INTRODUCTION
CHAPTER 1

STATE OF UNSATURATED SOIL

1.1 UNSATURATED SOIL PHENOMENA

1.1.1 Definition of Unsaturated Soil Mechanics

To provide and agree upon a precise definition of unsaturated soil mechanics is an academic challenge in itself. Perhaps one can draw some areas and boundaries by revisiting the classical definition of soil mechanics posed by Karl Terzaghi some 60 years ago. In his seminal book of 1943, *Theoretical Soil Mechanics*, Terzaghi defined soil mechanics as “the application of the laws of mechanics and hydraulics to engineering problems dealing with sediments and other unconsolidated accumulations of solid particles produced by the mechanical and chemical disintegration of rocks, regardless of whether or not they contain an admixture of organic constituents.” In drawing this silhouette of soil mechanics, Terzaghi refers to three basic requirements: (1) earthen materials, (2) the principles of mechanics and hydraulics, and (3) engineering problems.

The emerging appreciation of unsaturated soil in geotechnical engineering practice and education requires refinement of Terzaghi’s basic definition. The earthen materials dealt with in problems of unsaturated soil mechanics are arguably the same as in Terzaghi’s soil mechanics, referred to as “soils,” but under a very specific “unsaturated” condition. The qualifier “unsaturated” bears the same meaning as its alternative “partially saturated” and simply indicates that the degree of pore water saturation is any value less than unity or, more specifically, that a third phase of matter is introduced into the two-phase, saturated soil system. In the modern educational and professional geotechnical engineering environment, where the emphasis has historically been
limited to the arena of saturated cohesive materials and completely dry or completely saturated cohesionless materials, the “unsaturated” qualifier is indeed significant.

In dealing with unsaturated soil, one requires not only the principles of mechanics and hydraulics but also of fundamental interfacial physics. Physics in this regard refers primarily to the thermodynamic principles describing equilibrium among gas, solid, and liquid phases, the transition of matter from one phase to another, and the adsorption or desorption of one phase of matter onto or from an adjacent phase of different matter. The forces and energies associated with these multiphase interactions by their very nature separate unsaturated soil behavior from saturated soil behavior. In many practical problems, where the hydrologic and stress-strain behavior of natural or engineered systems comprised of soil is strongly influenced by the presence, absence, or changes in these interfacial interactions, the traditional saturated soil mechanics framework often fails to satisfactorily describe or predict the behavior of the system.

Terzaghi’s reference to engineering problems was developed in the wake of a period of great uncertainty in the basic understanding of soil behavior. His formalization of soil mechanics provided a rational basis for tackling many of the pressing engineering problems of the day, most notably bearing capacity, consolidation and settlement, slope stability, lateral earth pressure, and seepage-related problems. In addition to these traditional geotechnical engineering problems, the practical problems of interest today might also include geo-environmental, seismic, land reclamation, and other challenges that have come to light over the past 30 years or so. These emerging problems have created important subdisciplines within the more general field of geotechnical engineering, which often benefit from a thorough understanding of the physical and thermodynamic principles governing unsaturated soil behavior.

Extending Terzaghi’s classical definition, therefore, unsaturated soil mechanics might be defined as “the application of the laws of mechanics, hydraulics, and interfacial physics to engineering problems dealing with partially saturated soils.” The spirit of this definition and the laws, concepts, and problems that characterize it will be addressed throughout this book. Of course, as new technical discoveries are made, as new and unforeseen types of problems emerge, and as the once distinct boundaries between the traditional engineering and science disciplines continue to blur, there is no doubt that this definition may one day also require refinement.

1.1.2 Interdisciplinary Nature of Unsaturated Soil Mechanics

The history of unsaturated soil mechanics is embedded in the history of hydrology, soil mechanics, and soil physics. Engineering problems involving unsaturated soil span numerous subdisciplines and practices within the general field of civil engineering. Hydrologists, for example, have long recognized that modeling of regional or local surface water and groundwater systems and
cycles must consider infiltration, evaporation, and transpiration processes occurring in the near-surface unsaturated soil zone. Quantitative evaluation of moisture flux at the atmosphere-subsurface boundary requires not only knowledge of the relevant soil and pore water properties but also the predominant environmental conditions at the soil-atmosphere interface. Unsaturated soil often comprises cover or barrier materials for landfills and hazardous waste storage facilities of interest to the geo-environmental community. Contaminant transport and leaching processes are often strictly unsaturated fluid transport phenomena, occurring in many cases as multiphase transport problems. As national and international policy with regard to the health of the natural environment is becoming increasingly more regulated, recognition of these types of geo-environmental issues and development of solutions from an unsaturated soil mechanics framework is becoming more and more common.

Many of the more traditional geotechnical engineering problems also fall wholly or partly into the category of unsaturated soil mechanics problems. Compaction, for example, a classical application involving unsaturated soil, has been routine practice for improving the mechanical and hydraulic properties of soil since far before the formation of civil engineering as a formal discipline in the mid-nineteenth century. Compacted soil comprising the many earthworks constructed all over the world is most appropriately considered from an unsaturated soils framework. It has long been recognized that expansive soils pose a severe threat to civil engineering infrastructure such as roads, housing, and transportation facilities nationally and internationally. Expansive soil formations in the United States alone are responsible for billions of dollars in damage costs each year, an amount exceeding that of all other natural hazards combined, including earthquakes, floods, fires, and tornados (Jones and Holtz, 1973). Expansive soils have been the subject, if not the driving force, of unsaturated soil research since the early stages in the formulation of unsaturated soil mechanics principles. Collapsing soils also pose a significant threat in many areas of the world. These problematic soils, which are typified by the massive loess deposits of the central United States, are marked by a structurally sensitive fabric weakly cemented by a small clay fraction. Upon wetting, usually occurring either as a sudden precipitation event or gradual process associated with urbanization and development, the cementation bonds are weakened and the initially loose fabric collapses and densifies, often resulting in dramatic and damaging settlement. Any fundamental approach to mitigating collapsing soil hazards requires insight into the role of pore water interactions on the microscopic scale of the solid-liquid-air interface, a hallmark of unsaturated soil mechanics.

Reconsideration of the traditional saturated soil mechanics approach in light of these types of problems began to emerge during the late 1970s and continues today. In the authors’ opinion, the soil mechanics community is far from achieving a comprehensive and satisfactory framework for approaching these and other unsaturated soil mechanics problems, but new insights and technical advances are continuously being made.
1.1.3 Classification of Unsaturated Soil Phenomena

While the development of theory and techniques in unsaturated soil mechanics requires principles drawn from mechanics, hydraulics, and interfacial physics, it is convenient to classify the various geotechnical engineering problems involving unsaturated soil into three general phenomena, specifically, flow phenomena, stress phenomena, and deformation phenomena. It should be noted, however, that generalization in this manner is mainly for understanding purposes and for convenience of presenting the principles, not to set up boundaries among different geotechnical problems. The majority of practical engineering problems generally involve all three phenomena concurrently and in coupled fashion. An effective theory describing the deformation behavior of expansive soil, for example, could well require application of the principles of stress, strain, and flow in highly deformable porous media.

**Flow Phenomena**  
Flow phenomena require mainly the application of hydraulics and interfacial physics principles. One well-known example falling into this class is capillary flow. The search for the driving force for capillary flow had once been the subject of research for many years. As early as the 1900s, Buckingham (1907) systematically studied capillary rise and drainage in laboratory soil columns such as that illustrated in Fig. 1.1. Early data provided evidence of the important effects of soil type, grain size, and pore size properties on capillary rise and pore water retention in unsaturated soil. As part of this early work, the terms capillary potential and capillary conductivity were introduced as the driving force and controlling material variable, respectively, for capillary fluid flow. Later, others recalled the more

![Figure 1.1](image-url)  
**Figure 1.1** Capillary rise and equilibrium moisture content distribution in vertically oriented soil column (data from Buckingham, 1907). The curves shown, which describe the relationship between suction head and moisture content, are commonly called soil-water characteristic curves.
1.1 UNSATURATED SOIL PHENOMENA

general term chemical potential, to include components of the pore water potential resulting from dissolved chemical species, gravity, capillarity, and short-range physicochemical effects occurring at the solid-liquid phase interface (e.g., Gardner and Widstoe, 1921; Richards, 1928; Russell, 1942; Edlefsen and Anderson, 1943). The chemical potential, or free energy, concept for soil pore water has been generalized by Sposito (1981) and others to include the mass of all three phases (gas, solid, and liquid), together with temperature and pressure as independent state variables. As a result, many seepage-related problems in unsaturated soil mechanics may be effectively treated through the application of thermodynamic potential theory with little or no involvement of solid mechanics.

