MULTISCALE MONTE CARLO STUDY OF EPIDERMAL GROWTH FACTOR RECEPTOR DIFFUSION AND DIMERIZATION

by

Stuart D. Collins

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Chemical Engineering

Fall 2009

MULTISCALE MONTE CARLO STUDY OF EPIDERMAL GROWTH FACTOR RECEPTOR DIFFUSION AND DIMERIZATION

by

Stuart D. Collins

Approved: _____
Dionisios G. Vlachos, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Norman J. Wagner, Ph.D.
Chair of the Department of Chemical Engineering

Approved: _____
Michael J. Chajes, Ph.D.
Dean of the College of Engineering

Approved: _____
Debra Hess Norris, M.S.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

Dionisios G. Vlachos, Ph.D. for his academic support, responsiveness, drive, and belief in me.

My professional friends and colleagues, who supported and helped me throughout my graduate education.

The source material from which copyrighted material has been reprinted by permission.

The Department of Energy (DE-FG02-05ER25702), which partially supported this research.

This manuscript is dedicated to:

Mom and Dad, for what has been and continues to be a quality job of parenting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The ErbB family of receptors is dysregulated in a number of cancers, and the signaling pathway of this receptor family is a critical target for several anti-cancer drugs. Therefore, a detailed understanding of the mechanisms of receptors activation is critical. However, despite a plethora of biochemical studies and single particle tracking experiments, the early molecular mechanisms involving epidermal growth factor (EGF) binding and EGF receptor (EGFR) dimerization are not as well understood. Due to the large disparity of time and length scales involved in receptor dimerization reactions, we adapt the coarse-grained Monte Carlo (CGMC) simulation framework to enable the simulation of in vivo receptor diffusion and dimerization.

Using the CGMC method, spatial modeling of ligand-mediated membrane receptor dimerization reaction dynamics was performed. Furthermore, the simulations demonstrate the importance of spatial heterogeneity in membrane receptor localization. Mathematical models, especially one that takes into account spatial heterogeneity, show mechanistic understanding of receptor activation that may in turn enable improved future cancer treatments.

CHAPTER 1

INTRODUCTION

EGFR Background:

Abnormal EGRF behavior is highly correlated with cancerous growth and it is extensively studied [1, 2], and EGFR irregularities have been implicated in breast, lung and prostate cancers, among others[3]. Cancerous cells often exhibit an intense overpopulation of EFGR molecules, compared to healthy counterparts[1]. Receptor aggregation and endocytosis rates are amplified nonlinearly with increasing receptor counts and change the ultimate signal sensitivity and fate decisions of the cell in cancerous cells. By understanding the mechanism of cluster formation, directed research could be put into disrupting the overzealous cluster formation observed in cancerous cells [2].

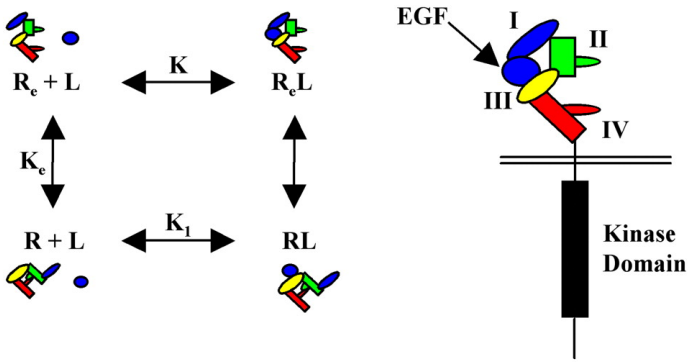

Figure 1.1 - Domains of the epidermal growth factor receptor (EGFR) and their basic functions. While unbound to ligand, domains II and IV interact to 'lock-out' dimerization with steric hindrances. Binding of the EGF ligand opens the receptor for dimerization, leading to cross-phosphorylation of the kinase domain and initiation of the signaling cascade. Reprinted from [4] © 2004 National Academy of Sciences, USA.

Although groups of EGFRs deliver discrete signals and form complex organizations, the fundamental, first tier of EGFR's behavior (that of a single EGFR molecule) contains only a few basic reactions. Each receptor is capable of surface diffusion, dimerization with a partner (and dissociation), ligand binding, and surface interaction (Figure 1.1). Membrane-bound molecules have been observed to diffuse across the surface on multiple time scales, supposedly as different domains of the membrane are occupied, altering the micro and macro diffusion rate [5, 6]. In order to initiate signaling, two receptors must form a dimer, bringing their intra-cellular kinase domains into proximity and starting cross-phosphorylation [7, 8]. Finally, each receptor is capable of binding a single extra-cellular ligand, which modifies the receptor's dimerization propensity thus regulating the initiation of the intra-cellular cascade [4, 9-11].

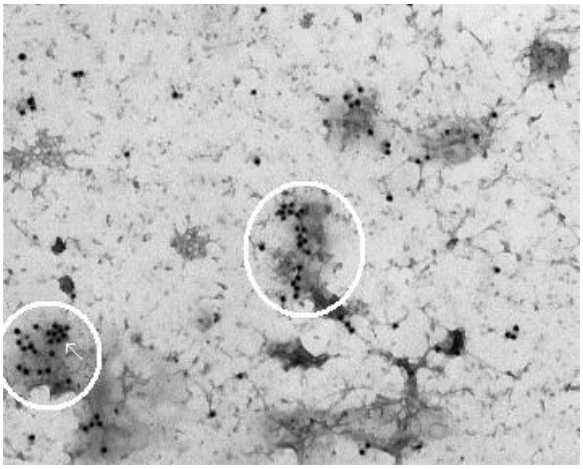

Figure 1.2 - TEM image of a cell membrane with receptors tagged by gold particles (dark dots). Receptors preferentially localize in selective membrane areas, seen here as darker than surrounding areas. The coordinate locations of receptors may be extracted via image analysis and used to analyze organization characteristics. Reprinted from [12], with permission from Elsevier

The next tier of EGFR's behavior is beyond single monomer-monomer binding to many-molecule clustering. Besides bringing kinase domains together to initiate the signaling cascade, these localized and highly concentrated clusters of up to hundreds of members amplify the signal passed into the intracellular space[13, 14] and have also been observed to bud off into the cell in endosomes, bringing the activated receptors off the surface and *into* the cytoplasm. Interaction with other receptors and hydrophobic or hydrophilic regions of the membrane influence this cluster formation and introduction of EGF (EGFR's ligand) further increases EGFR clustering, implying a ligand-regulated modification of the dimerization rates and clustering propensity [15-18]. A notable characteristic of the clusters visible in TEM images (Figure 1.2) is the tendency to gather in darker areas – cholesterol rich clatherin pits [19]. The exact reasons for this specific localization of membrane proteins are not entirely clear, but the formation of the pits themselves is a phase separation of hydrophobic cholesterol-rich membrane components from the more hydrophilic phospholipids predominantly comprising the lipid bilayer.

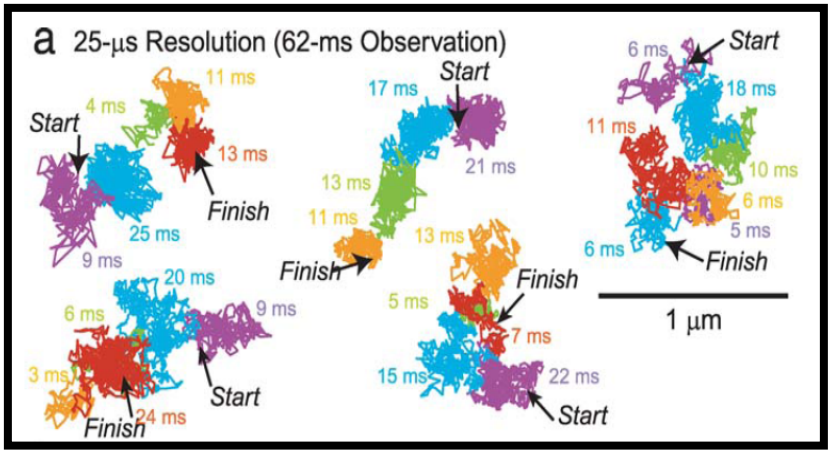

Figure 1.3 – Single particle tracking of individual membrane proteins. Reprinted, with permission, from [20] © 2005 by Annual Reviews www.annualreviews.org.

The classic fluid mosaic model of the cell membrane, by which membrane proteins diffuse freely and react in a two-dimensional space, is insufficient to apply to the EGFR system. Diffusion coefficients for proteins in artificial membranes are higher than those of a natural membrane, and the diffusion rate of dimers drop considerably lower than what would be expected. Single particle tracking experiments explain these observations by revealing compartments, or 'corrals', in natural membranes (Figure 1.3). Within each corral proteins freely diffuse as they would in the fluid mosaic model, but hopping from one corral to another is impeded by 'fence' barriers. The fences themselves are created by steric hindrance from the underlying cytoskeleton [21] or by membrane proteins bound to the cytoskeleton (protein picket model [20]). Since dimer fence hopping is impeded compared to monomers, we propose that the longevity of EGFR clusters may be explained by the combination of high-density areas and the kinetic diffusion barriers of corrals.

Previous simulations aimed at reproducing EGFR's surface clustering have used general attraction potentials to form oligomers [22]. A complementary explanation for

membrane cluster creation is a receptor-receptor attraction independent of the cholesterol rafts. Simulations have demonstrated that small groups of receptors may be held together by reversible dimerization 'partner-switching'[23]. Other experimental observations suggest that dimers may form a weak tetramer (weak in relation to the dimer bond) promoting clustering upon dimerization [16].

In this work we apply the CGMC method and introduce adaptive coarse graining multiscale simulation concepts to accelerate stochastic simulation of spatial systems. Since membrane diffusion exhibits such a range of scales (diffusion within a corral is fast relative to diffusion across corral interfaces), traditional KMC simulation is limited to small length and time scales. CGMC assumes that diffusion within a CG cell is infinity fast while cell-to-cell diffusion is not. By representing corrals as individual CG cells, the CGMC simulation could conceivably describe diffusion accurately while expanding the simulation's scale limits tremendously.

Background of CGMC and KMC:

Traditional kinetic Monte Carlo (KMC) simulations have enjoyed impressive success in the engineering and computational scientific community due to their ability to capture, among others, noise, out-of-equilibrium processes, and complex particle interactions [24-31]. The KMC method allows the simulation of spatial heterogeneous systems with nanoscopic variation. However, many systems are too computationally demanding for KMC simulation due to multiple reasons. KMC simulations are capable of capturing roughly $10^4 – 10^6$ lattice points (100 x 100 nm^2 to 1000 x 1000 nm^2, assuming a lattice constant of 1 nm), putting many systems of interest with correlation lengths at or above the micrometer range out of reach. Long timescale events, such as pattern formation

and aggregation, are also often intractable by KMC simulation [32-34]. Diffusion-controlled systems pose a particular difficulty for KMC simulation due to the hydrodynamic slowdown from the overwhelming number of small diffusion jumps that must be simulated [35]. Calculating long-distance interactions consumes a large fraction of CPU time. Finally, systems with large reaction networks involve too many individual processes for KMC to track, store and search through. This problem, termed as combinatorial complexity in this work, arises in many applications. Examples include biology, due to the huge number of conformations proteins can take [36-38], and epitaxy of metals, due to numerous diffusion barriers arising from different local atomic environments [39-41].

To extend the capabilities of the KMC method to longer time and length scales, the Coarse-Grained Monte Carlo (CGMC) method has recently been developed [30, 35, 42-47]. In our approach, neighboring microscopic sites are grouped together into 'coarse-grained' (CG) cells and a closure is applied at the stochastic level to resident atoms or molecules (here after termed adparticles) to describe their distribution in the cell [35, 46, 47]. In the simplest closure, the local mean-field (LMF) approximation, adparticles within cells are assumed to be well-mixed [35, 46, 47]. Other closures are explored in Ref. [30] and strong interactions in Ref. [48]. The adparticles of each cell are then allowed to interact with, react with, and diffuse to nearby cells.

The CGMC method efficiently addresses many of the stated weaknesses of the traditional KMC method. First, the grouping of microscopic sites simply reduces the number of lattice nodes to be individually tracked and the number of processes to be simulated. Second, as the size of the CG cells increases, the interaction potential length (relative to the CG lattice constant) shrinks, leading to a much faster calculation of the coarse interaction

potential. Third, simulated diffusion jumps are much larger, overcoming (in part) the hydrodynamic slowdown in diffusion-controlled systems.

Thesis Organization

The organization of this thesis is as follows. First the theory of coarse graining for various processes on a heterogeneous surface, followed by numerical examples, is presented in Chapter 2. Next, the performance of the CGMC method is quantified for reaction-diffusion systems with a wide variety of diffusion rates, reaction rates, and lattice coverage in Chapter 3. Chapter 4 brings together the CMGC method from Chapter 2 and the diffusion reaction system analysis of Chapter 3 to apply the CGMC method to the EGFR system. Lastly, conclusions are drawn.

REFERENCES

1. Santon, J.B., et al., *Effects of Epidermal Growth-Factor Receptor Concentration on Tumorigenicity of A431 Cells in Nude-Mice.* Cancer Research, 1986. **46**(9): p. 4701-4705.

2. Henson, E.S. and S.B. Gibson, *Surviving cell death through epidermal growth factor (EGF) signal transduction pathways: Implications for cancer therapy.* Cellular Signalling, 2006. **18**(12): p. 2089-2097.

3. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nature Reviews Molecular Cell Biology, 2001. **2**: p. 127-137.

4. Klein, P., et al., *A structure-based model for ligand binding and dimerization of EGF receptors (vol 101, pg 929, 2004).* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(30): p. 11176-11176.

5. Zidovetzki, R., et al., *Rotational Mobility of High-Affinity Epidermal Growth-Factor Receptors on the Surface of Living A431 Cells.* Biochemistry, 1991. **30**(25): p. 6162-6166.

6. Niehaus, A.M.S., *Simulations of Epidermal Growth Factor Receptor Dynamics on Corralled Membrane Surfaces*, in *Department of Chemical Engineering*. 2007, University of Delaware: Newark, DE, USA. p. 124.

7. Ogiso, H., et al., *Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains.* Cell, 2002. **110**(6): p. 775-787.

8. Fujioka, A., et al., *Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes.* Journal of Biological Chemistry, 2006. **281**(13): p. 8917-8926.

9. Lemmon, M.A., et al., *Two EGF molecules contribute additively to stabilization of the EGFR dimer.* The EMBO Journal, 1997. **16**: p. 281-294.

10. Ferguson, K.M., et al., *EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization.* Molecular Cell, 2003. **11**(2): p. 507-517.

11. Jorissen, R.N., et al., *Epidermal growth factor receptor: mechanisms of activation and signalling.* Experimental Cell Research, 2003. **284**(1): p. 31-53.

12. Zhang, J., et al., *Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy.* Micron, 2006. **37**(1): p. 14-34.

13. Eisinger, J., J. Flores, and W.P. Petersen, *Milling crowd model for local and long-range obstructed lateral diffusion.* Biophys. J., 1986. **49**: p. 987-1001.

14. Ichinose, J., et al., *EGF signaling amplification induced by dynamic clustering of EGFR.* Biochem. Biophys. Res. Comm., 2004. **324**: p. 1143-1149.

15. Gadella, T.W.J. and T.M. Jovin, *Oligomerization of Epidermal Growth-Factor Receptors on A431 Cells Studied by Time-Resolved Fluorescence Imaging Microscopy - a Stereochemical Model for Tyrosine Kinase Receptor Activation.* Journal of Cell Biology, 1995. **129**(6): p. 1543-1558.

16. Clayton, A.H.A., et al., *Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis.* Journal of Biological Chemistry, 2005. **280**(34): p. 30392-30399.

17. Haigler, H., J. McKanna, and S. Cohen, *Direct visualization of the binding and internalization of a ferritin conjugate of epidermal growth factor in human carcinoma cells A-431.* J. Cell Biol., 1979. **81**(2): p. 382-395.

18. Vanbelzen, N., et al., *Direct Visualization and Quantitative-Analysis of Epidermal Growth Factor-Induced Receptor Clustering.* Journal of Cellular Physiology, 1988. **134**(3): p. 413-420.

19. Lillemeier, B.F., et al., *Plasma membrane-associated proteins are clustered into islands attached to the cytoskeleton.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(50): p. 18992-18997.

20. Kusumi, A., et al., *Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules.* Annual Review of Biophysics and Biomolecular Structure, 2005. **34**: p. 351-U54.

21. Kusumi, A. and Y. Sako, *Cell surface organization by the membrane skeleton.* Current Opinion in Cell Biology, 1996. **8**(4): p. 566-574.

22. Gullick, W.J., et al., *Computational simulation of growth factor receptor clustering.* British Journal of Cancer, 2002. **86**: p. S10-S10.

23. Woolf, P.J. and J.J. Linderman, *Self organization of membrane proteins via dimerization.* Biophys. Chem., 2003. **104**: p. 217–227.

24. Binder, K., ed. *Monte Carlo methods in Statistical Physics*. Vol. 7. 1986, Springer-Verlag: Berlin.

25. Landau, D.P. and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics*. 2000, Cambridge: Cambridge University Press.

26. Keil, F.J., R. Krishna, and M.O. Coppens, *Modeling of diffusion in zeolites.* Reviews in Chemical Engineering, 2000. **16**(2): p. 71-197.

27. Auerbach, S.M., *Theory and simulation of jump dynamics, diffusion and phase equilibrium in nanopores.* Int. Rev. Phys. Chem., 2000. **19**(2): p. 155-198.

28. Gilmer, G., *Computer models of crystal growth.* Science, 1980. **208**: p. 355-363.

29. Gilmer, G.H., et al., *Lattice Monte Carlo models of thin film deposition.* Thin Solid Films, 2000. **365**(2): p. 189-200.

30. Chatterjee, A. and D.G. Vlachos, *Coarse-grained kinetic Monte Carlo models for complex lattices and multicomponent systems.* 2007: p. in preparation.

31. Voter, A.F., ed. *Introduction to the Kinetic Monte Carlo Method*. Radiation Effects in Solids, ed. K.E. Sickafus and E.A. Kotomin. 2006, Springer, NATO Publishing Unit: Dordrecht, The Netherlands. in press.

32. Kapur, S.S., et al., *Role of configurational entropy in the thermodynamics of clusters of point defects in crystalline solids.* Physical Review B, 2005. **72**(1).

33. Glotzer, S.C., E.A. Di Marzio, and M. Muthukumar, *Reaction-controlled morphology of phase-separating mixtures.* Physical Review Letters, 1995. **74**(11): p. 2034-2037.

34. Starke, J., et al., *Fluctuation-induced pattern formation in a surface reaction.* Europhysics Letters, 2006. **73**(6): p. 820-825.

35. Katsoulakis, M.A. and D.G. Vlachos, *Coarse-grained stochastic processes and kinetic Monte Carlo simulators for the diffusion of interacting particles.* J. Chem. Phys., 2003. **119**(18): p. 9412-9428.

36. Borisov, N.M., et al., *Signaling through Receptors and Scaffolds: Independent Interactions Reduce Combinatorial Complexity.* Biophys. J., 2005. **89**: p. 951–966.

37. Borisov, N.M., et al., *Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains.* BioSystems, 2006. **83**: p. 152–166.

38. Hlavacek, W.S., et al., *Rules for Modeling Signal-Transduction Systems.* Science STKE, 2006: p. 1-18.

39. Sprague, J.A., et al., *Simulation of growth of Cu on Ag(001) at experimental deposition rates.* Physical Review B, 2002. **66**(20): p. art. no.-205415.

40. Voter, A.F., *Classically exact overlayer dynamics: Diffusion of rhodium clusters on Rh(100).* Physical review B, 1986. **34**(10): p. 6819-6829.

41. Feibelman, P.J., *Diffusion path for an Al adatom on Al(001).* Physical Review Letters, 1990. **65**(6): p. 729-732.

42. Chatterjee, A. and D.G. Vlachos, *Multiscale spatial Monte Carlo simulations: Multigriding, computational singular perturbation, and hierarchical stochastic closures.* Journal of Chemical Physics, 2006. **124**(6): p. 0641101-06411016.

43. Ismail, A.E., G. Stephanopoulos, and G.C. Rutledge, *Wavelet-accelerated Monte Carlo sampling of polymer chains.* Journal of Polymer Science Part B-Polymer Physics, 2005. **43**(8): p. 897-910.

44. Katsoulakis, M.A. and D.G. Vlachos, *Coarse-grained stochastic processes and kinetic Monte Carlo simulators for the diffusion of interacting particles.* Journal of Chemical Physics, 2003. **119**(18): p. 9412-9427.

45. Katsoulakis, M.A., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems.* Journal of Computational Physics, 2003. **186**(1): p. 250-278.

46. Katsoulakis, M., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes for microscopic lattice systems.* Proc. Natl. Acad. Sci., 2003. **100**(3): p. 782-787.

47. Katsoulakis, M.A., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems.* J. Comp. Phys., 2003. **186**: p. 250-278.

48. Dai, J.G., W.D. Seider, and T. Sinno, *Coarse-grained lattice kinetic Monte Carlo simulation of systems of strongly interacting particles* J. Chem. Phys., 2008. **128**(19): p. 1947051-19470517.

CHAPTER 2

COARSE-GRAINED KINETIC MONTE CARLO MODELS: COMPLEX LATTICES, MULTICOMPONENT SYSTEMS, AND HOMOGENIZATION AT THE STOCHASTIC LEVEL

Abstract

On-lattice kinetic Monte Carlo (KMC) simulations have extensively been applied to numerous systems. However, their applicability is severely limited to relatively short time and length scales. Recently, the Coarse-Grained MC (CGMC) method was introduced to greatly expand the reach of the lattice KMC technique. Herein, we extend the previous spatial CGMC methods to multicomponent species and/or site types. The underlying theory is derived and numerical examples are presented to demonstrate the method. Furthermore, we introduce the concept of homogenization at the stochastic level over all site types of a spatially coarse-grained cell. Homogenization provides a novel coarsening of the number of processes, an important aspect for complex problems plagued by the existence of numerous microscopic processes (combinatorial complexity). As expected, the homogenized CGMC method outperforms the traditional KMC method on computational cost while retaining good accuracy.

Introduction

Previous CGMC simulations have focused on uniform surfaces comprised of a single type of microscopic site with a single type of adparticle. In this chapter, we extend our previous CGMC method [2-4] to an arbitrary number of site types and/or adparticle species. Lattice-based simulations of multiple site type and adparticle species systems have been performed previously using a mean-field (MF) estimate[5]. This extension allows CGMC to be applied to a much wider range of systems of interest, such as catalytic reaction systems [6] (where various site types may represent different elements or lattice positions), diffusion on surfaces and in nanoporous materials [7-9], and biological signaling [10-12] (where site types may represent distinct areas of a cell membrane). In order to overcome the problem of combinatorial complexity, the concept of homogenization at the stochastic level is introduced.

The organization of this chapter is as follows. First, the theory of coarse graining for various processes including adsorption, desorption, isomerization, reaction, and diffusion is described. Next, a simple MC algorithm used to implement the theory is presented, followed by numerical examples. The concept of homogenization is then outlined and illustrated with an example from diffusion in zeolites. Conclusions are finally drawn.

Multicomponent and Multisite Coarse Grained Systems

Coarse Lattice Description

This model follows previously published coarse graining procedures. For a description of the underlying microscopic model and mechanisms, refer to Ref.[13]. Major

nomenclature is summarized in Table 2.1. As in the previous CGMC method, microscopic lattice sites are grouped into CG cells forming a coarse lattice L_c. The CG cells are denoted as C_k ($k = 1,...,m$). While the number of occupants within each CG cell is recorded, exact locations are not. Instead, the local mean field (LMF) approximation is assumed to hold within CG cells. Let N_{sttype} be the number of microscopic site types. Each cell C_k contains $q_{\varphi k}$ microscopic sites of type φ, $\varphi = 1,..., N_{sttype}$, such that $q_{\varphi k}$ is a non-negative integer number. For a non-uniform mesh, each cell C_k may have a different $q_{\varphi k}$ value. The total number of sites in C_k is given by $\sum_{\varphi=1}^{N_{sttype}} q_{\varphi k}$. In the limit of no coarse-graining, i.e., when $\sum_{\varphi=1}^{N_{sttype}} q_{\varphi k} = 1$, the microscopic lattice description is recovered. The number of different surface atomic or molecular species is N_{sp}. Each adparticle type is referred to by an index α, α = 1, …,N_{sp}+1, and a site vacancy is the $(N_{sp}+1)^{th}$ surface species, referred to as $\alpha = N_{sp} + 1 = \phi$ for convenience.

Table 2.1 – Major Nomenclature

Term	Type	Description
$\varphi = 1,\ldots,N_{\text{st type}}$	Integer	Lattice site type
$\alpha = 1,\ldots,N_{sp};\alpha = \phi$	Integer	Surface species type; Vacancy
$\eta_{\varphi k}^{\alpha}$	Integer	Number of α adparticles on φ type sites in CG cell C_k
$q_{\varphi k}$	Integer	Number of φ type sites in CG cell C_k
$\theta_{\varphi k}^{\alpha}$	Real	Coverage of species α on φ type sites in CG cell C_k
$\theta_{\varphi k,\alpha'\varphi'}^{\alpha}$	Real	Coverage of species α on φ type sites in CG cell C_k, given that one α' adparticle is on a φ' type site
$\bar{U}_{\varphi k}^{\alpha}$	Real	The CG interaction energy of species α on φ type sites in CG cell C_k
Γ	Real	Microscopic event frequency
$\bar{\Gamma}$	Real	Averaged event frequency
$\bar{\Gamma}$	Real	Coarse-grained event transition probability rate
Γ_c	Real	Lumped coarse-grained transition probability rate

The CG occupation function for species α, $\alpha = 1,\ldots,N_{sp}+1$, chemisorbed on φ type of sites in cell C_k is given by

$$\eta_{\varphi k}^{\alpha} = \sum_{j \in C_k} \sigma_{\varphi j}^{\alpha} \quad . \tag{2.1}$$

Furthermore, the exclusion principle requires that

$$\sum_{\alpha=1}^{N_{sp}+1} \eta_{\varphi k}^{\alpha} = q_{\varphi k} \quad . \tag{2.2}$$

The CG occupancy tensor is defined as

$$\boldsymbol{\eta} = \{\eta_{\varphi k}^{\alpha}\}_{\varphi=1,\ldots,N_{sttype},k=1,\ldots,m,\alpha=1,\ldots,N_{sp}+1} \quad . \tag{2.3}$$

Energetics and Hamiltonian

The LMF approximation is employed throughout this chapter to derive all coarse-grained quantities in terms of the coarse observable $\boldsymbol{\eta}$ in a closed form. The coarse interaction potential is a projection over the two-body microscopic interaction potentials $J_{\varphi\varphi'}^{\alpha\alpha'}(\mathbf{r})$ using Haar wavelets[4]. For all expressions, the microscopic interaction energy is recovered in the limit of zero coarse-graining.

The CG interaction potential for two chemisorbed species on sites of the same type in Ck is given by

$$\bar{J}_{\varphi k,\varphi k}^{\alpha\alpha'} = \frac{1}{q_{\varphi k}(q_{\varphi k}-1)} \sum_{j\in C_k} \sum_{\substack{\mathbf{r} \\ j'\in C_k}} J_{\varphi\varphi}^{\alpha\alpha'}(\mathbf{r}) \quad \varphi = \varphi' \qquad q_{\varphi k} > 1 \quad , \quad \text{, when} \tag{2.4}$$

and $\bar{J}_{\varphi k,\varphi k}^{\alpha\alpha'} = 0$ otherwise when $q_{\varphi k} \leq 1$. Herein, overbars denote coarse variables(an average across all possible configurations within the k^{th} cell that satisfy the constraints regarding the $\boldsymbol{\eta}$ values). The term $q_{\varphi k} - 1$ in the denominator of Eq. (2.4) arises

because of the exclusion principle. Based on the position of the j^{th} site, the position vector $\mathbf{r}$ automatically specifies the position of site j'.

The CG interaction potential between two species on sites of different types in C_k is given by

$$\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k} = \frac{1}{q_{\varphi k} q_{\varphi' k}} \sum_{j\in C_k} \sum_{\substack{\mathbf{r} \\ j'\in C_k}} J^{\alpha\alpha'}_{\varphi\varphi'}(\mathbf{r}) \quad \varphi \neq \varphi' \qquad q_{\varphi k}, q_{\varphi' k} \geq 1 \quad \text{, when} \tag{2.5}$$

and $\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k} = 0$ otherwise.

