

Reverse VSP and Crosswell Seismic Imaging at the Savannah River Site

by

R. J. Cumbest

Westinghouse Savannah River Company
Savannah River Site
Aiken, South Carolina 29808

J. O. Parra

Southwest Research Institute
TX USA

B. J. Zook

Southwest Research Institute
TX USA

C. Addington

Tomoseis Institute
TX USA

V. Price

MASTER

A document prepared for SOCIETY OF EXPLORATION GEOPHYSICS EXTENDED ABSTRACTS GEOPHYSICS at Denver from 04/15/96 - 04/15/96.

DOE Contract No. DE-AC09-89SR18035

This paper was prepared in connection with work done under the above contract number with the U. S. Department of Energy. By acceptance of this paper, the publisher and/or recipient acknowledges the U. S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper, along with the right to reproduce and to authorize others to reproduce all or part of the copyrighted paper.

ds
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22161.

REVERSE VSP AND CROSSWELL SEISMIC IMAGING AT THE SAVANNAH RIVER SITE

by

J. O. Parra
B. J. Zook
C. Addington
R. J. Cumbest
V. Price

February 1996

ABSTRACT

Analysis of crosswell and three-component seismic data integrated with well logs have produced information on the distribution of subsurface heterogeneities below the In-Tank Precipitation facility at the Savannah River Site (SRS). The travel time P-wave tomogram and reflection imaging delineate lateral and vertical structural details of the formations. In particular, the high-resolution P-wave tomogram captures a low-velocity zone within the carbonates. This zone is surrounded by reflection events between depths of 150 and 200 ft in the reflection imaging. The reflections are caused by the acoustic impedance contrast between the low velocity zone of "soupy" sand mixtures of unconsolidated materials and the more rigid and dense competent surrounded medium. The time-frequency analysis of full waveforms particle velocity identifies guided waves in form of leaky and normal modes at the depths of about 138 to 150 ft. This resulting change in lithology associated with the presence of guided waves is consistent with a velocity low observed in the vertical velocity profile determined from the inversion of three-component seismic data. This low-velocity zone intercepted by the wells H-BOR-34 and H-BOR-50 correlates with the conductive Griffins Landing Member, which is located above the carbonates. The result of the experiments demonstrate that the present high-resolution crosswell seismic measurement technique (using frequencies up to 1500 Hz) meets the resolution requirements to map geological and geotechnical targets in the vicinity of the In-Tank Precipitation facility at the Savannah River Site.

INTRODUCTION

Adequate characterization techniques are needed to meet the ultimate goal of evaluating the seismic and structural safety of existing and planned facilities at the Savannah River Site (SRS). In particular, methodologies and technologies are needed to satisfy the resolution requirements for specific geologic and geotechnical targets in the vicinity of the In-Tank Precipitation/Extended Sludge Processing Facility and H-Tank Farm areas of SRS.

The geologic setting at the SRS is in unlithified to poorly lithified upper Coastal Plain sediments in South Carolina. The carbonate bodies of interest are stratigraphically confined, typically discrete bodies from 80 to 150 ft depth. The hydrologic confining units in this setting are typically non-fissile clay strata. These units may or may not be faulted. In particular, a technique is needed to resolve the extent and boundaries of the carbonate bodies in addition to the presence of internal voids and high porosity zones.

Crosswell seismic techniques hold promise of investigating the sediment properties (e.g. porosity, permeability, density, etc.) while also depicting its heterogeneous structural features (e.g. boundaries, faults, internal voids, etc.). High-resolution seismic measurements between two or more boreholes may be an efficient tool for resolving the geological structures at SRS. Traveltime tomography inversion application can be used in estimating the compressional wave velocity distribution between boreholes. Direct arrival time data selected from interwell seismic measurements represent only a small portion of the seismic events contained in a seismic section. Reflection events observed in full waveforms may be appropriate for resolving the layered structure in the target zone of interest at the SRS. In fact, interwell seismic reflection data acquired with a piezoceramic source have been analyzed and processed using state-of-the-art mapping techniques to produce high-

resolution of the subsurface geology at the Devine test site, near San Antonio, Texas, as reported by Lazaratos, et al (1994).

Substantial developments have been made recently in seismic sources, receivers and processing techniques (Hardage, 1992; Chen et al., 1990; Kennedy et al., 1988; Laurent et al., 1990; Chen, 1993; Balogh et al., 1988; and Becquey et al., 1992). The geological conditions at the Savannah River Site (SRS) present unique geotechnical problems for the applicability of high-resolution interwell seismic techniques (i.e., crosswell tomography and reverse vertical seismic profiling). Because of the potential use of high-resolution seismic techniques for mapping heterogeneous geological formations, it is anticipated that this technology may be appropriate for the characterization of the subsurface geology for the evaluation of existing and planned facilities. The objective of this work is to evaluate the applicability of a 1200-joule arc discharge borehole seismic source and a piezoceramic source to conduct crosswell experiments and reverse vertical seismic profiling at the SRS and to select the appropriate technology/methodology to image geological and geotechnical features beneath SRS facilities in support of engineering studies. The principal application of these seismic measurements is to investigate the propagation characteristics of seismic waves in a target zone formed by carbonate bodies, clay beds and loosely consolidated sands at the SRS. The processing efforts included the inversion of travel-time data to be used to produce compressional wave velocity, distributions (tomograms) from multiple offset seismic measurements using the arc-discharge source and the piezoceramic source. In addition, reflection imaging from multiple offsets seismic measurements is produced to determine the presence of interfaces or thin beds within the target zone of interest. Furthermore, reverse VSP velocity distributions are produced and integrated with the geology and compared with the reflection imaging to evaluate the borehole seismic technology.

Geological Setting

The in-tank precipitation facility (ITP) is in the Savannah River site, which is located about 20 miles southeast of the fall-line within the Upper Atlantic Coastal Plain of South Carolina and Georgia (see Figure 1). The stratigraphic formations that are encountered in the upper 300 ft of sediments in the ITP area are shown in Figure 2. In ascending order this cross section includes: the Congaree Formation, Santee Formation, Griffins Landing member of the Dry Branch Formation, Dry Branch Formation and the Tobacco Road Formation.

The Congaree Formation is located above the Paleocene strata. The upper Congaree consists of moderately to well-sorted, fine to coarse quartz sands. Thin clay laminae are present in places. According to Snipes et al. (1993) going upward across the Williamsburg/Congaree contact, the sands become cleaner, and clay bed thickness decreases. Above the Congaree formation is the middle Eocene Santee Limestone Formation which is composed of lithofacies. For example, in the siliclastic facies (going upward from the Santee/Congaree contact) the grain size decreases, the green clays become more common, and heavy minerals become abundant. A pebbly zone occurs at the base in places, which is referred as the "green clay" at the SRS. In the southeastern part of the site, Santee carbonates interfinger with laminated calcilutite, calcarenite, and calcareous silt and clay.

The Dry Branch Formation and Tobacco Road Formation constitute the upper Eocene at SRS. Part of the Dry Branch Formation consists of the calcareous Griffins Landing Member, composed of calcilutite, calcarenite, bioclastic and biomoldic limestone, calcareous sand, and shelly and calcareous clay. The upper Dry Branch Formation is made up of the Irwington Sand Member. It is composed of moderately sorted quartz sand, with interlaminated and interbedded clays, typically tan, abundant in places. Pebbly layers and zones rich in clay clasts occur. Irwington sands are

generally coarser than those of the underlying Santee siliciclastics, and glauconite and heavy minerals are less abundant.

The Late Eocene Tobacco Road Formation conformably overlies the Dry Branch Formation. The base of the unit is marked by a coarse layer that in places can contain flat quartz pebbles. The formation consists of moderately to poorly sorted quartz sands. The Tobacco Road Formation crops out over much of the southwestern South Carolina Plain and is widely exposed at the SRS.

The Borehole Layout and the Well Logs

In order to conduct crosswell seismic experiments WSRC drilled five 300 ft deep boreholes which surround the Tank 51 area. The wells penetrate the Tertiary sands and limestones. The location of these wells is given in Figure 3. The well logs available for this project are from wells H-BOR-34, H-BOR-50, H-BOR-44, and HTF-B-1.

The well logs recorded in wells H-BOR-34, H-BOR-50, and H-BOR-44 are correlated with the geologic cross section shown in Figure 2, using the above lithologic descriptions. The resistivity logs between 250-300 ft. given in Figure 4 show that there is an increase in resistivity going upward across the Congaree/Williamsburg contact. At the same time the gamma ray count (shown in Figure 5) decreases, which corresponds to much cleaner sands. This high resistivity anomaly is shown in the three wells. In general, the resistivity of wells H-BOR-44 and H-BOR-50 are very similar, which indicates that the geologic units between those wells are connected. Alternatively, the difference of resistivity signatures with respect to well H-BOR-34 indicates that the Santee Limestone is not intercepted by well H-BOR-34.

In addition, the well logs in the three wells show that the Santee/Congaree contact is associated with a low-resistivity zone, which corresponds to the green clay layer. This thin conductive zone correlates with a high gamma ray count. The Santee Formation is associated with

a large low-resistivity anomaly having higher resistivity than that associated with the green clay layer. A second low-resistivity zone is observed in the three wells above the Santee Formation which is associated with the tan clay thin layer which separates the Dry Branch Formation from the Santee Limestone Formation. In addition, the tan clay layer correlates with a low-velocity anomaly observed in wells H-BOR-44 and H-BOR-50.

In summary, the low-resistivity signatures associated with the tan clay (above the Santee) and the green clay (below the Santee) can be observed in the short normal, long normal and guard resistivity logs recorded in the three wells. In addition the tan clay layer correlates with a low compressional wave velocity observed in well logs recorded in wells H-BOR-44 and H-BOR-50. The presence of this low-velocity layer is not evident in the compressional wave velocity log of well H-BOR-34 (Parra et al., 1995).

The Geological Problem

Underneath the ITP facility there are soft zones associated with the apparent dissolution of calcareous sediments. The purpose of the geophysical measurements is aimed at characterizing the subsurface below Tank 51 based on the following two main issues: (1) The carbonate is not a continuous geologic unit in its lateral extent. In this case the objective is to delineate the edge of the transition zone. Geologists believe the transition zone may be a fairly rapid termination, possibly due to faulting or dissolution. The carbonate exists at about an elevation of 150-200 feet (MSL) and is about 20-50 feet thick. The limestone has a P-wave velocity of about 10,000 ft/s, and the surrounding material has a P-wave velocity of 5000 ft/s; and (2) soft zones or cavities associated with the calcareous sediments are present in the formation. The successful application of crosswell seismic measurements technique will be to demonstrate if the soft zones/cavities in the calcareous sediments can be detected and characterized

FIELD EXPERIMENTS AT THE SAVANNAH RIVER SITE

The main objective of the field experiments was to record crosswell seismic and reverse VSP data simultaneously between a source borehole and at least two detector boreholes. A diagram of the proposed experiments for crosswell tomography and a reverse VSP is shown in Figure 3. Multiple source-detector measurements were conducted using different source and receiver instrumentation. Crosswell measurements between wells H-BOR-34 and well H-BOR-50 were recorded with the TomoSeis piezoceramic source. As a result one crosswell profile was produced in two days of work (12 hour shifts). A second experiment was done by placing a 3-component shuttle sonde (the borehole shuttle is a wall locking sonde containing three geophones) in well H-BOR-50, leaving the arc-discharge source in well H-BOR-34.

The operation started by conducting noise test and spectral analysis of a few traces. After reviewing the data quality recorded with the hydrophones and the 3-component geophones, we decided that a total of eight stacks of the signal will be appropriate for accomplishing our objectives. The 3-component shuttle sonde was placed in well H-BOR-50 at the depth of 200 ft. The control unit to operate the shuttle was installed and operated in the recording truck.

Since the borehole casings were made of metal at the SRS, the magnetic unit that rotates the shuttle before it is clamped was not installed. Alternatively, we have determined the orientation of the particle motion at each source position using a numerical algorithm. The application of the technique is given in the data processing section of this report.

The 3-component data was acquired by firing the source eight times and moving the source every 1/2 m. Once the source was moved 147 times in well H-BOR-34, the 3-component shuttle was moved 10 feet. A total of eight detector positions were recorded once the experiments were completed.

DATA PROCESSING AND ANALYSIS

Processing of 3- Component Interwell Seismic Data

The 3-component detector data recorded using the 1200-Joule arc-discharge source was analyzed for the presence of shear waves. The 3-component data was processed by rotating the seismograms recorded at the Savannah River Site using several processing methods (see Parra, 1995). Also F-K filtering was applied to some of the 3-component seismic data to suppress undesirable events that obstruct the view of more important reflections.

The analysis of the 3-component seismic data suggested that converted and direct shear waves are strongly attenuated by the presence of fluid in the formation. As a consequence, direct shear wave energy may be required for the propagation of shear wave in the formations at the SRS. We believe that a shear source will be more appropriate to use for the generation of direct shear waves. If shear waves propagate in the fluid-filled porous saturated formation a 3-component detector tool coupled to the formation should be used to capture the particle motion of the shear wave polarization in the different directions.

After the rotation of 3-component common detector seismograms we have observed trapped energy (or channel waves) for the detector placed above the Santee Limestone Formation at a depth of 140 feet. The trapped energy (or channel waves) can be used as an indicator of the connectivity or continuity of low-velocity zones. As a consequence the waveforms recorded within these low-velocity zones were analyzed, in particular time-frequency and group velocity contour plots were analyzed to identify the events that may be associated with leaky and normal modes.

Time-Frequency Analysis

The interwell seismic data contains several different events, which can be seen as different arrivals in the full waveforms. These various events, such as P-waves, shear waves, and tube waves,

generally have very different spectral content as well. One way of analyzing signals is to examine how the spectral content of the signal varies in time (within the signal). We do this using time-frequency analysis, which shows the spectrum as a function of time. This technique is particularly useful in detecting guided waves, which exhibit velocity dispersion (velocity is a function of frequency). Such dispersion is apparent in time-frequency plots.

There are many time-frequency methods. The one we have used in our analysis of the Savannah River triaxial data is the spectrogram, which is just the squared magnitude of the short-time Fourier transform. Descriptions of other techniques can be found in the literature (Cohen, 1989; Hlawatsch and Boudreaux-Bartels, 1992; Zook, 1994).

We are using time-frequency analysis to examine the data for evidence of trapped guided waves, indicating continuous low-velocity layers. Locating such channels helps us determine the lithology as well as identify layers of interest (e.g. porous and fluid flow transport layers or impermeable layers), (Parra and Xu, 1994).

The data analyzed in this section is the horizontal particle velocity component, which was recorded for the receiver stationed at a depth of 140 ft, and the source moving at increments of 1.5 ft., in a well H-BOR-34. A portion of this common receiver seismogram is shown in Figure 6. The time-frequency representation of all the traces were produced to identify regions in which time-frequency signatures are similar. Our primary assumption in this case is that these regions corresponds to lithology changes. Since space does not allow us to show time-frequency plots of all the traces, we present below those plots that are representative of three regions of interest.

The first region corresponds to source positions below 150 ft, the second region is for source positions between 150 ft and 138 ft, and the third region is for source positions above 138 ft. For example, when the source is stationed at 160.5 ft, we observed a direct event (P-wave) at about

25 ms in the frequency interval of 100 to 700 Hz (see Figure 7a). We also observed a low frequency event (200 Hz) arriving at about 100 ms. As the source moves up in the borehole, we selected a second trace which was recorded at the source position of 145.5 ft (see Figure 7b). The time frequency representation of this trace shows that the direct event is split in three parts, a low, an intermediate (about 500 Hz) and a high frequency contour. The high-frequency energy observed above 700 Hz corresponds to background noise. Alternatively, the energy captured below 200 Hz (which arrives after 30 ms) corresponds to reflection events and tube waves. In addition, Figure 7b shows two time-frequency contours of 500 Hz at 75 ms and 100 ms. A similar time-frequency signature (given in Figure 7c) is produced when the source is moved at 138 ft. In this case the contour (or an event) at 100 ms is much larger than that shown in the previous Figure 7b.

Next, if we focus in those three contours of about 500 Hz (as is shown in Figure 7d) associated with a source placed at a depth of 129 ft feet, the signatures arriving at 25 ms have merged into one contour, and the other two signatures arriving at 75 ms and 100 ms have changed their characteristics. This analysis suggests that the change in signature (i.e., when the signature split in two major contours), as the source moves upward is associated with a change in lithology that occurs approximately between the depths of 138 ft and 150 ft. This lithology of interest corresponds to a low-velocity zone that trapped energy as the waves propagate between the source and detector boreholes. In fact, the event arriving at 25 ms at the frequency of 500 Hz is a leaky mode and the event arriving at about 100 ms (having the same frequency of 500 Hz) is a shear wave followed by a normal mode. This last event shows dispersion in the range of 500-600 Hz, which proves a low-velocity region has been identified between the depths of 138-150 ft using time-frequency analysis. In addition, a boundary is detected in the logs at 150 ft, which is consistent with the bottom boundary of a velocity low in the vertical velocity distribution observed in Figure 8.

P-wave Tomography and P-wave Reflection Imaging

The field data was edited and the first arrival travel times were picked. These first arrival times were input to a travel time inversion algorithm to produce a P-wave velocity tomogram. The data was initially inverted by constraining the velocity function to one dimension. Additional iterations were performed allowing progressively more freedom by reducing the horizontal bin size to a point where artifacts of the inversion begin to dominate the velocity function. The inversion process was stopped at this point to produce the final P-wave velocity tomogram.

The data prepared for the velocity tomogram was used to generate a reflection image of the interwell area. A velocity model was generated by interactively raytracing and modifying a velocity model until the predicted first arrival times matched the first arrivals of the data within less than 1 ms. Coherent arrivals which do not represent reflection events were filtered from the data (wavefield separation), and the resulting reflection data were VSP-CDP mapped. These mapped data have their reflection energy positioned correctly in the interwell area. The mapped data are then sorted on reflection incidence angle, and stacked over the optimal angle range.

The difficulty of wavefield separation for this profile is the tube wave removal. As we can see from the full wavefield data displays that the P-wave reflection moveouts are very close to the tube wave moveout because the P-wave velocity is close to the tube wave velocity (see Figure 9). This causes some inseparable tube wave residuals left behind when the P-wave reflections are preserved.

INTERPRETATION

Integration of Geophysical, Geological and Well Log Data

Since resistivity and gamma ray logs are valuable information we combined these logs recorded in wells H-BOR-34, H-BOR-50, and H-BOR-44 to estimate the petrophysical boundaries.

The logs shown in Figures 4 and 5 were used to correlate the petrophysical boundaries at the borehole scale with the interwell seismic results. As we previously discussed, the well logs indicated that the Santee/Congaree contact is associated with a low-resistivity anomaly and a high gamma ray count which represents the effect of the green clay layer at a depth of about 210 ft. Furthermore, this geologic unit correlates with a compressional wave velocity low in the P-wave velocity profile (see Figure 8) which was produced from the 3-component seismic data using an algorithm developed by Jackson and Tweeton (1994).

In addition, a second low-resistivity anomaly is observed in the three wells above the Santee Formation at depths of 140 ft to 150 ft. This conductive zone separates the Dry Branch Formation from the Santee carbonates and correlates with a compressional wave velocity low in the P-wave velocity profile. We believe that this low-velocity zone corresponds to a conductive clay layer within the Griffins Landing Member intercepted by the wells H-BOR-34 and H-BOR-50. To determine if this conductive low-velocity target is connected between the wells H-BOR-34 and H-BOR-50, we analyzed the seismic waves recorded at the depths of 140 ft - 150 ft for the presence of trapped waves. Since there is a P-wave velocity contrast between the target of interest and the host medium, trapped seismic energy in the form of leaky modes was detected using time-frequency analysis. In fact, the group velocity plot calculated from waveforms records for the detector at 140 ft. where the source was placed at 140.5 ft exhibits P-wave spectral components which contain low- and high-frequency distributions (see Figure 7b).

In a similar manner we analyzed the data associated with the green clay layer to determine if this layer is connected between wells H-BOR-34 and H-BOR-50. In this case the common detector seismogram for the depth of 200 ft was considered for the analysis. Several waveforms recorded when the source was placed above, below and within the green clay layer were processed using time-

frequency representations. The group velocity contour plots of these waveforms did not show evidence of trapped seismic energy. The time-frequency analysis of the 3-component seismic data recorded at the detector position of 200 ft suggests that the conductive material (green clay) intercepted by H-BOR-34 and H-BOR-50 is not connected between these wells, or that the detector was not placed in the green clay layer. For example, a detector placed in the low-velocity zone at a depth of 205 ft may couple better to the formation than a detector placed at 210 ft to detect trapped seismic energy associated with the green clay layer. Indeed, to test this concept in more detail it will require to take more 3-component measurements at several detector positions above and below and within the green clay layer to thoroughly determine if guided waves can be detected in this zone.

An additional interpretation was made by comparing the velocity profiles given in Figure 8 with the resistivity and gamma ray logs given in Figures 4 and 5. The maximum resistivity anomaly observed between the depths of 150 ft to 180 ft in well log H-BOR-50 correlated with a local maximum P-wave velocity anomaly and a low-gamma count. In this zone, the natural gamma count is low and the resistivity and the P-wave velocity are high. These petrophysical characteristics are typical of those of carbonate rocks. In addition, the resistivity log of well H-BOR-34 shows a small resistivity anomaly which may correspond to a heterogeneity not associated with carbonate rock.

The top and bottom of the carbonate rock at a depth of about 170 ft correlates with reflections observed in the reflection image plots in Figure 10. This reflection image shows that the carbonate has been intercepted by sands and other lateral changes of the region of interest. A weak reflection is observed at the depth of 145 ft which correlates with the tan clay unit. In addition, in the upper portion of the reflection image, two reflections are observed. The first reflection is at a depth of about 55 ft, which is associated with the boundary between the Tobacco Road Formation

and the Upper Dry Branch Formation. The second reflection is observed at a depth of about 85 ft, and it correlates with the Lower Dry Branch Formation.

Integration of Travel Time Tomograms and Reflection Imaging

Travel times acquired from the H-BOR-34 and H-BOR-50 survey were inverted to produce a velocity tomogram. This tomogram has delineated the lateral and vertical heterogeneous conditions of the formation between the depths of 140 and 200 ft. In particular, the lateral contact between the carbonate (intercepted by well H-BOR-50) and the fully saturated sediments (intercepted by well H-BOR-34) is captured in the high-resolution travelttime tomogram given in Figure 11. These results are similar to those determined by Majer et al. (1995). Majer also inverted the seismic amplitudes recorded between wells H-BOR-50 and H-BOR-34 to produce an attenuation tomogram.

The carbonate is a higher velocity unit and the water saturated sands are much lower velocity zones. In addition, the carbonates are associated with a low quality factor and the water saturated sands are associated with a higher quality factor (see Majer et al. 1995). Since the heterogeneous carbonate unit is porous and vuggy the fluid will flow in and out of the porous when the waves propagate through the carbonate matrix. Under this conditions energy loss will occur by reducing the wave amplitude as the wave travels between wells. Alternatively, the low velocity zone formed by fully saturated sands will not allow the fluid to flow in and out of a more uniform sand matrix, as a consequence no energy loss will occur by the presence of the fluids in the sands. Indeed energy losses may occur by the presence of viscoelasticity in the sands or any other material present in the formation. In the present application we expect that the viscoelastic losses will be much less than those associated with fluid flow.

The integration of the velocity tomogram and the P-wave reflection imaging has yielded a final characterization of the subsurface in the region under study (see Figures 10 and 11). In these plots

the soft sands delineated by the velocity tomogram as a blue zone (between 22 to 140 ft in the horizontal direction) are surrounded by reflection events observed between depths of 150 ft and 200 ft in the reflection imaging. These reflection events were originated by the acoustic impedance contrast between the low velocity zones of "soupy" sand mixtures (unconsolidated materials) and the more rigid and dense (competent) surrounded medium .

CONCLUSIONS

The integration of well logs and the 3-component seismic data as well as the reflection image delineated a clay layer within the Griffins Landing Member, and the carbonate rock units. The rotation of the 3-component data indicated the presence of trapped seismic energy associated with the connectivity of the tan clay layer between wells H-BOR-34 and H-BOR-50. On the other hand, the results of the interpretation suggested that there is not sufficient data to prove whether the green clay unit is connected or not.

In general, the Santee sediments are associated with a large velocity low relative to the Dry Branch (above) and Congaree (below) high-velocity formations. In several, this low-velocity anomaly a P-wave velocity which correlates with the resistivity logs. A local P-wave velocity and resistivity maxima correlates with the carbonate rocks which are associated with low-gamma ray counts. In addition, the large velocity anomaly is consistent with the compressional wave velocity logs given in Figures 4 and 5. Compressional wave velocities less than the compressional wave velocity of the water (in shallow geophysical applications) is not unusual in low shear-wave velocity formations having bulk densities greater than the density of the water.

The processing and analysis of the seismic data suggested that additional 3-component seismic measurements should be conducted at shorter spacing to resolve the heterogeneities within the Santee

Limestone formation, and to determine whether the green clay unit is continuous between wells, and to predict other lithologic features of interest.

The combined P-wave reflection imaging and the travel time velocity tomogram has delineated a large soft zone of sand mixtures (in the inter-well area between wells H-BOR-50 and H-BOR-34) at a depth of 150-200 ft. The length of the soft zone is 120 ft between wells H-BOR-50 and H-BOR-34 , and the extension of this zone beyond well H-BOR-34 is unknown.

ACKNOWLEDGMENTS

This work was supported by the Westinghouse Savannah River Company, Subcontract number AB6239N. The conclusions and opinions of the authors do not necessarily reflect those of the United States Department of Energy.

REFERENCES

- Balogh, W.T., Owen, T.E., and Harris, J.M. (1988). "A new piezoelectric transducer for hole-to-hole seismic applications," Expanded Abstracts, Paper DEV2.5 58th Ann. Int'l. Meeting SEG; Anaheim, CA; Oct. 30-Nov. 3.
- Barr, F.J., Beasley, T.R., and Piggin, R.H. (1989). "Method and apparatus for seismic exploration of strata surrounding a borehole," U.S. Patent No. 4,873,675.
- Becquey, M., Bermet-Rollande, J.O., Laurent, J., and Noual, G. (1992). "Imaging reservoirs -- A cross-well seismic experiment, First Break (EAEG), Vol. 10, pp. 337-344.
- Chen, S.T., (1993). "A Single-well profiling tool and tube wave suppression system", Expanded Abstracts, Paper BG1.4, 63rd Ann. Int'l Meeting SEG, Washington, DC.
- Chen, S.T., Zimmerman, L.J., and Tugnait, J.K. (1990b). "Subsurface imaging using reverse VSP and cross-hole tomographic methods, "Geophys., Vol. 55, pp. 1478-87.
- Cohen, L., (1989). "Time-Frequency Distributions - A Review," *Proc. IEEE*, Vol. 77, No. 7, 941-981.
- Hardage, B.A. (1992). Cross-Well Seismology and Reverse VSP, Geophysical Press, Ltd. London.
- Hlawatsch, F., and Boudreaux-Bartels, G.F., (1992). "Linear and Quadratic Time-Frequency Signal Representations," *IEEE Signal Processing Mag.*, Vol. 9, 21-67.
- Jackson, M.J., and Tweeton, D.R., (1994). "MIGRATOM-Geophysical Tomography using Wavefront Migration and Fuzzy Constraints," *Report of Investigations #9497*, Bureau of Mines, United States Department of Interior.
- Journel, A.G. (1989), "Fundamentals of geostatistics in five lessons," American Physical Union, Washington, DC, Vol. 8.

- Kennedy, W.S., Wiggins, W., and Aronstam, P. (1988). "Swept-frequency borehole source for inverse VSP and cross-borehole surveying," Expanded Abstracts, Paper DEV2.6; 58th Ann. Int'l. Meeting SEG; Anaheim, CA, Oct. 30-Nov. 3..
- Laurent, J., Layotte, P.C., Meyneir, P. and Noual, G. (1990). "A mechanical wall-clamped borehole source and its use in inverse VSP and cross-hole seismic surveying," Abstracts, Paper 13-19; 52nd Meeting EAEG; Copenhagen, Denmark; 28 May - 1 June.
- Lazaratos, S.K., Langan, R., and Harris, J.M. (1994). "Shear-wave crosswell reflection imaging in West Texas," Expanded Abstracts, Paper DP3.2, Ann. Int'l Meeting SEG, Los Angeles, CA, October 23-27.
- Majer, E., Peterson, J., Hubbard, S., Daley, T., and Vasco, D., (1995). "Evaluation of the applicability of high resolution crosshole seismic imaging beneath the H-Tank area for geomechanical properties" Final Report, Lawrence Berkeley Laboratory, Berkeley, California.
- Majer, E., Peterson, J., Hubbard, S., Daley, T., and Vasco, D., (1995). "Evaluation of high resolution crosshole seismic imaging beneath the H-tank area for geomechanical properties," Final Report, Lawrence Berkeley National Laboratory, Berkeley, California.
- Mosher, C.C. and Mason, I.M., (1983). "Borehole-to-borehole seam-wave transmission," Expanded Abstracts, Paper C1.4, 53rd Ann. Int'l Meeting SEG, Las Vegas, NV, Sept. 11-15.
- Owen, T.E. and Shirley, D.J. (1985). "Cylindrical bender-type vibration transducer," U.S. Patent No. 4,525,645.
- Parra, J.O., and Xu, P-C., (1994), "Dispersion and attenuation of acoustic guided waves in layered fluid-filled porous media," *J. Acoust. Soc. Am.*, 95, 91-98.
- Parra, J.O., (1995). "A feasibility study of high-resolution reverse VSP and interwell seismic methods for hydrocarbon reservoir characterization," *The Log Analyst*, May-June, 64-81.

- Raytheon Engineers and Constructors, (1994). "Geophysical investigations, in-tank precipitation facility, Task 176", Final Report, Westinghouse Savannah River Company Project Engineering Service Contract.
- Rowe, P. and Mason, I.M., (1982). "Downhole recharged electrolytic sleeve exploder," Expanded Abstracts, Paper S14.8, 52nd Ann. Int'l Meeting SEG, Dallas, TX, Oct. 17-21.
- Staron, P., Arens, G., and Gros, P. (1988). Method of instantaneous acoustic logging within a borehole," U.S. Patent No. 4,718,048.
- Snipes, D.S., Fallaw, W.C., Price, V., and Cumbest, R.S., (1993). The pen branch fault: Documentation of late cretaceous-tertiary faulting in the coastal plain of South Carolina: Southeastern geology, 33, 195-218.
- Zook, B.J., (1994). "Using Guided Waves to Characterize Gas Reservoir Continuity, Phase I: Data Processing," *Topical Report #GRI-94/0384.2*, Gas Research Institute.

LIST OF CAPTIONS

- Figure 1. The upper Atlantic coastal plain of South Carolina and Georgia.
- Figure 2. A simplified cross-section of the In-Tank Precipitation site geology at the Savannah River Site.
- Figure 3. A plan view of the five wells at the In-Tank Precipitation site.
- Figure 4. Display of resistivity logs for wells H-BOR-34, H-BOR-50, and H-BOR-44.
- Figure 5. Display of gamma ray logs for wells H-BOR-34, H-BOR-50, and H-BOR-44.
- Figure 6. A common receiver particle velocity seismogram for a detector depth of 140 ft. The seismic data was produced using a 1200 Joule arc discharge source and a shuttle 3-component detector.
- Figure 7a. Time-frequency spectrogram of a trace recorded at the source depth of 160.5 ft, given in the seismogram of Figure 6.
- Figure 7b. Time-frequency spectrogram of a trace recorded at the source depth of 145.5 ft, given in the seismogram of Figure 6.
- Figure 7c. Time-frequency spectrogram of a trace recorded at the source depth of 138 ft, given in the seismogram of Figure 6.
- Figure 7d. Time-frequency spectrogram of a trace recorded at the source depth of 129 ft, given in the seismogram of Figure 6.
- Figure 8. Compressional wave velocity profile derived from three-component detector data.
- Figure 9. A common receiver pressure seismogram for a hydrophone depth of 205 ft. The seismic data was produced using a piezoelectric source and an array of hydrophones.
- Figure 10. Comparison of down going reflection image with resistivity and gamma ray logs from wells H-BOR-50 and H-BOR-34.
- Figure 11. P-wave velocity tomogram showing a low-velocity zone formed by soft materials.