**Stress Phenomena** Problems requiring consideration of both mechanical and chemical equilibrium are classified as stress phenomena. These include traditional geotechnical engineering problems such as lateral earth pressure, bearing capacity, and slope stability analysis. For each of these problems, the strength of the soil at its limit state is the primary concern. Analysis of the stress distribution within the soil mass and the corresponding bulk strength becomes critically important. Limit analysis developed extensively since the 1930s for saturated soil applications formed the basis for solving most of these types of problems. Developing elastoplastic theories for soil became the focus of much of the geomechanics research activity during the 1970s and 1980s. Powerful numerical methods to solve the governing partial differential equations for stress equilibrium under static or dynamic conditions have been developed and applied to many difficult foundation problems in the past 20 years or so.

It has become clear in recent years that improved solutions of many stress-related geotechnical engineering problems require not only sustained activities along the continuum-based solid mechanics approach but also new theories along a microscopic discontinuous approach for describing effective stress under multiphase conditions. Terzaghi’s effective stress, which is the cornerstone of soil mechanics under saturated conditions, becomes either ineffective or inappropriate for fully describing the stress distributions or failure conditions in unsaturated soil. It has been recognized that theories for describing the states of stress and failure in unsaturated soil require consideration of the thermodynamic properties of the pore water in terms of soil suction, material variables such as grain size and grain size distribution, state variables such as the degree of saturation, and the consequent interparticle forces such as suction-induced effective stress or suction stress.

**Deformation Phenomena** Physical processes characterized by large deformations or strains are classified as deformation phenomena. In unsaturated soils, these deformations are very often caused or governed by changes in the moisture condition of the soil. Important deformation phenomena include compaction, multiphase consolidation and compressibility, and collapsing soil
behavior. Arguably, the most notorious unsaturated soil deformation phenomenon is that of swelling or shrinking (i.e., expansive) soil. Figure 1.2, for example, illustrates several important mechanisms commonly occurring in near-surface deposits of expansive soil. Many of these mechanisms, such as heave or subsidence of the ground surface, swelling pressure generation under pavements or foundations, and tension cracking, fall into the general category of deformation phenomena. Others shown in the figure, such as infiltration, evaporation, and the corresponding seasonal fluctuation in the subsurface moisture profile, fall into the general category of unsaturated flow phenomena. The inherent coupling between volume change, pressure generation, and moisture transport in expansive soil demonstrates the importance of the combined roles of deformation, stress, and fluid flow phenomena in this and numerous other types of unsaturated soil mechanics problems.

1.2 SCOPE AND ORGANIZATION OF BOOK

1.2.1 Chapter Structure

*Unsaturated Soil Mechanics* is organized into four divisible but interrelated parts. The intent of this separation is to provide the reader with a format

![Figure 1.2](image.png)

**Figure 1.2** Deformation and fluid flow phenomena in a near-surface deposit of unsaturated expansive soil.
where particular concepts, theories, phenomena, or practical applications of interest may be directly accessed in a focused and concise manner. Each chapter concludes with a series of qualitative and/or quantitative problems. The 12 chapters of the book are organized as follows:

Introduction: Chapter 1
Part I: Fundamental Principles, Chapters 2 to 4
Part II: Stress Phenomena, Chapters 5 to 7
Part III: Flow Phenomena, Chapters 8 and 9
Part IV: Material Variable Measurement and Modeling, Chapters 10 to 12

Chapter 1, State of Unsaturated Soil, provides a general introduction to unsaturated soil mechanics. The relevant state variables, material variables, and constitutive laws for describing flow, stress, and deformation phenomena in three-phase unsaturated soil systems are introduced. The important differences between saturated and unsaturated soil systems in terms of subsurface moisture, pore pressure, and stress profiles are described. Common types of practical engineering applications that warrant an unsaturated soil mechanics approach are introduced. The important role of unsaturated soil in terms of naturally occurring phenomena such as the hydrologic cycle, global climatic changes, and soil formation is described. Finally, the important concepts of pore water potential, soil suction, and the constitutive relationship between soil suction and water content, the soil-water characteristic curve, are introduced.

Part I, Fundamental Principles, provides the necessary background for the remainder of the book. Chapter 2, Material Variables, introduces the relevant physical properties of air, water, and water vapor and evaluates their dependency on the state variables that are used to describe multiphase unsaturated soil systems. Relative humidity and surface tension are introduced and described with respect to their roles in the behavior and analysis of unsaturated soil systems. Cavitation phenomena are systematically described. Chapter 3, Interfacial Equilibrium, describes several fundamental concepts within the general realm of interfacial physics. Mechanical and chemical equilibrium for air-water-solid interfaces are described with the introduction of Kelvin’s law. Associated interfacial phenomena including vapor pressure lowering, capillary condensation, and the solubility of air in water are described and illustrated through a series of thought experiments and quantitative examples. Finally, the soil-water characteristic curve is introduced from a micromechanical perspective by considering mechanical and chemical equilibrium for idealized systems of unsaturated soil grains. Chapter 4, Capillarity, introduces the Young-Laplace equation for describing equilibrium at an air-water interface, the height and rate of capillary rise, and the estimation of pore size distribution using capillary theory. The concept of suction stress is formulated from a micromechanical perspective to serve as a link between the preceding in-
terfacial equilibrium concepts and the associated interparticle stresses in unsaturated soil systems.

Part II, Stress Phenomena, contains three chapters. Chapter 5, State of Stress, complements the interfacial equilibrium concepts introduced in Chapter 4 by providing a derivation of effective stress among idealized unsaturated soil particles. Mechanisms for hysteresis in the soil-water characteristic curve and suction stress characteristic curve are introduced and evaluated. Tensor notation and graphical representation for the independent stress state variable approach and the effective stress approach to describing the state of stress in unsaturated soil are introduced and explained using example problems. The concept of axis translation for controlling the stress state variables relevant to unsaturated soil is presented. Chapter 6, Shear Strength, describes several alternative theories for interpreting and analyzing shear strength in unsaturated soil. The extended Mohr-Coulomb failure criterion, the shear strength parameters describing it, and its advantages and limitations are introduced. Bishop’s effective stress parameter $\chi$ and its role in effective stress for unsaturated soil is described. Finally, a unified framework for interpreting and measuring shear strength characteristics in unsaturated soil is suggested.

Chapter 7, Suction and Earth Pressure Profiles, includes theoretical development of subsurface suction stress and water content profiles under steady-state infiltration, hydrostatic, and evaporation conditions. Corresponding lateral earth pressure profiles are derived for conditions at rest and under active and passive conditions. These new theories serve as an instructional vehicle to provide insight into the fundamental differences between the states of stress in saturated and unsaturated soil.

Part III, Flow Phenomena, is divided into Chapter 8, Steady Flows, and Chapter 9, Transient Flows. Together, these chapters provide an introduction to the governing principles and solutions for both liquid and gas flow in unsaturated soil systems. Governing flow equations are solved analytically and numerically and illustrated graphically through simple one-dimensional example problems. Capillary barriers for geo-environmental applications are described along with vapor phase transport and diffusion processes. Practical examples involving transient pore airflow by barometric pumping are provided to highlight the important impact of variations in the governing state variables (e.g., temperature and pressure) on pore fluid transport processes in unsaturated soil.

Part IV, Material Variable Measurement and Modeling, provides the practicing and research community with a reference source pertaining to suction and hydraulic conductivity measurement and modeling alternatives. Chapter 10, Suction Measurement, describes the general principles, technical aspects, and performance of many of the more common suction and soil-water characteristic curve measurement techniques. Chapter 11, Hydraulic Conductivity Measurement, describes several common steady-state and transient techniques for measuring the unsaturated hydraulic conductivity function. Finally, chapter 12, Suction and Hydraulic Conductivity Models, describes numerous meth-
1.2 SCOPE AND ORGANIZATION OF BOOK

1.2.2 Geomechanics and Geo-environmental Tracks

Unsaturated Soil Mechanics contains sufficient material for a one-semester course tailored to follow either a geomechanics or geo-environmental track. Figure 1.3 illustrates two suggested paths through the chapters of the book corresponding to a geomechanics track, which emphasizes the stress and

![Diagram](Image)

Figure 1.3 Recommended chapter sequences for geomechanics and geo-environmental learning tracks.
strength concepts described in Part II, and a geo-environmental track, which emphasizes the hydrology concepts described in Part III. Content along each track has been included such that each may generally stand on its own.

