid coefficient $b_H^\varepsilon(x) = \rho(x)a^\varepsilon(x) + (1 - \rho(x))A_H(x)$, where A_H is the approximation of A obtained by HMM or other types of numerical homogenization methods. In the framework of HMM, we solve several cell problems over cubes with length δ_0 to approximate A_H , and use $e(\text{HMM})$ measuring error caused by such approximation; e.g., for local periodic $a^\varepsilon(x) = a(x, x/\varepsilon)$ with $a(x, y)$ periodic in y , it is consistent with the elliptic cases ([E et al., 2005](#); [Du & Ming, 2010](#)) that $e(\text{HMM})$ is $\mathcal{O}(\delta_0 + \varepsilon/\delta_0)$. Denote $t_k = k\Delta t$, $f_k(x) = f(x, t_k)$, $u_k(x) = u(x, t_k)$ and $u_k^\varepsilon(x) = u^\varepsilon(x, t_k)$. To numerically solve the equation, we find $U_k \in X_{h,H}$, $k = 1, \dots, m$ such that $t_m = T$ and

$$(\delta U_k, V)_D + (b_H^\varepsilon \nabla U_k, \nabla V)_D = (f_k, V)_D \quad \text{for all} \quad V \in X_{h,H}, \quad (1.3)$$

where $\delta U_k = (U_k - U_{k-1})/\Delta t$, and the initial value $U_0 \in X_{h,H}$ is the solution of the following variational problem such that

$$(b_H^\varepsilon \nabla U_0, \nabla V)_D = (b_H^\varepsilon \nabla u_0, \nabla V)_D \quad \text{for all} \quad V \in X_{h,H}. \quad (1.4)$$

To judge whether U_m can successfully capture the global information u over D and the local information u^ε over the local region K_0 , we shall analyse this concurrent scheme. Unlike the multiscale elliptic equations ([Huang et al., 2018](#)), time evolution presents further analytical difficulties. As shown in

Ming & Zhang (2007); Eckhardt & Verfürth (2023), the HMM approximates the homogenization solution u for multiscale parabolic equations with the H^1 -norm error estimate $\mathcal{O}(\Delta t + H + e(\text{HMM})\Delta t^{-1/2})$. The factor $\Delta t^{-1/2}$ is frustrating since it indicates that the time step cannot be arbitrarily small. In Theorem 6, we eliminate this factor for time-independent coefficient a^ε , and the global H^1 -error reads as

$$\|u_m - U_m\|_{H^1(D)} \leq C(\lambda, \Lambda, T, u)(\Delta t + H + \eta(K) + e(\text{HMM})),$$

where $\eta(K)$ measures the pollution caused by the local region K , which also appears in multiscale elliptic equations (Huang *et al.*, 2018). We achieve this improvement with the aid of a Galerkin projection operator (3.4) onto the finite-element space for an auxiliary elliptic problem, not the projection for original parabolic equations (1.2) in Ming & Zhang (2007); Eckhardt & Verfürth (2023).

Since, we only concentrate on the microscopic information on a local region K_0 , we expect the convergence of this concurrent scheme to the multiscale solution u^ε in an interior energy norm. Studies for interior energy error estimates can be traced back to Nitsche & Schatz (1974); Bramble *et al.* (1975) for approximating second-order elliptic equations on quasi-uniform meshes. Later, similar results have been proven on shape-regular meshes in Demlow *et al.* (2011) using a novel superapproximation result. They found that the interior H^1 -error may be bounded by a local approximation term plus a global pollution term, like $\mathcal{O}(h + H^2/d)$. These results are useful to analyse several local and parallel algorithms such as two-grid methods for elliptic-type equations (Xu & Zhou, 2001; Wang & Du, 2023).

For parabolic equations, we discuss interior energy errors in space $L^2(0, T; H^1(K_0))$, or the discrete corresponding errors in space $\ell^2(0, T; H^1(K_0))$, with

$$\|f\|_{L^p(0, T; X)}^p := \int_0^T \|f(\cdot, t)\|_X^p dt \quad \text{and} \quad \|f\|_{\ell^p(0, T; X)}^p := \Delta t \sum_{k=1}^m \|f_k\|_X^p,$$

for a Banach space X . Thomée and co-workers (Bramble *et al.*, 1977; Thomée, 1979; Schatz *et al.*, 1998) have extended Nitsche's work (Nitsche & Schatz, 1974; Bramble *et al.*, 1975) for parabolic equations with semidiscrete schemes. Unfortunately, these works are not suitable for the multiscale problems, because they require the smoothness of the coefficient a^ε and the solution u^ε . We also note that the local H^1 -error for the two-grid algorithm in Xu & Zhou (2001) for transient Stokes equations (Shang & Wang, 2010) reads as $\mathcal{O}(\Delta t^{-1/2}(\Delta t + h + H^2))$. All these works hang on the interior energy stability as a bridge. We follow another way to derive the interior energy estimate by considering $\partial_t u^\varepsilon$ as an extra source term of the elliptic equation; see Theorem 10 and Remark 4 that

$$\begin{aligned} \|\nabla(u^\varepsilon - U)\|_{\ell^2(0, T; L^2(K_0))} &\leq C(\lambda, \Lambda, T, u^\varepsilon, u) \left(\inf_{V \in [X_{h, H}]^m} \|\nabla(u^\varepsilon - V)\|_{L^2(K)} \right. \\ &\quad \left. + d^{-1}(|K|^{1/2-1/p} + \Delta t + H^2 + \eta(K) + e(\text{HMM})) \right), \end{aligned}$$

with Meyer's regularity exponent $p > 2$ (Bensoussan *et al.*, 2011). Hence, only Meyer's regularity for u^ε required and we eliminate the frustrating factor $\Delta t^{-1/2}$ for the interior energy error over K_0 .

The rest of the paper is organized as follows. In Section 2, we study the H-limit of the mixing of the general time-dependent hybrid coefficient for parabolic equations, which may be of independent interest. In Section 3, we present the accuracy for retrieving the macroscopic and microscopic information, which are the main technique part of this work. In Section 4, we provide several numerical experiments which

are adopted from [Huang *et al.* \(2018\)](#); [Ming & Song \(2022\)](#), fitted theoretical predictions. Finally, Section 5 concludes the paper with a summary and potential future research directions.

Throughout this paper, we shall use standard Sobolev spaces $W^{m,p}(\Omega)$ for any bounded domain $\Omega \subset \mathbb{R}^n$. Without further explanation is provided, the constant C only depends on λ , Λ , the end time T and the domain D , and is independent of ε , the diam of the local region d , time step Δt and the mesh sizes h, H .

2. H-convergence for parabolic equations

Before analysing the convergence of the method, we firstly examine the homogenization of the mixing of coefficients, following the elliptic case in [Huang *et al.* \(2018, § 2\)](#). Under H-convergence, a time-independent coefficient yields the same effective limit as in the elliptic problem, which suffices for our purposes. Here, we slightly extend the setting and study the case of time-dependent coefficients. We keep the notation unchanged, as this does not constitute an abuse of notation and should not cause confusion. Specifically, if $a^\varepsilon(x, t) \in \mathcal{M}(\lambda, \Lambda; D_T)$, where $D_T = D \times [0, T]$ and

$$\mathcal{M}(\lambda, \Lambda; D_T) := \left\{ a \in [L^\infty(D_T)]^{n \times n} \mid a(\cdot, t) \in \mathcal{M}(\lambda, \Lambda; D) \text{ for all } t \in [0, T] \right\},$$

then by H-convergence theory, there exists a subsequence of the sequence $a^\varepsilon(x, t)$ that H-converge ([Zhikov *et al.*, 1981](#); [Dall'Aglio & Murat, 1997](#)) to certain $A(x, t) \in \mathcal{M}(\lambda, \Lambda; D_T)$. In continuous sense, let the hybrid coefficient $b^\varepsilon(x, t) = \rho(x, t)a^\varepsilon(x, t) + (1 - \rho(x, t))A(x, t)$ for some cut-off function $\rho(x, t) : \Omega \times [0, T] \rightarrow [0, 1]$. Similarly, $b^\varepsilon(x, t) \in \mathcal{M}(\lambda, \Lambda; D_T)$ and there exists a subsequence of b^ε that H-converge to certain $B(x, t) \in \mathcal{M}(\lambda, \Lambda; D_T)$.

In this part, we measure the difference between $A(x, t)$ and $B(x, t)$. For the sake of convenience, we still denote the subsequences by $a^\varepsilon(x, t)$ and $b^\varepsilon(x, t)$. Denote the $L^2(D_T)$ inner product by $(\cdot, \cdot)_{D_T}$, and the duality pairing between $L^2(0, T; H^{-1}(D))$ and $L^2(0, T; H_0^1(D))$ by $\langle \cdot, \cdot \rangle_{D_T}$.

PROPOSITION 1. There holds

$$|A(x, t) - B(x, t)| \leq 2\Lambda(\Lambda/\lambda + \sqrt{\Lambda/\lambda})\rho(x, t)(1 - \rho(x, t)) \text{ a.e. } (x, t) \in D_T. \quad (2.1)$$

Here $|\cdot|$ is the ℓ^2 -norm of a matrix.

The proof is essentially a combination of [Antonić & Vrdoljak \(2009, Theorem 1\)](#) and [Huang *et al.* \(2018, Proposition 2.1\)](#).

Proof. Define

$$\mathcal{W} := \{v \in L^2(0, T; H_0^1(D)) \mid \partial_t v \in L^2(0, T; H^{-1}(D))\},$$

and subspaces

$$\mathcal{W}_0 := \{v \in \mathcal{W} \mid v(x, 0) = 0\}, \quad \mathcal{W}_T := \{v \in \mathcal{W} \mid v(x, T) = 0\}.$$

For given $\phi_0 \in \mathcal{W}_0$ and $\psi_0 \in \mathcal{W}_T$, we define $\phi^\varepsilon \in \mathcal{W}_0$ as the unique solution of the initial-boundary value problem

$$\begin{cases} \partial_t \phi^\varepsilon - \nabla \cdot (a^\varepsilon \nabla \phi^\varepsilon) = \partial_t \phi_0 - \nabla \cdot (A \nabla \phi_0), & \text{in } D_T, \\ \phi^\varepsilon \in \mathcal{W}_0, \end{cases} \quad (2.2)$$

and $\psi^\varepsilon \in \mathcal{W}_T$ as the unique solution of the backward-in-time problem

$$\begin{cases} -\partial_t \psi^\varepsilon - \nabla \cdot ((b^\varepsilon)^\top \nabla \psi^\varepsilon) = -\partial_t \psi_0 - \nabla \cdot (B^\top \nabla \psi_0), & \text{in } D_T. \\ \psi^\varepsilon \in \mathcal{W}_T. \end{cases} \quad (2.3)$$

Here $(b^\varepsilon)^\top$ is the transpose of the matrix b^ε . By H-convergence theory, we have

$$\begin{aligned} \phi^\varepsilon &\rightharpoonup \phi_0, & \psi^\varepsilon &\rightharpoonup \psi_0 && \text{weakly in } L^2(0, T; H_0^1(D)), \\ a^\varepsilon \nabla \phi^\varepsilon &\rightharpoonup A \nabla \phi_0, & (b^\varepsilon)^\top \nabla \psi^\varepsilon &\rightharpoonup B^\top \nabla \psi_0 && \text{weakly in } [L^2(0, T; L^2(D))]^d, \\ \partial_t \phi^\varepsilon &\rightharpoonup \partial_t \phi_0, & \partial_t \psi^\varepsilon &\rightharpoonup \partial_t \psi_0 && \text{weakly in } L^2(0, T; H^{-1}(D)). \end{aligned}$$