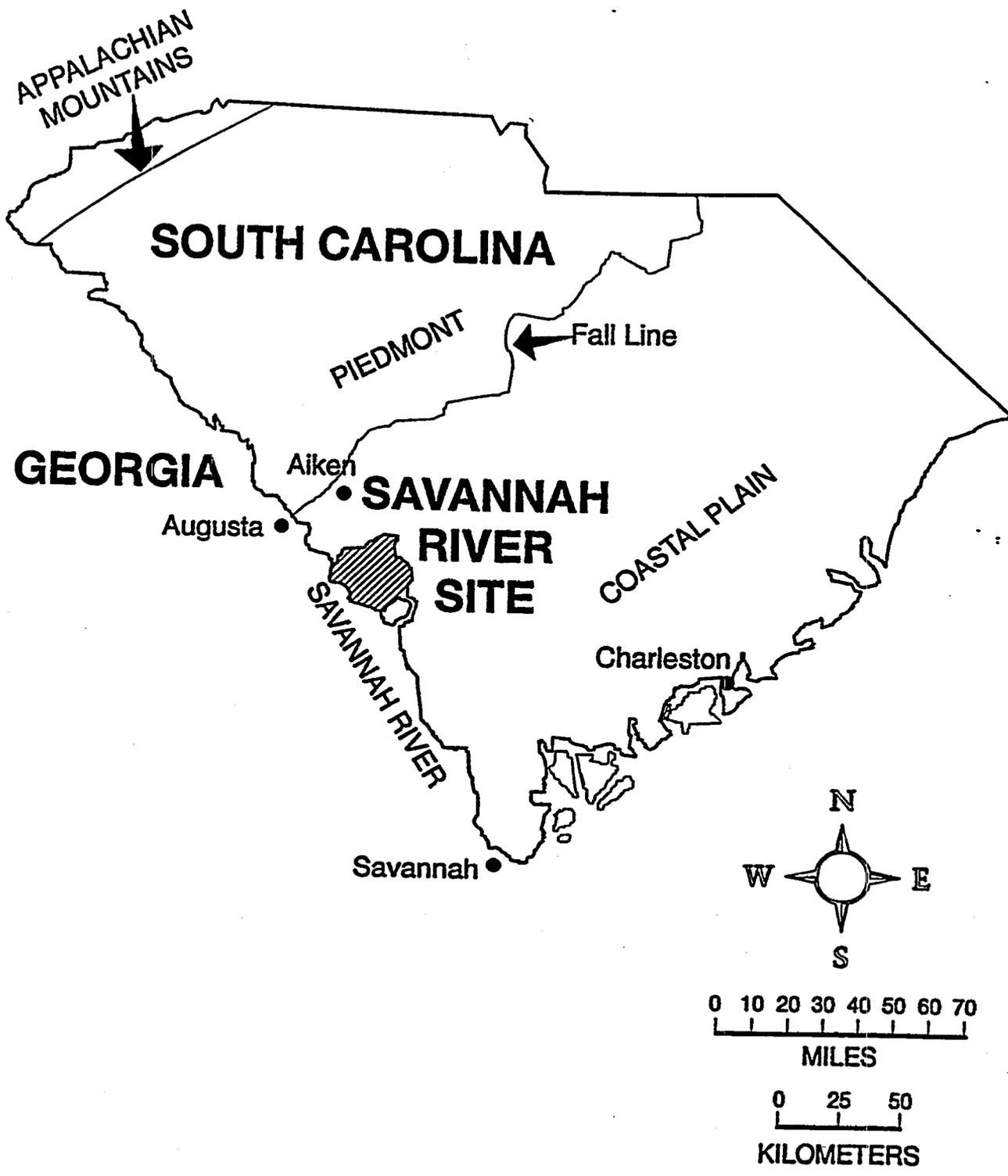


Figure 1

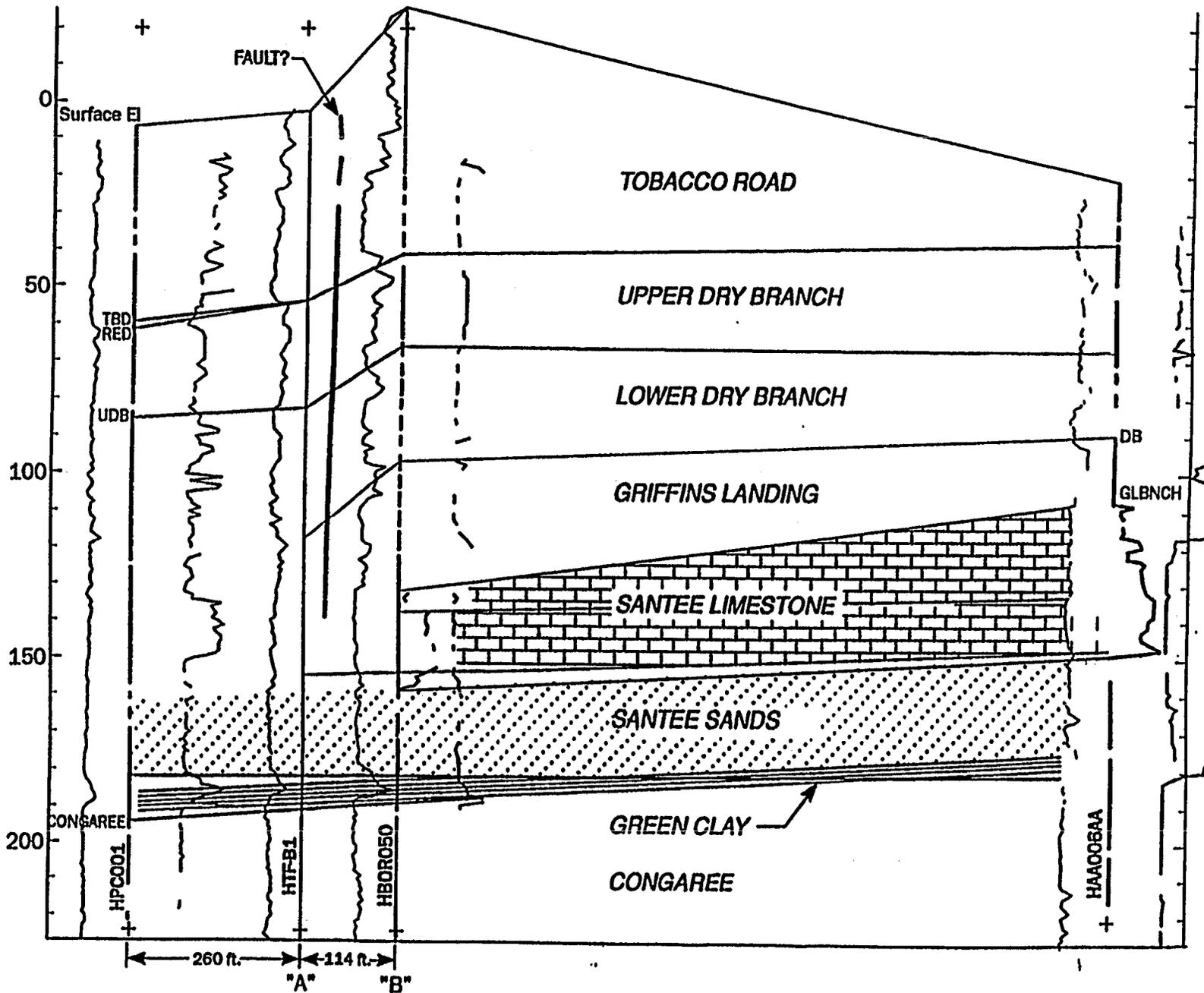


Figure 2

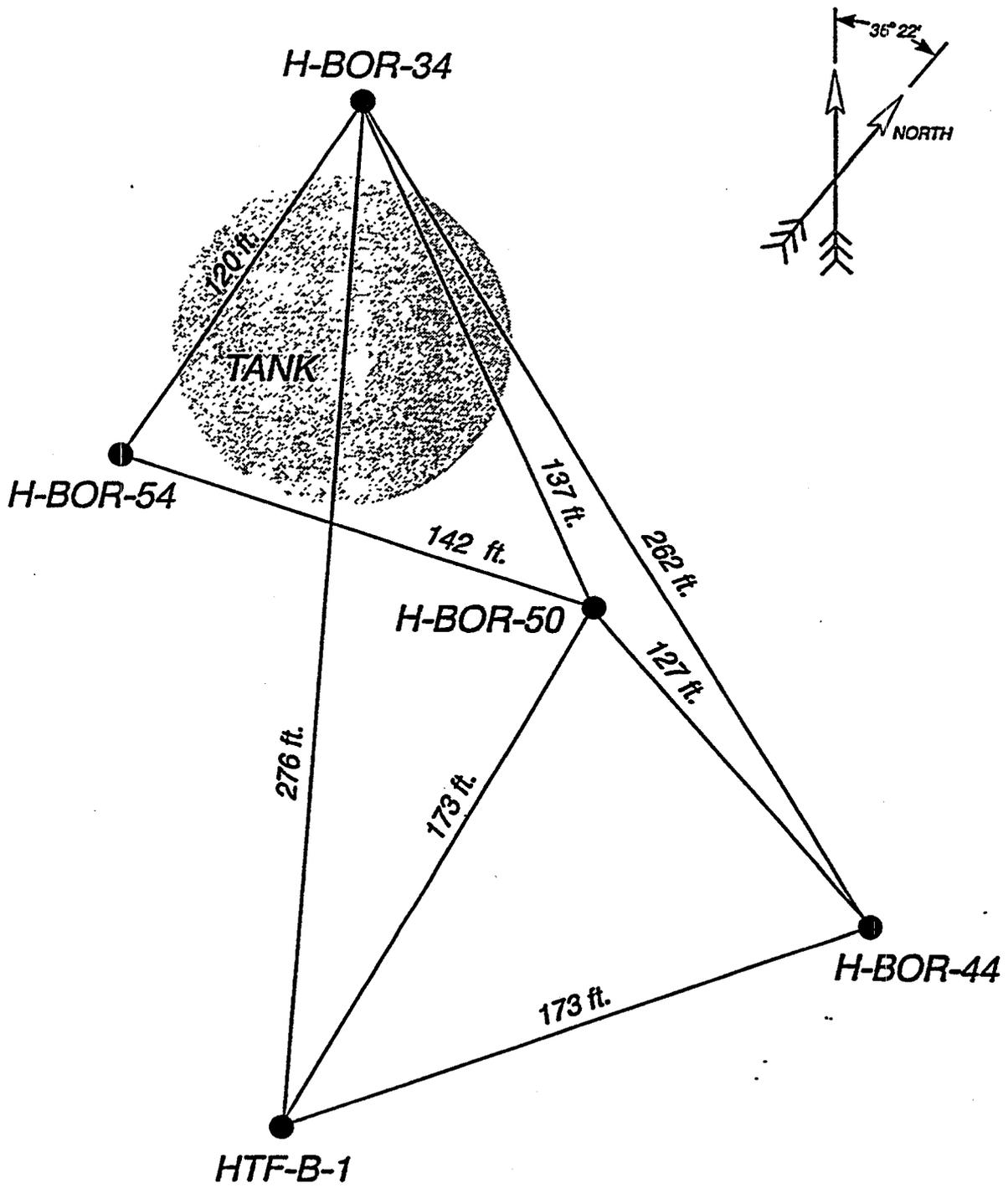


Figure 3

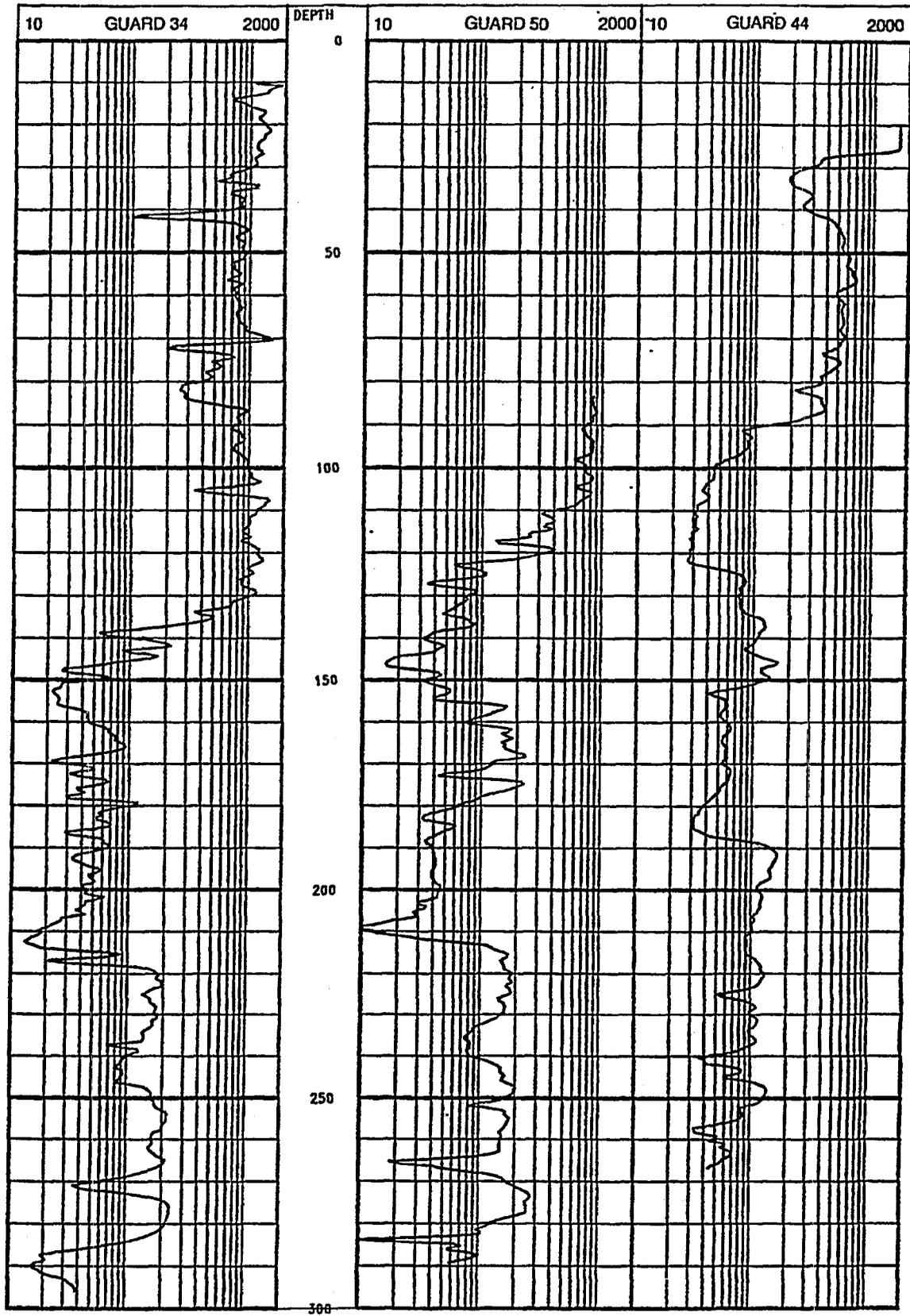


Figure 4

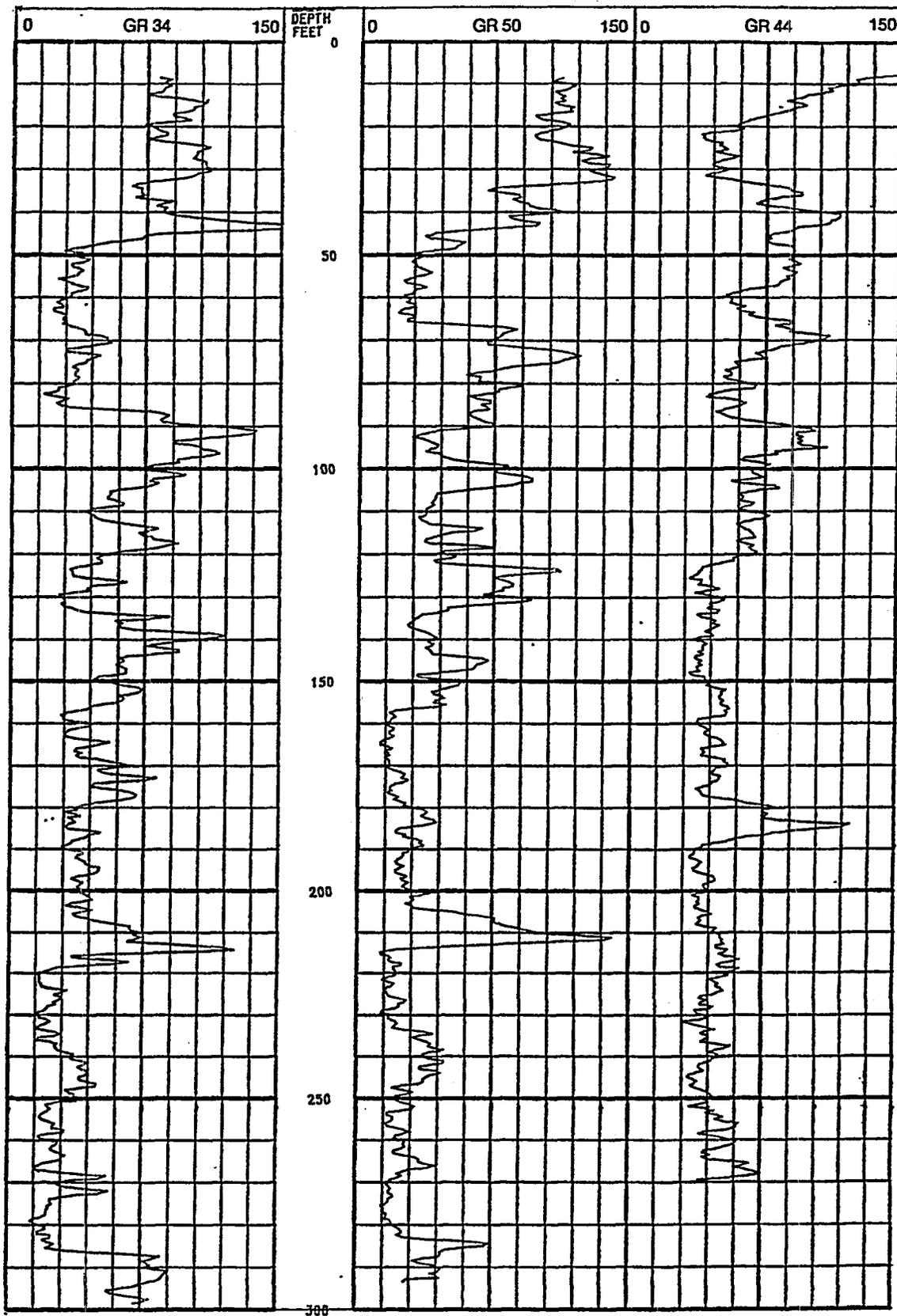


Figure 5

Savannah River Data

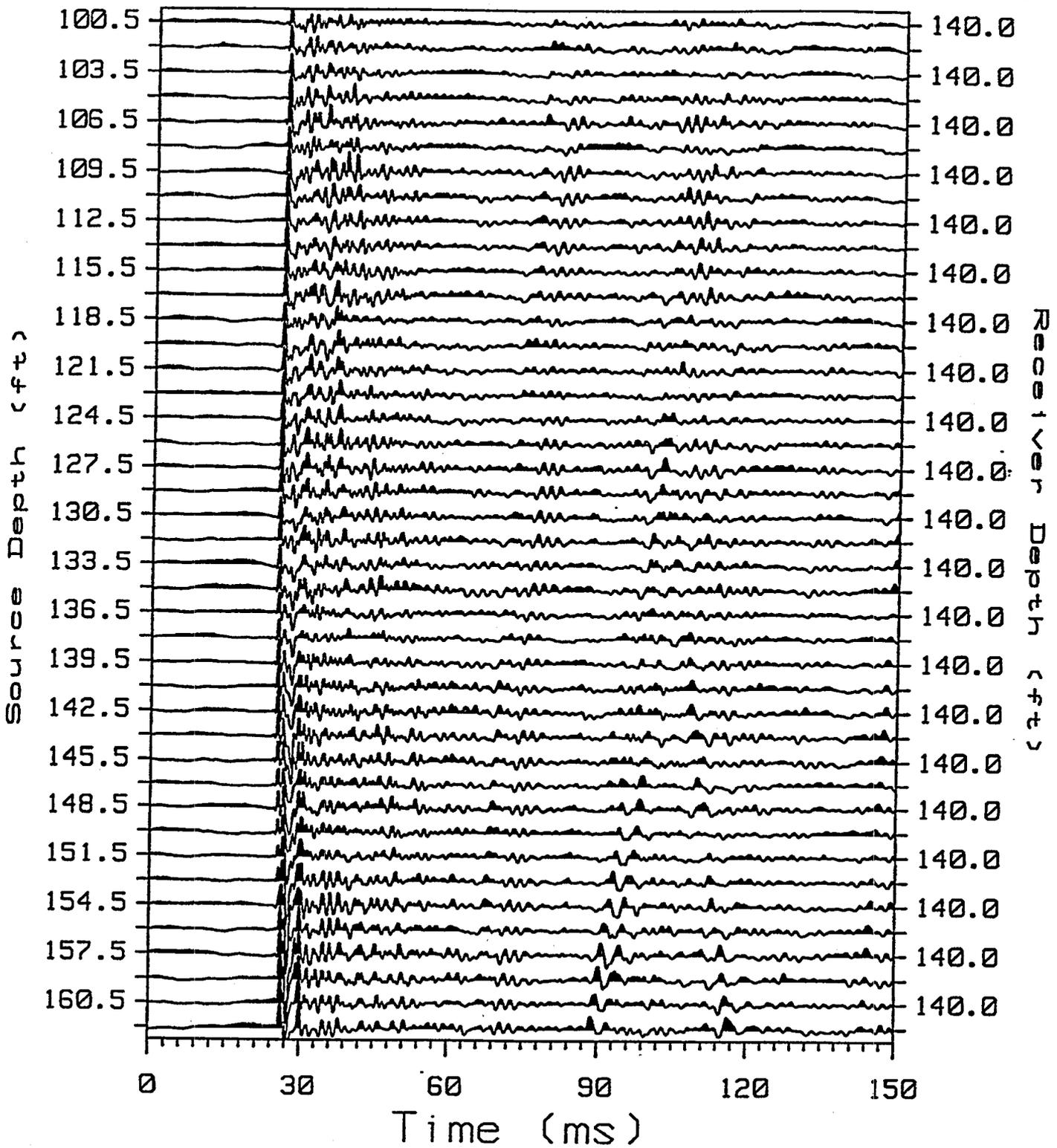


Figure 6

Fan #7-X, 160.5 ft (SPEC, 1.0E-3)

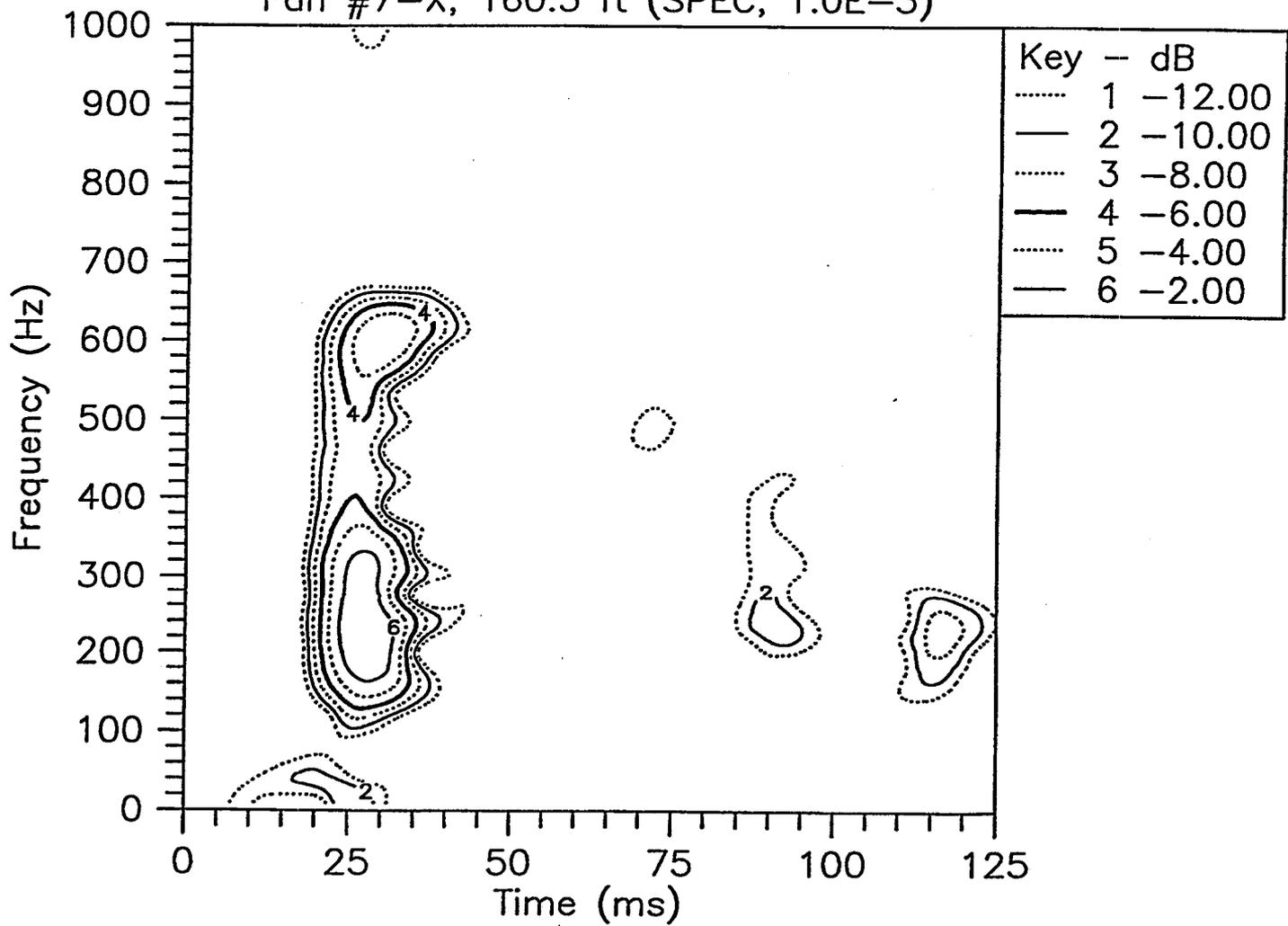


Figure 7a

Fan #7-X, 145.5 ft (SPEC, 1.0E-3)

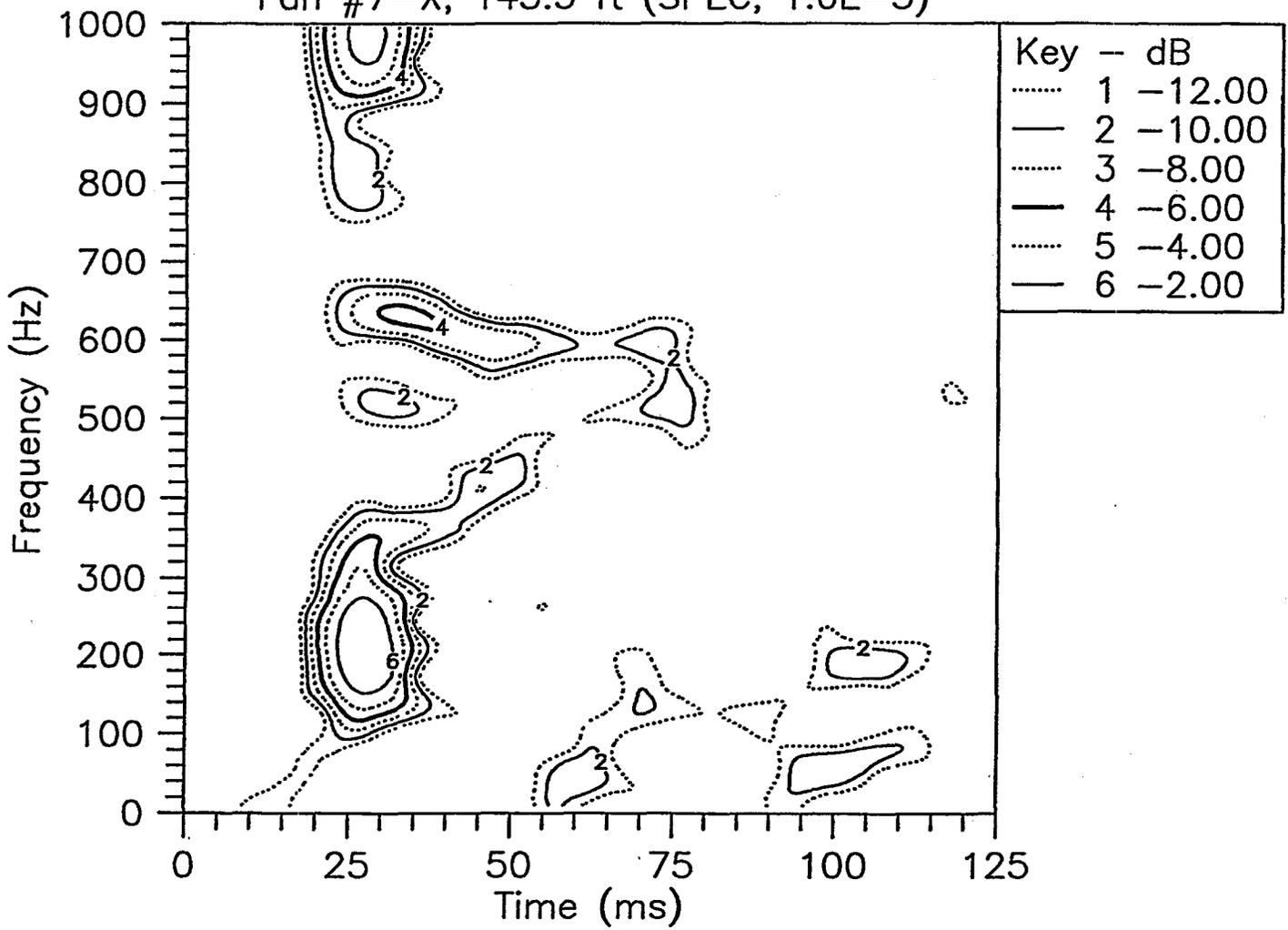


Figure 7b

Fan #7-X, 138.0 ft (SPEC, 1.0E-3)

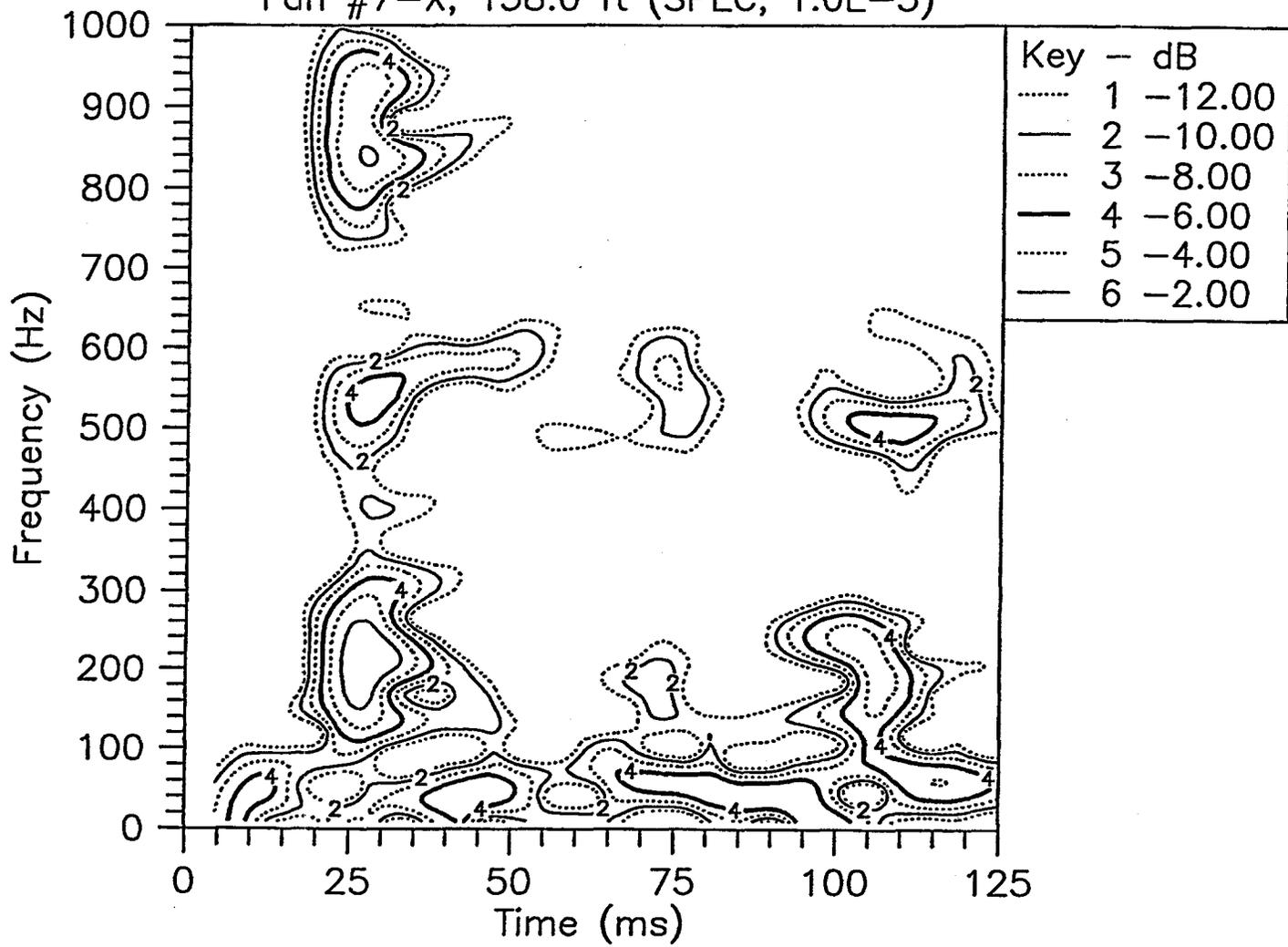


Figure 7c

Fan #7-X, 129.0 ft (SPEC, 1.0E-3)

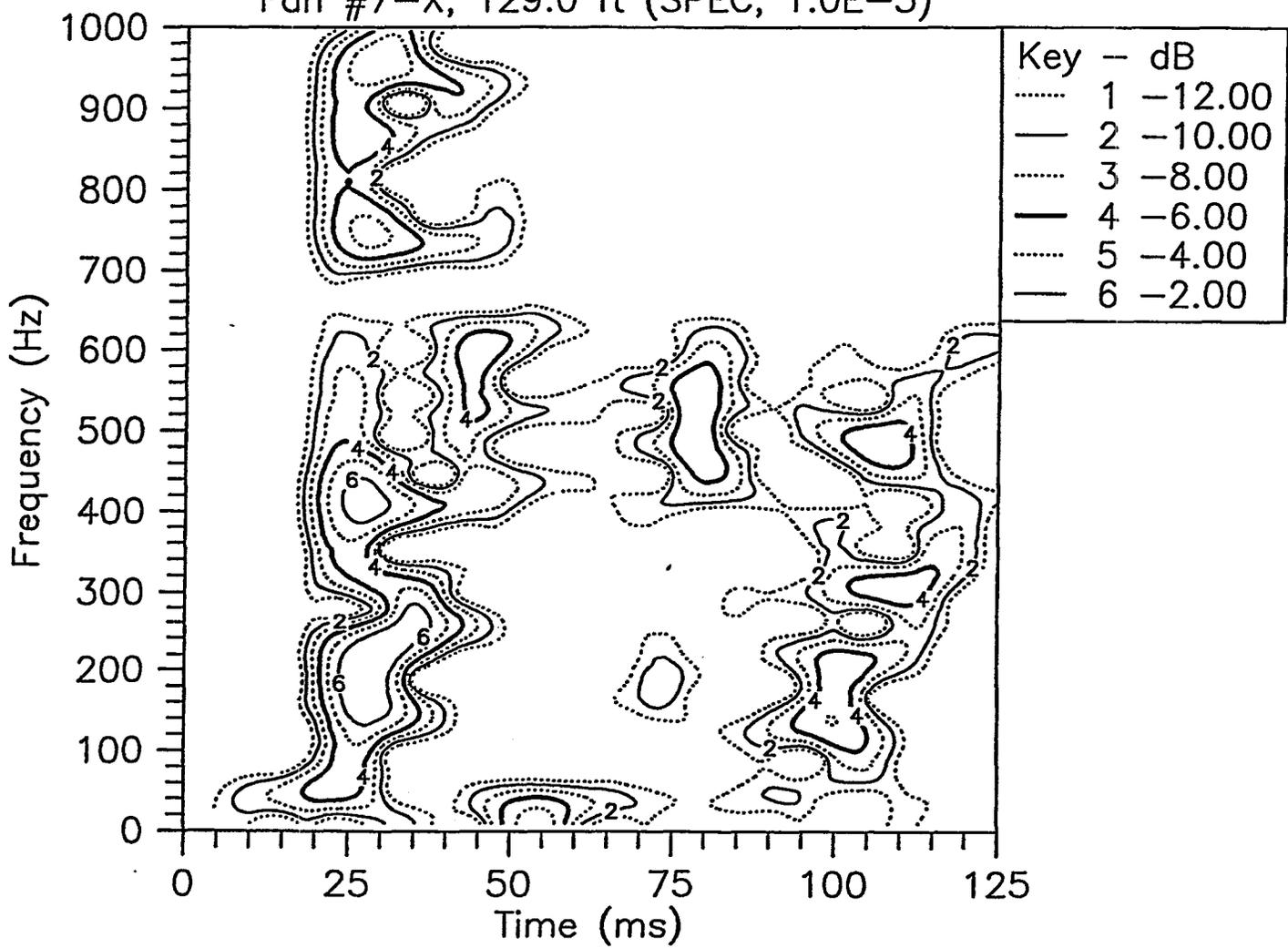
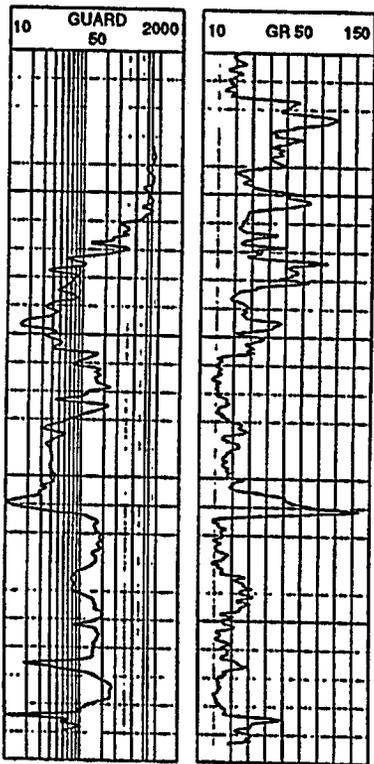


Figure 7d



Savannah River Triaxial Data
Layered Tomography - Fans #1,3,6,7

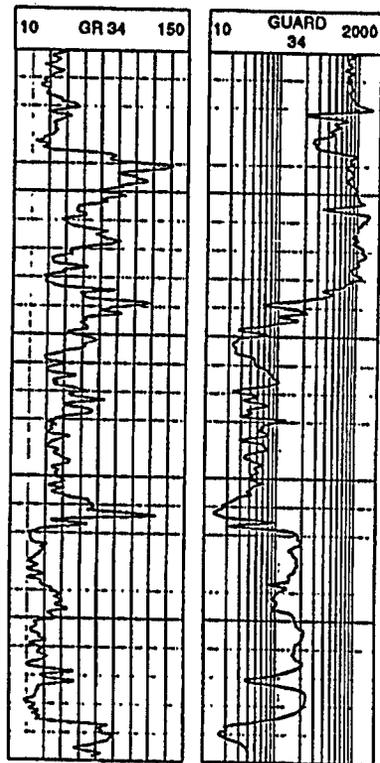
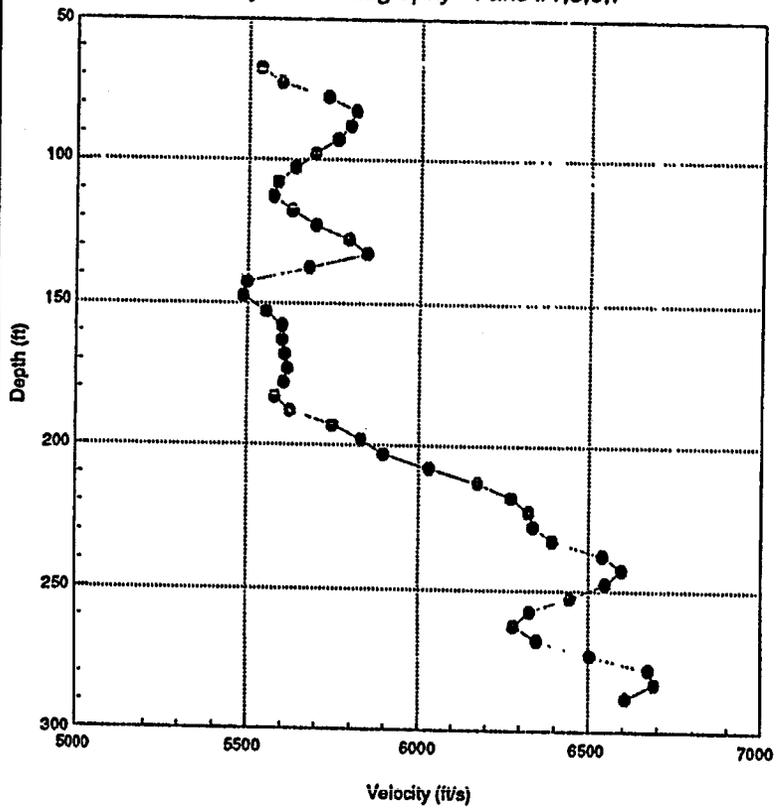


Figure 8

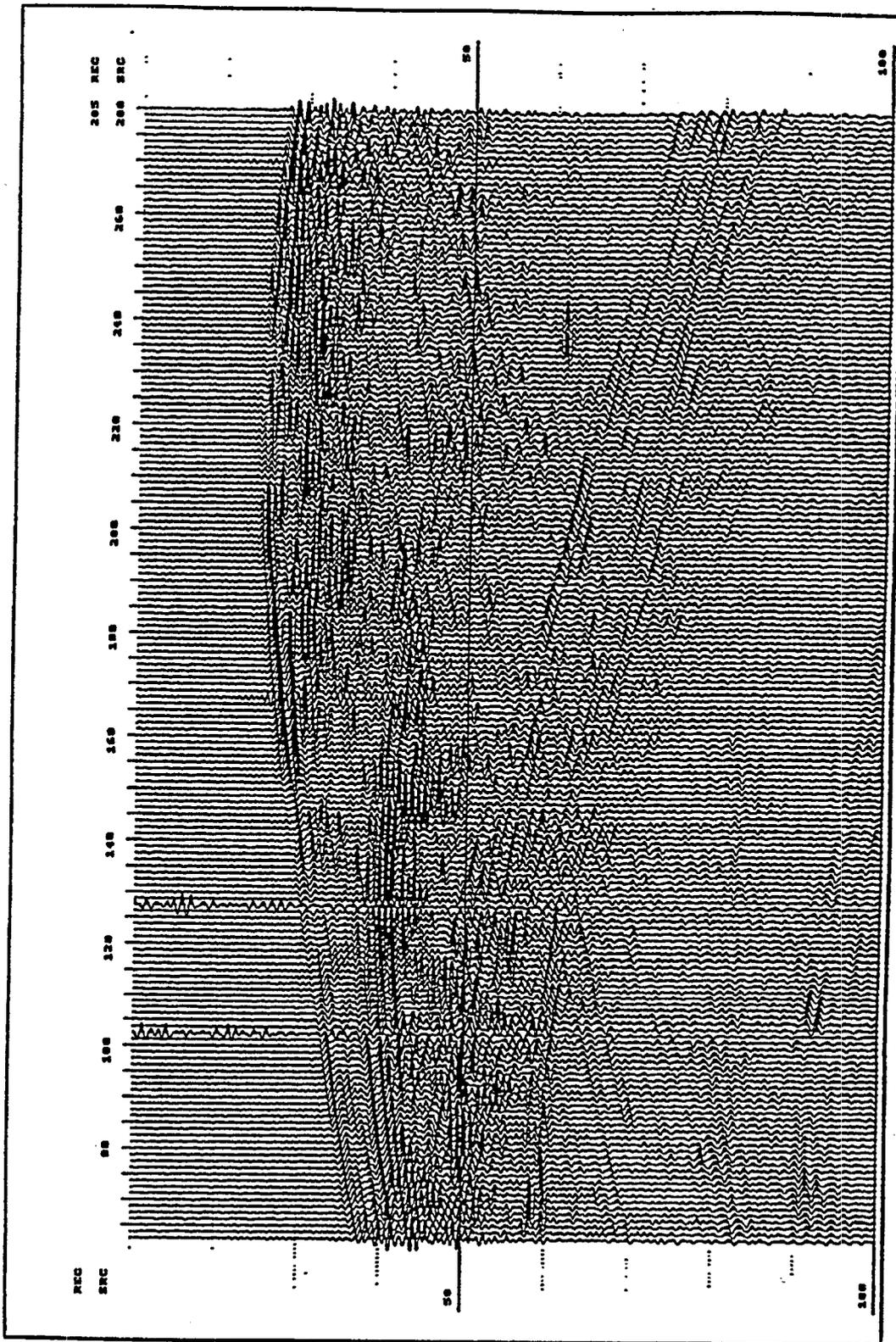


Figure 9

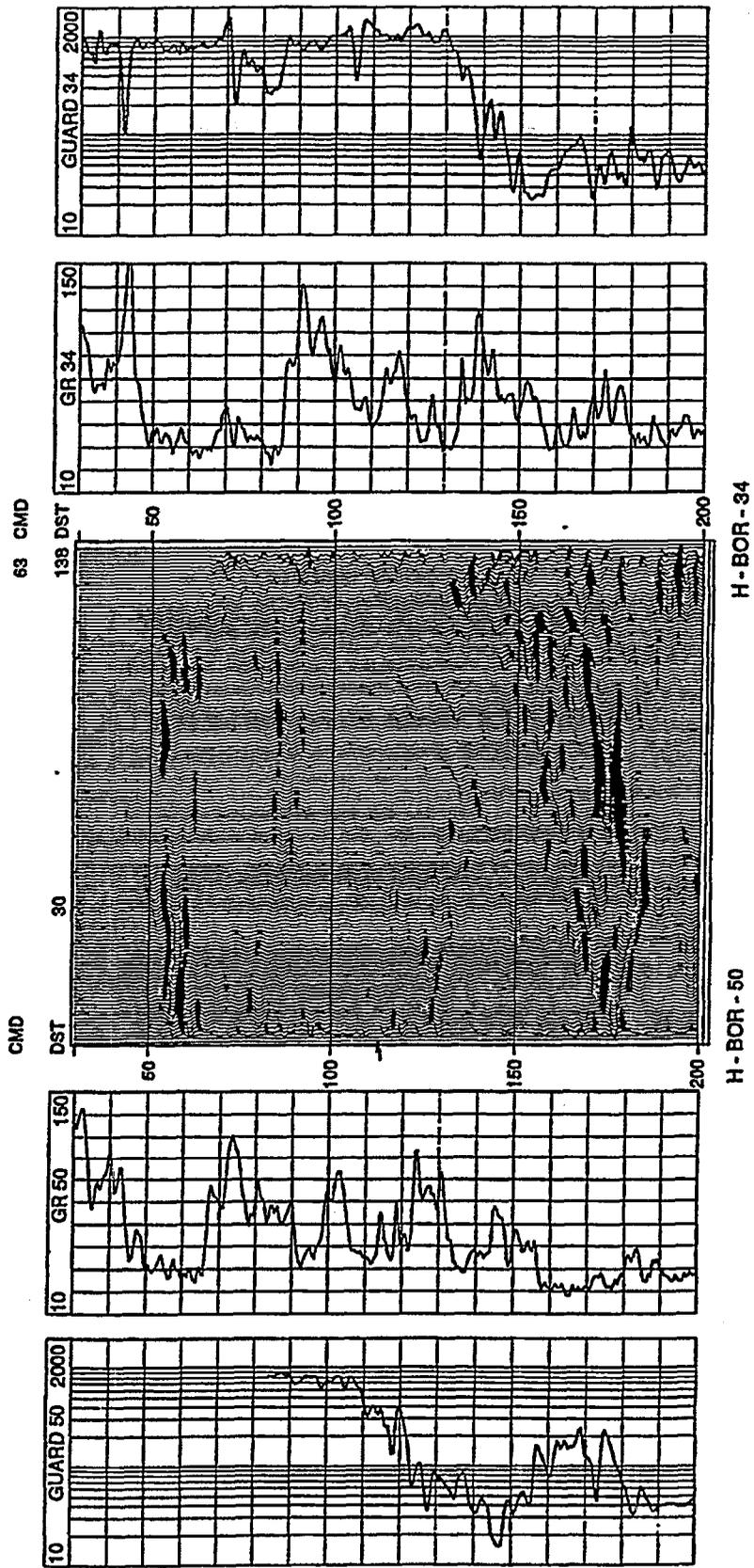


Figure 10

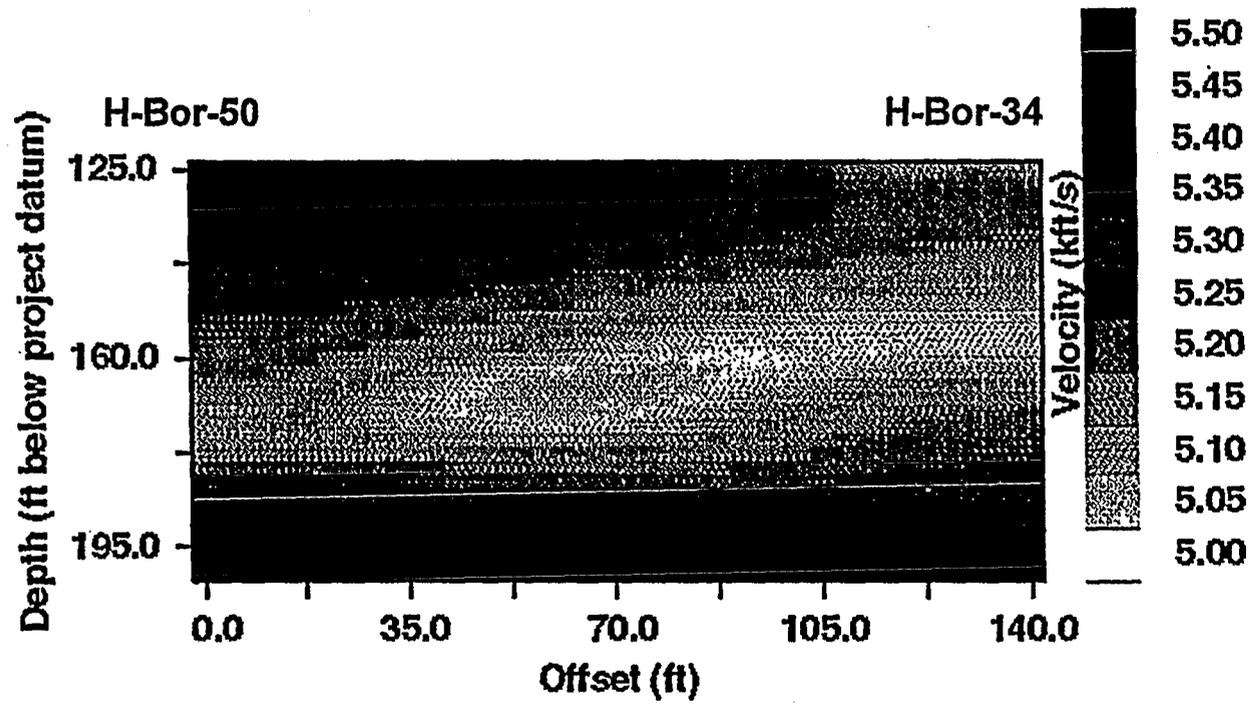
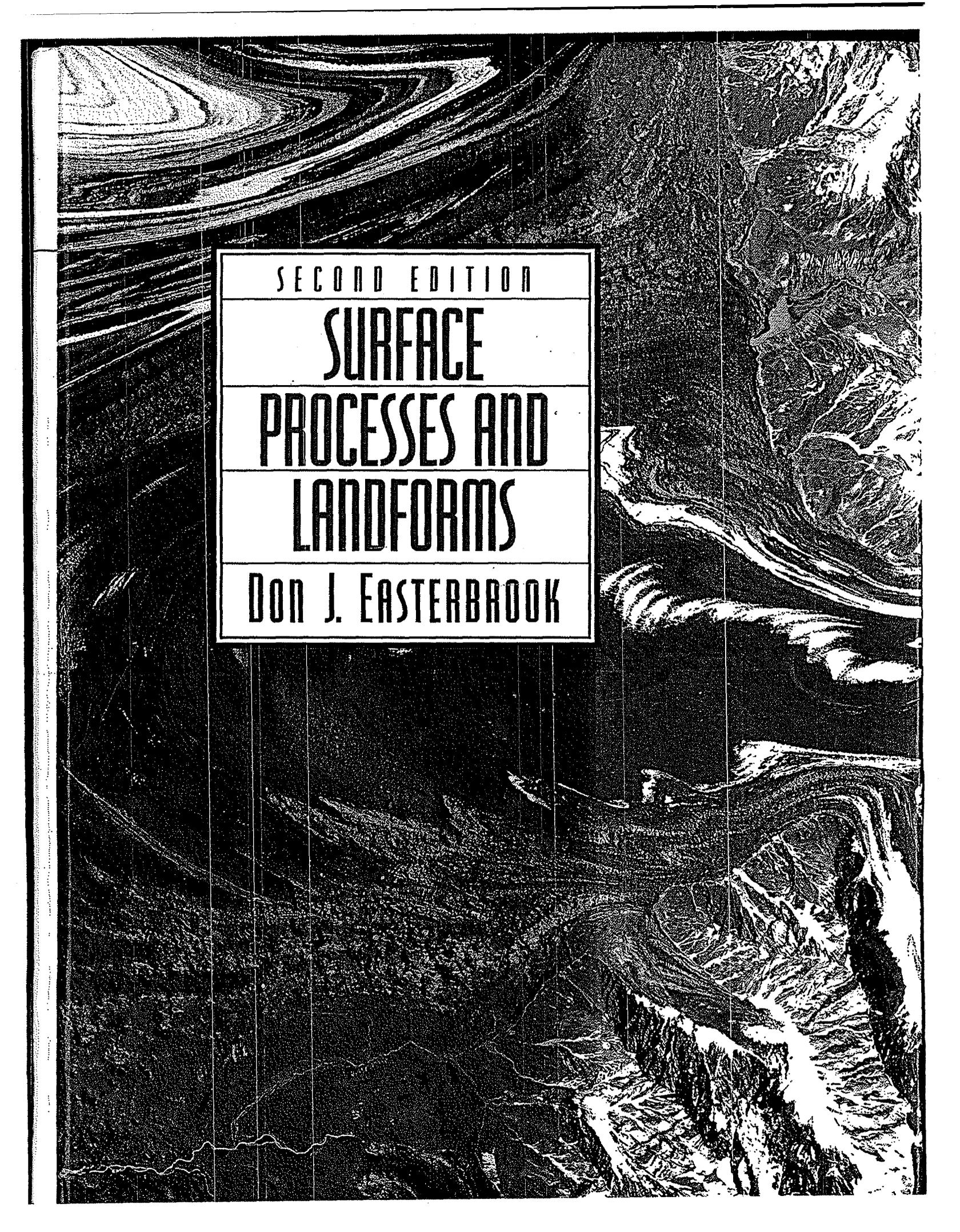


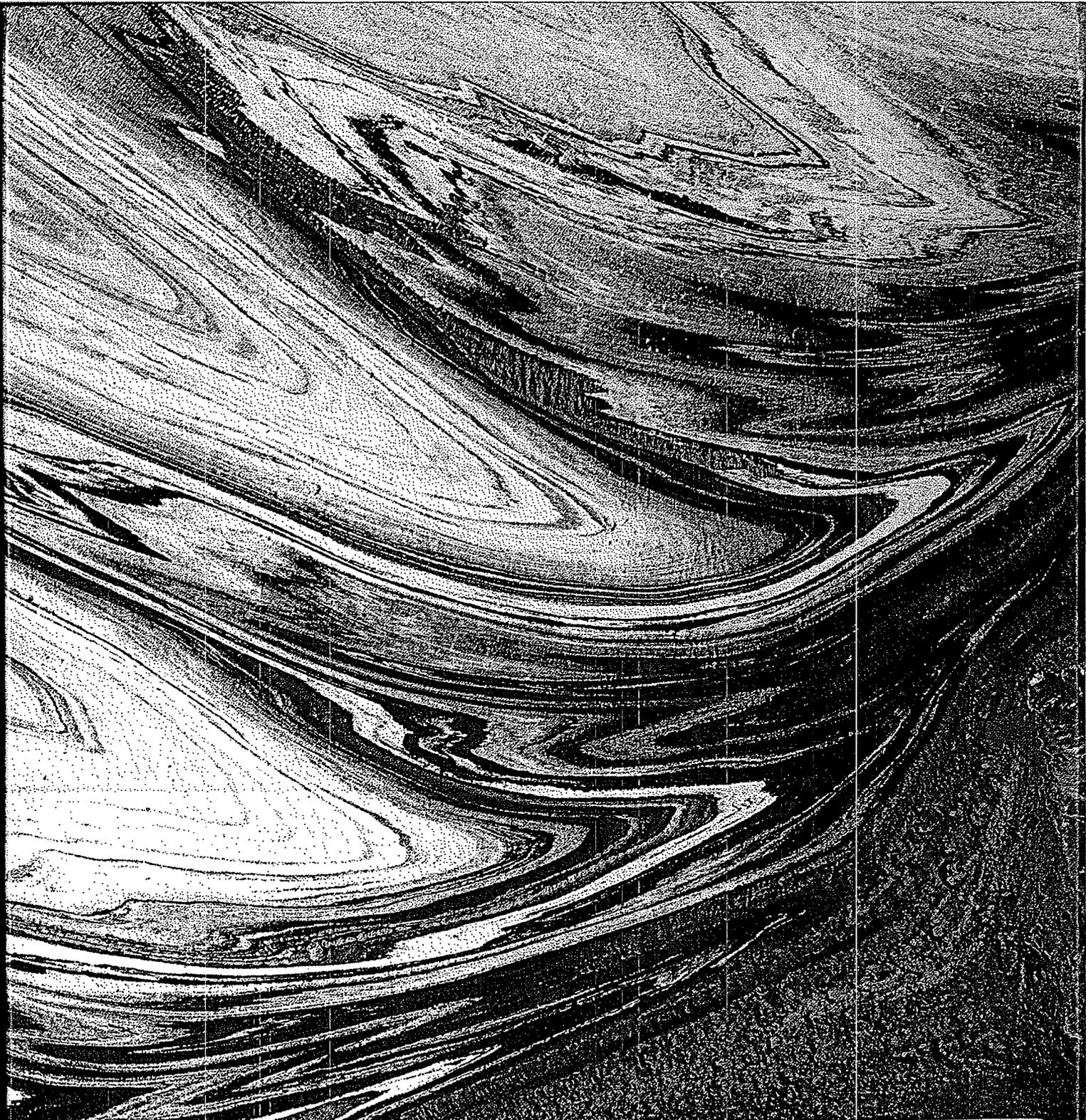
Figure 11



SECOND EDITION

**SURFACE
PROCESSES AND
LANDFORMS**

DON J. EASTERBROOK



PRENTICE HALL
Upper Saddle River, NJ 07458
<http://www.prenhall.com>

EASTERBROOK **USED**

SURFACE PROCESSES+LANDFORMS,
2ND '99 PH
978-0-13-860958-0



010003696801



2 980138 609589

39.8



Surface Processes and Landforms

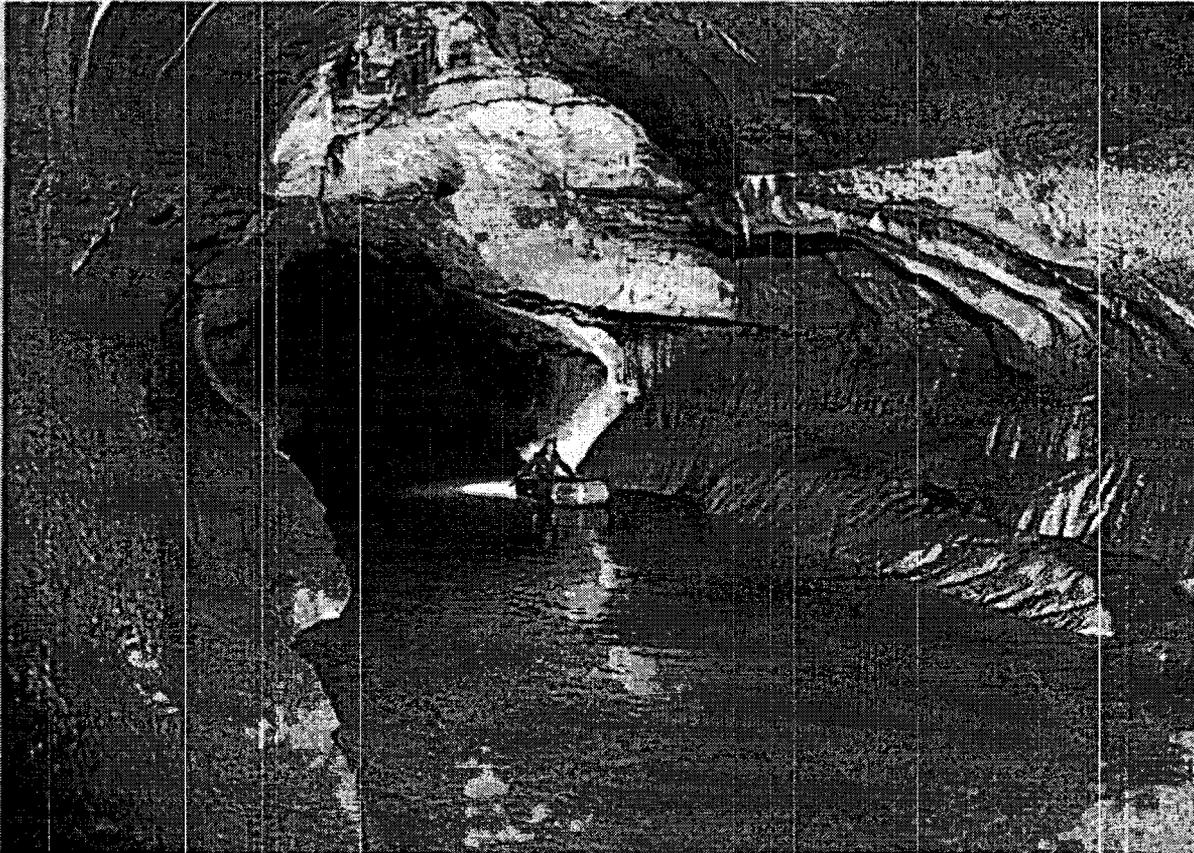
Second Edition

DON J. EASTERBROOK



Prentice Hall, Upper Saddle River, New Jersey 07458

Groundwater



Subterranean stream in a limestone cave, Krizna Jama, Slovenia. (Photo by A. N. Palmer)

IN THIS CHAPTER

Introduction

Porosity and Permeability

Porosity

Porosity in Unconsolidated Sediments

Porosity of Rock

Permeability

Specific Yield and Specific Retention

Hydraulic Conductivity

Permeability of Sediments

Permeability of Rocks

The Water Table

Aquifers

Unconfined Aquifers

Confined Aquifers

Thermal Springs and Geysers

Ground Subsidence Due to Withdrawal of Fluids

Karst

Karst Controls

Lithology and Structure

The Solution Process

Effect of Climate and Vegetation on Solution

Karst Groundwater Hydrology

Karst Landforms

Sinkholes

Uvalas

Poljes

Sinkhole Ponds and Karst Lakes

Solution Chimneys and Vertical Shafts

Solution along Fractures

Transition between Fluvial and Karst Drainage

Karst Valleys

Caves

Karst Windows

Origin of Karst Cave Systems

Tropical Karst

References

INTRODUCTION

Although groundwater lies beneath the Earth's surface where it is not easily observable, it constitutes an important part of the hydrological system. In addition to feeding surface streams, groundwater is also capable of widespread solution of carbonate rocks.