1.3 UNSATURATED SOIL IN NATURE AND PRACTICE

1.3.1 Unsaturated Soil in Hydrologic Cycle

Figure 1.4 shows a schematic diagram of the unsaturated soil environment and its role in the natural hydrologic cycle. The steady-state position of the water table is controlled by the general topography of the system, the soil properties, and the balance achieved among the natural mechanisms that act to either add or remove water to or from the subsurface. The scale of the corresponding hydrologic cycle could be either local or regional, extending from as small as a local engineering work site to as large as the continental or global scale. Globally, the amount of water in the unsaturated zone located between the water table and the ground surface represents only a small portion of the total water involved in the hydrologic cycle (less than 0.01%). However, because the unsaturated zone forms the necessary transition between the atmosphere and larger groundwater aquifers at depth, the movement of water within this small portion of the cycle is indeed significant.

1.3.2 Global Extent of Climatic Factors

The size and extent of the near-surface unsaturated soil zone are highly sensitive to perturbations in local or regional climate. Precipitation, evaporation,
and evapotranspiration are all important natural environmental mechanisms that act to influence the depth and extent of the unsaturated zone.

Figure 1.5 is a map delineating the estimated average annual precipitation on the global scale. The lightest zones on the figure indicate regions receiving less than 25 cm of annual precipitation. The darkest zones indicate regions receiving more than 200 cm annually. Regions receiving average annual precipitation less than 200 cm will generally have a depth to the water table greater than 200 cm. Figure 1.6 shows an estimate of net precipitation (precipitation minus evaporation) as a function of latitude, indicating that regions where evaporation exceeds precipitation are generally concentrated within 40° north and south of the equator.

Figure 1.7 is a global-scale map delineated in terms of the global humidity index (UNESCO, 1984). The global humidity index is based on the ratio of average annual precipitation and potential evaporation (P/PET), such that hyperarid zones fall into a category where P/PET < 0.05, arid zones indicate 0.05 < P/PET < 0.2, semiarid zones indicate 0.2 < P/PET < 0.5, dry subhumid zones indicate 0.5 < P/PET < 0.65, and humid zones indicate 0.65 < P/PET. Figure 1.8 shows a similar humidity index map for North America.

In the semiarid to arid regions of the western United States, the depth of the unsaturated zone may extend to as much as several hundred meters.

1.3.3 Unsaturated Zone and Soil Formation

The unsaturated soil zone plays a critical role in biological, physical, and chemical weathering processes that have occurred throughout the history of Earth. The history of soil formation is the history of the unsaturated zone. As a result of physical and chemical weathering processes largely controlled by environmental factors at the ground surface, parent rock weathers to a residual soil profile of distinct horizons and chemical composition. The evolutionary process from unweathered rock to mature soil is illustrated for a typical profile in Fig. 1.9. A simplified description of nomenclature for each soil horizon follows, based on a complete systematic description by Birkeland (1999).

**O Horizon** Surface accumulations of mainly organic material are subdivided on the degree of decomposition as measured by the fiber content.

**Oi Horizon** Least decomposed organic materials; the fiber content is greater than 40% by volume.

**Oe Horizon** Intermediate degree of decomposition; the fiber content is between 17 and 40% by volume.

**Oa Horizon** Most decomposed organic materials, the fiber content is less than 17% by volume.

**A Horizon** Accumulations of humidified organic materials mixed with dominant mineral fraction occur at the surface or below an O horizon.
Figure 1.5  Global average annual precipitation. (from Penman, 1970; reproduced with permission).
Figure 1.6  Net global precipitation and evaporation as a function of latitude. Positive values indicate latitudes where precipitation exceeds evaporation. Negative values (shaded regions) indicate latitudes where evaporation exceeds precipitation (after Peixoto and Kettani, 1973; illustration by Eric O. Mose).

**E Horizon** Accumulations usually underlie an O or A horizon and can be used for eluvial horizons within or between parts of B horizon, and are characterized by less organic materials and/or fewer compounds of iron and aluminum (sesquioxides) and/or less clay than the underlying horizon.

**B Horizon** Soil underlies an O, A, or E horizon, shows little or no evidence of original sediment or rock structure, and is recognized into several subhorizons based on the kinds of materials illuviated into them or residual concentrations of materials.

**Bh Horizon** Illuvial accumulation of amorphous organic materials-sesquioxide complexes that either coat grains or form sufficient coatings and pore fillings to cement the horizon.
Figure 1.7 Global humidity index map (adapted from GRID/UNEP, Office of Arid Lands Studies, University of Arizona).

Figure 1.8 Humidity index map of North America (adapted from GRID/UNEP, Office of Arid Lands Studies, University of Arizona).
Figure 1.9  Soil profile: (a) typical evolution of soil profile and (b) soil formed on marine-terrace deposits near San Diego, California [(b) from Birkeland, 1999; adapted by permission of Oxford University Press, Inc.].

**Bhs Horizon**  Illuvial accumulation of amorphous organic materials-sesquioxide complexes, and sesquioxide component is significant; both color value and chroma are three or less.

**Bk Horizon**  Illuvial accumulation of alkaline earth carbonates, mainly calcium carbonate; properties do not meet those for the K horizon.

**Bl Horizon**  Illuvial concentrations primarily of silt.

**Bo Horizon**  Residual concentration of sesquioxides, the more soluble materials having been removed.

**Bq Horizon**  Accumulation of secondary silica.

**Bs Horizon**  Illuvial accumulation of amorphous organic materials-sesquioxide complexes if both color value and chroma are greater than three.

**Bt Horizon**  Accumulation of silicate clay that has either formed in situ or is illuvial.

**Bw Horizon**  Development of color (redder hue or higher chroma relative to C horizon) or structure, or both, with little or no apparent illuvial accumulation of material.

**By Horizon**  Accumulation of secondary gypsum.

**Bz Horizon**  Accumulation of salts more soluble than gypsum.
**K Horizon**  Soil is so impregnated with carbonate that its morphology is determined by the carbonate.

**C Horizon**  Soil lacks properties of A and B horizons, excludes R horizon but includes materials in various stages of weathering.

**Cox and Cu Horizons**  Oxidized C horizon for Cox and unweathered C horizon for Cu.

**Cr Horizon**  Weathered rock formed in place.

**R Horizon**  Consolidated bedrock underlying soil.

The depth of the dynamic weathering zone described above is largely controlled by environmental factors including net precipitation and temperature, as illustrated in Fig. 1.10. If, for example, the weathering process occurs at a location where the net flux of water at the ground surface is downward (i.e., precipitation exceeds evaporation), the weathering front and associated dissolved minerals will extend relatively deep below the ground surface. If, on the other hand, the net flux of water from the ground surface is upward (i.e., evaporation exceeds precipitation), the weathering front will be relatively shallow and dissolved minerals may be deposited in horizons relatively near the ground surface. Near surface deposits of calcium carbonate (caliche) common to the arid regions of the western United States and Australia are an excellent example of the latter phenomenon. The depth and rate of pore water and pore vapor movement are largely controlled by the unsaturated hydrologic characteristics of the deposit, as quantified in Part III.

### 1.3.4 Unsaturated Soil in Engineering Practice

For many years, unsaturated soils were either ignored in civil engineering design and construction analyses or were approached inappropriately from the traditional framework of saturated soil mechanics. Rapid advancement in our understanding of unsaturated soil behavior over the last 30 to 40 years, however, has led today’s civil engineer to realize that there is now an opportunity to approach problems involving unsaturated soil on a much more rational basis. The expanding knowledge base on the fundamental principles of unsaturated soil mechanics is increasingly being incorporated into a diverse array of practical engineering problems. The following lists summarize several of the more common types of engineering problems involving predominantly unsaturated soils.

**Flow-Related Problems**

1. Water balance at the interface of soil and atmosphere
2. Net recharge rate to the saturated zone or aquifers
3. Design of final covers for underground waste storage and containment
4. Near-surface contaminant transport and remediation
5. Transient and steady seepage in unsaturated embankment dams
Figure 1.10 Relative depth of weathering and soil formation as related to environmental factors along a transect from the equator to the north polar region (Birkeland, 1999; reproduced by permission of Oxford University Press, Inc.).
Stress-Related Problems

1. Slope stability and land sliding under changing climatic conditions
2. Lateral earth pressure and stability of retaining structures
3. Excavation and bore hole stability
4. Bearing capacity for shallow foundations under moisture loading
5. Stress wave propagation in unsaturated soil

Deformation-Related Problems

1. Swelling and shrinkage of expansive soil
2. Desiccation cracking of clay
3. Collapsing soil
4. Consolidation and settlement of unsaturated soil
5. Soil compaction

1.4 MOISTURE, PORE PRESSURE, AND STRESS PROFILES

1.4.1 Stress in the Unsaturated State

Subsurface moisture, suction, and stress profiles depend on the soil and pore water properties as well as the prevalent environmental or atmospheric conditions. Soil type, particle size distribution, and pore size distribution all act to influence the equilibrium distribution and flow of pore water within the soil profile. Atmospheric conditions, which include relative humidity, temperature, wind speed, and precipitation, all act to influence transient changes in the flow and distribution of the subsurface pore water.

