Predicting colloid transport through saturated porous media: A critical review

Ian L. Molnar1, William P. Johnson2, Jason I. Gerhard1, Clinton S. Willson3, and Denis M. O’Carroll1,4

1Department of Civil and Environmental Engineering, University of Western Ontario, London, Ontario, Canada, 2Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah, USA, 3Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, Louisiana, USA, 4School of Civil and Environmental Engineering, Connected Water Initiative, University of New South Wales, Manly Vale, New South Wales, Australia

Abstract
Understanding and predicting colloid transport and retention in water-saturated porous media is important for the protection of human and ecological health. Early applications of colloid transport research before the 1990s included the removal of pathogens in granular drinking water filters. Since then, interest has expanded significantly to include such areas as source zone protection of drinking water systems and injection of nanometals for contaminated site remediation. This review summarizes predictive tools for colloid transport from the pore to field scales. First, we review experimental breakthrough and retention of colloids under favorable and unfavorable colloid/collector interactions (i.e., no significant and significant colloid-surface repulsion, respectively). Second, we review the continuum-scale modeling strategies used to describe observed transport behavior. Third, we review the following two components of colloid filtration theory: (i) mechanistic force/torque balance models of pore-scale colloid trajectories and (ii) approximating correlation equations used to predict colloid retention. The successes and limitations of these approaches for favorable conditions are summarized, as are recent developments to predict colloid retention under the unfavorable conditions particularly relevant to environmental applications. Fourth, we summarize the influences of physical and chemical heterogeneities on colloid transport and avenues for their prediction. Fifth, we review the upscaling of mechanistic model results to rate constants for use in continuum models of colloid behavior at the column and field scales. Overall, this paper clarifies the foundation for existing knowledge of colloid transport and retention, features recent advances in the field, critically assesses where existing approaches are successful and the limits of their application, and highlights outstanding challenges and future research opportunities. These challenges and opportunities include improving mechanistic descriptions, and subsequent correlation equations, for nanoparticle (i.e., Brownian particle) transport through soil, developing mechanistic descriptions of colloid retention in so-called “unfavorable” conditions via methods such as the “discrete heterogeneity” approach, and employing imaging techniques such as X-ray tomography to develop realistic expressions for grain topology and mineral distribution that can aid the development of these mechanistic approaches.

1. Introduction

The introduction of colloid filtration theory (CFT) [Yao et al., 1971] sparked four decades of research at the pore, macroscopic, and field scales to develop predictive capabilities for colloid transport and retention in porous media. Significant strides have been made in the development of mechanistic models and their upscaling that link pore-scale colloid transport processes to predictions of field-scale behavior. The fiftieth anniversary of Water Resources Research represents an ideal opportunity to review the major contributions in this evolving, highly relevant field. It is also an opportunity to critically reflect on how the accuracy of field-scale predictions may be constrained by the simplifying assumptions built into mechanistic models, correlation equations, and their relationship to our growing understanding of actual, pore-scale colloid behavior.

Prediction of colloid transport through the subsurface is important for a wide range of environmental and human-health-related risk scenarios. For example, predictive colloid transport models are necessary to assess and minimize the risk of pathogen transport in groundwater. Among numerous examples is the Walkerton Ontario, Canada tragedy [O’Connor, 2002] in which the town’s drinking water supply well was
contaminated by *E. coli* transported in groundwater flow through the subsurface. Colloid transport from septic systems, agricultural runoff, and other sources through near-shore soils may lead to elevated concentrations of pathogenic colloids in beach sands [e.g., Lipp et al., 2001] and transport across the groundwater/surface water interface may be the cause of frequent beach closures [e.g., Russell et al., 2012]. A wide array of research into pathogen transport has been undertaken including virus transport [Ryan et al., 1999; Schijven et al., 1999; Jin and Flury, 2002] and column-scale bacteria transport [e.g., Albinger et al., 1994; Baygents et al., 1998; Simoni et al., 1998; Torkzaban et al., 2008], bacteria transport in the field [Bales et al., 1989; Powelson et al., 1993; Harvey et al., 1995; DeBorde et al., 1999; Ryan et al., 1999; Schijven et al., 1999; Zhang et al., 2001a; Ryan et al., 2002; Blanford et al., 2005; van der Wiel et al., 2008; Scheibe et al., 2011], and phage persistence in soils [e.g., Yates et al., 1985] among many others; numerous reviews of pathogen transport in groundwater are available [Bitton and Harvey, 1992; Harvey, 1997; Taylor et al., 2004; Sen, 2011].

Engineered nanoparticles (ENPs), another class of colloids, are of significant interest for their transport in porous media. On the one hand, ENPs are an emerging environmental contaminant. Industrial production of engineered nanoparticles, such as carbon nanotubes and nanosilver, has increased significantly in recent years [Wiesner et al., 2006; Nowack and Bucheli, 2007; Christian et al., 2008; Majestic et al., 2010; Marambio-Jones and Hoek, 2010; Petersen et al., 2011; Gao et al., 2013] and it is expected that their presence in the environment will increase. This can occur through diffuse ENP release [Mueller and Nowack, 2008], including ENP dispersal in aerosols, in exhaust emissions, and in land application of wastewater sludge containing nanoparticles originating in consumer products. Point sources of ENPs include leakage from landfills and accidental industrial releases. Prediction of ENP transport in porous media is important for assessing the risk from ENP migration in aquifers and the contamination of wells used for drinking water supply. ENP studies have involved modeling [Cullen et al., 2010; Bai and Li, 2012; Taghavy et al., 2013], pore-scale experiments [Dunphy Guzman et al., 2006; May et al., 2012; May and Li, 2013; Molnar et al., 2014], and column experiments [Lecocanet et al., 2004; Wang et al., 2008; Jaisi and Elimelech, 2009; X. Y. Liu et al., 2009; Torkzaban et al., 2010; Uyusur et al., 2010; Lin et al., 2011; Zhang et al., 2012; Liang et al., 2013; Ren and Smith, 2013; Neukum et al., 2014]. Numerous review papers exist, including Petersen et al. [2011] who present a review of the transport and environmental risks of carbon nanotubes.

ENPs are also being considered as a remediation strategy for sites contaminated with industrial organic liquids. For example, nano-zero-valent iron (nZVI) particles are injected into the subsurface at contaminated sites for the in situ reduction of chlorinated solvents in groundwater [Johnson et al., 2013; O’Carroll et al., 2013; Tosco et al., 2014a]. This novel approach has been evaluated at both the laboratory [e.g., Berge and Ramsburg, 2009; Phenrat et al., 2009; Raychoudhury et al., 2010; Kocur et al., 2013] and field scales [Johnson et al., 2013; Kocur et al., 2014]. Recent advances include the tailored design and synthesis of novel ENPs including bimetallic composites and polymer coatings [He and Zhao, 2008; Bennett et al., 2010; Sakulchaichareon et al., 2010; Gastone et al., 2014; Tosco et al., 2014b], which can influence surface charge, fluid viscosity, and other properties atypical of traditional colloid suspensions. The design and optimization of these ENP-based remediation schemes require accurate predictive models of ENP transport in porous media, and users of existing CFT need to understand how and where it is possible to change parameters appropriately for nontraditional suspensions.

Colloids can also enhance the transport of dissolved contaminants in groundwater via their sorption onto colloid surfaces. Colloid-facilitated transport has been studied for a wide range of contaminants including radionuclides [e.g., Kersting et al., 1999; Novikov et al., 2006], hydrocarbons [e.g., Qi et al., 2014], and pesticides [e.g., de Jonge et al., 1998; Sprague et al., 2000]. Research in this area includes modeling the reduction in contaminant retardation [Corapcioglu and Jiang, 1993; Johnson et al., 1995a; Flury and Qiu, 2008]. Several reviews and special journal sections have provided reviews of colloid-associated contaminant transport [McCarthy and Zachara, 1989; Ouyang et al., 1996; Ryan and Elimelech, 1996; de Jonge et al., 2004; Sen and Khilar, 2006]. Further understanding of colloid transport mechanisms are required to predict the risk posed by colloid-enhanced transport mechanisms.

The predictive tools developed to explore column and field-scale colloid transport phenomena are primarily continuum-based numerical models that solve the advection-dispersion equation for solute transport in porous media. Terms employing rate coefficients to drive colloid mass transfer to the stationary (solid) phase and back are used to describe the macroscopically observed transport behavior of colloids as described by breakthrough-elution concentration histories and profiles of retained concentration as a function of distance. Rate constants for colloid transfer to the stationary phase may be independently determined using upscaled
predictions from mechanistic models that employ relevant force and torque balances to determine colloid trajectories in flow fields corresponding to representative pore (collector) geometries. Relevant forces include fluid drag, diffusion, gravity, as well as colloid-surface forces including van der Waals and electric double layer as described by DLVO theory [Derjaguin and Landau, 1941; Verwey and Overbeek, 1948] and its derivatives. The upscaled mechanistic predictions provide good prediction of colloid transport at the continuum scale for spheroidal colloids in uniform media when colloid-collector repulsion is absent (so-called “favorable” conditions), as is typically the case for oppositely charged surfaces [e.g., Elimelech, 1991].

In the environment, however, the typical condition is that both colloids and collector surfaces carry negative charge. Colloids exhibiting net negatively charged surfaces under environmental pH conditions include bacteria [e.g., McClaine and Ford, 2002; Bradford et al., 2006b; Foppen and Schijven, 2006], protozoa [e.g., Ruohola et al., 2012; Y. Liu et al., 2009], viruses [e.g., Ryan et al., 1999; Mondal and Sleep, 2013], and engineered nanoparticles [e.g., Torkzaban et al., 2010; Jiang et al., 2012a; Liang et al., 2013; Kocur et al., 2014; Molnar et al., 2014]. Common minerals exhibiting negatively charged surfaces at typical pH include quartz, silica, feldspars, mica, and certain clays [Molnar et al., 2011, Table 1]. The combination of both colloids and collectors carrying a negative charge leads to electric double layer repulsion (as well as steric repulsion from surface macromolecules), thereby yielding “unfavorable” conditions for attachment.

Repulsion between colloids and porous media collectors results in complex transport behavior that is mediated by the topology of the pore domain and the flow field, via such features as low-flow (diffusion-limited) zones that may be generated in nonospherial or nonuniform media, such as many environmental sediments and soils. In recognition of these complexities, mechanistic models have been adapted to address more complex scenarios [Long and Hilpert, 2009; Ma et al., 2009; Long et al., 2010; Ma and Johnson, 2010; Ma et al., 2011, 2013; Pazmino et al., 2014b]. However, upscaling of these types of models for implementation into advection-dispersion or other transport equations represents an important ongoing research opportunity.

Whereas mechanistic predictions regarding the above complexities are not yet readily implemented into continuum-scale transport models, advances have occurred in descriptive continuum-scale colloid transport models describing these complexities [e.g., Tufenkji and Elimelech, 2004a, 2005b, 2005a; Simu̇nek and van Genuchten, 2008; Bradford et al., 2009, 2011b; Leij and Bradford, 2013; Katzourakis and Chrysikopoulos, 2014]. These models employ parameters fitted to column-scale breakthrough-elution concentration histories and provide a method to infer potential pore-scale mechanisms driving complex transport behavior. There is clearly an important ongoing research opportunity associated with obtaining high resolution, pore-scale information on the behavior of colloids and its generalization into continuum-scale models in order to further improve predictive approaches to incorporating colloid interaction with soil and flow heterogeneity.

This paper provides a critical review of colloid transport literature with a specific focus on linking pore-scale processes to column-scale and field-scale observations as well as continuum-scale transport models across a range of relevant subsurface scenarios. This review exclusively discusses colloid transport through water-saturated media and aims to help distinguish and clarify the roles of different pore-scale and continuum-scale modeling approaches, focusing on post-1990 colloid transport work. The paper is structured so as to (i) provide an introduction to the fundamentals and review existing knowledge and knowledge gaps for those who are relatively new to the field and (ii) provide a more detailed, critical assessment of the state of the art. In section 2, a brief review is provided of mean-field DLVO interactions and experimentally observed colloid transport phenomena, focusing on the contrasting transport behaviors observed under favorable versus unfavorable conditions for attachment. Section 3 describes strategies by which kinetic parameters have been used in continuum models to infer mechanisms from experimental observations. Section 4 reviews the current and emerging ability of mechanistic approaches to independently predict kinetic parameters used in continuum models (e.g., retention coefficients) in both favorable and unfavorable conditions. Section 5 examines upscaling of mechanistic predictions to rate constants, explores the role of topology in upscaling strategies, and highlights opportunities for future research.

2. Mean-Field DLVO Interactions and Experimental Observations

2.1. Mean-Field DLVO Forces
As mentioned, the mechanistic models that underpin predictive tools for colloid transport depend upon DLVO interactions. Classic DLVO theory [Derjaguin and Landau, 1941; Verwey and Overbeek, 1948] attempts
to describe the colloid-collector interactions in terms of adhesive and repulsive forces. By summing the adhesive (negative) and repulsive (positive) energies between a colloid and collector over a range of separation distances, an interaction energy profile—the net energy versus separation distance—is constructed.

Figure 1 presents two typical, classic DLVO interaction energy profiles for favorable (silver colloid-iron oxide) and unfavorable conditions (silver colloid-quartz). The figure illustrates a number of distinct features or regions that are typical of colloid-surface interaction profiles. In unfavorable conditions, a repulsive barrier exists that either limits colloid attachment to the surface (experimental observations) or prevents attachment (for energy barriers exceeding several kT in mechanistic simulations).

Outward from this repulsive barrier exists a small region of attraction where colloids may accumulate in a secondary minimum at separation distances tens of nm from the collector surface. In situations where colloids either possess enough energy to overcome the barrier or where no barrier exists (i.e., favorable conditions), they can enter the primary minimum (Figure 1) and physically attach to the surface.

The “near-surface fluid domain” is defined, for the purposes of this review, as the region between the collector surface and the distance over which van der Waals attraction is significant, i.e., secondary minimum and closer (see Figure 1). The “bulk fluid domain” is defined as the area beyond the near-surface fluid domain beyond which colloid-collector DLVO interactions are insignificant. For typical colloid-collector systems, the interface between the bulk fluid domain and the near-surface domain occurs in the range of several to hundreds of nm (e.g., Johnson and Hilpert, 2013).

It is worthwhile mentioning that the energy barrier also influences colloid detachment from the primary minimum. In classic DLVO theory, the primary minimum is an infinitely deep well and the height of the energy barrier is inconsequential to colloids entrapped within the minimum. However, short-range forces such as Born repulsion serve to limit the depth of the well to finite values (Ruckenstein and Prieve, 1976). In this finite-depth scenario, the energy barrier now contributes to the effective depth of the energy well and acts as a barrier to both colloid attachment and detachment (Ryan and Elimelech, 1996). Increasing the
height of the energy barrier via perturbations in solution chemistry will deepen the energy well holding the colloid on the surface of the collector [Hahn et al., 2004].