We will first look at how subsurface water moves through various types of material, then consider how dissolving of subsurface carbonate rocks develops karst topography.

POROSITY AND PERMEABILITY

Precipitation that is not intercepted or does not run off on the surface infiltrates downward into the subsurface through pore spaces between mineral grains, cracks, solution cavities, or vesicles. The physical properties of subsurface material that control the volume and movement of groundwater are porosity—the percentage of the total volume of the rock consisting of voids—and permeability—the capacity of a rock to transmit fluids.

Porosity

Porosity is the percentage of rock or soil that consists of void space. It is defined by the equation

$$\text{porosity} = \frac{\text{volume of void space}}{\text{total volume}} \times 100$$

The nature of the void space varies considerably from one type of material to another. The most common types of pore spaces are:

1. Void spaces between mineral grains
2. Fractures
3. Solution cavities
4. Vesicles

Porosity in Unconsolidated Sediments

The porosity of unconsolidated sediments depends mostly on grain size distribution and grain packing. Surprisingly, fine-grained sediments (silt and clay) contain much more pore space than granular material (sand and gravel). Pore space in sand and gravel deposits may vary between about 12 to 45 percent, whereas pore space in clay/silt deposits may reach as high as 80 percent.

The particle size distribution is an important factor in determining the porosity of unconsolidated sediments. If many different grain sizes are present, the finer material can fit into some of the void space between larger grains, and the porosity is greatly reduced (Figure 7-1). The shape and packing of grains are also significant. The porosity of spherical grains packed together as in Figure 7-2 is about 47 percent, but if the grains are shifted by one-half diameter, the porosity drops to about 26 percent.

The filling of voids with cementing material greatly reduces porosity, as does compaction of unconsolidated sediments by the weight of overlying material. The porosity of

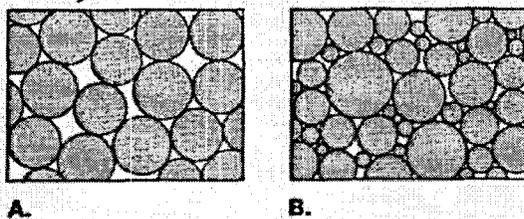


FIGURE 7-1
Comparison of porosities of (A) well-sorted sediments and (B) poorly sorted sediments.

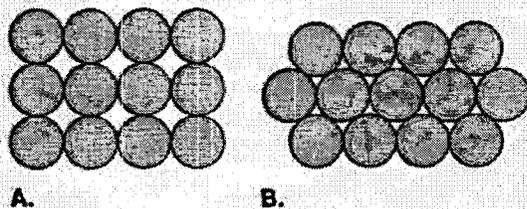


FIGURE 7-2
Effect of packing of spherical grains on porosity.

silt and clay overridden by continental glaciers during the Pleistocene is measurably less than similar sediment not compacted by the weight of overriding ice (Easterbrook, 1964).

Porosity of Rock

Virtually all rocks are cut by fractures, which may constitute the only significant pore space. Even dense granite may be made porous by sets of intersecting joint planes. If the fractures are numerous and interconnected, a surprisingly large volume of water may be contained in otherwise very dense rock (Figure 7-3). Some limestones have high porosity as a result of solution by water moving along joints and bedding planes. The moving water removes soluble material and progressively expands larger and larger conduits that may eventually become caves. Vesicles, formed by gas bubbles trapped in lava, can significantly affect the porosity of volcanic rocks (Figure 7-3). Vesicles are usually concentrated near the top of a lava flow, where they greatly increase porosity, and the vesicles may be interconnected with columnar joints or cinders and rubble at the top and base of the flow.

Porosity of Sedimentary Rocks Groundwater occurs in the pore spaces between grains (primary porosity) of sedimentary rocks, as well as in fractures (secondary porosity). Sedimentary rocks are invariably fractured to some degree, ranging from widely spaced joints to intense fracturing that may almost completely shatter the rock. Some fine-grained, cohesive sediments (high in silt/clay) contain shrinkage cracks that develop when the sediment dries and contracts. The porosity of sedimentary rocks varies from 3 to 30 percent in clastic rocks and 1 to 30 percent in limestones and dolomites (Manger, 1963).

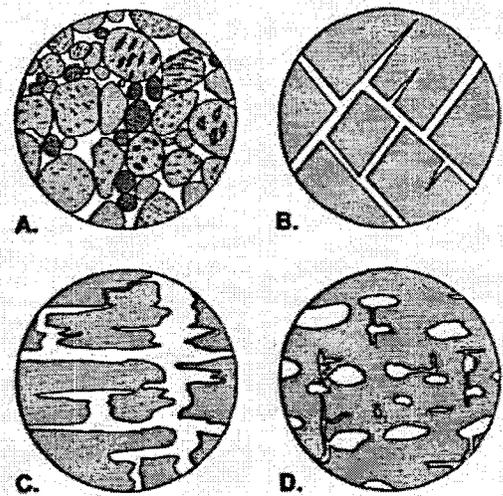


FIGURE 7-3
(A) Porosity resulting from spaces between grains.
(B) Porosity resulting from fracture occurs in most rocks.
(C) Porosity resulting from solution activity is common in limestone.
(D) Porosity resulting from vesicles is common in lava flow

Limestone, dolomite, gypsum, and evaporite deposits are formed by chemical precipitation or biochemical accumulation of calcium carbonate, calcium-magnesium carbonate, calcium sulfate, or sodium chloride. All may be strongly affected by groundwater percolating through pore spaces, fractures, and bedding planes, dissolving and enlarging them to give the rock high secondary porosity.

Porosity of Crystalline Plutonic and Metamorphic Rocks Interlocking crystals of plutonic rocks (formed by igneous intrusions) and metamorphic rocks typically give these rocks very low primary porosities, but intersection of fractures may allow significant secondary porosities. Joint sets in crystalline rock usually occur in three mutually perpendicular directions which may increase the porosity of crystalline rocks by about 2 to 5 percent. Porosity due to jointing is concentrated along the fracture zones and increases with the width of the joints. Rocks in fault zones may be extensively fractured, giving them high porosity. Weathering of plutonic and metamorphic rocks can increase their porosities from 30 to 60 percent.

Porosity of Volcanic Rocks Volcanic rocks cool and crystallize more rapidly than plutonic rocks; thus, volcanic rocks commonly have much higher porosity. Lava flows cool rapidly at the surface where escaping gases produce vesicles that may give the rock high porosity, although the vesicles may not be interconnected. The high vesicularity of pumice can produce porosities as high as 87 percent. More significant columnar joints, formed by shrinkage of the lava as it cools, can produce relatively high porosities with effective permeabilities. Additional porosity may be imparted by the following

1. The rubble tops of lava flows, formed by the breaking of crusted-over lava as it continues to flow beneath

2. Lava tubes through which molten lava once poured
3. Stream gravels trapped between lava flows

Pyroclastic deposits, consisting of ash and cinders thrown into the air during eruptions, occur in loose, unconsolidated deposits that can have porosities ranging from 15 to 50 percent. Weathering of such volcanic deposits can increase the porosity to more than 60 percent.

Permeability

Permeability, the ease with which a material can transmit a fluid, varies with the fluid's viscosity and hydrostatic pressure, the size of void space, and the degree to which the open spaces are interconnected. Porosity and permeability are not synonymous—a material can have high porosity but low permeability. Porosity is a measure of how much water a rock can hold, whereas permeability is how easily water can be transmitted; thus, permeability is dependent on the number of connected voids in a material rather than on the total volume of voids.

Sediments or rocks with high porosity do not necessarily also have high permeability. For example, clay has very high porosity, typically 50 to 80 percent, but it has very low permeability because water has difficulty moving through the small pore spaces. Voids are very small and typically not connected, and the molecular attraction between particles and the thin films of water further inhibit movement. Sand has considerably lower porosity than clay, commonly less than 30 to 40 percent, but it is more permeable than clay because the pores are larger and are interconnected (Matsch and Denny, 1966).

Flow velocities of groundwater typically vary from about 1 m/day to 1 m/yr. The highest rate of percolation measured in the United States, in unusually permeable material, is only 250 m/day (820 ft/day). In special cases, such as the flow of water in caves, the velocity of groundwater may approach that of surface streams.

Specific Yield and Specific Retention

Specific yield is the ratio of the volume of water that drains by gravity from saturated sediment or rock to the total volume of the material. The surface tension of water molecules causes them to cling to grain surfaces. The gravitational force on a film of water surrounding a mineral grain will cause some of the water to pull away, in spite of the surface tension, and drip downward, leaving a thinner film of water with greater surface tension, until the gravitational force is equal to the surface tension. The moisture clinging to mineral grains by surface tension is known as **hygroscopic water**. If two samples have the same porosity but different grain sizes, the total grain surface area of the finer sample will be larger, allowing more water to be held as hygroscopic moisture.

The **specific retention** of a sediment or rock is the ratio of the volume of water a sample can retain against gravity drainage to the total volume of the rock. Specific yield S_y is the volume of water that a sediment or rock will yield by gravity drainage, and specific retention S_r is the remaining volume. Thus, the sum of the two is equal to porosity.

$$\text{porosity} = S_y + S_r$$

Because specific retention is greater for samples with small grain sizes, clay may have a porosity of 50 percent and a specific retention of 48 percent, yielding only a 2-percent specific yield. Maximum specific yield occurs in sediments having medium-to-coarse sand sizes (0.5–1.0 mm) (Johnson, 1967).

Hydraulic Conductivity

Darcy (1856) demonstrated that the discharge of a saturated system is proportional to the hydraulic head between two points, to the cross-sectional area, and to the permeability of the material. This relationship is now commonly expressed as **Darcy's law**.

$$Q = PIA$$

where

- Q = discharge
- P = a permeability coefficient sometimes called the **hydraulic conductivity**
- I = **hydraulic gradient**, represented by the change in hydraulic head between two points
- A = cross-sectional area

Permeability of Sediments

The most copious producers of groundwater are commonly coarse-grained, unconsolidated sediments that are noted for their high permeability. Conversely, clay has very low permeability. Mixtures of the two produce a wide-ranging continuum of permeability values for unconsolidated sediments.

The permeability of sediment depends largely on the size of pore spaces, which is directly related to grain size. The smaller the grain size the larger the overall grain surface area and the greater the resistance to flow. The relationship between permeability and grain size includes the following factors:

1. Permeability increases as the median grain size increases, due to larger pore openings that are more likely to be connected.
2. Permeability decreases for a given median grain diameter as the standard deviation of particle size increases (indicating a more poorly sorted sediment), because the finer material can occupy the voids between larger fragments.
3. Coarse-grained sediments show a greater decrease in permeability with an increase in standard deviation of particle size distribution than do fine-grained sediments.

4. Sediments having unimodal particle size distribution have greater permeability.

Hydraulic permeability may be measured in the field, based on grain-size analysis of sediments and hydraulic conductivity tests of monitoring wells. Aquifer pumping tests may be used to determine the hydraulic conductivity of sediments in the field, providing an integrated, average permeability over a large area.

Permeability of Rocks

The permeability of rocks depends on the primary void space generated when the rock was formed and the secondary void space created after the rock was formed. Clastic sedimentary rocks typically have primary permeability properties somewhat similar to those of unconsolidated sediments, only with reduced pore space caused by cementation and compaction, which can substantially reduce permeability even without a large change in primary porosity. Primary permeability may be enhanced by sedimentary structures, such as bedding. Conglomerate, sandstone, and some limestone are the sedimentary rocks most likely to have high permeability. Conglomerate and sandstone typically have high primary permeability, whereas permeability in limestone is usually secondary, from fractures and solution cavities. Shale has low permeability because of its small void spaces.

Crystalline igneous, metamorphic, and chemical sedimentary rocks typically have low porosity and low primary permeability because of tightly interlocking crystals, which allow very few openings. However, high secondary permeability can develop in crystalline rocks as a result of fracturing, and some volcanic rocks have high primary porosities because of large numbers of vesicles. If the vesicles are large and interconnected, high permeabilities result. Lava flows also may have high permeabilities because of extensive columnar jointing and because the tops of flows may consist of rubble from the broken crusts of flows.

THE WATER TABLE

As water infiltrates soil at the ground surface, it is impelled downward by gravity through pore spaces that are only partially saturated. As long as the water continues to encounter

empty pore spaces, movement of the water is dominantly vertical (Figure 7-4). Eventually, however, all of the pore spaces become completely filled with water, leaving no more space for continued vertical percolation, and the material becomes totally saturated, producing a subsurface region known as the **zone of saturation**. The upper surface of the zone of saturation is the **water table**. Once water reaches the zone of saturation, it can no longer continue to move vertically downward because all of the pores beneath are already filled, so it moves laterally in the direction of the slope of the water table. The water table is not horizontal—it rises higher beneath hills and slopes in the direction of valleys where groundwater is discharged into streams, approximately mimicking the surface topography. The slope of the water table (h/l) represents the hydraulic gradient whose inclination depends on the permeability of material and the rate at which water is added to the zone of saturation.

The water table can be mapped from the level of water observed in wells and from the movement of water measured by injection of dyes and other tracers. It may be only a meter or so beneath the surface along flood plains in humid regions, but it may be hundreds of meters below the surface in deserts.

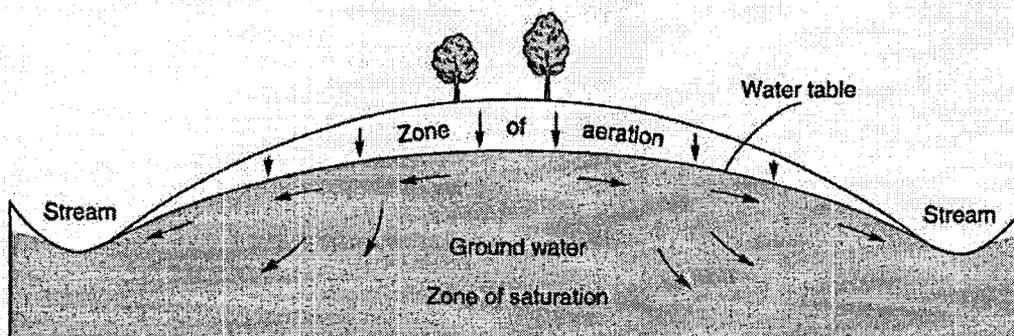
AQUIFERS

An aquifer is a geologic unit that can store and transmit water. Both unconsolidated sediments and rock have wide ranges of hydraulic conductivities. The permeability of aquifers is generally above about 10^{-2} darcy. Unconsolidated sand and gravel, sandstone, limestone, lava flows, and fractured plutonic and metamorphic crystalline rocks are examples of typical aquifers.

Unconfined Aquifers

Aquifers that extend continuously from the land surface downward through material of high permeability are known as **unconfined aquifers**. Recharge to the aquifer may be from percolation downward through the unsaturated zone, from lateral groundwater flow, or from upward seepage through underlying material.

FIGURE 7-4
Distribution and movement of subsurface water.



Aquifer transmissivity is a measure of the amount of water that can be transmitted horizontally by the full, saturated thickness of the aquifer under a hydraulic gradient of 1. The transmissivity T is the product of the hydraulic conductivity K and the saturated thickness t of the aquifer.

$$T = tK$$

Groundwater discharges wherever the water table intersects the land surface, usually along stream channels, marshes, and lakes where a link is established between groundwater reservoirs and surface elements of the hydrologic system. The discharge of groundwater into surface drainage systems provides a substantial amount of water for streamflow and keeps many streams from drying up between precipitation events.

Confined Aquifers

Subsurface material sometimes contains confining layers, having very low permeability, typically less than about 10^{-2} darcy. Groundwater may move through confining layers, but the rate of movement is usually very slow. Confining layers can be subdivided into aquitards, aquicludes, and aquifuges. An aquitard is a leaky confining layer of low permeability that can store groundwater and transmit it sluggishly from one aquifer to another. An aquiclude also has low permeability but is situated in a position to form the upper or lower bound-

ary of a groundwater flow system. An aquifuge is an almost totally impermeable body of rock or unconsolidated material.

Artesian water is confined groundwater that is under high hydrostatic pressure. The geologic conditions under which an artesian water system may develop include the following:

1. A permeable aquifer confined between impermeable layers
2. Surface infiltration to recharge the aquifer
3. Precipitation and infiltration adequate to fill the aquifer and maintain a hydrostatic head (Figure 7-5)

Artesian aquifers are overlain by an aquiclude. Recharge to them may occur where the aquifer is exposed at the surface, or by slow, downward percolation through an aquitard. Water from such an aquifer will rise above the top of the confined aquifer where it is intersected by a well or other conduit as a result of hydraulic head. The level to which water will rise in a well is the potentiometric surface (Figure 7-6). A potentiometric surface is inclined away from the recharge area because of frictional losses as water moves through the pores in an aquifer and because pressure is lost through leakage in fractures. If the potentiometric surface of an aquifer is above the land surface, a flowing artesian well is developed in which water flows from the well without pumping.

Where a layer of low-permeability material occurs as a lens within more permeable material, water moving downward through the unsaturated zone is intercepted and will collect on top of the impermeable lens. As saturation develops

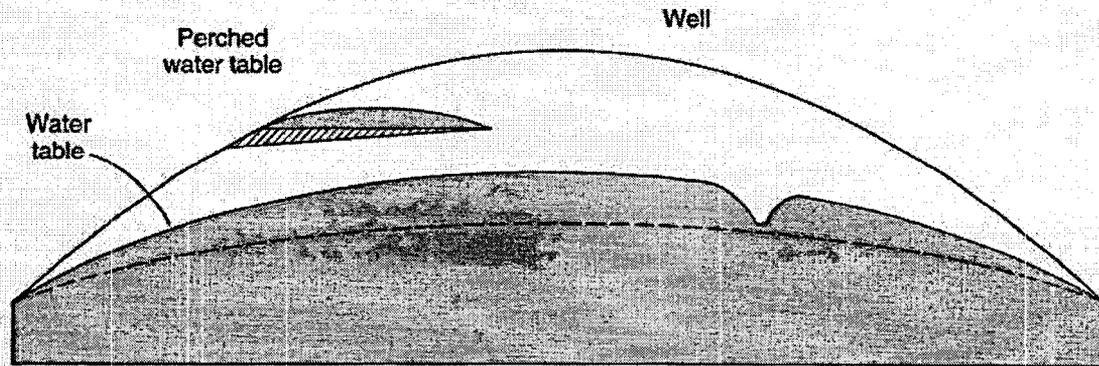


FIGURE 7-5
Regional and perched water tables.

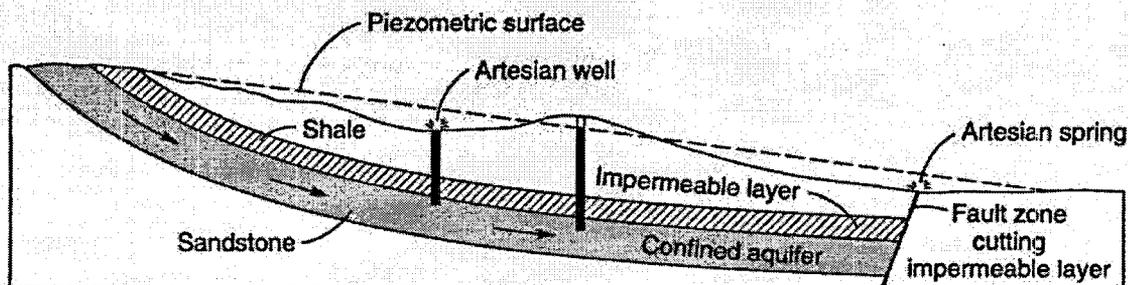


FIGURE 7-6
Artesian water resulting from a confined aquifer.

ove the impermeable lens, a perched aquifer is formed above the main water table (Figure 7-5). Water moves laterally above the low-permeability layer, much like a miniature expression of the regional water table, either until it spills outward over the edge of the lens toward the main water table or until it is intercepted by the ground surface where a spring will form (Figure 7-7).

A water well can withdraw groundwater by digging or drilling through the zone of aeration into the zone of saturation where water flows from pore spaces into the well. Without pumping the well, water will rise to the level of the water table. In order to extract the water, pumping is usually needed to bring it to the surface. However, under pumping, the removal of water causes the water table to be drawn down around the well, producing a cone of depression in the water table (Figure 7-5). If the rate of withdrawal of water by pumping is greater than the inflow of groundwater to the well, the cone of depression will grow increasingly large, flattening the hydraulic gradient in the vicinity of the well and progressively decreasing the rate at which water can be withdrawn, until ultimately the well goes dry.

THERMAL SPRINGS AND GEYSERS

In regions of unusually high geothermal gradients, especially those of recent volcanic activity, groundwater becomes heated to high temperatures, creating thermal springs (Figure 7-8) and geysers (Figures 7-9 and 7-10; Plate 5A) when the water discharges to the surface. If the water is hot enough, geysers may form when the water flashes into steam near the surface. In order for this to happen, rock temperatures just beneath the surface must be high, and groundwater must have access to the heated rocks through fracture systems. Geysers erupt when groundwater is heated to the boiling point, but because water near the base of the water column is under greater pressure than the water above, its temperature can rise above the boiling point of water at the surface. Boiling of water at the surface causes water to rise slightly, releasing pressure on the water column below. If the temperature of the water below is close to boiling, the water may flash into steam under the new reduced pressure, and the geyser will erupt (Figure 7-10).

GROUND SUBSIDENCE DUE TO WITHDRAWAL OF FLUIDS

Withdrawal of large quantities of groundwater or other fluids, such as petroleum, decreases the fluid pore pressure between grains. The decrease in pore pressure allows tighter grain-to-grain contact and leads to sediment compaction and subsidence of the ground surface. The amount of such subsi-

dence can be measured by precision leveling relative to stable reference points. Such measurements have led to some startling results.

One of the most spectacular examples of ground subsidence due to groundwater withdrawal occurs in Mexico City, which is built on unconsolidated, water-saturated, alluvial and lacustrine sediments. Most of Mexico City's water supply is from wells located within the city. As groundwater is pumped for domestic and industrial use, pore pressures decrease, causing widespread, slow subsidence of the land surface. Between 1891 and 1959, the maximum subsidence was 7.5 m (25 ft), and all of the city except the western edge subsided at least 4 m (13 ft) (Figure 7-11). The opera house (weighing 54,000 metric tons) has settled more than 3 m (10 ft), and half of the first floor is now below ground level.

Another example of ground subsidence caused by withdrawal of subsurface fluids occurs in Long Beach, California, which has subsided about 9 m (27 ft) as the result of 40 years of oil production. This subsidence has caused almost \$100 million worth of damage to wells, buildings, pipelines, transportation facilities, and harbor installations.

Compaction and subsidence also present serious problems in areas of recently deposited sediments. In New Orleans, for example, large areas of the city are now 4 m (13 ft) below sea level, a drop due largely to the pumping of groundwater. As a result, the Mississippi River flows some 5 m (16 ft) above parts of the city. Rainwater must be pumped out of the city at considerable cost, and water lines and sewers are damaged as the ground subsides.

KARST

Although groundwater is rarely seen, its effect on topography may be profound as a result of the dissolving of soluble rocks and the transportation of the dissolved material in solution. The effects of solution produce unique landforms that distinguish solution topography from other landforms.

Groundwater is typically slightly acidic and reacts chemically with subsurface rocks, especially those that are highly soluble, such as carbonate rocks (limestone and marble), dolomite (magnesium carbonate), gypsum (calcium sulfate), and some evaporites. Extensive solution of such rocks produces extraordinary landforms grouped under the general category of karst, characterized by numerous closed depressions, caves, various collapse features, and diversion of drainage underground.

The word karst stems from the German form of the Slavic word Krs or Kras and the Italian word carso, meaning "barren, stony ground," or "a bleak waterless place" (Monroe, 1970). Such features are typical of both the classical karst region on the high plateau near the Adriatic Sea between Italy and Yugoslavia and the Dinaric karst region to the south. This region is renowned for its large caves, irregular topography with many closed depressions, interrupted stream valleys, and

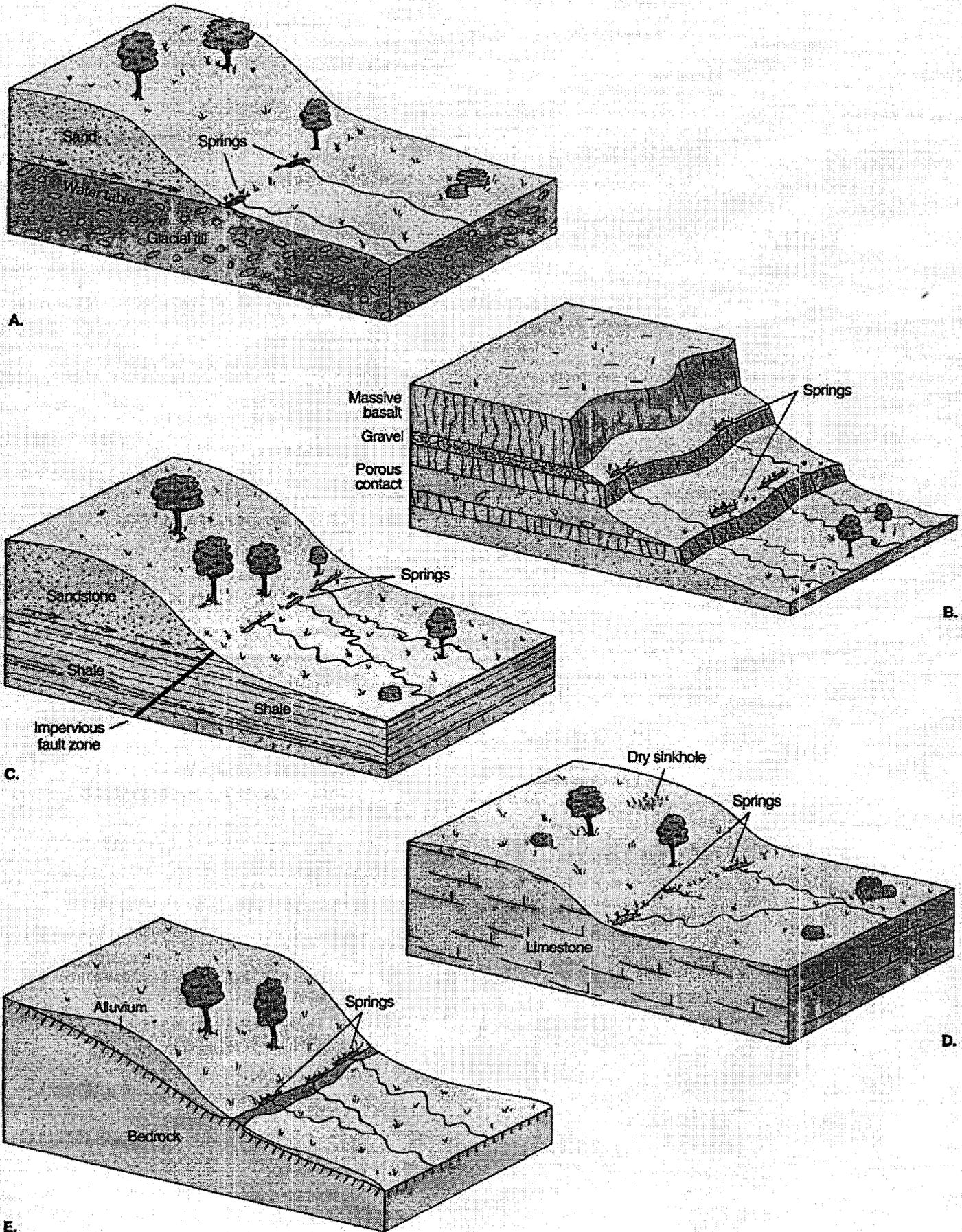


FIGURE 7-7

Springs (A) from the intersection of perched water and the land surface, (B) from interlayered permeable and impermeable rocks, (C) from water rising to the surface as a result of blockage of groundwater flow by an impermeable fault zone, (D) from intersection of solution cavities the surface, or (E) from rock protruding through alluvium.

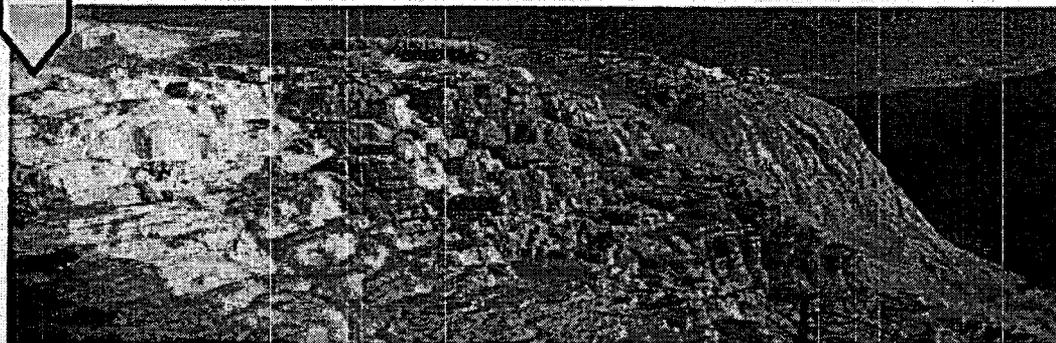


FIGURE 7-8
Mammoth hot springs,
Yellowstone National Park.
(Photo by W. K. Hamblin)

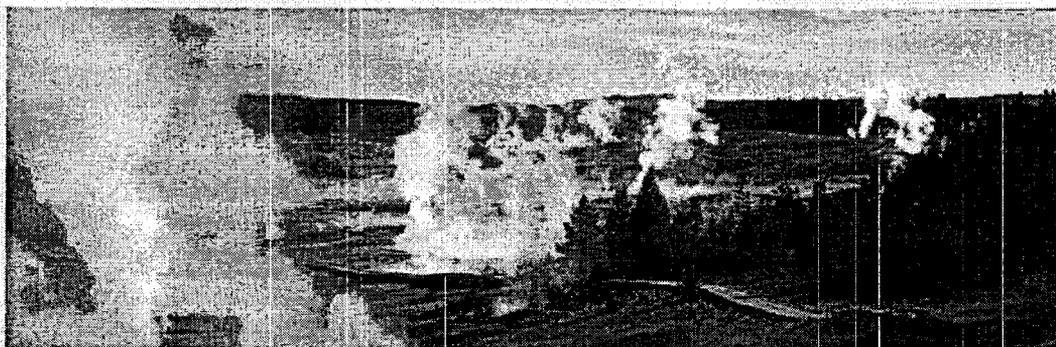


FIGURE 7-9
Thermal geysers, Yellowstone
National Park. (Photo by W. K.
Hamblin)

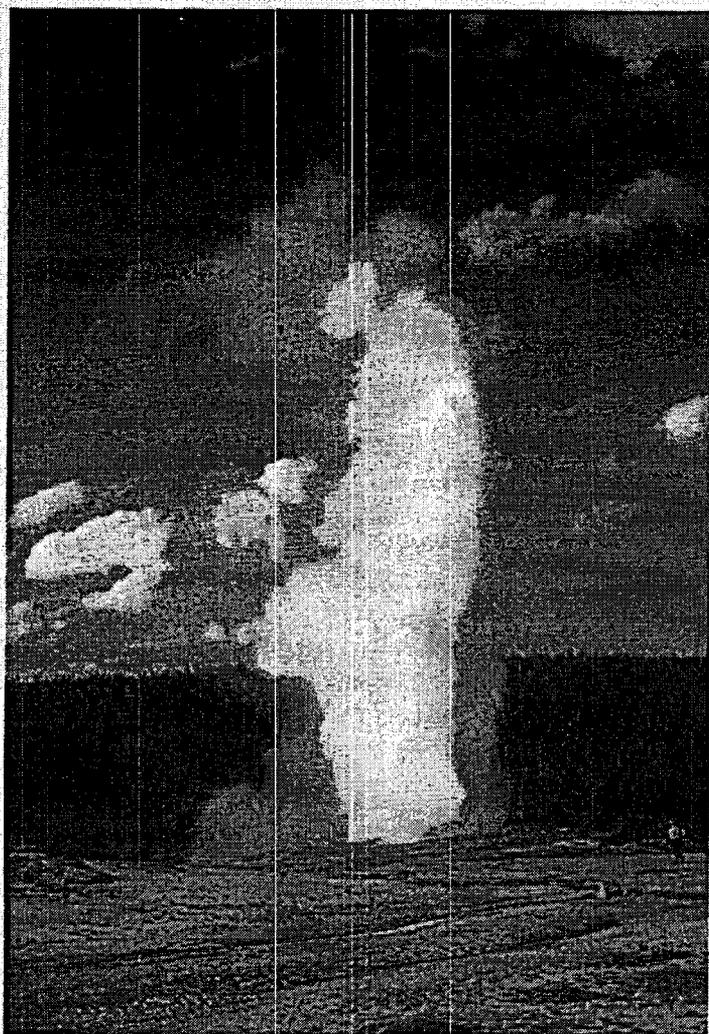


FIGURE 7-10
Old Faithful geyser, Yellowstone National Park.

an assortment of enclosed and subsurface-drained hollows (Figure 7-12).

Early studies of karst date back to the end of the eighteenth century (Roglic, 1972) when the collapse origin of solution depressions was recognized (Gruber, 1791; Hacquet, 1778; Virlet, 1834; Foumet, 1852; Prestwich, 1854). Owen (1856) and Cox (1876) first expressed the view that sinkholes in Kentucky and Indiana were formed by solution of limestone by infiltration of water.

Because the solution of limestone requires abundant water and dissolved carbon dioxide, the formation of karst landforms is affected by climate, especially the amount and distribution of rainfall. Variations in the calcium carbonate content of groundwater in karst terrains suggest that limestone solution may be different in various climates (Corbel, 1959). In arid climates, the water table is usually deep below the surface, greatly inhibiting solution activity. Some karst features do form in arid or semiarid climates, but the karst topography is quite subdued, and features such as collapse sinkholes and deranged surface drainage are not as conspicuous. In some instances, arid karst features formed when the climate was more humid than at present, or they may have formed due to acidic fluids rising upward into the rocks from deep-seated sources such as at the famous Carlsbad Cavern (Egemeir, 1981; Hill, 1981). In high-latitude cold regions water, when present, is often frozen, severely retarding free circulation and microbiologic activity. Karst is therefore quite rare in arctic and antarctic regions. Humid tropical climate however, provides an optimal combination of temperature and precipitation to facilitate solution processes. Chemical reactions take place more rapidly at higher temperatures, an luxuriant vegetation, together with intense microbiotic activity

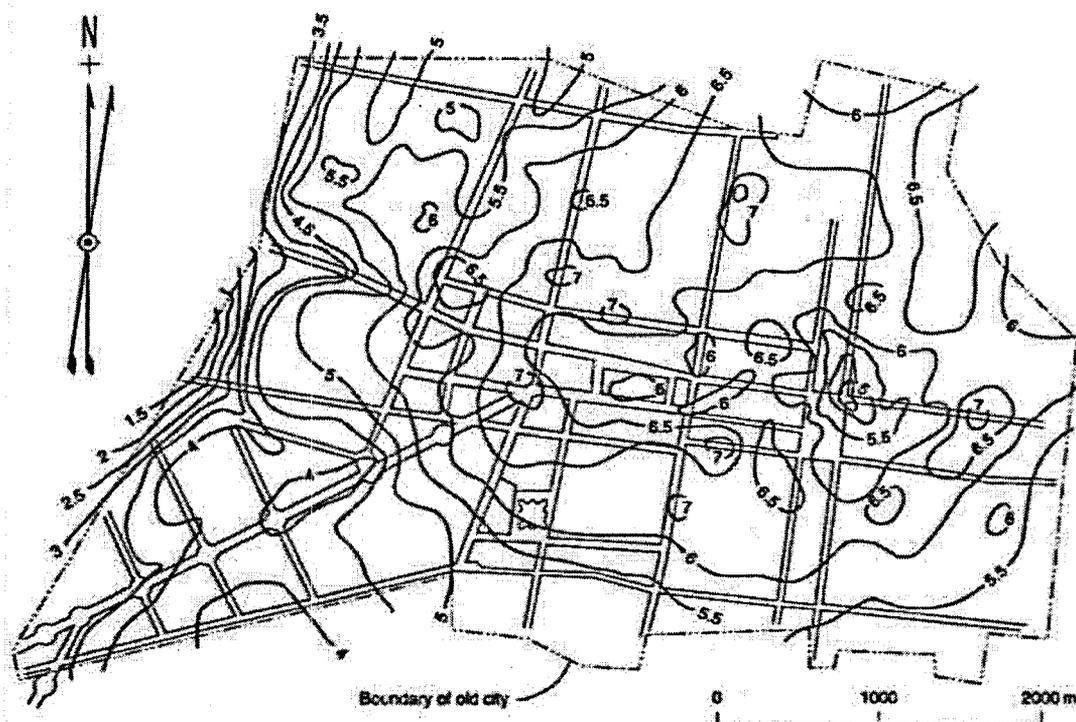


FIGURE 7-11
Subsidence of Mexico City, 1891–1959, shown by lines of equal subsidence (meters). (After Comision Hidrologica de la Cuenca del Valle de Mexico, 1961)

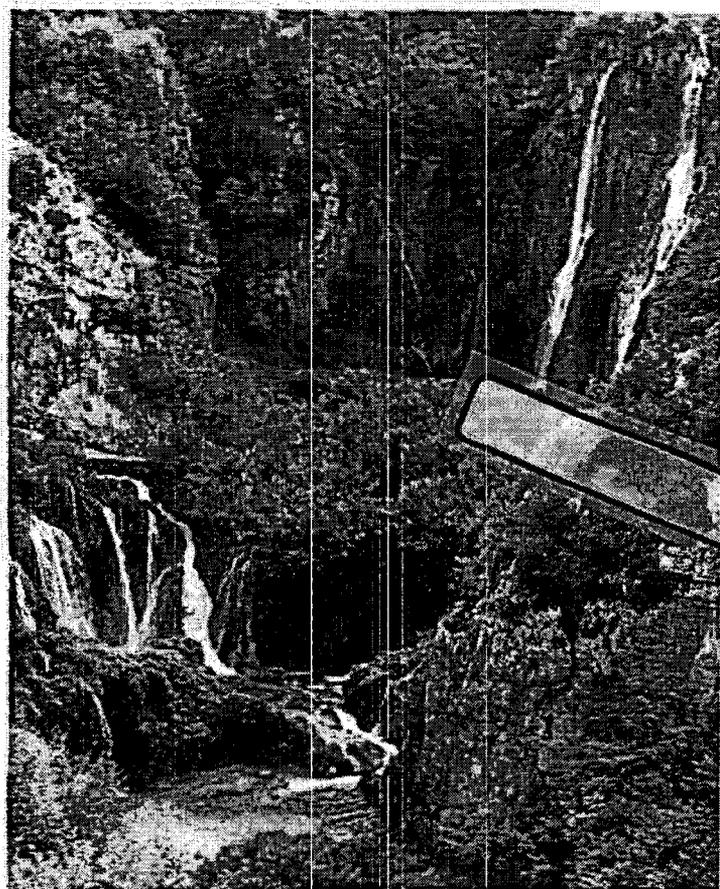


FIGURE 7-12
Solution valley, Plitvice, Yugoslavia.

ity, imparts high partial pressures of carbon dioxide in tropical groundwater, making it exceptionally effective as an agent of solution and producing strongly developed karst features (Jennings and Bik, 1962; Monroe, 1976). Generation of bio-

genic CO_2 is perhaps the most important aspect of tropical regions that enhances solution (Smith and Atkinson, 1976; Trainer and Heath, 1976; Drake, 1980; Brook et al., 1983).

Karst Controls

Lithology and Structure

Although karst develops primarily on carbonate rocks (mostly limestones), not all carbonate rocks possess the proper combination of physical and chemical properties that are conducive to generation of karst topography. Geologic structure is an important factor in karst development (Palmer, 1977).

Most of the world's karst regions are developed on limestone, which by definition consists of at least 50 percent calcite and/or aragonite (CaCO_3). Isomorphous substitution of magnesium for calcium in the carbonate mineral structure forms the mineral dolomite (CaMgCO_3). If more than 50 percent of the rock is composed of dolomite, the rock is a dolomite (or dolostone). Because substitution of magnesium for calcium in the mineral structure may occur in virtually any amount, a considerable variation in composition of a carbonate is possible. In general, the purer the limestone is in CaCO_3 , the greater is its proclivity to form karst. Some evidence suggests that about 60 percent CaCO_3 is necessary to form karst, and about 90 percent may be necessary to fully develop karst (Corbel, 1957). However, even pure limestones may not produce karst because other important factors may be lacking. Some karst features may form on dolomite, but their permeability is typically lower than that of limestone (Herman and White, 1985), so the occurrence of karst in dolomites is usually relatively minor. Rock salt (halite, NaCl)

and gypsum (CaSO_4) are even more soluble than limestone, but they are not widely distributed.

Optimum conditions for karst development include more than just the mineralogy of carbonate rocks. Porosity and permeability also play important roles, especially the ease with which water moves through the rock. Primary porosity, the void spaces in the rock between mineral grains, commonly decreases with time because pore space is diminished by recrystallization of calcite and dolomite, precipitation of calcite/dolomite cement, and changes in mineralogy. The older the rock, the greater the tendency for increase in calcite grain size by recrystallization and replacement of calcite by dolomite, both of which decrease the primary porosity.

Primary porosity is generally considered much less important than secondary porosity and permeability produced by fractures and bedding plane partings. The type and distribution of secondary openings are commonly regarded as the most important factors in karst formation because they permit rocks to hold more water and facilitate groundwater circulation through the system by increasing permeability, which is more important than porosity in the formation of karst. Limestone with high primary porosity seldom develops karst without secondary permeability through fractures and bedding planes. Even rocks with high primary porosity may not develop strong karst features if circulation is limited. The most important structural feature of carbonate rocks consists of joints, which act as conduits for groundwater circulation. Joints in sedimentary rocks typically consist of sets that commonly intersect at angles of 70° to 90° , allowing free circulation between and among fractures. The spacing of joint sets is also important. If intersecting joint planes are spaced too far apart, circulation is impeded; but if they are too closely spaced, the rock may be too structurally weak to support karst features, even though the rock may be highly permeable.

Faults can also transmit water, but their spacing is usually much greater than that of joints, so they are not as effective in developing karst. Some fault zones that have fine-grained fault gouge or secondary mineralization along the fault plane have low permeability, which deters karst formation. If secondary crystallization along the fault plane includes sulfide minerals, oxidation may generate sulfuric acid, which aids the solution process (Pohl and White, 1965; Moorhouse, 1968).

Considering the above-mentioned factors, optimum lithologic and structural conditions for full karst development include the following:

1. Thick, pure calcite, crystalline limestone uninterrupted by insoluble beds
2. Intersecting joint sets that allow free circulation of groundwater along discrete flow paths with enough discharge to create or enhance significant solution openings

The Solution Process

The process of carbonate solution forms the basis for development of karst. The solution process consists of a series of steps:

1. Neither groundwater, nor the rainwater from which it was derived, consists of pure water (H_2O). Both contain carbon dioxide (CO_2), which is soluble in pure water, and the CO_2 reacts with the water to form carbonic acid.



The amount of CO_2 dissolved in water depends on the partial pressure of CO_2 at the air-water interface in the atmosphere or in pore space, and on the temperature of the water. Dissolved CO_2 in the water increases with increase in the partial pressure of CO_2 of the air and with decrease of water temperature. Cold water will hold more dissolved CO_2 than warm water at any given partial pressure of CO_2 . Abnormally high dissolved CO_2 is commonly found in soil water (Jennings, 1971) as a result of biogenic CO_2 formed from the decomposition of organic material. This increased CO_2 becomes an important factor in the solution process. The carbonic acid dissociates readily into its ionic state according to the equation



3. Calcite dissociates into its ionic state as follows:



4. The hydrogen atom of equation (2) combines with the carbonate ion to form another bicarbonate ion



During this process, the Ca^{++} ion is released from the calcite crystal into the water.

These steps are illustrated in Figure 7-13. Although the net process can be written as

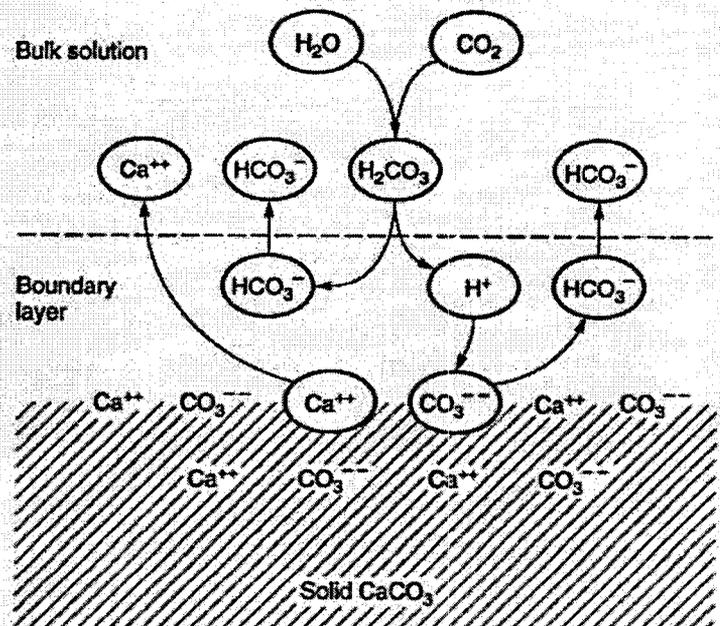
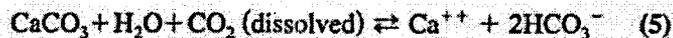


FIGURE 7-13

Steps in the solution of calcite in CO_2 -bearing water. (From W. B. White, 1988, *Geomorphology and Hydrology of Karst Terrains*. Copyright 1988 by Oxford University Press, Inc.)



the system is in reality exceedingly complicated, involving a number of mutually interdependent, reversible reactions. The bicarbonate ions (HCO_3^-), derived from the dissociation of CaCO_3 and the dissolving of CO_2 in water, introduce disequilibrium between the partial pressure of CO_2 in the air and the partial pressure of CO_2 in the water, which causes more CO_2 to dissolve in the water and drives more solution of calcite (Thraillkill, 1972; Thraillkill and Robi, 1981; Weyl, 1958). The amount of CO_2 dissolved in water is very important to the solution process. Where dissolved CO_2 is high, the water vigorously dissolves calcite. Soil water, with its high dissolved CO_2 from organic material, is effective in the solution of limestone (Drake and Wigley, 1975).