After multiplying the above equations by $\varphi \psi^\varepsilon$ and $\varphi \phi^\varepsilon$, respectively, for some $\varphi \in C_c^\infty(D_T)$, we obtain

$$\langle \partial_t \phi^\varepsilon, \varphi \psi^\varepsilon \rangle_{D_T} + (a^\varepsilon \nabla \phi^\varepsilon, \nabla (\varphi \psi^\varepsilon))_{D_T} = \langle \partial_t \phi_0, \varphi \psi^\varepsilon \rangle_{D_T} + (A \nabla \phi_0, \nabla (\varphi \psi^\varepsilon))_{D_T}$$

and

$$-\langle \partial_t \psi^\varepsilon, \varphi \phi^\varepsilon \rangle_{D_T} + ((b^\varepsilon)^\top \nabla \psi^\varepsilon, \nabla (\varphi \phi^\varepsilon))_{D_T} = -\langle \partial_t \psi_0, \varphi \phi^\varepsilon \rangle_{D_T} + (B^\top \nabla \psi_0, \nabla (\varphi \phi^\varepsilon))_{D_T}.$$

By subtracting these equalities, after using the identity $\langle \partial_t \phi^\varepsilon, \varphi \psi^\varepsilon \rangle_{D_T} = -\langle \psi^\varepsilon \partial_t \varphi + \varphi \partial_t \psi^\varepsilon, \phi^\varepsilon \rangle_{D_T}$ and $\langle \partial_t \phi_0, \varphi \psi^\varepsilon \rangle_{D_T} = -\langle \psi^\varepsilon \partial_t \varphi + \varphi \partial_t \psi^\varepsilon, \phi_0 \rangle_{D_T}$ we have

$$\begin{aligned} &-(\phi^\varepsilon, \psi^\varepsilon \partial_t \varphi)_{D_T} + (a^\varepsilon \nabla \phi^\varepsilon, \psi^\varepsilon \nabla \varphi + \varphi \nabla \psi^\varepsilon)_{D_T} - ((b^\varepsilon)^\top \nabla \psi^\varepsilon, \phi^\varepsilon \nabla \varphi + \varphi \nabla \phi^\varepsilon)_{D_T} \\ &= \langle \partial_t \psi_0, \varphi \phi^\varepsilon \rangle_{D_T} - \langle \partial_t \psi^\varepsilon, \varphi \phi_0 \rangle_{D_T} - (\phi_0, \psi^\varepsilon \partial_t \varphi)_{D_T} \\ &+ (A \nabla \phi_0, \psi^\varepsilon \nabla \varphi + \varphi \nabla \psi^\varepsilon)_{D_T} - (B^\top \nabla \psi_0, \phi^\varepsilon \nabla \varphi + \varphi \nabla \phi^\varepsilon)_{D_T}. \end{aligned}$$

Now we can pass to the limit as $\varepsilon \rightarrow 0$; the only terms left after the cancellation are

$$\lim_{\varepsilon \rightarrow 0} (\varphi (a^\varepsilon - b^\varepsilon) \nabla \phi^\varepsilon, \nabla \psi^\varepsilon)_{D_T} = (\varphi (A - B) \nabla \phi_0, \nabla \psi_0)_{D_T}.$$

Let $\varphi \geq 0$, and we define

$$X := \lim_{\varepsilon \rightarrow 0} (\varphi (b^\varepsilon - a^\varepsilon) \nabla \phi^\varepsilon, \nabla \psi^\varepsilon)_{D_T} = (\varphi (B - A) \nabla \phi_0, \nabla \psi_0)_{D_T}.$$

It is clear that

$$X \leq \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) |(A - a^\varepsilon) \nabla \phi^\varepsilon|, |\nabla \psi^\varepsilon|)_{D_T} \leq 2\Lambda \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) |\nabla \phi^\varepsilon|, |\nabla \psi^\varepsilon|)_{D_T}.$$

For any $\alpha > 0$, we bound X as

$$\begin{aligned} X &\leq \frac{2\Lambda}{\lambda} \left(\alpha \lambda \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho), |\nabla \phi^\varepsilon|^2)_{D_T} + \frac{\lambda}{4\alpha} \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho), |\nabla \psi^\varepsilon|^2)_{D_T} \right) \\ &\leq \frac{2\Lambda}{\lambda} \left(\alpha \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) a^\varepsilon \nabla \phi^\varepsilon, \nabla \phi^\varepsilon)_{D_T} + \frac{1}{4\alpha} \limsup_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) b^\varepsilon \nabla \psi^\varepsilon, \nabla \psi^\varepsilon)_{D_T} \right). \end{aligned} \quad (2.4)$$

After multiplying (2.2) by $(1 - \rho)\varphi\phi^\varepsilon$, we obtain

$$\begin{aligned} \langle \partial_t \phi^\varepsilon, (1 - \rho)\varphi\phi^\varepsilon \rangle_{D_T} + (a^\varepsilon \nabla \phi^\varepsilon, \nabla[(1 - \rho)\varphi\phi^\varepsilon])_{D_T} &= \langle \partial_t \phi_0, (1 - \rho)\varphi\phi^\varepsilon \rangle_{D_T} \\ &\quad + (A \nabla \phi_0, \nabla[(1 - \rho)\varphi\phi^\varepsilon])_{D_T}. \end{aligned}$$

We pass to the limit as $\varepsilon \rightarrow 0$, and obtain

$$\lim_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) a^\varepsilon \nabla \phi^\varepsilon, \nabla \phi^\varepsilon)_{D_T} = (\varphi(1 - \rho) A \nabla \phi_0, \nabla \phi_0)_{D_T}.$$

Similarly, we obtain

$$\lim_{\varepsilon \rightarrow 0} (\varphi(1 - \rho) b^\varepsilon \nabla \psi^\varepsilon, \nabla \psi^\varepsilon)_{D_T} = (\varphi(1 - \rho) B \nabla \psi_0, \nabla \psi_0)_{D_T}.$$

Substituting above two limits into (2.4), we obtain

$$\begin{aligned} X &\leq \frac{2\Lambda}{\lambda} \left(\alpha (\varphi(1 - \rho) A \nabla \phi_0, \nabla \phi_0)_{D_T} + \frac{1}{4\alpha} (\varphi(1 - \rho) B \nabla \psi_0, \nabla \psi_0)_{D_T} \right) \\ &\leq \frac{2\Lambda^2}{\lambda} \left(\alpha (\varphi(1 - \rho), |\nabla \phi_0|^2)_{D_T} + \frac{1}{4\alpha} (\varphi(1 - \rho), |\nabla \psi_0|^2)_{D_T} \right), \end{aligned}$$

which implies that,

$$|(B - A) \nabla \phi_0 \cdot \nabla \psi_0| \leq \frac{2\Lambda^2}{\lambda} (1 - \rho) \left(\alpha |\nabla \phi_0|^2 + \frac{1}{4\alpha} |\nabla \psi_0|^2 \right), \quad \text{a.e. } (x, t) \in D_T,$$

due to φ is arbitrary. Optimizing α , we obtain that

$$|(B - A) \nabla \phi_0 \cdot \nabla \psi_0| \leq \frac{2\Lambda^2}{\lambda} (1 - \rho) |\nabla \phi_0| |\nabla \psi_0|, \quad \text{a.e. } (x, t) \in D_T.$$

Finally, by the definition of the ℓ^2 -norm of a matrix, it gives

$$|A(x, t) - B(x, t)| \leq \frac{2\Lambda^2}{\lambda} (1 - \rho(x, t)) \quad \text{a.e. } (x, t) \in D_T, \quad (2.5)$$

since ϕ_0 and ψ_0 are arbitrary.

Next, we prove

$$|A(x, t) - B(x, t)| \leq 2\Lambda\sqrt{\Lambda/\lambda}\rho(x, t) \quad \text{a.e. } (x, t) \in D_T. \quad (2.6)$$

The proof of (2.6) is similar with the one that leads to (2.5). Motivated by $b^\varepsilon - A = \rho(a^\varepsilon - A)$, we define

$$Y := \lim_{\varepsilon \rightarrow 0} (\varphi (b^\varepsilon - A) \nabla \phi_0, \nabla \psi^\varepsilon)_{D_T} = (\varphi (B - A) \nabla \phi_0, \nabla \psi_0)_{D_T} \quad \text{for } \varphi \geq 0.$$

It is clear that

$$Y \leq \limsup_{\varepsilon \rightarrow 0} (\varphi \rho |A - a^\varepsilon| \nabla \phi_0, |\nabla \psi^\varepsilon|)_{D_T} \leq 2\Lambda \limsup_{\varepsilon \rightarrow 0} (\varphi \rho |\nabla \phi_0|, |\nabla \psi^\varepsilon|)_{D_T}.$$

For any $\alpha > 0$, we bound Y as

$$\begin{aligned} Y &\leq 2\Lambda \left(\alpha (\varphi \rho, |\nabla \phi_0|^2)_{D_T} + \frac{1}{4\alpha} \limsup_{\varepsilon \rightarrow 0} (\varphi \rho, |\nabla \psi^\varepsilon|^2)_{D_T} \right) \\ &\leq 2\Lambda \left(\alpha (\varphi \rho, |\nabla \phi_0|^2)_{D_T} + \frac{1}{4\alpha\lambda} \limsup_{\varepsilon \rightarrow 0} (\varphi \rho b^\varepsilon \nabla \psi^\varepsilon, \nabla \psi^\varepsilon)_{D_T} \right) \\ &\leq 2\Lambda \left(\alpha (\varphi \rho, |\nabla \phi_0|^2)_{D_T} + \frac{1}{4\alpha\lambda} (\varphi \rho B \nabla \psi_0, \nabla \psi_0)_{D_T} \right) \\ &\leq 2\Lambda \left(\alpha (\varphi \rho, |\nabla \phi_0|^2)_{D_T} + \frac{\Lambda}{4\alpha\lambda} (\varphi \rho, |\nabla \psi_0|^2)_{D_T} \right), \end{aligned}$$

which implies that

$$|(B - A) \nabla \phi_0 \cdot \nabla \psi_0| \leq 2\Lambda \rho \left(\alpha |\nabla \phi_0|^2 + \frac{\Lambda}{4\alpha\lambda} |\nabla \psi_0|^2 \right), \quad \text{a.e. } (x, t) \in D_T.$$

Optimizing α , we obtain

$$|(B - A) \nabla \phi_0 \cdot \nabla \psi_0| \leq 2\Lambda\sqrt{\Lambda/\lambda}\rho |\nabla \phi_0| |\nabla \psi_0| \quad \text{a.e. } (x, t) \in D_T.$$

Thus, we obtain (2.6). Finally, we use the convex combination of (2.5) and (2.6) as

$$|A(x, t) - B(x, t)| = \rho(x)|A(x, t) - B(x, t)| + (1 - \rho(x))|A(x, t) - B(x, t)|,$$

this leads to (2.1). □

If we replace A by A_H in the definition of b^ε , we may slightly generalize the above theorem as

COROLLARY 2. Let $A_H \in \mathcal{M}(\lambda', \Lambda'; D_T)$, and we define $b_H^\varepsilon = \rho a^\varepsilon + (1 - \rho)A_H$. Denote by B_H the H-limit of b_H^ε , then for a.e. $(x, t) \in D_T$, there holds

$$|B_H(x, t) - \rho(x, t)A(x, t) - (1 - \rho(x, t))A_H(x, t)| \leq (\Lambda + \tilde{\Lambda})\sqrt{\tilde{\Lambda}/\tilde{\lambda}}(\sqrt{\Lambda/\lambda} + 1)\rho(x, t)(1 - \rho(x, t)),$$

where $\tilde{\Lambda} = \Lambda \vee \Lambda'$ and $\tilde{\lambda} = \lambda \wedge \lambda'$.

REMARK 1. If the transition function $\rho = \chi_K$; i.e., the indicator function of K then the H-limit of the concurrent coefficient b^ε is exactly the effective matrix A .

3. Error estimates

In this part, we derive the error bounds for the proposed method, including the global error between U and u over domain D , and the interior energy error between U and u^ε over the defect domain K_0 .

3.1 Accuracy for retrieving the macroscopic information

We firstly present the well-known stability results (Thomée, 2007) of the scheme (1.3). The proof is included here for the sake of completeness.