The forces acting on the colloids are the derivative of the energies with respect to separation distance, such that both energy and force profiles exhibit the same features: primary minimum, repulsive barrier, and secondary minimum. Using this sign convention, attractive forces are negative and repulsive forces are positive. The separation distances of these features differ slightly between the force and energy profiles, but this is not important to this discussion. This review refers to the features illustrated in Figure 1 without distinguishing specifically between force and energy. Traditional mechanistic colloid transport and retention models describe colloid-collector interactions (i.e., repulsion, physical contact, and retention in the secondary minimum) by considering the above-described DLVO theory. A shortcoming of this theory (as presented above) is that it considers the surfaces to be monolithic (homogeneous). It is worthwhile emphasizing that this monolithic DLVO approach does not represent a “mean interaction energy” of colloids, but rather it is an interaction calculated from the bulk surface properties of colloids and collectors. This so-called mean-field DLVO approach is an overgeneralization of the surface characteristics of colloids and collectors, since it does not account for microscale to nanoscale heterogeneity that is expected to exist on real surfaces (as discussed in section 4). Hence, while we review below the great utility of mean-field DLVO interactions for interpreting the transport behavior of colloids in porous media, we also discuss in subsequent sections the emerging opportunities offered by the incorporation of heterogeneity into colloid-surface DLVO interactions.

2.2. Experimental Observations: Influence of Favorable Versus Unfavorable Conditions

While DLVO forces are significant over small (nanoscale) separation distances (see near-surface fluid domain in Figure 1), the presence or absence of a repulsive energy barrier controls the transport and retention of colloids at column and field scales. A large number of column-scale transport experiments have demonstrated the implications of these small-scale interactions that yield field-scale consequences, as described below.

A note on terminology: “attachment” is herein defined as the immobilization, via physical contact, of the colloid onto the collector surface within the primary energy well. As continuum-scale experiments typically cannot distinguish attachment from other retention mechanisms (such as occupancy in the secondary energy minimum), the term “retention” will be employed to describe colloids that do not exit the system during the term of the experiment. Likewise, continuum-scale experiments that observe delayed exit of colloids from the experiment typically cannot distinguish between colloids that detached from a collector surface or colloids that were temporarily retained without attachment. Thus, the term “reentrainment” will be employed to avoid attributing a particular mechanism to the colloids’ delayed exit during these experiments. To be clear, the processes of attachment versus retention and detachment versus reentrainment are indeed distinguishable in many pore-scale direct observation experiments, as well as in pore-scale mechanistic simulations (they must be distinguished in the latter). In those contexts, we will refer specifically to attachment and detachment as appropriate.

Column experiments conducted under favorable conditions typically exhibit relatively simple colloid transport behavior: as seen in Figure 2 (right-most column), colloid breakthrough during injection reaches a steady state (i.e., temporally constant) plateau [Elimelech et al., 2000; Li et al., 2004; Tufenkji and Elimelech, 2004a; Li et al., 2005; Tufenkji and Elimelech, 2005a].

During the elution phase in favorable conditions, the concentration of colloids exiting the column decreases sharply with time indicating that there is negligible reentrainment of previously retained colloids [Li et al., 2004; Tufenkji and Elimelech, 2004a, 2005a]. The corresponding profiles of retained colloid concentrations with column length show a classic log linear decrease under favorable conditions (Figure 2, right column), as predicted by analytical solutions of the advective-dispersive-colloid transport equation (discussed in section 3) [Li et al., 2004; Tufenkji and Elimelech, 2004a, 2005a; Han et al., 2014]. These simple, well defined column-scale behaviors are due to the straightforward colloid-collector interactions at the pore scale under favorable conditions. Colloids that approach the collector experience no repulsion and come into physical contact with the collector surface (immobilize) [Kuznar and Elimelech, 2007; Johnson et al., 2010]. These colloids do not typically detach, since the strong attractive forces exceed fluid drag forces [Bergendahl and Grasso, 2000].
Column experiments conducted under unfavorable conditions exhibit significantly different behavior from those conducted under favorable conditions (Figure 2, left and middle columns). Depending on the experimental conditions, unfavorable breakthrough behavior may also reach a steady state plateau (Figure 2, left and middle of top row) [Elimelech et al., 2000; Li et al., 2005; Jiang et al., 2012a]. However, nonsteady state behavior may also occur. The effluent concentration may gradually increase over time, suggesting that a limited number of sites for colloid retention exist and are progressively filled (i.e., blocking) [Johnson and Elimelech, 1995; Camesano et al., 1999; X. Y. Liu et al., 2009; Lin et al., 2011; Mattison et al., 2011; C. Wang et al., 2012; Liang et al., 2013]. Decreasing concentration during breakthrough (called ripening) suggests that favorable colloid-colloid interactions allow already retained colloids (presumably due to inferred “favorable” heterogeneity as described in section 4) to serve as additional sites for further colloid retention [Tong et al., 2008; Jiang et al., 2012a].

Extended tailing may occur during the elution phase of unfavorable experiments (Figure 2, left and middle top row). Tailing is defined by the slow release of colloids from the column (illustrated in Figure 2) and is indicative of significant reentrainment, as observed for microbes in the laboratory [Fontes et al., 1991; Hornberger et al., 1992; Lindqvist et al., 1994; McCaulou et al., 1994; Johnson et al., 1995b; McCaulou et al., 1995; Hendry et al., 1997, 1999; Harter et al., 2000; Li and Johnson, 2005; Li et al., 2005], field [Scholl and Harvey, 1992; Harvey et al., 1995; DeBorde et al., 1999; Ryan et al., 1999; Schijven et al., 1999], and nonbiological colloids [Li et al., 2004; Tufenkji et al., 2004; Li et al., 2005; Tong et al., 2005; Johnson et al., 2007b]. The reentrainment behavior under unfavorable conditions is sensitive to solution chemistry and fluid flow as shown for reentrainment with perturbations, including variation in ionic strength [Ryan et al., 1999; Tufenkji and Elimelech, 2004a, 2005a; Shen et al., 2007; Mattison et al., 2011; Jiang et al., 2012a; Shen et al., 2012], variation in pH [Ryan et al., 1999; Tufenkji and Elimelech, 2004a, 2005a], and variation in fluid velocity [Shang et al., 2008; Pazmino et al., 2014a]. Traditionally, such release has been attributed to colloids that were retained in secondary minima, although more recent mechanistic simulations incorporating heterogeneity indicate that detachment from primary minima may contribute as well (as described in section 4).
The resulting retention profiles from unfavorable condition experiments (Figure 2, left and middle bottom row) have shown nonlog linear (e.g., nonmonotonic or hyperexponential) decreases in concentration as function of distance [Li et al., 2004; Tufenkji and Elimelech, 2004a; Tufenkji et al., 2004; Tufenkji and Elimelech, 2005a,b; Tong and Johnson, 2007; Liang et al., 2013]. Hyperexponential and nonmonotonic profiles have also recently been described for engineered nanomaterials [Liang et al., 2013; D. Wang et al., 2014].

Column-scale experiments have been described using continuum-scale models employing rate constants and other parameters. Trends in these parameters as a function of fluid velocity, solution ionic strength (IS), among others, allow inference of the mechanisms responsible for observed differences in colloid transport and retention in favorable versus unfavorable conditions, as described in the following section.

3. Continuum-Scale Models: Inferring Mechanisms From Kinetic Coefficients

Continuum-scale models of the transport of species in groundwater solve mass balance equations at the macroscopic scale. At this scale, microscopic properties are averaged over a representative elementary volume (REV) and the averaged property (e.g., porosity, fluid pressure, and colloid concentration) is treated as a value that represents the property throughout the REV [Bear and Cheng, 2010]. Continuum models necessarily blur microscopic details and employ simplifying assumptions in the averaging process. The result is partial differential equations that describe spatial and temporal rate of change of key macroscopic properties (e.g., fluid velocity and colloid concentration) at a practical scale. These equations are typically discretized and solved on a numerical grid that correlates REVs to nodes. The advection-dispersion equation (ADE), which describes the transport of a nonreactive solute tracer in groundwater, is a typical example (see Figure 3, tracer equation).

![Figure 3. Typical breakthrough curve profiles that illustrate how equilibrium partitioning (i.e., retardation) (solid green line) and filtration (i.e., kinetic removal) (dashed blue line) impact breakthrough behavior relative to a conservative tracer (dotted red line). The variables in the equations are defined as $C$, effluent concentration; $v$, advective velocity; $t$, time; $x$, position; $l$, porosity; $D$, dispersivity; $K_d$, partitioning coefficient; $\rho_b$, bulk density; $R$, retardation coefficient; $k$, kinetic retention coefficient; $C_o$, initial (injection) concentration; and $L$, travel length.](image)

Column-scale experiments have been described using continuum-scale models employing rate constants and other parameters. Trends in these parameters as a function of fluid velocity, solution ionic strength (IS), among others, allow inference of the mechanisms responsible for observed differences in colloid transport and retention in favorable versus unfavorable conditions, as described in the following section.

The resulting retention profiles from unfavorable condition experiments (Figure 2, left and middle bottom row) have shown nonlog linear (e.g., nonmonotonic or hyperexponential) decreases in concentration as function of distance [Li et al., 2004; Tufenkji and Elimelech, 2004a; Tufenkji et al., 2004; Tufenkji and Elimelech, 2005a,b; Tong and Johnson, 2007; Liang et al., 2013]. Hyperexponential and nonmonotonic profiles have also recently been described for engineered nanomaterials [Liang et al., 2013; D. Wang et al., 2014].

Column-scale experiments have been described using continuum-scale models employing rate constants and other parameters. Trends in these parameters as a function of fluid velocity, solution ionic strength (IS), among others, allow inference of the mechanisms responsible for observed differences in colloid transport and retention in favorable versus unfavorable conditions, as described in the following section.

3. Continuum-Scale Models: Inferring Mechanisms From Kinetic Coefficients

Continuum-scale models of the transport of species in groundwater solve mass balance equations at the macroscopic scale. At this scale, microscopic properties are averaged over a representative elementary volume (REV) and the averaged property (e.g., porosity, fluid pressure, and colloid concentration) is treated as a value that represents the property throughout the REV [Bear and Cheng, 2010]. Continuum models necessarily blur microscopic details and employ simplifying assumptions in the averaging process. The result is partial differential equations that describe spatial and temporal rate of change of key macroscopic properties (e.g., fluid velocity and colloid concentration) at a practical scale. These equations are typically discretized and solved on a numerical grid that correlates REVs to nodes. The advection-dispersion equation (ADE), which describes the transport of a nonreactive solute tracer in groundwater, is a typical example (see Figure 3, tracer equation).

Additional continuum-scale (averaged) terms are incorporated as additional mechanisms, processes, and reactions and are necessary to describe the behavior of a solute or colloid at this scale (e.g., partitioning to surfaces, filtration, degradation; examples for colloids provided in Figure 3 and further discussed below). The values (or functions) assigned to the coefficients of these terms depend on the specific solute/colloid-soil-groundwater conditions being considered and are often determined through batch studies (e.g., for sorption) or model fitting to experimental column effluent results (e.g., colloid retention).
The treatment of colloid retention in continuum-scale models differs significantly from that of solutes due to their interactions with the surface being kinetic (colloids) rather than equilibrium (solutes) [Schijven and Hassanizadeh, 2000; Tufenkji, 2007]. However, it is worth noting that solute adsorption is not always strictly equilibrium based as there is evidence that adsorption can, in some circumstances, also be a kinetic process [Bohr and Rubin, 1987; Fujikawa and Fukui, 1991; Cvetkovic and Dagan, 1994; van Kooten, 1996; Espinoza and Valocchi, 1997; Zhang and Selim, 2006; Dusek et al., 2015]. In many scenarios, the high diffusion rates of solutes allow them to reach surfaces (adsorb) and leave (desorb) readily, at rates that are typically high relative to groundwater fluid velocities (i.e., low Peclet numbers). Thus, solute-surface interactions can be represented as a local equilibrium process that can often be represented using a linear proportionality (partition constant). Partition constants are empirical (experimentally measured), independently determined from continuum-scale experiments, and the one representing distribution of solute between water and sediment is often referred to as $K_d$ (defined as $C_s / C_w$ where $C_s$ is the solid-phase concentration and $C_w$ is the aqueous phase concentration).

In contrast, colloids exhibit relatively low diffusion, which limits their ability to reach surfaces (but also makes the likelihood of colloids reaching surfaces mechanistically predictable as described in section 4). Colloid transfer to surfaces (filtration) is therefore described using kinetic retention coefficients [Harvey and Garabedian, 1991; Schijven and Hassanizadeh, 2000; Tufenkji, 2007]. The resulting breakthrough behaviors of solutes versus colloids are quite distinct (Figure 3) in that partitioning (specifically the proportionality between concentrations in water and sediment) retards transient breakthrough of the solute relative to a conservative tracer and yields complete steady state breakthrough (effluent and influent concentrations equivalent, $C_s = C_d$). In contrast, filtration (and other predominantly one-way processes such as degradation) yields reduced steady state breakthrough ($C < C_d$), and transient breakthrough is not generally retarded relative to a conservative tracer (Figure 3) [e.g., Foppen and Schijven, 2006]. It is noted that transient breakthrough of colloids may occur earlier than for tracers if Taylor dispersion (intrapore fluid velocity variation), or size exclusion of colloids from small pores (interpore fluid velocity variation), yields enhanced advection of colloids relative to solutes [e.g., Bales et al., 1989; Powelson et al., 1993; Zhang et al., 2001b; Keller et al., 2004; Chrysikopoulos and Syngouna, 2014].

The classic ADE for a nonreactive solute tracer (see Figure 3, tracer equation) is commonly modified for solute partitioning behavior (see Figure 3, partitioning equation) as well as colloid transport (see Figure 3, filtration equation). The modified-ADE commonly employed for colloid transport incorporates a single kinetic retention coefficient for colloid retention. The single-rate description of colloid transport yields continuum-scale behavior that, in the absence of colloid aggregation, generally provides an excellent match to the colloid transport and retention behavior discussed in section 2 for favorable conditions (Figure 2) [e.g., Li et al., 2004; Tufenkji and Elimelech, 2004a, 2005a].

In unfavorable conditions, the single-rate description of colloid transport often fails to capture experimentally observed breakthrough and retention behavior (the nonsteady state breakthrough and nonlog linear retention profiles described in section 2; see Figure 2). The hyperexponential retention profiles often observed in unfavorable conditions suggest that there is a high deposition rate near the porous medium inlet followed by a region with low deposition rates. A number of studies have inferred that this depth-dependent retention rate is due to a distribution of retention rates that are present within the experimental system [Schijven and Hassanizadeh, 2000; Schijven and Simunek, 2002; Tufenkji et al., 2003; Li et al., 2004; Tufenkji and Elimelech, 2004a, 2005b, 2005a; Foppen and Schijven, 2006; Foppen et al., 2007; Tong and Johnson, 2007; Chatterjee et al., 2011]. To describe this observation, a second, additional kinetic retention coefficient is often employed—termed “dual-deposition” or “two-site” models—to better fit the experimental retention profiles (discussed in section 2). The two kinetic retention coefficients represent (a) a “fast” kinetic retention coefficient to describe the early, sharp decrease in retained concentration and (b) a “slow” kinetic retention coefficient to describe the later, more gradual decline in retained concentration [Schijven and Simunek, 2002; Tufenkji and Elimelech, 2004a, 2005b, 2005a; Chatterjee et al., 2011; Mattison et al., 2011; Mondal and Sleep, 2013]. This results in an improved model fit to experimental hyperexponential profiles (Figure 2).