The CG interaction potential between two species in two different CG cells C_k and $C_{k'}$ (regardless of site types) is given by

$$\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'} = \frac{1}{q_{\varphi k} q_{\varphi' k'}} \sum_{j\in C_k} \sum_{\substack{\mathbf{r} \\ j'\in C_{k'}}} J^{\alpha\alpha'}_{\varphi\varphi'}(\mathbf{r}) \quad k \neq k' \qquad q_{\varphi k}, q_{\varphi' k'} \geq 1 \quad \text{, when} \tag{2.6}$$

and $\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'} = 0$ otherwise. Note that the CG interaction potential is symmetric, i.e., $\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi k} = \bar{J}^{\alpha'\alpha}_{\varphi k,\varphi k}$ and $\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'} = \bar{J}^{\alpha'\alpha}_{\varphi k,\varphi' k'} = \bar{J}^{\alpha\alpha'}_{\varphi' k',\varphi k}$.

The CG Hamiltonian is given by

$$\bar{H}(\boldsymbol{\eta}) = -\frac{1}{2}\sum_k \sum_\alpha \sum_\varphi \eta^\alpha_{\varphi k} \left(\bar{J}^{\alpha\alpha}_{\varphi k,\varphi k}\left(\eta^\alpha_{\varphi k} - 1\right) + \sum_{\substack{\alpha' \\ \alpha'\neq\alpha}} \bar{J}^{\alpha\alpha'}_{\varphi k,\varphi k}\eta^{\alpha'}_{\varphi k} + \sum_{k'}\sum_{\alpha'}\sum_{\varphi'} \bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'}\eta^{\alpha'}_{\varphi k} \right) + \sum_k \sum_\alpha \sum_\varphi \bar{h}^\alpha \eta^\alpha_{\varphi k}, \tag{2.7}$$

where $\bar{h}^\alpha$ is the external field interaction per particle of type α. The CG interaction energy $\bar{U}^\alpha_{\varphi k}$ of species α on site φ in CG cell C_k is written as a sum of contributions from within and between CG cells, such that

$$\bar{U}^{\alpha}_{\varphi k} = \bar{U}^{\alpha}_{\varphi k,k} + \sum_{\substack{k' \in L_c \\ k' \neq k}} \bar{U}^{\alpha}_{\varphi k,k'} \tag{2.8}$$

where the within-cell contribution is given by

$$\bar{U}^{\alpha}_{\varphi k,k} = \bar{J}^{\alpha\alpha}_{\varphi k,\varphi k}(\eta^{\alpha}_{\varphi k} - 1) + \sum_{\substack{\alpha' \\ \alpha' \neq \alpha}} \bar{J}^{\alpha\alpha'}_{\varphi k,\varphi k} \eta^{\alpha'}_{\varphi k} + \sum_{\alpha'} \sum_{\substack{\varphi' \\ \varphi' \neq \varphi}} \bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k} \eta^{\alpha'}_{\varphi' k} \tag{2.9}$$

and the CG cell interactions between two different CG cells are given by

$$\bar{U}^{\alpha}_{\varphi k,k'} = \sum_{\varphi'} \sum_{\alpha'} \bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'} \eta^{\alpha'}_{\varphi' k} \quad k \neq k'$$

, . (2.10)

Coarse-grained (CG) process dynamics

CG transition probability rates are derived for various processes, listed below. In the limit of no coarse-graining, the microscopic transition probability rates are obtained.

Adsorption

The CG transition probability rate for adsorption on sites of type φ in cell C_k is derived as a function of the expected value of the microscopic adsorption rates as

$$\sum_{j \in C_k} \left\langle \Gamma^{\alpha}_{ads}(\boldsymbol{\sigma}, \varphi, j) \right\rangle = \sum_{j \in C_k} \left\langle \Gamma^{\alpha}_{ads,\varphi j} \sigma^{\phi}_{\varphi j} \right\rangle \approx \left(\frac{1}{q_{\varphi k}} \sum_{j \in C_k} \Gamma^{\alpha}_{ads,\varphi j} \right) \eta^{\phi}_{\varphi k} = \bar{\Gamma}^{\alpha}_{ads,\varphi k} \eta^{\phi}_{\varphi k}. \tag{2.11}$$

In (2.11), $\bar{\Gamma}^{\alpha}_{ads,\varphi k}$ is the average adsorption frequency for C_k, a function of the microscopic adsorption frequency, Γ, and the number of vacant sites in C_k. The approximation sign indicates that the local adsorption rate constant is replaced with a spatial

average if there is variation in the adsorption frequency between sites. Summarizing, the CG adsorption transition probability rate is given by

.

$$\overline{\Gamma}^{\alpha}_{ads}(\boldsymbol{\eta},\varphi,k)=\overline{\Gamma}^{\alpha}_{ads,\varphi k}\eta^{\phi}_{\varphi k} \tag{2.12}$$

The CG dissociative adsorption transition probability rates have a form similar to the corresponding CG transition probability rates for surface disproportionation reactions derived below.

Desorption and Isomerization Reaction

The CG transition probability rate for desorption from sites of type φ in cell C_k is derived as

$$\sum_{j\in C_k}\left\langle\Gamma^{\alpha}_{des}(\boldsymbol{\sigma},\varphi,j)\right\rangle=\sum_{j\in C_k}\left\langle\Gamma^{\alpha}_{des,\varphi j}\sigma^{\alpha}_{\varphi j}e^{-\beta U^{\alpha}_{\varphi j}}\right\rangle\approx\left(\frac{1}{q_{\varphi k}}\sum_{j\in C_k}\Gamma^{\alpha}_{des,\varphi j}\right)\sum_{j\in C_k}\left\langle\sigma^{\alpha}_{\varphi j}e^{-\beta U^{\alpha}_{\varphi j}}\right\rangle$$
$$\approx\left(\frac{1}{q_{\varphi k}}\sum_{j\in C_k}\Gamma^{\alpha}_{des,\varphi j}\right)\eta^{\alpha}_{\varphi k}e^{-\beta\overline{U}^{\alpha}_{\varphi k}}=\overline{\Gamma}^{\alpha}_{des,\varphi k}\eta^{\alpha}_{\varphi k}e^{-\beta\overline{U}^{\alpha}_{\varphi k}}. \tag{2.13}$$

The same assumption regarding the desorption frequencies, mentioned above for adsorption, is employed here as well. Summarizing, the CG desorption transition probability rate is given by

$$\overline{\Gamma}^{\alpha}_{des}(\boldsymbol{\eta},\varphi,k)=\overline{\Gamma}^{\alpha}_{des,\varphi k}\eta^{\alpha}_{\varphi k}e^{-\beta\overline{U}^{\alpha}_{\varphi k}}. \tag{2.14}$$

Likewise, the CG isomerization reaction transition probability rate is given by

$$\bar{\Gamma}^{\alpha}_{\text{iso}}(\boldsymbol{\eta}, \varphi, k) = \bar{\Gamma}^{\alpha\to\alpha'}_{\text{iso},\varphi k}\,\eta^{\alpha}_{\varphi k}. \tag{2.15}$$

Here,

$$\bar{\Gamma}^{\alpha\to\alpha'}_{\text{iso},\varphi k} = \sum_{j\in C_k} \Gamma^{\alpha\to\alpha'}_{\text{iso},\varphi j} / q_{\varphi k} \tag{2.16}$$

is the average reaction frequency for a non-activated isomerization reaction process.

Surface Disproportionation Reaction

The CG transition probability rate for the disproportionation reaction $\alpha + \alpha' \to \alpha'' + \alpha'''$ in cell C_k is partitioned into reactions inside C_k and reaction at the boundaries with neighboring cells, i.e.,

$$\sum_{j\in C_k} \Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{\text{rxn},\varphi\varphi'}(\boldsymbol{\sigma}, \varphi, j, \varphi', j') = \frac{1}{2}\sum_{j\in C_k}\sum_{\substack{j'\in C_k \\ j'\neq j}} \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{\text{rxn},\varphi\varphi'}}{n_{\varphi\varphi'}}\sigma^{\alpha}_{\varphi j}\sigma^{\alpha'}_{\varphi' j'} + \sum_{j\in C_k}\sum_{\substack{k' \\ k'\neq k}}\sum_{j'\in C_{k'}} \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{\text{rxn},\varphi\varphi'}}{n_{\varphi\varphi'}}\sigma^{\alpha}_{\varphi j}\sigma^{\alpha'}_{\varphi' j'}, \tag{2.17}$$

where j and j' are adjacent sites, and $n_{\varphi\varphi'}$ gives the number of adjacent φ' sites around a φ site. This is divided by $n_{\varphi\varphi'}$ to arrive at the reaction frequency per pair of reactants. The first term in the R.H.S. of (2.17) gives the sum of reaction transition probability rates on sites inside the cell, while the second term gives the sum of reaction transition probability rates when one of the reacting sites belongs to cell C_k and other one belongs to cell $C_{k'}$. The factor of ½ prevents double-counting of the same pair of sites within

C_k. Due to the exclusion principle, the first term is absent in the microscopic disproportionation reaction transition probability rate expression ($q_{\varphi k} = 1$).

Using the LMF approximation inside a CG cell, the average species α' coverage on φ' sites in C_k is given by $\eta_{\varphi' k}^{\alpha'} / q_{\varphi' k}$. To use this to find CG reaction transition probability rates boundary corrections and the exclusion principle need to be accounted for in the case of a finite-size CG cell. The average number of $\alpha - \alpha'$ reacting pairs within cell C_k is given by

$$(n_{\varphi\varphi'} q_{\varphi k} - \sum_{k' \in \text{neighbor of } k} q^{B}_{\varphi k, \varphi' k'}) \frac{\eta^{\alpha}_{\varphi k} \eta^{\alpha'}_{\varphi' k}}{q_{\varphi k} q_{\varphi' k}} \text{, and} \tag{2.18}$$

$$(n_{\varphi\varphi} q_{\varphi k} - \sum_{k' \in \text{neighbor of } k} q^{B}_{\varphi k, \varphi k'}) \frac{\eta^{\alpha}_{\varphi k} \eta^{\alpha'}_{\varphi k}}{q_{\varphi k} (q_{\varphi k} - 1)} , \tag{2.19}$$

for $\varphi' \neq \varphi$ and $\varphi' = \varphi$, respectively. The difference in the denominator of Eq. (2.19) reflects that when both reactants use the same site type, the first reactant necessarily occupies one of the potential sites of the second reactant, restricting the number of potential host sites within a finite CG cell for the second reactant by 1. Here $q^{B}_{\varphi k, \varphi' k'}$ gives the number of $\varphi' - \varphi$ pair sites at the boundary $B_{kk'}$ between cells C_k and $C_{k'}$, such that the φ' sites lie in cell $C_{k'}$. The term in the parenthesis indicates the number of pairs of sites within cell C_k. Using Eqs. (2.17), (2.18), and (2.19) the CG reaction transition probability rate inside the CG cell C_k is given by

$$\bar{\Gamma}^{\alpha'', \alpha'''}_{\text{rxn}}(\boldsymbol{\eta}, \varphi, k, \varphi', k) = \bar{\Gamma}^{\alpha' \to \alpha'' \alpha'''}_{\text{rxn}, \varphi\varphi'} \eta^{\alpha}_{\varphi k} \eta^{\alpha'}_{\varphi' k} , \tag{2.20}$$

where

$$\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi k\varphi' k'} = \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi\varphi'}}{n_{\varphi\varphi'}}\left(\frac{n_{\varphi\varphi'}q_{\varphi k} - \sum_{k'} q^{B}_{\varphi k,\varphi' k'}}{q_{\varphi k}q_{\varphi' k}}\right) \quad \text{or} \tag{2.21}$$

$$\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi k\varphi k} = \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi\varphi}}{n_{\varphi\varphi}}\left(\frac{n_{\varphi\varphi}q_{\varphi k} - \sum_{k'} q^{B}_{\varphi k,\varphi k}}{q_{\varphi k}(q_{\varphi k}-1)}\right) \tag{2.22}$$

is the CG hopping frequency. Note that the factor of ½, present in Eq. (2.17) to prevent double-counting of pairs is absent from Eq. (2.20) because the number of pairs when $\varphi' \neq \varphi$ or $\alpha' \neq \alpha$ is given by Eqs. (2.18) and (2.19). However, when $\alpha' = \alpha$ and $\varphi' = \varphi$, a modified form of Eq. (2.20) is obtained due to the exclusion principle

$$\bar{\Gamma}^{\alpha\alpha\to\alpha''\alpha'''}_{rxn}(\eta,\varphi,k,\varphi,k') = \frac{1}{2}\Gamma^{\alpha\alpha\to\alpha''\alpha'''}_{rxn,\varphi\varphi}\eta^{\alpha}_{\varphi k}(\eta^{\alpha}_{\varphi k}-1) \,. \tag{2.23}$$

As in Eq. (2.17), the factor of ½ prevents double-counting of the same pair of reactants within C_k. In arriving at Eq. (2.21), it has been tacitly assumed that all sites of the same type have the same frequency for reaction.

The CG disproportionation reaction transition probability rate for reactions between the CG cell C_k and $C_{k'}$ accounts for the missing boundary reaction terms in Eq. (2.17 – 2.23). Starting with the second term in the R.H.S. of Eq. (2.17), one gets a CG disproportionation reaction transition probability rate for each pair of cells $C_k - C_{k'}$ given by

$$\sum_{j\in C_k}\sum_{j'\in C_{k'}} \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi_j\varphi'_{j'}}}{n_{\varphi_j\varphi'_{j'}}}\sigma^{\alpha}_{\varphi j}\sigma^{\alpha'}_{\varphi' j'} = \frac{\Gamma^{\alpha\alpha'\to\alpha''\alpha'''}_{rxn,\varphi\varphi'}}{n_{\varphi\varphi'}}\sum_{j\in C_k\cap B_{kk'}}\sum_{j'\in C_{k'}\cap B_{kk'}}\sigma^{\alpha}_{\varphi j}\sigma^{\alpha'}_{\varphi' j'} \,. \tag{2.24}$$

The summation occurs over sites common to a CG cell along the boundary with the neighboring CG cell. Defining the average reaction frequency at the $C_k - C_{k'}$ boundary as

$$\bar{\Gamma}^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} = \frac{\Gamma^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn,\varphi\varphi'}}{n_{\varphi\varphi'}} \frac{q^{B}_{\varphi k,\varphi' k'}}{q_{\varphi k} q_{\varphi' k'}} \quad k \neq k' \quad , \quad , \tag{2.25}$$

the CG disproportionation reaction transition probability rate at the $C_k - C_{k'}$ boundary is given by

$$\bar{\Gamma}^{\alpha'',\alpha'''}_{rxn}(\boldsymbol{\eta}, \varphi, k, \varphi', k') = \bar{\Gamma}^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} \eta^{\alpha}_{\varphi k} \eta^{\alpha'}_{\varphi' k'} \quad k \neq k' \quad . \tag{2.26}$$

One can define an effective CG disproportionation reaction transition probability rate for cell C_k by combining Eq. (2.20) or (2.23) and Eq. (2.26) given by

$$\bar{\Gamma}^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn}(\boldsymbol{\eta}, \varphi, k, \varphi') = \begin{cases} (\bar{\Gamma}^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k} \eta^{\alpha'}_{\varphi' k} + \frac{1}{2} \sum_{\substack{k' \\ k' \neq k}} \bar{\Gamma}^{\alpha\alpha' \to \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} \eta^{\alpha'}_{\varphi' k'}) \eta^{\alpha}_{\varphi k}, & \alpha' \neq \alpha \\ (\bar{\Gamma}^{\alpha\alpha \to \alpha'' \alpha'''}_{rxn,\varphi\varphi} (\eta^{\alpha}_{\varphi k} - 1) + \frac{1}{2} \sum_{\substack{k' \\ k' \neq k}} \bar{\Gamma}^{\alpha\alpha \to \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} \eta^{\alpha}_{\varphi' k'}) \eta^{\alpha}_{\varphi k}, & \text{otherwise} \end{cases} . \tag{2.27}$$

Herein, the contributions from reactions at the cell boundaries are equally split between two adjoining cells, and hence the factor of ½ in Eq. (2.27). One can easily see the possibility for a very large number of transition probability rates with relatively few molecular species combining in many ways as reactants or products on various hosting site types. This explosion in the number of processes that must be tracked and searched through

can be very challenging for a simulation. To address this problem we introduce the concept of CG cell homogenization in a later section.

Diffusion

A CG hopping transition probability rate for an adparticle of species type α on site of type φ in cell C_k hopping to site of type φ in cell $C_{k'}$ is postulated and is given by

$$\overline{\Gamma}^{\alpha}_{mig}(\eta, \varphi, k, k') = \overline{\Gamma}_{mig,\varphi k \to \varphi k'} \frac{\eta^{\alpha}_{\varphi k} \eta^{\phi}_{\varphi k'}}{q_{\varphi k} q_{\varphi k'}} e^{-\beta \overline{U}^{\alpha}_{\varphi k}} \quad . \tag{2.28}$$

An expression for the unknown CG hopping frequency $\overline{\Gamma}_{m,\varphi k \to \varphi' k'}$ is derived next such that Eq. (2.28) reduces to the microscopic one in the limit of no coarse-graining. The cell coverages are defined in terms of an ensemble average $c^{\alpha}_{\varphi}(\vec{r}) = \left\langle \eta^{\alpha}_{\varphi k} \right\rangle / q_{\varphi k}$ and the mesoscopic diffusion flux equations are derived using Taylor expansions.

As an example, a Taylor expansion in the α species coverage at φ sites is written for cell $C_{k'}$ about the spatial coordinates of cell C_k and is given by

$$\frac{\left\langle \eta^{\alpha}_{\varphi k'} \right\rangle}{q_{\varphi k'}} = \left[c^{\alpha}_{\varphi} + \sum_{\xi} (\vec{a}_{\varphi k \varphi k'} . \vec{\xi}) \partial_{\xi} c^{\alpha}_{\varphi} + \frac{1}{2} \sum_{\xi} (\vec{a}_{\varphi k \varphi k'} . \vec{\xi})^2 \partial^2_{\xi^2} c^{\alpha}_{\varphi} + (\vec{a}_{\varphi k \varphi k'} . \vec{x})(\vec{a}_{\varphi k \varphi k'} . \vec{y}) \partial^2_{xy} c^{\alpha}_{\varphi} \right]_k \tag{2.29}$$

where $\vec{\xi} = \{\vec{x}, \vec{y}\}$ are the Cartesian coordinates of the diffusion hop's direction vector. This framework is easily extended to systems of any number of dimensions. Here, $\vec{a}_{\varphi k \varphi' k'}$ is the center-to-center vector from $C_{k'}$ to C_k. In this case, the center of C_k is the average spatial location of φ sites in C_k. For large cells one can typically use the geometric cell center. Likewise, the Arrhenius factor is expanded to give

$$\left\langle e^{-\beta \bar{U}^{\alpha}_{\varphi k'}} \right\rangle = e^{-\beta \left\langle \bar{U}^{\alpha}_{\varphi k'} \right\rangle} =$$

$$\left[e^{-\beta \bar{U}^{\alpha}_{\varphi}} \begin{pmatrix} 1 - \sum_{\xi} (\mathbf{a}_{\varphi k \varphi k'} . \boldsymbol{\xi}) \beta \partial_{\xi} \bar{U}^{\alpha}_{\varphi} + \frac{1}{2} \sum_{\xi} (\mathbf{a}_{\varphi k \varphi k'} . \boldsymbol{\xi})^2 (\beta^2 (\partial_{\xi} \bar{U}^{\alpha}_{\varphi})^2 - \beta \partial^2_{\xi^2} \bar{U}^{\alpha}_{\varphi}) \\ + (\mathbf{a}_{\varphi k \varphi k'} . \mathbf{x})(\mathbf{a}_{\varphi k \varphi k'} . \mathbf{y})(\beta^2 \partial_x \bar{U}^{\alpha}_{\varphi} \partial_y \bar{U}^{\alpha}_{\varphi} - \beta \partial^2_{xy} \bar{U}^{\alpha}_{\varphi}) \end{pmatrix} \right]_k . \quad (2.30)$$

Using Eqs. (2.29) and (2.30) the net rate of α adparticles diffusing from C_k to $C_{k'}$ along the positive $\boldsymbol{\xi}$ axis is given by

$$\mathbf{N}.\boldsymbol{\xi} = \Gamma^{\alpha}_{mig,\varphi k' \to \varphi k} \frac{\eta^{\alpha}_{\varphi k'} \eta^{\phi}_{\varphi k}}{q_{\varphi k'} q_{\varphi k}} e^{-\beta \bar{U}^{\alpha}_{\varphi' k'}} - \Gamma^{\alpha}_{mig,\varphi k \to \varphi' k'} \frac{\eta^{\alpha}_{\varphi k} \eta^{\phi}_{\varphi k'}}{q_{\varphi k} q_{\varphi k'}} e^{-\beta \bar{U}^{\alpha}_{\varphi k}} , \quad (2.31)$$

and matching the resulting terms with the corresponding continuum equation obtained from the microscopic hopping transition probability rate, one gets

$$\Gamma^{\alpha}_{mig,\varphi k' \to \varphi k} = \Gamma^{\alpha}_{mig,\varphi k \to \varphi k'} = \Gamma^{\alpha}_{mig,\varphi k k'} . \quad (2.32)$$

Using Eqs. (2.31) and (2.32), the CG flux is given by

$$\mathbf{N}.\boldsymbol{\xi} = \sum_{k'} \mathbf{N}_{k'} = -\frac{e^{-\beta \bar{U}^{\alpha}}}{2} \sum_{k'} \frac{\Gamma^{\alpha}_{mig,\varphi k k'}}{q^{B}_{\varphi k \varphi' k'}} (\mathbf{a}_{\varphi k \varphi k'} . \boldsymbol{\xi})(\eta^{\phi}_{\varphi} \partial_{\xi} c^{\alpha}_{\varphi} - \eta^{\alpha}_{\varphi} \partial_{\xi} c^{\phi}_{\varphi} - \beta \eta^{\alpha}_{\varphi} \eta^{\phi}_{\varphi} \partial_{\xi} U^{\alpha}_{\varphi}) \quad (2.33)$$

when only terms $\mathbf{a}_{\varphi k \varphi k'} . \boldsymbol{\xi} > 0, \forall k'$ are included in the summation (diffusion with a positive $\boldsymbol{\xi}$ component), and

$$\mathbf{N}.\boldsymbol{\xi} = \sum_{k'} \mathbf{N}_{k'} = \frac{e^{-\beta \bar{U}^{\alpha}}}{2} \sum_{k'} \frac{\Gamma^{\alpha}_{mig,\varphi k k'}}{q^{B}_{\varphi k \varphi' k'}} (\mathbf{a}_{\varphi k \varphi k'} . \boldsymbol{\xi})(\eta^{\phi}_{\varphi} \partial_{\xi} c^{\alpha}_{\varphi} - \eta^{\alpha}_{\varphi} \partial_{\xi} c^{\phi}_{\varphi} - \beta \eta^{\alpha}_{\varphi} \eta^{\phi}_{\varphi} \partial_{\xi} U^{\alpha}_{\varphi}) \quad (2.34)$$

when only terms $\hat{a}_{\varphi k \varphi k'} \cdot \hat{\xi} < 0, \forall k'$ are included in the summation (diffusion with a negative $\hat{\xi}$ component). Comparing the terms in the CG and microscopic diffusion flux expression, one gets

$$\hat{\Gamma}^{\alpha}_{\text{mig},\varphi k k'} = \frac{\Gamma^{\alpha}_{\text{m},\varphi \to \varphi} q^{\text{B}}_{\varphi k \varphi k'}}{n_{\varphi \varphi}} \frac{\sum_{j'} (\hat{a}_{\varphi j \to \varphi j'} \cdot \hat{\xi})}{\mathbf{a}_{\varphi k \varphi k'} \cdot \boldsymbol{\xi}}, \tag{2.35}$$

such that $\hat{a}_{\varphi j \to \varphi j'} \cdot \hat{\xi} > 0, \forall j'$ and $\hat{a}_{\varphi k \varphi k'} \cdot \hat{\xi} > 0$ (only microscopic diffusion hops and CG cell hops, respectively, with a positive $\hat{\xi}$ component are considered), and

$$\hat{\Gamma}^{\alpha}_{\text{mig},\varphi k k'} = 0, \text{ if } \hat{a}_{\varphi k \varphi k'} \cdot \hat{\xi} = 0. \tag{2.36}$$

The microscopic hopping transition probability rate is recovered from Eq. (2.35) in the limit of no coarse-graining.

For the case of hopping between different site types, φ, φ', $\varphi' \neq \varphi$, $k' \neq k$, Eq. (2.35) is adapted to give

$$\hat{\Gamma}^{\alpha}_{\text{mig},\varphi k \varphi' k'} = \frac{\Gamma^{\alpha}_{\text{m},\varphi \to \varphi'} q^{\text{B}}_{\varphi k \varphi' k'}}{n_{\varphi \varphi'}} \frac{\sum_{j'} (\hat{a}_{\varphi j \to \varphi' j'} \cdot \hat{\xi})}{\mathbf{a}_{\varphi k \varphi' k'} \cdot \boldsymbol{\xi}} \tag{2.37}$$

with the same constraints applying as in Eq. (2.35). Eq. (2.37) assumes that the microscopic transition probability rate for diffusion from φ to φ' $\left(\hat{\Gamma}^{\alpha}_{\text{m},\varphi \to \varphi'}\right)$ is representative of the average microscopic transition probability rate for diffusion within each cell. A summary of transition probability rates for CGMC is given in Table 2.2.

Table 2.2 – Summary of transition probability rates for CGMC (without homogenization).