LEMMA 3. If $U = \{U_k\}_{k=1}^m$ is the solution of (1.3) then

$$\|U_m\|_{L^2(D)} \leq \|U_0\|_{L^2(D)} + C\|f\|_{\ell^2(0,T;H^{-1}(D))}, \quad (3.1)$$

and

$$\|\delta U\|_{\ell^2(0,T;L^2(D))} + \|\nabla U_m\|_{L^2(D)} \leq C(\|\nabla U_0\|_{L^2(D)} + \|f\|_{\ell^2(0,T;L^2(D))}). \quad (3.2)$$

Proof. Let $V = U_k$ in (1.3), we have

$$(U_k - U_{k-1}, U_k)_D + \Delta t(b_H^\varepsilon \nabla U_k, \nabla U_k)_D = \Delta t(f_k, U_k)_D.$$

By the identity $a(a - b) = \frac{1}{2}(a - b)^2 + \frac{1}{2}(a^2 - b^2)$, we obtain

$$\frac{1}{2}\|U_k - U_{k-1}\|_{L^2(D)}^2 + \frac{1}{2}\|U_k\|_{L^2(D)}^2 - \frac{1}{2}\|U_{k-1}\|_{L^2(D)}^2 + \Delta t(b_H^\varepsilon \nabla U_k, \nabla U_k)_D = \Delta t(f_k, U_k)_D.$$

There exists C depends on D such that

$$\begin{aligned} \frac{1}{2}\|U_k\|_{L^2(D)}^2 + \lambda \Delta t \|\nabla U_k\|_{L^2(D)}^2 &\leq \frac{1}{2}\|U_{k-1}\|_{L^2(D)}^2 + C \Delta t \|f_k\|_{H^{-1}(D)} \|\nabla U_k\|_{L^2(D)} \\ &\leq \frac{1}{2}\|U_{k-1}\|_{L^2(D)}^2 + \frac{\lambda}{2} \Delta t \|\nabla U_k\|_{L^2(D)}^2 + C \Delta t \|f_k\|_{H^{-1}(D)}^2, \end{aligned}$$

i.e.,

$$\|U_k\|_{L^2(D)}^2 + \lambda \Delta t \|\nabla U_k\|_{L^2(D)}^2 \leq \|U_{k-1}\|_{L^2(D)}^2 + C \Delta t \|f_k\|_{H^{-1}(D)}^2.$$

Summing up the above inequality from 1 to m , we obtain (3.1).

Let $V = \delta U_k$ in (1.3), we have

$$\|\delta U_k\|_{L^2(D)}^2 + (b_H^\varepsilon \nabla U_k, \nabla \delta U_k)_D = (f_k, \delta U_k)_D.$$

By the identity $a(a-b) = \frac{1}{2}(a-b)^2 + \frac{1}{2}(a^2 - b^2)$, we obtain

$$\begin{aligned} \Delta t \|\delta U_k\|_{L^2(D)}^2 + \frac{1}{2} (b_H^\varepsilon \nabla U_k, \nabla U_k)_D &\leq \frac{1}{2} (b_H^\varepsilon \nabla U_{k-1}, \nabla U_{k-1})_D + \Delta t \|f_k\|_{L^2(D)} \|\delta U_k\|_{L^2(D)} \\ &\leq \frac{1}{2} \left((b_H^\varepsilon \nabla U_{k-1}, \nabla U_{k-1})_D + \Delta t \|f_k\|_{L^2(D)}^2 + \Delta t \|\delta U_k\|_{L^2(D)}^2 \right). \end{aligned}$$

Henceforth,

$$\Delta t \|\delta U_k\|_{L^2(D)}^2 + (b_H^\varepsilon \nabla U_k, \nabla U_k)_D \leq (b_H^\varepsilon \nabla U_{k-1}, \nabla U_{k-1})_D + \Delta t \|f_k\|_{L^2(D)}^2.$$

Since b_H^ε is time-independent, summing up the above inequality from 1 to m , we obtain (3.2). $\square$

We make the following two assumptions, which are reasonable because the homogenization solution (1.2) is essentially smooth.

Assumption I. The elliptic boundary value problem

$$\begin{cases} -\nabla \cdot (A \nabla \phi_g) = g, & \text{in } D, \\ \phi_g = 0, & \text{on } \partial D, \end{cases}$$

and the solution of the dual problem

$$\begin{cases} -\nabla \cdot (A^\top \nabla \phi_g) = g, & \text{in } D, \\ \phi_g = 0, & \text{on } \partial D, \end{cases} \quad (3.3)$$

satisfies the regularity estimate

$$\|\phi_g\|_{H^2(D)} \leq C \|g\|_{L^2(D)}.$$

Assumption II. $u_0 \in H^2(D) \cap H_0^1(D)$ and $u \in H^1(0, T; H^2(D) \cap H_0^1(D)) \cap H^2(0, T; L^2(D))$.

REMARK 2. Assumption I holds when $A \in C^{0,1}(D)$. For Assumption II, a sufficient condition to ensure $u \in H^1(0, T; H^2(D)) \cap H^2(0, T; L^2(D))$ is $A \in C^{1,1}(D)$, $u_0 \in H^3(D)$ and $f \in L^2(0, T; H^2(D)) \cap$

$H^1(0, T; L^2(D))$. See, e.g., [Evans \(2010, § 6.3, Theorem 4 and § 7.1, Theorem 6\)](#). These conditions are sufficient to ensure assumptions, but they are not necessary for the certain solution u under consideration.

One useful tool is the Galerkin projection operator $R_H : H_0^1(D) \rightarrow X_{h,H}$ defined by

$$(b_H^\varepsilon \nabla R_H v, \nabla V)_D = (A \nabla v, \nabla V)_D \quad \text{for all } V \in X_{h,H}. \quad (3.4)$$

To derive the error bound to the Galerkin projection R_H , we recall the following auxiliary result.

LEMMA 4 ([[Huang et al., 2018](#), Lemma 3.1]). For any $v \in H^1(D)$ with $D \subset \mathbb{R}^n$ a Lipschitz domain, and for any subset $K \subset D$, then

$$\|v\|_{L^2(K)} \leq C \eta(K) \|v\|_{H^1(D)}, \quad (3.5)$$

where

$$\eta(K) = \begin{cases} |K|^{1/2} |\log |K||^{1/2}, & n = 2, \\ |K|^{1/n}, & n \geq 3, \end{cases}$$

with $|K|$ the measure of K , and C is independent of $|K|$.

This estimate is the special case of [Huang et al. \(2018, Lemma 3.1\)](#) with $s = 1$ provided that u satisfies Assumption II. If u enjoys higher regularity, e.g., $u \in H^1(0, T; W^{1,\infty}(D))$ then $\eta(K) = |K|^{1/2}$ in any spatial dimension n . This rate is sharp, as demonstrated by the example in [Huang et al. \(2018, Appendix A\)](#). Further discussion on the factor $\eta(K)$ may be found in [Huang et al. \(2018\)](#) and we do not elaborate on this for brevity.

The error estimate for the above Galerkin projection may be found in

LEMMA 5. Under Assumption I, and if $v \in H_0^1(D) \cap H^2(D)$ then

$$\|\nabla(v - R_H v)\|_{L^2(D)} \leq C(H + \eta(K) + e(\text{HMM})) \|v\|_{H^2(D)}, \quad (3.6)$$

and

$$\|v - R_H v\|_{L^2(D)} \leq C(H^2 + \eta(K)^2 + e(\text{HMM})) \|v\|_{H^2(D)}, \quad (3.7)$$

where

$$e(\text{HMM}) = \max_{x \in D \setminus K} |(A - A_H)(x)|.$$

Particularly, if the initial value U_0 is the solution of (1.4) then

$$\|\nabla(U_0 - R_H u_0)\|_{L^2(D)} \leq C(\eta(K) + e(\text{HMM})) \|u_0\|_{H^2(D)}, \quad (3.8)$$

and

$$\|U_0 - R_H u_0\|_{L^2(D)} \leq C \left(H^2 + \eta(K)^2 + e(\text{HMM}) \right) \|u_0\|_{H^2(D)}. \quad (3.9)$$

Proof. Proceeding along the same line that leads to [Huang et al. \(2018, Theorem 3.1\)](#), we obtain

$$\begin{aligned} \|v - R_H v\|_{L^2(D)} &\leq C \left(\inf_{V \in X_{h,H}} \|\nabla(v - V)\|_{L^2(D)} + \|\nabla v\|_{L^2(K)} \right) \\ &\quad \times \sup_{g \in L^2(D)} \frac{1}{\|g\|_{L^2(D)}} \left(\inf_{V \in X_{h,H}} \|\nabla(\phi_g - V)\|_{L^2(D)} + \|\nabla \phi_g\|_{L^2(K)} \right) \\ &\quad + Ce(\text{HMM}) \|\nabla v\|_{L^2(D)}, \end{aligned} \quad (3.10)$$

and

$$\|\nabla(v - R_H v)\|_{L^2(D)} \leq C \left(\inf_{V \in X_{h,H}} \|\nabla(v - V)\|_{L^2(D)} + \|\nabla v\|_{L^2(K)} + e(\text{HMM}) \|\nabla v\|_{L^2(D)} \right), \quad (3.11)$$

where ϕ_g is the solution of (3.3).

As to the first term,

$$\inf_{V \in X_{h,H}} \|\nabla(v - V)\|_{L^2(D)} \leq \|\nabla(v - \Pi v)\|_{L^2(D)} \leq CH \|v\|_{H^2(D)},$$

where $\Pi : H_0^1(D) \rightarrow X_{h,H}$ is the linear Lagrange interpolation, and also

$$\inf_{V \in X_{h,H}} \|\nabla(\phi_g - V)\|_{L^2(D)} \leq CH \|\phi_g\|_{H^2(D)} \leq CH \|g\|_{L^2(D)},$$

because of Assumption I.

As to the second term, using Lemma 4, we obtain

$$\|\nabla v\|_{L^2(K)} \leq C\eta(K) \|v\|_{H^2(D)},$$

and

$$\|\nabla \phi_g\|_{L^2(K)} \leq C\eta(K) \|\phi_g\|_{H^2(D)} \leq C\eta(K) \|g\|_{L^2(D)}.$$

Substituting all the above inequalities into (3.10) and (3.11), we obtain (3.7) and (3.6).

Next, if U_0 is the solution of (1.4) then

$$(b_H^\varepsilon \nabla(U_0 - R_H u_0), \nabla V)_D = ((b_H^\varepsilon - A) \nabla u_0, \nabla V)_D \quad \text{for all } V \in X_{h,H}.$$

Hence,

$$\begin{aligned} ((b_H^\varepsilon - A)\nabla u_0, \nabla V)_D &\leq C\left(\|\nabla u_0\|_{L^2(K)} + e(\text{HMM})\|\nabla u_0\|_{L^2(D)}\right)\|\nabla V\|_{L^2(D)} \\ &\leq C(\eta(K) + e(\text{HMM}))\|u_0\|_{H^2(D)}\|\nabla V\|_{L^2(D)}. \end{aligned}$$

Choosing $V = U_0 - R_H u_0$, we obtain (3.8).