The hyperexponential profiles (initially observed for bacteria and protozoa) were attributed to heterogeneity among the bacterial population such that “stickier” individuals were retained upgradient of “less sticky” individuals [Albinger et al., 1994; Harvey et al., 1995; Hendry et al., 1997; Boygents et al., 1998; Simoni et al., 1994; and Bohm et al., 2000].
1998; Bolster et al., 1999; Schijven et al., 1999; Bolster et al., 2000]. Distributions in colloid size, surface charge, coatings, and hydrophobicity have been inferred to yield heterogeneity among colloid populations, and hyperexponential retention profiles as illustrated in Figure 2 [Simoni et al., 1998; Schijven and Hassanizadeh, 2000; Schijven and Šimůnek, 2002; Tufenkji et al., 2003; Li et al., 2004; Foppen et al., 2007; Tong and Johnson, 2007; Chatterjee et al., 2011].

Other studies have attributed the inferred “fast” and “slow” dual-deposition behavior to differences in interactions energies between colloids and collectors emanating partially from localized soil heterogeneities as well as deep secondary energy minima, the latter yielding “fast” retention, and the former yielding “slow” retention [Tufenkji and Elimelech, 2004a, 2005b, 2005a] on the basis that the former mechanism requires overcoming the repulsive energy barrier prior to retention. Additional studies have also suggested that soil heterogeneities, such as heterogeneities of attractive iron oxyhydroxide situated within bulk repulsive silica, can create “fast” and “slow” deposition rates in the favorable and unfavorable regions [Schijven and Hassanizadeh, 2000, 2002].

Whereas the above studies implicate soil heterogeneity in generating fast versus slow retention rates, they do not articulate how this would generate preferential upgradient versus downgradient retention (hyperexponential profiles) [Johnson and Li, 2005]. Assuming that inferred heterogeneity is distributed throughout the column (not predominantly located near the column inlet), then the kinetic retention coefficient across the column would be uniformly increased or decreased across the column by the presence of heterogeneity. Based on similar reasoning, a number of studies have concluded that soil heterogeneity alone is likely not the primary cause of this behavior [Schijven and Šimůnek, 2002; Li et al., 2004; Foppen et al., 2007; Tong and Johnson, 2007].

The above review demonstrates the utility of continuum-scale models for inferring colloid transport, retention, and reentrainment mechanisms; however, it should be noted that the kinetic parameters used in these models tend to be descriptive rather than predictive. For example, currently only the filtration rate constant under favorable conditions can be independently predicted (for spheroidal uniform colloids and media). No independently derived mechanistic basis currently exists for predicting “fast” and “slow” rate coefficients present in these dual-deposition models; rather, they require fitting to experimental breakthrough and retention profile data. Thus, dual-deposition rate models cannot be used to predict colloid transport without conducting detailed experiments a priori. The corollary is that, while this dual-deposition approach successfully produces hyperexponential profiles, the underlying mechanism(s) can only be inferred, and are not proven by a fit to data. For example, the observed decrease in attachment rate coefficient with increasing transport distance has been attributed to both (1) straining [Bradford et al., 2002, 2003, 2004, 2006a] and (2) heterogeneity among the colloid population [e.g., Tong and Johnson, 2007] which are very different mechanisms for achieving this phenomenon.

Another common example of mechanisms inferred from fitting to continuum-scale data are the dual-region or multiporosity models. The classic ADE for a one-dimensional system employs a single average pore water velocity. This single-velocity approach relies on the assumption that the velocity distribution within the porous medium can be approximated by a single, or volume-averaged, velocity term that predicts, in the absence of retardation, a classic breakthrough curve with $C/C_0 = 0.5$ at the location of the advective front. It also predicts a symmetrically sharp decrease in concentration during the elution phase of the experiment. As discussed in section 2, this approach provides an excellent description of colloid transport through uniform, homogeneous and simple (i.e., glass bead) porous media in favorable deposition conditions.

However, the anomalous early breakthrough and extended tailing behavior discussed in section 2 suggests that this average-velocity approach may not be appropriate for a wide range of water-saturated soils in both favorable and unfavorable conditions. These behaviors have been modeled at the continuum-scale by employing a distribution of permeabilities (e.g., dual-porosity or dual-permeability models). While these models are commonly employed for solute transport behavior in the presence of preferential pathways/low permeability regions and fractures/rock matrix [e.g., Haggerty and Gorelick, 1995; Haggerty et al., 2000; Šimůnek et al., 2003; Rotter et al., 2008; Šimůnek and van Genuchten, 2008], they have also been adapted for use in colloid transport, although to a relatively limited extent. Specifically, these studies have suggested that early-time colloid breakthrough and bimodal breakthrough curves may be linked to preferential flow [Leij and Bradford, 2013; Subramanian et al., 2013] which may occur in both physically heterogeneous [e.g.,...
Continuum-scale solute modeling studies have suggested that extended tailing is linked to the presence of low permeability zones and is described via dual-porosity (i.e., dual-region) or multiporosity models [e.g., Haggerty and Gorelick, 1995; Haggerty et al., 2000]. Colloid transport models typically describe tailing by assuming that a fraction of colloids are temporarily (i.e., reversibly) retained [Johnson et al., 1995b; Zhang et al., 2001a; Schijven and Hassanzadeh, 2002; Schijven and Šimůnek, 2002; Landkamer et al., 2013; Mondal and Sleep, 2013]. Conceptually, this treatment is similar to dual-region models to describe solute tailing as the retention still occurs in regions with below-average velocity, although for colloid transport in uniform media this is typically assumed to be the near-surface zone [Johnson and Hilpert, 2013]. Attribution of temporary colloid retention to residence in the near-surface fluid domain (i.e., secondary minima) can be inferred but is not proven by the match of continuum-scale models to column experiment observations. Notably, recent pore-scale modeling studies have suggested that below-average velocity regions in the bulk pore space may also temporarily retain colloids [Cardenas, 2008; Torkzaban et al., 2008; X. Q. Li et al., 2010; Z. Li et al., 2010; Li et al., 2012] suggesting that the near-surface fluid zone may not be the only contributor to extended tailing.

Mass transfer coefficients used in continuum models are often fitted parameters, determined by inverse modeling of experimental data [Köhne et al., 2009a, 2009b; Y. Wang et al., 2014b]. Results from these continuum models are powerful tools for identifying potential mechanisms and demonstrating the influence of physical heterogeneity (mobile versus immobile zones) and solution chemistry (e.g., favorable versus unfavorable attachment conditions). For example, trends in kinetic parameters describing colloid retention and reentrainment show qualitative agreement with DLVO predictions (barrier height and secondary minimum depth) as a function of ionic strength [e.g., Elimelech and O’Melia, 1990a, 1990b; Elimelech, 1991; Petosa et al., 2010]. Although quantification of rate constants through experimentation is useful, a priori determination of mechanisms, and independent derivation of rate constants, would enable prediction of colloid transport in a range of scenarios but requires mechanistic modeling as will be discussed in the next section.

4. Mechanistic Prediction of Retention in Favorable and Unfavorable Conditions

Differences between molecular (e.g., solute) and colloidal transport processes in porous media drive differing modeling approaches to describe their transport. Specifically, solutes exhibit much greater random motion and lack a deterministic trajectory so their likelihood of reaching a surface is determined empirically by laboratory analyses (e.g., batch equilibration tests). By contrast, colloids exhibit relatively limited diffusive (i.e., Brownian) motion and possess largely deterministic trajectories, enabling mechanistically based prediction of the likelihood of colloids reaching surfaces. Although for nanosized colloids, the distinction between deterministic colloidal trajectories and random diffusive solute behavior is blurred as these colloids exhibit a large degree of Brownian motion. The simulation of deterministic trajectories as well as quantification of the likelihood of colloids reaching surfaces is encompassed in CFT.

Most CFT approaches have two major components: (i) a mechanistic force/torque model that describes colloid trajectory and attachment (described in section 4.1) and (ii) correlation equations that approximate the results of the mechanistic models (described in section 4.2). A typical predicted model parameter, be it from a mechanistic model or correlation equation, is the “collector contact efficiency,” $\eta$, defined as the fraction of colloids entering the mechanistic model geometry that contacts the collector. Both mechanistic models and the resultant correlation equations have largely been successful in predicting $\eta$ for micron-sized colloids in “favorable” conditions [e.g., Yao et al., 1971; Rajagopalan and Tien, 1976; Tufenkji and Elimelech, 2004b]. It is also important to note that these approaches are subject to error arising separately at the mechanistic model and correlation equation level. In the case of mechanistic models, errors may result from employing incorrect physicochemical parameters and numerical approximations. In the case of correlation equations, errors may be due to the extent to which they are able to approximate mechanistic model results. Mechanistic models and correlation equations for unfavorable conditions, the subject of considerable recent research, are discussed in section 4.3. The following section will discuss model assumptions and their implications for the prediction of colloid and nanoparticle transport. For a conceptual and
mathematical summary of the major mechanistic models discussed, including all of the relevant force/torque equations employed by each model, the reader is referred to Tables (A1–A6) in Appendix A.

4.1. Mechanistic Simulations: Favorable Conditions

There are more than 10 different CFT mechanistic models, each employing different environmental conditions, model geometries, or force/torque mechanisms [Yao et al., 1971; Payatakes et al., 1974a, 1974b; Rajagopalan and Tien, 1976; Paraskeva et al., 1991; Burganos et al., 1992, 1994; Cushing and Lawler, 1998; Tufenkji and Elimelech, 2004b; Long and Hilpert, 2009; Ma et al., 2009; Nelson and Ginn, 2011]. Each of these mechanistic models is associated with one or more approximating correlation equations. Given the large number of mechanistic models and correlation equations, some of which are simple extensions or variations of previous models and correlation equations (discussed below), there may be some confusion related to which mechanistic model or correlation equation is applicable in a given scenario.

4.1.1. The Geometry of Mechanistic Models

The first CFT model proposed by Yao et al. [1971] considered colloids approaching a single, isolated collector that was perfectly spherical and surrounded by an infinite fluid. This approach, as noted by Yao et al. [1971], produced velocity distributions that were likely poor representations of realistic porous media. The isolated-sphere approach was subsequently modified to employ a Happel sphere-in-cell geometry [Happel, 1958] that included porosity, as well as constricted tube geometry [Payatakes et al., 1974a, 1974b; Paraskeva et al., 1991; Burganos et al., 1992, 1994]. The most popular, the Happel sphere-in-cell approach, considers the collector, otherwise isolated from the influence of other collectors and perfectly spherical, to be surrounded by an envelope of fluid that is associated with the collector. Tables A1 and A2 in Appendix A illustrate some of the model geometries that are employed in the most popular mechanistic models. This CFT-Happel approach was first employed by Rajagopalan and Tien [1976] and has been employed in more recent CFT models [Tufenkji and Elimelech, 2004b; Nelson and Ginn, 2011] as it is conceptually straightforward and simple to upscale the results. Figure 4, adapted from I. L. Molnar et al. (The impact of immobile zones on the transport and retention of nanoparticles in porous media, submitted to Water Resources Research, 2015), illustrates the Happel sphere-in-cell geometry along with the boundary conditions that would be employed by Eulerian mechanistic models [e.g., Tufenkji and Elimelech, 2004b]. The flow field around the collector is assumed to be equivalent to Stokes (i.e., creeping) flow around a sphere and undisturbed by the presence of nearby collectors (approximate velocity vectors are illustrated in the figure). $v_0$ and $C_0$ are the fluid velocity and colloid concentration upstream of the collector, respectively, and are considered uniform over the projected area upstream of the collector ($A_1$ in the figure).

Because an actual porous medium includes grain-grain contacts, it may yield flow phenomena inconsistent with Stokes flow around a Happel collector, including recirculation/vortex zones [Cardenos, 2008; Torkzaban et al., 2008; Z. Li et al., 2010] and low-flow zones [X. Q. Li et al., 2010; Li et al., 2012]. While Stokes flow incorporates low fluid velocities adjacent to the collector’s surface, the low/recirculating flow phenomena induced by grain-grain contacts likely extends this characteristic further outward. Thus, the flow field illustrated in Figure 4 and used in many mechanistic models is simplified relative to the expected actual flow field within a porous medium. It is important to understand the conditions under which this simplification is acceptable and when it may result in poor predictions of the rate at which colloids contact the surface.

Predictions from mechanistic models for medium-to-larger sized colloids (e.g., $>100$ nm) are generally in excellent agreement with experiments performed for favorable conditions [Yao et al., 1971; Rajagopalan and Tien, 1976; Tufenkji and Elimelech, 2004b; Tong and Johnson, 2006; Nelson and Ginn, 2011]. However, there is evidence that smaller colloids (e.g., $<100$ nm, including viruses and engineered nanoparticles) are potentially impacted by grain-grain contact flow phenomena in even favorable conditions [Long and Hilpert, 2009; Boccardo et al., 2014]. Mechanistic single-collector models have been observed to overpredict $\eta$ for Brownian particle transport experiments in favorable conditions [e.g., Elimelech and O’Melia, 1990a; Tong and Johnson, 2006; Long and Hilpert, 2009; Nelson and Ginn, 2011]. Hypothesizing that this is due to the simplified geometries employed by single-collector CFT models, Long and Hilpert [2009] and Long et al. [2010] explicitly incorporated the impact of grain-to-grain contacts by performing Lattice Boltzmann (LBM) colloid transport and retention simulations in randomly generated collector packings. The Long and Hilpert [2009] mechanistic model produced $\eta$ predictions for Brownian particles that were in better agreement with experimental results relative to single-collector models [Nelson and Ginn, 2011]. This improved prediction may
result from including grain-to-grain contacts and the resulting low-flow or recirculation/vortex zones that increase the distance particles must diffuse to reach certain surfaces.

The mechanistic model that incorporates grain-to-grain contact using a hemisphere-in-cell approach [Ma et al., 2009; Ma and Johnson, 2010] does still somewhat overpredict η [Nelson and Ginn, 2011]. This may be due to the orientation of the contacting grains. In the hemisphere-in-cell model, the flow direction is always perpendicular to the line connecting adjoining grain centers (see Figure 4). In a randomly packed porous medium, the flow direction occurs at many orientations relative to the line connecting adjoining grain centers, which is expected to yield a significantly larger low-flow/recirculation zone [Torkzaban et al., 2008].