Effect of Climate and Vegetation on Solution

Because carbonate solution is the driving force of karst formation, factors that enhance solution also promote karst growth. Thus, climatic factors are important to karst, especially the amount and distribution of precipitation, the temperature, and the evaporation rate. The optimum combination of precipitation and temperature to promote solution processes occurs in tropical humid climates where chemical reactions proceed rapidly and luxuriant vegetation combines with high levels of microbiotic activity to generate high partial pressures of carbon dioxide in groundwater. These conditions actively encourage solution and the proliferation of karst features (Monroc, 1976).

Perhaps an even more important aspect of climate than the amount of water supplied climatically is the high biogenic CO_2 values in groundwater from vegetation and biogenic activities that result from humid tropical climates (Smith and Atkinson, 1976; Trainer and Health, 1976; Drake, 1980; Brook et al., 1983). Areas of more intense solution within the same climatic zone occur as a result of vegetation differences (Woo and Marsh, 1977; Brook and Ford, 1982).

Arid or semiarid regions, where precipitation rates are low and temperature and evaporation rates are high, are less susceptible to karst development. Solution still occurs, but karst features are generally quite subdued, with fewer collapse features and fewer diversions of surface drainage. In cold arctic or subarctic climates (which may be arid to semiarid), karst development is further impeded by permafrost and by the lack of vegetation and biologic activity that is responsible for high biogenic CO_2 levels in groundwater (Smith, 1969). Much of the water that may be present is typically frozen, inhibiting free circulation necessary for solution. Karst topography is rare in such areas.

Karst Groundwater Hydrology

From the earliest days of research on karst phenomena, the nature of karst aquifers has been debated. Cvijic (1893), Grund

(1903), and others argued that both surface and subsurface cave streams drained downward to a karst water table, rather than an integrated drainage system. On the other hand, Katzer (1909), Martel (1910), and others believed that underground cave streams were not necessarily connected and did not flow at common levels or to a karst water table. Thus, two schools of thought have evolved:

1. Groundwater in karst is considered to be similar to that found in nonlimestone terrains, with downward movement through the unsaturated zone to a defined water table below.
2. Groundwater in karst is believed to be confined to interconnected cavities with no defined water table (Baker, 1976; Brucker et al., 1972).

Hydrostatic levels in karst terrain are commonly quite different, with dry cavities close to others filled with water, some conduits crossing the paths of others, and some conduits flooding while others remain dry. White (1988) suggests that the water table controversy can be resolved by considering the water table to be very irregular, with rapid response of the conduit system to hydraulic changes relative to the diffuse system.

Karst Landforms

Sinkholes

Sinkholes are small, shallow, circular to oval, closed depressions formed by downward solution of limestone from the surface or by collapse of the roof of a solution cavity. A synonymous term, *doline*, meaning "small valley" in Slavic, was defined by Cvijic (1893) as a shallow, closed depression; this term was widely used in the early days of karst research in the Adriatic region. Sinkholes comprise the most widely distributed forms in karst terrains, where they occur by the tens of thousands, giving such regions much of their characteristic pitted topographic expression (Figure 7-14 Plate 5B).

Sinkholes vary considerably in size, typically ranging from 10 to 100 m (~30 to 300 ft) in diameter (average ~50 m, or 160 ft) and from 2 to 100 m (~7 to 330 ft) in depth (average ~10 m, or 33 ft), although they can be more than 1000 m (3300 ft) in diameter and hundreds of meters deep. Three principal forms, linked by transitional forms, occur in karst terrains:

1. Bowl-shaped, with very shallow depth relative to diameter, and side slopes of 10° to 12° .
2. Funnel-shaped, with diameter two to three times depth, and slope angles of 30° to 45° . They are much less common than the bowl form.
3. Cylinder-shaped, with depth greater than diameter and steep to vertical sides. They are rare relative to the bowl and funnel shapes.

Sinkholes can occur singly, but more typically they form in large numbers, commonly 40 to 50 in a square kilometer and



FIGURE 7-14
(A) Topographic map of karst features near Mammoth Cave, Kentucky (Mammoth Cave quadrangle, U.S. Geological Survey).



FIGURE 7-14 cont.
(B) air photo of same region (U.S. Geological Survey).

by the hundreds of thousands in a single karst terrain (Cvijic, 1893). The spatial distribution of sinkholes is generally relatively random, although in places sinkholes may develop in rows along the strike of a joint or lithologic contact.

Most of the early studies of karst features in the classic Adriatic areas emphasized the role of roof collapse of underground cavities in the formation of sinkholes, and caves were

regarded as a precondition for formation. However, some noted that the collapse of roofs of solid rock should result in irregular and angular forms, rather than the round, funnel-shaped ones typical of most karst terrains, and that many small sinkholes were not connected with caves. This observation led to the conclusion that the great majority of the sinkholes could not possibly have originated from a collapsed roof.

Early English and American studies of karst differed from the generally prevailing collapse theory, finding instead that sinkholes terminated in small fissures attributed to solution of limestone along joints and cracks by atmospheric water (Prestwich, 1854). Cvijic (1893) argued vigorously against the collapse origin of all sinkholes on the basis that large sinkholes connected to caves by openings less than 1 meter could not have been formed by collapse.

Thus, two primary origins of sinkholes are now recognized: by solution from the surface downward or by collapse into solution cavities. The former are known as **solution sinkholes**, the latter as **collapse sinkholes**.

Solution Sinkholes. Water seeping into joints and fissures in limestone dissolves the calcium carbonate from the surface downward, enlarging the conduits until surface water is diverted into funnel-shaped fissures or bowl-shaped, closed depressions known as **solution sinkholes**, which continue to enlarge about the central sink (Figure 7-15).

The most common form of solution sinkholes is approximately circular. Other forms of sinkholes are probably determined by variations in the progress of decomposition. Round sinkholes occur around central fissures where solution of the limestone proceeds from the center outward toward the periphery. Oval sinkholes form by solution along elongate fissures or major fractures.

Several factors play important roles in the development of solution sinkholes:

1. **Slope.** Sinkholes commonly occur in greatest numbers on level or gently sloping areas, and they are much less numerous on steep slopes. They are especially abundant on elevated plateaus where most of the precipitation seeps into solution fissures. On steep slopes, much of the water runs off on the surface.
2. **Lithology and fracturing.** Dense, highly fractured limestones are more susceptible to solution than massive porous

limestone. Limestones composed of pure calcite dissolve more readily than argillaceous limestone or dolomite.

3. **Soil and vegetation.** Increased CO_2 saturation from organic material speeds up solution.

Collapse Sinkholes are produced by solution beneath the surface and enlargement of underground cavities until the roof collapses and a closed depression is formed on the surface. In the early days of karst investigation in Europe, nearly all sinkholes were thought to originate in this way, but later work has shown that although it is an important mechanism for solution formation, the collapse mode of origin is now considered less than ubiquitous.

In karst areas where limestone is overlain by weathered residual clay, loess, or other unconsolidated sediment, solution of the limestone enlarges subsurface cavities with roofs of unconsolidated material supported only by ever-decreasing limestone or, finally, only by natural arching in the sediment. When gravity overcomes the strength of the arching support, the roof collapses instantaneously, producing a depression at the surface, as shown in Figure 7-16. Another example of this type of collapse origin of sinkholes is shown in Figure 7-17 near Orlando, Florida, where a hole suddenly appeared at the surface and quickly developed into a major collapse feature.

Collapse may originate at varying depths below the surface and may include significant stoping of ceiling rock as well as direct solution. Stopping may play an especially important role in deep-seated collapse, where the initial solution cavity stopes its way upward to the surface, leaving the bottom of the cavity filled with rubble from the stoping, along with the collapsed roof rubble. Deep-seated collapse sinkholes are likely to have steep-sided walls, forming subsidence shafts. If groundwater circulates freely through the lower portions of collapse sinkholes or subsidence shafts, stoped blocks may be removed by solution or mechanical transport, thus facilitating further stoping. Deep-seated stoping may proceed from subsurface solution cavities through substantial thickness of overlying resistant rock. For example, Dante's Descent in

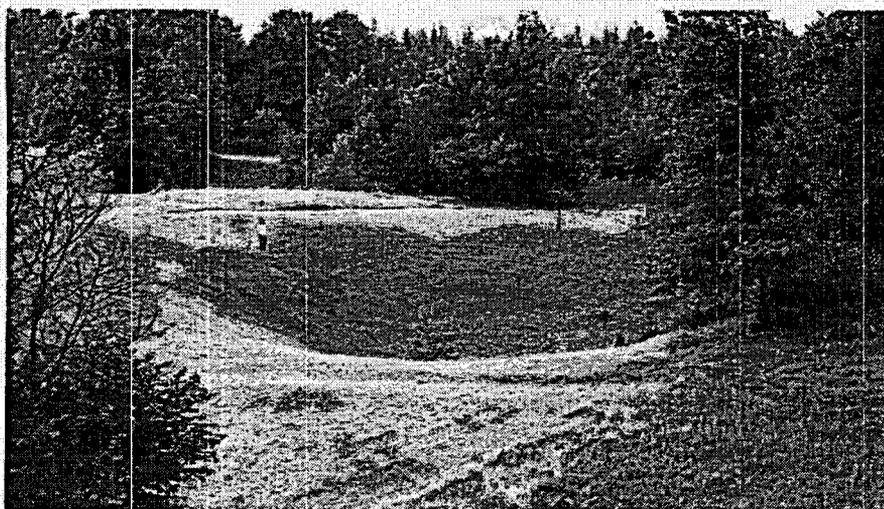


FIGURE 7-15
Sinkhole (doline) caused by subsidence into solution cavity in Cretaceous limestone near Planina, Yugoslavia. (Photo by A. N. Palmer)



FIGURE 7-16
Roof collapse into a solution cavity near Bloomington, Indiana.

northern Arizona is a vertical-sided pit 120 m (400 ft) deep made in Supai sandstone by upward stoping from the underlying Redwall limestone.

Evidence from sudden collapses in Florida and elsewhere suggests that collapsing may be brought about by lowering of the water table (Kemmerly, 1980 a, b). Ironically, some of the most spectacular collapses in the karst area around Orlando, Florida, have occurred following periods of drought when the water table has been substantially drawn down, allowing near-surface sand to collapse into solution cavities below.

Uvalas

Uvalas are essentially compound sinkholes made by the enlargement and coalescence of smaller, individual sinkholes.

They are typically shallow and somewhat irregular in shape as a result of the merging of multiple sinkholes of different diameter (Jennings, 1967). Their size may span several square kilometers in area, and their depth may vary from a few meters to 200 m.

Poljes

The term *polje* is Slovene for “field” and is used in karst literature to mean a large, closed depression with an alluvial flat floor, bounded by steep sides. The flat alluvial of poljes floors comprise one of their distinguishing characteristics. Poljes are perhaps best developed in the classic Adriatic karst region where they are typically 1 to 5 km (0.6 to 3 mi) wide and up to 60 km (36 mi) long. Their floors are flat and are usually filled with alluvium, making a sharp boundary with the surrounding limestone hills. Water enters the poljes from springs, flows across the flat floors as surface streams, and then disappears into sinkholes and caves at the opposite side. Water moves from one polje to another at successively lower elevations until it discharges as large springs along the Adriatic Coast. Poljes may flood annually from seasonally increased flow from springs, and the flat floors of the poljes may be entirely inundated to become shallow lakes, which later dry up.

The Adriatic poljes are formed in limestones that are part of a complex tectonic area, and the axes of elongation of the poljes seem to parallel structural trends in the region, suggesting that tectonism may play a significant role in their origin (Cvijic, 1960). Some of the poljes occur along faults, but where flat polje floors truncate bedding, factors other than structure must be important. Grund (1903) contended that the flat polje floors had been planed by solution to the level of the



FIGURE 7-17
Sinkhole that began in the backyard of a neighborhood in Winterpark, Florida and eventually swallowed up several new Porsches from a car dealership, several buildings, a swimming pool, a recreational vehicle, and a camper. (Photo by Geophoto)

water table. Louis (1956) and Roglic (1957) pointed out that the alluvium making up the polje floors comes from erosion of the adjacent mountains and that as the floors of the poljes become filled with sediment, local perching of groundwater occurs, which facilitates growth of the polje by lateral solution of the polje walls. Geophysical studies provide evidence to support both tectonic and karst origins for the poljes (Mijatovic, 1983). The alluvial fills of some poljes are quite shallow, simply mantling a thin veneer of sediment over solution-planated limestone floors. Other sediment fills are considerably deeper, extending below sea level in areas where active faults and earthquakes are common, suggesting a tectonic origin. The poljes of the Adriatic are thus a combination of intense solution processes and active block faulting.

Similar large, closed depressions with hydrologic features comparable to the Adriatic poljes occur in Canada (Brook and Ford, 1977), Cuba (Gradzinski and Radomski, 1965), Borneo (Sunartadirdja and Lehmann, 1960), Jamaica (Sweeting, 1958), and Ireland (Williams, 1970).

Sinkhole Ponds and Karst Lakes

Sinkhole ponds and karst lakes [Plate 5-B] are sinkholes that have filled with water because:

1. The central drain of the sinkhole has become plugged with clay and silt washed in from surface drainage or left behind as a residue during solution (Figure 7-18)
2. The sinkhole intersects the water table, and the level of the sinkhole pond or lake is the top of the water table (Figure 7-19)

The water surface of the lakes and ponds in Figure 7-19 is remarkably consistent in elevation, even though the surrounding topography has more than 30 m (100 ft) of relief, suggesting that the lakes and ponds represent the water table. Figure 7-20 is a cenote (sinkhole) in Yucatan, Mexico, which intersects the water table. The water in the sinkhole can be



FIGURE 7-18
Sinkhole pond.

observed to flow slowly in a consistent direction, down the slope of the regional water table.

Solution Chimneys and Vertical Shafts

Solution chimneys are formed by the dissolving of limestone walls along fissures or bedding planes, which are structurally controlled. Some are developed along a single fracture (Raines, 1972), whereas others follow steeply dipping bedding planes.

Vertical shafts are circular cylinders with vertical walls cutting across bedding. Their shape is generally controlled by groundwater hydraulics and is independent of structure and bedding (Pohl, 1955). They are most common where solution-resistant caprock prevents solution along vertical planes and solution is centered in restricted areas.

Vertical shafts are distinguished from solution chimneys primarily on the basis of hydraulic control versus structural control of solution morphology. Vertical shafts may be completely roofed over or may exit at the surface as a result of collapse of their roof. Shafts filled with rubble may appear as small, closed depressions. Other shapes of shafts and pits may act as subsurface drains, connecting closed depressions with cavities and caves below, or they may simply feather out at depth.

Solution along Fractures Solution of carbonates is usually enhanced along joint planes as water circulates through the fractures, leading to etching out of the joints (Figure 7-21). Cutters are planar notches in limestone made by vertical solution along joint planes (Howard, 1963). They are usually a few centimeters to several meters wide, a meter to tens of meters deep, and tens of meters or more long (Figure 7-22). Cutters are among the most common features in karst regions, but they generally do not have much surface expression because they taper downward and are typically filled with soil. They are most observed in road cuts or other vertical sections in horizontal to gently dipping limestone. Cutters occur in folded limestones as well as in horizontal beds. They are oriented along the strike of the bedding and make elongate, parallel limestone ledges. After prolonged solution, the limestone may have the appearance of Swiss cheese (Figure 7-23).

Cutters grade into larger and deeper features known as solution fissures, similar to cutters but with much greater depths than widths; solution corridors, having widths of tens of meters; and solution canyons, having widths and depths of tens to hundreds of meters and lengths of a kilometer or more (Monroe, 1976; Jennings and Sweeting, 1963). These forms typically have linear or rectilinear patterns as a result of structural control of fracture development.

Transition between Fluvial and Karst Drainage

Among the unique features of karst topography is the disruption of surface drainage and diversion underground

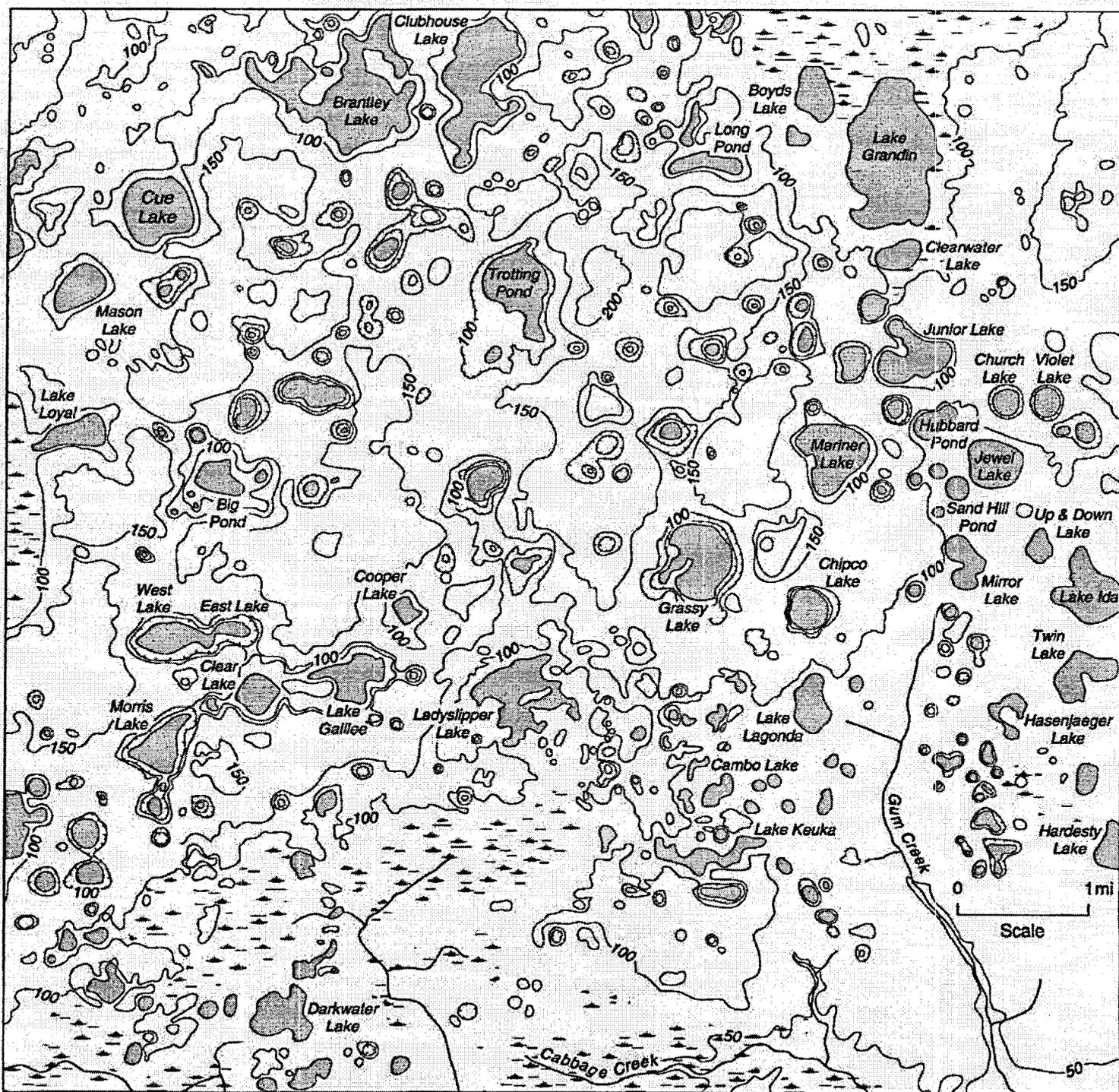


FIGURE 7-19

Sinkhole lakes and ponds representing the regional water table. (Data from Interlachen, Florida, quadrangle, U.S. Geological Survey)

through interconnected solution cavities. Surface streams end abruptly at sinkholes and swallow holes, and they emerge just as abruptly from karst springs as high-discharge streams. Thus, drainage in karst areas has both surface and subsurface components. Which of the two components predominates is a function of the thickness of limestone relative to the total thickness of exposed rocks and the surface area underlain by limestone relative to the area of the drainage basin.

In some of the classic Adriatic karst region, where thick limestones are exposed over extensive drainages, the drainage is entirely karstic, and although such areas are rare, their significance is such that they have become known as *holokarst* (Cvijic, 1960). The more common situation, with both karst and fluvial features, is referred to as *fluviokarst*. The transition from fluvial to karst drainage is best seen where normal fluvial drainage, developed on clastic beds overlying lime-

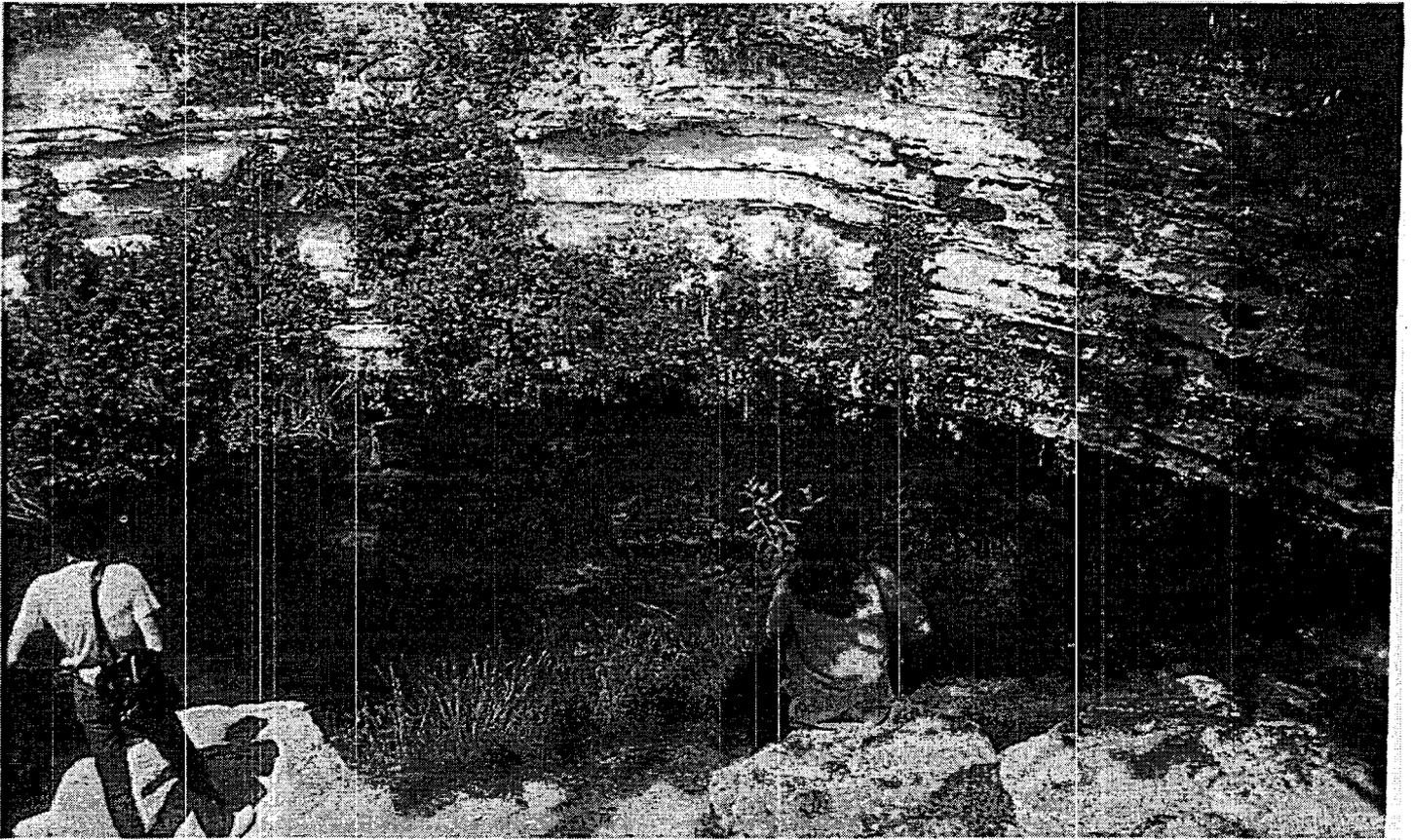


FIGURE 7-20

Cenote (sinkhole) in Yucatan, Mexico. The water level corresponds to the top of the regional water table, which can be observed to flow from right to left.



FIGURE 7-21

Joints in Carboniferous limestone widened by solution, Yorkshire, England. (Photo by A. N. Palmer)

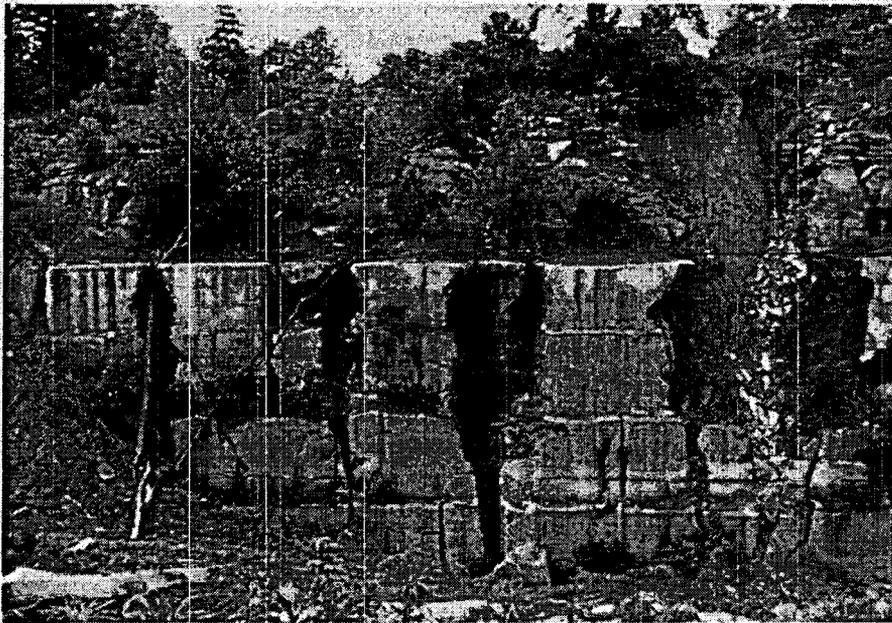


FIGURE 7-22
Cutters in quarry face of Salem limestone (Mississippian), Bedford, Indiana. (Photo by A. N. Palmer)



FIGURE 7-23
Solution of limestone, Gullin, China. (Photo by W. K. Hamblin)

stone, is let down onto the underlying limestone as the streams incised their channels. As the stream channels encounter the limestone, water becomes diverted underground through solution cavities, and the surface streams eventually disappear. As the clastic rocks are consumed by erosion, exposing more and more limestone, subterranean drainage becomes dominant over surface drainage, and the original fluvial drainage pattern ultimately vanishes. However, as in the case of the overlying clastic beds, solution and erosion eventually consume the limestone, and the topography returns once again to a normal fluvial system.

A variation of this scenario occurs when clastic segments of sedimentary sequences, whose weathering products must be mechanically transported, overwhelm subsurface streams, blocking underground channels, forcing streams to flow over a mantle of insoluble alluvium, and once again reverting to surface drainage.

Surface streams in a region of extensive sinkhole development rarely flow for long distances. Typically, surface streams that flow on noncarbonate rocks disappear into sinkholes or swallow holes once they reach carbonate rocks, and the drainage is diverted underground (Palmer, 1972). Such streams are known as **sinking creeks** or **disappearing streams** (Figures 7-24 and 7-25). The point where a sinking stream disappears underground is known as a **swallow hole** (or **swallet**). Swallow holes vary greatly in size and shape, and they may be sinkholes, pits, shafts, various forms of solution fissures, or cave entrances.

Karst Valleys

The transition from a fluvial valley to a karst valley takes place as the subterranean drainage system develops in the underlying limestone and surface flow becomes increasingly di-



FIGURE 7-24
Sinking stream, West Kingsdale, Yorkshire, England. Water runs off peat-covered uplands and drops 72 m (235 ft) into a cave through a solutionally widened joint in Carboniferous limestone. (Photo by A. N. Palmer)

verted underground. The valley retains its fluvial characteristics (that is, shape, gradient, and surface channel) throughout the sequence of development as more and more surface flow disappears underground. When all of the surface flow is finally diverted underground, the valley becomes a dry valley or a solution valley, and surface stream flow occurs only during unusually high discharges. The former surface stream channels gradually become erased by further solution, weathering, and mass wasting. Sinkholes develop in the old channel floor, disrupting the former stream-valley profile, to be replaced by an irregular surface pitted with sinkholes. Woolsey Hollow in the upper part of Figure 7-14A shows topography developed on sandstone beds underlain by limestone. Incision of streams into the sandstone produced normal, fluvial, V-shaped valleys, but upon cutting into the underlying limestone, carbonate solution produced sinkholes in the valley floors into which surface drainage was diverted.

Sinking creeks that discharge into sinkholes over a long period of time erode their valley floors below the original level of the sinkholes, leaving the valley ending abruptly against the steep walls of the distal side of the sinkhole, forming a blind valley (Figure 7-25).

Streams emerging from springs in a karst area may also incise valleys and produce the opposite of a blind valley,

namely, a pocket valley in which the surface stream flows away from a blind headwall rather than into it. Pocket valleys typically have steep headwalls kept nearly vertical by spring sapping that undermines the overlying rock. Their size varies depending on the discharge of the springs and on the physical and chemical characteristics of the rocks. Some pocket valleys reach dimensions of 1 km (0.6 mi) wide, 8 km (5 mi) long, and 300 to 400 m (990 to 1300 ft) deep (Sweeting, 1973). Some streams may even emerge from the blind headwall of a pocket valley and flow for some distance on the surface, only to disappear again into the sinkhole of a blind valley.

Caves

Caves are elongate cavities in limestone produced by solution, aided by mechanical erosion of subterranean flowing water (chapter opening photo). They form along paths of greatest groundwater solution and discharge where subsurface flow is great enough to remove dissolved limestone and keep undersaturated water in contact with the soluble cave walls. Cave formation usually requires a network of openings along which water can flow from recharge to discharge areas. Flow conduits that gain increasing discharge with time accelerate in growth, while the growth of others lags behind. As discharge increases, the maximum rate of cave wall retreat typically reaches about 0.01 to 0.1 cm/yr (0.004 to 0.04 in/yr), determined primarily by solution kinetics and nearly unaffected by further increase in discharge (Palmer, 1991). The time required to reach the maximum rate of enlargement varies directly with groundwater flow distance and temperature, and it varies inversely with initial fracture width, discharge, gradient, and partial pressure of CO_2 (Palmer, 1991). Most caves reach sizes large enough for humans to explore in about 104 to 105 years.

The morphology of caves is controlled by:

1. The spatial distribution of carbonate rocks,
2. The location of recharge and discharge points
3. The geologic structure
4. The distribution of vadose and phreatic flow
5. The geomorphic history of the region (Palmer, 1991)

The sizes and shapes of caves are usually best described by their diameter, length, and spatial pattern. The components of a cave system may be broken down into entrances, terminations, passages, and rooms.

Entrances. Subterranean cave systems are usually quite complex, consisting of both continuous and discontinuous cavities, conduits, and smaller openings in the limestone. Many cave systems are interconnected, serving as drainage ways for subsurface movement of water. Some are filled with water, others are dry; some are accessible from the surface, others are inaccessible because no openings reach to the sur-



FIGURE 7-25
Topographic map of karst area. Note the numerous sinkholes, blind valley, sinking stream, and residual knob. (Smith Grove quadrangle, U.S. Geological Survey)

face. The number of accessible caves in most karst regions is only a small percentage of the total system.

The entrance to a cave may be largely fortuitous. The most common entrances are found as swallow holes, spring mouths, upward sloping of cave passages, sinkhole collapse, intersection of vertical shafts with cave passages, or simply incision of surface valleys that intersect underlying solutional cavities connected to the cave. However, the entrance need not be related to the sources and discharges of the water that formed the cave. Some cave entrances are formed by the downcutting of surface streams, which uncover subsurface cavities. However, truncation and dissection may not form entrances if thick soil covers or weathered bedrock slump into and obliterate the opening. Other entrances are formed by processes that randomly uncover subterranean passages. Most caves have no natural entrance at all and remain virtually undetected from the surface.

Terminations. Even the longest of caves, such as Mammoth Cave, Kentucky, with over 50 km (30 mi) of continuous subterranean passages, eventually end. They may be terminated by collapse of the ceiling, thus filling with debris, or by gradual tapering of the conduit to smaller and smaller sizes. Because virtually all karst caves are made by solution of carbonates, the water that created a cave must exit somewhere.

The courses of subsurface karst streams generate profiles and gradients generally similar to those of normal surface streams. The subsurface drainage may be controlled by structural and stratigraphic conditions, but underground streams seem to emerge from the karst areas at elevations not unlike those of surface streams in drainage basins.

Passages and Rooms. The conduits of a cave system are so intricate as to almost defy description. Many long, winding passageways branch into cavities that enlarge into rooms and chambers, or into a maze of anastomosing corridors and channels along intersecting joint and bedding systems. Such complex systems would be difficult enough to follow in two dimensions, but in karst areas, multiple levels of cave networks are common, substantially increasing the degree of complexity. Figure 7-26 shows some of the nomenclature used to distinguish various patterns.

Passages are segments of caves that are longer than they are wide or high. Thus, caves can be thought of as an assemblage of passages of different sizes and shapes. They can be divided into two broad categories—single-conduit passages and maze passages—each with further subdivisions (Palmer, 1975; White, 1960) (Figure 7-26):

Single-conduit Passages

Linear passages
Angulate passages
Sinuous passages

Maze Passages

Network mazes
Anastomosing mazes

Spongework mazes

The type of maze formed depends on the nature of fracture systems in the limestone and groundwater recharge (Palmer, 1975). Rooms are enlarged conduits of various size caused by

1. The intersection of several passages
2. The coalescence of cavities due to localized vigorous solution
3. The collapse of several cavities

Cave Patterns. Palmer (1991) describes caves as branchwork, network, anastomotic, spongework, and ramiform. The pattern formed depends primarily on the type of groundwater recharge.

Branchwork caves consist of passages that join downstream as tributaries, much like surface dendritic stream patterns. They form where groundwater recharge is from sinkholes, and they are by far the most common cave pattern (Palmer, 1991). Each branch serves as a channel for water from a separate recharge source, and each converges into successively higher-order channels downstream.

Maze caves, characterized by many closed passage loops, form where steep passage gradients and undersaturation of CO₂ allow many passages to enlarge at the same time.

Angular network caves are developed from angular systems of intersecting fissures shaped by the solution widening of principal fractures in carbonate rock. Cave passages are relatively straight, high, and narrow with widespread closed loops. Some simple angular networks consist of a set of angular, discontinuous fissures with infrequent closed loops.

Anastomotic caves consist of curving passages that coalesce and separate, much like the pattern of a braided stream, with many closed loops. They typically form along two-dimensional planes, usually parallel to bedding planes or low-angle fractures. Three-dimensional anastomotic cave systems sometimes form along more than one geologic structure, but they are relatively rare.

Spongework caves consist of three-dimensional, linked, solution cavities of diverse size in a more or less haphazard pattern, resembling a sponge. They typically form by the random joining of individual irregular solution cavities, commonly associated with mixing of chemically diverse water sources.

Ramiform caves consist of irregular, three-dimensional passages, rooms, and galleries, shaped like uneven splotches in plan (Palmer, 1991). Interconnected branches spread outward from the principal solution cavities, grading imperceptibly into spongework and network caves. Sequential outward branches of ramiform caves form by rising hydrothermal water enriched in H₂S.

Karst Windows

Caving in of portions of the roof of subterranean streams creates karst windows through which some of the underground stream is visible from the surface. When all but a few remnants of the roof of a cave have collapsed, the remaining portions form natural bridges.

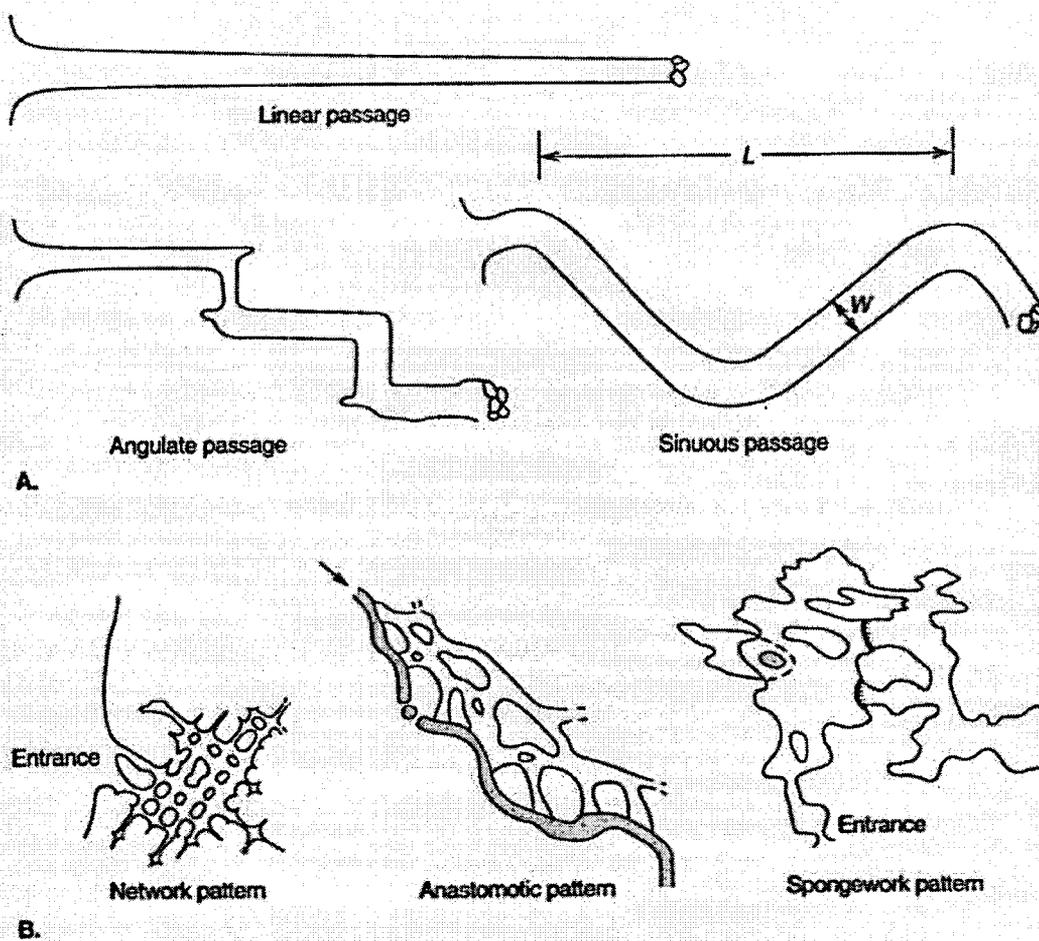


FIGURE 7-26
Types of caves: (A) single-conduit caves with linear, angulate, and sinuous patterns; (B) maze patterns, showing network, anastomotic, and spongework patterns. (From W. B. White, 1988, *Geomorphology and Hydrology of Karst Terrains*. Copyright 1988 by Oxford University Press, Inc.)

Origin of Karst Cave Systems

Theories of the origin of karst caves have evoked strong disagreements from the earliest days of karst investigation, and after more than 100 years of debate, the issue remains in doubt. Explanations for the origin of caves in the classic karst areas of Europe date back to the 1800s. Following the earlier work in the mid-1800s, great interest was displayed around the turn of the century (Cvijic, 1893; Penck, 1894, 1900, 1904; Grund, 1903; Katzer, 1909). Some of the controversy began to emerge even then (Roglic, 1972; Watson and White, 1985; White, 1988). Grund (1903) thought that caves were developed in the zone of active water circulation above the regional water table and that little if anything happened below the water table. A contrasting view was taken by Katzer (1909), who believed that cave development was independent of the water table, a view also favored by Martel (1921). Many of the earlier European ideas were applied to the karst areas of North America in the 1920s and 1930s (Martel, 1921; Malott, 1921; Davis, 1930; Piper, 1932; Gardner, 1935; Swinerton, 1932), leading eventually to several contrasting models of cave evolution.

The principal theories of cave evolution in carbonate rocks all center on the relationship of circulating groundwater to the water table (Figure 7-27):

1. Caves form above the water table by solution by vadose water.

2. Caves form beneath the water table by deep circulation of phreatic water.
3. Caves form at the water table or in the shallow phreatic zone, often associated with fluctuations of the water table itself.

The concept of cave formation above the water table by vadose water arose in the early part of the twentieth century. Caves were thought to be created by subsurface streams flowing at or above the water table as infiltrating water descended downward toward the water table (Martel, 1921; Malott, 1921; Piper, 1932; Gardner, 1935). In addition to carbonate solution by vadose water, subterranean streams were thought to enlarge caves by abrasion (Martel, 1921). Malott (1929, 1932) and Woodward (1961) emphasized the role of solution by vadose water, and Gardner (1935) stressed the importance of water percolating down the dip of limestone beds.

The idea of the origin of caves by deep circulation of phreatic water dates back to the work of Cvijic (1893) and Grund (1903); it was later developed by Davis (1930, 1931) into a two-cycle process. Davis (1930) and later Bretz (1942, 1949, 1953) were impressed by the amount of travertine recently deposited in caves, and they inferred that such caves must have formed during an earlier phase of solution below the water table and had only recently risen above the water table where active carbonate deposition had begun. Davis envisioned that the solution of limestone caves occurred along curving flow lines deep below the surface in the phreatic zone

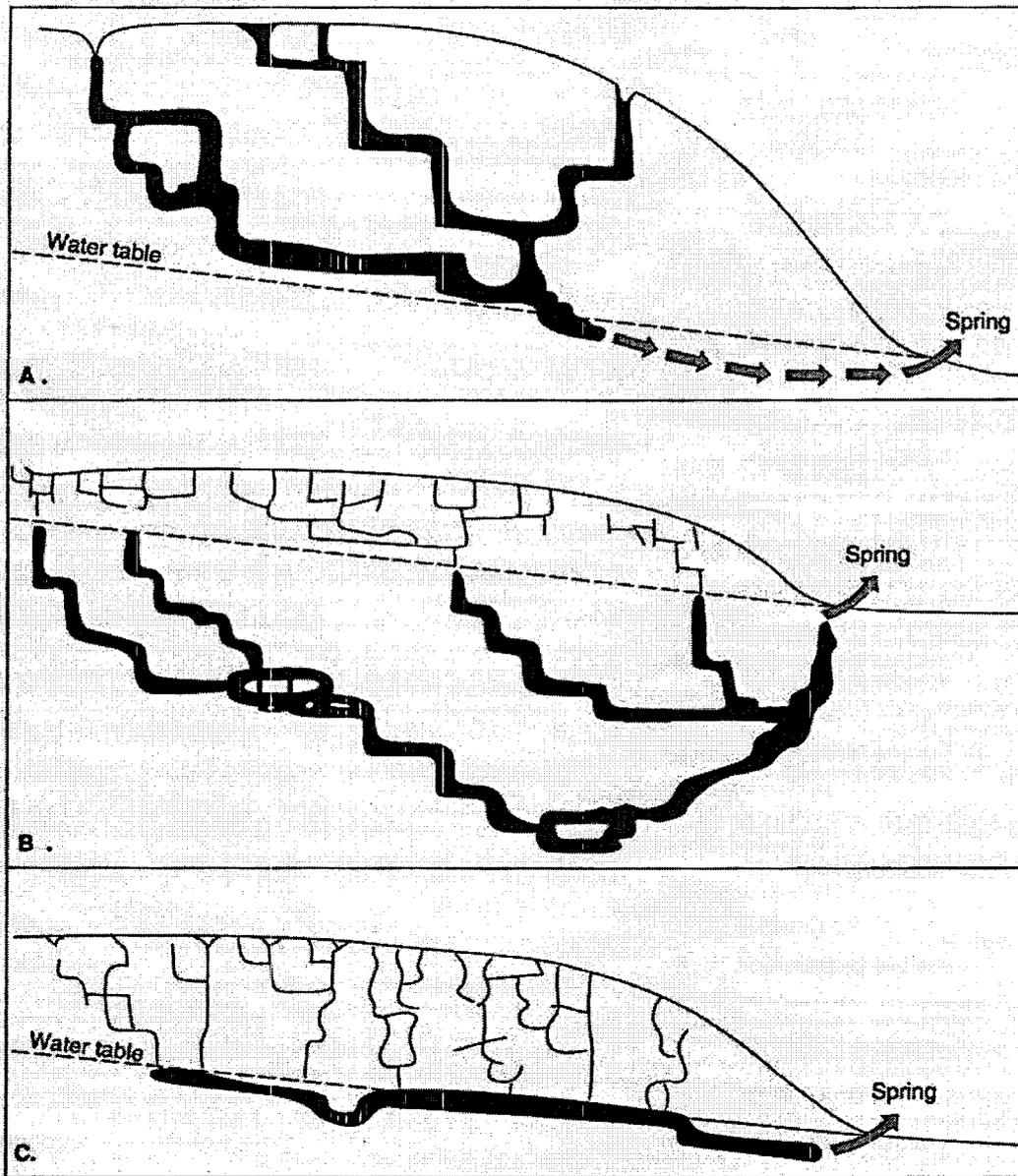


FIGURE 7-27
Zones of maximum cavern development, according to the (A) vadose, (B) deep phreatic, and (C) shallow phreatic theories. (From W. B. White, 1988, *Geomorphology and Hydrology of Karst Terrains*. Copyright 1988 by Oxford University Press, Inc.)

(Figure 7-27), with discharge emerging under the main river in the area. Subsequent rejuvenation of the region by uplift or base-level change causes incision of surface streams and lowering of the water table, leaving the caves high and dry to begin accumulating dripstone.

The shallow phreatic theory of cave formation calls for limestone solution within a zone where the water table fluctuates seasonally. Swinnerton (1932) noted that a great many caves are nearly horizontal and some seem to be stacked, one upon another. The common occurrence of horizontal caves in highly folded limestones suggests that they must be related to the water table. Swinnerton believed that cave development took place in the shallow phreatic zone between the high and low positions of the water table as it fluctuated with the season. Support for this origin of caves came from Sweeting (1950), White (1960), Wolfe (1964), and Thrailkill (1968). Davies (1960) proposed a four-stage, shallow phreatic origin of caves, beginning with random solution at depth, proceeding to integration of solution fissures and cavities

along the top of the water table, followed by deposition of clastic material and uplift and erosion to raise the cave above the water table.