Process	Coarse transition probability rate $\left(\overline{\Gamma}(\eta, \varphi, k, \varphi', k')\right)$	Coarse rate constant $\left(\Gamma^{\%}\right)$
Adsorption $\left(\alpha + \phi \cdot \varphi \rightarrow \alpha \cdot \varphi\right)$	$\overline{\Gamma}^{\alpha}_{ads} = \Gamma^{\%}_{ads,\varphi k} \dfrac{\eta^{\phi}_{\varphi k}}{q_{\varphi k}}$	$\Gamma^{\%}_{ads,\varphi k} = \sum_{j \in C_k} \dfrac{\Gamma^{\alpha}_{ads,\varphi j}}{q_{\varphi k}}$
Desorption $\left(\alpha \cdot \varphi \rightarrow \alpha + \phi \cdot \varphi\right)$	$\overline{\Gamma}^{\alpha}_{des} = \Gamma^{\%}_{des,\varphi k} \dfrac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}} e^{-\beta \overline{U}^{\alpha}_{\varphi k}}$	$\Gamma^{\%}_{des,\varphi k} = \sum_{j \in C_k} \dfrac{\Gamma^{\alpha}_{des,\varphi j}}{q_{\varphi k}}$
Isomerization Reaction $\left(\alpha \cdot \varphi \rightarrow \alpha' \cdot \varphi\right)$	$\overline{\Gamma}^{\alpha}_{iso} = \Gamma^{\%}_{iso,\varphi k}{}^{\alpha \rightarrow \alpha'} \dfrac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}}$	$\Gamma^{\%}_{iso,\varphi k}{}^{\alpha \rightarrow \alpha'} = \sum_{j \in C_k} \dfrac{\Gamma^{\alpha \rightarrow \alpha'}_{iso,\varphi j}}{q_{\varphi k}}$
Surface Hopping (Diffusion) $\left(\alpha \cdot \varphi + \phi \cdot \varphi' \rightarrow \phi \cdot \varphi + \alpha \cdot \varphi'\right)$		
$k \neq k'$	$\overline{\Gamma}^{\alpha}_{mig} = \Gamma^{\%}_{m,\varphi k \rightarrow \varphi' k'} \dfrac{\eta^{\alpha}_{\varphi k} \eta^{\phi}_{\varphi' k}}{q_{\varphi k} q_{\varphi' k}} e^{-\beta \overline{U}^{\alpha}_{k}}$	$\Gamma^{\%}_{m,\varphi k \rightarrow \varphi' k'} = \dfrac{\Gamma^{\alpha}_{m,\varphi k \rightarrow \varphi' k'} q^{B}_{\varphi k \varphi' k'}}{n_{\varphi \varphi'}} \dfrac{\sum_{j'} (\vec{a}_{\varphi j \rightarrow \varphi j'} \cdot \vec{\xi})^{*}}{\vec{a}_{\varphi k \varphi k'} \cdot \vec{\xi}}$
Disproportionation reaction $\left(\alpha \cdot \varphi + \alpha' \cdot \varphi' \rightarrow \alpha'' \cdot \varphi + \alpha''' \cdot \varphi'\right)$		
$k = k', \varphi = \varphi', \alpha = \alpha'$	$\overline{\Gamma}^{\alpha\alpha' \rightarrow \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} = \Gamma^{\%}_{rxn,\varphi k \varphi' k'}{}^{\alpha' \rightarrow \alpha'' \alpha'''} \dfrac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}} \dfrac{\left(\eta^{\alpha}_{\varphi k} - 1\right)}{\left(q_{\varphi k} - 1\right)}$	
$k = k', \varphi = \varphi', \alpha \neq \alpha'$	$\overline{\Gamma}^{\alpha\alpha' \rightarrow \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} = \Gamma^{\%}_{rxn,\varphi k \varphi k}{}^{\alpha' \rightarrow \alpha'' \alpha'''} \dfrac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}} \dfrac{\eta^{\alpha'}_{\varphi k}}{\left(q_{\varphi k} - 1\right)}$	$\Gamma^{\%}_{rxn,\varphi k \varphi' k'}{}^{\alpha' \rightarrow \alpha'' \alpha'''} = \dfrac{\Gamma^{\alpha\alpha' \rightarrow \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'}}{n_{\varphi \varphi'}} \left(n_{\varphi \varphi'} q_{\varphi k} - \sum_{k' \neq k} q^{B}_{\varphi k, \varphi' k'} \right)$
$k = k', \varphi \neq \varphi'$	$\overline{\Gamma}^{\alpha\alpha' \rightarrow \alpha'' \alpha'''}_{rxn,\varphi k \varphi' k'} = \Gamma^{\%}_{rxn,\varphi k \varphi' k'}{}^{\alpha' \rightarrow \alpha'' \alpha'''} \dfrac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}} \dfrac{\eta^{\alpha'}_{\varphi' k}}{q_{\varphi' k}}$	

$k \neq k'$	$\overline{\Gamma}^{\alpha\alpha' \to \alpha''\alpha'''}_{rxn,\varphi k\varphi' k'} = \Gamma^{\alpha\alpha' \to \alpha''\alpha'''}_{rxn,\varphi k\varphi' k'} \frac{\eta^{\alpha}_{\varphi k}}{q_{\varphi k}} \frac{\eta^{\alpha'}_{\varphi' k'}}{q_{\varphi' k'}}$	$\Gamma^{\alpha\alpha' \to \alpha''\alpha'''}_{rxn,\varphi k\varphi' k'} = \frac{\Gamma^{\alpha\alpha' \to \alpha''\alpha'''}_{rxn,\varphi k\varphi' k'}}{n_{\varphi\varphi'}} q^{B}_{\varphi k,\varphi' k'}$

* - Subject to the constraints of Eq. (2.35)

Monte Carlo Algorithm

A flowchart of the CGMC algorithm is shown in Figure 2.1. Initially, a microscopic lattice (no coarse-graining) is constructed with user-specified site types and initial adparticle placement. Spatially CG cells are then constructed given the number of microscopic sites they contain. The LMF is applied to all sites within the same CG cell. To execute a fully microscopic simulation (traditional KMC), each cell is given a size of one (q = 1).

Once the attributes of each cell are established, the number of adparticles is recorded in the main observable of the simulation, $\eta_{\varphi k}^{\alpha}$ (the concentration of adparticle type α on site type ϕ in the kth CG cell). Using the $\eta_{\varphi k}^{\alpha}$ values, the interaction potentials and all transition probability rates are calculated and cataloged into the arrays U and Γ, respectively.

As happens with a microscopic KMC simulation[14], different algorithms can be used to implement the CGMC method. Herein, a null-event selection process is used. At the beginning of each simulation loop, a CG cell, C_k, is randomly selected. Another random number is generated and multiplied by the normalization rate, $\Gamma_{\max}$, ensuring that all probabilities are less than 1. The normalization rate is the maximum expected value for the sum of all transition probability rates associated with all CG cells. The product of $\Gamma_{\max}$ and the random number is used to select a process corresponding to C_k from **Γ**, otherwise, a null event occurs.

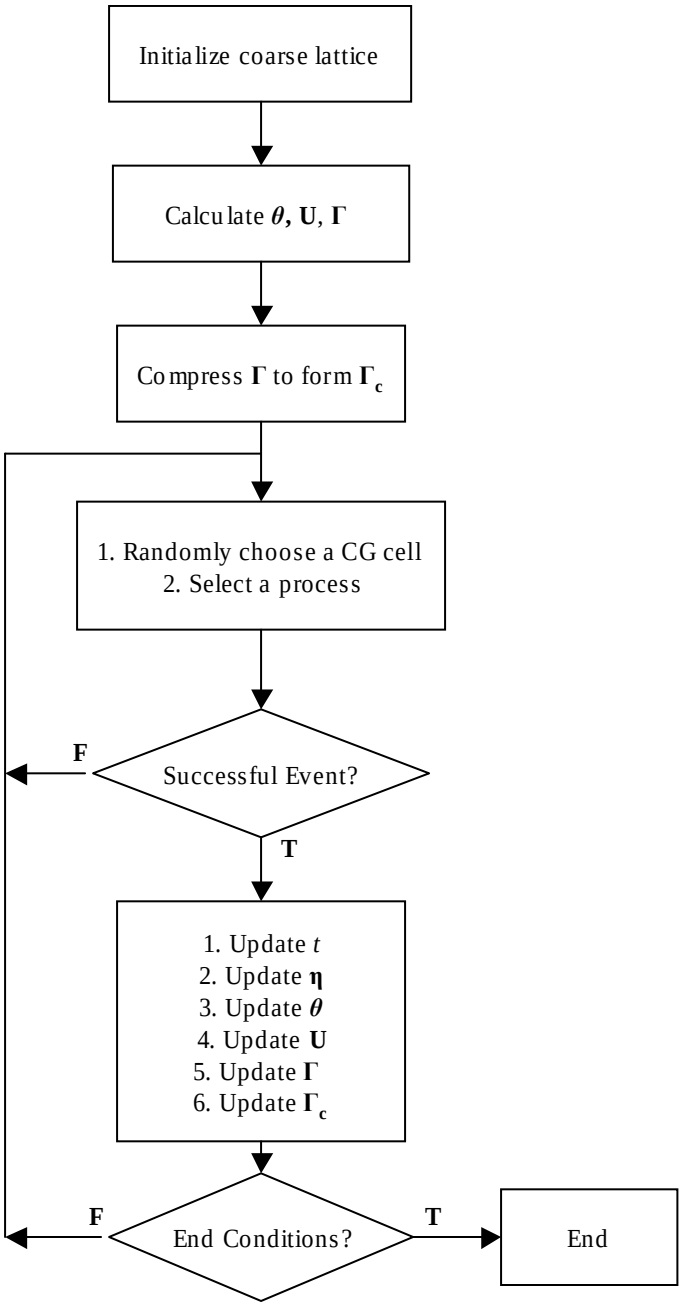

Figure 2.1 - Flowchart of null-event CGMC algorithm. The cell coverage θ and cell coarse-grained transition rate Γ_c will only be used if CG cell homogenization is applied.

Upon a successful step, the simulation's average real time is updated using $\Delta t = 1/P$ where *P* represents the sum of all transition probability rates. The $\eta_{\varphi k}^{\alpha}$ values are then updated. Only affected interaction potentials and transition probability rates are recalculated (a local update algorithm) using the new $\eta_{\varphi k}^{\alpha}$ values and cataloged in **U** and **Γ**, respectively.

Numerical Examples

Example A: Adsorption/Desorption on a (100) Surface with Multiple Site Types

To demonstrate the CGMC method with multiple types of microscopic sites and adparticle-adparticle interactions, the following example is adapted from Ref.[14]. A molecular species, ***A***, adsorbs on either a top (ϕ_{T}) or a bridge (ϕ_{B}) site of a (100) crystal plane (a schematic of the lattice is shown in Figure 2.2a), forming surface species $\mathrm{A}\cdot\phi_{\mathrm{T}}$ or $\mathrm{A}\cdot\phi_{\mathrm{B}}$, respectively. Species A from both site types are allowed to desorb. The various processes are summarized in Table 2.3. A microscopic lattice of size 40*a* by 40*a* is used, where *a* is the lattice constant, representing 1600 top (T) sites and 3200 bridge (B) sites. Initially the lattice is empty. Periodic boundary conditions are implemented.

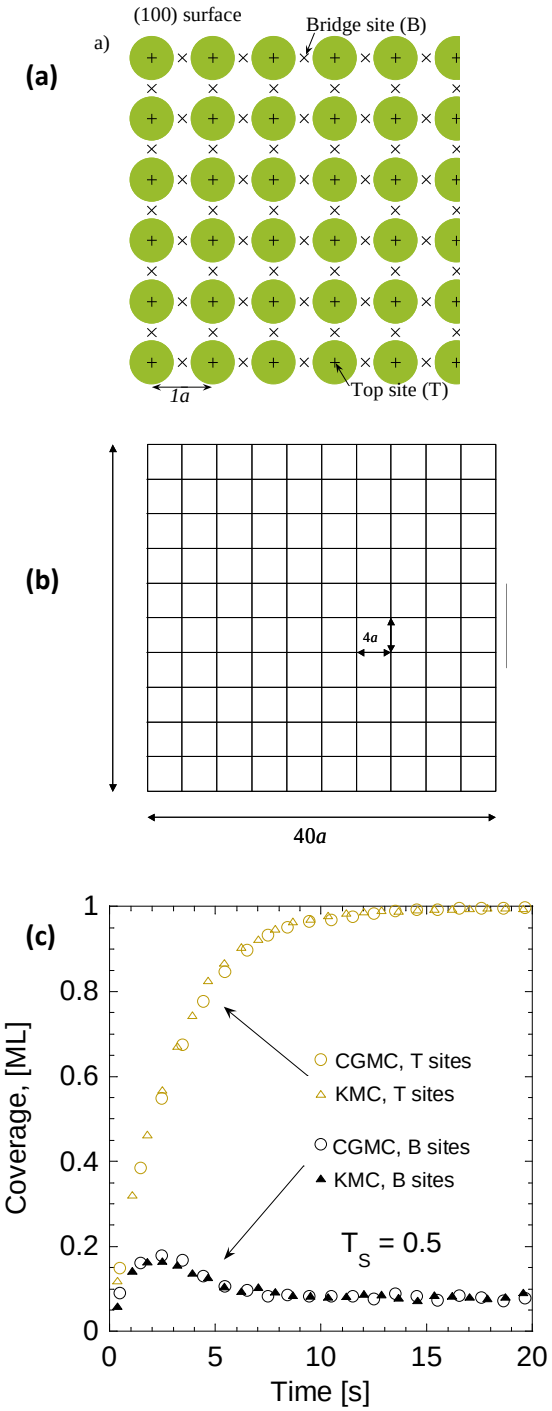

Figure 2.2 (a) Schematic of Top (T) and Bridge (B) sites of a (100) surface (Reprinted with permission from Ref. [14]. © 2007, American Institute of Physics). (b) Layout of CG cells over the microscopic

lattice in Example A. CG cells are uniform squares with a side length of 4 lattice constants. Periodic boundary conditions are used. (c) Coverage evolution against simulated time.

Microscopic Description

The microscopic transition probability rates for adsorption are:

$$\Gamma_{ads}(\sigma, T, j) = \Gamma_{ads,T}(1-\sigma_{T,j}) \quad , \tag{2.38}$$

$$\Gamma_{ads}(\sigma, B, j) = \Gamma_{ads,B}(1-\sigma_{B,j}) \quad , \tag{2.39}$$

where $\sigma_{T,j}$ and $\sigma_{B,j}$ denote the occupancy of the j[th] top or bridge site, respectively. The presence of a vacancy is denoted using the exclusion principle $(1-\sigma_{T,j})$. The transition probability rate prefactors for adsorption are $\Gamma_{a,T} = 0.4\ s^{-1}$ and $\Gamma_{a,B} = 0.2\ s^{-1}$.

The transition probability rates for desorption are

$$\Gamma_{des}(\sigma, T, j) = \Gamma_{des,T}\sigma_{T,j}e^{-\beta U_{Tj}} \quad , \tag{2.40}$$

$$\Gamma_{des}(\sigma, B, j) = \Gamma_{des,B}\sigma_{B,j}e^{-\beta U_{Tj}} \quad . \tag{2.41}$$

Competitive attractive and repulsive adparticle interactions can introduce complex phase behavior. Following Ref [14] with ω = 2, the top-top (TT), top-bridge (TB), and

bridge-bridge (BB) adparticle-adparticle interactions are given by $J_{TT}(r)/k_BT_s$ = 0.0605, $J_{TB}(r)/k_BT_s = J_{BT}(r)/k_BT_s$ = -0.0155, $J_{BB}(r)/k_BT_s$ = 0.00415, respectively, when $0 < r \leq 3$ and zero otherwise (fairly short interactions). Here r is the dimensionless distance between two interacting adsorbed particles (in terms of lattice constants). Each top site has 28 TT and 52 TB neighbor interactions, whereas each bridge site has 26 BT and 60 BB neighbor interactions. The difference in the number of neighbors is due to bridge sites having twice the density of the top sites.

Table 2.3 - Summary of processes and transition probability rates for Example A.

Process		Microscopic (KMC) - Γ	Coarse Grained - $\overline{\Gamma}$
Top site adsorption:	$A + \phi_T \rightarrow A \cdot \phi_T$	$\Gamma_{a,T}(1-\sigma_{T,j})$	$\overline{\Gamma}_{a,Tk}\eta^{\phi}_{Tk}$
Bridge site adsorption:	$A + \phi_B \rightarrow A \cdot \phi_B$	$\Gamma_{a,B}(1-\sigma_{B,j})$	$\overline{\Gamma}_{a,Bk}\eta^{\phi}_{Bk}$
Top site desorption:	$A \cdot \phi_T \rightarrow A + \phi_T$	$\Gamma_{d,T}\sigma_{T,j}e^{-\beta U_{Tj}}$	$\overline{\Gamma}_{d,Tk}\eta^{A}_{Tk}e^{-\beta \overline{U}^{A}_{Tk}}$
Bridge site desorption:	$A \cdot \phi_B \rightarrow A + \phi_B$	$\Gamma_{d,B}\sigma_{B,j}e^{-\beta U_{Bj}}$	$\overline{\Gamma}_{d,Bk}\eta^{A}_{Bk}e^{-\beta \overline{U}^{A}_{Bk}}$

CGMC Description

A schematic of the CG lattice is shown in Figure 2.2b. Uniform CG cells of size 4a by 4a (16 T sites, 32 B sites) are used. Adparticles on T (B) sites are considered to be spatially well-mixed over all T (B) sites in the same CG cell. The coverage on T sites and B sites within a CG cell are tracked independently.

The CGMC equations describing the adsorption and desorption transition probability rates are given by Eqs. (2.12) and (2.14). For this example, they become

$$\overline{\Gamma}^{A}_{ads}(\boldsymbol{\eta}, T, k) = \overline{\Gamma}^{0}_{ads,Tj}\eta^{\phi}_{Tk} \quad , \tag{2.42}$$

$$\overline{\Gamma}^{A}_{ads}(\boldsymbol{\eta}, B, k) = \overline{\Gamma}^{0}_{ads,Bj}\eta^{\phi}_{Bk} \quad , \tag{2.43}$$

$$\overline{\Gamma}^{A}_{des}(\boldsymbol{\eta}, T, k) = \overline{\Gamma}^{0}_{des,Tk}\eta^{A}_{Tk}e^{-\beta\overline{U}^{A}_{Tk}} \quad , \tag{2.44}$$

$$\overline{\Gamma}^{A}_{des}(\boldsymbol{\eta}, B, k) = \overline{\Gamma}^{0}_{des,Bk}\eta^{A}_{Bk}e^{-\beta\overline{U}^{A}_{Bk}} \quad , \tag{2.45}$$

corresponding to the four processes (see also Table 2.3). Similarly, the coarse-grained interaction potentials are obtained from Eqs. (2.4)-(2.10).

Figure 2.2c compares the KMC and CGMC simulation results. The agreement is very good for all trajectories. The small discrepancy between KMC and CGMC is attributed to small clusters seen in KMC (see Ref. [14]). The LMF assumption in CG cells does not resolve these small clusters well, as small-scale spatial detail is lost as the coarse-grained cell size increases. Despite this, the CGMC method performs very well as a consequence of the relatively large number of interacting neighbors that render the coarse-graining fairly accurate.

Example B: Diffusion and Reaction of Multiple Species on a Single Crystal

Diffusion and reaction of multiple surface species is demonstrated in the presence of spatial gradients and adparticle-adparticle interactions with the following example adapted from Ref. [14]. On a (100) crystal plane, adparticles are allowed to reside on hollow sites, creating a lattice with a square pitch and one available site type. The lattice is 50*a* x 50*a* in the x and y directions (2500 lattice nodes). Periodic boundary conditions are employed. Two adparticle species, A and B, initially reside in stripes on the surface. The probability of finding species A at a spatial location (x,y) at time t=0 is given by

$$p(x, y, t = 0) = \frac{1}{2}\left[\tanh\left(\frac{x - 15}{2} \right) - \tanh\left(\frac{x - 35}{2} \right) \right], \quad (2.46)$$

where x is the dimensionless position in the x dimension with respect to the hop distance (*a*). Species B initially occupies all sites not occupied with species A. Note that there are no gradients in the y direction.

Each adparticle is allowed to diffuse to an adjacent vacancy according to the respective transition probability rate, which is a function of adparticle-adparticle interactions. Additionally, an A adparticle may react with an adjacent B adparticle, followed by immediate desorption of the product, leaving behind two vacant sites. The processes are summarized in Table 2.4.

No vacancies are initially available. As the simulation proceeds, a local population of vacancies at the interface of the two species develops, allowing reactants to

diffuse to the interface. This leads to moving reaction fronts as areas predominantly occupied by a single species begin to diffuse to and react with the other species.

Table 2.4 - Summary of processes and transition probability rates for Example B.

Process		Microscopic (KMC) - Γ	Coarse Grained - $\overline{\Gamma}$
Species A diffusion:	$A\cdot\phi+\phi\rightarrow\phi+A\cdot\phi$	$\frac{\Gamma_m^A}{n}\sigma_j^A\sigma_{j'}^{\phi}e^{-\beta U_j}$	$\Gamma_{m,k\rightarrow k'}^{0A}\frac{\eta_k^A\eta_{k'}^{\phi}}{q_k q_{k'}}e^{-\beta\overline{U}_k^A}$
Species B diffusion:	$B\cdot\phi+\phi\rightarrow\phi+B\cdot\phi$	$\frac{\Gamma_m^B}{n}\sigma_j^B\sigma_{j'}^{\phi}e^{-\beta U_j}$	$\Gamma_{m,k\rightarrow k'}^{0B}\frac{\eta_k^B\eta_{k'}^{\phi}}{q_k q_{k'}}e^{-\beta\overline{U}_k^B}$
Reaction of A and B:	$A\cdot\phi+B\cdot\phi\rightarrow AB+2\phi$	$\frac{\Gamma_{rxn}^{AB}}{n}\sigma_j^A\sigma_{j'}^B$	$\Gamma_{rxn,k,k'}^{0AB}\eta_k^A\eta_{k'}^B$

Microscopic Description

Adparticles hop in the x and y directions according to the microscopic rate:

$$\Gamma(\sigma,j,j')=\frac{\Gamma_{mig}^{\alpha}}{n}\sigma_j(1-\sigma_{j'})e^{-\beta U_j},\quad \sigma_j=\begin{Bmatrix}1\,if\ \ site\ \ j\ occupied\\ 0\,otherwise\end{Bmatrix} \tag{2.47}$$

n represents the number of possible hop directions from site *j*, and $\hat{\Gamma}_m^{\alpha}$ represents the hopping frequency of an adparticle on site *j*. For our example on the (100) surface, *n = 4*, reflecting that each site has 4 neighbors. The hopping frequencies of A and B are taken to be $\Gamma_{mig}^{A}=\Gamma_{mig}^{B}=4$. The adparticle-adparticle interaction potentials defined for AA, AB, BA, and BB pairs are: $J^{AA}(r)=0.0214k_BT_s$,

$J^{AB}(r) = J^{BA}(r) = J^{BB}(r) = 0.0107 k_B T_s$, when $0 < r \leq 3$ and zero otherwise. Microscopic interactions for the j[th] site are calculated with each occupied site according to:

$$U_j^{\alpha} = \sum_{\mathrm{r}} \sum_{\alpha'} J^{\alpha\alpha'}(\mathrm{r}) \sigma_{j'}^{\alpha'} \tag{2.48}$$

An *A* adparticle reacts with a neighboring *B* adparticle with a transition probability rate of

$$\Gamma_{rxn}^{AB}(\sigma, j, j') = \frac{\Gamma_{rxn}^{AB}}{n} \sigma_j^A \sigma_{j'}^B, \tag{2.49}$$

where the reaction frequency is $\Gamma_{rxn}^{AB} = 0.1$.

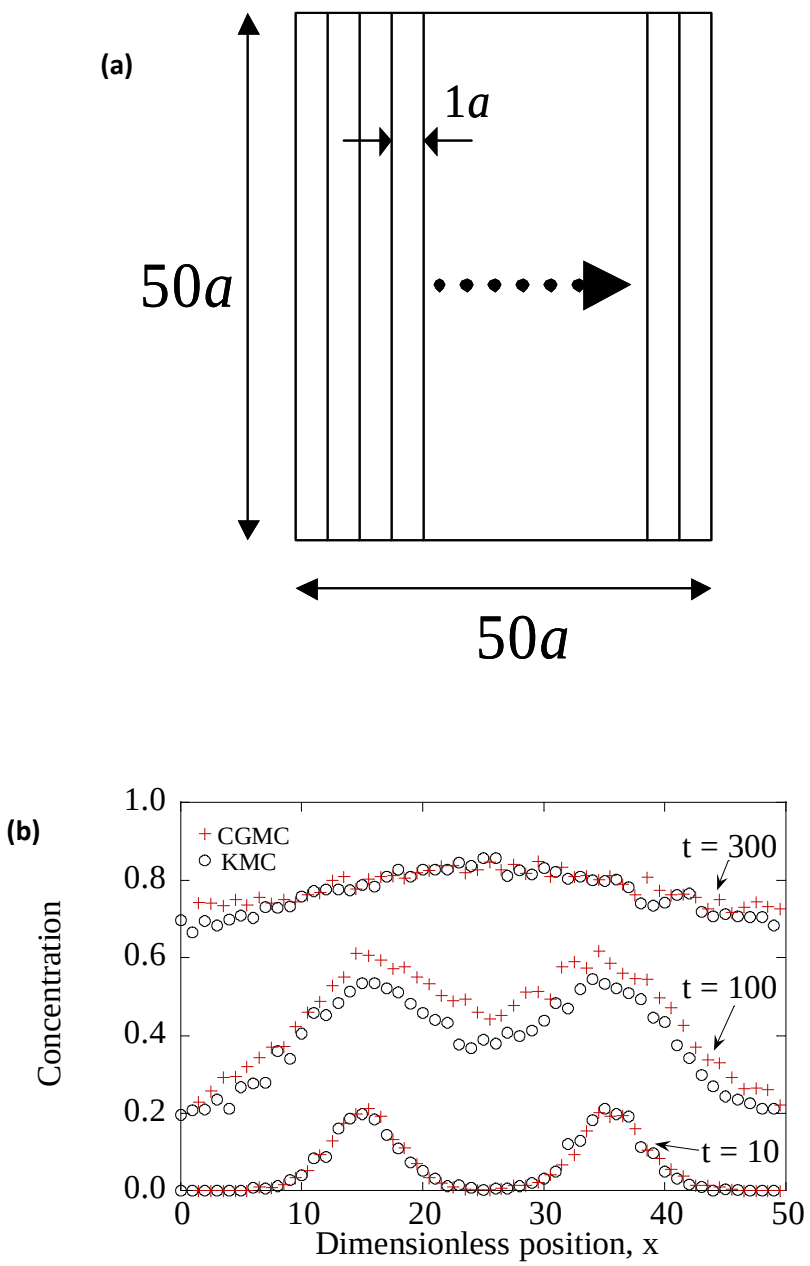

Figure 2.3 (a) Schematic of the CG mesh. Each CG cell encompasses 1x50 microscopic sites. Periodic boundary conditions are used. (b) Coverage of vacancies at increasing times (t=10, t=100, t=300). CGMC is compared against traditional KMC. The slight difference is attributed to the well-mixed assumption along the y-axis.

CGMC Description

For this example, the microscopic lattice is divided into 50 uniform CG cells. Each CG cell is 50*a* by 1*a*, effectively creating a 1D simulation, shown in Figure 2.3a. The reaction/desorption transition probability rate, Eq. (2.49), is spatially coarse grained using Eqs. (2.20) and (2.25) to

$$\overline{\Gamma}^{A,B}_{rxn}(\boldsymbol{\eta},k,k') = \hat{\Gamma}^{AB}_{rxn,k,k'}\eta^{A}_{k}\eta^{B}_{k'} \tag{2.50}$$

$$\hat{\Gamma}^{AB}_{rxn,k,k} = \frac{\Gamma^{AB}_{rxn}}{n}\left(\frac{nq_k - \sum_{k'} q^{Boundary}_{k,k'}}{q_k(q_k-1)}\right) \tag{2.51}$$

$$\hat{\Gamma}^{AB}_{rxn,k,k'} = \frac{\Gamma^{AB}_{rxn}}{n}\left(\frac{q^{Boundary}_{k,k'}}{q_k q_{k'}}\right) \quad k \neq k' \tag{2.52}$$

, .

Similar to the microscopic equations, $n = 4$. All cells have 50 microscopic sites ($q_k = 50$) and 50 site pairs along each boundary ($q^{Boundary}_{k,k'} = 50, k' \text{ neighbor of } k$).

The diffusion transition probability rate, Eq. (2.47), is spatially coarse grained using Eqs. (2.28) and (2.35) to

$$\overline{\Gamma}^{\alpha}_{mig}(\boldsymbol{\eta},k,k') = \hat{\Gamma}^{\alpha}_{mig,k\to k'}\frac{\eta^{\alpha}_{k}\eta^{\phi}_{k'}}{q_k q_{k'}}e^{-\beta\overline{U}^{\alpha}_{k}} \text{, with} \tag{2.53}$$

$$\hat{\Gamma}^{\alpha}_{mig,\varphi kk'} = \frac{\Gamma^{\alpha}_{mig}q^{Boundary}_{kk'}}{n}, \tag{2.54}$$

$$\text{since } \frac{\sum_{j'} (\vec{a}_{j\to j'} \cdot \vec{\xi})}{\vec{a}_{kk'} \cdot \vec{\xi}} = 1.$$

As in Example A, the CG interaction energy is calculated using Eqs. (2.4)-(2.10).

Figure 2.3b shows that the agreement between CGMC and KMC simulations is reasonably good. CGMC successfully predicts the diffusion and reaction of multiple species in spatial gradients. Deviations observed between the KMC and CGMC simulations are attributed to the LMF assumption in the CGMC simulation that in the case of nonlinear chemical reactions is not as accurate, as found also in previous work [7].

Homogenization over Multiple Site Types

Homogenization Concept and Coarse-Graining

As noted for the surface disproportionation reaction, Eq. (2.27), relatively few molecular species can combine in many ways as reactants or products on available site types, creating a combinatorial explosion. The term 'combinatorial complexity' refers to the proliferation of transition probability rates that must be tracked and searched through when many microscopic reactants, products, and host site types must be considered. Combinatorial explosion can lead to large memory and CPU requirements. In deterministic modeling, combinatorial explosion in the numbers of species or conformations amounts to writing a huge number of differential equations that have to be solved. At the partial differential equation (PDE) (macroscopic) scale, variation of a parameter, such as the diffusivity, over nanoscopic length scales imposes a computational challenge due to

separation of length scales, i.e., one needs nanoscopic resolution to simulate macroscopic length scales, obviously an impossible task. This latter problem has been solved at the PDE level using the now well-established homogenization theory [15-17], whereby an effective PDE is written with a CG diffusivity that depends on the microscopic variation of diffusivity over a unit cell. The resulting CG PDE 'blurs' microscopic details and provides the effective concentration field and flux, which are of practical interest. Herein we introduce such a method at the *stochastic level* that does not currently exist for CGMC or KMC simulation.