Nelson and Ginn [2011] employed a Happel sphere mechanistic model to achieve a similarly good match to experiments for Brownian particles. This was likely due to an improved correlation equation (discussed in section 4.2) that better approximated the mechanistic model, notably in the low fluid velocity (i.e., diffusive) regime, as opposed to employing a better mechanistic model. Noting the improved prediction accuracy, Ma et al. [2013] subsequently adjusted their correlation equation to extend to the diffusion-dominated velocity regime.
4.1.2. Modeling Diffusion in Mechanistic Models

Complexities associated with describing, or predicting, colloid trajectories that exhibit significant Brownian motion are also likely a contributing factor to overprediction of $\eta$ by mechanistic models. The size of small nanoparticles (e.g., 5–50 nm) and the mechanisms governing their transport fall between that of molecules and larger nanoparticles (e.g., >50 nm); thus, they exhibit large Brownian motion relative to colloids but possess trajectories that are more deterministic than solutes. While the trajectories of larger, non-Brownian, colloids have been adequately described since the early numerical description by Rajagopalan and Tien (1976), the treatment of Brownian motion has varied considerably. Yao et al. (1971) solved for diffusive flux by employing the Levich (1962) solution of the convective-diffusive equation for an isolated sphere in an infinite fluid. The Levich (1962) solution is a simplified description of diffusion as it does not incorporate mechanisms such as anisotropic diffusion. Anisotropic diffusion describes the decrease in colloid diffusivity that occurs within several particle diameters of the surface. As such, Brownian motion perpendicular to the surface becomes small relative to the tangential motion [Eral et al., 2010]. Mechanistic models that ignore diffusive mechanisms such as anisotropic diffusion or hydrodynamic retardation (defined as the decrease in colloid velocity near the collector surface due to resistance from the near-surface fluid) could also yield overestimates of $\eta$ for Brownian particles as the models cannot account for the decrease in particle velocity that occurs near collector surfaces.

Rajagopalan and Tien (1976), despite using a numerical solution to solve for the trajectories of large colloids (i.e., a Lagrangian approach), solved diffusion by employing an analytical method similar to that of Yao et al. (1971) but specific to the Happel sphere-in-cell geometry [Cookson, 1970] which ignores anisotropic diffusion. Tufenkji and Elimelech (2004b) employed the Eulerian numerical model of Elimelech (1994) which solves the advective diffusive equation and included anisotropic diffusion and hydrodynamic retardation. Long and Hilpert (2009) similarly solved diffusion numerically via the ADE but did not include anisotropic diffusion. Nelson and Ginn (2011) added a scaled randomly oriented translation of the colloid to mimic Brownian motion. Ma et al. (2009) employed a scaled Brownian force so that diffusive movement was subjected to the effects of hydrodynamic retardation equivalent to other governing forces. Tables A1–A6 in Appendix A present the equations employed to describe Brownian motion in the most commonly used mechanistic models.

4.1.3. Eulerian and Langrangian Numerical Solution Frameworks

Strategies to numerically solve the mechanistic models fall broadly into two categories: Eulerian and Lagrangian frameworks. The exception is the Rajagopalan and Tien (1976) model which employed both approaches, depending on whether the trajectory of the colloid was mostly deterministic (Lagrangian approach) or Brownian (Eulerian approach). Early models tended to be Eulerian [Elimelech, 1992; Tufenkji and Elimelech, 2004b], and with increased computational power, Lagrangian mechanistic models have been adopted [Ma et al., 2009; Ma and Johnson, 2010; Nelson and Ginn, 2011; Ma et al., 2013].

While both categories of mechanistic models have successfully described colloid transport and attachment in certain scenarios, Eulerian mechanistic models have limitations, especially within the Happel sphere geometry. For Happel sphere mechanistic models employing an Eulerian or analytical framework [Rajagopalan and Tien, 1976—diffusion only; Tufenkji and Elimelech, 2004b], a constant concentration boundary condition is specified at the outer fluid envelope surrounding the collector. The concentration at this boundary is assumed to be the concentration within the bulk fluid upstream of the collector (i.e., $C_0 = C_\text{up}$, Figure 4). This constant concentration boundary condition represents the distance at which the collector ceases to exert an influence on colloid distribution in the pore space.

For Brownian colloid systems with very high removal rates and low Peclet numbers, the constant concentration boundary condition on the downstream half of the Happel fluid envelope may be inappropriate [Song and Elimelech, 1992; Nelson and Ginn, 2011]. In these circumstances, the colloid concentration at the downstream boundary would be lower than that of the upstream boundary. As a result, there would be an implicit discrepancy between the mechanistic model and what would occur in reality. This would result in overestimates of $\eta$ for Brownian colloids [Song and Elimelech, 1992].

The concentration boundary conditions also implicitly assume a concentration gradient of $\frac{C_\text{up} - C_0}{r}$ (see Figure 4) which describes diffusive flux toward the collector surface. The gradient imposed by these boundary conditions may not be representative of the actual concentration gradient within a realistic porous medium as
$C_b$ might not be equivalent to $C_o$. The pore-scale flow phenomena resulting from grain-to-grain contacts (discussed earlier) may yield larger low velocity zones than predicted by Stokes flow. This extended low-flow zone may yield distances between the collector surface and boundary $C = C_b$ that are larger than the thickness of the Happel sphere fluid envelope. If the distance is larger in reality than in the Happel sphere model, the resulting concentration gradient would be smaller than is predicted. As such the diffusive flux toward the collector surface may be lower than predicted by Eulerian mechanistic models.

With Lagrangian mechanistic models, issues related to boundary conditions are largely avoided by introducing individual particles on the upstream outer fluid envelope boundary. The trajectories of the individual colloids (i.e., attachment or exiting the system) are determined without imposing a concentration at the downstream boundary of the fluid envelope.

### 4.1.4. Stokes Flow in Mechanistic Simulations

As discussed above, mechanistic CFT models assume that the flow regime around a collector can be modeled by assuming Stokes, or creeping, flow [Yao et al., 1971; Rajagopalan and Tien, 1976; Tufenkji and Elimelech, 2004b; Nelson and Ginn, 2011] (Tables A1 and A2 in Appendix A detail the governing flow equations for the most commonly used models). Stokes flow occurs at very low Reynolds numbers (Re $<< 1$) where inertial forces become insignificant relative to viscous forces. This assumption is likely appropriate at the low flow rates typically associated with filtration and ambient groundwater (i.e., no pumping) in nonfractured systems.

The application of Stokes flow results in streamlines that are mirrored on the forward and rear flow sides of the collector. At low flow rates, this results in downstream velocity distributions that are approximately similar to the upstream flow distribution (uniform $v_o$ over area $A_1$ in Figure 4). However, in situations of high flow where Re $> 1$ [Potter and Wiggert, 2002], the flow is no longer strictly within the Stokes regime. Under these conditions, the point at which flow separates from the rear side of the collector will change (i.e., it will produce a wake) and the rear flow stagnation point will widen relative to the forward flow stagnation point [Potter and Wiggert, 2002]. This could yield discrepancies between experimental results and mechanistic simulations of colloid transport and retention.

Non-Stokes flow is likely in laboratory experiments conducted at high flow rates. Surveying the literature, and assuming reasonable parameters when values were not available (i.e., viscosity $= 1 \times 10^{-3}$ Pa s, $T = 293$ K), suggests that many experiments were conducted with Re close to or exceeding 1 [e.g., Yao et al., 1971; Elimelech and O’Melia, 1990a; Elimelech, 1991; Lecoanet et al., 2004; Lecoanet and Wiesner, 2004; Phenrat et al., 2007, 2010a]. Given this finding, conclusions derived from studies where a CFT conceptual model was adopted, but Re was also high, need to be considered with caution.nZVI experiments typically employ high flow rates with the justification that velocities near the injection well may be very large [Kocur et al., 2013]. Given that non-Stokes flow may occur near the injection well, CFT mechanistic models and correlation equations for prediction of nZVI transport must be used with caution. Overall, this suggests that careful consideration of Reynolds numbers and flow regime is required when designing column-scale transport experiments, interpreting model fits to the data, or modeling field-scale nZVI injection.

### 4.2. Correlation Equations Derived Under Favorable Conditions

The above discussion focuses on the mechanistic models underlying CFT. However, these models are time consuming to run, and so require approximation via phenomenological expressions to allow easy implementation. To this end, mechanistic models are run under a range of environmental conditions (see Table 1). Results from simulations conducted using the mechanistic models are then fit to a phenomenological expression composed of dimensionless groups of relevant physicochemical parameters. This expression serves as a correlation equation thereby serving as a simple predictor for $\eta$. Such correlation equations currently exist only for favorable attachment scenarios since the underlying mean-field mechanistic models predict no attachment in unfavorable scenarios; attempts to adapt mechanistic models for unfavorable conditions are discussed in section 4.3.

The correlation equations divide colloid transport to the surface into three distinct dimensionless numbers representing interrelated mechanisms: interception via fluid drag interactions alone, diffusion-enhanced interception, and sedimentation-enhanced interception. A separate contribution to the overall $\eta$ is determined for each dimensionless number (i.e., $\eta_I$, $\eta_D$, and $\eta_S$, respectively), such that the overall collector efficiency is described as $\eta = \eta_I + \eta_D + \eta_S$. In the absence of diffusion or sedimentation, $\eta_I$ is greater for larger colloids,
since the number of streamlines that approach the surface to within one colloid radius increases as the colloid size increases. Diffusion enhances interception since it allows colloids to move across streamlines and come into contact with the collector. Because Brownian motion is greater for smaller colloids, \( \eta_D \) is greater for smaller colloids. Because settling allows colloids to move across the streamlines and come into contact with the collector, \( \eta_G \) is greater for larger colloids. The superimposed contributions from these interrelated mechanisms yield a trend of \( \eta \) with colloid size that matches mechanistic simulations and experimental observations wherein a minimum value of \( \eta \) occurs corresponding to colloid diameters of ~1 \( \mu m \), due to smaller colloids undergoing greater diffusion, and larger colloids undergoing greater settling. As an example, the correlation equation of Rajagopalan and Tien [1976] is (corrected version presented in Logan et al. [1995])

\[
\eta \approx \gamma^2 \left[ 4A_1 \frac{1}{3} N_{Pe}^{-2/3} + A_2 N_{LO}^{1/8} N_R^{15/8} + 0.00338 A_3 N_{G}^{1.2} N_R^{-0.4} \right]
\]

(1)

where the dimensionless parameters (i.e., \( N_R \), \( N_{Pe} \), \( N_{LO} \), and \( N_{G} \)) are defined in Table 2, and \( A_i \) is a porosity-dependent parameter defined as

\[
A_i = \frac{2(1-\gamma^2)}{-2\gamma^3 + 3\gamma^5 - 2}\gamma^6
\]

(2)

where \( \gamma = (1-\theta)^{1/3} \). \( A_i \) accounts for the influence of neighboring collectors on the fluid flow field in the Happel sphere-in-cell geometry and \( \theta \) represents the porosity of the porous medium. Tables A1 and A2 in Appendix A list the correlation equations for the six most commonly used models.

Figure 5 illustrates how well the correlation equations listed in Table 1 predict the experimental kinetic retention coefficient \( k \). Figure 5 compares experimental and predicted values of \( k \) (the upscaled \( \eta \)) since the definition of \( \eta \) differs among the different collector geometries (Appendix A, Tables A1 and A2 provide the equation for \( k \) for each model). As can be seen, the currently existing correlation equations generally provide very good predictions of \( \eta \) for micron-sized colloids in the range of conditions for which the correlations were developed [e.g., Tong and Johnson, 2006; Nelson and Ginn, 2011]. However, they generally overpredict \( \eta \) for Brownian particles (Figure 5, the “nanoparticles” region). While a number of possible causes for this overprediction were identified in the discussion related to mechanistic models, another possible source of error may be the approximating correlation equations themselves. Many of the flow and transport scenarios for Brownian particles involve parameter values that are outside the limited range of conditions employed to derive the correlations [e.g., Quinn et al., 2005; Kocur et al., 2013; Krol et al., 2013; O’Carroll et al., 2013; Gastone et al., 2014; Kocur et al., 2014; Tosco et al., 2014b].

This is often the case for the prediction of engineered nanoparticle transport. For example a limited range of viscosities, similar to water, have been used in mechanistic model simulations used in the development of correlation

| Table 1. Range of Parameter Values Employed to Derive Recent \( \eta \) Correlation Equations |

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle Diameter, ( d_p (\mu m) )</td>
<td>Collector Diameter, ( d_r (\mu m) )</td>
</tr>
<tr>
<td>TE 2004</td>
<td>0.01–10</td>
</tr>
<tr>
<td>LH 2009</td>
<td>0.1–1</td>
</tr>
<tr>
<td>Long et al. [2010]</td>
<td>0.02–1</td>
</tr>
<tr>
<td>MPFJ 2009</td>
<td>0.05–10</td>
</tr>
<tr>
<td>MHJ 2013</td>
<td>0.01–10</td>
</tr>
<tr>
<td>NG 2011</td>
<td>0.01–10</td>
</tr>
</tbody>
</table>


| Table 2. List of Dimensionless Parameters in Predicting Colloid Filtration |

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_R )</td>
<td>( a_p/a_s )</td>
<td>Aspect ratio</td>
</tr>
<tr>
<td>( N_{Pe} )</td>
<td>( U_d/D_{BM} )</td>
<td>Peclet number</td>
</tr>
<tr>
<td>( N_{LO} )</td>
<td>( H/\left[ 9\pi \eta \mu U \right] )</td>
<td>London number</td>
</tr>
<tr>
<td>( N_G )</td>
<td>( 2a_p^2 \left[ \rho_p/\rho_f - 1 \right]/g \left[ \eta \mu U \right] )</td>
<td>Gravity number</td>
</tr>
</tbody>
</table>

\( ^{\ast} a_p \) and \( a_s \) are the colloid and collector radii; \( d_p \) is the collector diameter, \( U \) is the approach velocity; \( D_{BM} \) is the bulk diffusion coefficient (described by Stokes-Einstein equation); \( \mu \) is the fluid viscosity; \( \rho_f \) and \( \rho_p \) are the fluid and colloid densities, respectively; \( g \) is the gravitational acceleration constant.
Long and Hilpert [2009] and Tufenkji and Elimelech [2004b] derived their correlation equation using a single value of water kinematic viscosity at 25°C (0.831026 m²/s) and Nelson and Ginn [2011] employ a range of viscosities, but for a range of water temperatures from 5 to 30°C. However, ENPs are often stabilized in a polymer solution to prevent aggregation and settling [e.g., Hotze et al., 2010; Phehrat et al., 2010b, 2010c; Kocure et al., 2013]. These stabilized ENP solutions tend to be much more viscous than water: 2.31023 to 1.31022 Pa s [Krol et al., 2013], 71023 Pa s [Kocur et al., 2014], 61023 Pa s [Sakulchaicharoen et al., 2010], 1.21022 to 8.81022 Pa s [Gastone et al., 2014; Tosco et al., 2014b] and upward of 1.942 Pa s [Quinn et al., 2005]. Since current η correlation equations were not derived for this range of viscosities, caution should be used in applying existing correlations to engineered nanoparticle subsurface applications. Extending CFT models to include a wider range of viscosities should be considered for future research.

Similarly, η correlation equations, with the exception of Ma et al. [2013], were derived for a relatively small range of colloid densities (i.e., 1–1.8 g/cm³) (Table 1). While viruses, bacteria, and protozoa likely fall within this near-neutrally buoyant range, aggregates of metallic ENPs may have significantly higher densities. No correlation equation has been derived using a mechanistic CFT model that explicitly considers high-density colloids such as aggregates of metallic ENPs. Nor have existing correlation equations been experimentally validated for high-density colloids.