Thrailkill (1968) noted that water table flow can be nearly horizontal and that water just beneath the water table flows along very shallow flow routes. Water near the water table surface mixes with vadose water and becomes under-saturated when it is cooled (dissolves more CO_2) or when new vadose water is added from back-flooding of nearby streams. Supporting evidence for this scenario came from White (1960), Davies (1960), White and White (1974), and Sweeting (1950), who showed that some caves seemed to be related to the elevation of nearby river terraces. White (1988) suggests that many caves show evidence of shallow phreatic solution with nearly horizontal caves even in highly folded limestones, and that three-dimensional cave networks typical of deep phreatic circulation are rare.

In contrast to the concept of cave solution in the phreatic zone and later enlarged by vadose water, Palmer (1991)

pointed out that presently active solution in caves is well adjusted to surface recharge and is not dependent on an earlier phreatic stage. Cave development proceeds in the vadose and phreatic zones at the same time, rendering the arguments of Davis (1930), Swinnerton (1932), and Bretz (1942) irrelevant.

Cave Deposits. Calcium carbonate dissolved by groundwater can be precipitated in caves as travertine, often called **dripstone**. The solubility of calcium carbonate depends largely on the CO_2 content of the groundwater, which is controlled by climate, temperature, nature of the soil, vegetative cover, and other variables. As water enters the cave from fractures in the ceiling and drips toward the floor, some of the water evaporates, increasing the concentration of calcium carbonate and changing the partial pressure of CO_2 . When supersaturation of calcium carbonate occurs, a small amount is precipitated. Each ensuing drop precipitates a little more calcium carbonate, which builds up as an icicle-shaped form, known as a **stalactite**, hanging down from the ceiling (Figure 7-28). The same thing happens to the water that drips off the ceiling to the floor directly below, precipitating additional calcium carbonate. Thus, stalactites are usually matched by similar forms, known as **stalagmites**, growing up from the floor. Given sufficient time, stalactites and stalagmites eventually meet and grow together to form columns. Pools of water that collect on the cave floor evaporate, depositing calcium carbonate as **travertine terraces**.

Tropical Karst

The common karst landforms associated with limestone terrains in humid temperate climates contrast rather sharply with the typical karst topography developed in humid tropical climates (Cooke, 1973; Day, 1976, 1978; Monroe, 1960, 1968). Although virtually all landforms recognized in karst regions of temperate climates also occur in tropical karst terrains, some notable differences exist. One of the more striking is that in tropical karst areas, residual hills, rather than the sinkholes so typical of temperate karst, dominate the topography. Residual steep-sided, cone-shaped hills are separated by cockpits, giving rise to **cone karst** and **tower karst** (Lehmann, 1936; Sweeting, 1958; Gerstenhauer, 1960; Wilford and Wall, 1965; Jennings, 1971; McDonald, 1976, 1979; Monroe, 1960, 1976; Silar, 1965; Williams, 1987). The slopes of towers may be steep-sided to vertical or overhanging, and the towers may stand several hundred meters above the adjacent lowland. The steepness of the residual hills of tower karst distinguishes it from **cockpit karst** (Figure 7-29 and 7-30), which is dominated by depressed cockpits.

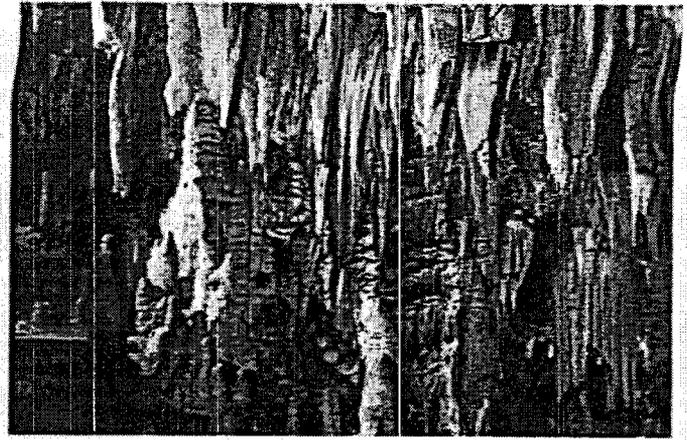


FIGURE 7-28
Stalactites hanging from the ceiling of a limestone cave in the Spanish Pyrenees.

Cone karst and tower karst are common in Central America, China, and parts of the South Pacific. The cone karst and tower karst of Cuba, Puerto Rico, and the Dominican Republic are characterized by short, stubby hills known as **mogotes** (Monroe, 1976). The exact mechanism for the development of mogotes rising above plains of blanket sand is not well understood. Differential solution of limestone, the collapse of caves, and stream erosion have all been suggested (Lehmann, 1954; Nunez-Jimenez, 1959; Panes and Stelcl, 1968; Monroe, 1976; Miotke, 1973). The mogotes of Puerto Rico are composed of the same limestone that underlies the blanket sands of the adjacent plains. However, the mogotes seem to have been more strongly indurated by calcium carbonate cementation.

Extensive tower karst occurs in southern China, where individual towers are relatively narrow with vertical sides. In southeast Asia, tower karst occurs in Java, the Celebes, Malaysia, New Guinea, and Borneo (Lehmann, 1936; Sunartadirdja and Lehmann, 1960; McDonald, 1976; Williams, 1971; Verstappen, 1964).

Cockpits are large, bowl-shaped sinkholes up to about 1 km (0.6 mi) in diameter that occur in thick limestones of tropical climates (Figure 7-30). The name comes from the Cockpit region of Jamaica where a 1000-m (3300-ft) thick limestone has been eroded into a karst belt 30 km (19 mi) wide and 90 km (56 mi) long (Sweeting, 1958). Cockpits are so large that secondary channel systems commonly develop on the cockpit walls, giving them a star-shaped form, in contrast to the circular or elliptical shapes of most sinkholes. Individual cockpits are typically separated by sharp ridgelines, and the rocks left at the intersection of several cockpits form pyramid-shaped hills.

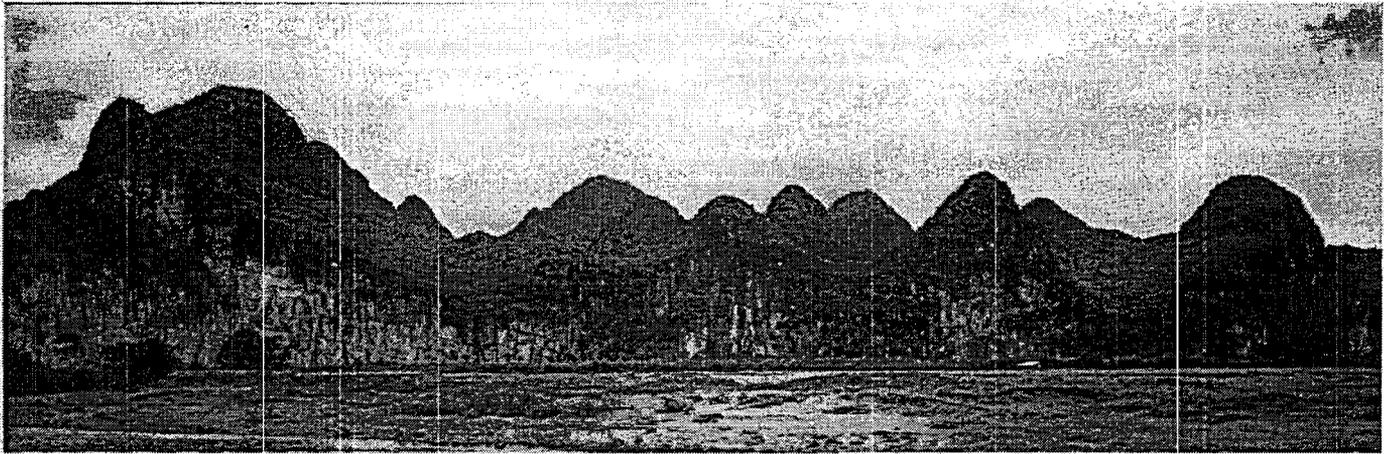


FIGURE 7-29
Tower karst, Guilin, China. (Photo by W. K. Hamblin)

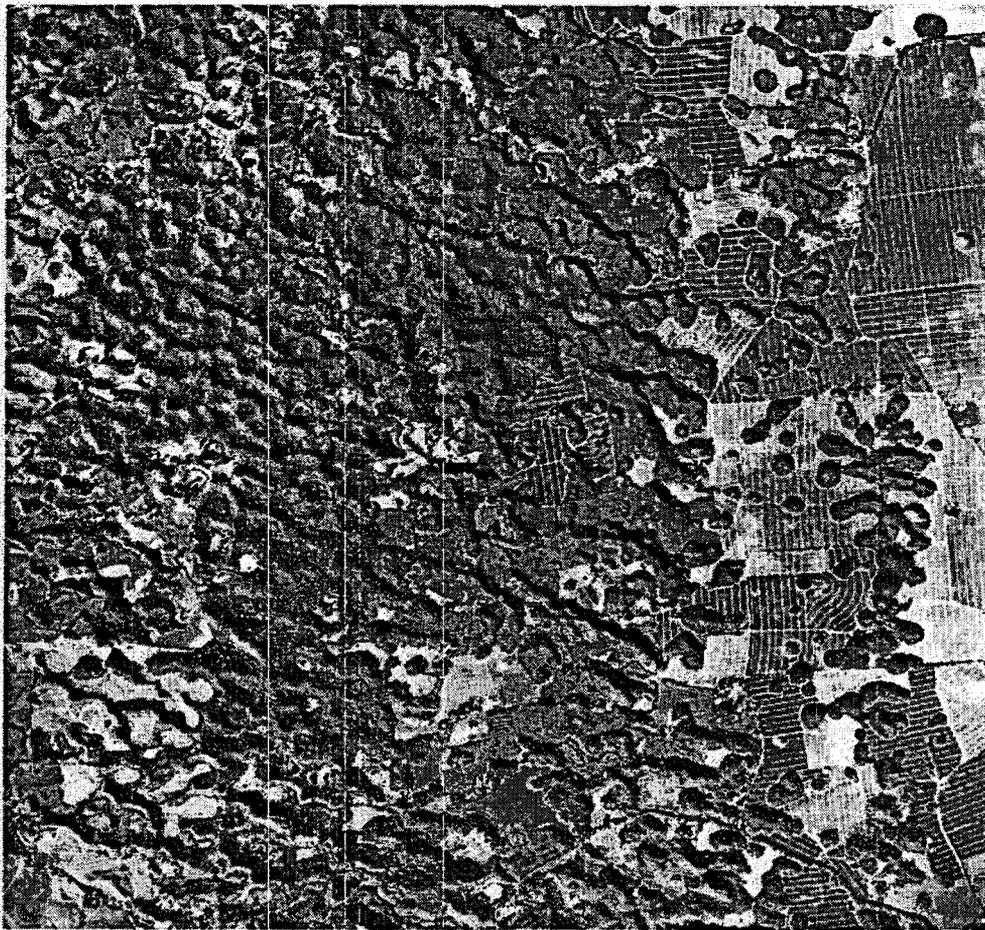


FIGURE 7-30
Cockpit karst, Puerto Rico.
(Photo by U.S. Geological Survey)

ARCHITECTURE OF AIR-FILLED CAVES WITHIN THE KARST OF THE BROOKSVILLE RIDGE, WEST-CENTRAL FLORIDA

LEE J. FLOREA

Department of Geology, University of South Florida, Tampa, FL 33620, lflorea@chuma1.cas.usf.edu

Air-filled caves surveyed in the Brooksville Ridge of west-central Florida provide insight into the organization of karstic permeability within the unconfined portions of the Upper Floridan Aquifer. The morphology of the passages that compose these caves in geologically young, high-permeability limestones is strikingly different from caves found in ancient carbonates far from the influence of the coast. Cave passages in west-central Florida are laterally extensive and tiered. Principal horizons of cave development occur between +3 m and +5 m, +12 m and +15 m, and +20 m and +22 m above modern sea level. The primary guide of cave passage orientations within these cave levels is widespread fractures oriented approximately NE-SW and NW-SE. Cave passages of human dimensions form at the intersection of the laterally extensive cavities and fractures and often acquire a characteristic plus-sign shape. The walls of cave passages in west-central Florida are porous and complex, with small-scale solution features such as pockets and tafoni structures extending into the host bedrock. Additionally, these cave passages often end in blind pockets, ever-narrowing fissures, sediment fills, and collapses. The passages do not appear to represent an integrated system of conduits between aquifer inputs and outputs.

INTRODUCTION

Cavers and karst scientists have long appreciated and recorded information concerning the morphology of passages in caves. These data about caves are important to understanding the flow of water in karst aquifers, which cover approximately 15% of the land surface and provide water to approximately one-fifth of the world's population (Ford and Williams, 1989). For example, compilations of cave maps reveal patterns in both the organization of passages in a cave and the shape of individual passage cross-sections that are a direct consequence of hydrogeological conditions within karst aquifers (e.g., Palmer, 2000; White 1988). To date, these observations are drawn primarily from experiences in caves formed far from the influence of the coast and within ancient carbonate rocks that are remarkably different from carbonate rocks that are geologically recent or are forming today. This paper presents a case study of the morphology of caves within the coastal karst aquifers of west-central Florida.

GEOLOGIC FRAMEWORK OF THE BROOKSVILLE RIDGE AND THE UPPER FLORIDAN AQUIFER

The Tertiary limestones that compose the highly productive Upper Floridan Aquifer are intensely karstified in regions that experience active groundwater circulation (e.g., Lane, 1986; Stringfield and LeGrand, 1966), particularly in the portion of west-central Florida where the Upper Floridan Aquifer is semi-confined to unconfined. This region, characterized by 33 springs with average discharge greater than $2.8 \text{ m}^3 \text{ s}^{-1}$ (e.g., Scott *et al.*, 2004; Roseneau *et al.*, 1977; Meinzer, 1927), stretches from the panhandle near Tallahassee in the north to Tampa in peninsular Florida (Fig. 1A) and encompasses several physiographic provinces including the Brooksville Ridge (White, 1970).

The Brooksville Ridge, a linear, positive-relief topographic feature extending from northern Citrus County, through Hernando County, and into southern Pasco County (White, 1970), is bounded by coastal lowlands to the west and south and wetlands of the Withlacoochee River to the east and north. The ridge system is a consequence of a localized geologic high termed the Ocala Platform by Scott (1988), who attributed this topographic feature to a westward tilt of thickened Eocene strata. Elevations in the Brooksville Ridge range from five to more than 75 m above sea level (Fig. 1B). The topography is rolling with internal drainage (Fig. 2). Upland mesic-hardwood hammocks separate sinkhole lowlands that are mostly occupied by wetlands or lakes. The Withlacoochee State Forest manages more than 525 km² (157,000 acres) in the region, including the 100-km² (30,000 acre) Citrus Tract that includes much of the study area. Pasture land and lime-rock quarries compose the remaining land uses. The city of Brooksville lies in the heart of the Brooksville Ridge (Fig. 1A).

Upper-Eocene and Oligocene carbonates (42–33 Mya) compose the Upper Floridan Aquifer, which is semi-confined to unconfined in the Brooksville Ridge. The strata of the Upper Floridan Aquifer thicken to the south along a regional dip that averages less than half of one degree (Scott *et al.*, 2001; Miller *et al.*, 1986). Miocene-age sands and clays of the Hawthorn Group thicken to more than 150 m in northern and southern Florida where the Upper Floridan Aquifer is confined (Scott, 1988). The Hawthorn Group is thin to missing in the center of the Brooksville Ridge in northern Hernando and southern Citrus Counties (Fig. 3).

The Suwannee Limestone, a pale-orange, partially recrystallized limestone that is extensively quarried in northern Hernando County, is more than 30 m thick to the south. In the up-dip sections of the northern Brooksville Ridge of Citrus County, the Suwannee Limestone is thin to nonexistent as a

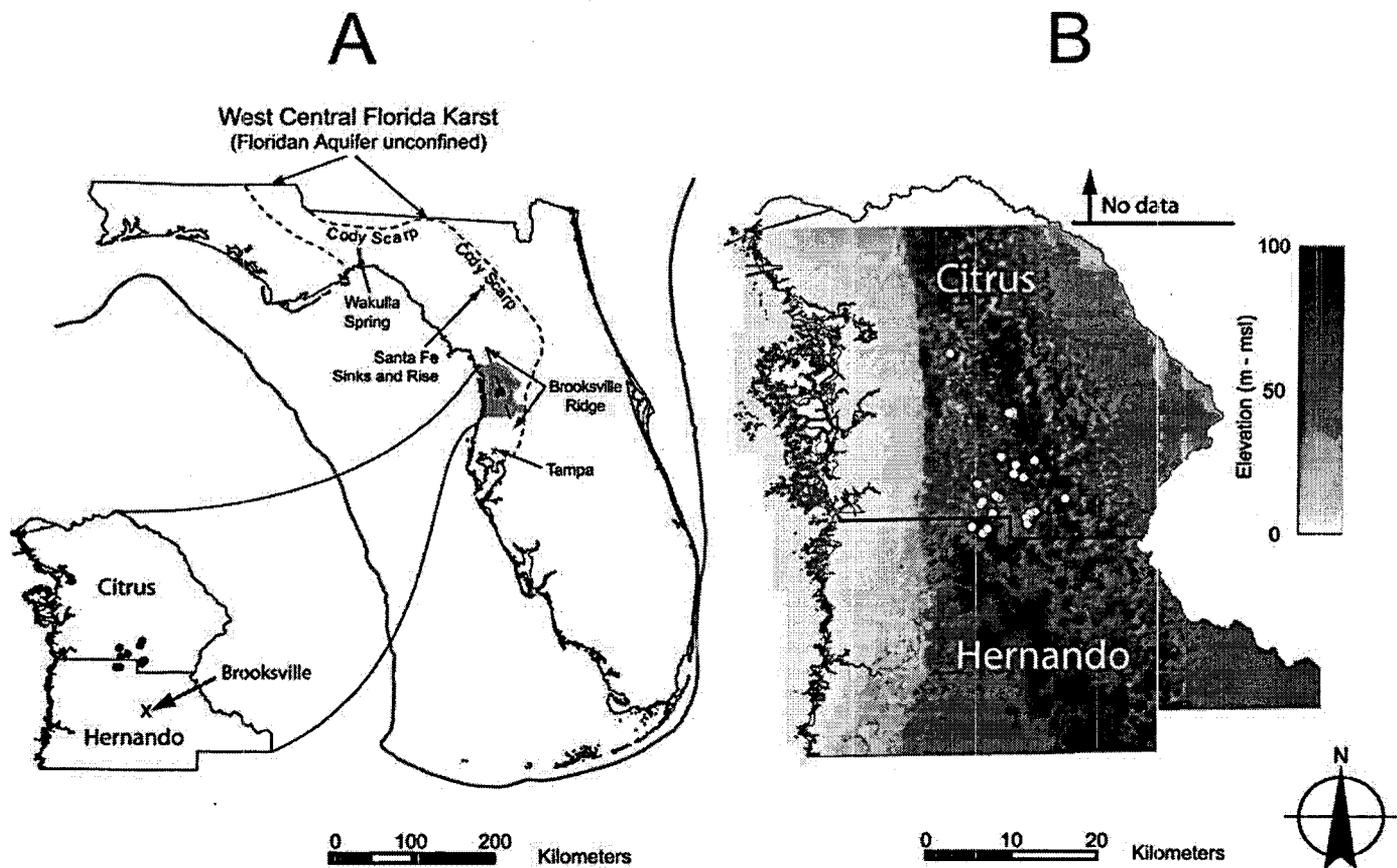


Figure 1. Data locations and topographic elevations. A) The grey line surrounding Florida is the -120 m bathymetric contour on the continental shelf. Inset is included for Citrus and Hernando Counties. Air-filled caves surveyed in this study are indicated by black dots. An “x” indicates the location of the city of Brooksville. B) Elevations for the Brooksville Ridge in Citrus and Hernando Counties are generated using GIS topographic data. Known air-filled caves in the Brooksville Ridge are indicated by white circles.

result of post-Oligocene exposure and erosion (Yon and Hendry, 1972). As a result, the Suwannee Limestone is thickest beneath the topographic highs and missing in many topographic lows (Yon *et al.*, 1989). Paleokarst filled with Miocene-age siliciclastics pierces the Suwannee Limestone throughout the Brooksville Ridge (Yon and Hendry, 1972). These paleokarst sinkholes indicate a period of intense karstification during the end-Oligocene exposure.

An irregular exposure surface with chert lenses, clay-rich marls, and a transition to non-recrystallized limestone marks the boundary between the Oligocene carbonates and the Ocala Limestone of late Eocene age. The Ocala Limestone is cream to white, soft, friable, and very porous in the Brooksville Ridge. It ranges in thickness from 30 m north of the study area to more than 120 m south of the Brooksville Ridge (Miller, 1986). Petrographic investigations of the Ocala Limestone by Loizeaux (1995) demonstrate three 3rd-order cycles of deposition. Shallow-water, high-energy facies, such as cross-bedded, low-mud grainstones and mixed-skeletal packstones, dominate all three cycles of the Ocala Limestone in the Brooksville Ridge.



Figure 2. Gentle rolling topography of the Brooksville Ridge near the city of Brooksville. An upland mesic-hardwood hammock is visible in the background. The foreground is a sinkhole lowland (photo by author).

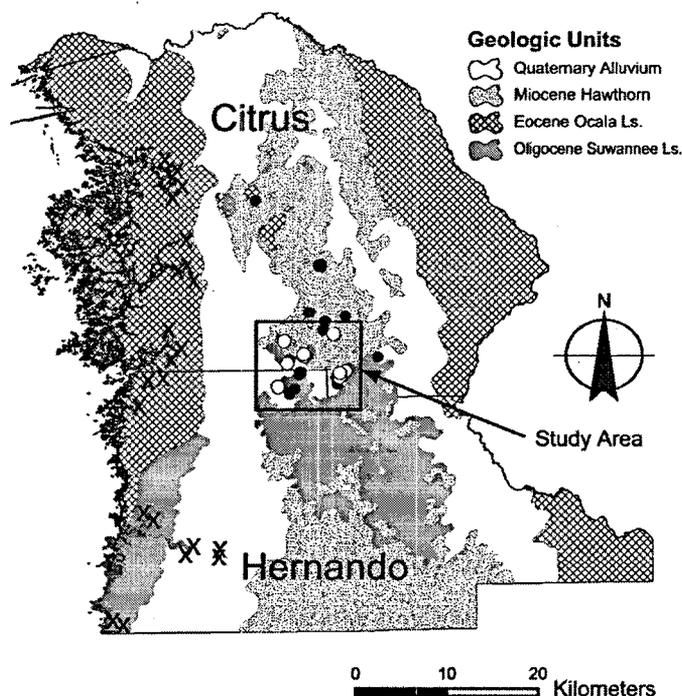


Figure 3. Geologic map of Citrus and Hernando Counties. Geologic units generally dip and thicken to the south. The Miocene Hawthorn Group is thin to non-existent in northern Hernando and southern Citrus Counties. The Oligocene Suwannee Limestone occupies only the topographic highs in the study area. Air-filled caves surveyed in this study are indicated by white circles. Additional air-filled caves known in the region are indicated by black dots. Springs are indicated with a black "X."

The geologically young carbonates of the Upper Floridan Aquifer retain much of their original porosity and permeability, which is highly heterogeneous and facies-dependent (Budd and Vacher, 2004). Measurements during this study from cave and core samples from the Brooksville Ridge indicate that the matrix permeability of the Ocala Limestone averages $10^{-12.7}$ m², which compares to an estimated value of $10^{-17.7}$ m² for the much older Paleozoic limestones of the Mammoth Cave region of Kentucky (Worthington *et al.*, 2000).

KARST OF THE BROOKSVILLE RIDGE

Historically, exploration of air-filled caves in Florida has been concentrated in portions of the panhandle near Florida Caverns State Park (Lane, 1986) and along the Cody Scarp in north-central Florida (*e.g.*, issues of the *Florida Speleologist*, published by the Florida Speleological Society). In west-central Florida, the emphasis of karst research has surrounded the first-magnitude springs concentrated near the Gulf of Mexico (Meinzer, 1927) (Fig. 3). These large springs, such as Weeki-Wachee, Crystal River, Chassahowitzka, and Homosassa, discharge several hundred million gallons of water per day (Scott

et al., 2004). The known underwater caves near these springs, such as Eagle's Nest, Twin-Dees, and Diepolder, are famous in the popular press for their large passages, great depths (in excess of 100 m), and technical diving challenges.

Less is known about the caves within the watersheds of the large springs along the coast in west-central Florida. These watersheds cover hundreds of square kilometers and include portions of the coastal lowlands and the Brooksville Ridge.

In the coastal lowlands, most caves are currently underwater because the depth to the water table is less than 15 m. Thick Quaternary sediments mantle karst features, subduing their surface expression (Tihansky, 1999). In contrast, the depth to the water table exceeds 45 m in the uplands of the Brooksville Ridge, and Quaternary sediments are thin to non-existent. Air-filled caves in the Brooksville Ridge have been known for decades; *e.g.*, the Dames Cave complex of southern Citrus County (Brinkmann and Reeder, 1994). However, there has been only limited exploration or scientific documentation of these caves until this study. The restricted number of natural, human-sized cave entrances contributes to the lack of exploration.

Beginning in 2001, local cave explorers located several previously unknown caves of significant size in the uplands of the Brooksville Ridge (*e.g.*, Turner, 2003). These newly-found caves are the focus of this study. Many of the discoveries were fortuitous; for example, otherwise hidden passages were revealed after structural collapses of cave roofs below abandoned lime-rock quarries. Such air-filled caves provide insight into the architecture of cave-scale porosity in the Upper Floridan Aquifer and greatly expand our perception of karst features in west-central Florida.

DATA COLLECTION

The data for this study are largely from surveys of seven caves within a study area in northern Hernando and southern Citrus Counties in west-central Florida (Fig. 3, Table 1). Maps of additional air-filled caves in the Brooksville Ridge were acquired from the archives of the Florida Cave Survey. The seven surveyed cave sites are in the central portion of the Brooksville Ridge where Miocene siliciclastics are thin and the Suwannee Limestone occupies only the upland hammocks. The Withlacoochee State Forest manages five of the seven sites; private landowners own the other two.

At each of the seven caves, I established elevation control using established data where available or by using an Ashtech Z-Extreme RTK (real-time kinematic) GPS base station and rover unit operated by the Coastal Research Group at the University of South Florida. I used a NOAA-HARN benchmark for our base station. The elevation of each in-cave survey station above mean sea level is based upon these control points. Subsequent survey from the control points, using a fiberglass tape and a hand-held compass and clinometer, is accurate to one-degree per station; this error propagates through the survey. In most of the surveyed caves, the number

Table 1. Caves surveyed in this study.

| Cave Name | County | Length (m) | n _(sta) ^a | n _(az) ^b |
|-------------------|----------|------------|---------------------------------|--------------------------------|
| Big Mouth Cave | Citrus | 96 | 13 | 14 |
| Blowing Hole Cave | Citrus | 257 | 50 | 54 |
| BRC Cave | Hernando | 1,033 | 276 | 281 |
| Football Cave | Citrus | 142 | 29 | 31 |
| Legend Cave | Citrus | 44 | 12 | 12 |
| Morris cave | Citrus | 92 | 12 | 13 |
| Werner Cave | Citrus | 561 | 105 | 115 |
| Totals | | 2,225 | 497 | 520 |

^a Number of survey stations.^b Number of azimuth readings.

of azimuth readings exceeds the number of survey stations (Table 1), because some stations were located at passage junctions where multiple azimuth readings were required to accommodate play shots or loop surveys.

I generated detailed maps of each cave in Adobe Illustrator and ESRI ArcGIS software using a combination of detailed sketches and the cave survey data. These maps were used to assess the overall cave morphology in plan and profile view, including height-width ratios of the passages, length-weighted rose diagrams of passage orientations, and a histogram of all the survey-station elevations.

RESULTS AND ANALYSIS

The data include more than 2.2 km of new cave survey (Table 1). Small-scale maps of the caves are presented in plan view in Figure 4. Of the caves surveyed, BRC Cave is by far the longest with more than a kilometer of mapped passage

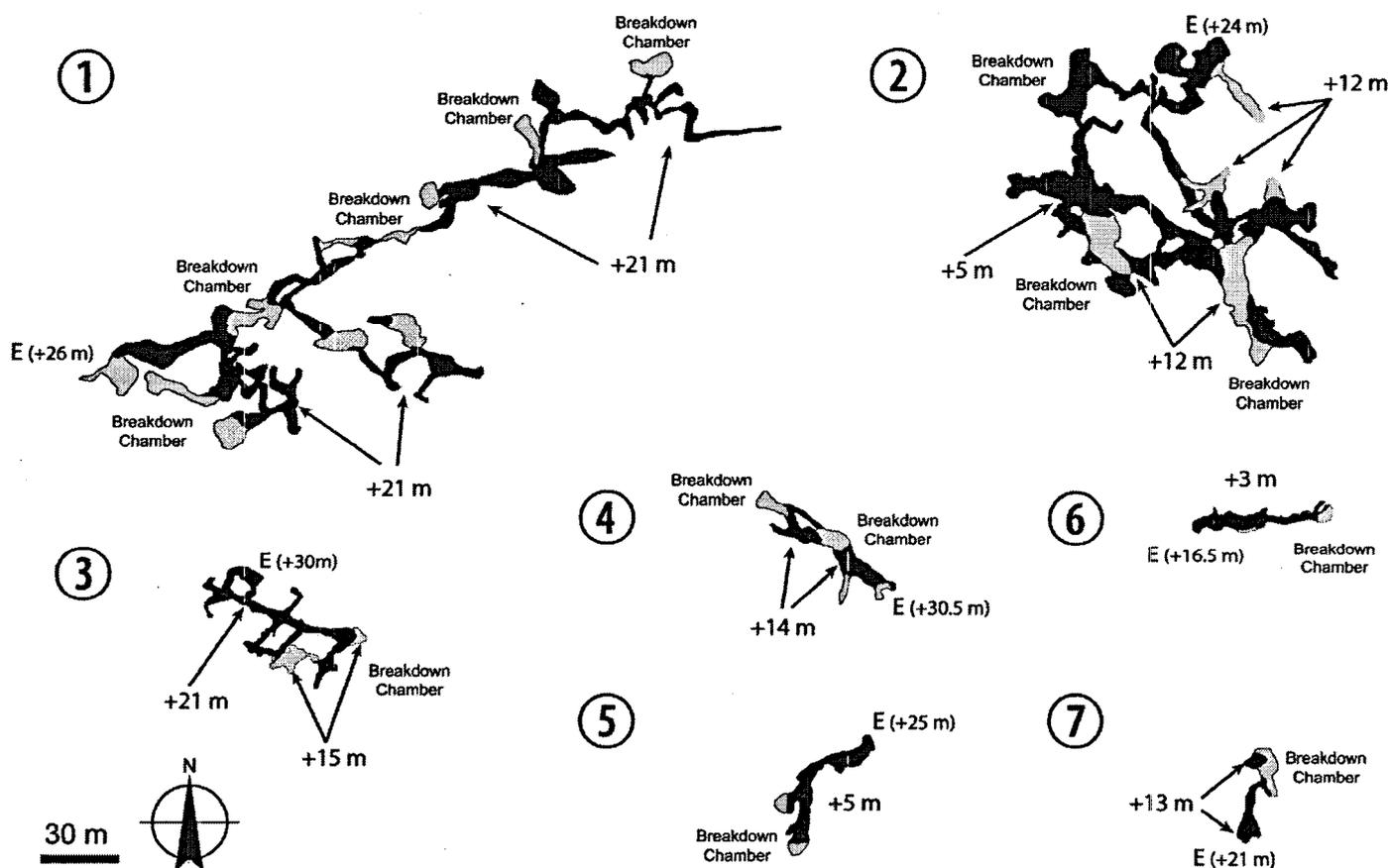


Figure 4. Index maps from air-filled caves surveyed during this study. 1 – BRC Cave, 2 – Werner Cave, 3 – Blowing Hole Cave, 4 – Football Cave, 5 – Big Mouth Cave, 6 – Morris Cave, 7 – Legend Cave. The cave passages occur on distinct levels. For instance, Werner Cave, Big Mouth Cave, and Morris Caves contain passages near the present-day water table between +3 m and +5 m. Werner Cave, Blowing Hole Cave, Football Cave, and Legend Cave all have passages between +12 m and +15 m. BRC Cave and Blowing Hole Cave both have extensive passages at +21 m. The entrances to every cave surveyed are above the level of passage development. Only Blowing Hole Cave and Football Cave have natural entrances that are fractures enlarged by dissolution that are several meters deep. All of the caves surveyed contain collapse features.

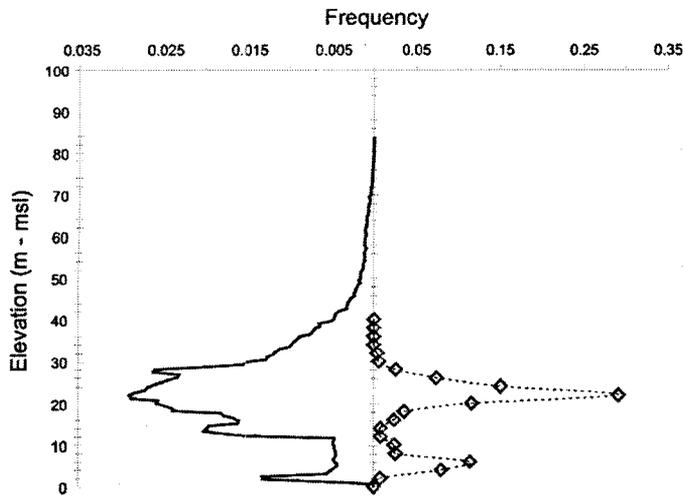


Figure 5. Frequency of data of land elevations in Citrus-Hernando Counties (left) compared to elevations of cave-survey stations in this study (right). Modes in the cave-survey data correspond with modes in the elevation data set from Citrus and Hernando Counties and with known marine terraces.

(Table 1); Werner Cave, (561 m, Table 1), together with Blowing Hole Cave (257 m, Table 1), round out the longest three caves in the study.

The entrances to all seven caves surveyed in this study, as well the entrances to other air-filled caves in the Brooksville Ridge, are at a higher elevation than the level of passages in the cave (Fig. 4). Football Cave and Blowing Hole Cave have natural entrances that are fractures enlarged by dissolution that are several meters deep. The entrance to Legend Cave is a small hole in a rock choke at the edge of a small lime-rock quarry. Werner, Big Mouth, and Morris Caves did not have natural entrances. Rather, a quarry operation intersected structural collapses within the cave.

Figure 5 collects elevation data for all caves surveyed in this study and compares the data to a frequency plot of elevations for Citrus and Hernando Counties from Figure 1B. Figure 6 presents a frequency plot of passage dimensions. Figure 7 presents the length-weighted rose diagrams of passage orientations and compares this data to a similar dataset from 14 caves in Marion County 40-50 km to the north and east of the study area.

Upon first inspection, all of the caves within the study area are strikingly similar in their appearance. For instance, natural solution walls, ceilings, and floors of all caves of the study area, as well as many caves throughout west-central Florida, contain cusped, pocket-like, or even tafoni features (Fig. 8). The passages in the caves of Figure 4 terminate in blind pockets, ever-narrowing fissures, sediment fills, and collapses. Development of cave passages along fractures is visible from cave maps in plan view (Fig. 4), and individual caves demonstrate a preferred orientation of passages (BRC, Werner, and

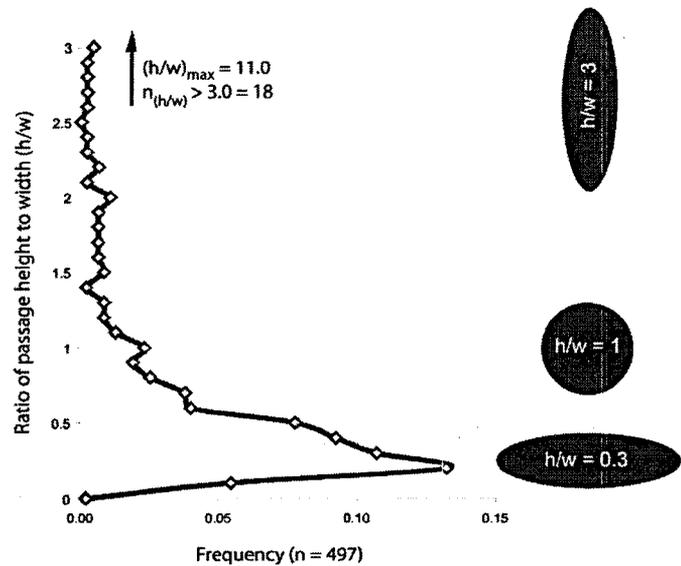


Figure 6. Frequency of passage height-width ratios at all survey stations in this study. Almost 15% of measured passages are more than four-times wider than they are tall, and 47% of measured passages are more than twice as wide as they are tall.

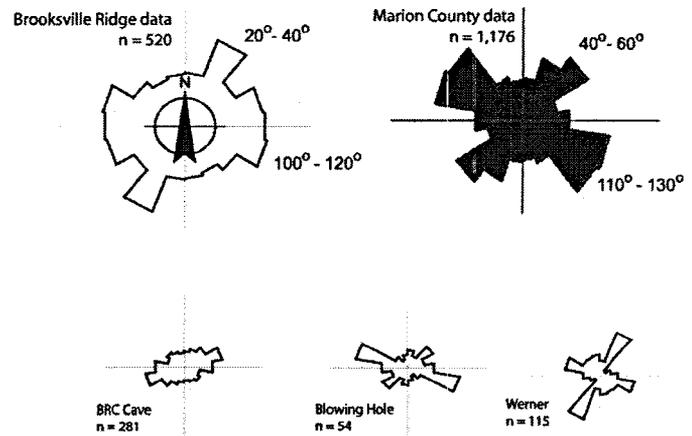


Figure 7. Length-weighted rose diagrams for the orientation of all segments of cave survey obtained during this study and from 14 caves in Marion County to the north and east of the study area. The data from this study reveal a regional WNW-ESE (100°-120°) and NNE-SSW (20°-40°) pattern of passages similar to the data from Marion County. Both are related to regional fracture sets. Individual caves have a preferred orientation to cave passages.

Blowing Hole Caves, Fig. 7). The cumulative length-weighted rose diagram of passage directions reveals a WNW-ESE (100°-120°) and NNE-SSW (20°-40°) pattern of passages (Fig. 7).

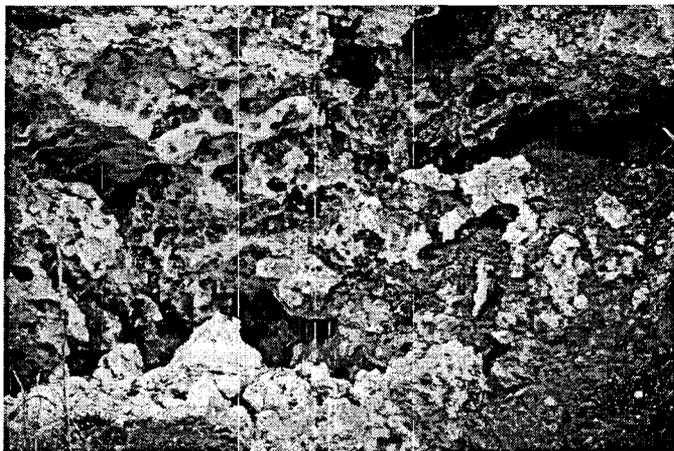


Figure 8. Spongework-like features present in the walls of an exposed cavity in the Haile Quarry near Gainesville, Florida. Height of cavity is approximately 40 cm (Figure 12 of LaFrenz *et al.*, 2003).

Observations from quarry highwalls in the study area and throughout west-central Florida reveal laterally extensive cavities (Fig. 9). These laterally extensive cavities occur at particular elevations throughout the study area (Figs. 4 and 5). The elevations of cave survey stations cluster between +3 m and +5 m and between +20 m and +22 m (Fig. 5) above mean sea level. The individual cave maps reveal a third, less-pervasive level of passages between +12 m and +15 m (Fig. 4) which is not visible in Figure 5 because it is masked by the scatter in the survey data for the higher-elevation peak.

Human-scale passages within these cavities often occur where they intersect fractures enlarged by solution. Each cave presented in Figure 4 is a group of these human-scale cavities.

Passages formed along fractures in the caves of the Brooksville Ridge often develop “fissure” morphologies. In contrast, passages not associated with fractures acquire a “tabular” morphology. The cave-survey data demonstrate the latter

Land Surface ~ 27.5 m

Cavernous Zone
~20.5 m

Water Table ~13.5 m

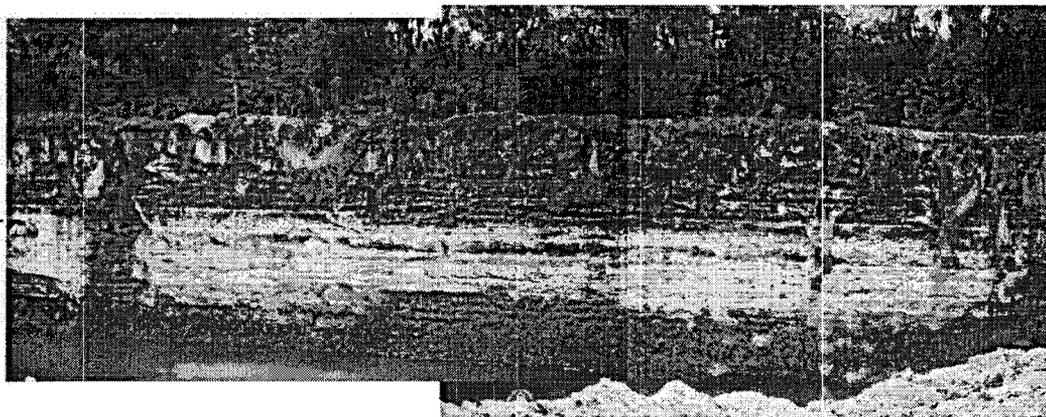


Figure 9. Photo of highwall at Haile Quarry near Gainesville in north-central Florida. The highwall is approximately 14 m tall, and the land surface is approximately 27.5 m above mean sea level. Note the laterally continuous cavernous zone 7 m below the top of the highwall at +20.5 m (Figures 5a and 9 of LaFrenz *et al.*, 2003).

to be more common; 47% of the surveyed stations are more than twice as wide as they are tall (Fig. 6). Commonly, passages combine fissure and tabular morphologies into a signature “plus-sign” cross-section.

DISCUSSION

Caves in the young, high-permeability, coastal karst aquifers of west-central Florida differ substantially from those of the traditional, textbook perspective (*e.g.*, White, 1988; Ford and Williams, 1989) of caves in ancient, low-permeability limestones of inland karst regions. The differences in cave morphology were anticipated by Palmer (2000) and briefly examined using examples of caves from the panhandle and north-central Florida by Palmer (2002).

The common conception of caves within the ancient limestones of the mid-continent is that water generally enters at discrete sites, travels through conduits, and discharges at springs. Caves in these settings have predictable geometries. According to Palmer (2003, p. 2):

Within karst aquifers, most of the dissolution porosity consists of conduits, usually arranged in dendritic patterns in which tributaries join each other to produce fewer but larger conduits in the downstream direction.

In such caves, the porosity tends to form “continuous conduits rather than isolated voids” Palmer (2003, p. 2).

The current perception of karst aquifers in the young carbonates of Florida is similar to this sinking-stream, spring model. For example, when speaking about the evolution of karst landscapes in Florida, Lane (1986, p. 14) states:

Continuing dissolution... will divert more of the surface water into the underground drainage. Eventually, all of the surface drainage may be diverted underground, leaving dry stream channels that flow only during floods, or disappearing streams that flow down swallow holes... and reappear at distant points to flow as springs or resurgent streams.

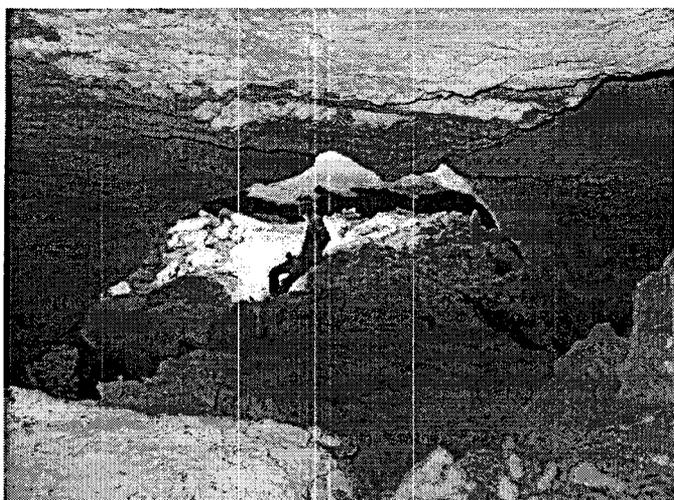


Figure 10. Breakdown chamber in Werner Cave. Such collapse features are common in the caves of the study area. Main level of passages is approximately 3 m below the top of the breakdown (photo by Tom Turner).

Certainly there are many examples of underground river caves in Florida that follow this model. In fact, most major surface streams that cross the Cody Scarp in the Florida panhandle and north-central Florida sink into the Upper Floridan Aquifer (Upchurch, 2002). The water from several of these sinking streams travels through conduits and returns to the surface as major springs (Scott, *et al.*, 2004). Well-studied examples include the Santa Fe River Sinks and Rise (Martin and Dean, 2001) and the Wakulla-Leon Sinks Cave System (Loper *et al.*, 2005; Lane, 1986).

On the other hand, the Cody Scarp is just one physiographic feature in an otherwise large karst region, and the underground river caves associated with the Cody Scarp account for only a small fraction of the nearly 1,500 known caves in the current Florida Cave Survey database. The Brooksville Ridge is not related to the Cody Scarp and it contains many caves that are not of the underground river type. What do the caves in the Brooksville Ridge look like? How do they differ from the caves of the mid-continent, and what do these caves reveal about the hydrogeology of the Upper Floridan Aquifer in west-central Florida?

To answer these questions, I will inspect the cave architecture documented from my cave-survey data from four viewpoints: passage cross-section, directionality, horizontality, and connectivity.

PASSAGE CROSS-SECTION

Many passages in the caves of the Brooksville Ridge and throughout west-central Florida are wider than they are tall (Fig. 6). These low, wide cavities can be laterally extensive (Fig. 9). Interspersed in the tabular voids created by the later-



Figure 11. Plus-sign passage in Roosevelt Cave in Marion County, Florida. Note that the vertical extension of the passage visually correlates to a fracture. Also note the laterally continuous horizon of passage approximately 1 m above the water table (photo by Sean Roberts).

ally extensive cavities are pillars of rock that have not dissolved (Fig. 4). As in an underground coal mine, these pillars hold the ceiling intact. Structural collapse of the ceiling is common between these rock pillars, predominantly where rock pillars are widely spaced or where ceiling blocks are bounded by fractures. These collapses are a mixed blessing to exploration, because, while they often create large rooms in the otherwise low, wide cave (Fig. 10), they also impede progress by blocking access (Fig. 4) to cave beyond the breakdown.