The mechanical stability of any point in the subsurface depends on the strength parameters of the soil and the state of stress at that point. In saturated soil, the state of stress can be described by total stress and pore pressure, unified under the concept of effective stress. Effective stress, which is the difference between total stress and pore pressure, is the stress experienced by soil’s solid phase, or skeleton. The state of effective stress controls whether or not a given soil mass is under a state of stability or a state of failure. Soil strength is an intrinsic material property that generally depends on the soil mineralogy, particle morphology, and interparticle arrangement. Macroscopic description of these controlling factors often leads to empirical material parameters, most notably cohesion and internal friction angle. These material parameters, together with the stress state variables, define the boundaries controlling whether soils are in stable or failure conditions.

Total stress can be considered as an external stress and is due to either surcharge load or the soil’s self-weight. Pore pressure in saturated soil is generally compressive and isotropic. All pore pressure in saturated soil contributes to total stress according to the effective stress principle. Pore pressure
in unsaturated soil, on the other hand, is generally tensile. The contribution of pore pressure to total stress depends on the degree of saturation and pore size distribution. This contribution is not always 100%, making analysis of the state of stress in unsaturated soil far more complicated than the relatively simple case for saturated conditions. A detailed micromechanical analysis of the origin and behavior of this stress follows in Chapter 4. The following sections contain several conceptual examples to illustrate the differences in the states of stress for saturated and unsaturated soil.

1.4.2 Saturated Moisture and Stress Profiles: Conceptual Illustration

Consider a homogeneous soil layer that is initially saturated and free of surcharge loading as shown in Fig. 1.11a. The water table is at the ground surface and the soil layer is bounded by a layer of bedrock below. The total vertical stress profile within the soil layer due to self-weight is a function of depth as follows:

\[ \sigma_z = \gamma z \]  

(1.1)

where \( \gamma \) is the bulk (total) unit weight of the soil and \( z \) is the depth from the ground surface. Horizontal stresses (\( \sigma_x \) and \( \sigma_y \)) may be estimated from the vertical stresses under the at-rest, or \( K_0 \), condition:

\[ \sigma_x = \frac{\mu}{1 - \mu} \sigma_z + \frac{1 - 2\mu}{1 - \mu} u_w \]  

(1.2)

where \( \mu \) is Poisson’s ratio and \( u_w \) is the pore water pressure. The pore water pressure profile under both the saturated and unsaturated hydrostatic condition is as follows:

\[ u_w = \gamma_w z_w \]  

(1.3)

Figure 1.11 Conceptual profiles of water content, stress, and effective stress in a homogeneous, saturated soil layer.
where $\gamma_w$ is the unit weight of water (9.8 kN/m$^3$) and $z_w$ is the distance from the water table to the point of interest and is positive for points located below the water table. The vertical effective stress under the saturated condition is

$$\sigma'_z = \sigma_z - u_w = \gamma z - \gamma_w z_w$$  \hspace{1cm} (1.4a)

and the horizontal effective stress under the $K_0$ condition is

$$\sigma'_x = \frac{\mu}{1 - \mu} \sigma_z'$$  \hspace{1cm} (1.4b)

Conceptual profiles of volumetric water content ($\theta = V_w/V_t$), total vertical stress $\sigma_z$, pore pressure $u_w$, and effective vertical and horizontal stress ($\sigma'_z$ and $\sigma'_x$) for the saturated soil layer shown in Fig. 1.11a are plotted in Figs. 1.11b, 1.11c, and 1.11d. Because the soil is saturated, the saturated volumetric water content is a constant equal to the soil porosity, $n$ (i.e., $\theta_s = V_{w_s}/V_t = V_{c_s}/V_i = n$). For a quantitative analysis, consider a 10-m-thick, homogeneous, saturated sand layer with Poisson’s ratio equal to 0.35, bulk unit weight equal to 18.8 kN/m$^3$, and porosity equal to 30% (Figure 1.12a). The corresponding volumetric water content profile is shown in Fig. 1.12b, the vertical total stress and pore pressure profiles are shown in Fig. 1.12c, and the vertical and horizontal effective stress profiles are shown in Fig. 1.12d. Each of these profiles is a linear function with depth.

### 1.4.3 Unsaturated Moisture and Stress Profiles: Conceptual Illustration

Unsaturated soils in the field are characterized by a water table located at some depth below the ground surface. If, for example, the water table in the preceding example drops 10 m to the interface of the soil layer and bedrock,
1.4 MOISTURE, PORE PRESSURE, AND STRESS PROFILES

1.4.4 Illustrative Stress Analysis

Comparison of the effective stress profiles under saturated and unsaturated conditions for the preceding examples (Figs. 1.12d and 1.13d) reveals that...

![Figure 1.13](image)

Figure 1.13 Profiles of water content, stress, and effective stress in a sandy soil layer under unsaturated conditions.
both the horizontal and vertical effective stresses increase considerably upon
desaturation over the entire depth of the soil layer. The fundamental question
for geotechnical engineers is: Which effective stress profile, saturated or un-
saturated, is more representative? Considering the traditional Mohr-Coulomb
failure criterion provides one answer to this question.

The Mohr-Coulomb failure criterion can be written in terms of cohesion
c’ and effective internal friction angle \( \phi’ \) in the space of shear stress at failure
\( \tau_f \) and normal effective stress \( \sigma_n’ \) as follows:

\[
\tau_f = c' + \sigma_n' \tan \phi'
\] (1.6)

For loose, un cemented sand that is either completely dry or completely
saturated, the cohesion term in eq. (1.6) may be considered essentially equal
to zero. For wet or moist sand, however, considerable cohesive strength may
exist. This “apparent cohesion” unique to unsaturated soil arises from neg-
ative pore water pressure and surface tension effects occurring at the interface
of the pore water, pore air, and soil solids among the unsaturated soil grains.

Consider a point in the sand layer from the preceding examples located 8
m from the ground surface. Assume a friction angle equal to 35° for both the
saturated case (where the water table is at the surface) and the unsaturated
case (where the water table is 10 m from the surface). For the saturated case
(Fig. 1.14a), where \( z = z_w \), and \( \gamma = 18.8 \text{ kN/m}^3 \), the vertical and horizontal
effective stresses at \( z = 8 \text{ m} \) are

\[
\sigma'_z = (\gamma - \gamma_w)z = (18.8 - 9.8)(8) = 72 \text{ kPa}
\] (1.7a)
\[
\sigma'_x = \frac{\mu}{1 - \mu} \sigma'_z = \frac{0.35}{1 - 0.35} (72) = 39 \text{ kPa}
\]

For the unsaturated case (Fig. 1.14b), where \( z = 8 \text{ m}, z_w = -2 \text{ m}, \) and \( \gamma =
15.0 \text{ kN/m}^3 \), the effective stresses are

\[
\sigma'_z = (\sigma_z - u_a) + \chi(u_a - u_w) = [(15)(8) - 0] + (0.5)[0 - (9.8)(-2)]
= 130 \text{ kPa}
\]
\[
\sigma'_x = \frac{\mu}{1 - \mu} \sigma'_z = \frac{0.35}{1 - 0.35} (130) = 70 \text{ kPa}
\] (1.7b)

If the unsaturated soil at 8 m depth experiences an apparent cohesion of
50 kPa due to capillarity, Mohr’s circles for the saturated and unsaturated
states of effective stress and the corresponding failure envelopes are shown
on Fig. 1.14c. Note that the saturated state of stress falls relatively close to
the saturated failure envelope. The unsaturated state of stress, however, be-
comes far more stable because the envelope shifts upward and the Mohr circle
shifts towards the right.