4.3. Mechanistic Simulations: Unfavorable Conditions

Mechanistic force/torque balance simulations that use mean-field DLVO interactions predict zero attachment even under conditions of modest repulsion (e.g., interaction energy > several kT, see Figure 1) between colloids and grain surfaces [Elimelech and O’Melia, 1990b]. Yet as discussed earlier, unfavorable conditions are observed in numerous environmental scenarios. As a result, we have until recently lacked mechanistic theory to predict colloid attachment and transport in environmental contexts, and there as yet exists no easily implemented approximating correlation equation for mechanistically based prediction of colloid attachment under unfavorable conditions.
The inability of mean-field force/torque balances to predict colloid attachment in unfavorable conditions has driven the development of semiempirical expressions to predict attachment [Elimelech, 1992; Bai and Tien, 1996, 1999; Chang et al., 2009]. Like the approximating correlation equations for favorable conditions described above, these semiempirical correlations are composed of dimensionless groups of relevant physicochemical parameters that were fit via coefficients and powers. An example correlation from Elimelech [1992] is provided:

\[
\eta_{\text{unf}} = 0.0257 (N_{\text{col}})^{1.19} \eta_{\text{fav}}
\]

where \( \eta_{\text{unf}} \) and \( \eta_{\text{fav}} \) are the collector efficiencies in the presence and absence of repulsion, respectively, and \( N_{\text{col}} \) represents the ratio of attractive to repulsive colloid-collector interactions and is defined as follows:

\[
N_{\text{col}} = \frac{A_{132}}{\varepsilon_0 k \varepsilon_r \varepsilon_c k^{-1}}
\]

where \( A_{132} \) is the combined Hamaker constant for the colloid, water, and porous media, \( \varepsilon_0 \) and \( \varepsilon_r \) are the absolute and relative permittivities in vacuum and water, respectively. The parameters \( \varepsilon_p \) and \( \varepsilon_c \) are the mean-field potentials of the colloid and collector surfaces, and \( k^{-1} \) is the inverse Debye length. The reported fitted coefficient (0.0257) and power (1.19) correspond to the general set of experiments reported by Elimelech [1992]. Please note that throughout this document, we do not use subscripting of \( \eta \) to denote whether it represents favorable versus unfavorable conditions, except in circumstances where direct comparison of these two conditions is required.

Semiempirical correlation coefficients and powers were fit directly to experimental data rather than to mechanistic simulations. They are therefore limited to the experimental conditions from which they were developed, and their relative performance across a larger range of conditions is generally poor when applied outside the experimental conditions under which they were developed. In contrast, the approximating correlation equations for favorable conditions typically show relative errors less than a factor of three (i.e., the error is less than 3X the estimated \( \eta \)) [Nelson and Ginn, 2011; Ma et al., 2013]. The relatively greater success of correlation equations developed for favorable conditions may reflect their having been fit to mechanistic simulations, which ideally extends their utility to a larger range of conditions, but may also reflect the greater complexity of colloid transport under unfavorable conditions.

The ratio \( \frac{\eta_{\text{unf}}}{\eta_{\text{fav}}} \) provides a measure of the effect of repulsion and is called the collision or attachment efficiency \( \alpha \). The value of \( \alpha \) should therefore range from 0 to unity. Investigators can obtain \( \alpha \) by comparing experiments conducted under unfavorable versus favorable conditions, or as more often is the case, by substituting favorable condition correlation equation predictions of \( \eta_{\text{fav}} \). Whereas the latter approach yields reasonable estimates of \( \alpha \) for many conditions, it may yield values of \( \alpha > 1 \) for less unfavorable (nearly favorable) conditions. This reflects the fact that, as described above, the approximating correlation equations for favorable conditions are just that, approximations of the underlying mechanistic models, and so they may introduce error via the approximation. Furthermore, even the underlying mechanistic models may introduce error if the experimental conditions “violate” assumptions in the underlying mechanistic models (e.g., uniformity, spheroidal colloids, and collectors) as is also true for predictions under favorable conditions.

### 4.3.1. Secondary Minimum Association

Whereas mean-field mechanistic models do not predict immobilization (attachment) of colloids in the presence of significant repulsion, they do predict the accumulation of colloids in the near-surface fluid domain due to secondary minimum attraction under unfavorable conditions [e.g., Johnson et al., 2007a]. The experimentally observed “skimming” behavior of colloids in the near-surface fluid domain under unfavorable conditions [e.g., Johnson et al., 2010] is predicted by mean-field mechanistic models, for example, in dense and simple cubic packing structures [e.g., Johnson et al., 2007a] and impinging jet systems where flow is directed normal to a planar surface and spreads radially from the flow stagnation axis [Johnson and Hilpert, 2013].

The influence of secondary minimum interactions is accounted for in mean-field force/torque models underlying CFT but this has received limited attention. The reason being that these models (often
performed in the Happel unit cell) were usually run under favorable conditions (absence of repulsive barrier yields absence of secondary minimum) but would need to be run under unfavorable conditions in order for the influence of the secondary minimum to be observed. Furthermore, when these models were run under unfavorable conditions, the primary purpose was to demonstrate lack of immobilization (lack of attachment) for significant repulsion [e.g., Elimelech and O’Melia, 1990b].

The possibility that experimentally inferred colloid retention at the column scale may include secondary minimum-associated colloids prompted the development of easily adopted approaches to predict potential colloid retention via this mechanism. Foremost among these is the Maxwell approach [Franchi and O’Melia, 2003; Hahn and O’Melia, 2004; Shen et al., 2007; Tosco et al., 2009; Sang et al., 2013], which determines the kinetic energy distribution among the colloid population from the Maxwell frequency distribution of their velocities. Retained colloids are those with insufficient kinetic energy to escape from the secondary energy minimum into the bulk fluid domain. The Maxwell approach is a simplified version of CFT in that it does not account for fluid flow, collector geometry, or mechanistic force/torque balance, such that the subsequent fate (e.g., exit from the collector) of colloids associated with the surface via secondary minimum interactions is not actually determined.

The utility of the Maxwell approach lies in its relative ease of application relative to a full force/torque balance. It must be stressed that although both the Maxwell or mean-field mechanistic force/torque balance approaches predict accumulation in the secondary energy minima, they predict no colloid attachment to (i.e., immobilization on) surfaces (except for energy barriers less than several kT).

4.3.2. Immobilization

As mentioned above, mean-field mechanistic models do not predict the experimentally observed immobilization of a significant fraction of near-surface colloids [e.g., Johnson et al., 2010] on bulk repulsive surfaces. Instead, as stated above, colloids are predicted to skim (spin and translate) outboard of the energy barrier in the near-surface fluid domain. As such, the challenge in prediction of colloid attachment (immobilization) is how to achieve physical contact with the surface when mean-field prediction indicates that it is prevented by the repulsive barrier. A number of strategies have been recently employed to mechanistically predict this immobilization. These strategies can be categorized as (1) macroscopic heterogeneity, (2) implied physical heterogeneity, and (3) representative discrete heterogeneity.

Heterogeneity has, for at least 20 years, been acknowledged as a potential explanation for the discrepancy between mechanistic predictions and experimental observations in environmental (unfavorable) conditions [Ryan and Elimelech, 1996]. Since then, research has confirmed the potential role of heterogeneity in colloid retention in environmental conditions. While our understanding of the processes governing colloid-heterogeneity interactions has evolved considerably, there remain a number of questions regarding the role and significance of heterogeneity at macroscopic and microscopic scales as described below.

4.3.2.1. Macroscopic and Implied Physical Heterogeneity

Macroscopic surface heterogeneity (i.e., readily measured surface properties) arises from, for example, the presence of multiple mineral surfaces with different isoelectric points such that some surfaces are positively charged, and some are negatively charged (e.g., iron oxyhydroxides versus silica) at circum-neutral pH. For porous media having macroscopic geochemical heterogeneity, the unfavorable fraction of the surface can be approximated as having $\alpha = 0$, and the favorable fraction of surface can be approximated as having $\alpha = 1$. This approach is successful in that a simple linear combination of the binary $\alpha$ values weighted by the corresponding fractions of favorable versus unfavorable surfaces in the porous media yields reasonable prediction of observed retention in macroscopically heterogeneous media [Song et al., 1994; Johnson et al., 1996; Chen et al., 2001; Loveland et al., 2003].

The primary limitation of this approach is that while macroscopic surface heterogeneities are widespread in environmental porous media, there is significant evidence that these heterogeneities may only minimally influence colloid transport in soils. Natural organic matter (NOM) has been shown to sorb strongly to positively charged (i.e., favorable) mineral surfaces such as ferric and aluminum oxyhydroxides and switch their surface charges from positive to negative [Amirbahman and Olson, 1993; Gu et al., 1995; Johnson and Logan, 1996; Pieper et al., 1997; Mosley et al., 2003; Mylon et al., 2004; Foppen et al., 2008; Y. Liu et al., 2009; Abudalo et al., 2010; D. Wang et al., 2012]. The sorption of NOM to surfaces that are favorable for colloid attachment
will potentially yield unfavorable attachment conditions. Given the widespread prevalence of NOM throughout environmental media, many macroscopic heterogeneities may be coated by organics that yield unfavorable conditions thereby limiting the relevance of macroscopic heterogeneities in environmental media. A large number of colloid transport studies have demonstrated this NOM “masking effect” in which the presence of organic matter can increase colloid transport [Amirbahman and Olson, 1993; Johnson and Logan, 1996; Peiper et al., 1997; Mylon et al., 2004; Foppen et al., 2008; Pelley and Tufenkji, 2008; Abudalo et al., 2010; Morales et al., 2011; Jiang et al., 2012b; D. Wang et al., 2012].

Environmental media coated with NOM cannot be represented by this “patchwise” linear approximation since the entire media would be represented by $\gamma = 0$; whereas significant colloid retention is experimentally observed in such media [Elimelech and O’Melia, 1990b; Li et al., 2004; Tong and Johnson, 2006; Johnson et al., 2010]. The discrepancy suggests the presence of microscale to nanoscale heterogeneity (i.e., not readily measurable surface properties) on the bulk repulsive surface, which may allow attachment to occur [e.g., Elimelech and O’Melia, 1990b] as described below in section 4.3.2.2.

The inferred presence of microscale to nanoscale heterogeneity on surfaces has spurred the development of a variety of mechanistic approaches to account for nanoscale to microscale physical and chemical heterogeneity. One approach that not only capitalizes on the predicted accumulation of colloids in secondary minima but also provides a mechanism to immobilize a subset of these secondary minimum-associated colloids is herein called the “implied physical heterogeneity approach” [Torkzaban et al., 2008]. The approach extends surface friction, which is traditionally considered to act at the primary minimum [e.g., Johnson et al., 1971; Bergendahl and Grasso, 2000], out to the secondary minimum, which is traditionally considered “non contact” such that the relevant friction in the secondary minimum arises from fluid viscosity [e.g., Rajagopalan and Tien, 1976]. Extending surface friction to the secondary minimum replaces the traditionally assumed relatively weak friction from fluid viscosity with stronger surface friction that immobilizes colloids in zones of low fluid drag (e.g., low velocity or recirculation zones). The approach effectively treats the near-surface fluid domain (see Figure 1) as a highly viscous fluid yielding a friction coefficient equal to that of the surfaces in contact. This method is justified on the basis that there can exist roughness, macromolecules, etc. that may effectively extend surface friction to the attractive secondary well [Johnson et al., 2009].

The approach is successful in that it identifies zones on the surface where colloids can be immobilized due to drag forces being less than the extended friction force. However, an important limitation of the approach is that it has not been reconciled with existing work that considers secondary minimum interactions as non-contact forces [Israelachvili, 2011], and it does not build on the formal approach provided by the previously existing body of work accounting for steric forces associated with water structure, roughness, and macromolecules in the near-surface domain [Israelachvili, 2011].

With respect to predictions, the implied physical heterogeneity approach predicts attachment only in zones of low fluid drag [Torkzaban et al., 2007, 2008; Bradford et al., 2011a]; whereas experiments demonstrate colloid deposition in high fluid drag zones such as forward flow stagnation zones [Johnson et al., 2010; Pazmino et al., 2014b] under unfavorable conditions. Forward flow stagnation zones are regions where flow impinging normal to the collector surface is converted to tangential flow. Tangential flow is zero on the forward flow stagnation axis that is defined as the center of the impinging flow. Normal flow on this axis decays with decreasing separation distance from the surface to a minimum (taken to be zero) at the forward flow stagnation point on the collector surface. Adjacent to the collector surface outside the forward flow stagnation point, tangential flow and tangential drag force on colloids rapidly increase. As a result, the forward flow stagnation zones are not zones of low fluid drag except at the forward flow stagnation point (identified in Figure 4). The same principle applies to the rear flow stagnation zone and point, where tangential flow converging on the flow stagnation axis is converted to normal flow away from the grain surface. Simulations utilizing implied physical heterogeneity also predict that colloids will immobilize (attach) at long distances (e.g., multiple tens of nm) from the surface, in which case detachment will occur in response to very slight increases in fluid drag [Pazmino et al., 2014a]. In contrast, direct observation experiments show that detachment requires fluid velocity perturbations significantly higher than the loading velocity [Pazmino et al., 2014a], as further described below.

### 4.3.2.2. Microscale to Nanoscale Heterogeneity

The existence of microscale to nanoscale heterogeneity (herein called discrete heterogeneity) on bulk repulsive surfaces is inferred from the observed attachment of colloids under unfavorable conditions lacking...
macroscopic heterogeneity. The rationale for this inference is that repulsion experienced by colloids from a bulk like-charged surface may be locally eliminated or reduced by the presence of physical asperities (protrusions associated with roughness) and charge heterogeneity [Bhattacharjee et al., 1998; Hoek and Agarwal, 2006]. As illustrated in Figure 6, physical asperities reduce repulsion because interaction forces scale directly with radius of curvature [Israelachvili, 2011], and the asperity reduces (locally) the radius of curvature associated with colloid-surface interaction. Charge heterogeneity creates local zones of opposite charge (attraction) that scales with the size of these zones, as illustrated in Figure 6.

Whereas macroscopic heterogeneity (e.g., metal oxyhydroxide coatings on silica) may be removed via acid-base cleaning, nanoscale to microscale heterogeneity from roughness and cation substitution in the mineral matrix are not removed, and in fact may be enhanced, by cleaning approaches [Litton and Olson, 1993; Johnson et al., 1996; Tong and Johnson, 2006; X. Y. Liu et al., 2009; DiCarlo et al., 2010; Molnar et al., 2011; Jiang et al., 2012a]. The sometimes-invoked assumption that colloid retention on cleaned surfaces must occur via straining (due to presumed absence of heterogeneity) has been discussed [Johnson et al., 2011] (see section 5.3 for a discussion on successes and opportunities in straining research).

4.3.2.3. Representative Discrete Heterogeneity

Another approach to achieve physical contact between the colloid and the collector surface in spite of significant repulsion between the bulk surfaces is to replace the mean-field colloid-collector interactions with interactions that explicitly recognize the existence of discrete nanoscale to microscale physical and chemical heterogeneity on the collector (or potentially colloid) surface where colloid-collector repulsion is locally reduced or eliminated (herein referred to as heterodomains) [e.g., Duffadar and Davis, 2007; Bendersky and
Heterodomains differ from the previously discussed macroscopic surface heterogeneities as they are not readily measurable surface properties. These heterodomains, illustrated in Figure 6, can bring colloids into physical (e.g., primary minimum) contact with the surface, where surface friction is more generally accepted to be applicable [Johnson et al., 1971; Bergendahl and Grasso, 2000], as described in detail in Pazmino et al. [2014b].