Tall, narrow passages in the caves of the Brooksville Ridge and throughout west-central Florida are always associated with fractures. Human-scale passages commonly occur where fractures and the laterally extensive cavities intersect, producing a characteristic plus-sign passage morphology (Fig. 11).

Walls of the cave passages in this study have complex, small-scale solution features (Fig. 8). These cusped, pocket-like, or tafoni structures are an indication of water-filled conditions during at least part of the cave-forming period. However, these features are not flow indicators as are scallops in caves within ancient carbonates of the mid-continent. Rather, they closely resemble spongework features found in the caves of young carbonate islands such as in the Bahamas (Myroie *et al.*, 1995) and caves of hypogenic settings such as in the Guadalupe Mountains of New Mexico (Hill, 1987).

PASSAGE DIRECTIONALITY

Caves in west-central Florida, regardless of cross-section, exhibit a preferred orientation of passages along fractures in the aquifer (Figs. 4 and 7). The datasets from the Brooksville Ridge and from Marion County are similar; both generally reveal a regional NW-SE and NE-SW pattern of passages.

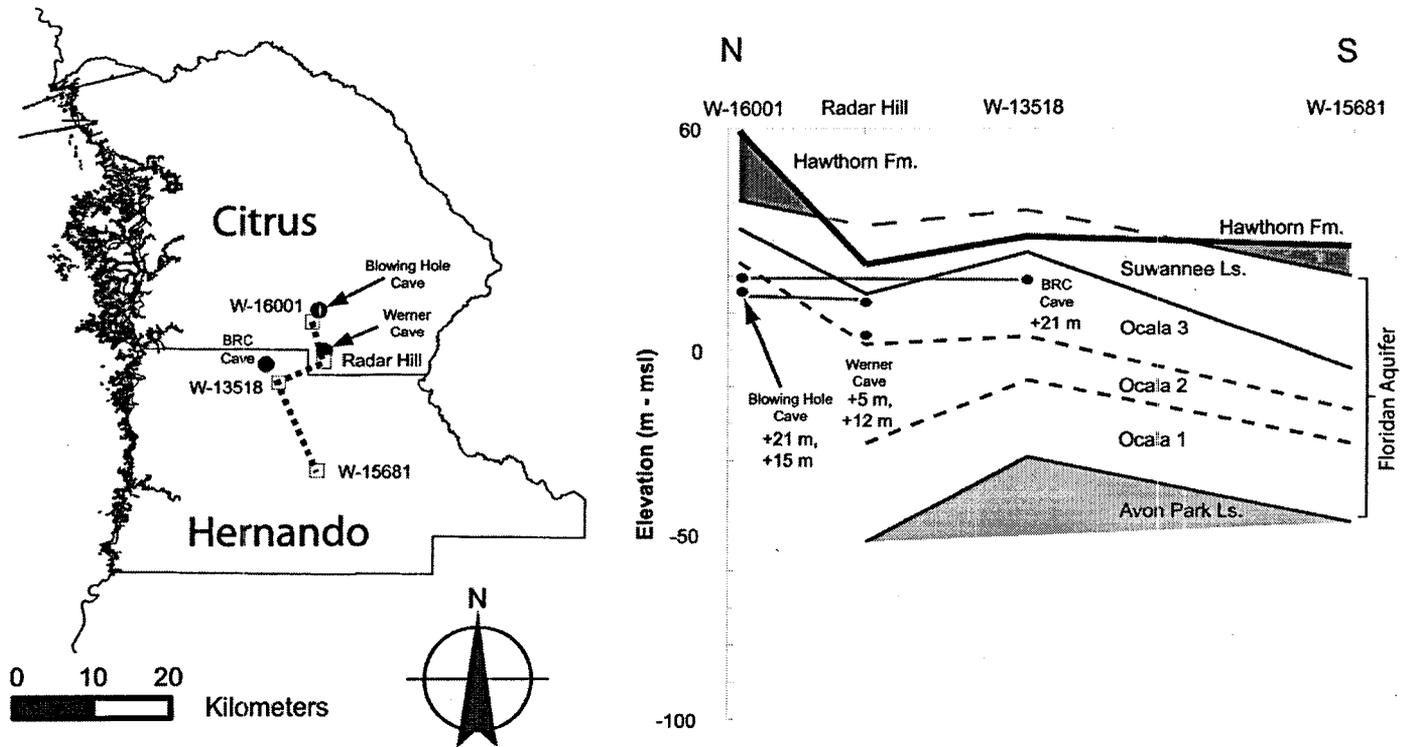


Figure 12. North-south cross-section through the study area in the Brooksville Ridge. Dashed line on the map at left indicates the location of the cross-section. White squares are the wells used for lithologic identification. Black dots are the caves from this study near the line of cross-section. Note that the levels within these caves do not occur in the same geologic units throughout the study area.

Vernon (1951), who looked at topographic and physiographic features (such as linear segments of the Withlacoochee River), and Littlefield *et al.* (1984), in a detailed study of sinkhole alignments in west-central Florida, identified a large number of photo-linear features attributed to fractures that follow this NW-SE and NE-SW pattern. The widespread nature of this pattern is a manifestation of a pervasive cause of the fractures that is not yet identified.

Individually, the rose diagrams of passage orientations vary amongst the caves in the study area and in the caves in Marion County (Fig. 7). However, these data do not provide credible evidence that explains the reason for the variation. For instance, it is unclear whether the passages surveyed in a particular cave are a representative subset of all passages in the vicinity of that cave. What is clear is that the passages are some measure of the anisotropy of the aquifer at the time the cave formed.

PASSAGE HORIZONTALITY

Cave passages in west-central Florida are not only laterally expansive, they occur at particular elevations much like the levels of cave passages within ancient limestones, such as at Mammoth Cave (Palmer, 1987). At Mammoth Cave, cave levels formed near the water table as the elevation of the Green

River experienced staged base-level lowering during glacial-interglacial cycles (Granger *et al.*, 2001). In Florida, the origin of cave levels may also result from changing positions of the water table, but one must also consider the role of lithology and, more specifically, variations in matrix permeability.

This second option, variations in matrix permeability, is often ignored in the study of caves in ancient limestones. However, the matrix permeability of the young carbonates that comprise the Upper Floridan Aquifer may be more than 10⁵ times more permeable than the ancient limestones of the mid-continent. Additionally, matrix permeability in the Upper Floridan Aquifer is facies-dependent and spans three orders of magnitude (Budd and Vacher, 2004). Such variations would provide preferred horizons of ground-water flow (Vacher *et al.*, 2006).

If the cave levels in Florida are related to lithologic units with high matrix permeability, the elevations of these cave levels would change in accordance with the geologic structure. However, the widespread levels of cavities do not follow the geologic structure; the cave levels are at the same elevation even though the lithologic units dip to the south (Fig. 12). Therefore, lithologic variability does not exert the first-order influence on the locus of cave development.

There is, however, some correspondence between the cave levels in the study area and modes in the histogram of topographic data for Citrus and Hernando Counties (Fig. 5). The

modes in the topographic data manifest the classic marine terraces identified in Florida by Cooke (1945) and later Healy (1975) including the Silver Bluff (+2.4 m), Talbot (+12.8 m), Penholoway (+21.3 m), and Wicomico (+30.5 m) terraces. These marine terraces are directly related to previous elevations of sea level.

In this near-coastal setting, the position of sea level has a direct influence on the position of the water table. Since the elevations of cave levels in the survey data generally correspond to the elevation of marine terraces, it appears that the development of air-filled caves in west-central Florida may be related to positions of the water table, and thus sea level, when they were higher than present.

PASSAGE CONNECTIVITY

Of the seven caves in the Brooksville Ridge surveyed during this study, none contain continuous conduits that connect sites of recharge to points of discharge within the Upper Floridan Aquifer. Neither do passages in the surveyed caves comprise a dendritic network of conduits with tributary passages. Only one cave, BRC Cave, receives occasional water from a sinking stream and contains natural indicators of localized directional flow such as sediment ripples and pebble imbrication. Three other caves, Big Mouth, Morris, and Werner, receive recharge from artificial sinking streams created during quarry reclamation. Discharge for the water that enters all seven caves rises some 15–20 km to the west at the large springs along the coast.

Connections between the caves and the surface are limited in the karst of west-central Florida. Many caves in the Brooksville Ridge, including four of the caves in this study (BRC, Big Mouth, Morris, and Werner), had no known human-scale entrance prior to lime-rock mining. In fact, most air-filled caves that are known in the karst of west-central Florida were discovered by human alteration of the land, in particular lime-rock quarries that excavate to the level of the cave passages. The subdued topography of Florida contributes to the lack of entrances by restricting the natural intersection of the land surface with the horizontal cave passages. The implication is that there are many more caves in west-central Florida than are currently known. The burgeoning sinkhole insurance industry in Florida is a manifestation of this fact.

Surveyed passages within the air-filled caves of west-central Florida do not extend long distances. Tabular passages pinch into low cavities. Fissure-type passages thin into increasingly-narrowing fractures. Quaternary-age siliciclastic sediments and structural collapse features are pervasive, and further segment the caves. The connections between human-scale passages at the same level, therefore, are small, and additional exploration requires excavation by dedicated cavers (Turner, 2003). Vertical exploration in the caves is achieved where structural collapse features or solution-enlarged fractures connect multiple levels (Fig. 4).

POSSIBLE HYDROLOGIC IMPLICATIONS

Data from the air-filled caves in the Brooksville Ridge of west-central Florida contradict the notion of an integrated network of conduits above the modern water table. If the observations from this study are representative of conditions below the present water table, then connectivity between input and output points within the Upper Floridan Aquifer may be limited.

It also appears that caves in west-central Florida do not follow the sinking stream-spring model so widely accepted by karst scientists who study the ancient limestones of the mid-continent. Rather, water in the karst aquifers of west-central Florida may travel through a maze of passages, fractures, sediment fills, and rock matrix at several horizons.

Available data support this conjecture of multi-level discontinuous mazes. For instance, maps of underwater caves reveal passages throughout west-central Florida that occur at specific depths up to 120 m below the water table (Florea and Vacher, in review). Furthermore, Quaternary-age siliciclastic sediments infiltrate these underwater caves, and these sediments are commonly recovered from cavities encountered during well construction (e.g., Hill and DeWitt, 2004).

Disjunct or occluded underwater passages in the Upper Floridan Aquifer would impede ground-water flow, resulting in higher elevations of the water table and steep hydrologic gradients. These are both observed within the karst of west-central Florida. As one example, a regional, finite-difference ground-water model that includes the northern portions of the Brooksville Ridge, developed for the Southwest Florida Water Management District by GeoTrans (1988), concluded that model calibration to known elevations of the water table is possible only if fractures or solution features are not regionally extensive or hydraulically connected. If the opposite case were true (i.e., if solution features were regionally extensive or hydraulically connected), the gradient of the water table would reduce to near-zero and the elevation of the water table would equilibrate near sea level. The coastal, carbonate aquifers in the Yucatán Riviera of Mexico, with more than 400 km of mapped underwater cave and water-table gradients of less than 0.00001 (Worthington *et al.*, 2000), illustrates this possibility. This hydrogeologic contrast between the great peninsulas of Florida and Yucatán, and its relation in part to the presence of infiltrating clastics in the case of Florida, was pointed out more than 30 years ago by Back and Hanshaw (1970).

CONCLUDING REMARKS

This study of air-filled caves in the Brooksville Ridge of west-central Florida offers an improved understanding of cave-scale porosity in the Upper Floridan Aquifer. How does the architecture of these caves compare with that of other cave systems? It is instructive to review summaries from two contrasting geologic settings, the caves of ancient low-permeability limestones of the mid-continent (Palmer, 2003) and the

caves of small islands composed of Pleistocene limestone (Myroie *et al.*, 1995).

The first example, the caves of the mid-continent, is important because it is the paradigm view of near-surface caves. Palmer (2003, p. 2) uses the following description for such caves:

Most accessible caves are surrounded by rock in which the vast majority of openings have hardly enlarged at all. The conduits are not surrounded by porous zones, with walls like a sponge, where progressively smaller openings extend indefinitely into the cave wall. The conduits are quite discrete.

Cave passages in the young carbonates of west-central Florida do not fit this description. Tabular passages are laterally extensive, and fissure-type passages thin into increasingly-narrowing fractures; both extend beyond the limits of human exploration. The walls of the passages are porous and complex, with small-scale solution features such as pockets and tafoni structures extending into the host bedrock, which itself has high permeability. Cave passages in the Brooksville Ridge are not discrete conduits, and they do not connect together into a dendritic-style drainage system as described by Palmer (1991). Ground water in the Upper Floridan Aquifer may readily exchange between the cave and the rock matrix (Martin and Dean, 2001).

The second example, from the young carbonate islands, is important because it is the paradigm for caves in young limestone. These flank margin caves, which form by mixing at the water table and at the freshwater-saltwater interface, are summarized as follows by Myroie and Carew (1995, p. 252-253):

Typically these caves are dominated by large globular chambers that are broad in the horizontal plane but vertically restricted...At the rear of the chamber there is usually a series of smaller chambers that change into tubular passages...Commonly there are many cross-connections between adjacent chambers and passages that give the caves a maze-like character. The passages...end abruptly. The chamber and passage walls are often etched into a variety of dissolution pockets and tubes...Flow markings, such as ablation scallops, are absent.

Many of the features found in the caves of the Brooksville Ridge are remarkably similar to this description. Laterally extensive cavities contain bedrock pillars and cusped dissolution features, and the passages often terminate in blind pockets. Flow indicators are generally not present. However, there are distinct differences between caves of west-central Florida and caves on young, carbonate islands. Whereas flank margin caves, for example, are composed of amorphous voids and rudimentary, spongework mazes (Palmer, 1991), the caves in west-central Florida contain passages with a sense of directionality imposed by fractures in the rock matrix. The result is maps that resemble network maze caves in plan view, such as those in the Black Hills of South Dakota (Palmer, 1991).

In conclusion, caves in west-central Florida do not fit existing models of cave architecture. They represent a style of cavern development important within coastal karst aquifers composed of young carbonates.

These west-central-Florida caves that lie above the water table demonstrate the extreme heterogeneity of permeability within the unconfined Upper Floridan Aquifer that lies below. This study offers the following insights to the architecture of cave-scale porosity in this critical-use aquifer: 1) cave-scale porosity is widespread but often composed of isolated or partially connected passages; 2) cave passages are generally restricted to specific elevations within the aquifer framework, and 3) the direction of cave passages in these levels occurs along a NE-SW and NW-SE system of fractures.

ACKNOWLEDGMENTS

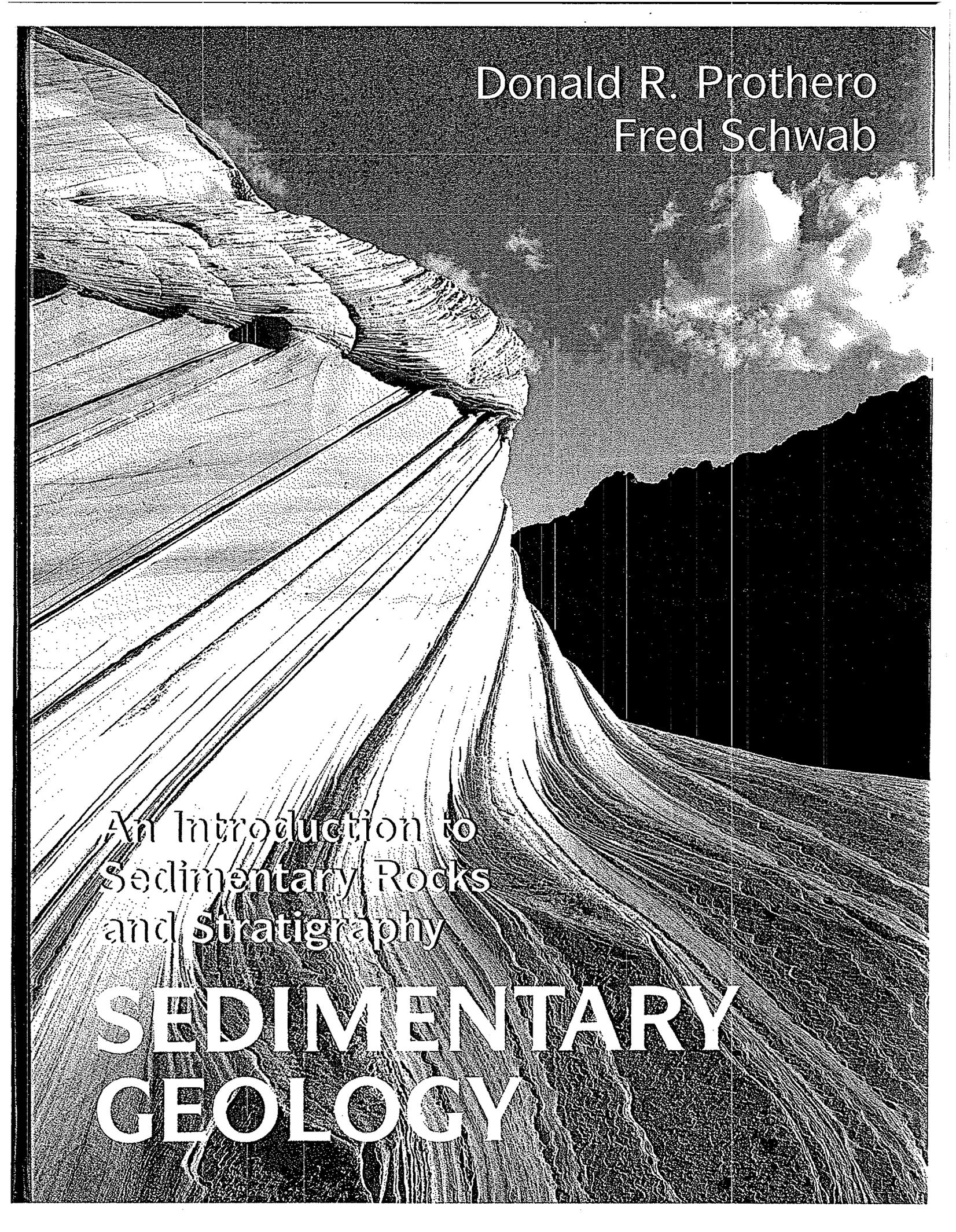
This project is indebted to the hard work of cavers and cave surveyors in Florida. I extend personal gratitude to members of the Florida Cave Survey, the Florida Speleological Society, the Tampa Bay Area Grotto, and the Withlacoochee State Forest – in particular Robert Brooks, Sean Roberts, Dan Straley, Tom Turner, Bill Walker, and Colleen Werner – for their assistance and cooperation during this project. I also thank other members of the Karst Research Group at the University of South Florida for their time, field assistance and ideas. Funding for this and other related work is provided by monetary and equipment grants from the Geological Society of America, Gulf Coast Association of Geological Sciences, Society for Sedimentary Petrology, National Speleological Society, Southwest Florida Water Management District, and the Florida Studies Center. Two anonymous reviewers helped to guide the presentation of data and ideas.

REFERENCES

- Back, W., and Hanshaw, B.B., 1970, Comparison of chemical hydrology of the carbonate peninsulas of Florida and Yucatan: *Journal of Hydrology*, v. 10, pp. 330–368.
- Brinkmann, R., and Reeder, P., 1994, The influence of sea-level change and geologic structure on cave development in west-central Florida: *Physical Geography*, v. 15, p. 52–61.
- Budd, D.A., and Vacher, H.L., 2004, Matrix permeability of the confined Floridan Aquifer: *Hydrogeology Journal*, v. 12, no. 5, 531–549.
- Cooke, C.W., 1945, *Geology of Florida*: Florida Geological Survey Bulletin 29, 339 p.
- Florea, L.J., and Vacher, H.L., in review, Quaternary cave levels in peninsular Florida: *Quaternary Science Reviews*.
- Ford, D.C., and Williams, P.W., 1989, *Karst geomorphology and hydrology*, Unwin Hyman, London, 601 p.
- GeoTrans, 1988, Hydrologic investigation of the northern portion of the Southwest Florida Water Management District, northern district model project, phase II final report: 151 p.

- Granger, D.E., Fabel, D., and Palmer, A.N., 2001, Plio-Pleistocene incision of the Green River, Kentucky determined from the radioactive decay of cosmogenic ^{26}Al and ^{10}Be in Mammoth Cave sediments: GSA bulletin, v. 113, no. 7, p. 825–836.
- Healy, H.G., 1975, Terraces and Shorelines of Florida: Florida Bureau of Geology Map Series 71.
- Hill, M.E., and DeWitt, D.J., 2004, Drilling and Testing Report for the ROMP WW-2 and WW-3 Monitoring Well Sites, Hernando County, Florida: Southwest Florida Water Management District, 12 p.
- Hill, C.A., 1987, Geology of Carlsbad Caverns and other caves in the Guadalupe Mountains, New Mexico: New Mexico Bureau of Mines and Mineral Resources Bulletin 117, 150 p.
- LaFrenz, W.B., Bulmer, W.H., Jamilla, S.V., and O'Neal-Caldwell, M., 2003, Characteristics and development of shallow solution features in thinly mantled karst, Alachua and Levy Counties, Florida: *In* Florea, L., Vacher, H.L., and Oches, E.A., eds. Karst Studies in West Central Florida: USF Seminar in Karst Environments, Southwest Florida Water Management District, p. 21–37.
- Lane, E., 1986, Karst in Florida: Florida Geological Survey Special Publication no. 29, 100 p.
- Littlefield, J.R., Culbreth, M.A., Upchurch, S.B., and Stewart, M.T., 1984, Relationship of Modern Sinkhole Development to Large-Scale Photolinear Features: *In* 1st Multidisciplinary Conference on Sinkholes, Beck, B.F., ed., p. 189–195.
- Loizeaux, N.T., 1995, Lithologic and hydrogeologic framework for a carbonate aquifer: evidence for facies-controlled hydraulic conductivity in the Ocala Formation, West-Central Florida: [MS Thesis] University of Colorado, Boulder, Colorado, 125 p.
- Loper, D.E., Werner, C.L., Chicken, E., Davies, G., and Kincaid, T., 2005, Coastal carbonate aquifer sensitivity to tides: EOS Transactions, American Geophysical Union, v. 86, no. 39, p. 353, 357.
- Martin, J.B., and Dean, R.W., 2001, Exchange of water between conduits and matrix in the Floridan aquifer: *Chemical Geology*, vol. 179, p. 145–165.
- Meinzer, O.E., 1927, Large springs in the United States: United States Geological Survey Water Supply Paper 557, United States Geological Survey.
- Miller, J.A., 1986, Hydrogeologic Framework of the Floridan Aquifer System in Florida and in Parts of Georgia, Alabama, and South Carolina: United States Geological Survey Professional Paper 1403-B, 91p.
- Myroie, J.E., Carew, J.L., and Vacher, H.L., 1995, Karst development in the Bahamas and Bermuda: *In* Curran, H.A. and White, B., eds. Terrestrial and shallow marine geology of the Bahamas and Bermuda, Geological Society of America Special Paper, Boulder, Colorado, v. 300, p. 251–267.
- Palmer A.N., 2003, Patterns of dissolution and porosity in carbonate rocks: Speleogenesis and Evolution of Karst Aquifers, v. 1, p. 1–9.
- Palmer, A.N., 2002, Karst in Paleozoic rocks: How does it differ from Florida? *In* Hydrogeology and Biology of Post Paleozoic Karst Aquifers, Martin, J.B., Wicks, C.M., and Sasowsky, I.D., eds., Karst Frontiers, Proceedings of the Karst Waters Institute Symposium, p. 185–191.
- Palmer A.N., 2000, Hydrogeologic control of cave patterns: *In* Klimchouk, A., Ford, D., Palmer, A., and Dreybrodt, W., eds. Speleogenesis: evolution of karst aquifers, National Speleological Society, p. 77–90.
- Palmer, A.N., 1991, Origin and morphology of limestone caves: Geological Society of America Bulletin, v. 103, p. 1–21.
- Palmer, A.N., 1987, Cave levels and their interpretation: National Speleological Society Bulletin, v. 49, p. 50–66.
- Rosenau, J.C., Faulkner, G.L., Hendry Jr., C.W., and Hull, R.W., 1977, Springs of Florida, Florida Geological Survey Bulletin 31 (Revised).
- Scott, T.M., Campbell, K.M., Rupert, F.R., Arthur, J.A., Missimer, T.M., Lloyd, J.M., Yon, J.W., and Duncan, J.G., 2001, Geologic map of the State of Florida: Florida Geological Survey Map Series 146.
- Scott, T.M., 1988, The Lithostratigraphy of the Hawthorn Group (Miocene) of Florida: Florida Geological Survey Bulletin 59, 148 p.
- Scott, T.M., Means, G.H., Meegan, R.P., Means R.C., Upchurch, S.B., Copeland, R.E., Jones, J., Roberts, and T., Willet, A., 2004, Springs of Florida, Florida Geological Survey Bulletin 66, 377 p.
- Stringfield V.T., and LeGrand, H.E., 1966, Hydrology of Limestone Terranes in the Coastal Plain of the Southeastern United States: Geological Society of America Special Paper 93, 41 p.
- Tihansky, A.B., 1999, Sinkholes, West-Central Florida: *In* Land Subsidence in the United States, Galloway D., Jones, D.R., and Ingebritsen S.E, eds., United States Geological Survey Circular 1182, 177 p.
- Turner, T., 2003, Brooksville Ridge Cave: Florida's Hidden Treasure: NSS News, May 2003, p. 125–131, 143.
- Upchurch S.B., 2002, Hydrogeochemistry of a karst escarpment: *In* Hydrogeology and Biology of Post Paleozoic Karst Aquifers, Martin, J.B., Wicks, C.M., and Sasowsky, I.D., eds., Karst Frontiers, Proceedings of the Karst Waters Institute Symposium, p. 73–75.
- Vacher, H.L., Hutchings, W.C., and Budd, D.A., 2006, Metaphors and Models: The ASR Bubble in the Floridan Aquifer: *Ground Water*, v. 44, n. 2, p. 144–152.
- Vernon, R.O., 1951, Geology of Citrus and Levy Counties, Florida: Florida Geological Survey Bulletin 33.
- White, W.A., 1970, The Geomorphology of the Floridan Peninsula: Florida Bureau of Geology Bulletin 51.
- White, W.B., 1988, Geomorphology and Hydrology of Karst Terrains, Oxford University Press, New York, 464 p.

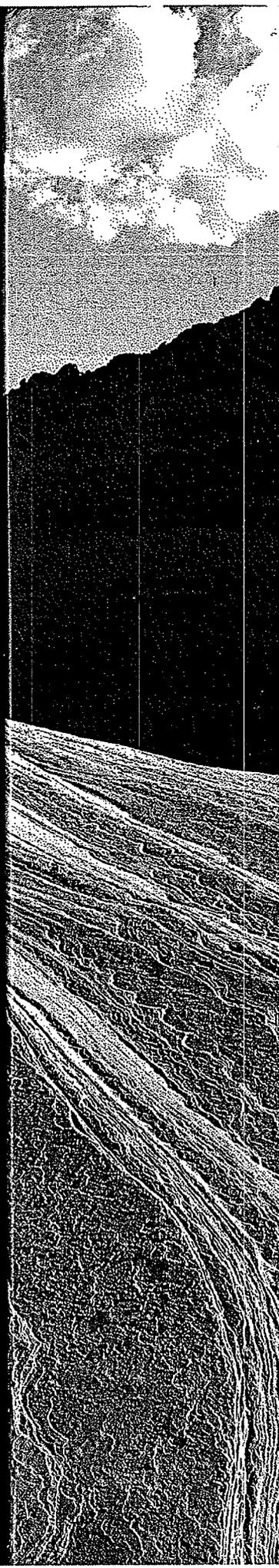
- Worthington, S.R.H., Ford, D.C., and Beddows, P.A., 2000, Porosity and permeability enhancement in unconfined carbonate aquifers as a result of solution. *In* Klimchouk, A., Ford, D., Palmer, A., and Dreybrodt, W., eds. *Speleogenesis: Evolution of karst aquifers*, National Speleological Society, p. 463–472.
- Yon, J.W., and Hendry, C.W., 1972, Suwannee Limestone in Hernando and Pasco counties, Florida; part I: Florida Bureau of Geology Bulletin 54, p.1–42.
- Yon, J.W., Waite, W.R., and Williams, C.T., 1989, Part II – Geology, Mining, and Reclamation at the Radar Hill Quarry, Citrus County, Florida: Florida Geological Survey Information Circular 105, p. 36–51.



Donald R. Prothero
Fred Schwab

An Introduction to
Sedimentary Rocks
and Stratigraphy

SEDIMENTARY GEOLOGY



SEDIMENTARY GEOLOGY

An Introduction to Sedimentary Rocks and Stratigraphy

Donald R. Prothero, *Occidental College*
Fred Schwab, *Washington & Lee University*

Donald R. Prothero and Fred Schwab combine their expertise to provide a solid introduction to sedimentary rocks and stratigraphy in one complete, up-to-date volume with numerous illustrations and photographs. This modern approach to sedimentary geology acknowledges recent changes in the field and in the way introductory courses are taught.

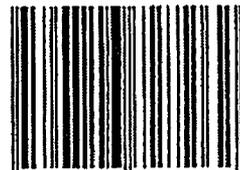
Keeping the needs of instructors and students in mind, Prothero and Schwab emphasize general principles instead of exceptions and special cases, with sections that cover sedimentary processes and products, siliciclastic sediments and environments, carbonates and other sediments, volcanogenic sediments, and stratigraphy (including sequence stratigraphy and seismic stratigraphy). Background principles in chemistry, physics, and structures are introduced as needed. Students gain an understanding of sedimentary structures and trace fossils, the strengths and weaknesses of various dating and correlation methods, and of the sandstones, shales, and limestones that make up to 99 percent of the sedimentary rocks they will encounter.

This balanced, current review, with its lucid writing, lively presentation, and outstanding graphics gives students a complete foundation in sedimentary geology while conveying the excitement and importance of a rapidly changing field.



W. H. Freeman and Company
41 Madison Avenue, New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England

ISBN 0-7167-2726-9



90000



9 780716 727262

SEDIMENTARY GEOLOGY

An Introduction to Sedimentary
Rocks and Stratigraphy

Donald R. Prothero
Occidental College

Fred Schwab
Washington & Lee University



W. H. Freeman and Company
New York

Carbonate Rocks



Many limestones are composed entirely of fossils and fossil fragments. This richly fossiliferous limestone from the Upper Ordovician Platteville Group, near Dickeyville, Wisconsin, includes abundant D-shaped strophomenids and other brachiopods, numerous branching bryozoans, massive bryozoans (*bottom center*), kidney-bean-shaped ostracodes, and even a fragment of a trilobite (*upper left corner*).
(Photo by D.R. Frothero.)

The earlier chapters described terrigenous sedimentary rocks, those composed mainly of siliciclastic material—fragments of pre-existing rocks and minerals transported and deposited as *grains*. This chapter covers limestone and dolomite, called carbonate rocks because they contain large amounts of carbonate (CO₃).

Carbonates are the most abundant nonterrigenous sedimentary rocks, constituting roughly one-tenth of the Earth's sedimentary shell (Table 11.1). In contrast to terrigenous rocks, they form chemically and biochemically. Dissolved *ions* carried from source to depositional site in solution eventually precipitate and form solid minerals. The distinction between biochemical and chemical sedimentary rocks is not always clear-cut. Organisms have a role in the formation of many sedimentary rocks. For example, they can extract dissolved components from seawater to manufacture shells or skeletons that later are incorporated into sedimentary rocks. Organisms can also be indirectly involved, such as when their metabolism modifies the geochemical setting enough to cause mineral precipitation.

The distinction between clastic sedimentary rocks and "crystalline" chemical and biochemical rocks such as carbonates can also be ambiguous. Besides broken or whole skeletal matter, many limestones contain composite grains formed of an inter-

locking mosaic of calcite or aragonite. These composite grains include ooids and intrabasinal detrital carbonate called *intraclasts*; such grains often experience a history of transport and deposition as *clasts*. Thus, some limestones can be considered clastic sedimentary rocks, but they are not siliciclastic. Finally, because most dolomite is the product of diagenetic alteration of limestone, dolomite is secondarily chemical sedimentary rock.

THE IMPORTANCE OF LIMESTONE

Limestone tells us much about the origin and evolution of the Earth. The texture and composition of limestone reveal as much about depositional setting as do the texture and composition of sandstone, conglomerate, and breccia. Limestones are the most important evidence remaining of shallow marine seas that covered much of the world during most of the Phanerozoic. Because they are almost invariably fossiliferous (many are made of nothing but fossils), they are probably the best documentation of organic evolution.

Limestones have an economic importance out of proportion with their abundance. Because they dis-

Table 11.1 Distribution through Time and Space of Types of Modern Carbonate Sediment and Ancient Limestone

| Type | Modern Carbonate | Ancient Limestone |
|---|---|---|
| 1. Precipitation | | |
| Direct inorganic precipitation | Rare travertine in caverns and tufa around springs | Both extremely rare; uncertain temporal extent. |
| Direct organic precipitation | Modern reefs, abundant in a band from 40°N to 40°S latitude in shallow marine areas where siliciclastic supply is low | Generally abundant throughout the Phanerozoic and back into late Proterozoic. Organic reef-makers differ as a consequence of evolution. Peak developments controlled by tectonism and organic evolution. |
| 2. Lithification of unconsolidated sediment | | |
| Nonmarine | Rare scattered shallow carbonate muds deposited on lake bottoms | Rare lacustrine marls of Phanerozoic and Proterozoic age. |
| Shallow marine | Abundant carbonate sediments of the neritic zones (intertidal, inner shelf, and outer shelf) circumscribing modern continents where siliciclastic sediment supply is low (the perimeter of the Florida Peninsula) or on isolated shallow marine plateaus (such as the Bahama Banks) | Abundant neritic carbonates of epicontinental (epeiric) sea deposits. Peak developments coincide with episodes of continental subsidence and/or ocean high stands. Abundant throughout the Phanerozoic; progressively less common back into the Proterozoic. No Archean examples. |
| Deep marine | Calcareous oozes abundant on abyssal plain areas that lie above the carbonate compensation depth, beneath surface waters where carbonate-secreting plankton thrive, and sediment masking is minimal | Rare examples exist. Sedimentary rock sequences of deep-sea origin are rarely incorporated into continental blocks. None are older than Jurassic because carbonate-secreting plankton were not abundant until the Jurassic. |
| 3. Miscellaneous | Scattered and rare unconsolidated limestone rubble, collapse rubble, and caliche | Very rarely produced or preserved. Unknown extent through time. |

solve easily, they are often porous in the subsurface and become reservoir rocks for petroleum and natural gas. Although only about 20% of the hydrocarbons in North America occur in carbonates, about 50% of the world's petroleum is recovered from carbonate reservoir rocks (most of the Persian Gulf's immense pools of oil is trapped in limestones). Because carbonate rocks are important aquifers, there is much research into how water flows through limestone caverns and fissures. High porosity also makes limestones excellent host rocks for ore-bearing solutions (such as the Mississippi Valley lead-zinc deposits). Cement is crushed limestone mixed with various clays and other components, and limestone is one of the most popular building materials in the world. For example, the Pyramids of Egypt and the

Empire State Building are made of limestone. Limestone is widely used in agriculture to neutralize soil acid, in metallurgy as a flux for smelting iron and steel, and in other industrial processes that require an inexpensive base to neutralize acid.

Limestone deposition in the oceans is partially responsible for Earth's unique atmosphere. Without this process and organic evolution, Earth's atmosphere would resemble those of Venus and Mars, which are composed mostly of carbon dioxide. Mercury, Venus, Earth, and Mars are small, dense bodies that lie close to the Sun. Planetary astronomers and historical geologists believe that these four terrestrial planets originally possessed identical atmospheres of carbon dioxide and water vapor degassed by volcanic activity from their inte-

rior. Mercury lost its atmosphere completely because it lies so close to the Sun and because of its low gravity. The carbon dioxide-rich atmospheres that mantle Venus and Mars are simply residues that remained after water vapor was condensed. On Earth, by contrast, deposition of limestone in the early seas and photosynthesis removed most of the carbon dioxide from the atmosphere, locking it up in sedimentary rock. These processes left a residual atmosphere composed of minor gases such as ammonia and nitrous oxide. Further alteration of these gases (by reaction with free oxygen produced from the photochemical dissociation of water vapor or as a by-product of photosynthesis) resulted in the present atmosphere of 80% nitrogen and 20% oxygen.

At present, limestone deposition is occurring at a much slower rate than it has in the geologic past. Two contrasting states of the Earth, icehouse and greenhouse, produce dramatic differences in seawater temperature and chemistry. During much of the Phanerozoic (for example, from the Ordovician through the Devonian and during the Jurassic and Cretaceous), warm greenhouse conditions prevailed. There were no polar ice caps, sea level was higher, the oceans were less stratified, and enormous volumes of limestone could be deposited. Under cold icehouse conditions (the Permo-Triassic, the late Cenozoic, and presently), sea level is low, there are extensive polar ice caps, ocean water is stratified, and little carbonate is formed. Modern shallow marine carbonates are now restricted to a few pathetic ghostlike remnants of a once extensive distribution. The mineralogy of carbonates also reflects ocean water conditions, with high-Mg calcite precipitated under greenhouse conditions and aragonite and low-magnesium calcite formed under icehouse conditions. Carbonate environments are discussed further in Chapter 12.

CARBONATE MINERAL CHEMISTRY

All carbonate minerals are formed by combining divalent cations (2+)—particularly calcium and magnesium, as well as minor amounts of iron, strontium, manganese, and barium—with carbonate anions (CO₃)²⁻. Although there are about 60 natural carbonate minerals, only 3 are abundant in the Earth's crust (or at the surface): calcite, aragonite, and dolomite. The first two are polymorphs of CaCO₃. Calcite is a soft mineral that readily fizzes in acid. Aragonite (the "mother of pearl" lining mol-

lusk shells and pearls) is less stable than calcite under most conditions and tends to alter into calcite during diagenesis.

Calcite is the rhombohedral form of calcium carbonate; aragonite has an orthorhombic lattice. This difference is important. The orthorhombic lattice of aragonite is much more open, so cations can be substituted for one another more easily than is possible in the tighter rhombohedral lattice of calcite. Seawater is supersaturated with calcium carbonate, so organisms have no trouble removing it to make their skeletons. They often use cations other than calcium, none of which fits the lattice perfectly. Calcium ions (Ca²⁺) have a radius of 0.99 Å; ions of Mg²⁺ (0.66 Å), Fe²⁺ (0.74 Å), and Mn²⁺ (0.80 Å) can fit in the calcium sites. But Ba²⁺ (1.32 Å) and Sr²⁺ (1.12 Å) are too large to fit in the calcite lattice. They can substitute for calcium in the less constricted aragonite lattice, however. Because magnesium is too small to fit in the large spaces available in the aragonite lattice, magnesium rarely replaces calcium in aragonite.

The free substitution of magnesium for calcium in calcite is common. Indeed, high-Mg (greater than 5% Mg) calcite is often treated as a distinct mineral, because many organisms (especially calcareous sponges, imperforate foraminiferans, octocorals, coralline algae, many crustaceans, and echinoderms) secrete high-Mg calcite (with as much as 15%–25% magnesium replacing calcium) (Table 11.2). Other organisms (most corals, pteropod mollusks, *Halimeda* calcareous algae, and the pearly layer on mollusk shells) make their skeletons out of aragonite. Because aragonite is unstable, however, fossils of these organisms now consist of calcite that was transformed from aragonite. In some cases, the instability of aragonite actually decreases the chances of fossilization. This is particularly true of pteropods, tiny planktonic snails with cap- or cone-shaped shells. Although pteropods are abundant in some surface waters, they are seldom preserved in deep-sea sediment. Their aragonitic shells dissolve too quickly.

Other organisms effectively resist any substitutions and from the outset form skeletons of only stable, low-Mg calcite. These animals include articulate brachiopods, mollusks (except for the pearly nautilus layer), most foraminiferans, and barnacles. Not surprisingly, these groups are well represented in the fossil record.

Dolomite, CaMg(CO₃)₂, is the third important carbonate mineral. The structure of dolomite consists of MgCO₃ and CaCO₃ alternating in a rhombohedral arrangement. Ideally, it should be 50–50 mole percent Ca and Mg, but the average dolomite is about 56% Ca and 44% Mg. A few large calcium ions are squeezed into smaller magnesium lattice spaces.

Table 11.2 Variation in Skeletal Carbonate Mineralogy among Organic Groups

| Taxon | Aragonite | Calcite mol % MgCO ₃ | | | | Both Aragonite and Calcite |
|------------------------------|-----------|---------------------------------|-----|-----|----|----------------------------|
| | | 0 | 10 | 20 | 30 | |
| Calcareous algae | | | | | | |
| Red | | | | x—x | | |
| Green | x | | | | | |
| Coccoliths | | x | | | | |
| Foraminifers | | | | | | |
| Benthonic | 0 | | x—x | | | |
| Planktonic | | | x—x | | | |
| Sponges | 0 | | | x—x | | |
| Coelenterates | | | | | | |
| Stromatoporoids ^a | x | | x? | | | |
| Milleporoids | x | | | | | |
| Rugose ^a | | | x | | | |
| Tabulate ^a | | | x | | | |
| Scleractinian | x | | | | | |
| Alcyonarian | 0 | | x—x | | | |
| Bryozoans | 0 | | x—x | | | |
| Brachiopods | | | xx | | | |
| Molluscs | | | | | | |
| Chitons | x | | | | | |
| Pelecypods | x | | x—x | | | |
| Gastropods | x | | x—x | | | |
| Pteropods | x | | | | | |
| Cephalopods (most) | x | | | | | |
| Belemnoids ^a | | | x | | | |
| Annelids (serpulids) | x | | x—x | | | |
| Arthropods | | | | | | |
| Decapods | | | x—x | | | |
| Ostracods | | | x—x | | | |
| Barnacles | | | x—x | | | |
| Trilobites ^a | | | x | | | |
| Echinoderms | | | x—x | | | |

Note: x, common; 0, rare

^a Not based on modern forms

Source: Scholle, P. A. 1978, Carbonate rock constituents, textures, cements, and porosities. *Amer. Assoc. Petrol. Geol. Mem.* 27; reprinted by permission of AAPG, Tulsa, Okla.

Ferrous (Fe²⁺) iron also substitutes for the magnesium. If enough ferrous iron is present, it forms ferroan dolomite or ankerite, with the composition Ca(MgFe)(CO₃)₂.

CARBONATE GEOCHEMISTRY

To better understand calcium carbonate precipitation, a brief review of chemical reactions, especially chemical reaction equilibrium constants, is appropriate.

Equilibrium Constants: A Case Study of the System H₂O = H⁺ + OH⁻

Consider the chemical reaction A + B = C + D. A and B are referred to as the *reactants*; C and D are the *products*. An equal sign is used to indicate that the reaction proceeds with equal speed in both directions. This type of chemical reaction is the exception in nature. Most chemical reactions are displayed with unequal-sized arrows pointing in opposite di-

rections to convey symbolically the effect of the reaction's equilibrium constant.

The equilibrium constant for a reaction (*k*) equals the product of the reaction product concentrations divided by the product of the reactant concentrations. For example, the equilibrium constant (*k*) for the reaction A + B (the reactants) to produce C + D (the products) is:

$$k_1 \text{ (equilibrium constant)} = \frac{(C)(D)}{(A)(B)}$$

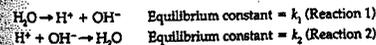
Like all chemical reactions, however, this reaction can occur in the reverse direction (C + D reacting to form A + B), for which a second equilibrium constant (*k*₂) is needed:

$$k_2 \text{ (equilibrium constant)} = \frac{(A)(B)}{(C)(D)}$$

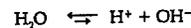
Although these two reactions occur simultaneously, they normally proceed at different rates, reflecting the different values of the two equilibrium constants. For example, if *k*₁ is larger than *k*₂, the reaction A + B → C + D will proceed from left to right (as indicated by the arrow). Conversely, if *k*₂ is larger than *k*₁, the reaction A + B ← C + D proceeds from right to left.

An overall equilibrium constant, *k*₃, describes the reaction A + B → C + D as well as A + B ← C + D. It is calculated by dividing *k*₁ by *k*₂. The relative speed by which the two reactions proceed simultaneously is shown by using arrows of different sizes. For example, if *k*₁ > *k*₂, then reaction A + B ⇌ C + D proceeds from left to right.

An example of this is the reaction by which water (H₂O) dissociates into cations of hydrogen (H⁺) and anions of hydroxyl (OH⁻). The overall reaction is made up of two distinct reactions that occur simultaneously:



Equilibrium constant *k*₁ describes reaction 1, the dissociation of water into ions of hydrogen and hydroxyl. Equilibrium constant *k*₂ describes reaction 2, the combining of hydrogen and hydroxyl anions to form water. The overall reactions should be shown correctly as



The force driving the reaction from left to right, *k*₁, is proportional to the concentration of H₂O. The force driving the reaction from right to left, *k*₂, is propor-

tional to the concentration of hydrogen and hydroxyl. Again, the two reactions occur simultaneously but at differing rates. The overall equilibrium constant for the two simultaneous reactions, *k*₃, is calculated by dividing *k*₁ by *k*₂. The value of *k*₃ is 1 × 10⁻¹⁴. This small value means that very little water dissociates into hydrogen and hydroxyl ions.

A practical consequence of this value of *k*₃ is the standard scale of pH that describes acidic or basic (alkaline) solutions. The pH of a solution is the negative log to the base 10 of the hydrogen cation concentration. When the concentration of hydrogen cations equals the concentration of hydroxyl anions, the solution is described as neutral. The pH is 7, and the hydrogen cations and hydroxyl anions each have a concentration of 1 × 10⁻⁷. Acidic solutions (pH from 1 up to 7) have more hydrogen cations than hydroxyl anions. Basic solutions (pH from more than 7 up to a maximum of 14) have more hydroxyl anions than hydrogen cations.

Carbonate Precipitation: Reactions and their Equilibrium Constants

Table 11.3 lists the four reactions that occur simultaneously in seawater to form limestone.

In reaction 1, atmospheric carbon dioxide combines with water to form carbonic acid. This reaction's small equilibrium constant means that very little carbonic acid is produced when rain falls through the atmosphere or when atmospheric carbon dioxide is absorbed into seawater.

In reaction 2, the carbonic acid produced by reaction 1 spontaneously dissociates into hydrogen cations (H⁺) and bicarbonate anions (HCO₃⁻). The equilibrium constant for reaction 2 is much smaller than that for reaction 1. This means that very few hydrogen cations are produced.

In reaction 3, the bicarbonate produced by reaction 2 dissociates into more hydrogen cations and

Table 11.3 Carbonate Geochemistry: Reactions and Equilibrium Constants

| | |
|--|--|
| 1. CO ₂ (gas) + H ₂ O (water) ⇌ H ₂ CO ₃ | <i>k</i> = 10 ^{-1.43} |
| 2. H ₂ CO ₃ ⇌ H ⁺ + HCO ₃ ⁻ | <i>k</i> = 10 ^{-6.40} |
| 3. HCO ₃ ⁻ ⇌ H ⁺ + CO ₃ ²⁻ | <i>k</i> = 10 ^{-10.33} |
| 4. Ca ²⁺ + (CO ₃) ²⁻ ⇌ CaCO ₃ (limestone) | <i>k</i> = 10 ^{-8.55} (aragonite) and 10 ^{-8.48} (calcite) |

Note: By convention, relative arrow lengths in each set of reactions indicate in which direction each normally proceeds.

carbonate anions (CO_3)⁻². The equilibrium constant for reaction 3 is even smaller than that for reaction 2.