To obtain the L^2 -estimate, we use the Aubin–Nitsche’s dual argument (Nitsche, 1968). For any $g \in L^2(D)$, let ϕ_g be the solution of (3.3), we have

$$\begin{aligned} (g, U_0 - R_H u_0)_D &= (A\nabla(U_0 - R_H u_0), \nabla\phi_g)_D \\ &= (A\nabla(U_0 - R_H u_0), \nabla(\phi_g - \Pi\phi_g))_D + \left((A - b_H^\varepsilon)\nabla(U_0 - R_H u_0), \nabla\Pi\phi_g\right)_D \\ &\quad + (b_H^\varepsilon\nabla(U_0 - R_H u_0), \nabla\Pi\phi_g)_D. \end{aligned} \quad (3.12)$$

The first term in (3.12) is bounded by

$$\begin{aligned} (A\nabla(U_0 - R_H u_0), \nabla(\phi_g - \Pi\phi_g))_D &\leq CH(\eta(K) + e(\text{HMM}))\|u_0\|_{H^2(D)}\|g\|_{L^2(D)} \\ &\leq C\left(H^2 + \eta(K)^2 + e(\text{HMM})^2\right)\|u_0\|_{H^2(D)}\|g\|_{L^2(D)}. \end{aligned}$$

As to the second term in (3.12), we have

$$\begin{aligned} ((A - b_H^\varepsilon)\nabla(U_0 - R_H u_0), \nabla\Pi\phi_g)_D &\leq C\left(\|\nabla(\phi_g - \Pi\phi_g)\|_{L^2(D)} + \|\nabla\phi_g\|_{L^2(K)} + e(\text{HMM})\|\nabla\phi_g\|_{L^2(D)}\right) \\ &\quad \times \|\nabla(U_0 - R_H u_0)\|_{L^2(D)} \\ &\leq C\left(H^2 + \eta(K)^2 + e(\text{HMM})^2\right)\|u_0\|_{H^2(D)}\|g\|_{L^2(D)}. \end{aligned}$$

Finally, the last term in (3.12) may be bounded as

$$\begin{aligned} (b_H^\varepsilon\nabla(U_0 - R_H u_0), \nabla\Pi\phi_g)_D &= ((b_H^\varepsilon - A)\nabla u_0, \nabla\Pi\phi_g)_D \\ &\leq C\left(\|\nabla u_0\|_{L^2(K)}\|\nabla\Pi\phi_g\|_{L^2(K)} + e(\text{HMM})\|\nabla u_0\|_{L^2(D)}\|\nabla\Pi\phi_g\|_{L^2(D)}\right) \\ &\leq C\left(H^2 + \eta(K)^2 + e(\text{HMM})\right)\|u_0\|_{H^2(D)}\|g\|_{L^2(D)}. \end{aligned}$$

Substituting the above three inequalities into (3.12), we obtain (3.9). $\square$

Based on the above lemma, we derive the error bound for the approximation of the homogenized solution.

THEOREM 6. Under the Assumptions I and II, there holds

$$\begin{aligned} \|u_m - U_m\|_{L^2(D)} &\leq C \left(H^2 + \eta(K)^2 + e(\text{HMM}) \right) (\|u_0\|_{H^2(D)} + \|\partial_t u\|_{L^2(0,T;H^2(D))}) \\ &\quad + C\Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))}, \end{aligned} \quad (3.13)$$

and

$$\begin{aligned} \|\nabla(u_m - U_m)\|_{L^2(D)} &\leq C(H + \eta(K) + e(\text{HMM})) (\|u_0\|_{H^2(D)} + \|\partial_t u\|_{L^2(0,T;H^2(D))}) \\ &\quad + C\Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))}, \end{aligned} \quad (3.14)$$

and

$$\begin{aligned} \|\partial_t u - \delta U\|_{\ell^2(0,T;L^2(D))} &\leq C \left(H^2 + \eta(K) + e(\text{HMM}) \right) (\|u_0\|_{H^2(D)} + \|\partial_t u\|_{L^2(0,T;H^2(D))}) \\ &\quad + C\Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))}. \end{aligned} \quad (3.15)$$

Proof. We write

$$u_m - U_m = u_m - R_H u_m + R_H u_m - U_m.$$

By Lemma 5, we have

$$\|u_m - R_H u_m\|_{L^2(D)} \leq C \left(H^2 + \eta(K)^2 + e(\text{HMM}) \right) \|u_m\|_{H^2(D)}, \quad (3.16)$$

and

$$\|\nabla(u_m - R_H u_m)\|_{L^2(D)} \leq C(H + \eta(K) + e(\text{HMM})) \|u_m\|_{H^2(D)}, \quad (3.17)$$

where

$$\|u_m\|_{H^2(D)} \leq \|u_0\|_{H^2(D)} + \int_0^T \|\partial_t u(\cdot, t)\|_{H^2(D)} dt \leq \|u_0\|_{H^2(D)} + \sqrt{T} \|\partial_t u\|_{L^2(0,T;H^2(D))}.$$

On the one hand, we write

$$\begin{aligned} (\delta(U_k - R_H u_k), V)_D + (b_H^\varepsilon \nabla(U_k - R_H u_k), \nabla V)_D &= (f_k, V)_D - (\delta R_H u_k, V)_D - (A \nabla u, \nabla V)_D \\ &= (\partial_t u_k - \delta R_H u_k, V)_D = (\partial_t u_k - R_H \delta u_k, V)_D, \end{aligned}$$

where R_H and δ are changeable because A and b_h^ε are time-independent. Using the stability estimates (3.1) and (3.2), we obtain

$$\|U_m - R_H u_m\|_{L^2(D)} \leq \|U_0 - R_H u_0\|_{L^2(D)} + C \|\partial_t u - R_H \delta u\|_{\ell^2(0,T;H^{-1}(D))}, \quad (3.18)$$

and

$$\begin{aligned} \|\delta(U - R_H u)\|_{\ell^2(0,T;L^2(D))} + \|\nabla(U_m - R_H u_m)\|_{L^2(D)} &\leq \|\nabla(U_0 - R_H u_0)\|_{L^2(D)} \\ &+ C\|\partial_t u - R_H \delta u\|_{\ell^2(0,T;L^2(D))}. \end{aligned} \quad (3.19)$$

Estimates for the initial value term have been shown in Lemma 5. As to the source term, we write

$$\partial_t u_k - R_H \delta u_k = (\partial_t u_k - \delta u_k) + (\delta u_k - R_H \delta u_k). \quad (3.20)$$

For the first term in (3.20), we have

$$\begin{aligned} (\partial_t u_k - \delta u_k)(x) &= \frac{1}{\Delta t} \int_{t_{k-1}}^{t_k} (\partial_t u(x, t_k) - \partial_t u(x, t)) \, dt = \frac{1}{\Delta t} \int_{t_{k-1}}^{t_k} (t - t_{k-1}) \partial_t^2 u(x, t) \, dt \\ &\leq \sqrt{\Delta t} \|\partial_t^2 u(x, \cdot)\|_{L^2(t_{k-1}, t_k)}. \end{aligned}$$

Therefore,

$$\|\partial_t u - \delta u\|_{\ell^2(0,T;L^2(D))}^2 \leq \Delta t^2 \sum_{k=1}^m \|\partial_t^2 u\|_{L^2(t_{k-1}, t_k; L^2(D))}^2 = \Delta t^2 \|\partial_t^2 u\|_{L^2(0,T;L^2(D))}^2.$$

As to the second term in (3.20), we obtain

$$\begin{aligned} (\delta u_k - R_H \delta u_k)(x) &= \frac{I - R_H}{\Delta t} \int_{t_{k-1}}^{t_k} \partial_t u(x, t) \, dt = \frac{1}{\Delta t} \int_{t_{k-1}}^{t_k} (\partial_t u - R_H \partial_t u)(x, t) \, dt \\ &\leq \frac{1}{\sqrt{\Delta t}} \|(\partial_t u - R_H \partial_t u)(x, \cdot)\|_{L^2(t_{k-1}, t_k)}. \end{aligned}$$

Therefore, we have

$$\|\delta u - R_H \delta u\|_{\ell^2(0,T;L^2(D))}^2 \leq \sum_{k=1}^m \|\partial_t u - R_H \partial_t u\|_{L^2(t_{k-1}, t_k; L^2(D))}^2 = \|\partial_t u - R_H \partial_t u\|_{L^2(0,T;L^2(D))}^2.$$

Combining above two inequalities with the triangle inequality, we get

$$\begin{aligned} \|\partial_t u - R_H \delta u\|_{\ell^2(0,T;H^{-1}(D))} &\leq \|\partial_t u - R_H \delta u\|_{\ell^2(0,T;L^2(D))} \\ &\leq \|\partial_t u - \delta u\|_{\ell^2(0,T;L^2(D))} + \|\delta u - R_H \delta u\|_{\ell^2(0,T;L^2(D))} \\ &\leq \Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))} + \|\partial_t u - R_H \partial_t u\|_{L^2(0,T;L^2(D))} \\ &\leq \Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))} + C \left(H^2 + \eta(K)^2 + e(\text{HMM}) \right) \|\partial_t u\|_{L^2(0,T;H^2(D))}. \end{aligned}$$

Substituting the above inequalities into (3.18) and (3.19), we obtain

$$\begin{aligned} \|U_m - R_H u_m\|_{L^2(D)} &\leq C \left(H^2 + \eta(K)^2 + e(\text{HMM}) \right) (\|u_0\|_{H^2(D)} + \|\partial_t u\|_{L^2(0,T;H^2(D))}) \\ &\quad + \Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))}, \end{aligned} \quad (3.21)$$

and

$$\begin{aligned} \|\delta(U - R_H u)\|_{\ell(0,T;L^2(D))} + \|\nabla(U_m - R_H u_m)\|_{L^2(D)} &\leq \Delta t \|\partial_t^2 u\|_{L^2(0,T;L^2(D))} \\ &\quad + C \left(H^2 + \eta(K) + e(\text{HMM}) \right) (\|u_0\|_{H^2(D)} + \|\partial_t u\|_{L^2(0,T;H^2(D))}). \end{aligned} \quad (3.22)$$

Combining (3.16) and (3.17) with the triangle inequality, we obtain estimates of $u_m - U_m$ in (3.13) and (3.14). Moreover,

$$\|\partial_t u - \delta U\|_{\ell(0,T;L^2(D))} \leq \|\partial_t u - R_H \delta u\|_{\ell(0,T;L^2(D))} + \|\delta(U - R_H u)\|_{\ell(0,T;L^2(D))},$$

which gives (3.15). $\square$

REMARK 3. In Ming & Zhang (2007, Theorem 1.1), for the general numerical homogenization solution U calculated by HMM, the term $\|\nabla(u_m - U_m)\|_{L^2(D)}$ is bounded by $\mathcal{O}(\Delta t + H + e(\text{HMM})\Delta t^{-1/2})$. The above estimate eliminates the factor $\Delta t^{-1/2}$ when the diffusion coefficient is time-independent, which indicates that Δt can be arbitrarily small.

3.2 Accuracy for retrieving the local microscopic information

In order to prove the localized energy error estimate, we introduce some notation and results in Demlow *et al.* (2011). For a subdomain $B \subset D$, define $H_{<}^1(B) := \{u \in H^1(D) | u|_{D \setminus B} = 0\}$. Let G_1 and G be subsets of K with $G_1 \subset G$ and $\text{dist}(G_1, \partial G \setminus \partial D) = d > 0$. The following assumptions are assumed to hold:

A1: Local interpolant. There exists a local interpolant such that for any $u \in H_{<}^1(G_1) \cap C(G_1)$, $Iu \in X_{h,H} \cap H_{<}^1(G)$.

A2: Inverse properties. For each $\chi \in X_{h,H}$ and $\tau \in \mathcal{T}_{h,H} \cap K$, $1 \leq p \leq q \leq \infty$, and $0 \leq \nu \leq s \leq r$,

$$\|\chi\|_{W^{s,q}(\tau)} \leq Ch_{\tau}^{\nu-s+n/p-n/q} \|\chi\|_{W^{\nu,p}(\tau)}.$$

A3. Superapproximation. Let $\omega \in C^\infty(K) \cap H_{<}^1(G_1)$ with $|\omega|_{W^{j,\infty}(K)} \leq Cd^{-j}$ for integers $0 \leq j \leq r$, for each $\chi \in X_{h,H} \cap H_{<}^1(G)$ and for each $\tau \in \mathcal{T}_{h,H} \cap K$ satisfying $h_{\tau} \leq d$,

$$\|\omega^2 \chi - I(\omega^2 \chi)\|_{H^1(\tau)} \leq C \left(\frac{h_{\tau}}{d} \|\nabla(\omega \chi)\|_{L^2(\tau)} + \frac{h_{\tau}}{d^2} \|\chi\|_{L^2(\tau)} \right), \quad (3.23)$$

where the interpolant I is defined in A1. Assumptions A1–A3 are satisfied by standard Lagrange finite element defined on shape-regular grids (Demlow *et al.*, 2011).