The practical importance of accounting for the unsaturated condition in
geotechnical engineering applications can be demonstrated by estimating the
Figure 1.14 Conceptual stress analysis for a sandy soil layer under saturated and unsaturated conditions: (a) saturated soil profile, (b) unsaturated soil profile, and (c) states of stress at \( z = 8 \) m and Mohr-Coulomb failure envelopes.
ultimate bearing capacity of a strip footing assumed to be constructed on this soil. According to classical theory, the ultimate bearing capacity \( q_u \) is as follows:

\[
q_u = c'N_c + \gamma DN_q + \frac{1}{2} \gamma BN_\gamma
\]  

(1.8)

where \( D \) is the footing embedment depth, \( B \) is the footing width, and \( N_c, N_q, \) and \( N_\gamma \) are bearing capacity factors. For \( \phi = 35^\circ, N_c = 46.12, N_q = 33.30, \) and \( N_\gamma = 48.03. \) For \( D = 1 \text{ m}, B = 1 \text{ m}, \gamma = 18.8 \text{ kN/m}^3, \) and \( c' = 0 \) under the saturated condition, the ultimate bearing capacity is 1078 kPa. For \( \gamma = 15.0 \text{ kN/m}^3 \) and assuming an apparent cohesion 10 kPa for the unsaturated condition, the ultimate bearing capacity increases to 1321 kPa, an increase of 32%. For an apparent cohesion 50 kPa, the capacity increases to 3166 kPa, an increase of 220%.

Is the notion that unsaturated soil is in a more stable state than saturated soil generally true? Do the changes in degree of saturation and water content that occur in the field under natural and manmade influences such as precipitation, evaporation, irrigation, or water table lowering significantly affect the state of stress and consequent stability of near-surface soil? Does negative pore pressure in unsaturated soil entirely contribute to total stress as positive pore pressure does in saturated soil? How can the differences in the stress conditions for saturated and unsaturated soil and the consequent differences in strength be effectively formulated and quantified? These types of questions have extremely important bearing on stress and deformation problems involving unsaturated soil in geotechnical engineering practice and will be addressed throughout the remainder of this book.

1.5 STATE VARIABLES, MATERIAL VARIABLES, AND CONSTITUTIVE LAWS

1.5.1 Phenomena Prediction

Many physical phenomena are constantly occurring or changing in behavior as functions of both space and time. The requirement for variables to describe these different phenomena results largely from our desire to predict their occurrence or behavior in the future. State variables, material variables, and constitutive laws are commonly used for phenomena prediction.

The number of state and material variables and constitutive laws used to define a given phenomenon depends on the conceptualization. For example, the strength of saturated soil can be represented by the Mohr-Coulomb criterion, which uses normal stress and shear stress as state variables to define the state of stress. To evaluate the stability of the soil under this state of stress, a series of conjugate material properties describing the strength characteristics of the soil must also be introduced. For the Mohr-Coulomb criterion, these properties are friction angle and cohesion. Such properties are
usually referred to as *material variables* because they generally vary with the state variables, just as friction angle and cohesion may depend on whether soil is under drained or undrained state conditions.

State variables are used to describe phenomena occurring in nature and engineering practice. By general definition, state variables are those that are required to completely describe the state of the system for the phenomenon at hand (Fung, 1965). For example, to describe today’s weather, one may use terminology including temperature, pressure, relative humidity, or wind speed. These quantitative descriptors are the state variables defining the state of the weather conditions. Effective stress is the state variable required to describe strength and deformation in saturated soil. Pressure and temperature are often used as state variables to describe the thermodynamic state of a system.

Following a macroscopic or phenomenological formalism under the framework of continuum mechanics and thermodynamics, state variables are not required to be non–material dependent quantities, but have been done so traditionally. The non–material dependent formalism is most effective for phenomena such as heat transfer, mass transport, wave propagation, and chemical reactions occurring in single-phase media such as liquid, gas, or solid or in equivalent continuum media such as saturated soil. In multiphase systems, however, the relative amount of each phase comprising the system often directly controls physical processes such as flow, stress, and deformation phenomena. In unsaturated soil, for example, a decrease in the relative amount of the pore water phase (i.e., a desaturation process) implies a drier soil or a soil with a lower hydraulic head. Consequently, phenomena involving fluid flow, stress, or deformation are likely to occur. Because the amount of pore water corresponding to a given value of head is highly dependent on the type of soil (clearly a material property), describing the mechanical and hydrological behavior of the multiphase system without involving the characteristic relationship between water content and head is impossible. Thus, for multiphase systems such as unsaturated soil, the commonly used conceptualization defining state variables as independent of material variables may be limiting and ineffective. In such cases, according to continuum mechanics, using material variables in conjunction with state variables to describe the state of a multiphase system is necessary (Fung, 1965).

In soil mechanics, it is convenient to differentiate between stress state variables, deformation state variables, and flow state variables. Common stress state variables are the total stress tensor, pore pressure, the effective stress tensor, Coulombian shear stress, and the first, second, and third stress invariants. Commonly used deformation state variables are the strain tensor, the first, second, and third stress and strain invariants, and void ratio. Widely used flow state variables are the degree of pore water saturation, water content, and total hydraulic head.

Material variables are intrinsic properties that depend on the type of material, usually varying from one material to another material or from one state to another state. Examples of material variables are elastic modulus, perme-
ability, and compressibility. Material variables can also be functions of state variables such as pressure, temperature, and stress. The viscosity of water, for example, a material variable, decreases with increasing temperature, a state variable. Similarly, the hydraulic conductivity of unsaturated soil has been effectively conceptualized as a function of the state variables water content or matric suction. Depending on the type of problem (e.g., flow, stress, or deformation), a number of material variables for describing the physical behavior of unsaturated soil have been widely used, several of which are introduced in Chapter 2.

Constitutive laws or equations describe the interrelationships between or among state variables and material variables. A constitutive equation for a given system allows the prediction of one state variable from others. For example, for an elastic soil, one-dimensional stress can be predicted from one-dimensional strain if the elastic modulus is known. Similarly, discharge velocity in soil can be predicted from the hydraulic gradient using the material variable hydraulic conductivity.

A typical path from physical observation to behavior prediction is shown in Fig. 1.15. Here, physical observation and measurement provides a basis for defining and quantifying the state variables and material variables that describe the phenomenon of interest (e.g., hydraulic conductivity and total head are material and state variables required for predicting steady-state fluid flow behavior). Physical laws provide the connection between the state and material variables governing the physical process (e.g., Darcy’s law). Thermodynamic principles (e.g., the first and second laws of thermodynamics) are applied to the physical laws to arrive at governing equations for predicting behavior in space and time (e.g., quantifying a hydraulic potential field). Comparing the predicted behavior with observations of actual behavior provides a basis to refine the state and material variables for improving the prediction.

1.5.2 Head as a State Variable

Total head is often used as a state variable to describe flow phenomena in soil. The total head concept is generally applicable to both saturated and unsaturated conditions. Use of total head for describing fluid flow stems from consideration of thermodynamic law, which assumes that energy flows from a place of higher value to a place of lower value.

Fundamentally, total head is the potential of the water retained in the soil pores. For many geotechnical engineering applications occurring on a relatively macroscopic scale (e.g., larger than the particle scale), the total head $h_t$ responsible for the flow of pore water at a given point can be sufficiently represented by the summation of the elevation head $h_e$ and pressure head $h_p$ at that point as follows:

$$h_t = h_e + h_p \tag{1.9a}$$

or
Physical Observation

State Variables

Material Variables and Constitutive Laws

Physical Laws

Mass and Energy Conservation

Behavior Prediction

Refinement

Refinement

Figure 1.15  Iterative process leading from physical observation and measurement to behavioral prediction.

\[ h_i = z + \frac{u_w}{g \rho_w} \]  \hspace{1cm} (1.9b)

where \( z \) is the vertical coordinate distance from a prescribed datum (m), \( u_w \) is the pore water pressure (Pa), \( g \) is gravitational acceleration (m/s\(^2\)), and \( \rho_w \) is the density of water (kg/m\(^3\)).

Darcy’s law describes discharge velocity \( v \) as a function of the gradient of total head (a state variable) and hydraulic conductivity (a material variable) as follows:

\[ v = -k \frac{dh_i}{dx} \]  \hspace{1cm} (1.10)

where \( x \) is the coordinate in the direction of flow (m) and \( k \) is hydraulic conductivity (m/s). Darcy’s law for fluid flow, therefore, is a direct application
of thermodynamic law, that is, energy flows from a higher place to a lower place.

The major difference between total head in saturated soil and total head in unsaturated soil is that the pressure head governed by the pore pressure $u_w$ is positive (compressive) in saturated soil and negative (tensile) in unsaturated soil. The pressure head in unsaturated soil is also highly dependent on the degree of saturation or water content and type of soil. Examining pressure and head profiles for a soil layer under the hydrostatic ("no-flow") condition illustrates these major differences. Consider, for example, the homogeneous soil layer shown in Fig. 1.16a. The soil located above the water table can be conceptualized as three regimes: (1) a regime where the soil remains saturated under negative pore water pressure, often referred to as the capillary fringe, (2) an unsaturated regime characterized by a continuous water phase, or funicular regime, and (3) a residual or pendular regime characterized by an isolated, discontinuous water phase. The transitions between each of these regimes, which are by no means well defined in typical field settings, are largely controlled by the pore size and pore size distribution of the soil. As illustrated by the degree of saturation profile shown on Fig. 1.16b, the point where desaturation commences in the soil located above the water table is referred to as the air-entry point. The hydraulic head associated with this point is referred to as the air-entry head.