In reaction 4, dissolved calcium and dissolved carbonate are combined (with or without the assistance of organisms) to form solid calcium carbonate (aragonite or calcite).

Ignoring for the moment the equilibrium constants for reactions 1, 2, and 3, LeChatelier's principle (adding components to the left side of a reaction will cause the reaction to proceed from left to right, and vice versa) implies that calcium carbonate formation is favored by conditions that promote absorption of atmospheric carbon dioxide into water. But if the equilibrium constants for these reactions are taken into account, precisely the opposite is true. Because reaction 2 generates hydrogen cations at a far greater rate than reaction 3 generates carbonate anions, any carbonate anions produced by reaction 3 will be consumed by the far larger number of hydrogen cations produced by reaction 2. In other words, adding carbon dioxide to water actually decreases the concentration of carbonate; conversely, removing carbon dioxide from water increases the concentration of dissolved carbonate. This conclusion has far-ranging implications for limestone formation. The precipitation of limestone is promoted by any process that removes carbon dioxide from water.

CONTROLS ON CARBONATE DEPOSITION

We can now predict which physical conditions favor formation of limestone simply by considering which conditions remove carbon dioxide gas from seawater. Simple kitchen chemistry helps. Consider a can of capped soda or beer, for example. What promotes the escape of gas? Three obvious choices come to mind: uncapp the can to remove the pressure, heat it, or shake it. Carbon dioxide dissolved in seawater is analogous to the fizz in soda or beer. Three physical conditions control how much CO_2 can be dissolved: temperature, pressure (water depth), and degree of agitation.

1. **Temperature.** Raising the temperature of seawater promotes limestone deposition. Modern carbonate sediment and ancient limestone form more readily in tropical seas than in polar waters.

2. **Pressure.** Reducing the pressure (or depth) of seawater promotes limestone deposition. Modern carbonate sediment and ancient limestone form more readily in shallow water than in deep water. In set-

tings such as the modern Bahama Banks, the effects of temperature and pressure combine to promote carbonate deposition. The upwelling of cold, dense bottom water onto shallow shelf areas warms the water and reduces pressure. Carbon dioxide is removed and limestone is formed.

3. **Degree of agitation.** Breaking waves in the surf zone mix seawater with air. This agitation promotes limestone formation because additional carbon dioxide is absorbed by the atmosphere. For example, modern fringe reefs grow faster in directions that face breaking waves.

Warm temperature, shallow depth, and agitation greatly enhance the likelihood that limestone will form, but four other factors are also important: organic activity, sediment masking and clogging, light, and carbonate compensation depth.

4. **Organic activity.** Plants and animals either precipitate calcium carbonate directly or modify the geochemical environment enough for precipitation to occur. Animals such as clams, snails, brachiopods, and zooplankton extract their calcareous skeletons from seawater directly. Plants such as phytoplankton and algae promote calcium carbonate precipitation because they remove carbon dioxide from seawater by photosynthesis (combining it with water to produce organic tissue and energy).

5. **Sediment masking and clogging.** As we will discuss further in Chapter 14, the kind of sediment accumulating at any point in time and space reflects what isn't happening as much as what is; this is the sedimentary masking effect. Even where carbonate sediment is forming, if clay and silt are being supplied more rapidly, (calcareous) mudrock rather than limestone accumulates. Evidently, large accumulations of limestone can occur only when other kinds of sediment are being deposited at exceedingly slow rates. In addition, an influx of mud can clog the filter-feeding apparatus and gills of many marine organisms. When this happens, many invertebrates that promote carbonate deposition cannot survive.

6. **Light.** Because photosynthetic organisms (especially calcareous algae and hermatypic corals) require light for photosynthesis, most large carbonate accumulations form in water shallow enough (less than 20 m deep) for adequate light to penetrate. Muddy water further inhibits the growth of corals and algae that depend upon clear water for sufficient light to grow.

7. **Carbonate compensation depth (CCD).** As discussed in Chapter 10, the temperature and pressure

of very deep seawater control the areal distribution of calcareous ooze on the modern abyssal plains. Calcareous ooze consists of the unconsolidated shells of floating pelagic organisms that thrive in the photic zone that extends from the water surface to about 200 m. After organisms die, their shells settle to the seafloor, much as snowflakes fall through the atmosphere and accumulate on the ground. Several factors control the rate at which calcareous ooze accumulates. The rate at which shells are supplied from above is very important and reflects organic productivity. This rate varies with water temperature and is largely a function of latitude, water depth, and the supply of organic nutrients. Equally important is the rate at which shells are destroyed by dissolution (see Fig. 10.19). This is controlled by the carbonate compensation depth, as discussed in Chapter 10. The CCD is analogous to the permanent snow line in mountains, which is simply the elevation above which there is year-round snow cover. The snow line coincides with the precise elevation at which the total rate of snowfall and the total rate of snow melting are in balance. Above the snow line, more snow accumulates annually than melts; below it, more snow melts annually than accumulates. The CCD marks the water depth at which slowly falling calcium carbonate sediment dissolves at precisely the same rate as it is supplied from above (see Fig. 10.20). Calcareous oozes accumulate only above the CCD; below it, calcium carbonate dissolves at a faster rate than it is supplied (Fig. 10.21). The depth of the CCD varies with water temperature. Colder temperatures increase the rate of solution. In modern oceans, the depth of the CCD ranges from 4500 or even 5000 m in warm equatorial waters to 3000 m in polar waters.

LIMESTONE COMPONENTS AND CLASSIFICATION

The most useful schemes for classifying limestone recognize the diverse origin of carbonate rocks and the clastic aspect of their texture. Components are normally recognizable using thin sections and a number of standard reference guides facilitate the procedure. (For example, see studies published by Scholle, 1978; Milliman, 1972; and Adams et al., 1964.) Where distinctive clastic components are absent, a general descriptive scale of crystalline grain size can be used (Table 11.4).

One of the most popular classifications was devised by R. L. Folk (1959, 1962). It separates allochemical and orthochemical components. Allochemical

Table 11.4 Standard Scale for Describing Crystalline Grain Sizes

| Terminology | Crystal Grain Size (mm) ^a |
|--|--------------------------------------|
| Extremely coarsely crystalline | >4 |
| Very coarsely crystalline | 1-4 |
| Coarsely crystalline | 0.25-1.0 |
| Medium crystalline | 0.062-0.25 |
| Finely crystalline | 0.016-0.062 |
| Very finely crystalline | 0.004-0.016 |
| Aphanocrystalline or cryptocrystalline | <0.004 |

Source: Folk, 1962.

^a Carbonate rocks in which readily distinguishable allochemical components occur are broadly subdivided into major categories on the basis of the grain diameter of such components as follows: calcinulites (carbonate conglomerates and breccias) have clast diameters in excess of 1 mm; calcarenites (carbonate sandstones) have clast diameters from 0.062 to 1.0 mm, and calcilulites (carbonate mud) have clast diameters less than 0.062 mm.

(from the Greek *allos*, "elsewhere" or "from outside") components (or allochems) are any grains of calcium carbonate that, after formation, are transported and deposited as clasts. They are analogous to rock and mineral fragments in the framework fraction of terrigenous sandstone. There are different types of allochems, including coated grains (such as ooids) and skeletal fragments (bioclasts). We will discuss their origin and distinguishing characteristics shortly.

Orthochemical (from the Greek *orthos*, "straight" or "true") components (or orthochems) are not transported and deposited as clasts. Orthochems are found precisely where they formed or have been moved only a short distance. There are two kinds of orthochems: (1) microcrystalline calcite matrix, or micrite, is fine-grained (finer than 4 microns in diameter) carbonate mud, analogous to matrix in wacke sandstone; (2) microcrystalline sparry cement, or spar, is relatively clear interlocking crystals of calcium carbonate, analogous to cement in arenite sandstone.

Figure 11.1A shows the four principal limestone families: (1) sparry allochemical rocks, (3) microcrystalline allochemical rocks; (3) micrite; and (4) biolithite. Sparry allochemical rocks resemble arenites and consist chiefly of such allochems as skeletal fragments or ooids glued together by interstitial sparry cement. Microcrystalline allochemical rocks are similar to wackes. They consist of such allochems as skeletal fragments or ooids floating in or intermixed with fine-grained microcrystalline matrix (micrite). Micrite is analogous to mudrock. It consists principally of microcrystalline calcium carbonate matrix in which are scattered less than 10% allochems. Biolithites are rocks that were cemented

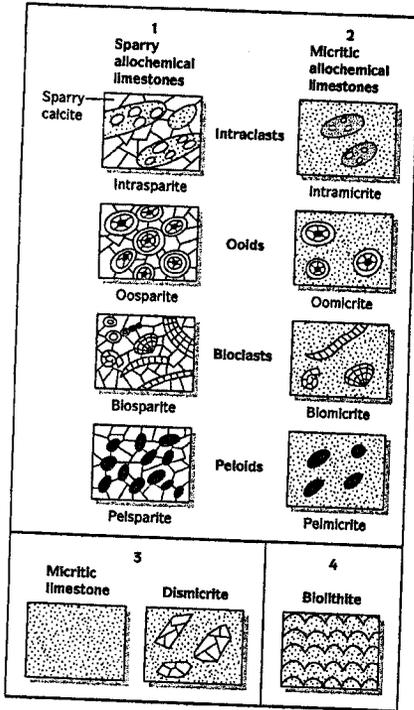


Figure 11.1

(A) R. L. Folk's (1959, 1962) classification of limestones, which uses prefixes to indicate the framework grains present (bio- for fossils, pel- for peloids, oo- for ooids, and Intra- for intracrysts) and stems to indicate whether the interstitial calcite is micritic or sparry. If the rock is originally bound together (as in a reef rock), it is a *biolithite*. (B) Textural maturity classification of limestones proposed by Folk (1962). Textural maturity classes are based on the percentage of allochems present, their degree of sorting, and the extent of rounding (a function of abrasion history). (After Folk, 1962, *Amer. Assoc. Petrol. Geol. Memoir 1*; reprinted by permission of the American Association of Petroleum Geologists, Tulsa, Okla.)

| Percent allochems | Over $\frac{2}{3}$ lime mud matrix | | | | Subequal spar and lime mud | Over $\frac{2}{3}$ spar cement | | |
|---------------------------|------------------------------------|-----------------------|------------------------------|-------------------|----------------------------|--------------------------------|-----------------------|---------------------|
| | 0-1 | 0-10 | 10-50 | Over 50 | | Sorting poor | Sorting good | Rounded and abraded |
| Representative rock terms | Micrite and dismicrite | Fossiliferous micrite | Sparse biomicrite | Packed biomicrite | Poorly washed biosparite | Unsorted biosparite | Sorted biosparite | Rounded biosparite |
| Terminology | Micrite and dismicrite | Fossiliferous micrite | Biomicrite | | Biosparite | | | |
| Terrigenous analogs | Claystone | Sandy claystone | Clayey or immature sandstone | | Submature sandstone | Mature sandstone | Supermature sandstone | |

Legend: Lime mud matrix Sparry calcite cement

together into limestone while the organisms that constitute them were still alive and growing. These include reef limestone and stromatolites.

Allochems: Characteristics and Genesis

Allochems are subdivided into (1) skeletal (biogenic) grains and (2) nonskeletal grains.

Skeletal Allochems. There are many types of skeletal grains. The various kinds present in a limestone reflect age and depositional setting. Geologic age is important because a variety of calcium carbonate-secreting organisms have evolved over the geologic past. Depositional setting dictates paleoecology. The specific biological community that exists in any given setting is determined by such factors as water depth, water temperature, salinity, and turbidity. Fossil components in limestone (and in sedimentary rocks generally) reflect both age and depositional environment.

Identifying organic remains is a complex task that is accomplished by using thin sections (Fig. 11.2). Thin sections randomly cut three-dimensional fossils in two dimensions. With experience and the use of published guides, a carbonate petrographer can identify specific skeletal remains by size, external shape, and internal organization (see, for example, Horowitz and Potter, 1971; Scholle, 1978).

Nonskeletal Allochems. Four major types of nonskeletal grains are recognized: (1) various coated grains (ooids, pisoids, and oncoids); (2) peloids; (3) clumped or aggregated grains (lumps, grapestones); and (4) limestone clasts (limeclasts); that is, fragments of pre-existing limestone derived from intrabasinal and extrabasinal sources. These clasts are ripped up, transported, redeposited and confined within limestone strata.

1. **Coated grains.** There are several types of coated calcium carbonate grains. All share a common spherical to subspherical shape, but they vary in size and degree of internal organization (Fig. 11.3). Ooids (also called ooliths) have a nucleus (often a skeletal fragment or a small clast of detrital quartz) around which concentric layers of calcium carbonate are wrapped (Fig. 11.3A, B). Ooids have diameters of less than 2 mm; most are the size of very fine sand (from 0.2 to 0.5 mm in diameter). Coarser coated grains that are identical to ooids in shape and internal organization are called pisoids (or pisolites) (Fig. 11.3C). Oncoids (or oncolites) are the same size as ooids or may be much larger, but they are irregularly

shaped (Fig. 11.3D). They contain no obvious nucleus; individual coated laminae vary in thickness and show irregular overlap.

Modern ooids form in marine settings at shallow water depths. Although ooids can form in water as deep as 15 m, most form at less than 5 m. Maximum development occurs at depths less than 2 m. Agitation by waves, tides, and storm currents promotes the growth of ooids in shoal areas around the perimeter of the Bahama Platform and in the tidal channels and deltas of the Persian Gulf. These modern ooids consist of acicular needles of aragonite. Some show tangential orientation with long axes subparallel with the ooid laminae. Radial microfabrics can also form, with aragonite needles fanning out in all directions from the nucleus, perpendicular to the laminae. Ancient ooids are composed of calcite rather than aragonite, and fibrous crystals typically fan out radially from the grain center. Recrystallization and dolomitization often obscure the original texture and mineralogy.

Ooids and pisoids can be biogenic, inorganic, or both. They apparently grow by simple accretion as wave, tidal, and storm currents sweep grain nuclei back and forth in shallow marine seawater supersaturated with dissolved calcium and carbonate. This is comparable to the growth of a snowball as it rolls across a slope blanketed with freshly deposited snow. The presence of radial microfabrics in modern ooids and pisoids, and the lack of any apparent mechanism for accreting successive discrete layers, however, makes many carbonate specialists skeptical about this mode of origin. Also, most modern ooids contain bits and pieces of organic matter that coat and permeate the carbonate laminae. This suggests that organisms play a role, but geologists disagree about whether the calcium carbonate is precipitated directly by organisms such as colonial algae, is a by-product of bacterial activity in organic matter, or occurs inorganically in a chemical milieu altered by photosynthesis.

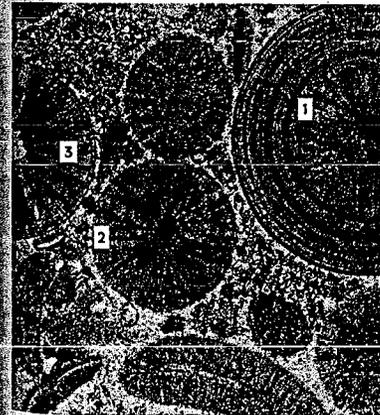
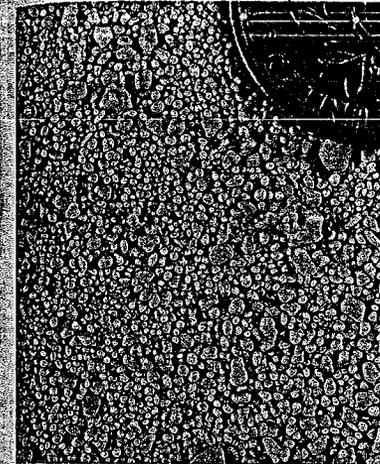
The irregular thickness of individual laminae and their tendency to overlap suggest that oncoids form under less uniform conditions than do ooids. Most are produced biogenically; modern examples have living encrustations of algae, foraminifers, and corals and form in various settings, including freshwater and brackish water reef complexes, shallow marine tidal flats and platforms, and (rarely) even deeper water carbonate settings. Many large oncoids appear to be made of pieces of ripped-up algal mat that have been rolled around into an irregular ball.

2. **Peloids.** Most peloids are fine-grained (0.1-0.5 mm) sand- to silt-sized clasts of microcrystalline carbonate that lack a coherent internal structure (Fig. 11.4A, B). The term implies neither size nor mode of

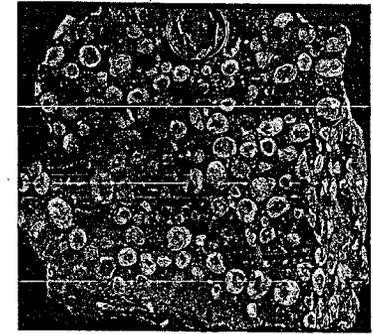
| Microstructure Common minerals (rare minerals in parenthesis) | Appearance of thin section in ordinary transmitted light | Appearance of thin section under crossed polars | Examples |
|--|--|--|--|
| Homogeneous prismatic Calcite (Aragonite) | No visible structure | Extinction in one direction; optic axes parallel and usually normal to surface of skeleton | Triobolites, ostracods |
| Granular Calcite Aragonite | Irregular grains (if fine and uniform in size, sometimes referred to as sugary or sucrosic) | Random orientation of optic axes | Foraminifers |
| Normal prismatic Calcite (Aragonite) | Polygonal prisms normal to outer surface Long Transverse | Each prism extinguishes as a unit Long Transverse | Punctate brachiopods <i>Inoceramus</i> |
| Foliated Calcite | Thin parallel leaves of calcite often having a wavy banded appearance | Variable orientation of optic axes of leaves | Bryozoans, pseudopunctate brachiopods worm tubes oysters |
| Nacreous Aragonite | Regular thin parallel leaves of aragonite (separated by organic films) | Parallel extinction | Mollusks |
| Single crystal Calcite | Coarse single calcite grain showing cleavage | Grain extinguishes as a unit | Echinoderms, sponge spicules |
| Crossed lamellar Aragonite (Calcite) | Layer of large lamellae, each lamella composed of small flat crystals, uniformly inclined in plane of larger lamella, giving herringbone pattern | Uniform orientation of optic axes in small crystals sometimes causes large lamellae to extinguish as a unit | Mollusks |
| Spherulitic fascicle Aragonite (Calcite) | Fibers radiating fanlike outward from a point center that is dark due to the concentration of very fine crystals | Each fiber extinguishes as a unit | Coelenterates |

Figure 11.2

(Facing page) the common types of skeletal microstructure in thin section as seen in plane-polarized light and crossed polars. (From T. Scoffin, 1987, *An Introduction to Carbonate Sediments and Rocks*, Table 4.2, p. 17; by permission of Blackie, Ltd., Glasgow.)



0 0.5 mm



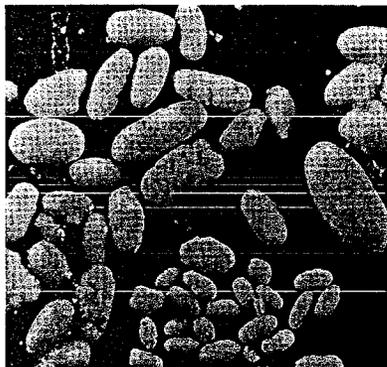
C



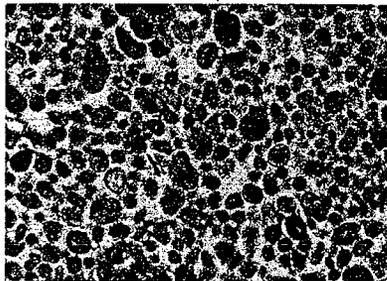
D

Figure 11.3

Several types of coated grains and concentrically banded particles occur in carbonate rocks. (A) Modern oolitic carbonate sand, Cat Cay, Bahamas. The dime shows the scale; most ooids are much smaller than 2 mm in diameter. (Photo by D. R. Prothero.) (B) Oolitic limestone with clear sparry cement showing the internal structure of ooids in thin section. This example is from the Cambrian Warrior Formation of Pennsylvania. 1, Radial-concentric ooids with peloidal core and purely radial center; diameter is 1.7 mm. 2, Purely radial ooids. 3, Radial ooids with two concentric coatings. Note that the diameters of the inner radial parts of the ooids is similar. (P. L. Heller et al., 1980, *J. Sed. Petrology* 50:944; by permission of the SEPM.) (C) Concentrically banded pisolites are much larger than ooids and usually form in calcareous soil horizons. Pisolites are also known to form in caves, lakes, and even in normal marine settings by the same process as ooids. (Photo by D. R. Prothero.) (D) These irregular, ovoid, concentrically laminated structures are oncolites (from the Lower Cambrian Chambliss Limestone, Marble Mountains, California; typical diameter is 1-2 cm.) They form when algal mats are ripped up and rolled about or encrusted around a nucleus. (Photo by D. R. Prothero.)



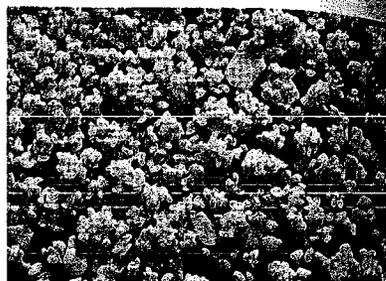
A



B

formation. Most peloids are rounded to subrounded, but they can be spherical, subspherical, ellipsoidal, or irregular in shape. Peloids form in shallow marine low-energy platform carbonate settings such as the lagoonal areas of the Bahama Banks. They originate in various ways. Many peloids are fecal pellets of waste matter generated by such organisms as fish and shrimp. Others are produced by the micritization of other kinds of allochems: ooids, oncoids, intraclasts (limeclasts), and abraded skeletal grains.

3. *Grain aggregates.* Grain aggregates form when such carbonate particles as ooids and peloids adhere to one another (Fig. 11.4C). Most are fine- to medium-grained sand-sized masses that are given specific descriptive names: grapestones, lumps, and botryoidal lumps (Illing, 1954). Though distinctive in appearance, these masses are not genetically different. Interstitial pores are filled with micrite, but



C

Figure 11.4

Pellets and peloids are a third type of carbonate framework grain. (A) Fecal pellets produced by various species of marine animals (10× magnification.) (Photo by P. E. Cloud, courtesy of U.S. Geological Survey.) (B) Peloidal limestone with clear sparry calcite cement, from the Jurassic of the Venetian Alps, Italy. Length of elongate peloid in the right center is 0.9 mm. (Photo courtesy of A. Bosellini.) (C) Cemented clusters of peloids or ooids are called grapestone. (From R. Siever, 1988, *Sand*, p. 18).

most aggregates are bound together by encrusting organisms. Grain aggregates forming today on the Bahama Banks occur in a band that separates shallow-water lagoonal areas in which micrite and peloidal micrites are accumulating. Water depths are very shallow (typically 3 m or less), and only weak bottom currents are present. These conditions permit the growth of thriving surface mats of living microbes that help to bind grains together.

4. *Limestone clasts.* Limestone clasts (or limeclasts) are ripped-up and transported fragments of pre-existing carbonate derived from extrabasinal and intrabasinal sources (see Fig. 5.6B). Extrabasinal limeclasts differ in composition, texture, and overall character from the limy material in which they occur. Intrabasinal clasts are eroded from the same stratigraphic unit, or even the same stratum,

which they are found now. Partially consolidated lime muds can be ripped up and transported short distances by currents generated by storm waves or tidal activity.

Orthochems: Characteristics and Genesis

Micrite is calcium carbonate mud that occurs as matrix in allochem-bearing limestone or by itself as limy mudrock (calclutite). It consists of silt and clay fragments (0.00625 mm or finer) of aragonite (in modern sediment) or calcite (in ancient limestone). Micrite is typically translucent under the microscope with a dull brown cast. It is dull and opaque in hand specimen.

Spar is carbonate cement. Crystals of spar are generally coarser than micrite. Crystal diameters typically range from 0.02–0.1 mm, but crystals of spar as fine as 0.001 mm (microspar) occur, so spar and micrite sizes overlap. Under the microscope, spar is crystal clear without the hazy brownish cast of micrite. Neomorphism—the various diagenetic processes of recrystallization and replacement, including changes in mineralogy—is very common in carbonate rocks. Yesterday's micrite can become today's spar, and vice versa.

Micrite is polygenetic and is most commonly formed by the physical disintegration of calcareous green algal masses within which biogenically produced needles of aragonitic are disseminated. Micrite also forms when allochems such as peloids and limeclasts disaggregate; it can also be the by-product of mechanical abrasion and bioerosion (organic burrowing and ingestion) of carbonate grains. It is also possible that some micrite is produced by direct chemical precipitation, although this is still disputed. Favorable conditions for this process are found in seawater across broad stretches of the Bahama Banks.

Spar forms as a simple, primary, pore-filling cement or is generated secondarily by recrystallizing micrite.

Limestone Classifications of Folk and Dunham Compared

R. J. Dunham (1962) proposed a second classification scheme that is used as much as Folk's (1959, 1962) scheme. Both classifications distinguish allochems, matrix or micrite, and sparry calcite cement, and

both emphasize texture. Both are so widely used, to the virtual exclusion of other limestone classification systems, that each warrants further discussion. Many geologists use the two schemes interchangeably, or side-by-side, so it is worthwhile to know them well.

The Folk Classification. Figure 11.1A shows Folk's four major limestone types. Sparry allochemical limestones and micritic allochemical limestones are subdivided on the basis of the kind and proportion of allochems and given composite names. The name stem indicates whether the interstitial material is spar or micrite. A prefix identifies the predominant allochem, and a suffix indicates allochem size as either finer (*micrite*) or coarser (*rudite*) than 1.0 mm. For example, a limestone with a sparry cement and ooids and shell fragments would be an oobiosparite; one with peloids and intraclasts surrounded by lime mud would be an intrapelmicrite. Micritic limestone contains less than 10% allochems; specific varieties are named based on the predominant allochem. The recrystallized, bioturbated micrite is called dismicrite. Biolithites are limestones that were crystallized directly from the activity of reef-building corals or algae.

Folk (1962) introduced the concept of limestone textural maturity, which is determined by measuring the grain-to-matrix ratio (GMR) (Fig. 11.1B). Textural maturity adds precision to limestone description and allows energy conditions at the depositional site to be implied. Stronger or more frequent currents (contingent in most instances on shallower depth) abrade away micrite; allochems become better sorted and, with continued abrasion, better rounded.

The Dunham Classification. The one major aspect of limestones not reflected in Folk's classification is whether the sparry calcite is primary cement or a secondary recrystallization of micrite. In other words, any limestone with a sparry calcite cement is a *sparite* even if it started out as a *micrite*. Since this determination is often hard to make, Folk's classification is much more descriptive and objective. Some geologists felt that the classification should also separate primary from secondary sparry limestones. The easiest way to do this is to incorporate grain support into the classification. Limestones in which all the grains touch others (grain-supported) were originally porous and were later cemented. This is *primary sparry cement*. Limestones with grains floating in spar (not grain supported) were probably originally composed of grains floating in micrite that has been recrystallized. This is *secondary spar*, a replacement after micrite. These features are important in the Dunham classification scheme (Fig. 11.5A).

| Depositional Texture Recognizable | | | | Original components bound together during deposition, as shown by intergrown skeletal matter, lamination contrary to gravity, or sediment-floored cavities roofed over by organic or questionable organic matter and too large to be interstices | Depositional Texture Not Recognizable |
|--|----------------------|-----------------|------------|--|---|
| Original components not bound together during deposition | | | | | |
| Contains mud (particles of clay and fine silt size) | | | | Lacks mud and is grain-supported | Crystalline Carbonate |
| Mud-supported | | Grain-supported | Packstone | | |
| Less than 10% grains | More than 10% grains | | | Grainstone | Boundstone |
| Mudstone | Wackestone | Packstone | Grainstone | Boundstone | (Subdivide according to classifications designed to bear on physical texture or diagenesis) |

A

| Allochthonous Limestone Original components not organically bound during deposition | | | | | Autochthonous Limestone Original components organically bound during deposition | | | | | |
|--|-------------------------|-----------------|-----------------------------------|---|--|------------------------------------|----------------------------------|------------|------------|-----------|
| Less than 10% >2 mm components | | | Greater than 10% >2 mm components | | By organisms that build a rigid framework | By organisms that encrust and bind | By organisms that act as baffles | Boundstone | | |
| Contains lime mud (<0.05 mm) | | No lime mud | Matrix-supported | Supported by grain components coarser than 2 mm | | | | Boundstone | | |
| Mud-supported | | Grain-supported | | | Packstone | Grainstone | Boundstone | | | |
| Less than 10% grains (>0.05 mm <2 mm) | Greater than 10% grains | | Packstone | Grainstone | | | Floatstone | Rudstone | Framestone | Bindstone |

B

Figure 11.5

Classification scheme for limestones proposed by Dunham (1962.) (A) An outline of the six original limestone varieties in Dunham's 1962 classification. (After Dunham, 1962, *Amer. Assoc. Petrol. Geol. Mem.* 1:117; reproduced by permission of the American Association of Petroleum Geologists, Tulsa, Okla.) (B) Modification of the Dunham classification by Embry and Kiovan (1972), adding five additional categories for reef rocks (three of these were boundstones in Dunham's original scheme.) (After Embry and Kiovan, 1972, *Geol. Rundschau* 61:676; reprinted by permission.)

The Dunham classification emphasizes limestone texture, especially grain (allochem) packing and the ratio of grains to matrix. Allochem type is ignored. Five types of limestones are identified: mudstone, wackestone, packstone, grainstone, and boundstone. All except boundstone accumulate as clastic carbonates; individual components are not bound together during deposition.

Mudstone, wackestone, and packstone contain mud, which Dunham defined as any silt- or clay-size grains, regardless of composition. Mudstone and wackestone are mud-supported. They are limestones in which allochems are scattered through a rock that

is basically micrite. Packstone contains less mud and is grain-supported. Grain-supported limestones typically have their allochems in tangential contact. Grainstone contains no mud, and allochem grains support one another. Limestones in which the components have been bound together from origin (such as reef rocks) are called boundstones (equivalent to Folk's biolithites). A sixth category, crystalline carbonate, refers to any limestone in which the original depositional texture is unrecognizable.

Embry and Kiovan (1972) further modified the original Dunham classification to provide niches for limestones that contain allochems coarser than 2 mm

(see Fig. 11.5B). Those with a matrix-supported texture are called floatstone. Those with a grain-supported texture coarser than 2 mm are called rudstone. Boundstones are further subdivided into framestone, bindstone, and bafflestone.

LIMESTONE DIAGENESIS

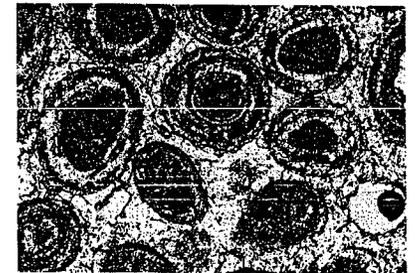
Because calcium carbonate is so soluble, diagenesis occurs more readily in limestones than in terrigenous siliciclastic rocks. In fact, diagenesis can take place at surface conditions through changes in the pore water chemistry. This is how beachrock is formed, as mentioned in Chapter 7. Almost any carbonate material that has been subjected to even moderate burial has undergone significant diagenesis, and in many cases the limestone is so recrystallized that its original texture is unrecognizable. Thus, nearly every study of ancient limestones involves the interpretation of both the diagenetic history and the original depositional texture. When metamorphism recrystallizes a limestone completely, it becomes a marble.

Diagenetic Textures

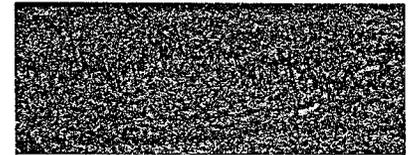
Diagenetic processes such as cementation occur as commonly in sandstones as they do in limestones, but other processes, such as dissolution and replacement, are much more common in carbonate rocks.

Dissolution produces pore space by dissolving pre-existing minerals. This process is particularly important in carbonate rocks because it often creates additional porosity that might serve as a hydrocarbon trap. Consequently, the petroleum industry has spent much time and effort on carbonate diagenesis research in recent years. Although high-Mg calcite and aragonite are stable in seawater, they are easily dissolved in other waters; high-Mg calcite is even more soluble than aragonite. For this reason, very little high-Mg calcite or aragonite survives in the pre-Cenozoic record, except in cases of unusual preservation.

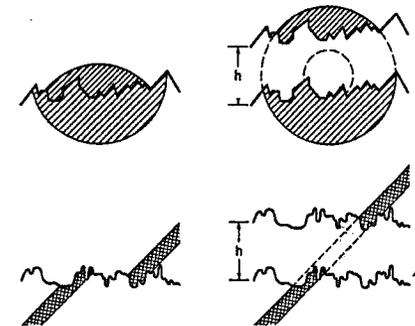
Dissolution can leave a variety of distinctive and interesting textures in a limestone. Often it produces voids where aragonitic or high-Mg calcitic fossils were selectively dissolved but the low-Mg calcitic matrix persisted. This produces secondary porosity (in contrast to primary porosity, the original void space between grains in a clean, well-sorted sand). Selective dissolution may etch out the centers of ooids, for example, and leave concentrically banded rinds (Fig. 11.6A). Selective dissolution is particu-



A



B



C

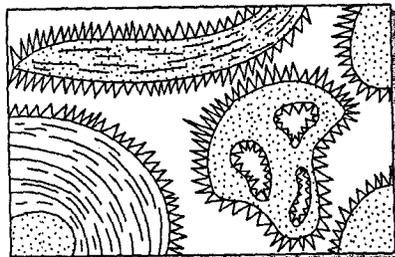
Figure 11.6

Diagenesis can be detected by its effects on the textures of limestone components. (A) The cores of these Pleistocene ooids from the Miami Limestone, Coral Gables, Florida, have been dissolved away, leaving secondary porosity. In some, the laminae have been selectively dissolved; in others, most of the ooid interior has been completely dissolved. These ooids are about 0.2 mm in diameter. (Photo courtesy of H. S. Chafetz.) (B) A stylolitic seam in a limestone showing the jagged boundary of dissolution highlighted by a residue of dark minerals. (Photo by D. R. Prothero.) (C) The volume of limestone dissolved along a stylolite can be estimated by the amount of material missing in an offset carbonate grain or seam or by the missing cross section of a circular crinoid stem.

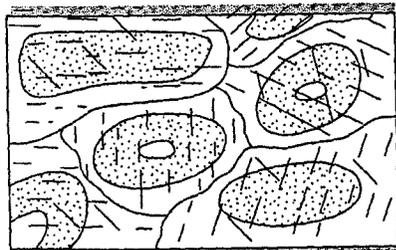
been dolomitized but the framework skeletal grains remain as calcite, then dissolution leaves a mass of dolomite with voids shaped like fossils.

Particularly distinctive are the dissolution features known as stylolites (Fig. 11.6B). These structures are jagged, irregular seams dividing the limestone into two parts that interpenetrate. Each side of the seam has toothlike projections that fit into the cavities of the opposite side. These interpenetrating tooth-and-socket structures are highlighted by a residue of insoluble opaque minerals (such as hematite) or organic matter, which shows that much limestone has been dissolved to leave so much residue behind.

Stylolites are formed by pressure solution. When limestones undergo pressure during burial, they first dissolve rather than deform. The direction of pressure is usually perpendicular to the plane of the stylolites. In some cases, the amount of dissolution can also be estimated. For example, if a carbonate grain of known shape is crossed and partially dissolved by



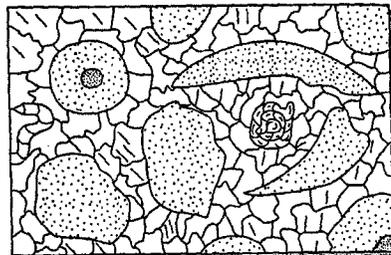
A Drusy cement



C Rim cement

particle before dissolution can be estimated. Studies of stylolites show that, typically, about 25% of the original rock is missing, and in some cases, as much as 90% of the original limestone has been dissolved away.

Cementation occurs when carbonate is precipitated into a pre-existing void space. It is one of the most common diagenetic processes, because limestones are easily dissolved and reprecipitated. In carbonate rocks, the cements often have distinctive habits and may reveal a sequence of cementing events. For example, the first generation of cement lining the voids often forms needlelike crystals that radiate away from the rim of the void and toward the center; these are known as drusy cement (Fig. 11.7). Other cements show larger, irregular, equant patches and mosaics of sparry calcite known as blocky cement. This cement often fills the remaining void space as a second generation of cement, following a first generation of drusy cement. In other limestones, cements form overgrowths around frame-



B Blocky cement

Figure 11.7

Cement types in carbonate rocks. (After Blatt et al., 1980, *Origin of Sedimentary Rocks*, 2nd ed., Fig. 14-5, p. 498; by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.)

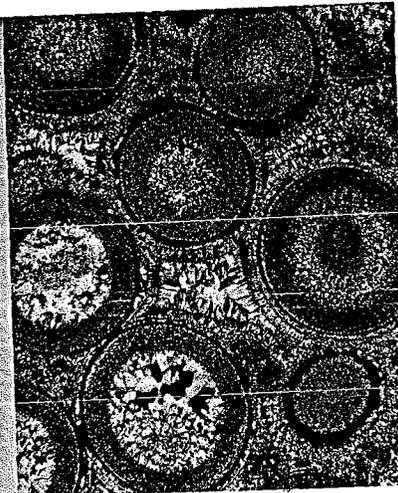


Figure 11.8

Some limestones can be so perfectly replaced by silica that the original texture is almost completely preserved. This is the famous State College Oolite from the Cambrian Mines Formation, State College, Pennsylvania. Except for some recrystallization of the cores of the ooids, the concentric structure of the ooids (1.0–1.5 mm in diameter), the drusy primary cement around the rims of the pores, and the blocky secondary cement in the centers of pores are well preserved, yet this limestone is completely silicified—no carbonate remains. (Photomicrograph by D. R. Prothero.)

work grains, resulting in rim cement. A syntaxial overgrowth consists of a framework grain composed of a single crystal of calcite (such as an echinoderm fragment) that has been overgrown by calcite cement precipitated in the same optical orientation. When thin sections of such grains are rotated on the microscope stage under crossed polars, the core grain and its overgrowth become extinct (go black) at the same time.

Replacement involves the simultaneous dissolution of original material and precipitation of a new mineral while preserving the original form. Many structures, such as ooids or fossils, can be replaced with all the fine details still intact, so there must have been slow, step-by-step dissolution and immediate cementation. Such a process suggests that the

pore waters were in a delicate balance between the dissolution of one mineral and the precipitation of its successor.

Replacement is recognized when a structure of known composition, such as a fossil or ooid, is clearly made of a mineral that is not original. Aragonitic fossils or ooids that are now calcite, and calcium carbonate fossils that are now dolomite, are two common examples. As discussed earlier, micrite is originally aragonite but is often replaced by coarsely sparry calcite (secondary or neomorphic spar). In addition to carbonates, other minerals can replace the components of limestone. The most spectacular examples occur when silica in the form of chert or chalcedony replaces carbonate. Many cherts preserve beautifully silicified fossils, ooids, peloids, or even the original calcite cementation texture of drusy and blocky cement (Fig. 11.8).

Diagenetic Environments

Because seawater is supersaturated with calcium carbonate, it takes just a slight change in water chemistry, or impurities and sites of nucleation, for marine cementation to occur. A wide variety of environmental settings (Fig. 11.9) can produce diagenesis in carbonate rocks. In the *intertidal zone*, for example, there is rapid fluctuation of the water chemistry during tides. The result is beachrock, which can form in a matter of years under the right conditions. The cements in beachrock are composed of drusy needles of aragonite and high-Mg calcite. Apparently, beachrock is precipitated when seawater evaporates and loses CO_2 , a reaction that favors carbonate precipitation, although some beachrocks occur in regions of frequent mixing of fresh and marine waters, which also produces rapid changes in water chemistry.

Cementation also takes place in the *shallow subtidal* region, filling the hollows in and between carbonate skeletal grains and within reef cavities. Again, most of the cements are aragonite and high-Mg calcite, often with a high strontium content; they typically have a crusty, bladed, or fibrous habit rather than the needlelike appearance of beachrock cement. In some cases, cements precipitate as coatings of micrite lining the cavities of either reef framework organisms or of detached but untransported shells. Obviously, this type of cementation must occur very soon after the organism died, but the exact mechanism by which such cementation occurs is not well understood. The availability of abundant sites for nucleation of calcite must be important. In the Bahamas, studies have shown that

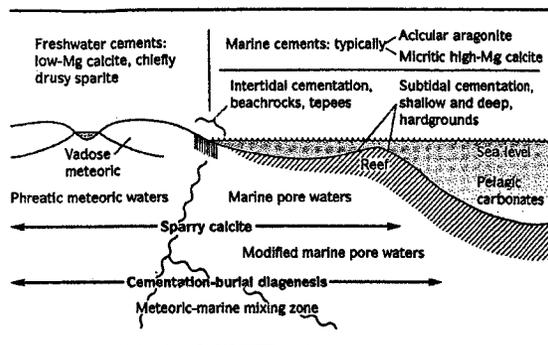


Figure 11.9

Principal environments of cementation in carbonate rocks. (After Tucker, 1981, *Sedimentary Petrology, An Introduction*, Fig. 4.26, p. 127; by permission of Blackwell Scientific Publishers, Oxford.)

dissolved CO_2 is driven off by heating and turbulence, resulting in carbonate precipitation in areas that are quiet enough to allow it.

The result of shallow subtidal cementation is a crusty surface known as a hardground. These surfaces are usually cemented most at the sediment-water interface, and cementation decreases within centimeters of the surface. They occur predominantly in shallow water, although they can occur as deep as 300 m. Hardgrounds are often found in ancient limestone and are considered evidence of interruptions in sedimentation, or hiatuses, that may span several thousand years (based on faunal discontinuities).

The most extensive areas of early carbonate sedimentation, however, occur above sea level, in zones where fresh meteoric (atmospheric) waters—that is, rain or runoff groundwater—flow through the pores of carbonate sediments. This area is known as the vadose zone, or zone of aeration above the water table. The phreatic zone, or zone of saturation below the water table, can also produce cementation. The most noticeable difference between marine and freshwater cements is that marine cements precipitate high-Mg calcite, whereas meteoric waters can only precipitate low-Mg calcite.

Some of the most detailed studies of vadose zone cementation focused on the Pleistocene sediments of the Caribbean, which were exposed to meteoric waters during the drop in sea level that occurred during glacial episodes. For example, Land et al. (1967) documented five stages of diagenesis in the limestones of Bermuda. After the primary sediment is deposited (stage I), a thick fringe of drusy low-Mg calcite needles forms on the rims of the voids, and syntaxial overgrowths occur (stage II). In stage III, the high-Mg calcite is replaced by low-Mg calcite

and aragonite. In stage IV, aragonitic skeletal grains are dissolved and their voids are filled with low-Mg calcite cement (although the internal structure of the fossils is lost). Stage V involves final cementation to fill the remaining voids, although these Pleistocene limestones of Bermuda may still have up to 20% porosity. Rarely do these limestones become cemented to the point that they have only 5% porosity or less, like a typical ancient limestone.

Other studies of meteoric cementation (James and Choquette, 1984) showed that vadose and phreatic cementation produce recognizably different textures. For example, vadose zone cements commonly include micritic envelopes around grains, thin layers of cement in the contact zone between grains (meniscus cement), and stalagmitic cements that hang down into the pores. Phreatic zone cements, on the other hand, are composed largely of drusy overgrowths of equal thickness around the rim of the void and secondary blocky cement. The fact that some of the cements clearly indicate which way was up may be valuable in beds that have been deformed and overturned.

All these cases are examples of early diagenesis, typically occurring at near-surface conditions. Later diagenesis occurs when carbonate sediments have undergone burial and compaction. In addition to stylolites, compaction can be demonstrated by fossils, ooids, or other grains that have been visibly crushed or distorted in clean skeletal limestones. Micritic limestones, however, seldom show crushing of skeletal grains, because the micrite can absorb about a 50% decrease in volume without causing distortion of the framework grains.

As burial proceeds, however, the porosity of the limestone decreases to nearly zero, which inhibits

DOLOMITE AND DOLOMITIZATION

the flow of diagenetic solutions. Yet the diagenesis of limestones becomes more intense at greater depths, even though there is almost no pore space left to transport fluids. Somehow, other processes, such as pressure solution, must operate at depth to allow diagenesis without the extensive flux of pore waters. Most deep carbonate cements are identified because they encase younger features—such as stylolites, crushed and healed coatings on grains, or dolomite crystals—that must have formed at depth. Later diagenetic cements also have distinctive chemical signatures (enriched in Fe and Mn and depleted in Sr compared to early diagenetic cements), and they are distinctive under cathodoluminescence.

Many ancient carbonate rocks consist in whole or in part of the mineral dolomite, $\text{CaMg}(\text{CO}_3)_2$. Technically, the name dolostone should be applied to carbonate rock made of the mineral dolomite, but sedimentary petrologists persist in using the mineral name for the rock out of force of habit. The context typically indicates whether the mineral or rock is meant. Dolomitization is the process by which calcium carbonate rock alters to dolostone. This process obscures the original texture, which makes classifica-

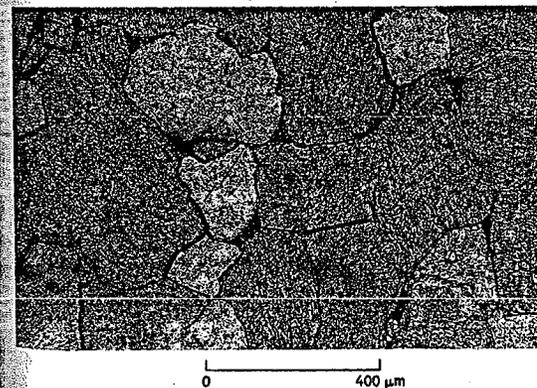


Figure 11.10

Dolomite can have a distinctive texture of euhedral rhombohedral crystals in thin section. (A) Dolomite rhombs have replaced ooids, preserving a fabric of ooids as "ghosts." The largest rhombs are 0.3 mm in diameter. (Photo courtesy of R. C. Murray.) (B) At higher temperatures, dolomite crystals become much coarser and more intergrown, losing their euhedral rhombic shape. This example is from the Devonian Lost Burro Formation, California. (Photo courtesy of D. H. Zenger.)

tion and genetic interpretation difficult. Where original textures and components survive, limestone classification is used and the degree of dolomitization is noted.

Dolomite is very similar to other carbonate minerals, so it is not always easy to identify in the field. It is less soluble than calcite and does not readily fizz in dilute acid. The standard field test is to powder a bit of the sample, which increases the surface area, and apply the acid; this causes a weak effervescence. In the laboratory, chemical stains readily separate dolomite from other minerals on a polished and etched surface or in a thin section. The stain known as Alizarin Red S colors calcite and aragonite deep red but does not affect dolomite.

Frequently, dolomite changes the texture of a rock in a noticeable way. For example, dolomitized rocks look "sugary," and in many instances dolostones have a golden-brown or tan color (in contrast to limestones, which are typically gray). In addition, the different solubilities of dolomite and calcite mean that dolostones weather differently. Often, the fossils in a limestone will be undolomitized, so that under chemical weathering and dissolution, the dolomitized matrix will remain and there will be voids where calcitic fossils have dissolved away. In other cases, dolomite selectively replaces the burrows in a micritic limestone, resulting in a mottled texture.