A method to compress this explosion of data involves homogenization within the CG cells such that the coverage of adparticles on different sites is no longer kept independently, allowing use of fewer coarse observables. In essence, this idea allows additional coarse-graining whereby we lump appropriate transition probability rates together (*coarsening the number of processes*). This is made possible by the application of the LMF over all site types within a CG cell (homogenization). By contrast, in Example A the LMF approximation was applied to site types T and B independently. The idea of lumping has previously been used in a different context, i.e., in classic chemical kinetics networks[18] and also for grouping reversible processes together to remove stiffness (the net-event MC method) [19, 20].

In the homogenization framework, the main coarse observable becomes η_k^α, the number of α adparticles in cell C_k:

$$\eta_k^\alpha = \sum_{\varphi}^{N_{sttype}} \eta_{\varphi k}^\alpha = \sum_{\varphi}^{N_{sttype}} \sum_{j \in C_k} \sigma_{\varphi j}^\alpha \quad . \tag{2.55}$$

$$\boldsymbol{\eta} = \{\eta_k^\alpha\}_{k=1,\ldots,m;\alpha=1,\ldots,N_{sp}+1} \quad . \tag{2.56}$$

All CG transition probability rates for events with identical adparticle stoichiometry, $\boldsymbol{\nu}$ (the reactants and products of a process without consideration of the underlying lattice site types), and participating CG cells, $\mathrm{C_k}$ and $\mathrm{C_{k'}}$, are lumped into a coarse transition probability rate:

$$\Gamma_{\mathbf{C}\,k,k'}^{\boldsymbol{\nu}} = \sum_{\varphi}^{N_{st\,type}} \sum_{\varphi'}^{N_{st\,type}} \overline{\Gamma}_{\varphi k,\varphi' k'}^{\boldsymbol{\nu}}, \forall k, \boldsymbol{\nu}. \tag{2.57}$$

For example, all diffusion processes of α adparticles from cell $\mathrm{C_k}$ to cell $\mathrm{C_{k'}}$ are lumped together *regardless of site type* (see Figure 2.4c). Lumping transforms $\boldsymbol{\Gamma}$ into the smaller, coarsened array $\boldsymbol{\Gamma_c}$. The algorithm then selects coarse processes from $\boldsymbol{\Gamma_c}$. The smaller size of $\boldsymbol{\Gamma_c}$ reduces memory requirements and provides CPU savings when searching through the transition probability rates to selecting a process.

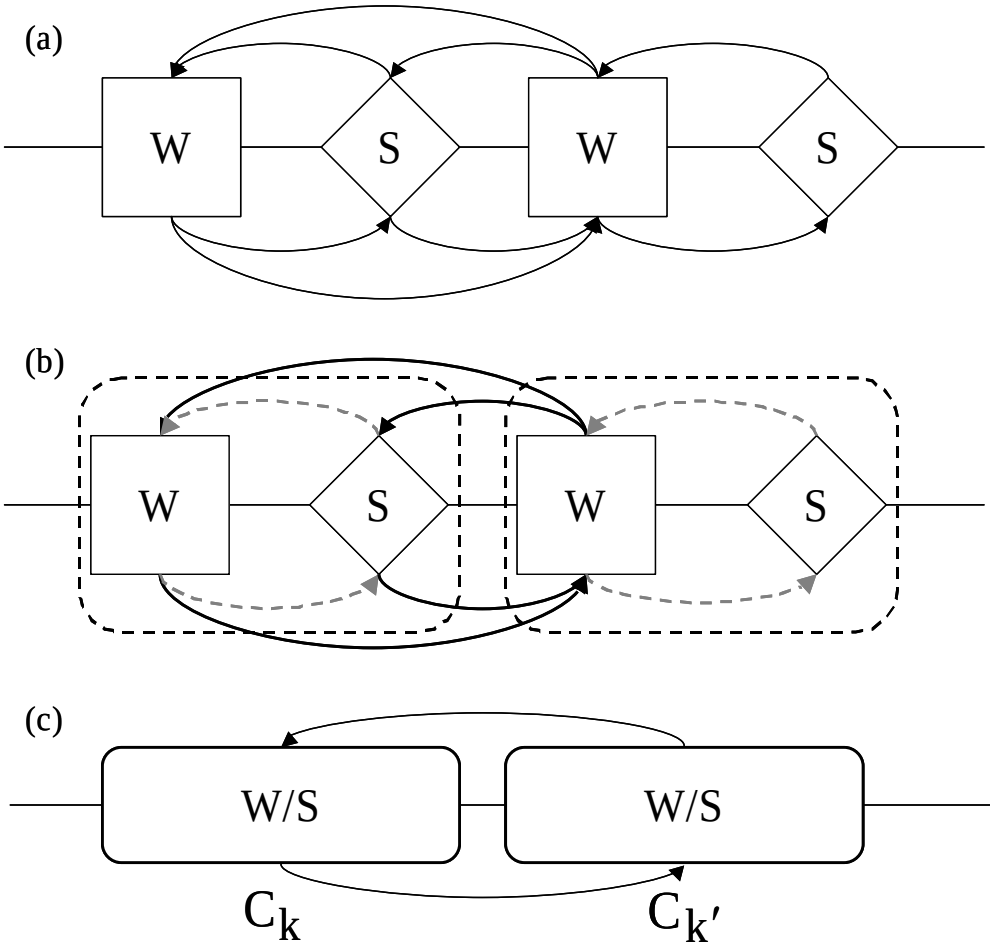

Figure 2.4 Schematic of 1D diffusion through window (W) and intercage (S) sites. Traditional KMC will store and simulate all potential events shown in (a). The CGMC method with q = 2 assumes neighboring W and S sites are in equilibrium with each other (b). The two diffusion rates between neighbors are lumped into one coarse process via homogenization (c).

Reconstruction

Since adparticle interactions and transition probability rates are usually functions of the occupancy on a particular site type, the coverage on each site type must be found to determine the corresponding transition probability rates $\boldsymbol{\Gamma}_c$ from the updated $\boldsymbol{\eta}$. Due to variation of coarse observables among cells, the coverages of adparticles on each site

type must then be solved in principle for all CG cells and at each time step. In essence, this task translates into *a reconstruction process*: given the coarse variable η_k^α, one needs to find $\theta_{\varphi k}^\alpha$ (the average coverage of α adparticles on φ site types in cell C_k) for all sites φ. The LMF approximation, inherent to the coarse-graining, can provide exactly this reconstruction, i.e., the partitioning of each species on all sites subject to the constraint of a fixed η_k^α. The main observable, $\boldsymbol{\eta}$, translates into

$$\boldsymbol{\theta} = \{\theta_{\varphi k}^\alpha\}_{\varphi=1,\ldots,N_{sttype},k=1,\ldots,m,\alpha=1,\ldots,N_{sp}+1}. \tag{2.58}$$

Finding $\boldsymbol{\theta}$ from $\boldsymbol{\eta}$ can be accomplished via a deterministic or a stochastic reconstruction approach. In the former method, one uses a nonlinear solver, such as the Newton's method, to compute the coverages on all site types given the constraint η_k^α and the local equilibrium (LMF) assumption within a coarse cell. Obviously, the resulting coarse variables $\theta_{\varphi k}^\alpha$ are non-integers. In the latter method, one can use short KMC simulations over small lattices (with respect to the CG cells) to compute the time-averaged equilibrium partitioning of each species on the various sites. This is a constrained KMC simulation at fixed η_k^α, i.e., a canonical ensemble simulation. The results from these KMC simulations can be tabulated and used for all cells and at all times. Alternatively, one can perform a KMC simulation in each CG cell on-the-fly, especially if cataloging each state is unreasonable. For example, if long-range spatial interactions are relevant the occupancy of nearby CG cells could be incorporated into each reconstruction process. This reconstruction idea was formally introduced in Ref. [13] (two grid approach) to improve accuracy of the CGMC method.

Since the concentration of adparticles on individual sites is now described using a real-type variable, $\boldsymbol{\theta}$ (no longer $\boldsymbol{\eta}$, an integer), the task of preventing adparticles from energetically interacting or reacting with themselves becomes more complex. Previously, this was accomplished using the exclusion principle (see Eqs. 2.9, 2.22, 2.23). However, if this same method is applied to $\theta_{\varphi k}^{\alpha}$, negative transition probability rates could potentially result. As an example, consider the adaptation of Eq. 23 by replacing $\eta_{\varphi k}^{\alpha}$ with $\theta_{\varphi k}^{\alpha} q_{\varphi k}$:

$$\overline{\Gamma}_{rxn}^{\alpha',\alpha'''}(\boldsymbol{\eta},\varphi,k,\varphi,k') = \frac{1}{2}\Gamma_{rxn,\varphi\varphi}^{\alpha\alpha\to\alpha''\alpha'''}\theta_{\varphi k}^{\alpha} q_{\varphi k}(\theta_{\varphi k}^{\alpha} q_{\varphi k} - 1) \quad . \tag{2.59}$$

If $0 < \theta_{\varphi k}^{\alpha} q_{\varphi k} < 1$, the resulting transition probability rate would be negative! This is clearly unacceptable. This in fact is a general problem with using the deterministic LMF approximation [21], which was also addressed in previous work in a different context (in removing stiffness from KMC simulations) discussed in Ref. [22].

To address this issue, when a transition probability rate of a multi-site reaction within a CG cell is calculated, the concentration of reactants must be carefully considered. The concentration of one reactant must be estimated with the other reactant's presence as a constraint. For instance, consider an arbitrary cell C_k with $\eta_k^{\alpha} = 2$ and $q_k = q_{\varphi k} = 4$ (N_{sttype} = 1). The probability of finding a reactant for Eq. 59 on any site is $p = \theta_{\varphi k}^{\alpha} = \frac{\eta_k^{\alpha}}{q_{\varphi k}} = \frac{2}{4}$. The probability of picking a site with a valid second reactant is now $p = \theta_{\varphi k,\alpha\varphi}^{\alpha} = \frac{\eta_k^{\alpha} - 1}{q_{\varphi k} - 1} = \frac{1}{3}$, assuming LMF and *knowing that one α adparticle already occupies a φ site* (a conditional probability). Therefore, Eq. 59 may be re-written as:

$$\bar{\Gamma}_{rxn}^{\alpha'',\alpha'''}(\boldsymbol{\eta},\varphi,k,\varphi,k') = \frac{1}{2}\bar{\Gamma}_{rxn,\varphi\varphi}^{\alpha\alpha\to\alpha''\alpha'''}\theta_{\varphi k}^{\alpha} q_{\varphi k}\cdot\theta_{\varphi k,\alpha\varphi}^{\alpha}(q_{\varphi k}-1)$$

. (2.60)

Here $\theta^{\alpha}_{\varphi k,\alpha'\varphi'}$ is the average coverage of α adparticles on type φ sites within CG cell C_k, given the constraint that one α' adparticle is already on a type φ' site.

Using the reconstruction approach, transition probability rates as functions of coverages for adsorption, desorption and isomerization are adapted from Eqs. (2.11)-(2.16) for surface disproportionation, from Eqs. (2.18)-(2.27), and for surface diffusion, from Eq. (2.28). The correction noted in Eq. (2.60) is included where appropriate. The two-body interactions presented in Eqs. (2.4)-(2.6) remain the same, while Eqs. (2.9) and (2.10) become

$$\bar{U}^{\alpha}_{\varphi k,k} = \bar{J}^{\alpha\alpha}_{\varphi k,\varphi k}\cdot\theta^{\alpha}_{\varphi k,\alpha\varphi}\cdot\left(q_{\varphi k}-1\right)+\sum_{\substack{\alpha'\\ \alpha'\neq\alpha}}\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi k}\cdot\theta^{\alpha'}_{\varphi k}\cdot q_{\varphi k}+\sum_{\alpha'}\sum_{\substack{\varphi'\\ \varphi'\neq\varphi}}\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k}\cdot\theta^{\alpha'}_{\varphi' k}\cdot q_{\varphi' k} \quad (2.61)$$

$$\bar{U}^{\alpha}_{\varphi k,k'} = \sum_{\varphi'}\sum_{\alpha'}\bar{J}^{\alpha\alpha'}_{\varphi k,\varphi' k'}\theta^{\alpha'}_{\varphi' k'}q_{\varphi' k'} \quad k\neq k'$$

, . (2.62)

Direct Stochastic Calculation

Here we briefly discuss an alternative approach to computing $\boldsymbol{\Gamma_c}$ directly from $\boldsymbol{\eta}$, bypassing reconstruction. Small KMC simulations similar to ones used for reconstruction are run allowing reactions and diffusion. The corresponding transition probability rates are computed from the number of reactions within a CG cell divided by time. As in the stochastic reconstruction process, the results can be tabulated and used for all cells at all times or simulations may be performed for each CG cell on-the-fly.

Calculation of adparticle interactions, **U**, requires resolution of specific site coverages in nearby cells, so $\boldsymbol{\theta}$ must still be recorded. Additionally, transition probability

rates for reactions and diffusion across CG cell borders are more complicated to stochastically calculate since adjacent CG cells are involved in the simulation. A massive number of combinations of η_k^{α}, $\eta_{k'}^{\alpha'}$, $q_{\varphi k}$, $q_{\varphi k'}$, $\bar{U}_{\varphi k}^{\alpha}$ and $\bar{U}_{\varphi k'}^{\alpha'}$ are possible, making simulation and storage demanding. A reconstructed $\boldsymbol{\theta}$ may be used in these cases while direct stochastic synthesis generates the rest of $\boldsymbol{\Gamma_c}$. It is worth noting that using the direct stochastic method for intra-cell transition probability rates negates the need for the correction noted in Eq. (2.60). A summary of transition probability rates for CGMC with homogenization is given in Table 2.5.

Table 2.5 - Summary of transition probability rates for CGMC (with homogenization).

Process	Coarse transition probability rate $\left(\overline{\Gamma}(\boldsymbol{\eta},\varphi,k,\varphi',k')\right)$	Coarse rate constant $\left(\overline{\Gamma}^{\%}\right)$
Adsorption $\left(\alpha+\phi\cdot\varphi\rightarrow\alpha\cdot\varphi\right)$	$\overline{\Gamma}^{\alpha}_{ads}=\overline{\Gamma}^{\%}_{ads,\varphi k}\theta^{\phi}_{\varphi k}$	$\overline{\Gamma}^{\%}_{ads,\varphi k}=\sum_{j\in C_k}\frac{\Gamma^{\alpha}_{ads,\varphi j}}{q_{\varphi k}}$
Desorption $\left(\alpha\cdot\varphi\rightarrow\alpha+\phi\cdot\varphi\right)$	$\overline{\Gamma}^{\alpha}_{des}=\overline{\Gamma}^{\%}_{des,\varphi k}\theta^{\alpha}_{\varphi k}e^{-\beta\overline{U}^{\alpha}_{\varphi k}}$	$\overline{\Gamma}^{\%}_{des,\varphi k}=\sum_{j\in C_k}\frac{\Gamma^{\alpha}_{des,\varphi j}}{q_{\varphi k}}$
Isomerization Reaction $\left(\alpha\cdot\varphi\rightarrow\alpha'\cdot\varphi\right)$	$\overline{\Gamma}^{\alpha}_{iso}=\overline{\Gamma}^{\%}_{iso,\varphi j}\theta^{\alpha}_{\varphi k}$	$\overline{\Gamma}^{\%}_{iso,\varphi k}=\sum_{j\in C_k}\frac{\Gamma^{\alpha\rightarrow\alpha'}_{iso,\varphi j}}{q_{\varphi k}}$
Surface Hopping (Diffusion) $\left(\alpha\cdot\varphi+\phi\cdot\varphi'\rightarrow\phi\cdot\varphi+\alpha\cdot\varphi'\right)$		
$k\neq k'$	$\overline{\Gamma}^{\alpha}_{mig}=\overline{\Gamma}^{\%}_{m,\varphi k\rightarrow\varphi' k'}\theta^{\alpha}_{\varphi k}\theta^{\phi}_{\varphi' k'}e^{-\beta\overline{U}^{\alpha}_{k}}$	$\overline{\Gamma}^{\%}_{m,\varphi k\rightarrow\varphi' k'}=\frac{\Gamma^{\alpha}_{m,\varphi k\rightarrow\varphi' k'}q^{B}_{\varphi k\varphi' k'}}{n_{\varphi\varphi'}}\frac{\sum_{j'}(\mathbf{r}_{\varphi j\rightarrow\varphi j'}\cdot\boldsymbol{\xi})^{*}}{\mathbf{r}_{\varphi k\varphi k'}\cdot\boldsymbol{\xi}}$
Disproportionation reaction $\left(\alpha\cdot\varphi+\alpha'\cdot\varphi'\rightarrow\alpha''\cdot\varphi+\alpha'''\cdot\varphi'\right)$		
$k=k'$	$\overline{\Gamma}^{\alpha\alpha'\rightarrow\alpha''\alpha'''}_{rxn,\varphi k\varphi' k'}=\overline{\Gamma}^{\%}_{rxn,\varphi k\varphi' k'}\theta^{\alpha}_{\varphi k}\theta^{\alpha'}_{\varphi' k,\alpha\varphi}$	$\overline{\Gamma}^{\%}_{rxn,\varphi k\varphi' k'}=\frac{\Gamma^{\alpha\alpha'\rightarrow\alpha''\alpha'''}_{rxn,\varphi k\varphi' k'}}{n_{\varphi\varphi'}}\left(n_{\varphi\varphi'}q_{\varphi k}-\sum_{k'\neq k}q^{B}_{\varphi k,\varphi' k'}\right)$
$k\neq k'$	$\overline{\Gamma}^{\alpha\alpha'\rightarrow\alpha''\alpha'''}_{rxn,\varphi k\varphi' k'}=\overline{\Gamma}^{\%}_{rxn,\varphi k\varphi' k'}\theta^{\alpha}_{\varphi k}\theta^{\alpha'}_{\varphi' k'}$	$\overline{\Gamma}^{\%}_{rxn,\varphi k\varphi' k'}=\frac{\Gamma^{\alpha\alpha'\rightarrow\alpha''\alpha'''}_{rxn,\varphi k\varphi' k'}}{n_{\varphi\varphi'}}q^{B}_{\varphi k,\varphi' k'}$

*- Subject to the constraints of Eq. (2.35)

Example C: Diffusion in a Zeolite-like Membrane

Homogenization is demonstrated with one-dimensional (1D) diffusion through a zeolite-like membrane, adapted from Ref. [7]. A 1D zeolite-like chain is comprised of alternating window (W) and intercage (S) sites. Only diffusion is considered. There are 3 separate diffusion processes (see Figure 2.4a, Table 2.6). Two diffusion processes are to a nearest neighbor (W→S and S→W), and to a next-nearest neighbor (W→W), skipping over the intermediate S site. Coarse cells are shown in Figure 2.4b. Each cell was homogenized over all site types (dotted line in Figure 2.4b). This allows the lumping of all transition probability rates from cell *k* to cell *k'* into a single rate (Figure 2.4c).

Two lattice sizes are considered: 101 and 1001 microscopic sites to allow both boundaries to be W sites. Dirichlet boundary conditions are implemented; specifically, the left boundary site is always occupied and the right boundary site is always empty. The lattice is initially empty.

Table 2.6 – Summary of processes and transition probability rates for Example C.

Process		**Microscopic (KMC) -** Γ	**Coarse Grained -** $\overline{\Gamma}$
W→W diffusion:	$A \cdot \phi_W + \phi_W \rightarrow \phi_W + A \cdot \phi_W$	$\frac{\Gamma^A_{mig,W\rightarrow W}}{n_{WW}} \sigma_{W\,j}(1-\sigma_{W\,j'}) e^{-\beta U_j}$	$\overline{\Gamma}_{\text{mig,Wk}\rightarrow\text{Wk}'} \theta^{\text{A}}_{\text{Wk}} \theta^{\phi}_{\text{Wk}'} e^{-\beta \overline{U}_0^{\text{A}}}$
W→S diffusion:	$A \cdot \phi_W + \phi_S \rightarrow \phi_W + A \cdot \phi_S$	$\frac{\Gamma^A_{mig,W\rightarrow S}}{n_{WS}} \sigma_{W\,j}(1-\sigma_{S\,j'}) e^{-\beta U_j}$	$\overline{\Gamma}_{\text{mig,Wk}\rightarrow\text{Sk}'} \theta^{\text{A}}_{\text{Wk}} \theta^{\phi}_{\text{Sk}'} e^{-\beta \overline{U}_0^{\text{A}}}$
S→W diffusion:	$A \cdot \phi_S + \phi_W \rightarrow \phi_S + A \cdot \phi_W$	$\frac{\Gamma^A_{mig,S\rightarrow W}}{n_{SW}} \sigma_{S\,j}(1-\sigma_{W\,j'}) e^{-\beta U_j}$	$\overline{\Gamma}_{\text{mig,Sk}\rightarrow\text{Wk}'} \theta^{\text{A}}_{\text{Sk}} \theta^{\phi}_{\text{Wk}'} e^{-\beta \overline{U}_0^{\text{A}}}$

Microscopic Description

The microscopic transition probability rate of adparticle diffusion is

$$\Gamma(\sigma, j, j') = \frac{\Gamma^{\alpha}_{mig}}{n_{\varphi_j \varphi_{j'}}} \sigma_j (1-\sigma_{j'}) e^{-\beta U_j}, \ \sigma_j = \begin{Bmatrix} 1 \, if \ \ site \ \ j \ occupied \\ 0 \, otherwise \end{Bmatrix}. \qquad (2.63)$$

Constants for the processes are presented in Table 2.7. The temperature is assumed to be 700 K. Adparticle-adparticle interactions are considered to be negligible.

Table 2.7 – Transition probability rate constants for Example C.

ϕ_j	ϕ_j	Γ_{mig} (s^{-1})	U_0 (kJ mol^{-1})
W	W	$6.0 \cdot 10^{11}$	1.1
W	S	$2.7 \cdot 10^{12}$	17.0
S	W	$1.6 \cdot 10^{13}$	44.8

CGMC Description

Simulations were performed at varying CG cell size with an adaptive mesh [23]. To implement the boundary conditions, the two boundary CG cells were of size *q = 1*, representing a single microscopic site of defined occupancy. Across the rest of the lattice, different levels of coarse graining were used (*q* = 2, 4, 8, and 16). Starting at the site following the left boundary cell, the first *q* microscopic sites were lumped, followed by the

next *q* microscopic sites, and so on. Since the 99 (or 999) intermediate lattice nodes did not evenly divide into uniform CG cells with these selections of *q*, upon reaching the right boundary the remainder of sites were grouped into a CG cell smaller than *q.*

All spatially coarse diffusion rates between cells were calculated using the homogenized equations in Table 2.5. For all site combinations, $\mathrm{n}_{\varphi_j\varphi_{j'}} = 2$, reflecting that each site has two potential reacting neighbors for each process. The center-to-center distance between cells, $\hat{\mathrm{a}}_{\varphi k\varphi k'}.\hat{\xi}$, depends on the CG cell size (*q*) chosen for the simulation: $\hat{\mathrm{a}}_{\mathrm{W}\to\mathrm{W}}.\hat{\xi} = 2$, whereas $\hat{\mathrm{a}}_{\mathrm{W}\to\mathrm{S}}.\hat{\xi} = \hat{\mathrm{a}}_{\mathrm{S}\to\mathrm{W}}.\hat{\xi} = 1$, reflecting the extra distance between pairs of window sites. For all cells, $\mathrm{q}^{\mathrm{B}}_{\varphi k\varphi' k'} = 2$. Transition probability rates for diffusion between the cells are lumped according to Eq. (2.57),

$$\Gamma_{\mathrm{C}\,k,k'}^{mig} = \sum_{\varphi,\varphi'\in I_{k,k'}} \overline{\Gamma}_{\varphi k,\varphi' k'}^{mig},$$

(2.64)

where $I_{k,k'}$ is the $\mathrm{C_k}$ to $\mathrm{C_{k'}}$ interface region of sites over which jumps are allowed. For example, only boundary sites would be included when only nearest-neighbor jumps are allowed. Applied to the present example, Eq. (2.64) becomes

$$\Gamma_{\mathrm{C}\,k,k'}^{mig} = \overline{\Gamma}_{Wk,Wk'}^{mig} + \overline{\Gamma}_{Wk,Sk'}^{mig} + \overline{\Gamma}_{Sk,Wk'}^{mig}$$

. (2.65)

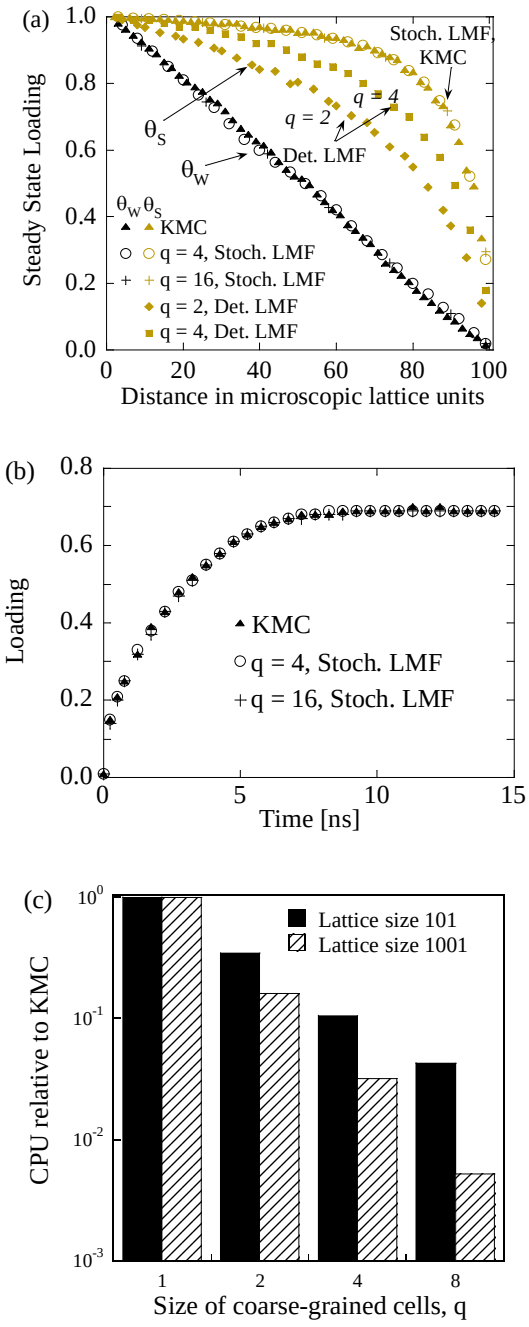

Figure 2.5 (a) Steady state loading profiles in a 1D zeolite membrane with Dirichlet boundary conditions. CGMC using a stochastic LMF closure with varying levels coarse graining compares well with traditional KMC. Steady state solution based on the deterministic LMF closure (ϑ_S only q = 2, 4 is shown for clarity; ϑ_W not shown for clarity). (b) Zeolite loading vs. time. (c) CPU relative to

traditional KMC of simulating the same amount of time at steady state with varying levels of spatial coarse graining on two lattice sizes.

Note that the terms on the RHS of Eq. (2.65) describe diffusion events among microscopic lattice sites close to or at the interface of the two cells. Some terms may be zero depending on what types of microscopic diffusion jumps are possible across the boundary from $\mathrm{C_k}$ to $\mathrm{C_{k'}}$. For example, for a $\mathrm{C_k}$ and $\mathrm{C_{k'}}$ corresponding to Figure 2.4b only the first and third terms are non-zero.

Figure 2.5a compares steady state loading solutions for KMC (filled triangles) and CGMC (open circles, q = 4 and crosses, q = 16) using a stochastic LMF reconstruction. In the stochastic LMF reconstruction, short KMC simulations were used to solve for $\theta_{\varphi k}^{\alpha}$ concentrations at various η_k^{α} occupancies and the results were tabulated. Figure 2.5b shows the loading against real time as the system approaches steady state. CPU comparison of simulating the same real time at steady state for the two lattice sizes is shown in Figure 2.5c. The CGMC method provides large CPU savings while retaining very good accuracy. Computational savings are larger the larger the level of coarse-graining q. Reducing combinatorial complexity by coarsening process rates succeeded without disrupting the accuracy of the solution.