According to thermodynamics, the total head in the soil profile under no-flow conditions must be the same everywhere at equilibrium. Accordingly, two piezometers, one placed in the saturated zone at point $A$ (Fig. 1.16a), and the other in the capillary fringe zone at point $B$, indicate identical total head values. The pressure head at $A$ is positive. The pressure head at $B$ is negative. Assigning an elevation datum to the bottom of the saturated soil layer, the pore pressure profile varies linearly from zero at the water table, to positive values below the water table, and to negative values above the water table. In the unsaturated zone, equilibrium requires the pore water potential to be equal among the three different phases comprising the soil (i.e., pore water, pore gas, and soil solids). This requirement provides the physical basis governing the height of the capillary fringe and the relative humidity of the pore gas in the unsaturated zone. Detailed thermodynamic treatment of these interfacial equilibrium concepts is provided in Chapters 3 and 4.

1.5.3 Effective Stress as a State Variable

Effective stress is considered a fundamental state variable for describing the state of stress in soil. For saturated soil, Terzaghi (1943) defined effective stress as the difference between the total stress and pore pressure. Physically, effective stress describes the stress acting on the soil skeleton. One can fully define the effective stress at any point of interest in saturated soil as long as the total stress and pore pressure are known.
Figure 1.16  Conceptual illustration of the unsaturated soil zone: (a) pore water regimes and (b) saturation, total head, and pore pressure profiles.
For unsaturated soil, the physical meaning of effective stress remains the same. However, two additional factors must be considered: (1) the stress acting through the air phase (i.e., the pore air pressure, $u_a$) and (2) the difference between the pore air pressure and the pore water pressure, or matric suction. Bishop’s (1959) widely cited effective stress approach for unsaturated soil expands Terzaghi’s classic effective stress equation as follows:

$$\sigma' = (\sigma - u_a) + \chi(u_a - u_w)$$ (1.11)

The difference $\sigma - u_a$ is referred to as the net normal stress, the difference $u_a - u_w$ is matric suction, and the effective stress parameter $\chi$ is a material variable that is generally considered to vary between zero and unity. For $\chi = 0$, corresponding to completely dry soil, and for $\chi = 1$, corresponding to fully saturated soil, eq. (1.11) reduces to Terzaghi’s classic effective stress equation for describing the behavior of saturated soil ($\sigma' = \sigma - u_w$).

The material variable $\chi$ is captured by its strong dependency on the degree of pore water saturation $S$:

$$\chi = \chi(S)$$ (1.12)

Determination of the effective stress parameter and its dependency on the amount of water in the system is essential in order to evaluate effective stress in unsaturated soil. Figure 1.17 illustrates this dependency on saturation for several types of soil. Measurement of this function is experimentally challenging, particularly near the low saturation range. For relatively high degree of saturation, indirect measurement of $\chi$ is possible via shear testing under
controlled suction conditions. Theoretical studies have also shown that \( \chi \) could exceed unity and is highly nonlinear in the transitional regime between conditions of isolated pore water menisci and continuous pore water. The nature of \( \chi \), as well as its determination by experimental techniques, are important and wide open subjects in unsaturated soil mechanics. A systematic description of the nature of \( \chi \) from both microscopic and macroscopic perspectives is provided in Chapters 5 and 6.

### 1.5.4 Net Normal Stresses as State Variables

Bishop’s original treatment of effective stress as a single-valued stress state variable for unsaturated soil [eq. (1.11)] has been challenged from theoretical, experimental, and philosophical perspectives. Jennings and Burland (1962), for example, explored the limitations of using the effective stress concept and suggested that it may not be adequate for describing deformation phenomena such as collapse upon wetting. Khalili et al. (2004) and others contend that this argument has been formulated within the context of a linear elastic framework and that nonrecoverable (plastic) deformations such as collapse can indeed readily be described within an effective stress framework using elastoplastic theories. The effective stress approach for unsaturated soil mechanics continues to be the subject of debate. Of specific issue is the necessity to include the material variable \( \chi \), which may or may not be readily determined, in defining effective stress.

In wrestling with these apparent difficulties, Coleman (1962) suggested the use of net normal stress \( \sigma - u_a \) and matric suction \( u_a - u_w \) as stress variables to describe stress-strain relations for unsaturated soil. Further work by Bishop and Blight (1963) illustrated some advantages of using net normal stress and matric suction as stress state variables. Fredlund and Morgenstern (1977) considered the approach from both experimental and theoretical standpoints and formally proposed the use of net normal stress and matric suction as independent stress state variables for unsaturated soil. Here, it was considered that any two of three stress variables \( \sigma - u_a, \sigma - u_w, \) or \( u_a - u_w \) may be used, that is, \( \sigma - u_a \) and \( u_a - u_w \), \( \sigma - u_w \) and \( u_a - u_w \), or \( \sigma - u_a \) and \( \sigma - u_w \). The proposed approach was supported by a series of “null” triaxial tests conducted earlier (Fredlund, 1973) where it was shown that the volume of unsaturated soil specimens remains relatively unchanged if changes in the proposed stress state variables were prohibited.

Studies conducted over the past two decades have demonstrated renewed interest in Bishop’s effective stress approach, as well as support for Fredlund and Morgenstern’s (1977) independent stress state variable approach. Identifying the most appropriate stress state variables for unsaturated soil remains a highly active area of research. Detailed introductions and analyses of each approach, their major differences, their very different roles in describing the shear strength of unsaturated soil, and reconciliation between these two approaches are provided in Chapters 5 and 6.
1.6 SUCTION AND POTENTIAL OF SOIL WATER

1.6.1 Total Soil Suction

Total soil suction quantifies the thermodynamic potential of soil pore water relative to a reference potential of free water. Free water in this regard is defined as water containing no dissolved solutes, having no interactions with other phases that impart curvature to the air-water interface, and having no external forces other than gravity. The physical and physicochemical mechanisms responsible for total soil suction are those that decrease the potential of the pore water relative to this reference state.

Neglecting temperature, gravity, and inertial effects, the primary mechanisms that decrease the potential of soil pore water include capillary effects, short-range adsorption (particle-pore water interaction) effects, and osmotic effects. The former mechanism is unique to unsaturated soil. The latter two may occur under either saturated or unsaturated conditions.

Capillary effects, which include curvature of the air-water interface and the associated negative pore water pressures in the three-phase unsaturated soil system, are described in detail in Chapters 3 and 4.

Short-range adsorption effects arise primarily from electrical and van der Waals force fields occurring within the vicinity of the solid-liquid (i.e., soil-pore water) interface and are most important for fine-grained soils. Electrical fields emanate from the net negative charge on the surface of clay minerals. van der Waals fields arise from atomic scale interactions between the molecules comprising the surface of the solid phase (i.e., the soil particles) and the molecules comprising the liquid phase (i.e., the pore water) and occur for all types of soil. The effect of each of these fields is most pronounced for water adsorbed by clay particles, which posses both significant net surface charge and relatively large surface area. The strength of electrical and van der Waals fields decays rapidly as the distance from the particle surface increases. Accordingly, short-range adsorption effects are most relevant at relatively low water content or degree of saturation when the adsorbed pore water is primarily in the form of thin films coating the particle surfaces.

Osmotic effects are the result of dissolved solutes in the pore water. Dissolved solutes may arise from two sources: as externally introduced solutes (e.g., through natural leaching processes), or as naturally occurring solutes adsorbed by the soil mineral surfaces (e.g., exchangeable cations adsorbed by clay particles). Hydration and solvation of such dissolved solutes and the associated structural ordering of neighboring water molecules reduces the chemical potential of the pore water to a degree dependent on the dissolved solute concentration.

Suction arising from the combined effects of capillarity and short-range adsorption is usually grouped under the more general term matric suction, which may be designated in units of pressure as $\psi_m$. The term matric reflects earlier usage of the term matrix, which was intended to describe the com-
ponent of suction arising from interactions between the pore water and the soil solids, or soil matrix. Suction arising from the presence of dissolved solutes is referred to as *osmotic* suction, or $\psi_o$. Total soil suction $\psi_t$ is generally considered the algebraic sum of the matric and osmotic components, which may be written as follows:

$$\psi_t = \psi_m + \psi_o$$  \hspace{1cm} (1.13)

### 1.6.2 Pore Water Potential

The thermodynamic potential of soil pore water is most rigorously described in terms of chemical potential. Chemical potential, typically designated $\mu$, has units of energy per unit mass, measured in either joules per mole (J/mol) or joules per kilogram (J/kg). Pore water chemical potential represents the amount of energy stored per unit mass of pore water. Potential in the latter units of joules per kilogram is often referred to as free energy per unit mass, or $E$. Chemical potential is the primary criterion for equilibrium with respect to the transfer of energy within any given phase of matter (e.g., water) or from one phase of certain matter to another phase of the same matter (e.g., from liquid water to water vapor). Equilibrium requires that energy be transferred from locales or phases of relatively high chemical potential to locales or phases of relatively low chemical potential. In a closed system at equilibrium, the chemical potential of the matter under consideration is the same at every point within each phase and among all of the phases.