Set $K_0 \subset K_1 \subset K$, with $\text{dist}(K_1, \partial K \setminus \partial D) = \text{dist}(K_0, \partial K_1 \setminus \partial D) = cd$ for a universal constant c . The following lemma is a generalization of Demlow *et al.* (2011, Lemma 3.3) by including the effect of the source term. Similar results have been proven in Nitsche & Schatz (1974); Xu & Zhou (2001) without clarifying the dependence on the distance d .

LEMMA 7. For $W \in X_{h,H}$ satisfies

$$(b_H^\varepsilon \nabla W, \nabla V)_D = (f, V)_D \quad \text{for all } V \in X_{h,H} \cap H_{<}^1(K_1),$$

then

$$\|\nabla W\|_{L^2(K_0)} \leq C \left(d^{-1} \|W\|_{L^2(K_1)} + \|f\|_{H^{-1}(K_1)} \right). \quad (3.24)$$

Proof. We set $\omega \in C^\infty(K_1) \cap H_{<}^1(K_1)$ a cut-off function with $\omega \equiv 1$ on K_0 and $|\nabla^j \omega| \leq Cd^{-j}$, $j = 0, 1$. A direct manipulation gives

$$\begin{aligned} (b_H^\varepsilon \nabla(\omega W), \nabla(\omega W))_D &= (b_H^\varepsilon \nabla W, \nabla(\omega^2 W))_D + (b_H^\varepsilon W \nabla \omega, W \nabla \omega)_D \\ &\quad + (b_H^\varepsilon W \nabla \omega, \nabla(\omega W))_D - (b_H^\varepsilon \nabla(\omega W), W \nabla \omega)_D \\ &\leq (b_H^\varepsilon \nabla W, \nabla(\omega^2 W))_D + \frac{\lambda}{4} \|\nabla(\omega W)\|_{L^2(K_1)}^2 + \frac{C}{d^2} \|W\|_{L^2(K_1)}^2. \end{aligned}$$

The first term may be reshape into

$$(b_H^\varepsilon \nabla W, \nabla(\omega^2 W))_D = (b_H^\varepsilon \nabla W, \nabla(\omega^2 W - I(\omega^2 W)))_D + (f, I(\omega^2 W))_D.$$

Using the superapproximation (3.23) and the inverse estimate, we get

$$\begin{aligned} (b_H^\varepsilon \nabla W, \nabla(\omega^2 W))_D &\leq C \|\nabla W\|_{L^2(K_1)} \|\nabla(\omega^2 W) - \nabla I(\omega^2 W)\|_{L^2(K_1)} + \|f\|_{H^{-1}(K_1)} \|I(\omega^2 W)\|_{H^1(K_1)} \\ &\leq C \left(d^{-1} \|W\|_{L^2(K_1)} + \|f\|_{H^{-1}(K_1)} \right) \left(\|\nabla(\omega W)\|_{L^2(K_1)} + d^{-1} \|W\|_{L^2(K_1)} \right), \end{aligned}$$

where we have assumed $h \leq d$.

Combining the above inequalities, there holds

$$\begin{aligned} \lambda \|\nabla(\omega W)\|_{L^2(K_1)}^2 &\leq (b_H^\varepsilon \nabla(\omega W), \nabla(\omega W))_D \\ &\leq \frac{\lambda}{2} \|\nabla(\omega W)\|_{L^2(K_1)}^2 + C \left(d^{-2} \|W\|_{L^2(K_1)}^2 + \|f\|_{H^{-1}(K_1)}^2 \right). \end{aligned}$$

This gives (3.24). $\square$

We are ready to give the localized energy error estimate. First, we define a local Galerkin projection operator $R_h : H_{<}^1(K) \rightarrow X_{h,H} \cap H_{<}^1(K)$ such that

$$(b_H^\varepsilon \nabla R_h v, \nabla V)_D = (a^\varepsilon \nabla v, \nabla V)_D \quad \text{for all } V \in X_{h,H} \cap H_{<}^1(K).$$

The interior estimate relies on the Meyers regularity for linear parabolic equations, which is given in Bensoussan *et al.* (2011, Chapter 2, Theorem 2.5).

LEMMA 8. Let D be a bounded domain, $f \in L^2(0, T; H^{-1}(D))$, $u_0 \in H_0^1(D)$ and $u \in L^2(0, T; H_0^1(D))$ be the solution of

$$\begin{cases} \partial_t u - \nabla \cdot (A \nabla u) = f, & \text{in } D_T, \\ u|_{t=0} = u_0, & \text{in } D, \end{cases}$$

where $A \in [L^\infty(D_T)]^{d \times d}$ is uniformly elliptic with corresponding parameters λ and Λ . There exists $p > 2$, depending on λ , Λ and D , such that if $u_0 \in W_0^{1,p}(D)$ and $f \in L^p(0, T; W^{-1,p}(D))$ then $u \in L^p(0, T; W_0^{1,p}(D))$. Furthermore,

$$\|u\|_{L^p(0,T;W^{1,p}(D))} \leq C \left(\|f\|_{L^p(0,T;W^{-1,p}(D))} + T^{1/p} \|u_0\|_{W^{1,p}(D)} \right),$$

where the constant $C > 0$ depends on λ , Λ and D .

THEOREM 9. If $u^\varepsilon \in \ell^2(0, T; W^{1,p}(K))$ for certain $p > 2$ then

$$\begin{aligned} \|\nabla(u^\varepsilon - U)\|_{\ell^2(0,T;L^2(K_0))} &\leq C \left(\inf_{V \in [X_{h,H} \cap H_{<}^1(K)]^m} \|\nabla(\omega u^\varepsilon - V)\|_{\ell^2(0,T;L^2(K_0))} \right. \\ &\quad + |K \setminus K_0|^{1/2-1/p} \left(d^{-1} \|u^\varepsilon\|_{\ell^2(0,T;L^p(K))} + \|\nabla u^\varepsilon\|_{\ell^2(0,T;L^p(K))} \right) \\ &\quad \left. + d^{-1} \|u^\varepsilon - U\|_{\ell^2(0,T;L^2(K))} + \|\partial_t u^\varepsilon - \delta U\|_{\ell^2(0,T;H^{-1}(K))} \right), \end{aligned} \quad (3.25)$$

where $\omega \in C^\infty(K) \cap H_{<}^1(K)$ with $\omega = 1$ on K_1 and $|\nabla^j \omega| \leq C d^{-j}$, $j = 0, 1$.

Proof. For arbitrary $V \in X_{h,H} \cap H_{<}^1(K)$,

$$\begin{aligned} (b_H^\varepsilon \nabla (R_h v - V), \nabla (R_h v - V))_D &= (a^\varepsilon \nabla v - b_H^\varepsilon \nabla V, \nabla (R_h v - V))_D \\ &= ((a^\varepsilon - b_H^\varepsilon) \nabla v, \nabla (R_h v - V))_D + (b_H^\varepsilon \nabla (v - V), \nabla (R_h v - V))_D. \end{aligned}$$

This gives

$$\|\nabla (R_h v - V)\|_{L^2(K)} \leq C \left(\|\nabla v\|_{L^2(K \setminus K_0)} + \|\nabla (v - V)\|_{L^2(K)} \right).$$

By the triangle inequality and the arbitrariness of V , it holds that

$$\|\nabla (v - R_h v)\|_{L^2(K)} \leq C \left(\inf_{V \in X_{h,H} \cap H_{<}^1(K)} \|\nabla (v - V)\|_{L^2(K)} + \|\nabla v\|_{L^2(K \setminus K_0)} \right).$$

Set $v = \omega u_k^\varepsilon$, we obtain

$$\begin{aligned} \|\nabla(I - R_h)(\omega u_k^\varepsilon)\|_{L^2(K)} &\leq C \left(\inf_{V \in X_{h,H} \cap H_{<}^1(K)} \|\nabla(\omega u_k^\varepsilon - V)\|_{L^2(K)} + \|\nabla(\omega u_k^\varepsilon)\|_{L^2(K \setminus K_0)} \right) \\ &\leq C \left(\inf_{V \in X_{h,H} \cap H_{<}^1(K)} \|\nabla(\omega u_k^\varepsilon - V)\|_{L^2(K)} \right. \\ &\quad \left. + |K \setminus K_0|^{1/2-1/p} (d^{-1} \|u_k^\varepsilon\|_{L^p(K)} + \|\nabla u_k^\varepsilon\|_{L^p(K)}) \right). \end{aligned}$$

Note that for all $V \in X_{h,H} \cap H_{<}^1(K_1)$, we have

$$(b_H^\varepsilon \nabla (U_k - R_h(\omega u_k^\varepsilon)), \nabla V)_D = (f_k - \delta U_k, V)_D - (a^\varepsilon \nabla u_k^\varepsilon, \nabla V)_D = (\partial_t u_k^\varepsilon - \delta U_k, V)_D.$$

By Lemma 7, we obtain

$$\begin{aligned} \|\nabla(U_k - R_h(\omega u_k^\varepsilon))\|_{L^2(K_0)} &\leq C \left(d^{-1} \|U_k - R_h(\omega u_k^\varepsilon)\|_{L^2(K_1)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K_1)} \right) \\ &\leq C \left(d^{-1} \|\omega u_k^\varepsilon - R_h(\omega u_k^\varepsilon)\|_{L^2(K)} + d^{-1} \|u_k^\varepsilon - U_k\|_{L^2(K_1)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K_1)} \right) \\ &\leq C \left(\|\nabla(I - R_h)(\omega u_k^\varepsilon)\|_{L^2(K)} + d^{-1} \|u_k^\varepsilon - U_k\|_{L^2(K)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K)} \right), \end{aligned}$$

where we have used the Poincaré inequality in the last line. By the triangle inequality,

$$\begin{aligned} \|\nabla(u_k^\varepsilon - U_k)\|_{L^2(K_0)} &\leq \|\nabla(I - R_h)(\omega u_k^\varepsilon)\|_{L^2(K)} + \|\nabla(U_k - R_h(\omega u_k^\varepsilon))\|_{L^2(K_0)} \\ &\leq C \left(\inf_{V \in X_{h,H} \cap H_{<}^1(K)} \|\nabla(\omega u_k^\varepsilon - V)\|_{L^2(K_0)} \right. \\ &\quad \left. + |K \setminus K_0|^{1/2-1/p} (d^{-1} \|u_k^\varepsilon\|_{L^p(K)} + \|\nabla u_k^\varepsilon\|_{L^p(K)}) \right. \\ &\quad \left. + d^{-1} \|u_k^\varepsilon - U_k\|_{L^2(K)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K)} \right). \end{aligned}$$

Sum up k from 1 to m , we obtain (3.25). $\square$

Under the same condition of Theorem 9, we may bound the last two terms in the right-hand side of (3.25).