Discrete heterogeneity on a collector surface yields different net interactions (attractive versus repulsive) for different colloid sizes [Pazmino et al., 2014b]. Due to curvature of the colloid surface and the rapid decay of colloid-collector interactions with increasing separation distance, the zone of significant colloid-collector interaction (ZOI) is a fraction of the projected area of the colloid. The ZOI area decreases with increasing ionic strength [Duffadar et al., 2009; Pazmino et al., 2014b]. The net colloid-collector interaction is determined by the combination of attractive and repulsive interactions across the ZOI, which is determined by the fraction of the ZOI that is occupied by heterodomain(s) [Pazmino et al., 2014b, 2014a]. Hence, the net colloid-collector interaction (attractive versus repulsive) depends on the interplay of colloid size, heterodomain size, and ionic strength, as shown in Figure 7 adapted from Pazmino et al. [2014b]. In the figure, a given heterodomain size and ionic strength condition yield a net interaction force that depends directly on colloid size, as shown by the force curves in Figure 7, where blue yields no repulsion and red yields the largest repulsion. This occurs because the radius of the ZOI ($R_{ZOI}$) increases with colloid size, as shown by spheres in Figure 7. The fractional coverage (of ZOI) by heterodomains decreases as the colloid size increases relative to a given heterodomain size, leading to net attractive interaction at all separation distances for the smallest (blue) colloid, and net strong repulsion at intermediate distances (repulsive barrier) for the largest (red) colloid.

Successes of the discrete heterogeneity approach, referring to the incorporation of nanoscale zones of attraction in the DLVO interaction, include (i) colloid attachment on the open surface and (ii) wedging in...
The outcomes that emerge from the force/torque simulations utilizing discrete heterogeneity match experimental observations across an array of colloid sizes, fluid velocities, and ionic strength values. The discrete heterogeneity approach, by developing a representative heterogeneity for a surface, provides a mechanism to represent differences in heterogeneity among different mineral types and different pH conditions. These changes are not captured by mean-field parameters such as bulk surface zeta potential [Elimelech et al., 2000]. Representative discrete heterogeneity provides a potential platform for mechanistic representation of blocking or ripening, both of which begin with a limited number of colloid attachment sites. Heterodomains (representative heterogeneity) would serve as these initial attachment sites, with subsequent decrease (blocking) or increase (ripening) in the number of attachment sites with increased colloid attachment, depending on whether colloid-colloid interactions are repulsive or attractive, respectively.

As explained above, for a given heterogeneous surface, the net interaction (attractive versus repulsive) depends on colloid size and surface properties. As such, slight differences in colloid size and/or surface properties may generate effective heterogeneity in “stickiness” among the colloid population. Demonstrating this effect may yield a mechanistic explanation for the apparent heterogeneity in retention rate coefficients among apparently uniform populations, which is thought to drive the observed nonlinear linear profiles of retained colloids observed under unfavorable conditions [Albinger et al., 1994; Harvey et al., 1995; Hendry et al., 1997; Baygents et al., 1998; Simoni et al., 1998; Bolster et al., 1999; Schijven et al., 1999; Bolster et al., 2000; Li et al., 2004; Tufenkji and Elimelech, 2004a; Tufenkji et al., 2004; Tufenkji and Elimelech, 2005a,b; Tong and Johnson, 2007; Liang et al., 2013; D. Wang et al., 2014].

5. Upscaling From Pore to Continuum Scale

The parameter $\eta$ describes the efficiency of colloid delivery to the surface (favorable conditions) or the near-surface fluid domain (unfavorable conditions) at the pore scale. Its implementation into continuum-scale models (upscaling) is performed under the simple assumption that the control volume considered in the continuum model is represented by a collection of identical pore-scale collectors comprising the same volume and having the same porosity as the control volume [Logan et al., 1995; Nelson and Ginn, 2011;
Johnson and Hilpert, 2013]. Under this assumption, the equation expressing $\eta$ as a rate constant ($k$) for the Happel sphere-in-cell geometry is

$$
\frac{3(1-\eta)^{1/3}}{2d_e} \ln(1-\eta)v
$$

(5)

where $d_e$ is the average collector diameter, $v$ is velocity, and $\theta$ is porosity. The functional form of $k$ is not the same for all $\eta$ correlations, it is a function of the conceptual model employed; the hemisphere-in-cell [Ma et al., 2009] and constricted tube [Paraskeva et al., 1991] geometries have different relationships for $k$.

5.1. Opportunities in Secondary Minimum Interactions

The $k$-$\eta$ relationship (equation (5)) describes colloid delivery to the near-surface fluid, which under favorable conditions is practically equivalent to attachment. Under unfavorable conditions, the fate of near-surface colloids is less straightforward. While a fraction of near-surface (secondary minimum-associated) colloids may attach in response to local attraction emanating from heterogeneity, the complement may translate across the surface until they return to the bulk fluid.

While equation (5) is specifically for porous media, it is worthwhile mentioning that it is possible to link the observed deposition rates in impinging jet systems to mechanistically predicted collector efficiencies in pore-scale models [Pazmino et al., 2014]. In contrast, comparison of $k$ between these systems requires relating the flow fields between them, and this likely requires incorporating discrete heterogeneity representations into Happel sphere-in-cell geometry.

Additionally, the presence of secondary minima in unfavorable conditions may yield an indefinite residence time of colloids in the near-surface fluid domain. Secondary minimum interactions will influence the average velocity of the colloid population to an extent depending on $\eta$ as well as the depth of the secondary minimum, since velocity in the near-surface fluid is far lower than the average pore water velocity. Furthermore, secondary minimum-associated colloids may be effectively retained in the control volume (e.g., packed column in experiments) if elution does not occur prior to termination of the observation. Currently, there exists no relationship to quantitatively account for the influence of secondary minimum association on the concentration and timing of colloid elution. Developing such a relationship will require identifying the frequency with which colloids transfer between the bulk and near-surface fluid domains, as discussed below.

5.2. Opportunities in the Role of Topology

The fact that mechanistic simulations and upscaling provide good prediction at the continuum scale under favorable conditions suggests that mixing of near surface and bulk fluid must occur between subsequent collectors (grains) in the control volume [Johnson and Hilpert, 2013]. This is inferred on the basis that under favorable conditions in the absence of mixing, the near-surface fluid would become depleted of colloids, increasingly so with each subsequent collector, yielding decreasing $k$ with increased transport distance [Johnson and Hilpert, 2013]. The log linear profiles of retained colloids that are typically observed under favorable conditions indicate that $k$ remains constant with transport distance. Hence, mixing of near-surface and bulk fluid between subsequent collectors is reasonably inferred. In contrast, absence of mixing between collectors under unfavorable conditions would yield increasing excess of colloids in the near-surface fluid (via secondary minimum attraction), increasingly so with each subsequent collector, such that $k$ would increase with increased transport distance [Johnson and Hilpert, 2013]. Whereas $k$ does vary with transport distance under unfavorable conditions (e.g., hyperexponential and nonmonotonic profiles of retained colloids), it does not strictly increase with transport distance.

Nevertheless, the association of distance-varying $k$ and extended tailing with unfavorable conditions suggests a relationship with colloid accumulation in the near surface under unfavorable conditions. Johnson and Hilpert [2013] speculated that incomplete mixing of bulk and near-surface fluid between subsequent grains may at least partially drive the observed extended tailing (and nonlog linear retained profiles) under unfavorable conditions. This, in addition to the above-described potential influence of discrete heterogeneity on amplifying heterogeneity among the colloid population, provide potent mechanisms to potentially explain these critical phenomena observed under unfavorable conditions.

Since colloid transfer from the bulk to the near-surface fluid occurs at forward flow stagnation points, and colloid transfer from the near-surface to the bulk fluid occurs largely at rear flow stagnation zones (less so...
for Brownian-dominated colloids [Johnson and Hilpert, 2013], the degree of mixing between collectors is a topological issue that depends on the alignment of the local flow field with the grain-to-grain contacts [Johnson and Hilpert, 2013]. Whereas experimental evidence that topology (degree of alignment of flow field and grain-to-grain contacts) governs retention at the continuum scale has been provided [Johnson and Hilpert, 2013], the opportunity exists for many additional studies to explore experimentally and theoretically the relationship of observed profiles of retained colloids to topology (as governed by packing structure). Such information can guide development of upscaling techniques for nonuniform media.

An additional mechanism potentially driving extended tailing under unfavorable conditions is provided by recent pore-scale modeling studies that suggest the possibility of indefinite colloid retention under unfavorable conditions in low-flow/recirculation zones associated with grain-grain contacts (discussed in section 4.1.1) [Cardenas, 2008; Torkzaban et al., 2008; X. Q. Li et al., 2010; Z. Li et al., 2010; Li et al., 2012]. Temporary hydraulic retention of colloids in these regions could drive long-term tailing as colloids slowly exit the immobile zones via Brownian motion (Molnar et al., submitted manuscript, 2015). More generally, solute transport demonstrates diffusive mass transfer between advective and nonadvective zones [e.g., Haggerty and Gorelick, 1995; Haggerty et al., 2000; Berkowitz et al., 2006; Scheibe et al., 2013]. There is opportunity to distinguish the influence of near-surface residence from the influence of mass transfer between bulk advective versus nonadvective zones in colloid transport. This is particularly true for nanoparticles, for which it is not yet demonstrated whether their diffusivity will allow them to enter bulk nonadvective domains and whether this entry would require skimming along the secondary minima (as required by micron-sized colloids) or direct diffusive mass transfer that could occur in both favorable and unfavorable conditions.

5.3. Opportunities in Straining

Another critical aspect of nonuniformity of porous media is the expectation that a fraction of the pore throats will be too small for colloids to pass (i.e., straining) [Bradford et al., 2002, 2003, 2004; Tufenkji et al., 2004; Bradford et al., 2006a, 2006b; Shen et al., 2008; Bradford et al., 2013]. The process of straining is not included in CFT because, while both filtration and straining lead to attachment, straining has no dependence on diffusion, sedimentation, and limited dependence on colloid-surface interaction. Early work on straining [Sakthiavadiel, 1966, 1969; Herzog et al., 1970; McDowell-Bayer et al., 1986] identified the ratio of mean particle:collector diameter \(d_p/d_c\) as the critical factor in determining straining. These early works presented a range of critical \(d_p/d_c\) ratios required for straining, such as >5% [Sakthiavadiel, 1966, 1969] and >18% [Matthes et al., 1985]. These large critical ratios suggest that only the largest (i.e., \(d_p \sim 10 \mu m\)) colloids in fine textured soils would be strained. However, there has recently been a large volume of research published inferring (see section 3 for a discussion on the limitations of inferred transport behavior) colloid straining as a significant removal mechanism for a wide range of colloid and collector sizes [Bradford et al., 2002, 2003, 2004; Tufenkji et al., 2004; Bradford et al., 2005; Poppen et al., 2005; Bradford et al., 2006a, 2006b; S. Xu et al., 2006; Gaillard et al., 2007; Shen et al., 2008; Xu et al., 2008; Xu and Saiers, 2009; Chowdhury et al., 2011; Porubcan and Xu, 2011; Sagee et al., 2012; Bradford et al., 2013; Du et al., 2013; Raychoudhury et al., 2014]. These studies have inferred that \(d_p/d_c\) ratios can be much lower than originally reported (as low as 0.01%) [Bradford et al., 2002, 2003, 2004; W. Xu et al., 2006; Shen et al., 2008; Xu et al., 2008; Tosco and Sethi, 2010; Raychoudhury et al., 2014]. In addition, a number of studies have suggested that straining is not strictly a function of pore throat and colloid size by suggesting that flow velocity [Bradford et al., 2006b; Du et al., 2013], input concentration [Bradford and Bettahar, 2006], ionic strength [Shen et al., 2008], and colloid shape [Xu et al., 2008] can influence straining behavior. However, many of these studies base their evidence for straining on a priori assumptions that may not be correct (e.g., that acid washed sand possesses no macroscopic surface impurities ergo straining is predominant) [Johnson et al., 2011]. These a priori assumptions have led to, for instance, a number of studies suggesting that even nanosized colloids (with \(d_p/d_c\) ratios orders of magnitude lower than suggested in the early literature) can be strained out of solution [Tosco and Sethi, 2010; Sagee et al., 2012; Raychoudhury et al., 2014].

Many of the above investigators relate straining to wedging; however, wedging is a mode of filtration for larger colloids, whereas straining is entrapment in pore throats too small to pass. Wedging is a mode of attachment (filtration) that involves mass transfer from the bulk fluid to the near-surface fluid and eventual contact with two surfaces simultaneously in order to create the adhesive resisting torque that is needed to arrest larger colloids which experience a large drag torque [Ma et al., 2011]. Thus, wedged colloids can be much smaller than the pore throats through which they pass [Johnson et al., 2010], and wedging is subject to the same force/
torque balance governing colloid attachment to the open surface [Johnson et al., 2007a; Ma et al., 2011]. In contrast, straining is purely a matter of colloid size being larger than the pore throat, and as such does not concern the mass transfer of colloids from the bulk fluid to the near-surface fluid domain. Instead, straining is a process that cannot be investigated at the single pore scale and instead requires investigation at the assemblage scale where pore throat size distribution can be determined and compared to colloid size distribution.

While the models developed to incorporate straining at the continuum-scale can accurately reproduce colloid retention behavior when fit to experimental data [Bradford et al., 2003, 2004; S. Xu et al., 2006; Flury and Qiu, 2008; Shen et al., 2008], they are subject to the same limitations of inferred mechanisms discussed in section 3 and elsewhere in the literature (Johnson et al., 2011). Developing mechanistic methods to predict colloid straining will require an understanding of the pore throat distributions and flow fields. Pore-scale visualization methods such as micromodels and X-ray computed tomography (XCT) are uniquely suited to this purpose.

### 5.4. Opportunities in Imaging and X-Ray Computed Microtomography

Flow and transport within porous media are governed by the pore size distribution and the topology, i.e., the way in which pores of different sizes are connected [Vogel and Roth, 2001]. Thus, while micromodels, Happel spheres, cubic packings, etc. are useful for studying fundamental colloid transport processes and mechanisms, the idealized or simple pore structures do not represent environmental media. Over the past two decades or so, X-ray computed tomography (XCT) has become a relatively standard approach for extracting grain and pore-scale features of unconsolidated porous media (see recent review by Wildenschild and Sheppard [2013]) due to its ability to "see inside" of three-dimensional opaque materials. XCT is a nondestructive, noninvasive technique for acquiring 3-D data sets where the fundamental unit, the volume element (voxel, has a value that represents the average X-ray absorption of the material within the voxel. This X-ray absorption value is a function of density, atomic number, and the energy of incident X-rays. In this way, it is possible to differentiate between different materials. Therefore, the primary information is simply the voxel gray value (mass linear attenuation, which roughly corresponds to absorbance for SXCM and CT# for conventional XCT). The major advantage of microtomography over conventional 2-D methods (e.g., thin sections, microscopy) is the level of quantitative information that can be derived from a full 3-D property map.