Describing the energy state of soil pore water is best accomplished by considering the change from a reference condition for free water. The total change in pore water potential $\Delta \mu_t$, resulting from the various physical and physicochemical suction mechanisms in unsaturated soil can be written as follows:

$$\Delta \mu_t = \Delta \mu_c + \Delta \mu_o + \Delta \mu_e + \Delta \mu_f$$  \hspace{1cm} (1.14)

where $\Delta \mu_c$ is the change in potential due to curvature at the air-water interface (i.e., capillary effects), $\Delta \mu_o$ is the change due to dissolved solute effects (i.e., osmotic effects), $\Delta \mu_e$ is the change to due the presence of electrical fields, and $\Delta \mu_f$ is the change due to van der Waals fields. Each term on the right-hand side of eq. (1.14) is a negative value, reflecting a decrease or decrement in chemical potential associated with each mechanism. Soil suction is a positive value because it describes this decrement relative to a reference potential for free water equal to zero.

As described in detail in Chapter 4, curvature at the air-water interface in unsaturated soil decreases the chemical potential (J/mol) of soil pore water an amount described by a form of the Young-Laplace equation as follows:
\[
\Delta \mu_c = -T_s \nu_w \left( \frac{1}{R_1} + \frac{1}{R_2} \right)
\] (1.15)

where \( T_s \) is the surface tension of the water (mN/m), \( R_1 \) and \( R_2 \) are principal radii describing the curvature of the air-water interface (m), and \( \nu_w \) is the partial molar volume of water (m³/mol). As the net curvature of the air-water interface located between and among unsaturated soil grains increases (i.e., as the soil desaturates and the pore pressure becomes more negative), the decrement in chemical potential becomes greater.

For ideal and dilute solutions, the decrement in chemical potential (J/mol) due to the presence of dissolved solutes may be approximated by a form of the van’t Hoff equation:

\[
\Delta \mu_o = -CRT \nu_w = -\pi \nu_w
\] (1.16)

where \( C \) is the molar concentration of the pore solute solution (mol/m³), \( R \) is the universal gas constant (J/mol \( \cdot \) K), and \( T \) is temperature (K). The product \( CRT \) in the above equation is commonly referred to as osmotic pressure, or \( \pi \). Under more general, nondilute conditions, osmotic pressure is described by the viral equation (e.g., Shaw, 1992):

\[
\pi = CRT \left( 1 + B_2 C^2 + B_3 C^3 + \cdots \right)
\] (1.17)

where \( B_2, B_3, \ldots \) are viral coefficients. As the concentration term approaches zero, the viral equation (1.17) approaches the van’t Hoff approximation (1.16). As shown in Fig. 1.18, as the concentration of dissolved solutes increases, the osmotic pressure increases. The corresponding chemical potential of the pore water solution decreases.

Because H₂O is a polar molecule, the physical consequence of short-range electrical fields emanating from soil particle (clay mineral) surfaces is to attract, align, and impart order into the molecular arrangement of neighboring pore water. Considering an interaction with a single particle surface, the corresponding decrement in chemical potential of the pore water is dependent on the location of the water relative to the particle surface and may be quantified as follows (e.g., Iwata et al., 1995):

\[
\Delta \mu_e = \int_0^D \frac{D \nu_w}{4\pi} \left( \frac{1}{\epsilon} - 1 \right) dD
\] (1.18)

where \( \epsilon \) is the partial dielectric constant of the pore water and \( D \) is the value of electric displacement at the point where the water exists. The value of \( D \) depends on the shape and size of the soil particle, its surface charge density, and the distance from the particle surface to the water under consideration.
1.6 SUCTION AND POTENTIAL OF SOIL WATER

Figure 1.18 Osmotic pressure as a function of dissolved solute concentration as predicted by eq. (1.16).

As water molecules move closer to the surface of charged soil particles, such as in the thin films surrounding clay particles at very low water content, their potential energy is reduced just as if they were falling in a gravity field or being driven along by a pressure gradient.

The decrement in chemical potential due to van der Waals fields is also dependent on the location of pore water molecules relative to the soil particle surface. The magnitude of this decrement generally depends on the assumed shape of the particle surface (e.g., planar, spherical), which may be captured by the following functional dependency:

$$\Delta \mu_f = -Af(z, \Gamma)$$  \hspace{1cm} (1.19)

where $A$ is Hamaker’s constant for the soil-water interaction, $z$ is the distance from the pore water molecule to the particle surface, and $\Gamma$ is a shape function.

Equations (1.18) and (1.19) dictate that the decrements in chemical potential due to electrical and van der Waals fields are much less at locations relatively far from particle surfaces than at locations near the surfaces. How-
ever, local thermodynamic equilibrium requires the chemical potential throughout the entire pore water phase to be the same. To satisfy this requirement, positive internal water pressure, which acts to increase the potential, builds up in the water films immediately adjacent to the soil particles. The change in potential at any point due to the internal pressure buildup is equal and opposite in magnitude to the change produced by the electrical field and van der Waals fields. Accordingly, the magnitude of the internal pressure decreases as the distance from the particle surface increases. Macroscopically, the positive pore pressure immediately adjacent to soil particles is manifested as a convexly curved film coating the particles (much like a film of oil coating a ball). Thus, the pore water pressure in unsaturated soil may actually be either positive or negative depending on the location of the pore water relative to the particle surface and the scale of the problem under consideration (e.g., Olson and Langfelder, 1965; Nitao and Bear, 1996). For most practical geotechnical engineering purposes, the problems of interest occur on a relatively macroscopic scale and equivalent negative pore water pressures are considered.

The total reduction in chemical potential associated with each of the above mechanisms defines the total suction of the soil-water system. The magnitude of the potential reduction is dependent on the amount of pore water in the system. The constitutive relationship that describes this dependency is referred to as the soil-water characteristic curve, or SWCC.

1.6.3 Units of Soil Suction

The potential of soil pore water may be expressed as an energy per unit mass, a chemical potential (i.e., J/kg or J/mol), as an energy per unit volume, a pressure potential (i.e., J/m$^3$ = N · m/m$^3$ = N/m$^2$ = Pa), or as an energy per unit weight, a head potential (i.e., J/N = N · m/N = m). Conversion among units of potential $\mu$, pressure $\psi$, and head $h$ may be achieved by considering the following equivalency:

$$\mu = \psi \nu_w = hg \omega_w$$  \hspace{1cm} (1.20)

where $g$ is gravitational acceleration (m/s$^2$), $\omega_w$ is the molecular mass of water (kg/mol), $\nu_w$ is the partial molar volume of water (m$^3$/mol), and chemical potential $\mu$ is in units of Joules per mole. Equations for direct conversion among these various units are summarized in Table 1.1.

Pore water potential in units of either head $h$ or pressure $\psi$ are preferred for describing flow, stress, and deformation phenomena in unsaturated soil mechanics. Use of the term suction head generally refers to pore water potential in units of head. The preferred term in geotechnical engineering practice, soil suction, refers to pore water potential in units of pressure. The International System of Units (SI) of suction pressure are pascals (Pa), which are typically described in terms of kilopascals (kPa) for the range of magni-
1.6 SUCTION AND POTENTIAL OF SOIL WATER

### TABLE 1.1 Conversion Chart for Pore Water Potential Terms

<table>
<thead>
<tr>
<th>Potential (J/mol)</th>
<th>Head (m)</th>
<th>Pressure (kPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>—</td>
<td>$h = \frac{\mu}{g \omega_w}$</td>
<td>$\psi = \frac{\mu}{\psi}$</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\mu = h \omega_w$</td>
<td>$h = \frac{\psi \omega_w}{g \omega_w} = \frac{\psi}{\rho_w g}$</td>
<td>$\psi = \frac{h g \omega_w}{\psi} = h g p_w$</td>
</tr>
</tbody>
</table>

$p_w = \omega_w / \nu_w$.