REMARK 4. Using Hölder inequality and the estimate (3.13), we obtain

$$\begin{aligned}\|u^\varepsilon - U\|_{\ell^2(0,T;L^2(K))} &\leq \|u^\varepsilon - u\|_{\ell^2(0,T;L^2(K))} + \|u - U\|_{\ell^2(0,T;L^2(D))} \\ &\leq C|K|^{1/2-1/p} \left(\|u^\varepsilon\|_{\ell^2(0,T;L^p(K))} + \|u\|_{\ell^2(0,T;L^p(K))} \right) \\ &\quad + C(u) \left(\Delta t + H^2 + \eta(K)^2 + e(\text{HMM}) \right).\end{aligned}$$

Note that for any $0 < t \leq T$ and for $p > 2$, denoting by p' the conjugate index of p , we get

$$\begin{aligned}\|\partial_t(u^\varepsilon - u)(\cdot, t)\|_{H^{-1}(K)} &= \sup_{\varphi \in H_0^1(K)} \frac{|(\partial_t(u^\varepsilon - u)(\cdot, t), \varphi)|}{\|\nabla \varphi\|_{L^2(K)}} = \sup_{\varphi \in H_0^1(K)} \frac{|(a^\varepsilon \nabla u^\varepsilon - A \nabla u)(\cdot, t), \nabla \varphi|}{\|\nabla \varphi\|_{L^2(K)}} \\ &\leq \Lambda \sup_{\varphi \in H_0^1(K)} \frac{(\|\nabla u^\varepsilon(\cdot, t)\|_{L^p(K)} + \|\nabla u(\cdot, t)\|_{L^p(K)}) \|\nabla \varphi\|_{L^{p'}(K)}}{\|\nabla \varphi\|_{L^2(K)}} \\ &\leq \Lambda |K|^{1/p'-1/2} (\|\nabla u^\varepsilon(\cdot, t)\|_{L^p(K)} + \|\nabla u(\cdot, t)\|_{L^p(K)}) \\ &= \Lambda |K|^{1/2-1/p} (\|\nabla u^\varepsilon(\cdot, t)\|_{L^p(K)} + \|\nabla u(\cdot, t)\|_{L^p(K)}).\end{aligned}$$

Using (3.15) and note that $u^\varepsilon \in \ell^2(0, T; W_0^{1,p}(K))$, we get

$$\begin{aligned}\|\partial_t u^\varepsilon - \delta U\|_{\ell^2(0,T;H^{-1}(K))} &\leq \|\partial_t u^\varepsilon - \partial_t u\|_{\ell^2(0,T;H^{-1}(K))} + \|\partial_t u - \delta U\|_{\ell^2(0,T;H^{-1}(K))} \\ &\leq C(u^\varepsilon, u) |K|^{1/2-1/p} + C(u) \left(\Delta t + H^2 + \eta(K) + e(\text{HMM}) \right).\end{aligned}$$

Theorem 9 relies on Meyers' regularity estimate, which introduce an additional factor $|K \setminus K_0|^{1/2-1/p}$. This factor cannot be ignored, as u^ε may exhibit severe oscillation. The next theorem shows that choosing $\rho = \chi_K$ removes this factor, thereby eliminating the pollution originating from outside K_0 . This observation clarifies the poor performance of $\rho = \chi_{K_0}$ observed in the elliptic-case experiments of Huang *et al.* (2018). In practice, for continuous transition functions, the pollution outside K_0 remains acceptable.

THEOREM 10. If $\rho = \chi_K$ then

$$\begin{aligned}\|\nabla(u^\varepsilon - U)\|_{\ell^2(0,T;L^2(K_0))} &\leq C \left(\inf_{V \in [X_{h,H}]^m} \|\nabla(u^\varepsilon - V)\|_{\ell^2(0,T;L^2(K))} \right. \\ &\quad \left. + d^{-1} \|u^\varepsilon - U\|_{\ell^2(0,T;L^2(K))} + \|\partial_t u^\varepsilon - \delta U\|_{\ell^2(0,T;H^{-1}(K))} \right). \quad (3.26)\end{aligned}$$

Proof. Set $\chi_k = \arg \min_{V \in X_{h,H}} \|\nabla(u_k^\varepsilon - V)\|_{L^2(K)}$ with $\int_K \chi_k \, dx = \int_K u_k^\varepsilon \, dx$. Define a cut-off function $\omega \in C^\infty(K) \cap H^1_{<}(K)$ with $\omega = 1$ on K_1 and $|\nabla^j \omega| \leq C d^{-j}$, $j = 0, 1$.

$$\begin{aligned} \|\nabla(u_k^\varepsilon - U_k)\|_{L^2(K_0)} &\leq \|\nabla(\omega u_k^\varepsilon - \omega \chi_k) - \nabla R_h(\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K)} \\ &\quad + \|\nabla(U_k - \chi_k) - \nabla R_h(\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K_0)}. \end{aligned}$$

By $\int_K (u_k^\varepsilon - \chi_k) \, dx = 0$ and Poincaré inequality on K , we have

$$\begin{aligned} \|\nabla(\omega u_k^\varepsilon - \omega \chi_k) - \nabla R_h(\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K)} &\leq C \|\nabla(\omega u_k - \omega \chi_k)\|_{L^2(K)} \\ &\leq C \left(\|\nabla(u_k^\varepsilon - \chi_k)\|_{L^2(K)} + d^{-1} \|u_k^\varepsilon - \chi_k\|_{L^2(K)} \right) \\ &\leq C \|\nabla(u_k^\varepsilon - \chi_k)\|_{L^2(K)} \\ &\leq C \inf_{V \in X_{h,H}} \|\nabla(u_k^\varepsilon - V)\|_{L^2(K)}, \end{aligned} \tag{3.27}$$

where we use the relation

$$\|\nabla R_h v\|_{L^2(K)} \leq \frac{\Lambda}{\lambda} \|\nabla v\|_{L^2(K)} \quad \text{for all } v \in H^1_{<}(K)$$

in the first step. Noting that $\omega = 1$ and $b_H^\varepsilon = a^\varepsilon$ on K_1 , for all $V \in X_{h,H} \cap H^1_{<}(K_1)$, we obtain

$$\begin{aligned} &(b_H^\varepsilon \nabla (U_k - \chi_k - R_h(\omega u_k^\varepsilon - \omega \chi_k)), \nabla V)_D \\ &= (f_k - \delta U_k, V)_D - (b_H^\varepsilon \nabla \chi_k, \nabla V)_D - (a^\varepsilon \nabla (\omega u_k^\varepsilon - \omega \chi_k), \nabla V)_D \\ &= (f_k - \delta U_k, V)_D - (a^\varepsilon \nabla u_k^\varepsilon, \nabla V)_D = (\partial_t u_k^\varepsilon - \delta U_k, V)_D. \end{aligned}$$

By Lemma 7 and invoking the Poincaré inequality again, we obtain

$$\begin{aligned} &\|\nabla (U_k - \chi_k) - \nabla R_h (\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K_0)} \\ &\leq C \left(d^{-1} \|U_k - \chi_k - R_h (\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K_1)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K_1)} \right) \\ &\leq C \left(d^{-1} \|\omega u_k^\varepsilon - \omega \chi_k - R_h (\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K)} + d^{-1} \|u_k^\varepsilon - U_k\|_{L^2(K_1)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K_1)} \right) \\ &\leq C \left(\|\nabla(\omega u_k^\varepsilon - \omega \chi_k) - \nabla R_h (\omega u_k^\varepsilon - \omega \chi_k)\|_{L^2(K)} + d^{-1} \|u_k^\varepsilon - U_k\|_{L^2(K_1)} + \|\partial_t u_k^\varepsilon - \delta U_k\|_{H^{-1}(K_1)} \right). \end{aligned}$$

Substituting (3.27) into above inequality again and summing up from $k = 1$ to m , we obtain (3.26). $\square$

So far, we have not discussed the estimate of $\ell(\text{HMM})$, the so-called resonance error, which has been studied extensively in E *et al.* (2005) and Ming & Zhang (2007). We also refer to Gloria & Habibi (2016) for certain more advanced techniques to reduce the resonance error.

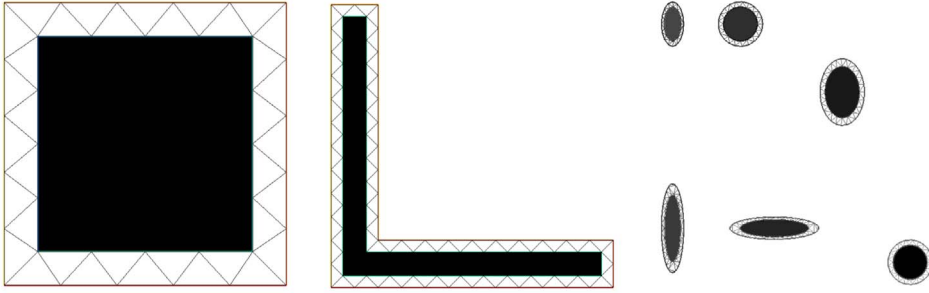


FIG. 1. Three defects. Left: well defect. Mid: L-shaped defect. Right: Porous defect.

4. Numerical examples

In this part, we present several numerical examples to validate the accuracy of the proposed method. In all cases, the domain is set as $D = [0, 1]^2$ with the source term $f = 1$ and the initial value $u_0 = x(1-x)y(1-y)$. We consider three representative defects: a well defect, an L-shaped defect, and a porous defect. These geometries are selected to model distinct physical scenarios: for instance, L-shaped defects often arise in problems involving channelized flow, while porous defects are typical of materials with distributed inclusions (Oden & Vemaganti, 2000). We note that, as suggested by the theoretical analysis, the method's performance may deteriorate when the defect region is excessively large.

The region K is constructed by generating a single-layer mesh surrounding K_0 . If K_0 consists of several disjoint subdomains, each subdomain is treated separately. Consequently, our method is equally applicable to cases with multiple defects. Figure 1 illustrates K_0 and K for three types of defects.

The transition function ρ on $K \setminus K_0$ is constructed via linear Lagrange interpolation on the single-layer mesh described above, ensuring continuity and straightforward implementation. This approach, adopted in Ming & Song (2022), has proven effective in practice. Although Theorem 10 suggests that the indicator function χ_K leads to better performance for local H^1 -estimates, our experiments show that the interpolated transition function performs comparably in local H^1 -estimates and yields improved results for local L^2 -estimates.

Relative errors are reported at $T = 1$. We consider the global quantities

$$e_0(D \setminus K) = \frac{\|U_m - u_m\|_{L^2(D \setminus K)}}{\|u_m\|_{L^2(D \setminus K)}}, \quad e_1(D \setminus K) = \frac{\|\nabla(U_m - u_m)\|_{L^2(D \setminus K)}}{\|\nabla u_m\|_{L^2(D \setminus K)}},$$

and the local quantities

$$e_0(K_0) = \frac{\|U_m - u_m^\varepsilon\|_{L^2(K_0)}}{\|u_m^\varepsilon\|_{L^2(K_0)}}, \quad e_1(K_0) = \frac{\|\nabla(U_m - u_m^\varepsilon)\|_{L^2(K_0)}}{\|\nabla u_m^\varepsilon\|_{L^2(K_0)}}$$

on defect K_0 , where $m = T/\Delta t$. We fix the time step $\Delta t = 0.02$, and vary h and H . The reference solution u^0 and u^ε are computed with $h = 1/3000$ and $\Delta t = 0.01$. All the numerical experiments are conducted on FreeFEM++ Hecht (2012).

TABLE 1 *Macroscopic and microscopic errors for two scales coefficient with well defect*

$1/H$	$e_0(D \setminus K)$	order	$e_1(D \setminus K)$	order	$1/h$	$e_0(K_0)$	order	$e_1(K_0)$	order
Macro errors with $h = 10^{-3}$					Micro errors with $H = 10^{-2}$				
20	5.54E-03		6.70E-02		200	3.72E-03		2.88E-01	
40	1.39E-03	1.99	3.30E-02	1.02	400	1.07E-03	1.80	1.44E-01	1.00
80	4.38E-04	1.67	1.64E-02	1.01	800	3.42E-04	1.63	6.26E-02	1.20
160	2.48E-04	0.82	8.20E-03	1.00	1600	2.36E-04	0.53	2.65E-02	1.24

4.1 Example 1: coefficient with two scales

The first two-scale diffusion coefficient is taken from [Huang et al. \(2018\)](#) with

$$a^\varepsilon(x) = \frac{(R_1 + R_2 \sin(2\pi x_1))(R_1 + R_2 \cos(2\pi x_2))}{(R_1 + R_2 \sin(2\pi x_1/\varepsilon))(R_1 + R_2 \sin(2\pi x_2/\varepsilon))} I. \quad (4.1)$$

A direct calculation gives the effective coefficient

$$A(x) = \frac{(R_1 + R_2 \sin(2\pi x_1))(R_1 + R_2 \cos(2\pi x_2))}{R_1 \sqrt{R_1^2 - R_2^2}} I.$$

Here, we select $\varepsilon = 0.01$, $R_1 = 2.5$ and $R_2 = 1.5$. We consider the well defect K_0 a 0.1×0.1 square, and K a 0.12×0.12 square in the centre of D ; i.e., $K_0 = [0.45, 0.55]^2$ and $K = [0.44, 0.56]^2$. Therefore, in this case, $|K_0| = 0.01$ and $|K| = 0.0144$.