While the most common use of XCT in porous media characterization is the segmentation of the volume into solid and void space (which has come to mean the identification of discrete materials in an image; e.g., the binarization of an image into solid and void), XCT can be used to separate and identify more than two phases. Figure 8 (left) is a vertical cross section from a raw XCT image data set of a partially water-drained column packed with two types of glass beads (75% of the beads (by mass) contained some lead; the other 25% did not have any lead). The presence of lead increases the X-ray absorption of the voxels within those glass beads and makes them distinguishable from the glass beads without lead (note: the voxels with lead show up brighter due to the higher absorption values) and from the water. Figure 8 (right) is the identical cross section after the phases have been segmented; i.e., each voxel has been uniquely identified as glass bead with lead (white); glass bead without lead (light gray); water (dark gray); and air (black).

A significant amount of research has gone into quantifying the geometrical and topological properties from XCT data sets and relating those to flow and transport processes [Vogel and Roth, 2001; Herring et al., 2015] as well as extracting physically representative pore network structures, i.e., a direct mapping of the pore bodies, throats, and connectivity [Lindquist et al., 1996; Bhattad et al., 2011]. Two specific items to note here: (1) the topological properties of the pore space will have a strong influence on the flow and transport, thus impacting the spatial distribution of colloids as well as their trajectories toward and around collector surfaces and (2) the direct mapping of the pore structure (e.g., pore body and throat sizes as well as connectivity) provides quantitative data that can be used to determine whether or not certain colloid transport mechanisms (e.g., straining, discussed in section 5.3) might be relevant.

There has been some pore-scale visualization work on straining in both saturated and unsaturated conditions [Crist et al., 2004; Auset and Keller, 2006; Bradford et al., 2006b; Han et al., 2006; Gaillard et al., 2007]. However, straining studies have yet to fully incorporate the rich, quantitative data that can be extracted from these methods such as coupling real pore throat distributions with FEM/LBM modeling of the flow field (discussed in section 5.4). The type of quantitative data relevant to straining that can be obtained from XCT can be demonstrated using one of the data sets found in Molnar et al. (The impact of immobile zones on the transport and retention of nanoparticles in porous media, submitted to Water Resources...
Research, 2015). In a domain of $3.4 \times 3.4 \times 4.4 \text{ mm}^3$ imaged at a resolution of $9.87 \mu \text{m/voxel}$, the methods described in Thompson et al. [2006] and Bhattad et al. [2011] were used to determine that there were 616 sand grains and the pore network structure consisted of 3422 unique pore bodies and 17,160 pore throats. The average size (and range) of pore bodies and throats were 62.4 (10.5–203) \mu m and 40.6 (9.87–156) \mu m, respectively. The sizes listed here are the pore and throat inscribed radii, which are defined as the radius of the largest spheres (pores) and circles (throats) that can be drawn within the identified geometries. In addition, the average pore coordination number was 5.1 with a range of 1–33 pore connections per pore.

The richness of the XCT data sets and the ability to resolve the pore space, combined with increases in computing power allows for the direct flow modeling at the pore scale in XCT-obtained domains of real pore networks using Finite Element Modeling (FEM), LBM, etc. In particular, the unstructured nature of the FEM allows one to capture the discontinuous and complex nature of the solid/void interface, highly relevant for capturing the near-surface flow field (within nm if necessary), low-flow/recirculating regions and stagnation points. While these simulation approaches are now at the point where detailed hydrodynamics and bulk flow properties (e.g., permeability) are representative, there are still challenges to obtaining representative bulk transport properties (e.g., breakthrough curves) due to the complex, time scale-dependent nature of the transport processes (e.g., diffusion (Brownian motion), surface-particle interactions, and gravity) and the number of particles that need to be simulated (thousands if not tens or hundreds of thousands). Nevertheless, the increasing sophistication of pore-level hydrodynamic modeling is providing critical insights into colloid transport through and around complex structures [X. Q. Li et al., 2010; Z. Li et al., 2010; Li et al., 2012].

XCT has been successfully used to examine some pore-scale colloid transport details, in particular the spatial location of retained colloids. Gaillard et al. [2007] and Chen et al. [2008] demonstrated the ability to identify the distribution of micrometer-sized colloids (of high X-ray attenuation) in a sandstone and Iltis et al. [2011] mapped the distribution of biofilm in porous media using silver microspheres. Both of these studies utilized micron-sized colloids and, due to their material properties, had high X-ray attenuation coefficients enabling relatively straightforward identification within the systems. Molnar et al.
[2014] developed a method that used absorption-edge XCT to extract silver nanoparticle concentrations within individual pores in static and quasi-dynamic (i.e., transport) systems. This approach was then used to assess how relatively immobile (e.g., low velocity) pore space regions near grain-grain contacts could yield deviations between experimentally observed and CFT-predicted retention rates (Molnar et al., submitted manuscript, 2015). While the large amount of spatial data of \( \mu m \)-scale nAg concentrations provides powerful statistical measures of the nAg distributions within the pore space, limitations due to X-ray diffraction across the grain/void interface prevents concentration measurements close to grain surfaces (termed “shadow zone” in Molnar et al. [2014] which, in the system studied, was \( \sim 18 \mu m \)). The development and application of techniques for eliminating this shadow zone will allow for better estimates of near-surface concentrations and gradients and potentially of the impact of surface heterogeneities.

XCT has also been applied to extract granular properties such as particle size, shape, and angularity as well as geomechanically relevant structural properties such as grain-grain contacts (see review by Moreno-Atanasio et al. [2010]). As mentioned above, the ability to separate the system into void and solid phases allows for detailed hydrodynamic simulations that can be used to better understand the particle trajectories and proximity to the collector surface, locations of straining and wedging.

However, several factors currently limit the use of XCT for fundamental colloid transport studies including the spatial resolution, sample sizes below that necessary to obtain representative properties, and image contrast. The true spatial resolution of a CT instrument also depends on the sharpness of the imagery [Ketcham and Carlson, 2001; Wildenschild et al., 2002]. The spatial resolution of most XCT systems limits the ability to capture small structural/surface features (or heterogeneities) that can influence colloid-surface interactions and retention. While increases in the spatial resolution are possible through the use of higher-resolution XCT systems or other imaging techniques (e.g., SEM, thin sections), the quantitative details extracted may not be representative of the bulk sample and, if not collected directly with (or concurrently with) the XCT data, need to be mapped back to the image data [e.g., Golab et al., 2012]. Quantitative delineation of the mineral phases, which strongly influence the spatial distribution of favorable/unfavorable attachment sites depends on the X-ray attenuation properties of the minerals (as well as the spatial resolution). Fluid phase separation using doping agents is a common approach [e.g., Al-Raoush and Willson, 2005; Schnaar and Brusseau, 2005; Prodanović et al., 2007; Porter and Wildenschild, 2010; Porter et al., 2010]; however, much less work has been done on the spatial distribution of mineralogy. See the recent review paper by Kyle and Ketcham [2015] that discusses the approaches and limitations to mineral identification and quantification (while they focus primarily on ores, this is a very good overview). An interesting but intensive approach that uses the 2-D-3-D registration of SEM and XCT was developed by a group at the Australian National University [Golab et al., 2012; Sheppard et al., 2014]. In addition to providing additional submicron level structural details, this technique allows for the coupling of mineralogy (from the SEM) with the XCT data.

A major limitation of an XCT-based imaging approach is the inability to identify coatings/films on solid surfaces that are thinner than several microns and/or have X-ray absorption values that are similar to the grain. This may limit or prevent direct mapping of transport-relevant heterogeneities onto the grain surface, resulting in the need for indirect “mapping” of chemical heterogeneities.

While XCT is allowing for more detailed pore-level and grain-level characterization and the increases in computing power are pushing the simulation details to higher levels, there are still a number challenges/opportunities that can have a major impact on our understanding of colloid transport such as developing robust techniques to, directly or indirectly, extract and map submicron heterogeneities. In addition, improvements in the spatial resolution are needed to ensure that the pore structure/topology and mineralogy distribution is representatively captured in complex, environmental media. These features, when combined with improved techniques in quantitatively mapping colloid concentrations and in the modeling of complex pore-level hydrodynamics and colloid transport mechanisms (such as mineral specific surface-particle forces), will greatly improve our understanding of colloid transport in natural systems.

6. Conclusions

In this paper, we have presented the many advances in our understanding of colloid transport in porous media in the past 25 years. There is an increasingly sophisticated understanding of the applications of
employing short-range DLVO forces to describe colloid transport through porous media. The favorable and unfavorable deposition conditions that arise from DLVO interactions yield significantly different experimental transport and retention behaviors. However, the experimentally observed behaviors often differ from mean-field DLVO predictions; this is especially true for colloid retention in unfavorable conditions. A large amount of recent research has successfully elucidated a number of causes for these discrepancies such as grain and colloid-heterogeneity, secondary minimum interactions, and site blocking. A large portion of our understanding of these mechanisms has arisen from column-scale transport experiments coupled with continuum-scale models that, by fitting various rate parameters, can successfully describe experimental transport behavior.

There have also been significant advances in colloid filtration theory, a two-component method for predicting the colloid attachment rate parameter, $k$. The first component, a detailed mechanistic model of colloid transport and attachment onto a unit-cell collector by way of a force-torque balance, has become increasingly advanced with modern computational power. Particularly, successes have occurred with describing Brownian motion and moving toward a mechanistic approach to predicting colloid retention (and detachment) in unfavorable conditions, the prevailing condition in the environment. The second component of CFT, correlation equations that summarize the results of the detailed mechanistic models have also improved significantly and can, for the most part, accurately predict retention of larger colloids in favorable conditions.

Despite the large number of advances and successes, there remain many gaps in our current understanding, limiting our ability to predict colloid transport in a range of porous media systems. Many of the mechanisms invoked to describe discrepancies between DLVO-predicted colloid transport behavior and experimental observations arise from the use of fitted kinetic retention parameters in continuum-scale models. However, fitted parameters can only infer mechanisms, not prove those mechanisms are responsible for the fit. As such, many of the mechanisms attributed to colloid transport and retention behavior remain inferred. There are substantial opportunities available to develop mechanistic approaches to validating, and predicting, the influence of these mechanisms. Relatively new tools, such as XCT, will prove invaluable in this as they are capable of quantitatively extracting pore and grain details (i.e., pore body/throat sizes and distributions, void topologies and grain shapes, and mineralogy) as well as directly observing colloid distribution in the pore space and coupling with high-fidelity simulations of the flow fields.

Mechanistic CFT models and their respective correlation equations still struggle to predict nanoparticle transport and retention. It is still unclear if the unit-cell CFT approach is valid for highly diffusive materials. While fitting the $z$ parameter effectively overcomes the discrepancy between theoretical predictions and experimental observations for nanoparticle transport, true prediction of nanoparticle transport and retention will require accurate mechanistic models and correlation equations. Likewise, existing correlation equations are only derived for a limited range of environmental conditions that are not applicable to many nanoparticle scenarios such as a viscous solution of nano-zero-valent iron stabilized with polymer.

Finally, while mechanistic models are now able to quantitatively predict colloid attachment and qualitatively predict detachment under unfavorable conditions, these are limited to idealized systems involving carboxylate modified polystyrene latex microspheres on silica (which reflect the vast majority of existing colloid transport experiments). Even in these simple systems, the need to account for influences such as roughness in addition to charge heterogeneity is well noted and warrants further research. Furthermore, experiments looking at nonsilica surfaces and nonideal colloids, as well as size-distributed porous media present major opportunities to address environmental conditions in a more comprehensive manner.

Together, the above listed challenges represent a significant opportunity for advances that will undoubtedly lead to more informed decisions and design regarding colloids in the environment for the protection of human and ecological health.

**Appendix A**

Overview of the geometries and equations of select CFT models (Tables A1–A6).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model type</td>
<td>Lagrangian</td>
<td>Eulerian</td>
<td>Lagrangian</td>
</tr>
<tr>
<td>Model Geometry</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Particle boundary conditions**
- Entry boundary (Lagrangian only)§
- Free exit boundary (Lagrangian only)
- Constant concentration \(C = C_0\) (Eulerian only)
- Perfect sink \(C = 0\)

**Fluid flow equation**
- Stokes (creeping) flow: \(\nabla p = \mu \nabla^2 v\)
- Continuity: \(\nabla \cdot v = 0\)

**Governing colloid transport equation**
- \(F_0 + F_1 + F_2 + F_3 + F_4 = 0\)
- \(t_0 + t_1 + t_2 + t_3 + t_4 = 0\)

**\(\eta\) correlation**
- \(\eta \approx \gamma [4A_{1/2}N_{v}^{2}N_{r}^{1/2} + A_{2/3}N_{v}^{2}N_{r}^{1/3} + 0.00336A_{v}N_{v}^{2}N_{r}^{-0.8}]\)

**Kinetic retention rate equation**
- \(k = \frac{\beta \gamma^{1-n} \mu^{1/4}}{4\pi} \eta^{3/2}\)
- \(k = \frac{\beta \gamma^{1-n} \mu^{1/4}}{4\pi} \eta^{3/2}\)
- \(k = \frac{\beta \gamma^{1-n} \mu^{1/4}}{4\pi} \eta^{3/2}\)

---

§Rajagopalan and Tien [1976] only employed a Lagrangian model for colloid interception and sedimentation. An analytical solution (not shown here) was employed for diffusion.

Assumes steady-state conditions (i.e., \(\frac{\partial v}{\partial t} = 0\)) and negligible external forces. Subject to boundary conditions outlined in each model.

The force/torque equations for each model are listed below in their respective tables.

The \(\eta\) correlation equation originally presented in Rajagopalan and Tien [1976] is incorrect. Here we are presenting the Logan et al. [1995] corrected version.

Trajectory simulations employ random starting locations at upstream (labeled 'injection') collector boundary.