A deterministic reconstruction method to solve the LMF approximation returns incorrect $\theta_{\varphi k}^{\alpha}$ values for CG cells in CGMC. In sufficiently large, well-mixed systems, equilibrium can accurately be described by solving the nonlinear set of equations describing the diffusion between sites:

$$k_{W\to S}\cdot\theta_W\cdot(1-\theta_S)=k_{S\to W}\cdot\theta_S\cdot(1-\theta_W) \tag{2.66}$$

$$q_W \cdot \theta_W + q_S \cdot \theta_S = \eta_W + \eta_S \ . \qquad (2.67)$$

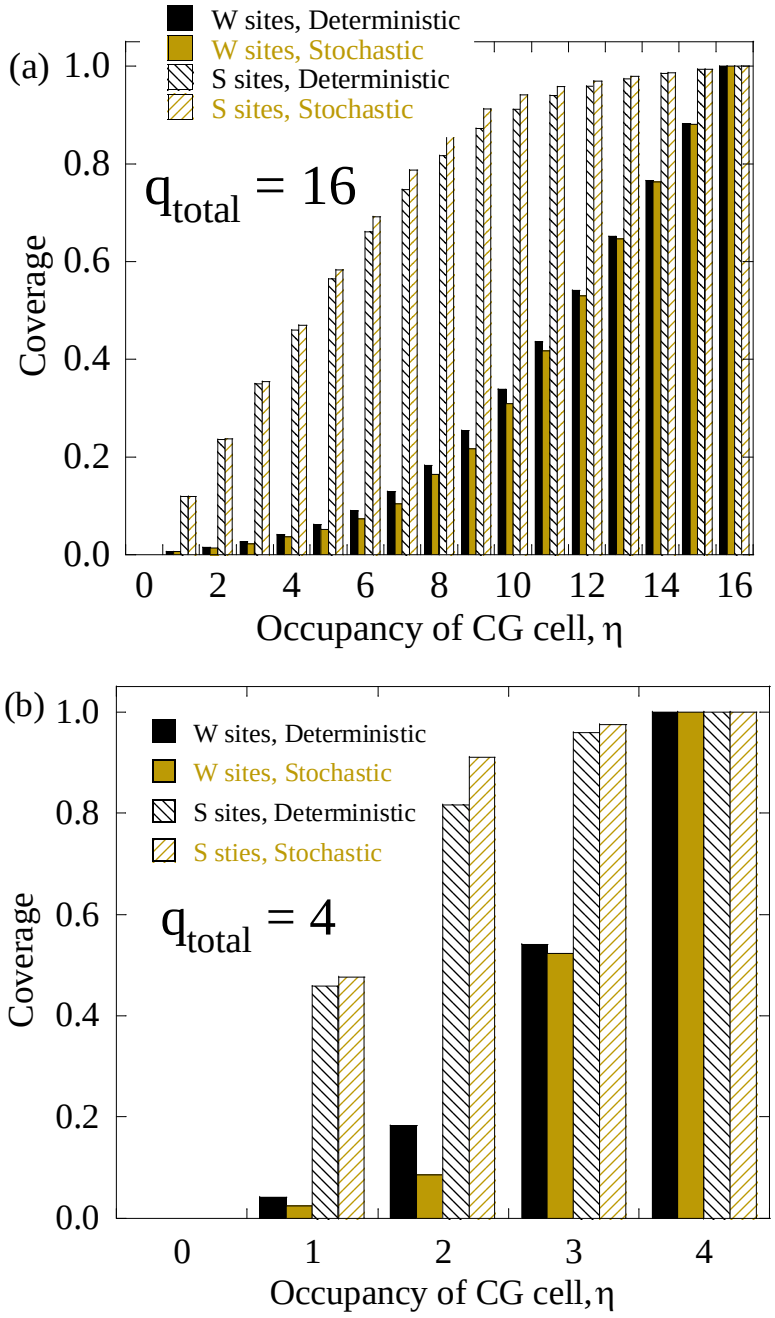

Figure 2.6 - Examples of mean-field coverages within CG cells using the deterministic and stochastic LMF approximation. The error in the deterministic LMF solution relative to the accurate stochastic LMF solution increases as the CG cell size, q, decreases. (a) qw= qs = 8, (b) qw= qs = 2.

This set may easily be solved using a nonlinear solver, such as the Newton's method. However, errors occur due to the significant covariance of the concentrations (ϑ_W, ϑ_S) when applied to small CG cells with a finite number of sites. For large cells the covariance decreases, and consequently lower error is observed.

In order to rationalize the error from the LMF approximation, Figure 2.6 shows the discrepancy in coverage between a deterministic nonlinear reconstruction and a stochastic reconstruction of specific site coverages of well-mixed adparticles on W and S sites. At larger cell sizes, the discrepancy between the deterministic and stochastic solutions is small (Figure 2.6a). As the coarse cell size decreases, the error increases (Figure 2.6b). This also is apparent in the actual simulation. Figure 2.5a shows two profiles for the S sites for q = 2 and q = 4 using a deterministic LMF reconstruction (W sites and q > 4 are not shown for clarity). As q increases, the deterministic LMF reconstruction's accuracy improves. While we have only demonstrated this q relationship for the present diffusion system and extension to other systems has not been tested, this does demonstrate that deterministic reconstruction methods deliver incorrect results.

Conclusions

In this chapter, a framework for on-lattice, coarse-grained Monte Carlo (CGMC) simulation with an arbitrary number of site types and/or surface species has been derived. This enables efficient Monte Carlo calculations of many realistic systems for the first time. Numerical examples demonstrated the ability to capture adsorption, desorption, diffusion,

and reaction systems on spatially inhomogeneous surfaces and/or with multicomponent mixtures. Large computational savings are achieved for multicomponent systems compared to the traditional kinetic Monte Carlo (KMC) method. A novel homogenization approach at the stochastic level was introduced. The homogenization approach enables reduction in the number of processes over multiple sites with concomitant decrease in the memory and CPU requirements. It is expected that the homogenization method will be particularly important for systems exhibiting combinatorial explosion.

Acknowledgements

The research was partially supported by the Department of Energy (DE-FG02-05ER25702).

REFERENCES

1. Collins, S.D., A. Chatterjee, and D.G. Vlachos, *Coarse-grained kinetic Monte Carlo models: Complex lattices, multicomponent systems, and homogenization at the stochastic level.* Journal of Chemical Physics, 2008. **129**(18).

2. Katsoulakis, M.A. and D.G. Vlachos, *Coarse-grained stochastic processes and kinetic Monte Carlo simulators for the diffusion of interacting particles.* J. Chem. Phys., 2003. **119**(18): p. 9412-9428.

3. Katsoulakis, M., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes for microscopic lattice systems.* Proc. Natl. Acad. Sci., 2003. **100**(3): p. 782-787.

4. Katsoulakis, M.A., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems.* J. Comp. Phys., 2003. **186**: p. 250-278.

5. Iyengar, V. and M.O. Coppens, *Dynamic Monte-Carlo simulations of binary self-diffusion in zeolites: effect of strong adsorption sites.* Chemical Engineering Science, 2004. **59**(22-23): p. 4747-4753.

6. Ziff, R.M., E. Gulari, and Y. Barshad, *Kinetic phase transitions in an irreversible surface-reaction model.* Physical Review Letters, 1986. **56**(24): p. 2553-2556.

7. Snyder, M.A., D.G. Vlachos, and M.A. Katsoulakis, *Mesoscopic modeling of transport and reaction in microporous crystalline membranes.* Chemical Engineering Science, 2003. **58**: p. 895-901.

8. Keil, F.J., R. Krishna, and M.O. Coppens, *Modeling of diffusion in zeolites.* Reviews in Chemical Engineering, 2000. **16**(2): p. 71-197.

9. Auerbach, S.M., *Theory and simulation of jump dynamics, diffusion and phase equilibrium in nanopores.* Int. Rev. Phys. Chem., 2000. **19**(2): p. 155-198.

10. Mayawala, K., D.G. Vlachos, and J.S. Edwards, *Computational modeling reveals molecular details of epidermal growth factor binding.* Bmc Cell Biology, 2005. **6**.

11. Woolf, P.J. and J.J. Linderman, *Self organization of membrane proteins via dimerization.* Biophys. Chem., 2003. **104**: p. 217–227.

12. Mahama, P. and J. Linderman, *A Monte Carlo study of the dynamics of G-protein activation.* Biophys. J., 1994. **67**(3): p. 1345-1357.

13. Chatterjee, A. and D.G. Vlachos, *Multiscale spatial Monte Carlo simulations: Multigriding, computational singular perturbation, and hierarchical stochastic closures.* Journal of Chemical Physics, 2006. **124**(6): p. 0641101-06411016.

14. Chatterjee, A. and D.G. Vlachos, *Coarse-grained kinetic Monte Carlo models for complex lattices and multicomponent systems.* 2007: p. in preparation.

15. Papanicolaou, G.C., *Diffusion in random media.* Surveys in Applied Mathematics, 1995. **1**: p. 205-253.

16. Lam, R., D.G. Vlachos, and M.A. Katsoulakis, *Homogenization of mesoscopic theories: Effective properties of model membranes.* AIChE Journal, 2002. **48**(5): p. 1083-1092.

17. Deshmukh, S.R., et al., *From density functional theory to microchemical device homogenization: Model prediction of hydrogen production for portable fuel cells.* International Journal on Multiscale Computational Engineering, 2004. **2**(2): p. 221-238.

18. *Chemical reactions in complex mixtures: the Mobil Workshop*. ed. F.K. A. J. Sapre. 1991, Van Nostrand Reinhold: New York.

19. Vlachos, D.G., *Stochastic modeling of chemical microreactors with detailed kinetics: Induction times and ignitions of H_2 in air.* Chemical Engineering Science, 1998. **53**(1): p. 157-168.

20. Snyder, M.A., A. Chatterjee, and D.G. Vlachos, *Net-event kinetic Monte Carlo for overcoming stiffness in spatially homogeneous and distributed systems.* Comput. Chem. Eng., 2005. **29**(4): p. 701-712.

21. Cao, Y., D.T. Gillespie, and L.R. Petzold, *The slow-scale stochastic simulation algorithm.* Journal of Chemical Physics, 2005. **122**: p. 014116-1-18.

22. Samant, A., B.A. Ogunnaike, and D.G. Vlachos, *A Hybrid Multiscale Monte Carlo Algorithm (HyMSMC) to Cope with Disparity in Time Scales and Species Populations in Well-Mixed Reaction Networks.* BMC Bioinformatics, 2007. **8**: p. No. 175.

23. Chatterjee, A., D.G. Vlachos, and M.A. Katsoulakis, *Spatially adaptive lattice coarse-grained Monte Carlo simulations for diffusion of interacting molecules.* Journal of Chemical Physics, 2004. **121**(22): p. 11420-11431.

CHAPTER 3

ERROR QUANTIFICATION IN COARSE-GRAINED MONTE CARLO SIMULATIONS OF CHEMICAL REACTIONS

Abstract:

The Coarse-grained Monte Carlo (CGMC) method was recently introduced and has demonstrated a considerable speedup over the traditional lattice-based kinetic Monte Carlo (KMC) method. However, little work has been devoted to studying the errors in CGMC simulations in the presence of chemical reactions. Density of reactants, ratio of diffusion to reaction, and the coarseness of the lattice all contribute to the accuracy of the CGMC method. In this chapter, we investigate the error of CGMC simulations for prototype reaction systems.

Introduction:

Prior studies using the coarse-grained MC (CGMC) method [1, 2] have only been applied to a limited number of systems. While the CGMC method offers large computational savings, the accuracy of the CGMC method is degraded by its (local mean-field) approximation. Unfortunately there is no straightforward way to quantitatively predict the error associated with replacing a KMC method with a CGMC method. By investigating and

quantifying simulation error, the CGMC method can be more confidently applied and techniques to correct for inaccuracies can be developed.

There has been great interest in developing models of protein-protein interactions in signaling networks [3-7]. The KMC method is well suited to many of these systems as they often feature a small number of protein copies and spatial diffusive transport. However, the large space (relative to the size of a protein) of the plasma membrane and the long time scales involved limit the KMC method to adequately simulate these systems in any reasonable amount of computational time. We are therefore greatly interested in applying the CGMC method to these systems to reach these large scales. Many of these protein interaction systems can be generally classed as reaction-diffusion systems [4, 8-10]. By investigating prototype reaction-diffusion systems, we can apply the results to these systems of immediate interest.

Herein, we focus on error quantification for the CGMC method for two prototype reaction-diffusion systems. Due to the interest in applying KMC to systems of small, discrete population (such as those found in biological system) we focus much of our simulations on these low-population situations. We also assess the accuracy of the CGMC method at different ratios of reaction rate and diffusivity, represented as the dimensionless Damköhler number (Da).

Methodology

Damköhler number

The accuracy of the CGMC against the KMC method is assessed at various values of the intrinsic reaction rate constant (k_f) and the diffusitivity (D), by employing a dimensionless number, the second Damköhler number, which is defined as the ratio of the time scale of a collision event and the time scale of a reaction event [11]. Hereafter, we drop the prefix 'second' for brevity. Precise calculation of the time for two reactants to diffuse to each other (the collision timescale) requires knowledge of the spatial arrangement of molecules, an output of KMC simulations. To avoid carrying out KMC simulations, we estimate the distribution using the mean field (MF) approximation and the macroscopic parameters available (k_f and D).

Two simple diffusion-reaction systems are used in this work. In the AA system, a single species diffuses upon the lattice and reacts with identical molecules. In the AB system, two species diffuse upon the lattice and react with the opposite species but are nonreactive with identical molecules. Following [11], the Da number is calculated as:

$$\mathrm{Da} = \frac{\text{time scale of collision event}}{\text{time scale of reaction event}} = \frac{\Gamma^{\mathrm{r}}_{\mathrm{MF-ODE}}}{\Gamma_{\text{MF-collision}}} \tag{3.1}$$

Here, $\Gamma^{\mathrm{r}}_{\mathrm{MF-ODE}}$ is the MF rate of reaction per unit site and $\Gamma_{\text{MF-collision}}$ is the MF rate of collision of two reactive species per unit site. $\Gamma^{\mathrm{r}}_{\mathrm{MF-ODE}}$ on a square lattice is calculated as

$$\Gamma^{\mathrm{r}}_{\mathrm{MF-ODE}} = k_{\mathrm{f}}\theta_A\theta_B \text{ or} \tag{3.2}$$

$$\Gamma^{\mathrm{r}}_{\mathrm{MF-ODE}} = k_{\mathrm{f}}\theta_A^2 \tag{3.3}$$

For AB and AA diffusion-reaction systems, respectively. $\Gamma_{\text{MF-collision}}$ on a square lattice is calculated as

$$\Gamma_{\text{MF-collision}} = \frac{4D}{s^2}\theta_A\theta_B = \Gamma^d_{AB}\theta_A\theta_B \text{ or} \tag{3.4}$$

$$\Gamma_{\text{MF-collision}} = \frac{4D}{s^2}\theta_A^2 = \Gamma^d_{AA}\theta_A^2 \tag{3.5}$$

for AB and AA diffusion-reaction systems, respectively [12]. Species lattice coverage is noted as θ_A and θ_B, the sum of the diffusivities of the reacting species is ***D*** (For the AB system, ***D*** = D_A+D_B; the AA system, ***D*** = 2D_A). Γ^d_{AA} and Γ^d_{AB} are the total transition rate of diffusion in all directions. The lattice constant (distance between lattice sites) is **s**. Inserting Eq. (3.2) & (3.4) into (3.1),

$$\mathrm{Da} = \frac{k_{\mathrm{f}}}{\Gamma^d_{AB}} = \frac{k_{\mathrm{f}}}{\frac{4D_{AB}}{s^2}} \tag{3.6}$$

$$\mathrm{Da} = \frac{k_{\mathrm{f}}}{\Gamma^d_{AA}} = \frac{k_{\mathrm{f}}}{\frac{4D_{AA}}{s^2}} \tag{3.7}$$

for the AB and AA systems, respectively.

Length Scales

The average expected distance between reactants and the characteristic length of coarse diffusion in CGMC are useful in analysis of the results. The distance traveled via diffusion is

$$L = \sqrt{4Dt}. \tag{3.8}$$

Here ***L*** is the diffusion distance, ***D*** is the sum of diffusivities of the colliding reactants, ***t*** is the time spent diffusing. Using the inverse of the MF approximation of the rate of collision events ($\Gamma_{\text{MF-collision}}$) we calculate the expected timescale for collisions per particle as

$$t_{A-AA\,model} \cong \frac{N_A}{2\Gamma_{\text{MF-collision}} N_{sites}} \text{ and} \tag{3.9}$$

$$t_{A-AB\,model} = t_{B-AB\,model} \cong \frac{N_A}{\Gamma_{\text{MF-collision}} N_{sites}} \tag{3.10}$$

for the AA system and AB systems, respectively. Since $N_A = N_B$ and $D_A = D_B$ in the AB system, the expected timescale of diffusion is the same for each reactant. The expected length scale of diffusion may then be estimated by inserting Eq. (3.10) into Eq. (3.8):

$$L_{AB} = \sqrt{4Dt_{diff}} = \sqrt{4D(t_A + t_B)} \cong \sqrt{\frac{4(2D_A)(2N_A)}{\Gamma_{\text{MF-collision}} N_{sites}}} \tag{3.11}$$

$$\frac{L_{AB}}{s} \cong \sqrt{\frac{2}{\theta_A}} = \sqrt{\frac{4}{\theta_T}} \tag{3.12}$$

For the AA system, Eq. (3.9) is used, resulting in

$$\frac{L_{AA}}{s} \cong \sqrt{\frac{1}{\theta_A}}. \tag{3.13}$$

$\frac{L_{AA}}{s}$ and $\frac{L_{AB}}{s}$ are the diffusion lengths divided by the lattice constant s, resulting in the dimensionless diffusion length in terms of lattice steps.

The characteristic length of coarse-graining is defined as the average length of a single coarse-grained hop. This lnytength as a function of q is straightforward on the uniform square coarse lattice used in these simulations:

$$\frac{L_{CG}}{s} = \sqrt{q} \tag{3.14}$$

where ***q*** is the number of sites in each square CG cell.

Quantification of Simulation Error

Simulation error is defined as the ratio of the reaction rates of the CGMC simulation method to the KMC method.

$$\xi = \frac{\mathrm{r}}{\mathrm{r_{KMC}}} \tag{3.15}$$

The KMC rate of reaction, r_{KMC}, is defined as the time average rate of microscopic events over some time at steady state. The rates of reaction of the Gillespie (hereafter MF) and CGMC methods are denoted respectively as r_{MF} and r. The further ξ is from 1, the greater the inaccuracy relative to the KMC method. Both the Gillespie and CGMC methods use MF approximation, which assumes infinitely fast diffusion, on some length scale. This neglects the diffusion time necessary to bring together reactants and leads to a ξ > 1.

Reaction-Diffusion Systems

The two reaction-diffusion systems (AA system with a single species diffusion and bimolecular reaction and the AB system with two-species diffusion and bimolecular reaction) are described in Table 3.1 and Table 3.2, respectively.

Table 3.1– The AA diffusion-reaction system and its parameters.

AA diffusion-reaction system	
Reactants	1 species (A). The number of adparticles, N_A, is held constant throughout the simulation.
Simulation Space	100x100 (N_{sites} = 10000) square lattice. Each site may host a single adparticle. All sites are identical. Periodic boundary conditions are used.
Diffusion	Diffusion is performed to immediate neighbors only (4 neighbors for each site). $$A^* + * \rightarrow * + A^*$$ Here, A* represents an A-occupied site and * represents a vacant site. The transition rate of diffusion for an adparticle from site *i* to neighboring site *j* is $$\Gamma^{d} = \frac{1}{4}\frac{D_A}{s^2}\sigma_i(1-\sigma_j)$$ D_A is the adparticle's diffusivity, σ is the discrete site occupancy value (1 if a site is occupied and 0 if vacant), and s is the lattice constant.
Reaction	Adjacent A adparticles may react and desorb via the mechanism: $$A^* + A^* \rightarrow AA + 2*$$ Here, A* represents an A-occupied site and * represents a vacant site. The transition rate of reaction is then $$\Gamma^{r} = \frac{k_f}{4}\sigma_i\sigma_j$$ k_f is the MF reaction rate constant, σ is the discrete site occupancy value

	(1 if a site is occupied and 0 if vacant). To maintain a constant number of adparticles on the surface, each reaction-desorption event is followed with the immediate adsorbtion of two A adparticles on randomly selected vacant lattice nodes.
Parameters	
Damköhler number	$\mathrm{Da} = {k_\mathrm{f}}\Big/{\frac{4(D_A + D_A)}{s^2}}$
Species Coverage	$\theta_A = \frac{\mathrm{N_A}}{\mathrm{N_{sites}}}$
Diffusion Length	$\frac{L_{AA}}{s} \cong \sqrt{\frac{1}{\theta_A}}$
Characteristic CG length	$\frac{L_{CG}}{s} = \sqrt{q}$
Reaction rate ratio	$\xi = \frac{\mathrm{r}}{\mathrm{r_{KMC}}}$

Simulations were performed of the AA system (Table 3.1) and compared to KMC results to assess accuracy. The Damköhler number was varied between (Da = 0.016-3.0). The total coverage of reactants on the lattice was varied from (θ_T = 0.0003-0.9). CGMC simulations were performed using q = 4, 25, 100, 625. Results of the simulations are shown in Figure 3.2-Figure 3.1.

Table 3.2 – AB diffusion reaction system

AB diffusion-reaction system	
Reactants	2 species (A and B). The number of each species, N_A and N_B, are held constant throughout the simulation.
Simulation Space	100x100 (N_{sites} = 10000) square lattice. Each site may host a single adparticle. All sites are identical. Periodic boundary conditions are used.
Diffusion	Diffusion is performed to immediate neighbors only (4 neighbors for each site). $$A^* + * \rightarrow * + A^*$$ Here, A* represents an A-occupied site and * represents a vacant site. The transition rate of diffusion for an adparticle from site *i* to neighboring site *j* is $$\Gamma^{d} = \frac{1}{4}\frac{D_A}{s^2}\sigma_i(1-\sigma_j)$$ D_A is the adparticle's diffusivity, σ is the discrete site occupancy value (1 if a site is occupied and 0 if vacant), and s is the lattice constant.
Reaction	Adjacent A and B adparticles may react and desorb via the mechanism: $$A^* + B^* = AB + 2*$$ Here, A* represents an A-occupied site and * represents a vacant site. The elementary, well-mixed reaction rate is $$\Gamma^{r} = \frac{k_f}{4}\sigma_i\sigma_j$$ k_f is the MF reaction rate constant, σ is the discrete site occupancy value (1 if a site is occupied and 0 if vacant). To maintain a constant number of adparticles on the surface, each reaction-desorption event is followed with the immediate adsorption of one A and one B adparticle on randomly selected vacant lattice nodes.
Parameters	

Damköhler number	$\mathrm{Da} = {k_f}\Big/{\frac{4(D_A + D_B)}{s^2}}$
Species Coverage	$\theta_A = \frac{\mathrm{N_A}}{\mathrm{N_{sites}}}, \theta_B = \frac{\mathrm{N_B}}{\mathrm{N_{sites}}}, \theta_T = \theta_A + \theta_B$
Diffusion Length	$\frac{L_{AB}}{s} \cong \sqrt{\frac{2}{\theta_A}} = \sqrt{\frac{4}{\theta_T}}$
Characteristic CG length	$\frac{L_{CG}}{s} = \sqrt{q}$
Reaction rate ratio	$\xi = \frac{\mathrm{r}}{\mathrm{r_{KMC}}}$

Simulations were performed of the AB system (Table 3.2) and compared to KMC results to assess accuracy. The Damköhler number was varied between (Da = 0.01-10.0). The total coverage of reactants on the lattice was varied from (θ_T = 0.001-0.90). CGMC simulations were performed using q = 4, 25, and 100. The individual species coverage was kept equal ($\theta_A = \theta_B = ½\ \theta_T$) and each reactant had the same diffusivity ($D_A = D_B$). Results of the simulations are shown in Figure 3.6-Figure 3.9.

Results of AA system

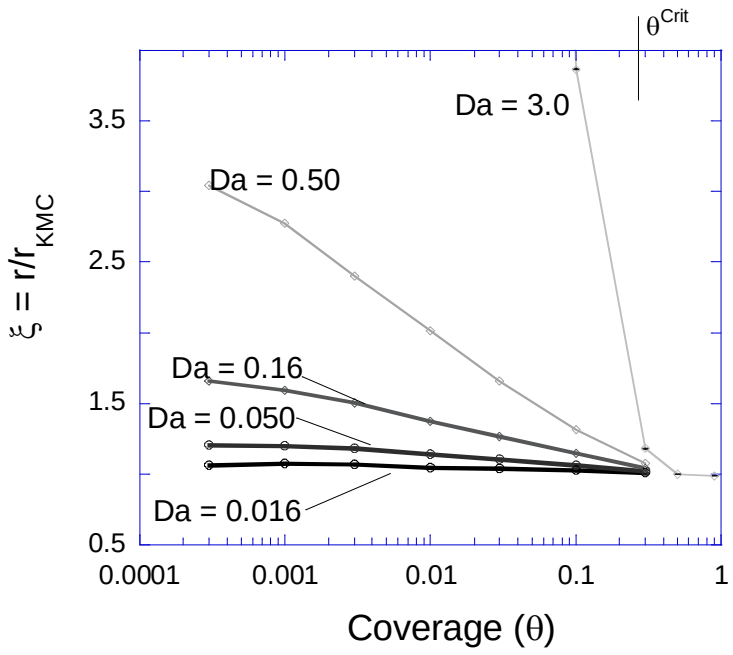

Figure 3.1– Error of CGMC method for AA system (Table 3.1), in the limit a single cell (Gillespie).

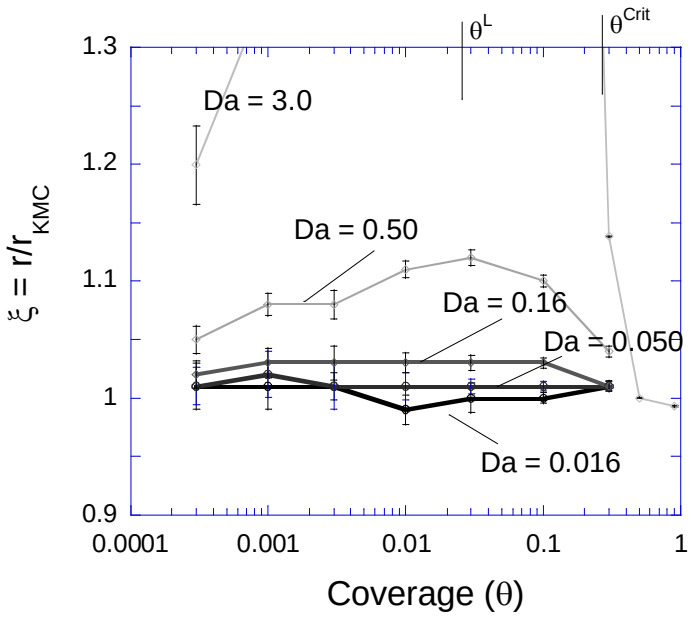

Figure 3.2 – Error of CGMC method for AA system (Table 3.1), q = 4.