To investigate the macroscopic error, we fix $\Delta t = 0.02$ and $h = 1/1000$ while varying H . The results, shown in the left columns of [Table 1](#), correspond to the case $A_H = A$, so that $e(\text{HMM}) = 0$. We observe that $e_1(D \setminus K)$ converges at first order in H . However, the convergence rate of $e_0(D \setminus K)$ drops sharply from approximately 2 to about 1, likely due to the combined influence of the defect region K and the time discretization step Δt in both the reference and numerical solutions. Overall, the experimental findings align well with the theoretical predictions.

As to the microscopic error, we vary h from $1/200$ to $1/1600$ fixing $\Delta t = 0.02$ and $H = 1/100$. As in the macroscopic case, the homogenization approximation error $e(\text{HMM}) = 0$. Moreover, since $H^2 = 10^{-4} \ll h$, the local mesh size h is the dominant error source for $e_1(K_0)$; see [Remark 4](#). The results are presented in the right columns of [Table 1](#), showing that $e_1(K_0)$ exhibits first-order convergence with respect to h , while $e_0(K_0)$ tends to a fixed value as h decreases.

4.2 Example 2: different transition functions.

In this example, we investigate the influence of different choices of the transition functions. Since numerical results for C^1 - and C^0 -transition functions have been reported in [Ming & Song \(2022\)](#), we focus here on the discontinuous selections, particularly on the indicator functions. We retain the two-scale diffusion coefficient defined in (4.1), with the well defect given by $K_0 = [0.45, 0.55]^2$ and $K = [0.44, 0.56]^2$.

TABLE 2 *Macroscopic and microscopic errors for different transition functions*

$1/H$	$e_0(D \setminus K)$	order	$e_1(D \setminus K)$	order	$1/h$	$e_0(K_0)$	order	$e_1(K_0)$	order
Macro errors with $\rho = \chi_{K_0}, h = 10^{-3}$					Micro errors with $\rho = \chi_{K_0}, H = 10^{-2}$				
20	5.51E-03		6.70E-02		200	2.92E-03		2.87E-01	
40	1.36E-03	2.02	3.29E-02	1.03	400	9.32E-04	1.65	1.54E-01	0.90
80	3.96E-04	1.78	1.64E-02	1.00	800	9.55E-04	-0.04	8.45E-02	0.87
160	2.01E-04	0.98	8.17E-03	1.01	1600	1.06E-03	-0.15	6.21E-02	0.44
Macro errors with $\rho = \chi_K, h = 10^{-3}$					Micro errors with $\rho = \chi_K, H = 10^{-2}$				
20	5.52E-03		6.72E-02		200	3.54E-03		2.91E-01	
40	1.38E-03	2.00	3.33E-02	1.01	400	8.58E-04	2.04	1.48E-01	0.98
80	4.39E-04	1.65	1.70E-02	0.97	800	8.52E-04	0.01	6.61E-02	1.16
160	2.55E-04	0.78	8.41E-03	1.01	1600	1.02E-03	-0.26	2.50E-02	1.40

We fix $\Delta t = 0.02$ and $h = 1/1000$ and summarize the macroscopic errors in the left columns of Table 2. While the results for different transition functions are comparable, $\rho = \chi_{K_0}$ performs slightly better. This observation aligns with Theorem 6. Specifically, the factor $\eta(K)$ should more precisely be expressed as $\eta(\text{supp } \rho)$. Consequently, when $\rho = \chi_{K_0}$, the area of its support equals to $|K_0| = 0.01$, which is slightly smaller than $|K| = 0.0144$, resulting in a smaller contribution from this term.

The behaviour of the microscopic errors is more nuanced, as shown in the right columns of Table 2. We fix $\Delta t = 0.02$ and $H = 1/100$. The indicator function χ_{K_0} yields the poorest performance, consistent with the findings reported in Huang *et al.* (2018). This can be explained by Theorem 9, which introduces a factor of $|K \setminus K_0|^{1/2-1/p}$ in this case. In contrast, χ_K performs best for $e_1(K_0)$, in line with Theorem 10. The linear Lagrange interpolation yields comparable results in practice, as shown in Table 1. This is reasonable because the principle underlying Theorem 10 is that the coefficient near, but outside K_0 , should remain close to a^ε , a condition that still holds for continuous ρ , albeit with mild pollution.

The performance of $e_0(K)$ is poor for the discontinuous transition functions, including both χ_{K_0} and χ_K . Although a specific theoretical analysis for this case is beyond the scope of the present work, this aligns with the general expectation that duality-based L^2 -estimates typically require the continuity of the coefficient. In the view of these observations, we henceforth adopt the linear Lagrange interpolation as the transition function in the subsequent experiments.

4.3 Example 3: different sizes of defects

In this example, we continue to use the two-scale coefficient given in (4.1) and examine the effect of varying the size of the well defect. For macroscopic errors, we fix $\Delta t = 0.02$ and $h = 1/1000$. For microscopic errors, we set $\Delta t = 0.02$ and $H = 1/100$. First, we consider a smaller defect $K_0 = [0.48, 0.52]^2$ and $K = [0.47, 0.53]^2$ with corresponding areas $|K_0| = 0.0016$ and $|K| = 0.0036$, substantially smaller than those in Table 1. We report the macroscopic and microscopic errors in the top rows of Table 3. The results show a noticeable improvement in the L^2 -estimates, confirming that the previously observed reduction in convergence rate was indeed caused by the larger defect K . Nevertheless, the errors eventually saturate around 10^{-4} , which is mainly attributable to the fixed time step Δt shared by both the reference and the numerical solutions.

Next, we consider a larger defect with $K_0 = [0.4, 0.6]^2$ and $K = [0.39, 0.61]^2$, the corresponding areas are $|K_0| = 0.04$ and $|K| = 0.0484$, much larger than those in Table 1. Macroscopic and microscopic

TABLE 3 *Macroscopic and microscopic errors for different sizes of defects*

$1/H$	$e_0(D \setminus K)$	order	$e_1(D \setminus K)$	order	$1/h$	$e_0(K_0)$	order	$e_1(K_0)$	order
Macro error with $K = [0.47, 0.53]^2, h = 10^{-3}$					Micro errors with $K_0 = [0.48, 0.52]^2, H = 10^{-2}$				
20	5.61E-03		6.81E-02		200	1.01E-03		2.54E-01	
40	1.39E-03	2.01	3.32E-02	1.04	400	3.79E-04	1.41	1.62E-01	0.65
80	3.78E-04	1.88	1.61E-02	1.04	800	1.42E-04	1.42	8.60E-02	0.91
160	1.42E-04	1.41	7.57E-03	1.09	1600	1.46E-04	-0.04	4.18E-02	1.04
Macro error with $K = [0.39, 0.61]^2, h = 10^{-3}$					Micro errors with $K_0 = [0.4, 0.6]^2, H = 10^{-2}$				
20	5.53E-03		6.38E-02		200	9.59E-03		2.85E-01	
40	1.53E-03	1.85	3.12E-02	1.03	400	3.06E-03	1.65	1.36E-01	1.07
80	6.29E-04	1.28	1.52E-02	1.04	800	9.65E-04	1.67	4.94E-02	1.46
160	4.67E-04	0.43	7.14E-03	1.09	1600	5.87E-04	0.72	2.51E-02	0.98

errors are shown in the bottom rows of Table 3. First-order convergence is maintained for both $e_1(D \setminus K)$ and $e_1(K_0)$. Although $e_0(K_0)$ initially appears to converge more rapidly, this impression stems mainly from its lower accuracy on coarse grids. Compared with the smaller well defects, the L^2 -error estimates deteriorate, highlighting the stronger influence of a larger defect region.

4.4 Example 4: different shapes of defects

In this example, we continue to employ the two-scale coefficient given in (4.1) and consider different defect geometries, as shown in Fig. 1. For the L-shaped defect K_0 , the vertices (listed counterclockwise) are

$$(0.4, 0.4), (0.76, 0.4), (0.76, 0.41), (0.41, 0.41), (0.41, 0.76) \text{ and } (0.4, 0.76).$$

The longest edge of K_0 measures 0.36 and the shortest 0.01, resulting in an aspect ratio of 36. This configuration can be used to model a thin channel.

The annular layer between K and K_0 has thickness 0.005, so the vertices of K are

$$(0.395, 0.395), (0.765, 0.395), (0.765, 0.415), (0.415, 0.415), (0.415, 0.765) \text{ and } (0.395, 0.765).$$

For K , the longest edge is 0.37 and the shortest 0.02.

The macroscopic and microscopic errors are reported in the top rows of Table 4. Following the setup of Example 1, we fix $\Delta t = 0.02$ and $h = 1/1000$ for the macroscopic errors, and $\Delta t = 0.02$ and $H = 1/100$ for the microscopic errors. The areas $|K_0| = 0.0071$ and $|K| = 0.0144$ are closed to those of the well defect in Example 1, thereby reducing size effects. The defect shape appears to primarily affect the L^2 -estimates, specifically $e_0(D \setminus K)$ and $e_0(K_0)$, while first-order convergence is still maintained for both $e_1(D \setminus K)$ and $e_1(K_0)$.

The following example illustrates the capability of the proposed method to handle multiple defects. The porous defect K_0 comprises six ellipses centred at

$$(0.2, 0.8), (0.2, 0.2), (0.4, 0.8), (0.5, 0.2), (0.7, 0.6), (0.9, 0.1).$$

TABLE 4 *Macroscopic and microscopic errors for different shapes of defects*

$1/H$	$e_0(D \setminus K)$	order	$e_1(D \setminus K)$	order	$1/h$	$e_0(K_0)$	order	$e_1(K_0)$	order
Macro errors with L-shaped defect, $h = 10^{-3}$					Micro errors with L-shaped defect, $H = 10^{-2}$				
20	5.27E-03		6.22E-02		200	1.96E-03		1.99E-01	
40	1.52E-03	1.79	3.04E-02	1.03	400	1.33E-03	0.56	1.65E-01	0.27
80	5.34E-04	1.41	1.54E-02	0.98	800	9.27E-04	0.52	7.75E-02	1.09
160	3.76E-04	0.51	8.12E-03	0.92	1600	8.65E-04	0.10	3.94E-02	0.98
Macro errors with porous defect, $h = H/2$					Micro errors with porous defect, $H = 2h$				
20	5.97E-03		6.17E-02		200	4.05E-03		3.23E-01	
40	3.53E-03	0.76	3.07E-02	1.01	400	2.46E-03	0.72	1.86E-01	0.80
80	2.46E-03	0.52	1.60E-02	0.94	800	2.39E-03	0.04	9.58E-02	0.96
160	1.02E-03	1.27	7.70E-03	1.06	1600	2.00E-03	0.26	5.35E-02	0.84

The semiaxes in the x - and y -directions are, respectively,

$$(R/2, R), (R/2, 2R), (R, R), (2R, R/2), (R, 3R/2), (R, R)$$

with $R = 0.023$. This configuration can be used to model materials with different type of inclusions. The outer set K consists of ellipses with the same centres, but with $R = 0.028$. Consequently, $|K_0| \approx 0.00997$ and $|K| \approx 0.01478$.

Because the porous defect induces very thin elements near the pores and much coarser elements elsewhere, resulting in mesh with poor element, we maintain a fixed ratio between coarse and fine mesh sizes by setting $H = 2h$ in this example. To mitigate the computational cost associated with such meshes, one could employ nonmatching grids and enforce interface conditions via Nitsche's method, as in [Ming & Song \(2022\)](#), but this lies beyond the scope of the present work. We report the macroscopic and microscopic errors in the bottom rows of [Table 4](#) and observe the performance is consistent with that of the L-shaped defect.