Pore water potential relevant to most practical unsaturated soil mechanics applications. Alternative units of (pF), which were common in the early literature, are defined as the logarithm of pore water potential in units of head in centimeters of water (Schofiel, 1935):

$$pF = \log(cm_{H_2O})$$  \hspace{1cm} (1.21)

For convenience, units of pF may be approximated in terms of kilopascals as follows:

$$10^{(pF-1)} \approx kPa$$  \hspace{1cm} (1.22)

for example, for pF = 4:

$$10^{(4-1)} \approx 10^3 \text{ kPa} = 1000 \text{ kPa}$$

An additional series of commonly used equivalent units for pore water potential is summarized in Table 1.2.

#### 1.6.4 Suction Regimes and the Soil-Water Characteristic Curve

The relative importance of the individual physical and physicochemical mechanisms responsible for soil suction depends on the water content of the unsaturated soil-water-air system. At relatively low values of water content and correspondingly high values of suction, where pore water is primarily in the form of thin films on the particle surfaces, the dominant mechanisms contributing to suction are the relatively short-range adsorption effects governed by the surface properties of the soil solids. On the other hand, at relatively high values of water content and correspondingly low values of suction, the dominant pore water retention mechanism becomes capillarity, governed primarily by the particle and pore structure and pore size distribution. Osmotic suction is constant over the entire range of water content unless the concentration of dissolved solutes changes.
TABLE 1.2 Equivalent Units for Describing Potential, Head, and Pressure

<table>
<thead>
<tr>
<th>Unit</th>
<th>Equivalent Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Units</td>
<td></td>
</tr>
<tr>
<td>J/kg</td>
<td>100</td>
</tr>
<tr>
<td>J/mol</td>
<td>1.8016</td>
</tr>
<tr>
<td>Head Units</td>
<td></td>
</tr>
<tr>
<td>m H₂O</td>
<td>10.2</td>
</tr>
<tr>
<td>cm H₂O</td>
<td>1020</td>
</tr>
<tr>
<td>inches H₂O</td>
<td>401.5</td>
</tr>
<tr>
<td>feet H₂O</td>
<td>33.42</td>
</tr>
<tr>
<td>mm Hg</td>
<td>750</td>
</tr>
<tr>
<td>inches Hg</td>
<td>29.53</td>
</tr>
<tr>
<td>pF</td>
<td>3.01</td>
</tr>
<tr>
<td>Pressure Units</td>
<td></td>
</tr>
<tr>
<td>Pa</td>
<td>1000000</td>
</tr>
<tr>
<td>kPa</td>
<td>100</td>
</tr>
<tr>
<td>MPA</td>
<td>0.1</td>
</tr>
<tr>
<td>bars</td>
<td>1.0</td>
</tr>
<tr>
<td>millibars</td>
<td>1000</td>
</tr>
<tr>
<td>atmospheres</td>
<td>0.987</td>
</tr>
<tr>
<td>lb/in² (psi)</td>
<td>14.5</td>
</tr>
<tr>
<td>lb/ft² (psf)</td>
<td>2088</td>
</tr>
<tr>
<td>US tons/ft² (tsf)</td>
<td>1.044</td>
</tr>
</tbody>
</table>

The transition between the high suction regime dominated by short-range adsorption mechanisms and the low suction regime dominated by capillary mechanisms is highly dependent on soil type. In fine-grained materials such as clays, for example, a much greater amount of pore water is required to satisfy the relatively large surface hydration energies associated with the high suction regime. In sands, however, very little water is adsorbed under initial surface hydration mechanisms and capillary effects dominate over the majority of the unsaturated water content range. The SWCC describes the corresponding constitutive relationship between soil suction and soil-water content. The general shape of the SWCC for various soils reflects the dominating influence of material properties including pore size distribution, grain size distribution, density, organic material content, clay content, and mineralogy on the pore water retention behavior. Understanding the general behavior of the SWCC and its relationship to the physical properties of the soil that it describes is a critical component of unsaturated soil mechanics. McQueen and Miller (1974) developed an instructive conceptual model based on empirical evidence for describing the general shape and behavior of the SWCC. As illustrated graphically in Fig. 1.19, it was suggested that any SWCC could be approximated as a composite of three straight-line segments on a semilog plot of suction versus moisture content ranging from zero to saturation. These line segments include one extending from 10⁶ to 10⁴ kPa designated the tightly adsorbed segment, a second extending from 10⁴ kPa to approximately 100
1.6 SUCTION AND POTENTIAL OF SOIL WATER

Soil Water Content, \( \theta \)

Air-Entry Suction

Saturated Water Content

Figure 1.19 Illustration of McQueen and Miller's (1974) conceptual model for general behavior of the soil-water characteristic curve.

kPa designated the adsorbed film segment, and a third extending from 100 to 0 kPa (saturation) designated the capillary segment. Each segment is characterized by a change in slope at the transition points.

Within the so-called tightly adsorbed regime, pore water is retained by molecular bonding mechanisms, primarily hydrogen bonding with exposed oxygen or hydroxyl on the surfaces of the soil minerals. Within the adsorbed film segment, water is retained in the form of thin films on the particle surfaces under the influences of short-range solid-liquid interaction mechanisms (e.g., electrical field polarization, van der Waals attraction, and exchangeable cation hydration). The amount of water adsorbed within the first two regimes (i.e., the slopes of the line segments) is a function of the surface area of the soil particles, the surface charge density of the soil mineral, and the type and valency of any adsorbed exchangeable cations. When the adsorbed films on the particle surfaces grow thick enough to extend beyond the range of influence of the short-range solid-liquid interaction effects, the characteristic curve enters a regime dominated by capillary pore water retention mechanisms. The amount of water adsorbed here is a function of the particle and pore size properties, terminating at the air-entry pressure where the capillary air-water interfaces begin to disappear as the system approaches saturation.
Figure 1.20 shows SWCCs representative of sand, silt, and clay that illustrate the general behavior of characteristic curves and their associated dependency on soil type. For sandy soil, the surface adsorption regime in the high suction range is generally very limited because the specific surface and surface charge properties of sand are relatively small. Capillarity is the dominant suction mechanism over the majority of the unsaturated water content range, terminating at a relatively low air-entry pressure controlled by the relatively large pore throats formed between and among the sand particles. The overall slope and shape of the capillary regime is controlled primarily by the pore size distribution of the material. Soils with a relatively narrow pore size distribution are marked by relatively flat characteristic curves in the capillary regime because the majority of pores are drained over a relatively narrow range of suction. Silty soil may adsorb a significantly greater amount of water under short-range adsorption mechanisms because the specific surface area of silt is much larger than sand. The air-entry pressure of silt is also larger as controlled by the relatively small pores. Clay has the highest capacity for water adsorption under short-range surface interaction effects because clay particles have charged surfaces and very high specific surface area. As illustrated by the experimental data in Fig. 1.21, highly expansive clays (e.g., smectite) are capable of adsorbing as much as 20% water by mass during the initial surface adsorption regime and may sustain extremely high suction over
a wide range of water content. Nonexpansive clays (e.g., kaolinite), on the other hand, adsorb much less water in the high suction regime. For expansive clays, the SWCC is more physically meaningful in terms of gravimetric water content since the volume is a variable during the sorption process.

**PROBLEMS**

1.1. Where are the regions in the United States where unsaturated soils are likely encountered to significant depth below the ground surface?

1.2. What kind of climatic conditions tend to lead to the formation of a thick unsaturated zone?

1.3. What is the fundamental difference between saturated soils and unsaturated soils in terms of pore water pressure?

1.4. Describe and illustrate the Mohr-Coulomb failure criterion.
1.5. When the state of stress (i.e., Mohr circle) in a soil reaches the Mohr-Coulomb criterion, what is the state of the stress called?

1.6. Give three examples of unsaturated soil mechanics problems in geo-technical engineering.

1.7. For a given unsaturated soil under either a dry or wet condition, which one has a higher suction?

1.8. What are state variables, material variables, and constitutive laws?

1.9. What are the principal differences between saturated and unsaturated soil profiles of pore water pressure, total stress, and effective stress?

1.10. According to Bishop’s effective stress concept, which state, saturated or unsaturated, has a higher effective stress? Why?

1.11. What is the shape of the pore pressure profile under the hydrostatic condition in saturated and unsaturated states, respectively?

1.12. If an unsaturated soil has a water potential of $-1000 \text{ J/kg}$, what is the equivalent soil suction value? If the soil at the air dry condition has a matric suction of 100 MPa, what is the soil water potential in joules per kilogram?

1.13. Three soils—clay, silt, and sand—are all equilibrated at the same matric suction, which soil has the highest water content and why?

1.14. Describe the major physical and physicochemical mechanisms responsible for soil suction.