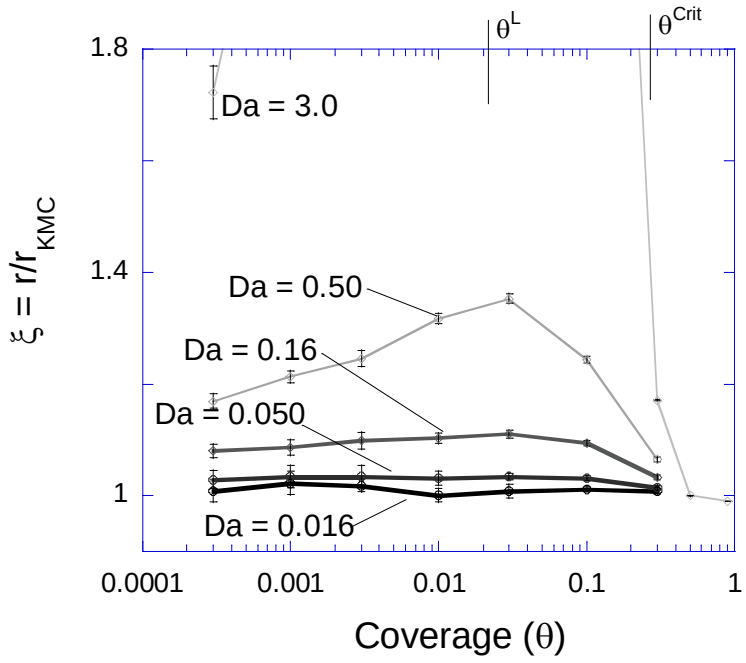

Figure 3.3 – Error of CGMC method for AA system (Table 3.1), q = 25.

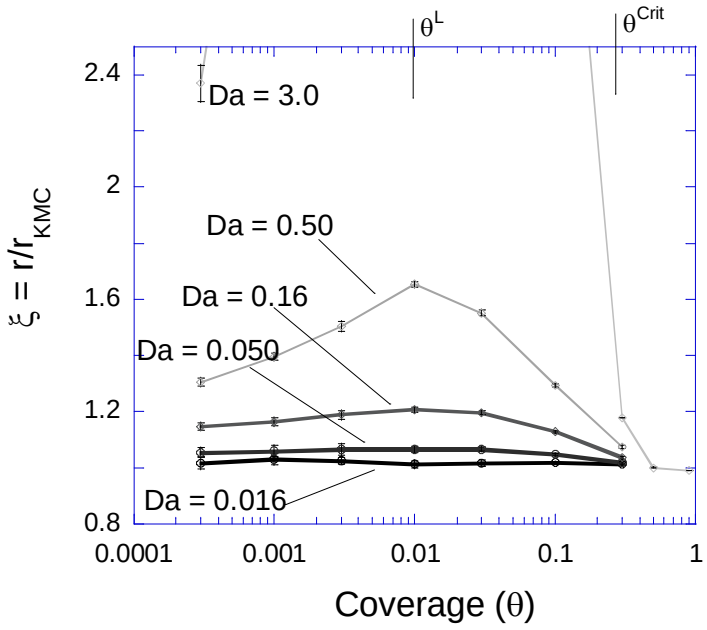

Figure 3.4 – Error of CGMC method accuracy for AA system (Table 3.1), q = 100.

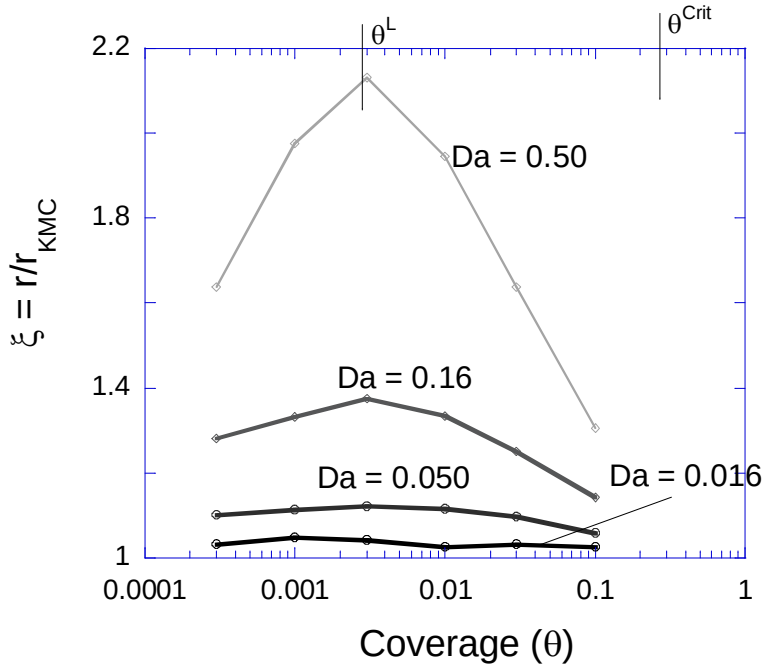

Figure 3.5 – Error of CGMC method for AA system (Table 3.1), q = 625.

As $\theta_A \rightarrow 0$, the distance reactants must travel to collide with another reactant increases ($L \rightarrow \infty$) and the travel time involved increases. Since the fully MF simulation (Figure 3.1) neglects this travel time, it becomes increasingly inaccurate as the coverage drops. As θ_A drops, ξ increases, reflecting that larger and larger diffusion times are being neglected by the infinite-diffusion assumption of the MF model. At the limit of $\theta_A \rightarrow 0$, $\xi \rightarrow \infty$.

As θ_A increases, reactants have less distance to travel and MF inaccuracies drop. Notably, L's relevance to transport is eliminated at a critical packing density ($\theta_A = \theta^{Crit} = \sim 0.3$). At this density, the lattice is packed such that the coverage of the nearest neighbors of each reactant approximates the average concentration at all times. For example, every site at θ^{Crit} has a ~75% chance of at least one adjacent reactant, assuming a random distribution. Alternatively, this can be thought of as a state where the availability of reacting neighbors is not sensitive to position, and the variation in availability from site to site is not extreme. Spatial position is much more important at $\theta_A < \theta^{Crit}$ where there are few sites which offer a reacting neighbor. Once the spatial location of a reactant is not important, the rate of new lattice arrangements (the diffusion rate) is not relevant. ξ becomes insensitive to Da at $\theta_A > \theta^{Crit} = 0.3$ for all CG levels (Figure 3.2-Figure 3.1).

For $\theta_A < \theta^{Crit}$, the lattice is sufficiently vacant such that L becomes relevant, diffusion is required to bring together reactants, and the value of the Da number impacts the accuracy of the simulation. As noted above, this is most clearly shown in the results of the single cell (fully MF simulation; Figure 3.1). This is also observed in the CGMC models (Figure 3.2-Figure 3.5). Within a single CG cell, the MF approximation is applied and travel time within the cell is negligible. Small transport distances are neglected by the local MF approximation, causing $\xi \neq 1$. Unlike the fully MF model, transport from one CG cell to a neighboring cell *is* accounted for and because of this the CGMC method is always more

accurate than a single cell (the fully MF) method. Additionally, the CG cell to cell hops capture large diffusion distances accurately. At the limit of low coverage, the relatively small distance within each cell neglected by local MF assumption is an insignificant fraction of the overall diffusive journey of a reactant. Therefore, $\xi \rightarrow 1$ as $\theta_A \rightarrow 0$ and $L \rightarrow \infty$ for all CGMC models.

The CGMC method is accurate ($\xi \rightarrow 1$) as $\theta_A \rightarrow 0$ and as $\theta_A \rightarrow \theta^{Crit}$, but between these two limits $\xi \neq 1$. This implies a maximum in ξ between $\theta_A=0$ and $\theta_A= \theta^{Crit}$ (observed in Figure 3.2-Figure 3.5) at $\theta_A = \theta^L$, where the diffusive transport transitions from primarily short distances to long distances relative to the characteristic length of CGing, L_{CG}. At $\theta = \theta^L$, L is on the order of L_{CG} (shown in Table 3.3). As the coverage is increased toward the critical packing coverage, ($\theta^L < \theta_A < \theta^{Crit}$; $L < L_{CG}$), the model is increasingly accurate ($\xi \rightarrow 1$) since the transport distances being neglected by local MF assumptions are shrinking. As the coverage is decreased toward low coverage ($0 < \theta_A < \theta^L$; $L > L_{CG}$), the model is increasingly accurate ($\xi \rightarrow 1$) as the long distance transport time is properly accounted for by CG cell to cell hops.

Table 3.3 – Comparison of the transport distance at the observed maximum error to the characteristic length scale of CGMC for the AA system.

q	θ^L (observed)	$\frac{L_{AA}}{s} \cong \sqrt{\frac{1}{\theta_A}}$	$\frac{L_{CG}}{s} = \sqrt{q}$
4	0.0025	6.3	2
25	0.013	7.5	5
100	0.018	8.8	10

625	0.025	20.0	25
∞ (MF)	→0	∞	∞

In summary, when densely packed ($\theta_A > \theta^{Crit}$), transport becomes irrelevant and all simulation methods are accurate. As θ_A is reduced below θ^{Crit} the average diffusion distance increases, and the MF assumption leads to larger inaccuracies. While the MF model's inaccuracies reach a maximum as $\theta_A \rightarrow 0$, the CGMC models have a maximum error at some $0 < \theta^L < \theta^{crit}$. θ^L is related to the average diffusion length, L, and the characteristic CG length, L_{CG}. As $\theta_A \rightarrow 0$ for $\theta_A < \theta^L$, the CGMC simulation converges on the KMC solution since the long diffusion distances are properly accounted for by the coarse diffusion hop.

Results of AB system

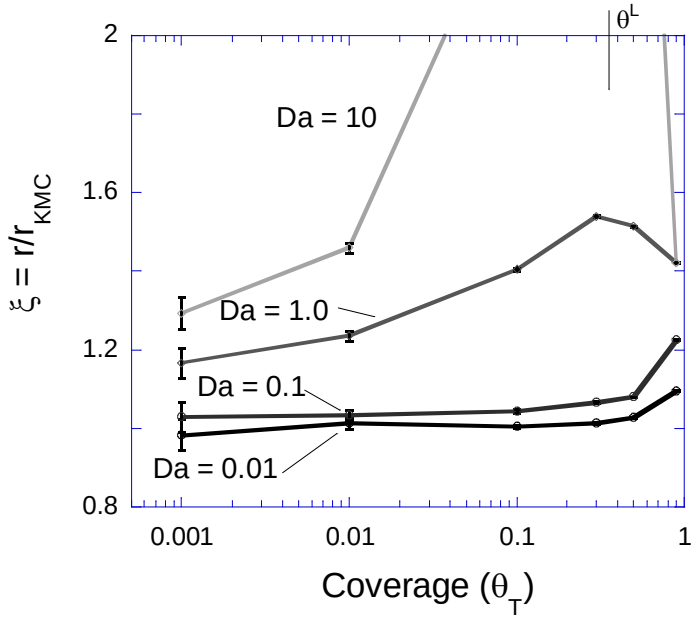

Figure 3.6 - Error of CGMC method for AB system (Table 3.2); q = 4.

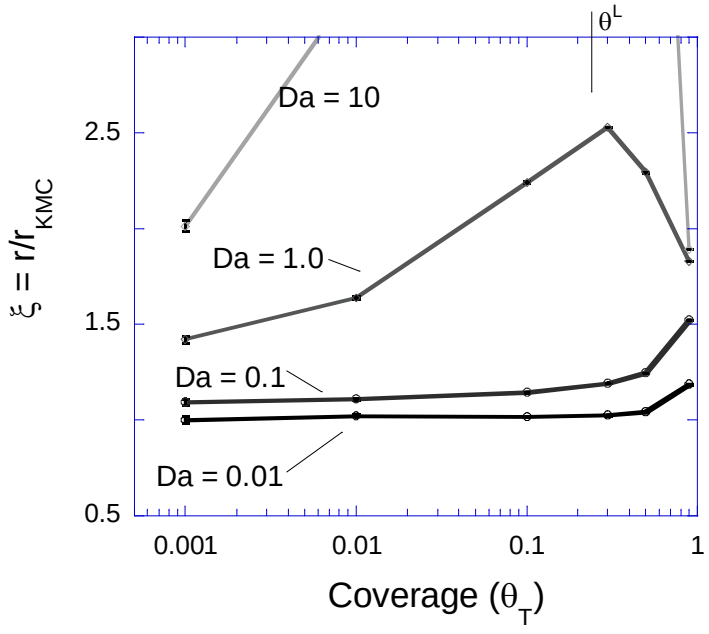

Figure 3.7– Error of CGMC method for AB system (Table 3.2); q = 25.

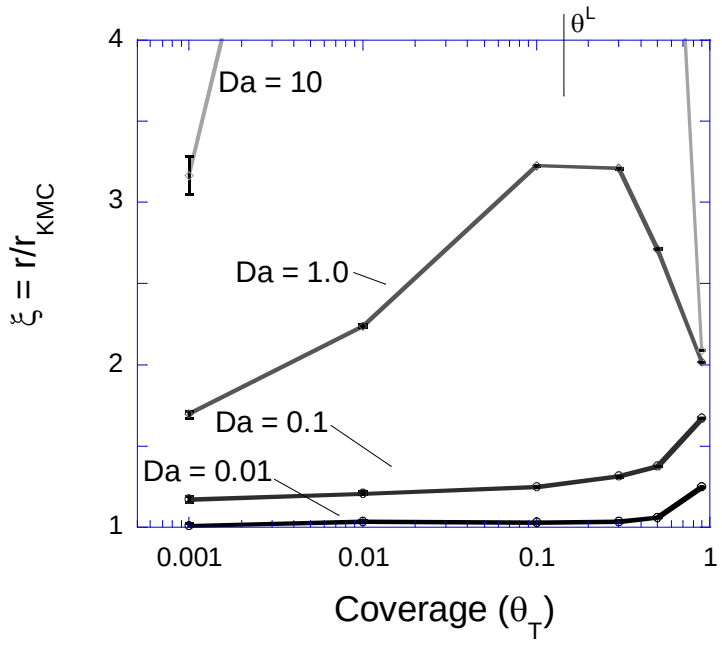

Figure 3.8– Error of CGMC method for AB system (Table 3.2); q = 100.

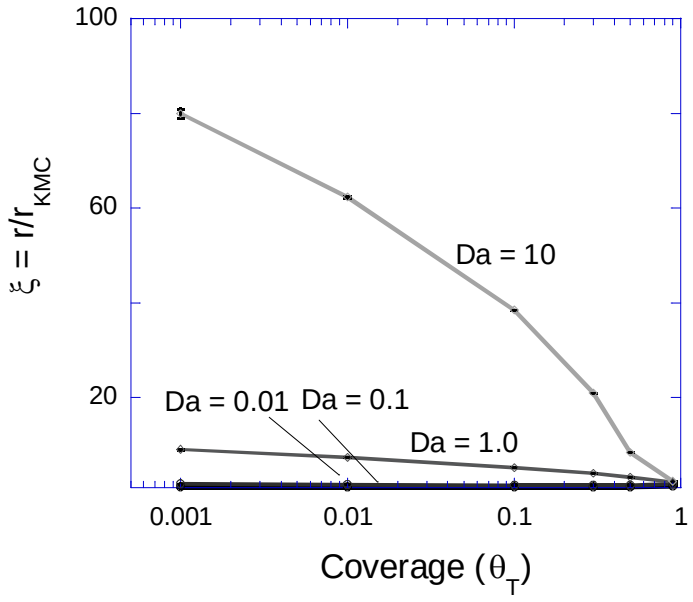

Figure 3.9– Error of a single cell (Gillespie) CGMC method for AB system (Table 3.2).

The AB reaction system is sensitive to Da at all levels of lattice coverage, including high lattice coverage. This indicates that diffusive transport is significant even when reactants are very closely packed in a mean field sense (unlike the AA reaction-diffusion system). Analogous to the AA system, the CGMC methods (Figure 3.6-Figure 3.8) have a maximum in ξ for Da = 1.0 and 10 corresponding to neglected transport distance. However, Da = 0.1 and 0.01 do not show the same distinctive maximum but errors are small due to the systems being nearly kinetically controlled.

As θ_T approaches the upper limit in Figure 3.6-Figure 3.9 ($\theta_T = 0.9$), the error is increasing for lower Da values and decreasing for higher Da values. This indicates that the reactants are now close enough together to not rely on long-distance transport but now have some amount of short-scale crowding resistance to transport. This reordering resistance is characterized by a tightly packed lattice where *mixing* of the reactants, not transport distance, becomes the rate limiting transport step. In our AB system, neighbors may be nonreactive (AA and BB pairs), unlike the AA system. Once the lattice is sufficiently crowded, transport time is spent shuffling the lattice on a short length scale. Local and global MF assumptions neglect this ordering of the lattice and overestimate the efficiency of mixing of the reactants. Even the smallest of CG cells (Figure 3.6, q = 4) miss this organization to some extent, indicating that the relevant spatial detail is on a very short length scale. It is possible that as $\theta_T \rightarrow 1$, the organization of the lattice will increase in length scale, resulting in behavior that cannot be extrapolated from Figure 3.6-Figure 3.8.

At $\theta_T << 1$, diffusion distance between reactants on the mostly vacant surface is the primary transport resistance. Accuracy in this regime is sensitive to Da and θ_T in the same manner as the AA system. Neglected transport distance, quantified by L in Eq. (3.12), is the dominant reason for inaccuracies of the MF approximation. As $\theta_T \rightarrow 0$, the fully MF

method of Figure 3.9 becomes increasing inaccurate due to neglecting larger and larger L, while the CGMC method, Figure 3.6-Figure 3.8, performs more accurately since large transport distances (relative to the size of CG cells) are well captured. At high Da, the CGMC method exhibits a local maximum in error at $\theta_T = \theta^L$. θ^L is roughly where $L \approx L_{CG}$, the characteristic length of coarse graining (Table 3.4).

Table 3.4 – Comparison of the transport distance at the observed maximum error to the characteristic length scale of CGMC for the AB system.

q	θ^L (observed)	$\frac{L_{AB}}{s} \cong \sqrt{\frac{4}{\theta_T}}$	$\frac{L_{CG}}{s} = \sqrt{q}$
4	0.35	3.4	2
25	0.25	4.0	5
100	0.15	5.2	10
∞ (MF)	→0	∞	∞

Recommendations:

For the simulation of a single species (AA system) only the distance between reactants must be considered when comparing KMC, CGMC and fully MF methods. If Da < 0.01 diffusion relative to reaction is fast enough that a single cell (fully MF) approach is appropriate at any coverage. The exception may be at extreme coverage ($\theta_A \approx 0$ and $\theta_A \approx 1$) where ξ_{MF} tends to diverge. If Da > 0.01, then diffusion is slow enough relative to reaction

that other methods should be utilized. At high-densities, $\theta_T > \theta^{Crit}$, transport distance is not significant and a fully MF method is appropriate. For low-densities, $\theta_T < \theta^{Crit}$, where the transport distance (L) between reactants is large, the fully MF model is no longer accurate. If $L >> L_{CG}$, then CGMC accurately describes the system and should be applied to reduce computational overhead. However, if $L \approx L_{CG}$ or $L < L_{CG}$, CGMC is too coarse and KMC will need to be used.

For the simulation of multiple species (AB system), the distance between reactants and the order of the lattice must be considered when comparing KMC, CGMC and single cell (fully MF) methods. If $Da \leq 0.01$ and $\theta_T < 0.3$ (Figure 3.9), then we may assume diffusion is fast enough relative to reaction that the fully MF method does well. At extremely low coverage ($\theta_T \approx 0$), the increased transport distance will require $Da << 0.01$ for fully MF methods to be correct. At this extreme, the CGMC method should be applied as it handles low coverages very well. Otherwise, the CGMC method is accurate for $Da \leq 0.1$, but does not produce very accurate reaction rates for $Da \geq 1.0$. Without improvements to the CGMC method, the KMC method appears be the method of choice for $Da \geq 1.0$. Also, at high coverage ($\theta_T > 0.3$), the short length scale detail of the lattice demands the high resolution of the KMC method.

Improvements to CGMC:

To compensate for an expected inaccuracy of the CGMC method, an 'effective' reaction rate constant could be calculated based on the macroscopic parameters of the simulation (θ and Da). The corrections as functions of the macroscopic parameters would ideally be known a priori but could also be simulated with short KMC simulations and recorded during the CG simulation. This effective rate constant would then be used in a

CGMC simulation to deliver the computational saving of the CGMC method with little loss of accuracy. To expand the effectiveness of this technique, individual CG cells could automatically adjust the local reaction rate constant based on local conditions. This would allow highly heterogeneous systems to operate on a single lattice.

All reaction-diffusion systems presented here were spatially homogeneous with all lattice nodes identical, and uniform and homogeneous CG meshes were applied. However, in a spatially heterogeneous system, a non-uniform CG mesh might be ideal. If one region demands the detail of KMC while other areas suggest CGMC, a hybrid of the two methods could conceivable be constructed. This 'adaptive' coarse-grain meshing would bring together the speed of CGMC where it is appropriate with the detail of KMC where it is needed. Conceivably, this would be applied to non-uniform lattices, where the location of high detail can be predicted. For systems like the high-density AB system, where clusters of single species form in unpredictable locations, a CGing method that applies adaptive meshes on high-detail areas as they form is conceivable. This on the fly adaptive CGing would need to detect local concentrations of reactants and automatically rearrange the lattice around them.

Application to real-world systems:

The results of this study of model systems can be used to suggest approaches to other diffusion-reaction systems. In the EGFR diffusion-reaction system of [13, 14] protein receptors participate in bimolecular dimerization reactions while in monomer form and do not while in dimer form. Because both species occupy the same lattice positions, dimers may block monomer-monomer collisions in high-density situations, which have been experimentally shown to exist on the membrane. Compared to the AA and AB systems

presented here, the EGFR system is a sort of hybrid: An AA system (monomer-monomer) with a second, obstructive lattice species (dimers), as seen in the AB system.

Assuming that diffusion control will be significant for the EGFR system (Da > 0.01) and considering the high and low density domains, we expect fully MF methods will be inaccurate (this has been demonstrated in [13]). When organization of the lattice species is important (high density areas where dimers will block monomers, the results of the AB system suggest that a high-resolution method should be applied. For low-density regions with $L < L_{CG}$, the AB and AA systems suggest that lower resolution is appropriate. This suggests an adaptive mesh approach: the clustered areas are simulated with highly detailed KMC while the low density areas and simulated using a coarse lattice.

REFERENCES

1. Chatterjee, A., D.G. Vlachos, and M.A. Katsoulakis, *Spatially adaptive lattice coarse-grained Monte Carlo simulations for diffusion of interacting molecules.* Journal of Chemical Physics, 2004. **121**(22): p. 11420-11431.
2. Collins, S.D., A. Chatterjee, and D.G. Vlachos, *Coarse-grained kinetic Monte Carlo models: Complex lattices, multicomponent systems, and homogenization at the stochastic level.* Journal of Chemical Physics, 2008. **129**(18).
3. Bhalla, U.S. and R. Iyengar, *Emergent properties of networks of biological signaling pathways.* Science, 1999. **283**: p. 381-387.
4. Asthagiri, A.R. and D.A. Lauffenburger, *Bioengineering Models of Cell Signaling.* Annual Review of Biomedical Engineering, 2000. **2**(1): p. 31-53.
5. Breitling, R. and D. Hoeller, *Current challenges in quantitative modeling of epidermal growth factor signaling.* Febs Letters, 2005. **579**(28): p. 6289-6294.
6. Orton, R.J., et al., *Computational modelling reveals feedback redundancy within the epidermal growth factor receptor/extracellular-signal regulated kinase signalling pathway.* Iet Systems Biology, 2008. **2**(4): p. 173-183.
7. Wiley, H.S., S.Y. Shvartsman, and D.A. Lauffenburger, *Computational modeling of the EGF-receptor system: a paradigm for systems biology.* Trends in Cell Biology, 2003. **13**(1): p. 43-50.
8. Bray, D., *Signaling complexes: Biophysical constraints on intracellular communication.* Annual Review of Biophysics and Biomolecular Structure, 1998. **27**: p. 59-75.
9. Ehrenberg, M., et al., *Systems biology is taking off.* Genome Research, 2003. **13**(11): p. 2377-2380.
10. Takahashi, K., S.N.V. Arjunan, and M. Tomita, *Space in systems biology of signaling pathways - towards intracellular molecular crowding in silico.* Febs Letters, 2005. **579**(8): p. 1783-1788.
11. Mayawala, K., D.G. Vlachos, and J.S. Edwards, *Spatial modeling of dimerization reaction dynamics in the plasma membrane: Monte Carlo vs. continuum differential equations.* Biophys. Chem., 2006. **121**: p. 194–208.
12. Torney, D.C. and H.M. McConnell, *Diffusion-limited reaction rate theory for two-dimensional systems.* proceedings of the Royal Society of London A, 1983. **387**: p. 147-170.
13. Mayawala, K., D.G. Vlachos, and B.J. Edwards, *Heterogeneities in EGF receptor density at the cell surface can lead to concave up Scatchard plot of EGF binding.* FEBS Letters, 2005. **579**: p. 3043-3047.
14. Niehaus, A.M.S., *Simulations of Epidermal Growth Factor Receptor Dynamics on Corralled Membrane Surfaces*, in *Department of Chemical Engineering*. 2007, University of Delaware: Newark, DE, USA. p. 124.

CHAPTER 4

APPLICATION OF ADAPTIVE COARSE GRAINED MONTE CARLO SIMULATION TO HETEROGENEOUS PLASMA MEMBRANES

Introduction

The Epidermal Growth Factor (EGF) receptor (EGFR) is a well-studied member of the ErbB family of receptor tyrosine kinases (RTKs), which are involved in cell fate decisions and are implicated in numerous human cancers [1]. Like other RTKs, EGF activates its receptor by altering the receptor's conformation and removing steric hindrances that prevent dimerization [2]. Upon activation, EGFR forms high-density membrane clusters presumably to amplify intra-cellular signaling and stimulate endocytosis [3]. A complex signaling cascade within the cell brings the signal to the nucleus for gene expression and signal response. Understanding the mechanisms involved in this entire process, and combining simulations of both the surface and the cytoplasm hold great potential to assist with the design of cancer related pharmaceuticals [4, 5].

The kinetic Monte Carlo (KMC) method is well-suited to simulate the EGFR system. With this method, individual receptor locations and different domains of the membrane are discretely represented, and the spatial heterogeneities in receptor density and membrane environment can easily be captured. Furthermore, tracking of receptor locations allows straightforward comparisons to single particle tracking experiments.

However, membrane protein aggregation covers wide time and space scales rendering the KMC method CPU intensive. A multiscale approach to bridge the separation of scales while preserving the attributes of KMC is needed.

In this chapter, the coarse-grained kinetic Monte Carlo (CGMC) method is applied to investigate the long time and space scale behavior of EGFR on the membrane surface. This model includes receptor diffusion, dimer formation, and membrane diffusion barriers. Additionally, an 'adaptive' coarsening technique is used to improve accuracy of the CGMC simulation in high density areas while keeping CPU cost low.

Methodology

CGMC Simulation Lattice

In order to simulate a small fenced-in high density area within a large low density region, a single rectangular corral of side length equal to 48 nm was placed in the simulation space. This side length is within the generally accepted clatherin pit size of 10-200 nm [6]. The coral represents a potentially 'high density' region of the membrane, which accounts for 4% of the entire simulation box. Thus, for this simulation, the 48x48 nm corral is enclosed in a 240x240 nm simulation box to which periodic boundaries were implemented. A lattice constant (the distance between lattice sites) of a = 6 nm was chosen following [7] for which the whole simulation space consists of 40x40 (1600) sites, and the corral of 8x8 (64) sites.

To initialize the simulation, a group of concentrated monomers is placed in the corral. This initial placement means most reactions will happen at short time scales in the high-density corralled region, with few in the extra-corral space.

We employ three on-lattice MC layouts: (1) A traditional **KMC** simulation, whereby all microscopic sites are fully resolved. (2) A Uniform CGMC (**UCGMC**) simulation, whereby the entire simulation space including the central corral is represented as 25 coarse-grained cells of size 8x8 sites (48 nm x 48 nm). (3) An Adaptive CGMC (**ACGMC**) simulation. In this approach, an adaptive mesh is applied. The central corral region is microscopically resolved like a KMC simulation, while the rest of the lattice is uniformly coarse-grained into 8x8 (48nm x 48nm) CG cells. This multiscale approach attempts to combine the efficiency of the uniform CGMC method with the accuracy of the KMC method.