The illustrated models typically considered downwards flow (with gravity).
Overview of the Geometries and Equations Governing Selected Mechanistic Non-Happel Sphere Models

<table>
<thead>
<tr>
<th>Model type</th>
<th>Model Name</th>
<th>Hemispher-in-Cell [Ma et al., 2013]*</th>
<th>Long and Hilpert [2009]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Geometry</td>
<td>Eulerian</td>
<td>Flow out</td>
<td>Flow out</td>
</tr>
<tr>
<td></td>
<td>Lagrangian</td>
<td>Plane of symmetry</td>
<td>Contact point</td>
</tr>
<tr>
<td></td>
<td>Perfect sink</td>
<td>Perfect sink replaced with</td>
<td>Perfect sink</td>
</tr>
<tr>
<td></td>
<td>Perfect sink</td>
<td>Perfect sink replaced with</td>
<td>Perfect sink</td>
</tr>
<tr>
<td></td>
<td>Injection boundary</td>
<td>Free exit boundary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dC/dz=0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Particle boundary conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant concentration (C = C₀)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfect sink (C = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfect sink (C = 0)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Governing colloid transport equation:

\[
\frac{\partial C}{\partial t} + \nabla \cdot \mathbf{q} + \nabla D \nabla C = D_D \nabla^2 C + \left(1 - \frac{\eta}{\rho_f}\right) \frac{m g}{3 \pi \eta D_f} \frac{\partial C}{\partial z}
\]

\[
\eta \approx \left(\frac{1}{2}\right)^2 + 2\left(\frac{n-3}{3n}\right)^2 + 0.9 \left(\frac{n-1}{n+1}\right)
\]

\[
\eta \approx \gamma \left[0.01A_0 + 0.02N_0 + 0.02N_0 + 0.02N_0 + 0.07N_0 + 0.07N_0 + 0.09N_0 + 0.9\right]^{-1}
\]

\[
k = \frac{\gamma}{2(1-n)\eta} \frac{1}{g x z}
\]

Required correlation:

\[
k = \frac{3(1-n)}{2(1-n)\eta} m \left(\frac{3-n}{3} - \frac{2(3-n)}{3(3-n)} \cos^{-1} \frac{3(3-n)}{3(3-n)} + \frac{2}{\sqrt{2(3-n)(3-n)}} \right) \frac{1}{2(1-n)\eta}
\]

\[
k = \frac{1}{2(1-n)\eta} \frac{1}{g x z}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Name</th>
<th>Hemispher-in-Cell [Ma et al., 2013]*</th>
<th>Long and Hilpert [2009]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eulerian</td>
<td>Flow out</td>
<td>Flow out</td>
</tr>
<tr>
<td></td>
<td>Lagrangian</td>
<td>Plane of symmetry</td>
<td>Contact point</td>
</tr>
<tr>
<td></td>
<td>Perfect sink</td>
<td>Perfect sink replaced with</td>
<td>Perfect sink</td>
</tr>
<tr>
<td></td>
<td>Perfect sink</td>
<td>Perfect sink replaced with</td>
<td>Perfect sink</td>
</tr>
<tr>
<td></td>
<td>Injection boundary</td>
<td>Free exit boundary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dC/dz=0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are a number of iterations of the Hemisphere-in-Cell model [Ma et al., 2009, Ma and Johnson, 2010; Ma et al., 2011, Ma et al., 2013]. Here we are specifically presenting the correlation for Ma et al. [2013]. Many of the model details relevant to these tables remain the same between the iterations and, unless noted otherwise, it is not necessary to differentiate between the specific iterations.

Assumes steady-state conditions (i.e., \(\frac{\partial C}{\partial t} = 0\)) and negligible external forces. Subject to boundary conditions outlined in each model.

The force/torque equations for each model are listed below in their respective tables. Yao 1971 does not include forces or torques therefore none are reported here.


In regimes of either pure diffusion or sedimentation mass transfer use the asymptotes described in Ma et al. [2013].

Trajectory simulations employ random starting locations at upstream (labelled `injection`) collector boundary.

The illustrated models typically considered downwards flow (with gravity).
Table A3. Force/Torque Equations for Rajagopalan and Tien [1976]

<table>
<thead>
<tr>
<th>Force</th>
<th>Torque</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inertia</td>
<td>$\mathbf{F}_I = m \left( \ddot{\mathbf{r}} + \mathbf{u} \cdot \nabla \mathbf{u} \right)$</td>
</tr>
<tr>
<td>Gravity</td>
<td>$\mathbf{F}_G = -\frac{4}{3} \pi \alpha \rho^2 (\rho_p - \rho)</td>
</tr>
<tr>
<td>London</td>
<td>$\mathbf{F}_L = -\frac{2}{3} \pi \rho \alpha \rho^2 \mathbf{e}_r$</td>
</tr>
<tr>
<td>Electrostatic layer</td>
<td>$\mathbf{F}_E = \frac{1}{2} \rho \alpha (2T - q^2) \left( \mathbf{e}_r \cdot \nabla \right) \mathbf{e}_r - \frac{1}{2} \rho \alpha (2T - q^2) \mathbf{e}_r \nabla \cdot \mathbf{e}_r$</td>
</tr>
<tr>
<td>Particle movement</td>
<td>Hydrodynamic drag (t = translation, r = rotation, m = movement of fluid around particle)</td>
</tr>
<tr>
<td>Hydrodynamic drag</td>
<td>$\mathbf{F}_D = -9 \pi \mu a^2 \mathbf{u} \cdot \mathbf{v}$</td>
</tr>
<tr>
<td>Particle velocity from streamline</td>
<td>$\mathbf{u}_t = \frac{1}{8} \left[ \mathbf{B}^t \cdot (1 + \delta^t) \bigg( \delta^t \bigg)^2 \right] \mathbf{q}$</td>
</tr>
<tr>
<td>Particle velocity with London van der Waals forces</td>
<td>$\mathbf{u}_l = f_L (\delta^l) \mathbf{u}_v$</td>
</tr>
<tr>
<td>Particle velocity with sedimentation</td>
<td>$\mathbf{u}_s = -\frac{1}{2} \left[ \frac{N_D \sin(\theta)}{\left( e^{r/2} \right)^2} \right] \mathbf{U}$</td>
</tr>
<tr>
<td>Particle velocity with hydrodynamic retardation</td>
<td>$\mathbf{u}_h = \frac{1}{8} \left[ \mathbf{B}^h \cdot (1 + \delta^h) \bigg( \delta^h \bigg)^2 \right] \mathbf{q}$</td>
</tr>
<tr>
<td>Total expression for particle velocity</td>
<td>$\mathbf{u}_p = \frac{1}{4} \left[ \mathbf{B}^p \cdot (1 + \delta^p) \bigg( \delta^p \bigg)^2 \right] \mathbf{q}$</td>
</tr>
<tr>
<td>Particle displacement due to Brownian force vector (F_B) (Cartesian)</td>
<td>$\mathbf{R}_p = \frac{\mathbf{n}_p \sqrt{2D_B \Delta t}}{\mathbf{n}_p \sqrt{2D_B \Delta t}}$</td>
</tr>
</tbody>
</table>

Table A4. Force/Torque Equations for Tufenkji and Elimelech [2004b]a

<table>
<thead>
<tr>
<th>Force</th>
<th>Torque</th>
</tr>
</thead>
<tbody>
<tr>
<td>External force in governing equation</td>
<td>$\mathbf{F} = \mathbf{F}_n + \mathbf{F}_D$</td>
</tr>
<tr>
<td>Colloidal forces</td>
<td>$\mathbf{F}_n = -\nabla \phi$</td>
</tr>
<tr>
<td>Gravity</td>
<td>$\mathbf{F}_G = -\frac{4}{3} \pi \alpha \rho^2 (\rho_p - \rho)</td>
</tr>
<tr>
<td>Hydrodynamic dragb,c</td>
<td>$\mathbf{F}_D = -8 \pi \mu a^2 \mathbf{u} \cdot \mathbf{v}$</td>
</tr>
<tr>
<td>Force/Torque Equations for Nelson and Ginn [2011]mb,c</td>
<td>$\mathbf{u}_I = \frac{1}{4} \left[ \mathbf{B}^r \cdot (1 + \delta^r) \bigg( \delta^r \bigg)^2 \right] \mathbf{q}$</td>
</tr>
</tbody>
</table>

aThe model equations are presented in Elimelech [1994], not Tufenkji and Elimelech [2004b].
bElimelech [1994] does not present hydrodynamic force equations. We have reproduced the particle velocity equations instead.
cD$_1$ and D$_2$ are referred to more generally as D$_{tar}$ in the governing colloid transport equation in Table A1.

Table A5. Force/Torque Equations for Nelson and Ginn [2011]mb,c

| Equation for a and r Component Particle Velocities |
|---|---|
| Particle velocity from streamline | $\mathbf{u}_r = \frac{1}{4} \left[ \mathbf{B}^t \cdot (1 + \delta^t) \bigg( \delta^t \bigg)^2 \right] \mathbf{q}$ |
| Particle velocity with London van der Waals forces | $\mathbf{u}_l = f_L (\delta^l) \mathbf{u}_v$ |
| Particle velocity with sedimentation | $\mathbf{u}_s = -\frac{1}{2} \left[ \frac{N_D \sin(\theta)}{\left( e^{r/2} \right)^2} \right] \mathbf{U}$ |
| Particle velocity with hydrodynamic retardation | $\mathbf{u}_h = \frac{1}{8} \left[ \mathbf{B}^h \cdot (1 + \delta^h) \bigg( \delta^h \bigg)^2 \right] \mathbf{q}$ |
| Total expression for particle velocity | $\mathbf{u}_p = \frac{1}{8} \left[ \mathbf{B}^p \cdot (1 + \delta^p) \bigg( \delta^p \bigg)^2 \right] \mathbf{q}$ |
| Particle displacement due to Brownian force vector (F_B) (Cartesian)d | $\mathbf{R}_p = \frac{\mathbf{n}_p \sqrt{2D_B \Delta t}}{\mathbf{n}_p \sqrt{2D_B \Delta t}}$ |

dMany of these equations are not presented in Nelson and Ginn [2011] but appear instead in Nelson and Ginn [2005].
eThe Nelson and Ginn [2005, 2011] papers do not present their force/torque equations. Instead, they present the contributions to particle velocity from the individual components. We have reproduced these velocity expressions here along with the final, complete equations for particle velocity.
f$D_{tar}$ in this table is a function of both $\delta^r$ and the classic Stokes-Einstein equation, please see Nelson and Ginn [2011] for full details.
## Table A6. List of Force Equations for the Hemisphere-In-Cell Model [Mo et al., 2009; Ma and Johnson, 2010; Ma et al., 2011, 2013]a,b

<table>
<thead>
<tr>
<th>Force Type</th>
<th>Force Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrodynamic drag</td>
<td></td>
</tr>
<tr>
<td>Gravity</td>
<td></td>
</tr>
<tr>
<td>Shear lift</td>
<td></td>
</tr>
<tr>
<td>Electrostatic double layer</td>
<td></td>
</tr>
<tr>
<td>van der Waals</td>
<td></td>
</tr>
<tr>
<td>Brownian force</td>
<td></td>
</tr>
</tbody>
</table>

F_r = \frac{6\pi\eta a_r v_r}{l} + \frac{6\pi\eta a_r v_t}{l}
F_g = \frac{2\pi\eta u_r (p_r - p_t)}{g}
F_t = \frac{6\pi\eta a_t v_t}{l}
F_{EDL} = 4\pi\varepsilon_0\kappa k_F \varepsilon_0 \left( \frac{1}{\exp(2a/k_B T) - 1} - \frac{1}{1 - \exp(-2a/k_B T)} \right)
F_{vDW} = \frac{\phi_{mol} (d_{min} / 12.33) \kappa}{
\begin{align*}
\exp \left( \frac{d_{min} (1 - \exp(-2a/k_B T))}{2\varepsilon_0} \right) \\
\exp \left( \frac{1 - \exp(-2a/k_B T)}{2\varepsilon_0} \right)
\end{align*}
}
F_{B} = \frac{2\pi\kappa}{a}

a The hemisphere-in-cell model employs torques that are equivalent to the Rajagopalan and Tien [1976] model. The result of this is that, for colloids in contact with the collector surface, all torque-induced rolling is assumed yield translation along the collector surface (i.e., no slipping or friction) in unfavorable conditions. The only difference is an additional surface friction torque that appears in the Ma et al. [2011] iteration of the hemisphere-in-cell model.

b Many of the terms in the hemisphere-in-cell model and the Rajagopalan and Tien [1976] model are similar, however, the hemisphere-in-cell flow field has to be solved numerically, this results in references to velocity (v) in the force equations as opposed to the A, B, and D coefficients employed by Rajagopalan and Tien [1976] from their analytical solution of the flow field.

### Notation for Appendix

A brief note on notation: The authors of these mechanistic model studies often use different symbols to represent the same parameter. Wherever possible, we have consolidated the symbols representing each parameter into a single consistent symbol. For instance, the separation distance between colloidal and collector surface has been variously referred to as δ, h, and H. In this notation, we use the symbol δ from the Rajagopalan and Tien [1976] paper.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>a_r</td>
<td>radius of colloid</td>
<td>L</td>
</tr>
<tr>
<td>a_t</td>
<td>radius of collector</td>
<td>L</td>
</tr>
<tr>
<td>a_h</td>
<td>porosity dependent parameter</td>
<td>N.A.</td>
</tr>
<tr>
<td>A, B, D</td>
<td>coefficients to correct fluid velocity. Refer to the individual papers for their exact values</td>
<td>N.A.</td>
</tr>
<tr>
<td>A^+, B^+, D^+</td>
<td>dimensionless versions of the A,B,D coefficients</td>
<td>N.A.</td>
</tr>
<tr>
<td>C</td>
<td>concentration of colloids</td>
<td>m/L^3</td>
</tr>
<tr>
<td>C_{mol}</td>
<td>concentration of colloids entering the domain</td>
<td>m/L^3</td>
</tr>
<tr>
<td>D_{diff}</td>
<td>Brownian diffusion coefficient for a sphere in an infinite fluid (from Stokes-Einstein equation: $D_{diff} = \frac{k_B T}{6\pi\eta a}$)</td>
<td>L^2/t</td>
</tr>
<tr>
<td>D_{diff}</td>
<td>diffusion tensor (from $D_{r}$, $D_{a}$)</td>
<td>L^2/t</td>
</tr>
<tr>
<td>D_r, D_a</td>
<td>diffusion coefficient in the radial and angular directions</td>
<td>L^2/t</td>
</tr>
<tr>
<td>d_t</td>
<td>diameter of collector</td>
<td>L</td>
</tr>
<tr>
<td>d_p</td>
<td>diameter of colloid</td>
<td>L</td>
</tr>
<tr>
<td>e_r, e_t, e_a</td>
<td>unit vectors in the radial and angular directions</td>
<td>N.A.</td>
</tr>
<tr>
<td>f_r, f_t, f_a, f_r, f_t, f_a</td>
<td>drag correction factors in RT 1976 and NG 2011. Refer to individual papers for exact values.</td>
<td>m/L^2</td>
</tr>
<tr>
<td>f_r, f_t, f_a, f_r, f_t, f_a</td>
<td>corrections for hydrodynamic interactions in TE 2004 and MFPJ. Refer to individual papers for exact values.</td>
<td>N.A.</td>
</tr>
<tr>
<td>g</td>
<td>gravitational acceleration, 9.81 m/s^2</td>
<td>L/T^2</td>
</tr>
<tr>
<td>g''</td>
<td>torque correction factors in RT 1976</td>
<td>m/L^2</td>
</tr>
<tr>
<td>H</td>
<td>Hamaker coefficient</td>
<td>N.A.</td>
</tr>
<tr>
<td>k_B</td>
<td>Boltzmann constant (~1.38 × 10^{-23} m^2 kg s^{-2} K^{-1})</td>
<td>L^2 m/(K T)</td>
</tr>
</tbody>
</table>

(Continued)
Data used in this paper can be requested from the corresponding author.

References


Bai, C., and Y. Li (2012), Modeling the transport and retention of nc60 nanoparticles in the subsurface under different release scenarios, J. Contam. Hydrol., 136–137(0), 43–55.


Bradford, S. A., S. Torkzaban, and A. Wiegemann (2011a), Pore-scale simulations to determine the applied hydrodynamic torque and colloid immobilization, Vadose Zone J., 10(1), 252–261.


Johnson, W. P., and X. Li (2005), Comment on breakdown of colloid filtration theory: Role of the secondary energy minimum and surface charge heterogeneities, Langmuir, 21(23), 10,895–10,895.


Wang, Y., S. A. Bradford, and J. Simůnek (2014a), Physicochemical factors influencing the preferential transport of Escherichia coli in soils, Vadose Zone J., 13(1).