4.5 Example 5: coefficient without scale separation

In previous examples, there is scale separation in the diffusion coefficient. This example is taken from [Abdulle & Jecker \(2015\)](#). The coefficient $a^\varepsilon = \chi_{K_0} \tilde{a} + (1 - \chi_{K_0}) \tilde{a}^\varepsilon$, where

$$\tilde{a}(x) = 3 + \frac{1}{7} \sum_{j=0}^4 \sum_{i=1}^j \frac{1}{j+1} \cos \left(\left\lfloor 8 \left(ix_2 - \frac{x_1}{i+1} \right) \right\rfloor + \lfloor 150ix_1 \rfloor + \lfloor 150x_2 \rfloor \right),$$

and

$$\tilde{a}^\varepsilon(x) = 2.1 + \cos(2\pi x_1/\varepsilon) \cos(2\pi x_2/\varepsilon) + \sin(4x_1^2 x_2^2).$$

We still consider the well defect with $K_0 = [0.45, 0.55]^2$ and $K = [0.44, 0.56]^2$, as in Example 1, and set $\varepsilon = 0.0063$ for the experiment. The effective matrix A_H is obtained using the online-offline method proposed in [Huang et al. \(2020\)](#).

TABLE 5 *Macroscopic and microscopic errors for no scale separation coefficient with well defect*

$1/H$	$e_0(D \setminus K)$	order	$e_1(D \setminus K)$	order	$1/h$	$e_0(K_0)$	order	$e_1(K_0)$	order
Macro errors with $h = 10^{-3}$					Micro errors with $H = 10^{-2}$				
20	5.15E-03		7.19E-02		200	5.02E-04		4.23E-02	
40	1.18E-03	2.13	3.42E-02	1.07	400	5.22E-04	-0.06	3.20E-02	0.40
80	3.14E-04	1.91	1.69E-02	1.02	800	5.43E-04	-0.06	1.87E-02	0.78
160	1.16E-04	1.43	8.05E-03	1.07	1600	5.51E-04	-0.02	1.08E-02	0.79

We set $\Delta t = 0.02$ and $h = 1/1000$ and report the macroscopic errors in the left columns of Table 5. The results are similar to those in Example 1. Let $\Delta t = 0.02$ and $H = 1/100$, the microscopic errors are shown in the right columns of Table 5. The quantity $e_0(K_0)$ remains nearly constant at approximately 5×10^{-4} , a behaviour attributable to the discontinuity of $\tilde{a}$ as in Example 2. Meanwhile, the convergent rate for $e_1(K_0)$ stays below 1, reflecting the reduced regularity of $\tilde{a}$ caused by the roughness of u^ε .

5. Conclusion

In this paper, we presented a concurrent global-local numerical method for solving multiscale parabolic equations in divergence form. The proposed method effectively integrates global and local information to accurately capture both macroscopic and microscopic behaviours of the solution, which has been confirmed both theoretically and experimentally. For multiscale parabolic equations with time-independent coefficient, we successfully eliminate the factor $\Delta t^{-1/2}$ for both macroscopic and microscopic accuracy. It is also interesting to study such problems with time-dependent coefficients, time-varying defects and time-varying boundary conditions.

Acknowledgements

The authors thank the anonymous reviewers for their valuable suggestions.

Funding

National Key R&D Program of China (2024YFA1012502); National Natural Science Foundation of China (12371438).

REFERENCES

- ABDULLE, A. & E, W. (2003) Finite difference heterogeneous multi-scale method for homogenization problems. *J. Comput. Phys.*, **191**, 18–39.
- ABDULLE, A., E, W., ENGQUIST, B. & VANDEN-EIJNDEN, E. (2012) The heterogeneous multiscale method. *Acta Numer.*, **21**, 1–87.
- ABDULLE, A. & JECKER, O. (2015) An optimization-based, heterogeneous to homogeneous coupling method. *Commun. Math. Sci.*, **13**, 1639–1648.
- ALBELLA, J., DHIA, H. B., IMPERIALE, S. & RODRÍGUEZ, J. (2019) Mathematical and numerical study of transient wave scattering by obstacles with a new class of Arlequin coupling. *SIAM J. Numer. Anal.*, **57**, 2436–2468.
- ANTONIĆ, N. & VRDOLJAK, M. (2009) Parabolic H-convergence and small-amplitude homogenization. *Appl. Anal.*, **88**, 1493–1508.

- APOUNG KAMGA, J.-B. & PIRONNEAU, O. (2007) Numerical zoom for multiscale problems with an application to nuclear waste disposal. *J. Comput. Phys.*, **224**, 403–413.
- BENSOUSSAN, A., LIONS, J.-L. & PAPANICOLAOU, G. (2011) *Asymptotic Analysis for Periodic Structures*. Providence, RI: American Mathematical Society.
- BRAHIM-OTSMANE, S., FRANCFORT, G. A. & MURAT, F. (1992) Correctors for the homogenization of the wave and heat equations. *J. Math. Pure. Appl.*, **71**, 197–231.
- BRAMBLE, J. H., NITSCHKE, J. A. & SCHATZ, A. H. (1975) Maximum-norm interior estimates for ritz-Galerkin methods. *Math. Comp.*, **29**, 677–688.
- BRAMBLE, J. H., SCHATZ, A. H., THOMÉE, V. & WAHLBIN, L. B. (1977) Some convergence estimates for semidiscrete Galerkin type approximations for parabolic equations. *SIAM J. Numer. Anal.*, **14**, 218–241.
- CHEN, S., E, W. & SHU, C.-W. (2005) The heterogeneous multiscale method based on the discontinuous Galerkin method for hyperbolic and parabolic problems. *Multiscale Model. Simul.*, **3**, 871–894.
- CIARLET, P. G. (2002) The finite element method for elliptic problems. *Classics in Applied Mathematics* (O'Malley, Jr. ed.), vol. 40. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- DALL'AGLIO, A. & MURAT, F. (1997) A corrector result for H -converging parabolic problems with time-dependent coefficients. *Ann. Scuola Norm.*, **25**, 329–373.
- DEMLOW, A., GUZMÁN, J. & SCHATZ, A. (2011) Local energy estimates for the finite element method on sharply varying grids. *Math. Comp.*, **80**, 1–9.
- DHIA, H. B. & RATEAU, G. (2005) The Arlequin method as a flexible engineering design tool. *Internat. J. Numer. Methods Engrg.*, **62**, 1442–1462.
- DU, R. & MING, P. (2010) Convergence of the heterogeneous multiscale finite element method for elliptic problems with nonsmooth microstructures. *Multiscale Model. Simul.*, **8**, 1770–1783.
- E, W. & ENGQUIST, B. (2003) The heterogenous multiscale methods. *Commun. Math. Sci.*, **1**, 87–132.
- E, W., ENGQUIST, B., LI, X., REN, W. & VANDEN-EIJNDEN, E. (2007) Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.*, **2**, 367–450.
- E, W., MING, P. & ZHANG, P. (2005) Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, **18**, 121–156.
- ECKHARDT, D. & VERFÜRTH, B. (2023) Fully discrete heterogeneous multiscale method for parabolic problems with multiple spatial and temporal scales. *Bit*, **63**, 35.
- EVANS, L. C. (2010) Partial differential equations. *Graduate Studies in Mathematics*, vol. 19, 2nd edn. Providence, RI: American Mathematical Society.
- FISH, J. (2009) *Multiscale Methods: Bridging Scales in Science and Engineering*. Oxford, UK: Oxford University Press.
- GERVASIO, P., LIONS, J.-L. & QUARTERONI, A. (2001) Heterogeneous coupling by virtual control methods. *Numer. Math.*, **90**, 241–264.
- GLORIA, A. & HABIBI, Z. (2016) Reduction in the resonance error in numerical homogenization II: correctors and extrapolation. *Found. Comput. Math.*, **16**, 217–296.
- GLOWINSKI, R., HE, J., RAPPAP, J. & WAGNER, J. (2004) A multi-domain method for solving numerically multi-scale elliptic problems. *C. R. Math. Acad. Sci. Paris*, **338**, 741–746.
- GORYNINA, O., LE BRIS, C. & LEGOLL, F. (2021) Some remarks on a coupling method for the practical computation of homogenized coefficients. *SIAM J. Sci. Comput.*, **43**, A1273–A1304.
- HECHT, F. (2012) New development in FreeFem++. *J. Numer. Math.*, **20**, 251–265.
- HOU, T. Y. & WU, X. (1997) A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, **134**, 169–189.
- HU, Q. & GUNZBURGER, M. D. (2000) A gradient method approach to optimization-based multidisciplinary simulations and nonoverlapping domain decomposition algorithms. *SIAM J. Numer. Anal.*, **37**, 1513–1541.
- HUANG, Y., LU, J. & MING, P. (2018) A concurrent global-local numerical method for multiscale PDEs. *J. Sci. Comput.*, **76**, 1188–1215.
- HUANG, Y., MING, P. & SONG, S. (2020) An efficient online-offline method for elliptic homogenization problems. *CSIAM Trans. Appl. Math.*, **1**, 664–692.

- LOZINSKI, A. & PIRONNEAU, O. (2011) Numerical zoom for advection diffusion problems with localized multiscales. *Numer. Methods Partial Differ. Equ.*, **27**, 197–207.
- MING, P. & SONG, S. (2022) A Nitsche hybrid multiscale method with non-matching grids. *J. Sci. Comput.*, **91**, Paper No. 44.
- MING, P. & ZHANG, P. (2007) Analysis of the heterogeneous multiscale method for parabolic homogenization problems. *Math. Comp.*, **76**, 153–177.
- NITSCHKE, J. (1968) Ein kriterium für die quasi-optimalität des ritzschen verfahrens. *Numer. Math.*, **11**, 346–348.
- NITSCHKE, J. A. & SCHATZ, A. H. (1974) Interior estimates for Ritz-Galerkin methods. *Math. Comp.*, **28**, 937–958.
- ODEN, J. T. & VEMAGANTI, K. S. (2000) Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials, I. Error estimates and adaptive algorithms. *J. Comput. Phys.*, **164**, 22–47.
- ODEN, J. T. & VEMAGANTI, K. S. (2001) Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials, II. A computational environment for adaptively modeling of heterogeneous elastic solids. *Comput. Methods Appl. Mech. Engrg.*, **190**, 6089–6124.
- SCHATZ, A. H., THOMÉE, V. & WAHLBIN, L. B. (1998) Stability, analyticity, and almost best approximation in maximum norm for parabolic finite element equations. *Comm. Pure Appl. Math.*, **51**, 1349–1385.
- SHANG, Y. & WANG, K. (2010) Local and parallel finite element algorithms based on two-grid discretizations for the transient stokes equations. *Numer. Algorithms*, **54**, 195–218.
- TARTAR, L. (2010). *The General Theory of Homogenization*. Lecture Notes of the Unione Matematica Italiana, vol. 7. Berlin: Springer.
- THOMÉE, V. (1979) Some interior estimates for semidiscrete Galerkin approximations for parabolic equations. *Math. Comp.*, **33**, 37–62.
- THOMÉE, V. (2007). *Galerkin Finite Element Methods for Parabolic Problems*, 2nd edn. Springer Series in Computational Mathematics, vol. 25. Berlin, Heidelberg: Springer Science & Business Media.
- WANG, X. & DU, G. (2023) Local and parallel stabilized finite element methods based on two-grid discretizations for the stokes equations. *Numer. Algorithms*, **93**, 67–83.
- XU, J. & ZHOU, A. (2001) Local and parallel finite element algorithms based on two-grid discretizations for nonlinear problems. *Adv. Comput. Math.*, **14**, 293–327.
- ZHIKOV, V. V., KOZLOV, S. M. & OLEINIK, O. A. (1981) G-convergence of parabolic operators. *Russian Math. Surveys*, **36**, 9–60.