KMC simulations have produced informative results [8], but are limited to a relatively short range of time scales. The necessity of simulating fundamental EGFR processes (diffusion, dimerization, etc), combined with the need to observe long-scale structures (clusters), challenge the capabilities of KMC simulations. Uniform mesh CGMC simulations have been shown to accelerate simulations of systems with a wide range of time and space scales [9, 10]. However, we expect spatial detail of the corral region to be critical, because the rate of monomer dimerization is on the order of the rate of monomer fence jumps (Table 4.2). Thus, we expect the UCGMC to inaccurately simulate the proper ratio of reactions to fence jumps in the corralled region. With the ACGMC method we overcome this problem by using a fully refined lattice on the high-reaction-density intra-corral region. In choosing a level of coarse graining of the extracorral region we consult the results of Chapter 3. The rate of diffusion is roughly 3 orders of magnitude faster than monomer dimerization ($Da < 10^{-3}$), and the concentration of monomers is expected to be low ($\theta < 0.01$). Spatial detail is therefore not important and any level of coarse graining is expected to return an accurate rate of reaction. For convenience, a CG cell size of 8x8 is chosen for the extracorral region.

CGMC Simulation for Corralled Membranes

In the CGMC method, the diffusion equations [10, 11] are accurate between coarse grained cells of different sizes when there is a single time scale of diffusion. However, these equations cannot be directly applied to the special case of corrals separated by fences, where the hop over the interface is slower than the intracorral diffusion rate. Both the fence and intracorral diffusion rate contribute to the effective diffusion rate between two cells. The effective diffusion rate between two cells separated by a single fence, $\Gamma_{effective}$, as a function of the fence and intracorral diffusion rates was formulated and tested in [12].

$$\boldsymbol{\Gamma_{effective} = \frac{\frac{1}{2}\Gamma_{IC}\Gamma_{F}}{\Gamma_{IC}L_{CG}+\Gamma_{F}L_{CG}(L_{CG}-1)}} \tag{4.1}$$

Here Γ_{IC} and Γ_{F} are the intracorral and fence diffusion rates, respectively, for a single microscopic site jump. L_{CG} is the one-dimensional CG cell center-to-center distance of the two relevant coarse-grained cells perpendicular to the cell boundary over which diffusion occurs. Eq. (4.1) was applied to corralled membrane EGFR simulations to test the accuracy of CGMC for the case of pure diffusion. Intracorral and fence diffusion occur for the single species on the lattice (Table 4.1). The simulation starts with concentrated receptors in the corralled area. Over time, the receptors diffuse out of the corral.

Table 4.1 – Diffusion model in CGMC simulations. Notation: EGF receptor (R), lattice vacancy (*), intracorral (X1), extracorral (X2)

Description	Mechanism	Rate Constant
Intracorral Diffusion	$R_1 + *_1 \rightarrow *_1 + R_1$	$2.50\ 10^5\ s^{-1}$
Extracorral Diffusion	$R_2 + *_2 \rightarrow *_2 + R_2$	$2.50\ 10^5\ s^{-1}$
Fence Diffusion	$R_1 + *_2 \leftrightarrow *_2 + R_1$	$2.50\ 10^2\ s^{-1}$

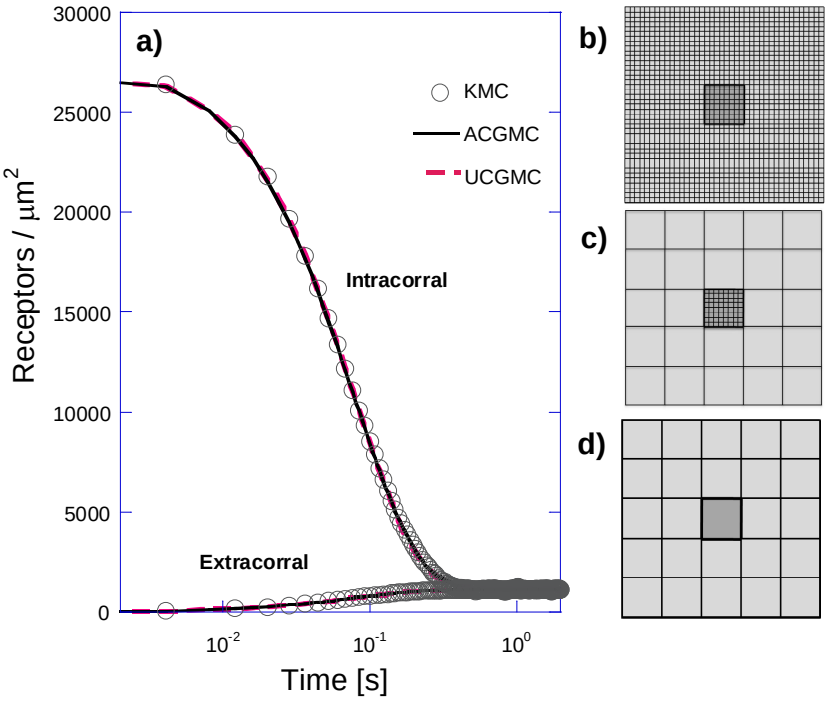

Figure 4.1 – Demonstration of effective diffusion in ACGMC and UCGMC simulation methods. The central 48 nm corral starts fully covered with a local concentration of ~26000 Receptors/µm^2. A fence barrier separates the intracorral area from the extracorral area. KMC (b), ACGMC (Adaptive) (c), and UCGMC (Uniform) (d) layouts. Both coarse-grained simulations perform accurately.

The ACGMC and UCGMC methods produce the same result as the traditional KMC method (Figure 4.1), confirming that Eq. (4.1) correctly describes the effective diffusion rate. Additionally, this shows that the CGMC method correctly handles diffusion for spatially heterogeneous systems with high-density areas separated by low-density areas (in the absence of intermolecular forces between receptors). This allows us to attribute CGMC errors in later high-density simulations to reactions.

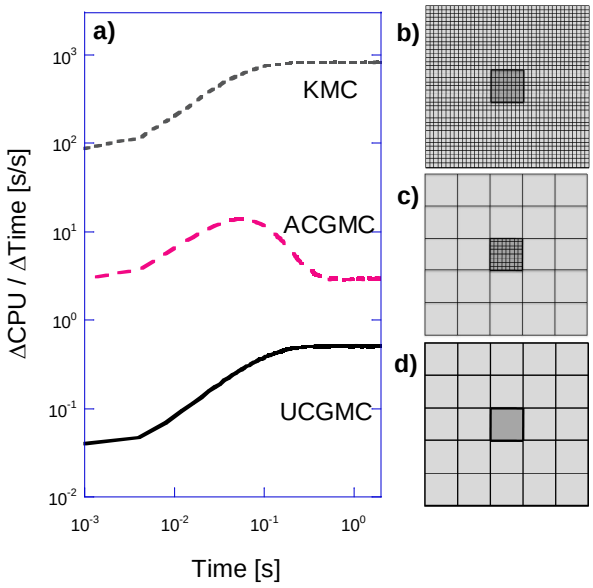

Figure 4.2 – CPU cost comparison of the KMC, ACGMC, and UCGMC methods in a diffusion-only system with 48 nm corral. This plot shows the instantaneous ratio of CPU time to simulated time. This cost ratio shifts differently for each CG layout and as the simulation progresses.

With regard to simulation cost, UCGMC simulations are cheaper than KMC by over three orders of magnitude, whereas the ACGMC method is cheaper by 0.5-2.5 orders of magnitude (Figure 4.2). The efficiency of the ACGMC method is especially pronounced at longer times when the majority of receptors have jumped out of the intracorral region. The two CGMC simulations are faster because they only simulate large coarse hops in the extracorral region, whereas the KMC simulation resolves all microscopic moves.

While the KMC and UCGMC simulations increase in cost until reaching a steady state, the ACGMC simulation reaches a maximum cost at a time that corresponds to the corral being ~1/2 covered (initially the corral is fully covered). At this point the number of molecules and vacancies on the fine-grid corral is equal, leading to the maximum frequency of expensive micro-diffusion hops. As receptors leave the corral and the coverage of the

corral falls below 0.5, the ACGMC cost drops by an order of magnitude. This is because there are fewer receptors in the expensive fine-grid corral, and more receptors in the coarse-grained space outside the corral.

Short Time EGFR Simulations

The previous section demonstrated that the CGMC simulations provide the same result as the KMC simulation at a much reduced cost for a diffusion only system. We expect similar CPU savings when extended to reacting systems, but the accuracy of the simulation comes into question. In the following, we investigate the performance and accuracy of the CGMC method for diffusion-reaction systems.

The reaction-diffusion model for the EGFR system is shown in Table 4.2. The rates are taken from Table 4.3 of [12] and represent a diffusion-controlled system. The diffusion of the dimer is taken to be half of that of the monomer. The diffusion of monomers over a fence is three orders of magnitude slower compared to the intra and extracorral monomer diffusion, whereas that of dimers is four orders of magnitude slower than the intra and extracorral dimer diffusion. All receptors start in the corral in monomer form.

Table 4.2 – EFGR diffusion-reaction model.

Notation: EGF receptor monomer (M), dimer (D), lattice vacancy (*), intracorral (X_1), extracorral (X_2). Dimers occupy a single lattice site. Monomerization and dimerization reactions do not occur over a fence.

Description	Mechanism	Rate Constant
Diffusion:		
Extracorral (M)	$M_1 + *_1 \rightarrow *_1 + M_1$	$2.50\ 10^{5}\ s^{-1}$
Intracorral (M)	$M_2 + *_2 \rightarrow *_2 + M_2$	$2.50\ 10^{5}\ s^{-1}$
Extracorral (D)	$D_1 + *_1 \rightarrow *_1 + D_1$	$1.25\ 10^{5}\ s^{-1}$
Inracorral (D)	$D_2 + *_2 \rightarrow *_2 + D_2$	$1.25\ 10^{5}\ s^{-1}$
Fence (M)	$M_1 + *_2 \leftrightarrow *_1 + M_2$	$2.50\ 10^{2}\ s^{-1}$
Fence (D)	$D_1 + *_2 \leftrightarrow *_1 + D_2$	$1.25\ 10^{-1}\ s^{-1}$
Reactions:		
Monomerization	$D + * \rightarrow M + M$	$1.70\ 10^{-2}\ s^{-1}$
Dimerization	$M + M \rightarrow D + *$	$5.67\ 10^{2}\ s^{-1}$

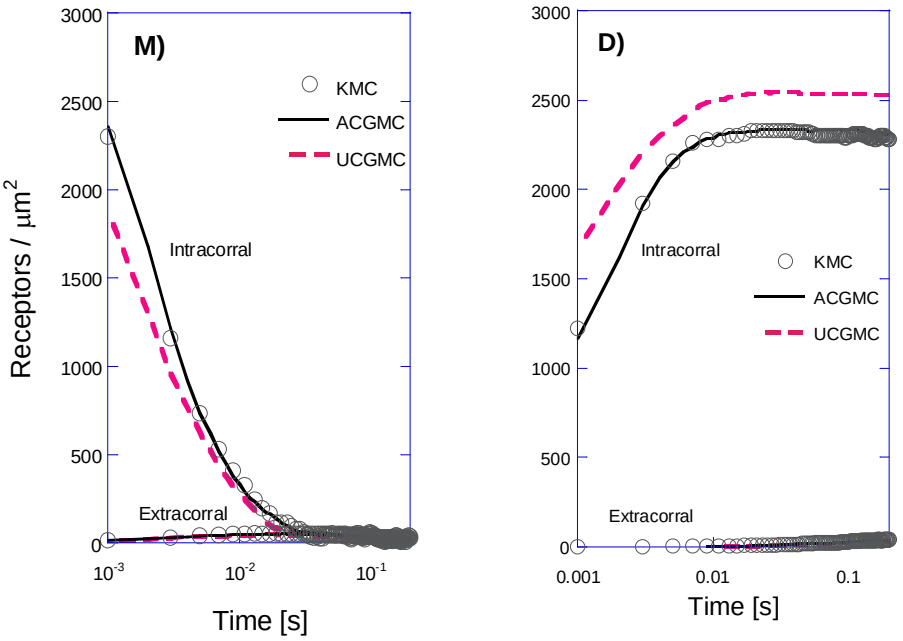

Figure 4.3 –Short-time (t = 1 – 100 ms) density of receptors in monomer (M) and dimer (D) form. Overall density of 150 receptors/µm^2, kinetic rates from Table 4.2.

Typical results for the diffusion –reaction model are shown in Figure 4.3. Notable is the error made by the UCGMC method, and the accuracy of the ACGMC method. This error is due to the reactions happening inside the corral, which the ACGMC simulation handles accurately. At very short times (Figure 4.3) the dimerization reaction and monomer fence jumps dominate. As a result, a fine grid is necessary for accurate simulations. At higher densities (the corral region) the fine-mesh of KMC and the ACGMC method results in accurate but expensive simulation. At low-densities (extracorral regions), the coarse mesh of ACGMC and UCGMC simulations leads to accurate results with low computational intensity.

Looking beyond simulation accuracy to the physical behavior of the system, we see that corrals can maintain clusters of receptors by stabilizing the dimer form inside a corral. Initially, all receptors are in the monomer form and within a short time (t < 0.01 s),

they either leave the corral or dimerize. Dimers have such a reduced rate of fence hopping that their formation essentially locks the dimerized receptors inside the corral. This highly concentrated area slowly dissipates via two slow mechanisms: i) dimers dissociate and the resulting monomers jump the fence before associating again; ii) receptors jump the fence in dimer form. Counting of jumping events shows that the contributions of both mechanisms (i) and (ii) mechanisms are on the same order of magnitude; yet, the contribution of mechanism (i) is more significant under our conditions.

Long Time EGFR Simulations

The short-time simulations of Figure 4.3 show only the creation of receptor clustering. To calculate the lifetime of these clusters and the factors controlling cluster longevity, long simulations were performed.

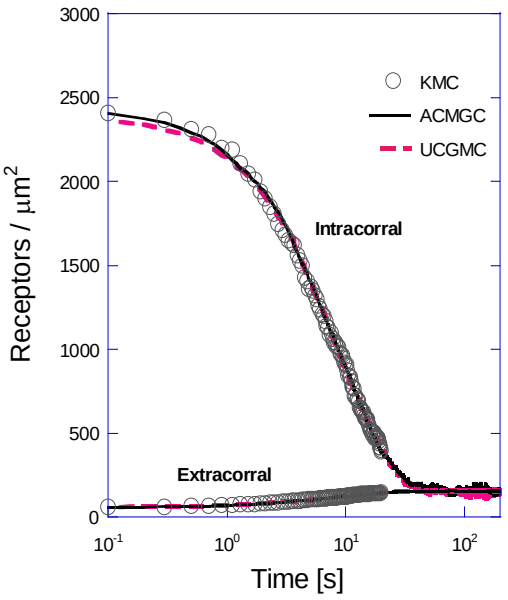

Figure 4.4 – Long-time (t = 0.1 – 200 s) profile of receptor density. Overall density of 150 Receptors/μm^2 and rates from Table 4.2. Dimer and monomers counts are combined.

Thus, the timecourses of receptor density simulated with the ACGMC and UCGMC methods (Figure 7) show that monomer coverage reaches quasi-steady-state (QSS) for t < 1 s, due to monomer fence diffusion between intracorral and extracorral regions.. On the other hand, dimer fence jumping and dimer dissociation is slow, and thus, the dimers are kinetically held in the corral and do not relax to a uniform density until t = ~30 s. Due to this effect, the ACGMC method becomes more efficient as the kinetically held clusters dissipate (fewer receptors in the expensive high-resolution corral) around t = 1 s (Figure 8). This simulation demonstrates high-density spatial receptor heterogeneity of receptors persisting on the order of seconds.

KMC is expensive to run at this timescale; as a result, KMC comparisons with the coarse-grained simulations were done up to only 20 s. These comparisons reveal that the ACGMC and UCGMC methods produce results that are in excellent agreement with those of the KMC simulation (Figure 4.4), and are able to easily reach the final steady state concentrations in reasonable CPU time (Figure 4.5). The accurate ACGMC simulation reduces the cost of simulation by 2-3 orders of magnitude (Figure 8) allowing us to obtain accurate statistics and perform a sensitivity analysis.

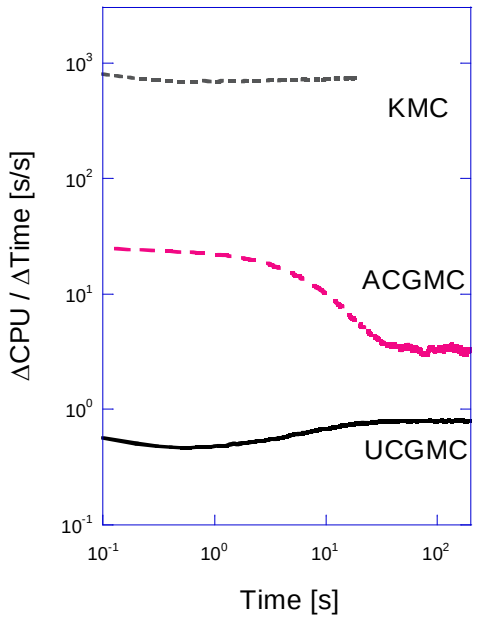

Figure 4.5 – Long-time (t = 0.1 – 200 s) CPU cost comparison of the KMC, ACGMC, and UCGMC methods in the reaction-diffusion system of Table 4.2 with a 48 nm corral. This plot shows the instantaneous ratio of CPU time to simulated time. This cost ratio shifts differently for each CG layout and as the simulation progresses. This plot begins approximately at the end time of Figure 4.2. The ACGMC method shows an additional gain in efficiency once the corralled cluster dissipates between 10^1 and 10^2 s. KMC simulations were only run to 20 s.

Hopkin's Test Statistic

It is common in the MC literature to define clusters based on a first- or second-nearest neighbor criterion. However, the exact spatial location of some molecules is not known in CGMC and this criterion cannot be easily implemented. Instead we employ a technique commonly applied to experimental microscopy images to test for clustering, the Hopkins spatial statistic [13, 14]. Given a spatial region (a microscopy image or a simulation lattice space), a set of particle coordinates are tested against randomly generated points for their proximity to other particles. A homogenous Poisson process has an expected value of

½, whereas clustered particles will return a value closer to 1. This procedure is repeated to construct a histogram which is compared to the expected distribution.

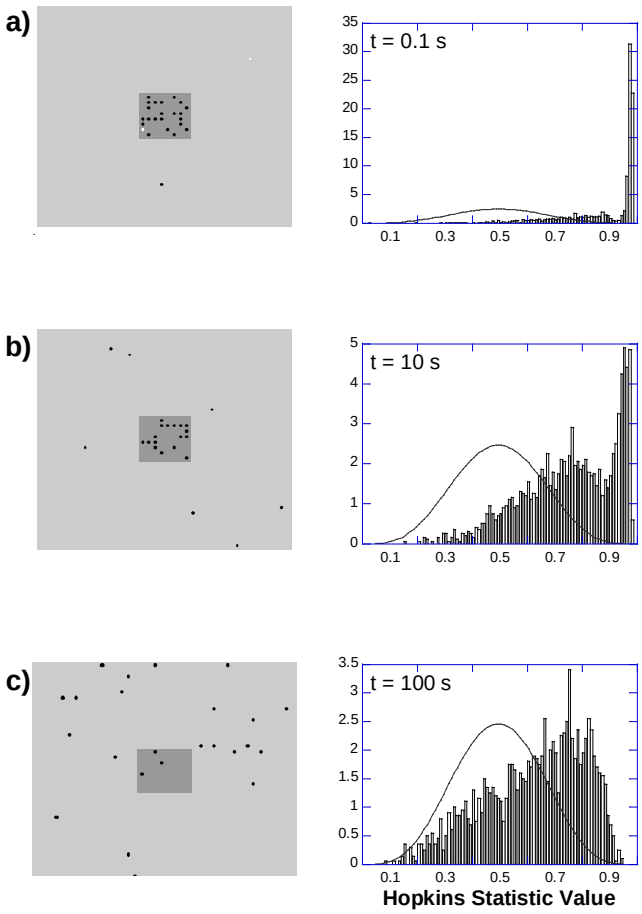

Figure 4.6 – Snapshots of receptor location and corresponding Hopkins statistic distributions. The distribution tends toward the expected value of ½ for a random distribution (shown by the solid line) as time goes on. Three times are shown from the simulation presented in Figure 4.4. White dots represent monomers, while black dots represent dimers.

Shown in Figure 4.6 are snapshots of receptor locations next to their corresponding Hopkins statistic distributions for three different times during the simulation described above for Figure 4.4 . The dissipation of the intracorral cluster can be clearly seen as the average value of the Hopkins statistic moves toward ½. Figure 4.6a shows distribution

heavily shifted to the right, indicating a high degree of clustering. Figure 4.4 indicates that the intracorral and extracorral receptor concentrations reach equilibrium at t < 100 s, which is seen in Figure 4.6c (t = 100 s), where the Hopkins distribution shows a very weak spatial correlation of the receptors.

Sensitivity Analysis

In order to investigate the dominant mechanisms controlling the properties of receptor clusters, a sensitivity analysis of the diffusion-reaction model of Table 4.2 was performed.

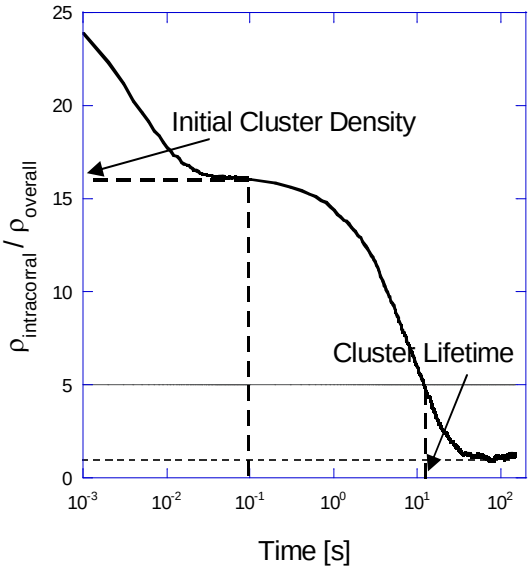

Figure 4.7 - Comparison metrics for clustering simulation.

To facilitate our analysis, we used two metrics, noted in Figure 4.7, which shows an example plot of the intracorral density of receptors (based on a weighted sum of monomers and dimers) normalized by the overall density ($\rho_{intracorral} / \rho_{overall}$) vs. time. These

metrics are: i) initial cluster density and ii) cluster lifetime. The initial cluster density is defined as $\rho_{intracorral} / \rho_{overall}$ at t = 0.1 s. At this time, the initial monomers have either dimerized or left the corral (Figure 4.3). This metric can also be thought of as the effectiveness of monomer trapping by the corral when dimerization partners are readily available. The cluster lifetime is defined as the time at which the concentration of receptors drops below 5 times the overall receptor density (namely $\rho_{intracorral} / \rho_{overall} = 5$). The cluster lifetime illustrates how effective the corral is at stabilizing receptors in the dimer form.

These metrics were calculated for a range of values for the following variables: corral size, overall receptors density, and dimer fence jumping propensity. The corral sizes (48 nm and 24 nm) were chosen within the observed 10-200 nm size range of clatherin pits on living cells [6]. The overall density of receptors was also varied since the dramatically different density of receptors in cancerous and normal cells is suspected to play a large role in the dysregulation of cell communication. Finally, dimer fence diffusion was disabled in some simulations to reflect that dimers may cross the fence with an extremely low probability.

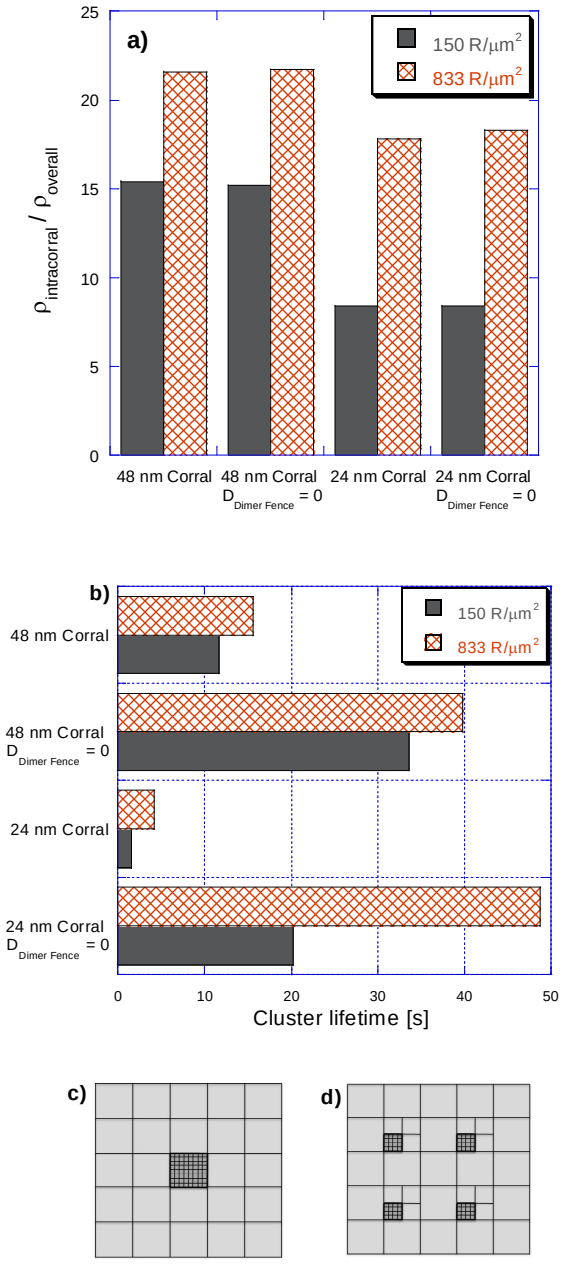

Figure 4.8 – Sensitivity of clustering to monomer-only fence diffusion, different corral sizes, and different overall receptor densities. (a) Initial cluster density ($\rho_{intracorral}$ / $\rho_{overall}$ at t = 0.1 s). (b) Cluster lifetime (time when $\rho_{intracorral}$ / $\rho_{overall}$ = 5). (c) ACGMC layout of 48 nm corral simulations. (d) ACGMC layout of 24 nm corral simulations. All receptors initially start at random locations in corrals.

The results of the sensitivity analysis on these metrics appear in Figure 4.8. The initial cluster density (Figure 4.8a) is most noticeably affected by the density of receptors in the simulation. For fixed corral sizes, higher overall density simulations (833 receptors/μm^2) exhibit more pronounced initial clustering relative to low density simulations. This result can be attributed to the higher receptor density increasing the likely number of dimerization partners adjacent to any given monomer (higher dimerization rate), while not affecting the probability of a monomer to be bordering the corral edge (fixed monomer fence jumping rate). Consequently, a higher proportion of receptors remain in the dimer form as the intracorral density of receptors in increase.

Smaller corral sizes have a higher circumference to area ratio and therefore result in a higher likelihood for a monomer to border a corral edge (increased monomer fence jumping rate). The diffusion time from the center of the corral to the circumference also drops. As the above logic would suggest, these smaller corrals have a lower initial cluster density at the same overall densities as larger corrals.

Disabling dimer fence diffusion (Figure 4.8a) has very little effect on the initial cluster density, reflecting that monomer fence diffusion is effectively the only path by which receptors leave the corral in the very early stages of the simulation.

On the other hand, when dimer diffusion is taken into account, the smaller corrals have a much weaker hold on the receptors (Figure 4.8b). Since the smaller corral has a higher circumference to area ratio, dimers capable of jumping have more contact with the fence and leave the corral at a faster rate than in larger corrals.

Disabling dimer diffusion extends cluster lifetimes (Figure 4.8b) by half an order of magnitude in 48 nm corrals and one order of magnitude in 24 nm corrals. This

disproportionate increase in small corrals is attributed to a higher chance of recombination of dissociated dimers. We assume that past the initial stage ($t > 0.01$ s of Figure 4.3) the probability of more than 2 monomers at a time existing in the corral is negligible (since 2 monomers will most likely dimerize or jump a fence long before another dimer breaks). Smaller corrals (24 nm) will hold the two monomers much closer together than the larger corral (48 nm) giving the monomers a higher chance of association before one monomer jumps a fence.

It appears that in large corrals, the receptor density plays a secondary role in the cluster lifetime (Figure 4.8b). This is in contrast to small corrals, where the density of receptors is a major factor for determining cluster longevity. This suggests that the sensitivity of the cluster lifetime to receptor density is correlated with the size of the corral. At large corral sizes, the corral size itself has a dominant effect on cluster lifetime, whereas at smaller corral sizes the cluster lifetime is primarily a function of receptor density. Manipulating the size of corrals together with the receptor density has a greater overall effect on the dispersion rate of EGFR clusters than changing each variable individually.

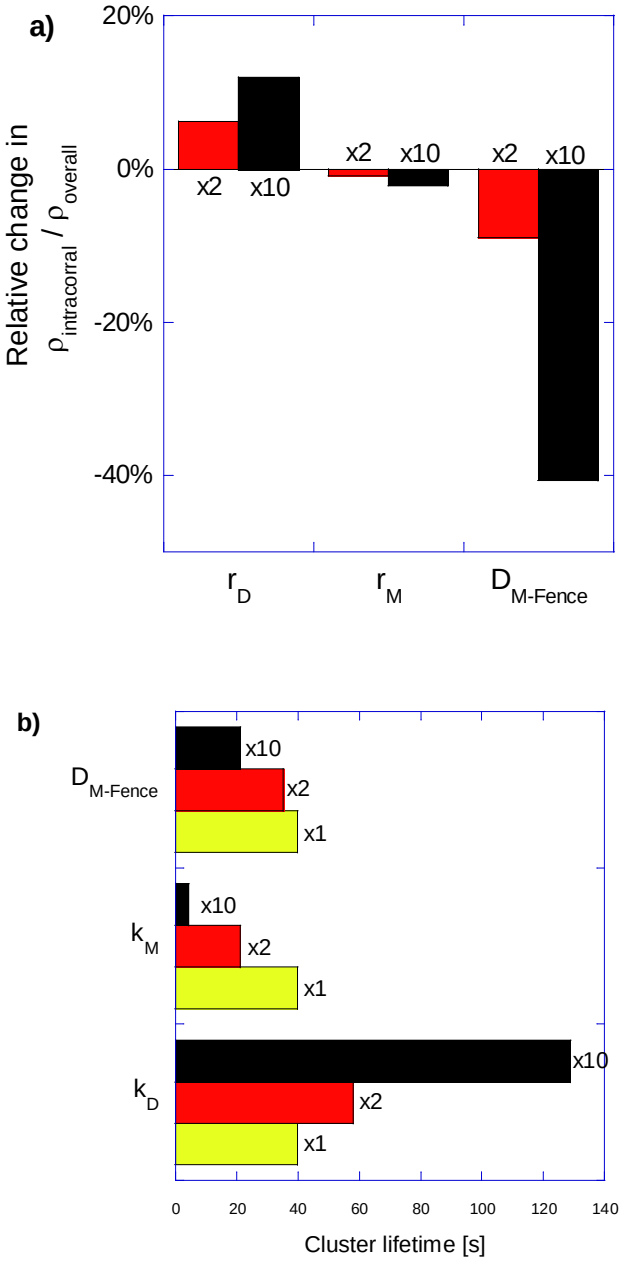

Figure 4.9 – Sensitivity of clustering to reaction and fence diffusion rate constants. (a) Initial cluster density (ρintracorral / ρoverall at t = 0.1 s). (b) Cluster lifetime (time when ρintracorral / ρoverall = 5).

In order to understand the effect of rate constants for monomerization, dimerization, and monomer fence jumping (k_M, k_D, and $D_{M\text{-}Fence}$, respectively), we defined a

new nominal case with an overall receptor density of ~833 receptors/μm^2, disabled dimer fence diffusion, and a corral size of 48 nm (Figure 4.8c). The results of this sensitivity analysis are shown in Figure 4.9. Each rate was increased by a factor of 2 and 10.

It was observed that the initial cluster density is most dramatically affected by increasing the rate at which monomers jump the fence. Initially, all receptors are in monomer form, and thus, an increase in the monomer fence jumping rate significantly decreases the initial cluster density. Similarly, higher dimerization rates result in higher initial cluster densities, since monomers lock into the dimer state faster. The rate of monomerization (dissociation of dimers) has little influence, which is to be expected, since at short times the primary reactions taking place are dimerization and fence jumps.

Moreover, all kinetic rate constants affect the cluster lifetime. Changes to the rate of monomerization and dimerization influence the lifetime more than changes to the monomer fence diffusion rate. Increasing the dimerization rate by an order of magnitude increases the kinetic lifetime of the cluster by about half an order of magnitude and well into the range of minutes. Increasing the rate of monomerization by an order of magnitude shortens the cluster lifetime by approximately an order of magnitude. The increase in the rate of monomerization seems to have a relatively *greater* effect on cluster lifetime than proportional changes to the rate of dimerization.

These results can be explained as follows. Clearly, more time spent in monomer form directly correlates with faster cluster dissolution. This is because two monomers are only capable of associating to form a dimer if they both reside in the corral long enough for a dimerization event (a bimolecular reaction) to occur. If either monomer resulting from dimer disassociation event (a monomolecular reaction) stochastically leaves the corral, the

remaining monomer will most likely leave the corral as well. This causes the overall probability of dimerization of two monomers to be a function both of the rate of monomer fence jumping and the dimerization reaction rate constant, whereas the rate of monomerization is only a function of monomerization reaction rate constant.

Given the uncertainty in kinetic and diffusion rate constants, it is quite possible that regions of high concentration of receptors could kinetically lock clusters over the period of minutes. Eventually, though, in the lack of a thermodynamic stabilization or regeneration mechanism, no clustering will be observed.

Large Scale Simulation

A major incentive to develop CGMC methods is the ability to reach long time and length scales. To demonstrate this, an ACGMC method simulation was performed on a lattice size 1024x1024, representing a membrane section size of ~38 μm^2. A receptor density of 5500 receptors per μm^2 was used following [15], which is a representative value for cancerous cells [16]. Reaction rates from Table 4.2 were used. The corrals covered 14% of the total surface (in contrast to the 4% used in the sensitivity analysis simulations) and were all 96 nm on a side (16x16 lattice sites). In one simulation, 70% of all receptors were initially placed in the corrals, while 50% started in the corrals in a second simulation. The intracorral (extracorral) space was resolved using q=4 (q = 256) CG cells.

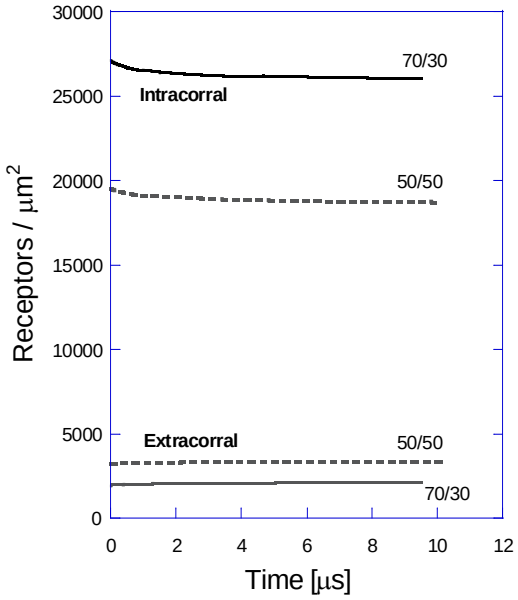

Figure 4.10 – Large length scale simulations

The simulation reached well into the μs timescale within a week of computational time despite the very large size (nearly 3 orders of magnitude increase), 14% high-resolution corral space, and high density of receptors. Figure 4.10 shows the same fundamental behavior as Figure 4.3, namely, an initial leak of receptors and approach to an intracorral and extracorral receptor density QSS, and thus establishes our smaller length scale simulations as representative of a large system. This simulation demonstrates that the ACGMC method is a feasible approach to spatially heterogeneous systems with length and time scales far beyond current KMC methods. While this simulation used a repeating surface corral pattern, it can easily be adapted to for similarly large simulations that incorporate surface feature patterns on the μm length scale.

Conclusions

Adaptive Coarse-Grained kinetic Monte Carlo (CGMC), a multiscale spatial stochastic simulation, was applied to the EGFR diffusion-reaction system. The novel Adaptive CGMC method was applied in order to properly capture the detailed spatial reactions in high receptor density corrals of the membrane while efficiently and accurately simulating the low density areas of the membrane with a low resolution method. Kinetically preserved clusters (as defined by the Hopkins statistic) were observed on the order of minutes. A sensitivity analysis of the density and longevity of these clusters was carried out.

Given the uncertainty in the kinetic parameters of receptor chemistry, kinetic stabilization of EGFR clusters by diffusion fences is quite plausible. EGFR clusters would only need to exist on the order of seconds or less to successfully pass on signals, or for endosomes to form around them.

Our work demonstrates that diffusion barriers of the membrane surface play a significant role in the clustering behavior of EGFR. We have shown that diffusion fences offer a clustering mechanism independent of others that have been theorized in the literature but were not included in this study, such as receptor oligomer formation (tetramers, etc), hydrophobic/hydrophilic membrane regions, and partner switching. In view of our results, insights into the effective treatment of cancerous cells exhibiting high receptor densities could be obtained by focusing on methods to break, strengthen, reorganize, or otherwise manipulate membrane fences that localize trans-membrane receptors.

The sensitivity studies in this work suggest a number of approaches to how one might control EGFR's large-scale behavior, assuming that there are high-density and low-density membrane domains. Since an increase in receptor density corresponds to the

conditions found in many cancerous cells, than we would conceivably be interested in weakening the clustering behavior of EGFR. Firstly, and most obviously, decreasing the number of receptors on the membrane weakens clustering behavior (Figure 4.8). However, there are other methods besides directly manipulating receptor density to affect clustering behavior. Large and small corrals with the same receptor density behave differently, and this changing sensitivity of EGFR's clustering behavior suggests that direct manipulation of corral sizes may be an effective control method. At relatively high receptor densities, smaller corrals exhibit stronger clustering than large corrals (Figure 4.8b). Thus, by forcing many small domains to coalesce into larger ones, the overall clustering behavior may be weakened. Nevertheless, at low receptor densities clustering is weaker in smaller corrals than larger ones. Under these conditions, dispersing large corrals to form smaller ones may effectively weaken clustering. These corral size targeting strategies would not require direct manipulation of the EGFR protein (by genetic modification or ligand blocking).

Regardless of the approach used, our results demonstrate that corral size must be considered in conjunction with receptor density, as they synergistically affect EGFR clustering behavior. It would thus be inappropriate to compare the behavior of cells of similar receptor densities but with different corral domain sizes or dispersions.

Regarding the simulation method presented, ACGMC successfully simulated the multiscale EGFR reaction-diffusion system by applying a KMC-resolution mesh on the high reaction density corralled area while applying an otherwise coarse grained mesh on the rest of the membrane space. This method is promising for a wide variety of multiscale and spatially heterogeneous problems, such as catalyst-support simulation (high resolution at interface of catalyst and support with low-resolution over the bulk support metal), membrane separations (high resolution for membrane, low resolution in bulk space), and

other large-scale spatially heterogeneous reaction-diffusion systems that are too computationally intense for KMC, yet require discrete microscopic detail. CGMC could also be conceivably expanded to three dimensions and applied to a whole new class of systems. Using the EGFR system as an example, the lifetime of an endosome, including receptor clustering, budding, and cytoplasmic transport may be simulated using 'thin' CG cells for the membrane and coarse 3D cells representing the cytoplasm. Adaptive CGMC would allow detailed spatial resolution for high-reaction locations (local areas of the membrane), with coarse and computationally cheap cells for diffusion-heavy processes (cytoplasmic transport).

REFERENCES

1. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nature Reviews Molecular Cell Biology, 2001. **2**: p. 127-137.
2. Ferguson, K.M., et al., *EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization.* Molecular Cell, 2003. **11**(2): p. 507-517.
3. Haigler, H., J. McKanna, and S. Cohen, *Direct visualization of the binding and internalization of a ferritin conjugate of epidermal growth factor in human carcinoma cells A-431.* J. Cell Biol., 1979. **81**(2): p. 382-395.
4. Henson, E.S. and S.B. Gibson, *Surviving cell death through epidermal growth factor (EGF) signal transduction pathways: Implications for cancer therapy.* Cellular Signalling, 2006. **18**(12): p. 2089-2097.
5. Tyson, J.J., K. Chen, and B. Novak, *Network dynamics and cell physiology.* Nature Reviews Molecular Cell Biology, 2001. **2**(12): p. 908-916.
6. Pike, L.J., *Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function.* Journal of Lipid Research, 2006. **47**(7): p. 1597-1598.
7. Ritchie, K., et al., *Detection of non-Brownian diffusion in the cell membrane in single molecule tracking.* Biophysical Journal, 2005. **88**(3): p. 2266-2277.
8. Mayawala, K., D.G. Vlachos, and B.J. Edwards, *Heterogeneities in EGF receptor density at the cell surface can lead to concave up Scatchard plot of EGF binding.* FEBS Letters, 2005. **579**: p. 3043-3047.
9. Katsoulakis, M.A., A.J. Majda, and D.G. Vlachos, *Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems.* Journal of Computational Physics, 2003. **186**(1): p. 250-278.
10. Chatterjee, A., D.G. Vlachos, and M.A. Katsoulakis, *Spatially adaptive lattice coarse-grained Monte Carlo simulations for diffusion of interacting molecules.* Journal of Chemical Physics, 2004. **121**(22): p. 11420-11431.
11. Collins, S.D., A. Chatterjee, and D.G. Vlachos, *Coarse-grained kinetic Monte Carlo models: Complex lattices, multicomponent systems, and homogenization at the stochastic level.* Journal of Chemical Physics, 2008. **129**(18).
12. Niehaus, A.M.S., *Simulations of Epidermal Growth Factor Receptor Dynamics on Corralled Membrane Surfaces*, in *Department of Chemical Engineering*. 2007, University of Delaware: Newark, DE, USA. p. 124.
13. Wilson, B.S., et al., *Markers for detergent-resistant lipid rafts occupy distinct and dynamic domains in native membranes.* Molecular Biology of the Cell, 2004. **15**(6): p. 2580-2592.
14. Jain, A.J., Dubes, R.C., *Algorithms for Clustering Data*. 1988, Englewood Cliffs, NJ: Prentice Hall.
15. Mayawala, K., D. Vlachos, and J. Edwards, *Computational modeling reveals molecular details of epidermal growth factor binding.* BMC Cell Biology, 2005. **6**(1): p. 41.
16. Wiley, H.S., *Anomalous binding of epidermal growth factor to A431 cells is due to the effect of high receptor densities and a saturable endocytic system.* J. Cell Biol., 1988. **107**(2): p. 801-810.

CHAPTER 5

CONCLUSIONS

On-lattice kinetic Monte Carlo (KMC) simulations have extensively been applied to numerous systems. In response to the severe limitations of the KMC method at long time and length scales, the Coarse-Grained MC (CGMC) method was introduced. CGMC method outperforms the traditional KMC method on computational cost while retaining good accuracy. Through the work presented here, the CGMC model was developed (Chapter 2), explored (Chapter 3), and applied to the EGFR system (Chapter 4).

The EGFR is an ideal system for the CGMC method, as the two domain structure (high-density corralled regions; low density extracorral regions) may be neatly mirrored by the adaptive CGMC method. Since membrane diffusion exhibits a wide range of scales (diffusion within a corral is fast relative to diffusion across corral interfaces), traditional KMC simulation is limited to small length and time scales. CGMC assumes that diffusion within a CG cell is infinitely fast while cell-to-cell diffusion is not. By representing corrals as individual CG cells, the CGMC simulation describes diffusion accurately while expanding the simulation's scale limits tremendously. This approach is further refined by applying adaptive meshing to increase simulation detail in high density areas of the membrane.

In Chapter 2, the CGMC method was expanded beyond single species simulations on a homogeneous surface. The framework for on-lattice, CGMC simulation with an arbitrary number of site types and/or surface species was derived and tested. This enabled efficient calculations of many realistic systems that are beyond the realm of the traditional KMC method. Numerical examples demonstrated the CGMC method's ability to capture many types of elementary mono and bimolecular surface reactions on spatially inhomogeneous surface. Large computational savings were achieved compared to the KMC method.

Additionally, a novel homogenization approach at the stochastic level was introduced. The homogenization approach reduces the number of processes in the algorithm, decrease memory and CPU requirements. The homogenization is expected to be particularly important for systems exhibiting combinatorial explosion, a concern for protein signaling networks such as the EGFR system due to the large number of surface heterogeneities, protein species, and protein conformations.

Chapter 3 investigated the accuracy of the newly introduced multicomponent CGMC method. Two general bimolecular reaction-diffusion systems were studied: the AA system (single species reaction-diffusion), and the AB system (a two-species reaction-diffusion system). The rates of diffusion relative to the rate of reaction, as well as the coverage of the lattice were varied and the reaction rate recorded for various levels of coarse graining. The results were compared to the KMC method to assess their accuracy.

In Chapter 4, the CMGC method was applied to the EGFR system. The Adaptive Coarse-Grained kinetic Monte Carlo (ACGMC), a multiscale spatial stochastic simulation was introduced and applied to the EGFR diffusion-reaction system. The ACGMC method properly

captures the detailed spatial reactions in high receptor density corrals of the membrane while efficiently and accurately simulating the low density areas of the membrane with a low resolution lattice.

Kinetically preserved clusters (as defined by the Hopkins statistic) were observed on the order of minutes. Sensitivity analysis of the density and longevity of these clusters was carried out. This chapter demonstrated that the clustering behavior of EGFR is heavily influenced by diffusion barriers of the membrane surface. In fact, diffusion fences offer a (transient) clustering mechanism independent of those theorized in the literature. The correlations seen in the sensitivity analysis demonstrated that corrals size and receptor density must be considered together and not independently. For instance, it is inappropriate to compare the behavior of cells of similar receptor densities but with different corral domain sizes or dispersions since the impact of each variable is dependent on the other.

Possibilities for future work go in a number of directions. Leaping off chapter 3, effective reaction rate constants could be found as a function of the local macroscopic parameters of the simulation (such as θ and Da) to compensate for expected inaccuracies of the CGMC method. The corrections as functions of the macroscopic parameters would ideally be known a priori but could also be simulated with short KMC simulations and recorded during the CG simulation. This effective local rate constant would then be used in a CGMC simulation to deliver the computational saving of the CGMC method with a reduced loss of accuracy. Individual CG cells would automatically adjust the local reaction rate constant based on local conditions, allowing highly heterogeneous systems to operate on a single lattice.

For high-density systems where clusters of single species form in unpredictable locations, a CGing method that applies adaptive meshes on high-detail areas as they form is conceivable. This on-the-fly adaptive CGing would detect local concentrations of reactants and automatically rearrange the lattice around them. Conceivably, this technique could be utilized for three dimensional CGMC and applied to a whole new class of systems. Using the EGFR system as an example, the lifetime of an endosome, including receptor clustering, budding, and cytoplasmic transport may be simulated using 'thin' CG cells for the membrane and coarse 3D cells representing the cytoplasm. On-the-fly adaptive CGMC would allow detailed spatial resolution for high-reaction locations as they form (i.e. endosome formation), with coarse and computationally cheap cells for long-scale diffusion processes (i.e. cytoplasmic transport).

APPENDIX

PERMISSIONS TO REPRINT

Permission to reprint Figure 1 of Chapter 1

Dear Dr. Collins,

Permission is granted for your use of the figure as described in your message below. Please cite the full journal references and "Copyright (copyright year) National Academy of Sciences, U.S.A."

Best regards,
Sara Spizzirri for
Diane Sullenberger
Executive Editor
PNAS
- Hide quoted text -
-----Original Message-----
From: Stuart Collins [mailto:collins.stu@gmail.com]
Sent: Thursday, June 18, 2009 4:18 PM
To: PNAS Permissions
Subject: Request for permission to reprint

Hello -
I'm writing to request permission to reprint a figure in my Masters of Science in Chemical Engineering thesis for the University of Delaware. I've included all requested information below. Please contact me if there are any questions.
-Stuart Collins

Original material information:
1. Your full name, affiliation, and title
Stuart D Collins, University of Delaware, Graduate Student
2. Your complete mailing address, phone number, fax number, and e-mail address

Stuart Collins
Department of Chemical Engineering
Colburn Laboratory
150 Academy Street
Newark, DE 19716

302-831-1247
collins.stu@gmail.com

3. PNAS volume number, issue number, and issue date
PNAS January 27, 2004 vol. 101 no. 4 929-934

4. PNAS article title
A structure-based model for ligand binding and dimerization of EGF receptors
5. PNAS authors' names
Peter Klein, Dawn Mattoon, Mark A. Lemmon, and Joseph Schlessinger
6. Page numbers of items to be reprinted
Page 2
7. Figure/table number or portion of text to be reprinted
Figure 1

1. Title of work in which PNAS material will appear
(Master's Thesis): Multiscale Monte Carlo Study of Epidermal Growth Factor Receptor Diffusion and Dimerization
2. Authors/editors of work
Stuart D Collins, Advised by Dr. Dionisios Vlachos
3. Publisher of work
University of Delaware
4. Retail price of work
None
5. Number of copies of work to be produced
Three
6. Intended audience
Internally at University of Delaware

7. Whether work is for nonprofit or commercial use
Nonprofit: Master's of Science in Dept of Chemical Engineering

Permission to reprint Figure 3 of Chapter 1

Dear Stuart,

Thank you for following up on your request for permission to use the figure cited in your e-mail. Permission is granted. Please use the following acknowledgment (preferably on the page(s) where this material appears):

"Reprinted, with permission, from the Annual Review of Biophysics and Biomolecular Structure, Volume 34 © 2005 by Annual Reviews www.annualreviews.org"

The fee for this usage has been waived. Any subsequent use of this material would require submission of a new permission request. If I can be of further assistance, please do not hesitate to contact me.

Best wishes for success with your thesis.

Sincerely,
Laura

Laura Folkner
Permissions Dept.
ANNUAL REVIEWS
A Nonprofit Scientific Publisher
4139 El Camino Way
Palo Alto, CA 94306
Ph: 650.843.6636
Fax: 650.855.9815
lfolkner@annualreviews.org
- Hide quoted text -

-----Original Message-----
From: Stuart Collins [mailto:collins.stu@gmail.com]
Sent: Wednesday, August 26, 2009 9:28 AM
To: permissions
Subject: Re: Permission to reprint

Hello-

I was just checking in on the status of this request. Was it received
and is there any other information you need?

-Stuart Collins

On Wed, Aug 19, 2009 at 8:33 PM, Stuart Collins<collins.stu@gmail.com> wrote:
> Hello -
> I'm writing to request permission to reprint a figure in my Masters of
> Science in Chemical Engineering thesis for the University of Delaware.
> I've included all requested information below. Please contact me if
> there are any questions.
> -Stuart Collins
>
>
> ---title, edition and copyright year of the Annual Reviews volume:
> Annual Review of Biophysics and Biomolecular Structure
> Vol. 34: 351-378
> June 2005
>
> ---author and name of article
> PARADIGM SHIFT OF THE PLASMA MEMBRANE CONCEPT FROM THE
TWO-DIMENSIONAL
> CONTINUUM FLUID TO THE PARTITIONED FLUID: High-Speed Single-
Molecule
> Tracking of Membrane Molecules
> Akihiro Kusumi, Chieko Nakada, Ken Ritchie, Kotono Murase, Kenichi
> Suzuki, Hideji Murakoshi, Rinshi S. Kasai, Junko Kondo, and Takahiro
> Fujiwara
>
> ---exact material, including page numbers and figure numbers, for
> which permission is requested
> Figure 1, Page C-1
>
> ---author and title of the work in which the material will appear
> (Master's Thesis): Multiscale Monte Carlo Study of Epidermal Growth
> Factor Receptor Diffusion and Dimerization
> Stuart D Collins, Advised by Dr. Dionisios Vlachos
>
> ---publisher and publication date of the work in which the material will appear
> University of Delaware, September 2009

>
> ---format/media of the new work
> Three paper copies, electronic
>

Permission to reprint Figure 2 of Chapter 1

ELSEVIER LICENSE
TERMS AND CONDITIONS

Sep 12, 2009

This is a License Agreement between Stuart D Collins ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Stuart D Collins
Customer address	4517 Spruce St, Apt 1 Philadelphia, PA 19139
License Number	2252710954880
License date	Aug 19, 2009
Licensed content publisher	Elsevier
Licensed content publication	Micron
Licensed content title	Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale

	electron-dense probes and electron microscopy
Licensed content author	Jun Zhang, Karin Leiderman, Janet R. Pfeiffer, Bridget S. Wilson, Janet M. Oliver and Stanly L. Steinberg
Licensed content date	January 2006
Volume number	37
Issue number	1
Pages	21
Type of Use	Thesis / Dissertation
Portion	Figures/table/illustration/abstracts
Portion Quantity	1
Format	Both print and electronic
You are an author of the Elsevier article	No
Are you translating?	No
Order Reference Number	
Expected publication date	Sep 2009
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD

Total 0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and
he permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic

version,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx , As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com

All content posted to the web site must maintain the copyright information line on the bottom of each image
You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**None

v1.6

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable

license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a check or money order referencing your account number and this license number 2252710954880.
If you would prefer to pay for this license by credit card, please go to http://www.copyright.com/creditcard to download our credit card payment authorization form.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2252710954880 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Permission to reprint Figure 5a of Chapter 2

AMERICAN INSTITUTE OF PHYSICS LICENSE
TERMS AND CONDITIONS

Sep 12, 2009

This is a License Agreement between Stuart D Collins ("You") and American Institute of Physics ("American Institute of Physics") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Institute of Physics, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2262091334970
License date	Sep 04, 2009
Licensed content publisher	American Institute of Physics
Licensed content publication	Journal of Chemical Physics
Licensed content title	Continuum mesoscopic framework for multiple interacting species and processes on multiple site types and/or crystallographic planes
Licensed content author	Abhijit Chatterjee
Licensed content date	Jul 21, 2007
Volume number	127
Issue number	3
Type of Use	Thesis/Dissertation
Requestor type	Student

Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Order reference number	02
Title of your thesis / dissertation	Multiscale Monte Carlo Study of Epidermal Growth Factor Receptor Diffusion and Dimerization
Expected completion date	Sep 2009
Estimated size (number of pages)	120
Total	0.00 USD

Terms and Conditions

American Institute of Physics -- Terms and Conditions: Permissions Uses

American Institute of Physics ("AIP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Institute of Physics." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.

2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP, with permission, from another source. Under no circumstances does AIP, purport or intend to grant

permission to reuse material to which it does not hold copyright.

3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.

5. AIP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP, American Institute of Physics, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.

6. AIP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP, American Institute of Physics, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.

7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP. This license may not be amended except in a writing signed by the party to be charged.

8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a

check or money order referencing your account number and this license number 2262091334970.
If you would prefer to pay for this license by credit card, please go to http://www.copyright.com/creditcard to download our credit card payment authorization form.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2262091334970 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Permission to reprint Chapter 2

AMERICAN INSTITUTE OF PHYSICS LICENSE
TERMS AND CONDITIONS

Sep 12, 2009

This is a License Agreement between Stuart D Collins ("You") and American Institute of Physics ("American Institute of Physics") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Institute of Physics, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2262100156316
License date	Sep 04, 2009
Licensed content publisher	American Institute of Physics
Licensed content publication	Journal of Chemical Physics
Licensed content title	Coarse-grained kinetic Monte Carlo models: Complex lattices, multicomponent systems, and homogenization at the stochastic level
Licensed content author	Stuart D. Collins, Abhijit Chatterjee, Dionisios G. Vlachos
Licensed content date	Nov 14, 2008
Volume number	129
Issue number	18
Type of Use	Thesis/Dissertation

Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Order reference number	03
Title of your thesis / dissertation	Multiscale Monte Carlo Study of Epidermal Growth Factor Receptor Diffusion and Dimerization
Expected completion date	Sep 2009
Estimated size (number of pages)	120
Total	0.00 USD

Terms and Conditions

American Institute of Physics -- Terms and Conditions: Permissions Uses

American Institute of Physics ("AIP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Institute of Physics." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.

2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP and does not contain the

copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP, with permission, from another source. Under no circumstances does AIP, purport or intend to grant permission to reuse material to which it does not hold copyright.

3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIP reserves all rights not specifically granted herein.

4. The foregoing license shall not take effect unless and until AIP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.

5. AIP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP, American Institute of Physics, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.

6. AIP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP, American Institute of Physics, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.

7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP. This license may not be amended except in a writing signed by the party to be charged.

8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a check or money order referencing your account number and this license number 2262100156316.
If you would prefer to pay for this license by credit card, please go to http://www.copyright.com/creditcard to download our credit card payment authorization form.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2262100156316 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.