In thin section, it is difficult to tell rhombohedral calcite crystals from dolomite rhombs, since both have a similar crystal habit. The most reliable test is to stain the rock with Alizarin Red S before gluing the cover slip on the thin section, and this is done routinely in most petrographic research that involves dolomite. Most carbonate rocks that show extensive development of rhombohedral crystals, which often overgrow the original limestone textures, can be assumed to be dolomite (Fig. 11.10A). Such crystals grow at temperatures between 50°C and 100°C. When large mosaics of blocky anhedral dolomite are found (Fig. 11.10B), they are apparently replacements of earlier dolostone that formed at temperatures greater than 100°C (Gregg and Sibley, 1984).

The process of dolomitization has been the subject of vigorous debate. Virtually all dolostones are composed of secondary (or replacement) dolomite, meaning that calcite and aragonite form first but are later converted to dolomite. This can be proved easily if relict textures of limestone are still visible despite the overprinting of dolomite. In some dolostones, there are no relict fossils, ooids, or other evidence of limestone parent material; there are just masses of rhombohedral dolomite crystals. Are these

dolostones just very completely replaced limestones, or did the crystals of dolomite grow directly from solution with no intervening stage as a calcium carbonate mineral? In other words, are they truly primary dolomite? Modern settings in which recently formed dolomite occurs are not necessarily sites of primary dolomite. Most sites reported as primary are actually locales in which the time interval between the formation of calcium carbonate minerals and their conversion to dolomite is exceedingly short.

Experimental Studies and Dolomitization Reactions

Our inability to produce primary dolomite in the laboratory under normal marine conditions suggests that little if any primary dolomite forms naturally. Although seawater is saturated with respect to dolomite and euhedral dolomite crystals have been precipitated from seawater artificially, they precipitate in the laboratory only under unrealistically high temperatures. This inability to produce dolomite experimentally under conditions that mimic modern seawater apparently reflects the kinetics of crystallizing the ordered lattice structure of dolomite. In dolomite, cations and anions occupy the eight corners of a rhombohedral unit cell. Precise sequences of alternative cation sites are systematically occupied by calcium and magnesium. To grow such crys-

Table 11.5 Chemical Reactions that Produce Dolomite

1. Direct (primary) precipitation of dolomite. This reaction produces crystals of dolomite directly from seawater.

$$\text{Ca}^{2+} + \text{Mg}^{2+} + 2(\text{CO}_3)^{2-} \rightarrow \text{CaMg}(\text{CO}_3)_2$$
2. Dolomitization involving replacement of individual calcium cations by magnesium cations on a one-for-one basis. This reaction requires the addition of magnesium to the system and the removal of calcium from the system.

$$\text{Mg}^{2+} + 2(\text{CaCO}_3) \rightarrow \text{CaMg}(\text{CO}_3)_2 + \text{Ca}^{2+}$$
3. Dolomitization involving dolomitizing fluids that provide magnesium cations and carbonate anions. This reaction eliminates the requirement of reaction 2 that calcium ions be removed from original calcite or aragonite.

$$\text{CaCO}_3 + \text{Mg}^{2+} + (\text{CO}_3)^{2-} \rightarrow \text{CaMg}(\text{CO}_3)_2$$

tals takes time. The requirement for ordering slows growth enough that it rarely occurs under natural conditions. Because seawater is simultaneously supersaturated with calcium carbonate, aragonite or calcite (minerals of simpler structure) form instead and grab the cations and anions away from any crystallizing dolomite. Natural precipitation of calcite (normal as well as high-Mg) and aragonite may reflect the tendency of metabolizing organisms to extract precisely equal amounts of magnesium and calcium.

Table 11.5 lists three reactions that produce dolomite. Reaction 1 is the direct precipitation of dolomite from seawater. Reactions 2 and 3 are two reactions that convert calcium carbonate to dolomite. Reaction 2 requires an open system in which a one-for-one replacement of calcium by magnesium occurs. Reaction 3 does not require calcium to be removed. The dolomitizing fluid supplies both carbonate and magnesium.

Insights from Alleged Primary Dolomite Sites

A number of modern sites in which dolomite forms have been studied. The dolomite at most of these sites is secondary rather than primary. It forms almost immediately after the precipitation of calcite or aragonite. Nevertheless, such localities provide clues about the conditions that favor dolomitization.

Figure 11.11 summarizes several models to explain dolomitization at these sites. All sites have a physical-chemical environment that produces dense brines with a magnesium-to-calcium ratio higher than 8, well above the 5.4 ratio of normal seawater. The brines are produced by evaporating seawater until Ca-rich, Mg-poor minerals such as gypsum and anhydrite form. It is this preferential removal of calcium that changes the Mg-to-Ca ratio.

Figure 11.11A shows one of the best known modern dolomite sites, the Coorong coastal plain on the southeastern coast of Australia. Apparently, primary dolomitic micrite forms within ephemeral alkaline lakes and lagoons developed in the interdune flat areas separating Quaternary beach barrier sands. During the summer when the region is arid, the pH in the lakes ranges from 8 to 10, and the Mg-to-Ca ratio in lake water varies from 1 to 20. Precipitation of gypsum and anhydrite preferentially removes Ca^{2+} and SO_4^{2-} . The fluid that remains is the dolomitizing agent, although high alkalinity and the presence of nearby Mg-rich groundwater (gen-

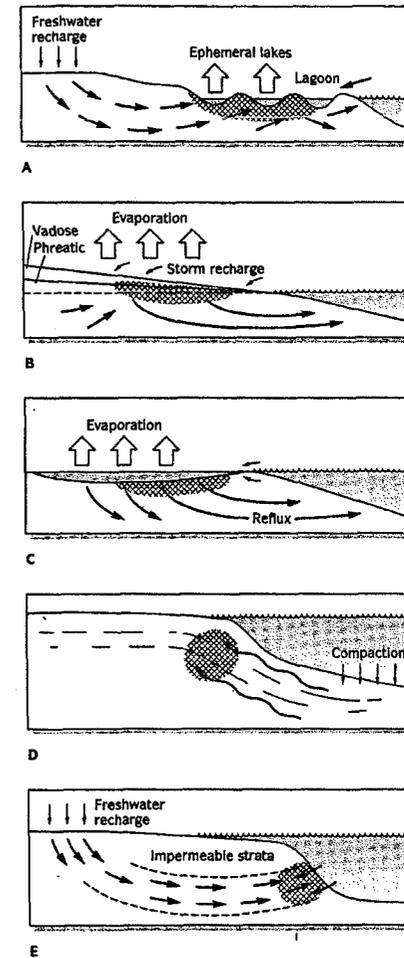


Figure 11.11

Some proposed models for dolomitization. (A) Coorong lagoon type. (B) Sabkha evaporation. (C) Seepage reflux. (D) Burial compaction. (E) Mixing zone. (Tucker and Wright, 1990, *Carbonate Sedimentology*, Fig. 8.1, p. 66; by permission of Blackwell Scientific Publishers, Oxford.)

erated during the leaching of subjacent volcanic rocks) may help. The exact process is uncertain. The dolomite mineral is actually protodolomite. It has the approximate chemical composition of dolomite, but the distinctive ordered crystal structure is absent. It is therefore unreasonable to apply this model widely.

Fluids that can dolomitize calcium carbonate minerals quickly are generated in hypersaline tidal and supratidal belts of the Caribbean region and upon sabkhas (or *sebkhas*, the Arabic word for salt flats) developed along the arid coastlines of the Persian Gulf region in the Middle East (Hardie, 1987). Figure 11.11B shows an example of this setting. Middle Eastern sabkhas have complex plumbing systems. Storms and abnormally high tides flood supratidal flats with seawater of normal salinity during the winter and spring. During the long, hot, arid summer, net evaporation leads to the crystallization of Ca-rich evaporite minerals. Upward evaporative pumping causes these dolomitizing fluids to react with aragonitic mud.

On the Bahama Banks, desiccated polygonally cracked crusts of dolomite mantle the upper surface of supratidal carbonate mud. The dolomitic crusts are apparently residues left by the evaporation of films of seawater that remained perched above the high tide level following occasional storm floods. Again, it is thought that the precipitation of gypsum raises the Mg-to-Ca ratio. The elevated salinity increases the density of the fluid, causing it to move downward and dolomitize underlying carbonate.

Another somewhat hypothetical mechanism that invokes hypersaline brines is seepage-reflux dolomitization, shown in Figure 11.11C. The best modern example of this possible mechanism was thought to be within a supratidal lake developed along the coast of Bonaire Island in the Netherland Antilles (Deffeyes et al., 1965). Seawater of normal salinity with a normal Mg-to-Ca ratio was believed to be seeping into the lake where the overall salinity and Mg-to-Ca ratio had been raised by evaporation. This was thought to cause the denser, Mg-enriched lake water to flush downward (the reflux process) and react with carbonate rock. Further study of Bonaire suggests that while the model might be valid, it does not describe what is going on there accurately. Nevertheless, evaporative reflux is invoked to explain classic dolostones such as those of the west Texas Permian (see p. 228).

Figure 11.11D illustrates burial dolomitization, a model that explains the many dolostones not intimately associated with evaporites. Burial dolomitization occurs when buried Mg-rich mudrock is compacted and dewatered. Expelled fluids rich in

magnesium then invade and dolomitize adjacent limestone. Mudrock does contain a large amount of magnesium that would be released (along with carbonate) during clay mineral diagenesis and burial metamorphism. But expelled fluids typically migrate vertically, not laterally. Furthermore, mass balance equations suggest that compacting even tremendous volumes of mudrock would generate dolostone equal to as little as 1% of the mudrock volume.

The mixing-zone dolomitization model (Folk and Land, 1975) is a second model developed to explain ancient dolostones that show no apparent relationship with hypersaline brine and evaporite (Fig. 11.11E). In this model, seawater of normal salinity and an Mg-to-Ca ratio of 5.2 is mixed with fresh groundwater, producing brackish groundwater near the coastline. The seawater ostensibly provides enough additional magnesium to supersaturate the brackish groundwater with dolomite. Groundwater circulation is thought to flush the brackish dolomitizing groundwater through subjacent limestone. The validity of this model has been questioned on geochemical grounds. There are no data indicating that significant volumes of dolostone form by mixing seawater with fresh groundwater. Typically, the salinity of fresh groundwater is low; its Mg-to-Ca ratio approaches 1.

The Origin of Dolostone: A Continuing Dilemma

Geochemists continue to develop models for dolomitization. Dense, Mg-rich fluids capable of dolomitization are easy to produce. The problem has always been how and when such fluids can be brought into contact with existing limestone to create regionally extensive belts of dolostone. Depositional settings that produce fluids capable of dolomitization are volumetrically minuscule. Many dolostones show none of the relict textural components (spar, micrite, allochems) associated with the sabkha, alkaline lake, and coastal tidal lagoon settings that most typically form dolomitizing fluids.

The temporal distribution of limestone and dolostone adds a curious twist. While about one-fifth of all carbonate rocks are dolostone, the proportion of carbonate rock that is dolostone increases progressively with age (Daly, 1907; Ronov, 1972). In Cretaceous and younger carbonate rocks, the limestone-to-dolostone ratio is 80:1; in Paleozoic carbonate rocks, it is 3:1; and in Precambrian carbonate

rocks, it is 1:3. These data can be used to support either of two arguments: (1) Depositional settings conducive to the formation of dolostone were more common early in Earth's history; or (2) dolostone is secondary, and older limestones have had more time to alter to dolostone.

There is also a correlation between dolomite abundance and periods of high sea level (Givens and Wilkinson, 1987). Most dolostones occur in rocks formed during greenhouse conditions (warm global temperature, high sea level), but little dolostone is found in rocks deposited during icehouse conditions (cold global temperature, low sea level). Dolomite abundances in deep-sea sediments show the same general pattern (Lumsden, 1985). The exact significance of this crude correlation is not known. Are there unusual characteristics of the planet during greenhouse conditions that promote dolomitization? Or is this pattern a result of the greater abundance of limestone during high sea levels and greenhouse conditions, and the scarcity of limestones during low sea levels and icehouse conditions?

Understanding dolomites, especially their origin, is an area of active research. Many unsolved problems continue to challenge us.

CONCLUSIONS

The number of sedimentary geologists who have devoted their careers to the examination, description, classification, and interpretation of limestone and dolostone closely matches the number of petrologists who have specialized in the study of siliciclastic sedimentary rocks. On the basis of the economic importance of carbonate rocks, this attention is clearly warranted. But even more important is the role that these rocks play in deciphering sedimentary environments. The texture and composition of limestone and dolostone are unmatched in their ability to provide reliable insights, both into the physical and chemical conditions that prevailed during deposition and into the geochemical settings that developed long after sedimentation.

FOR FURTHER READING

- Bathurst, R. G. C. 1975. *Carbonate Sediments and their Diagenesis*. 2nd ed. Amsterdam: Elsevier.
- Bathurst, R. G. C., and L. S. Land. 1986. Carbonate depositional environments. Part 5: Diagenesis. *Colorado School of Mines Quarterly* 81:1-41.
- Choquette, P. W., and N. P. James. 1987. Diagenesis #12. Limestones: The deep burial environment. *Geoscience Canada* 14:3-35.
- Givens, R. K., and B. H. Wilkinson. 1987. Dolomite abundance and stratigraphic age: Constraints on rates and mechanisms of Phanerozoic dolostone formation. *J. Sed. Petrol.* 57:1068-1078.
- Hardie, L. A. 1987. Dolomitization: a critical view of some current views. *J. Sed. Petrol.* 57:166-183.
- James, N. P., and P. W. Choquette. 1983. Diagenesis #6. Limestones: The sea floor diagenetic environment. *Geoscience Canada* 10:162-179.
- James, N. P., and P. W. Choquette. 1984. Diagenesis #9. Limestones: The meteoric diagenetic environment. *Geoscience Canada* 11:161-194.
- Land, L. S. 1985. The origin of massive dolomite. *J. Geol. Educ.* 33:112-125.
- Milliman, J. D. 1974. *Marine Carbonates*. Berlin: Springer-Verlag.
- Moore, C. H. 1989. *Carbonate Diagenesis and Porosity*. New York: Elsevier.
- Morse, J. W., and F. W. Mackenzie. 1990. *Geochemistry of Sedimentary Carbonates*. New York: Elsevier.
- Pray, L. C., and R. C. Murray. 1965. *Dolomitization and Limestone Diagenesis*. SEPM Special Publication 13.
- Scholle, P. A. 1978. *A Color Illustrated Guide to Carbonate Rock Constituents*. Amer. Assoc. Petrol. Geol. Memoir 27.
- Schneidermann, N., and P. M. Harris. 1985. *Carbonate Cements*. SEPM Special Publication 36.
- Scottin, T. P. 1987. *An Introduction to Carbonate Sediments and Rocks*. New York: Chapman and Hall.
- Tucker, M. E., and V. P. Wright. 1990. *Carbonate Sedimentology*. Oxford: Blackwell Scientific Publications.
- Zenger, D. H., J. B. Dunham, and R. L. Ethington, eds. 1980. *Concepts and Models of Dolomitization*. SEPM Special Publication 28.

Soil Behavior at the Microscale: Particle Forces

J. Carlos Santamarina¹

Abstract

Soils are particulate materials. Therefore, the behavior of soils is determined by the forces particles experience. These include forces due to boundary loads (transmitted through the skeleton), particle-level forces (gravitational, buoyant, and hydrodynamic), and contact level forces (capillary, electrical and cementation-reactive). The relative balance between these forces permits identifying various domains of soil behavior. Furthermore, the evolution of particle forces helps explain phenomena related to unsaturation, differences between drained and undrained strength under various loading modes (including the effect of plasticity), sampling disturbance, and fines migration during seepage. Generally accepted concepts gain new clarity when re-interpreted at the level of particle forces.

Introduction

The limitations with continuum theories for the analysis of soil behavior were recognized early in the twentieth century. Terzaghi wrote "... Coulomb... purposely ignored the fact that sand consists of individual grains, and ... dealt with the sand as if it were a homogeneous mass with certain mechanical properties. Coulomb's idea proved very useful as a working hypothesis for the solution of one special problem of the earth-pressure theory, but it developed into an obstacle against further progress as soon as its hypothetical character came to be forgotten by Coulomb's successors. The way out of the difficulty lies in dropping the old fundamental principles and starting again from the elementary fact that sand consists of individual grains" (Terzaghi 1920 - includes references to previous researchers).

The fundamental understanding of soil behavior begins by recognizing the particulate nature of soils and its immediate implications: the interplay between particle characteristics (e.g., size, shape, mineralogy), inter-particle arrangement and interconnected porosity, inherently non-linear non-elastic contact phenomena, and

¹ Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332. E-mail: carlos@ce.gatech.edu.

particle forces. While all these parameters are interrelated, the focus of this manuscript is on particle forces, their relative importance, and the re-interpretation of soil phenomena relevant to engineering applications with emphasis on processes studied by Prof. C. Ladd. Finally, the microscale analysis of forces leads to a re-interpretation of the effective stress principle and previously suggested modifications.

Particle Forces

Particle forces in soils were considered in the seminal paper by Ingles (1962), and later reviewed in Mitchell (1993) and Santamarina et al. (2001-a). Particle forces are classified herein in relation to the location of the generation mechanism:

- *Forces due to applied boundary stresses*: they are transmitted along granular chains that form within the soil skeleton. Capillary effects at high degree of saturation prior to air-entry fall under this category.
- *Particle-level forces*: includes particle weight, buoyancy and hydrodynamic forces. A particle can experience these forces even in the absence of a soil skeleton.
- *Contact-level forces*: includes capillary forces at low degree of saturation, electrical forces, and the cementation-reactive force. The first two can cause strains in the soil mass even at constant boundary loads. Conversely, the cementation-reactive force opposes skeletal deformation.

Mass-related magnetic forces (not relevant in most soils) and contact-level hydrodynamic squirt-flow type forces (that develop during dynamic excitation) are not considered in this review. The emphasis in this section is on recent developments in the understanding of particle forces, and includes simple, order-of-magnitude expressions to estimate these forces for the case of spherical particles.

Skeletal Forces (related to applied boundary stresses)

Early analyses based on spherical particles and regular packings (Deresiewicz 1973), photoelastic models (Durelli and Wu 1984), and the more recent developments in numerical micro-mechanics pioneered by Cundall (Cundall and Strack 1979) have provided unique insight into the distribution and evolution of inter-particle skeletal forces in soils.

Both normal N and tangential T forces develop at contacts when an effective stress σ' is applied at the boundary². The normal force N at a contact is related to the applied state of effective stress σ' and the particle diameter d . The first order approximation $N = d^2 \sigma'$ is appropriate for a simple cubic packing of equal size spheres. For a random packing of spheres, the mean normal contact force \underline{N} is related to the void ratio e through some empirical or semi-empirical correction functions, rendering expressions such as (for $0.4 < e < 1$)

$$\underline{N} = \sigma' d^2 \left[\frac{\pi(1+e)^2}{12} \right] \quad (1)$$

² For convenience, the more common equivalent continuum concept of "stress" is herein used to refer to the distributed force applied at the boundary.

The boundary stress is not supported uniformly by the skeletal forces N and Weibull or exponential distributions apply (Dantu 1968; Gherbi et al. 1993; Jaeger et al. 1996). Further insight is gained from photoelastic studies such as those shown in Figure 1:

- Chains of particles form columnar structures that resist the applied boundary stress (Drescher and de Josselin de Jong 1972; Oda et al. 1985). These chains resemble a fractal-type structure. The smallest scale of chains is a few particle diameters in size. Particles that form part of these chains are loaded in the direction of the applied principal stress.
- Particles that are not part of the main chains play the secondary yet very important role of preventing the buckling of the main chains. Hence, the main forces acting on these particles are transverse to the main chains (Radjai et al. 1998 - see Figure 1-a).
- There are many particles that sit within the granular medium and do not carry skeletal load. These are "movable" particles and, if smaller than the pore throats, may migrate when proper fluid flow conditions develop in the medium.
- When large pores are present, force chains arch around the pores. These arches tend to collapse during shear (see Figure 1-b).
- The stability of the columns is related to the direction of particle movement during loading, so that reversing the direction of loading promotes instability. Load history dependency is manifested even at small deformations as shown in Figure 1c (Duffy and Mindlin 1957): the stiffness contributed by the contact shear resistance would be lost if the loading direction is reversed.

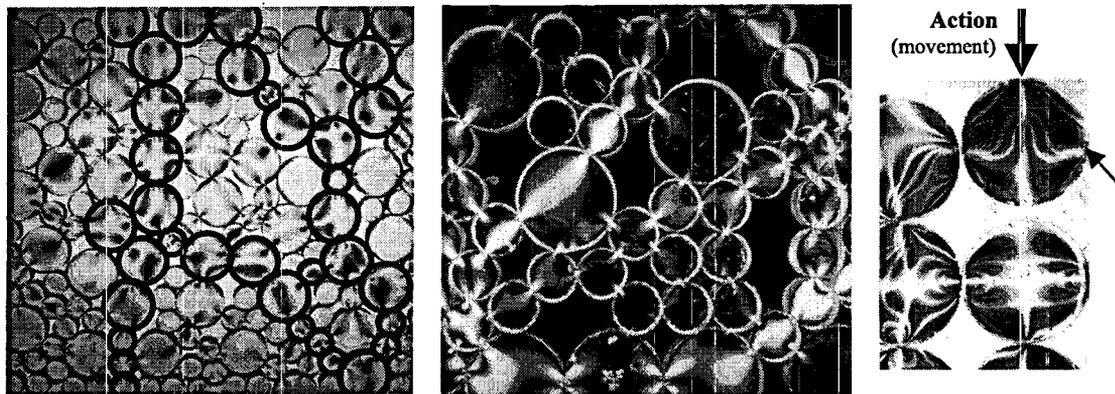


Figure 1: Skeletal force distribution – Photoelastic disks. (a) Random packing and force chains - different force directions along principal chains and in secondary particles (b) arches around large pores - precarious stability around pores. (c) Resistance mobilized during loading - contribution of shear stiffness (Courtesy of J. Valdes, M. Guimaraes, and M. Aloufi).

While the increase in mean stress promotes volume reduction and a higher coordination number (contacts) between particles, the increase in deviatoric load causes internal anisotropy in contact distribution. Inter-particle coordination and its anisotropy restrict the possible axis of rotation eventually leading to rotational arrest or frustration (Figure 2). In general, the probability for rotational frustration increases

with increasing coordination, i.e., with decreasing void ratio (the dense system shown on the right is frustrated for any rotation). Therefore, rotational frustration is overcome by either frictional slippage at contacts or by fabric changes that lead to fewer contacts among particles, i.e., decreasing coordination number which is often associated with local dilation. In fact, the higher interparticle friction, the lower the extent of frictional slippage at contacts, and the lower the coordination number that develops during shear (see Thornton 2000). While these concepts are illustrated with monosize spherical particles, rotational frustration is affected by the relative size among neighboring particles, by the ability to attain high densities in soils with high coefficient of uniformity, and by particle shape (sphericity, angularity and roughness).

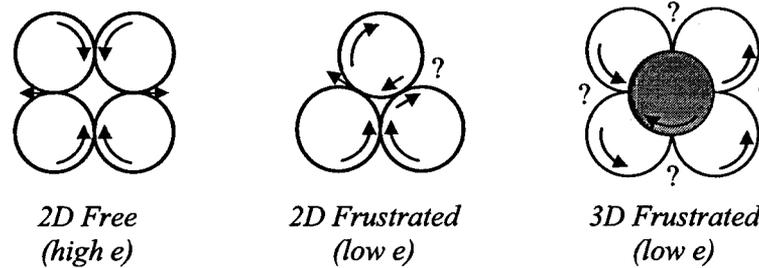


Figure 2. Rotational frustration. The lower the void ratio, the higher the number of contacts per particle and the higher the probability of rotational frustration.

These observations gain further relevance in the context of 3D micro-mechanical simulations of axial compression AC ($b=0$) and axial extension AE ($b=1$), where $b=(\sigma_2-\sigma_3)/(\sigma_1-\sigma_3)$. The distribution of contacts and average normal and shear contact forces in a given direction θ , herein denoted as $\underline{N}(\theta)$ and $\underline{T}(\theta)$, are depicted in Figure 3 (Chantawarangul 1993; see also Rothenburg and Bathurst 1989, and Thornton 2000). The following observations can be made:

- Contact normals during anisotropic loading become preferentially oriented in the direction of the main principal stress σ_1 , in agreement with observations made above (see also Oda 1972).
- The main reduction in inter-particle contacts takes place in the direction of the minor principal stress: σ_2 and σ_3 directions in AC, and σ_3 direction in AE. This situation allows for more degrees of freedom for particle rotation and for chain buckling in AC (even when the total coordination number at failure is about the same in both cases).
- Such volume-average microscale response provides insight into the observed effective peak friction angle (macroscale - numerical results presented in the lower frame of Figure 3): higher friction angle is mobilized in AE than in AC. Furthermore, the lack of particle displacement in the direction of plane strain hinders rearrangement and causes an even higher peak friction angle in plane strain loading. The critical state friction angle obtained in numerical simulations follows a similar trend, but with less pronounced differences.
- Results by Chantawarangul (1993 - not presented here) also show that early volume contraction before the peak strength, is more pronounced in AE than in AC tests – relevant to undrained strength.

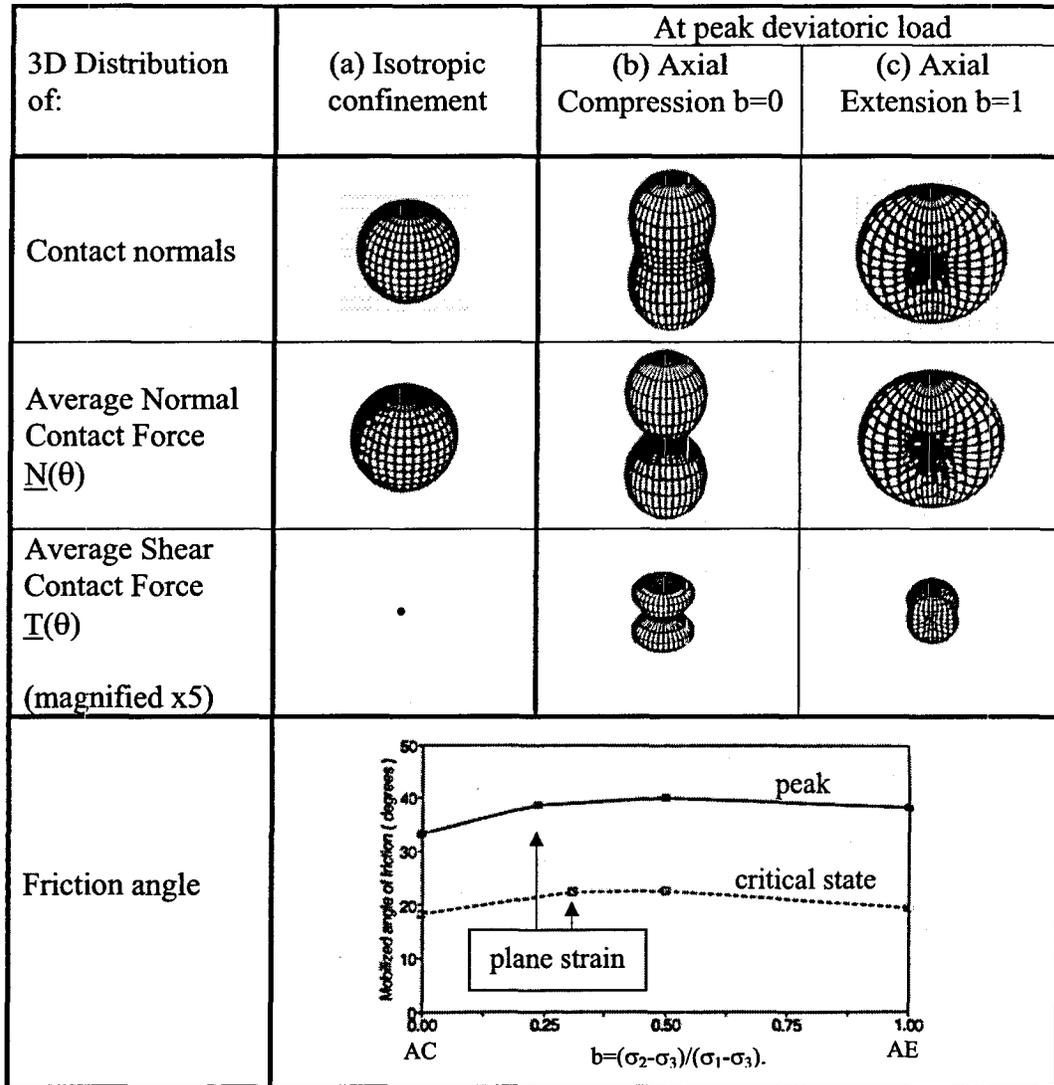


Figure 3. Numerical simulation: Evolution of inter-particle contacts and average normal and shear contact forces during axial compression and axial extension loading. Variation in friction angle with the intermediate stress σ_2 . Figure compiled from Chantawarangul (1993).

The evolution of anisotropy in contact normals and in contact forces $\underline{N}(\theta)$ and $\underline{T}(\theta)$ reflects the soil response to the anisotropic state of stress that is applied at the boundaries. Ultimately, *the shear strength of a soil is the balance between two competing trends: the reduction in coordination to minimize frictional resistance by freeing particle rotation, and the increase in coordination following the buckling of particle chains.* Therefore, the shear strength of a soil reflects the restrictions to particle motion due to either mutual frustration or boundary conditions.

Gravitational Force: Weight and Buoyancy (Particle-Level)

Newton's fourth law specifies the attraction force between two masses m_1 and m_2 at a distance r : $F_G = Gm_1m_2/r^2$, where $G = 6.673 \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2$. This force causes tides as the moon interacts with oceans, and gives rise to the weight of soil grains on the earth (e.g., m_1 is the mass of the earth and m_2 is the mass of a soil grain); for a spherical particle of diameter d is,

$$W = \frac{1}{6} \pi G_s \gamma_w d^3 \quad \text{Weight of a sphere} \quad (2)$$

where G_s is the specific gravity of the mineral that makes the particle and γ_w is the unit weight of water. The gravitational attraction F_G between two grains is much smaller than the weight of grains (about 10^{12} times smaller for millimetric particles).

Hydrostatic fluid pressure results from the weight of the fluid above the point under consideration. When a particle is submerged in water (or any other fluid), the water pressure is normal to the particle surface. The integral of the fluid pressure acting on the particle renders the buoyancy force. This force does not change, regardless of the submerged depth because the difference between the water pressure acting at the bottom and at the top of the particle remains the same, $\Delta u = \gamma_w d$ (rigid particles). In Archimedes' words, the buoyant force is equal to the weight of fluid the particle displaces, regardless of depth. For completeness,

$$U = \text{Vol} \cdot \gamma_w = \frac{1}{6} \pi \gamma_w d^3 \quad \text{Buoyant force} \quad (3)$$

The effective weight of a submerged particle becomes $W-U$.

A related experiment considers the case of two soil grains press together using a clamp, and submerged into a pond. The particles experience not only the same buoyant force but also the same inter-particle contact force due to the clamp at all depths (Figure 4). Hence, *the local pore fluid pressure around a particle does not alter the effective inter-particle skeletal force.*

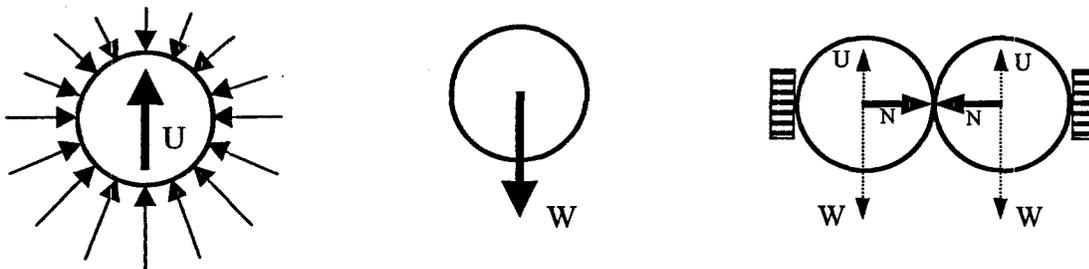


Figure 4 Hydrostatic fluid pressure, buoyant force U , weight W , and inter-particle skeletal forces N .

Hydrodynamic Force (Particle-Level)

The pore fluid moving along the interconnected pore network in the soil exerts viscous drag forces and forces resulting from velocity gradients (null average in straight macro-flow). Consider the pore fluid moving with velocity v relative to a particle of diameter d . The magnitude of the viscous drag force F_{drag} is proportional to the viscosity of the fluid μ , as predicted by Stoke's equation (applies to low Reynolds numbers - Graf 1984; Bear 1972):

$$F_{\text{drag}} = 3\pi\mu v d \quad (4)$$

where the viscosity of water at 20°C is $\mu \cong 1$ centiPoise = 0.001 N·s/m². The velocity of fluid moving through the pores in a soil is related to the hydraulic gradient i , the hydraulic conductivity k , and the porosity n of the soil, $v = ki/n$. Combining this relation with Equation 4 permits estimating the drag force experienced by a potentially movable particle sitting on a pore wall. Viscous drag also acts on the particles that form the skeleton, and together with the velocity gradient forces alters the effective stress acting on the soil (often referred to as the seepage force).

Capillary Force - Mixed Fluid Phase (Boundary to Contact-Level)

A molecule in a fluid experiences van der Waals attraction to neighboring molecules. As these forces act in all directions, they tend to cancel. However, this is not the case for water molecules at the surface of the fluid: these molecules feel an effective pull-in resultant force normal to the fluid surface. At the macroscale, this effect resembles a membrane that tries to shrink, creating a surface tension T_s which characterizes the interface ($T_s = 0.0727$ N/m for water-air at room temperature). Because this membrane tries to shrink, the fluid inside drops has positive pressure. This pore fluid pressure is computed using Laplace's equation,

$$u = T_s \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad (5)$$

where r_1 and r_2 are the curvature radii of the air-water interface. For a spherical drop, $r_1 = r_2$.

On the other hand, water tends to wet and hydrate hydrophilic mineral surfaces. When a saturated soil mass begins to dry, the gradually shrinking volume of water pulls the membrane in, while the membrane attempts to cling to the mineral surfaces around pore throats. Therefore, suction develops inside the pore fluid. The membrane tension is transmitted onto the skeleton in terms of effective stress, as in a triaxial specimen surrounded by a thin membrane and subjected to vacuum, hence, the force acting on particles at this stage develops at the *boundary*. As desiccation progresses, air gradually invades the specimen. If the fluid phase remains continuous, the medium is in the funicular regime. The average interparticle force $F_{\text{cap}} = \pi d^2 u / 4$, is

computed by invoking Equation 5. Radii r_1 and r_2 are related to the diameter d_{pore} of the largest pore at the air-water interface where the front is currently receding (the effective value for d_{pore} is about the diameter d of the particles surrounding the pore),

$$F_{\text{cap}} = \pi \frac{d^2}{d_{\text{pore}}} T_s \quad \text{Funicular regime} \quad (6)$$

At very low moisture content, disconnected fluid remains in the form of menisci at inter-particle contacts; this is the pendular regime. Figure 5 shows a sequence of microphotographs that capture the drying of water in the menisci between two spherical particles. As menisci dry, the negative pressure increases, and the cross section decreases. The contact-level capillary force computed from Laplace's equation at very low moisture content is (asymptotic solution for $w\% \rightarrow 0$; the value of F_{cap} is limited by water cavitation),

$$F_{\text{cap}} = \pi d T_s \quad \text{Pendular regime (contact-level)} \quad (7)$$

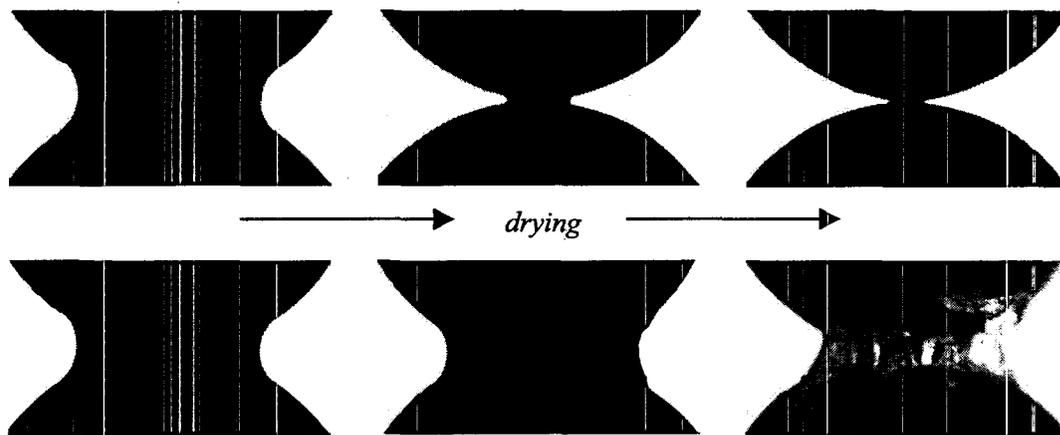


Figure 5. Evolution of unsaturation – Pendular regime. The lower spherical glass bead ($d=2\text{mm}$) is being held by the meniscus. Top sequence: de-ionized water. Lower sequence: water has salt in solution, salts eventually precipitate rendering inter-particle cementation (Gathered with D. Fratta).

Note that capillary forces in clays can be very high even at high degree of saturation, because d_{pore} is very small (Equation 6; in the presence of soluble salts and double layers, the total suction combines matric suction and osmotic suction - see Fredlund and Rahardjo 1993).

As the air-water interface gradually recedes during drying, it displaces first along the largest pores it encounters, then, the smaller pores are evacuated. Therefore, the negative pore fluid pressure increases as drying continues, and the evolution of inter-particle forces reflects the pore size distribution of the soil, as seen at the receding front. Pore size distribution is related to grain size distribution, thus, the suction-moisture plot resembles the grain size distribution of the soil (as observed in

Öberg 1997). Because of the stiffening effect of suction on soils, there is also a strong parallelism between the shear wave velocity vs. moisture plot and the grain size distribution plot (Cho and Santamarina, 2001 – continuously drying tests without remixing).

Electrical Forces (Contact-Level)

Can uncemented remolded soils exhibit cohesion? This can be tested by preparing a mud ball, submerging it in water to cancel capillary effects, and observing if the soil crumbles and reaches the angle of repose. This experiment must be carefully conducted to avoid seepage forces, entrapped air, diffusion and osmotic effects. Consider the following simple procedure to avoid these difficulties: a few grams of soil are mixed with some selected solution inside a test tube (diameter much larger than the particle size), the system is allowed to homogenize for 24 hr, a vacuum is applied to extract all the air while shaking the test tube, and the sediment is consolidated by subjecting the test tube to a high g-field in a centrifuge. Finally, the top is sealed with wax to avoid entrapping any air, and the tube is inverted. This is a very simple tension test. Kaolinite specimens prepared following these guidelines have been kept upside down for more than two years and no detachment has been observed (Figure 6). Clearly, the electrical attraction between particles is sufficient to support their buoyant weight.

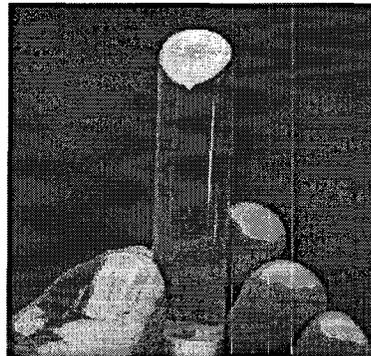


Figure 6. Electrical attraction forces are greater than the particle buoyant weight in fine soils (experimental details in the text – gathered with Y.H. Wang).

While the study of contact-level electrical forces started more than a century ago, molecular dynamic simulations and atomic force microscopy studies in the last 20 years have provided unprecedented information. The essence of all phenomena involved is the interplay between geometric compatibility, thermal agitation and Coulomb's electrical force. As a result, various repulsion and attraction forces develop, and the balance between these forces varies with inter-particle distance, rendering a highly non-linear force-distance relation. A brief review follows.

When the inter-particle distance exceeds $\sim 30\text{-}40 \text{ \AA}$, the response is well described disregarding the molecular nature of water molecules and the atomic nature of charges. Both, van der Waals attraction and electrical repulsion must be considered

(DLVO theory). The van der Waals attraction between two spherical particles of equal diameter d is (Israelachvili 1992),

$$A_{tt} = \frac{A_h}{24t^2}d \quad \text{van der Waals attraction} \quad (8)$$

where t is the separation between particles, and the Hamaker constant for silica-water-silica is $A_h=0.64 \cdot 10^{-20} \text{J}$. Repulsion results from thermal agitation and kinetic effects (also referred to as osmotic pressure or entropic confinement). The repulsion force between two spherical particles of diameter d is (derived after Israelachvili 1992 – applies to low surface potential and inter-particle distance $t > \vartheta$)

$$\begin{aligned} \text{Rep} &= 32\pi RTc_o d \vartheta e^{-\frac{t}{\vartheta}} && \text{Electrical repulsion force} && (9) \\ &\approx 0.0024 d \sqrt{c_o} e^{-10^8 t \sqrt{c_o}} \end{aligned}$$

where the gas constant $R=8.314 \text{ J}/(\text{mol}\cdot\text{K})$, T is temperature [$^{\circ}\text{K}$], c_o is the ionic concentration of the pore fluid [mole/m^3], and the double layer thickness or Debye-Huckel length in units of [m] is equal to $\vartheta=9.65 \cdot 10^{-9}(c_o)^{-0.5}$ for monovalent ions. The second simplified equation applies to monovalent ions, $T=298^{\circ}\text{K}$, distance t in [m] and the force in [N].

At smaller distances, the ionic concentration between the particles exceeds the pore fluid concentration c_o , the osmotic pressure becomes independent of c_o and it is only a function of the particle surface charge density. In turn, the effective surface charge density and the surface potential gradually decrease as particles come closer together due to ion binding (Delville, 2001). In this range, the electrostatic attraction force may exceed the van der Waals attraction, particularly when di-valent and tri-valent ions are present. Di-valent and tri-valent ions such as Ca^{2+} and Al^{3+} are most effective at shielding the electric field generated by the particle (lowering the surface potential), and if they bind to the particle, they may even render the particle positively charged. High-valence ions also interact among themselves so their positions are correlated, causing an additional attraction force between particles. Because of these effects, the presence of di-valent and tri-valent ions hinders the swelling of clays; these effects are not taken into consideration in the DLVO theory, therefore, this theory applies best to mono-valent ions.

When the inter-particle distance is less than $\sim 10\text{-}20\text{\AA}$, the discrete nature of ions and water molecules must be recognized (the size of a water molecule is $\sim 2.8 \text{\AA}$). In this range, the behavior of the particle-fluid-particle system resembles two plates with marbles in between: the molecular structure tends to be crystal-like ordered, friction is understood within the framework of thin film lubrication (Bhushan et al. 1995; Landman et al. 1996; Persson 1998), the water-ion system is organized reflecting the counterions' affinity for water and the density profile oscillates (Skipper et al. 1991; Delville 1995), inter-particle forces vary cyclically with distance with a periodicity of about one molecular diameter ("hydration force" - experimental data first reported in Horn and Israelachvili 1981), swelling progresses by discrete jumps

("hydration swelling"), and ionic mobility and diffusion are reduced (Skipper, 2001). Finally, at very high applied load, mineral to mineral interaction could develop and interpenetration is opposed by Born repulsion at the atomic level.

This framework explains most colloidal phenomena, including coagulation and swelling as a function of ionic concentration, as well as the effect of changes in fluid permittivity (changes in Hamaker constant and ion hydration). Yet, the direct application of these concepts could be misleading. For example, the sedimentation volume observed in test tubes does not decrease monotonically with ionic concentration, but often starts increasing at high concentration (Figure 7): either the form of aggregation of individual particles changes (e.g., from face-to-face "domains" to edge-to-face), or already formed domains flocculate forming open edge-to-face flocculation (Emerson 1959; Bennett and Hulbert 1986). Similar trends can be found in mechanical properties such as viscosity vs. concentration (e.g., van Olphen 1991). The transition concentration depends on clay mineralogy and pH.

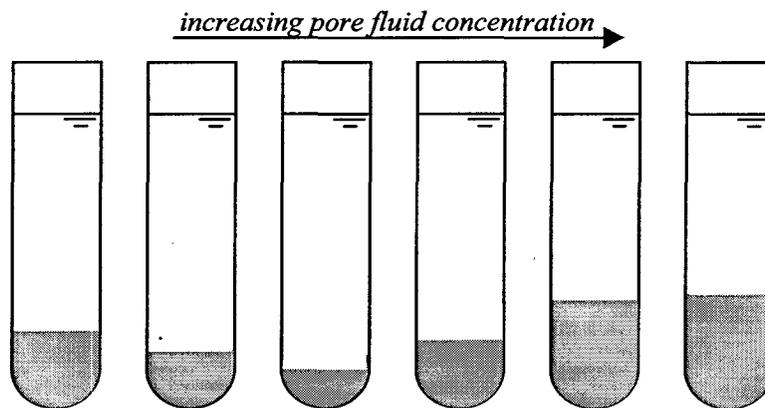


Figure 7. Electrical forces and fabric. The increase in ionic concentration above a characteristic value may render lower density fabrics (response varies with pH and mineralogy).

Cementation-reactive Force (Contact-Level)

There are many mechanisms leading to cementation. Figure 5 shows the evaporation of water in the meniscus between two particles and the precipitation of salt, forming crystals that bond the particles together. Some agents lithify the soil around particles and at contacts, while other processes change the initial physical-chemical structure (Mitchell 1993; Larsen and Chilingar 1979). Cementation is a natural consequence of aging and the ensuing diagenetic effects in soils. Most natural soils have some degree of inter-particle bonding (e.g., an ingenious device and data for London clay are presented in Bishop and Garga 1969).

Cementation is often accompanied by either shrinkage or swelling, and the ensuing changes in inter-particle skeletal forces. However, the most significant mechanical contribution of cementation is activated when strains are imposed onto the soil. To facilitate comparing this cementation-reactive force to other forces, the tensile force required to break the cement at a contact is computed herein. Consider a

homogeneous layer of cementing material of thickness t deposited all around particles of diameter d . The diameter of the cement across the contact is determined by the Pythagorean relation; for small cement thickness, $d_{\text{cont}}^2 = 4d \cdot t$. Then, for a cement with tensile strength σ_{ten} , the maximum tensile force that the contact may withstand is

$$T = \pi t d \sigma_{\text{ten}} \quad (10)$$

Even small amounts of cementation may produce significant changes in the behavior of soils if the confinement is relatively low.

Forces: Relative Relevance and Implications

The relative balance between particle forces gives rise to various regimes and phenomena in soils that affect geotechnical engineering practice. A few salient examples follow. Phenomena studied by Prof. C. Ladd or reported in two comprehensive reviews he co-authored are often invoked (Ladd et al. 1977; Jamiolkowski et al. 1985; Ladd 1991 Terzaghi Lecture).

Skeletal -vs- Contact-Level Forces

Skeletal, capillary, and van der Waals forces contribute to the normal compressive contact force (the contribution of particle weight is in the vertical downwards direction and may be compressive if sitting or tensile if the particle is hanging). Their relative contributions for different size spherical particles are depicted in Figure 8 (using previous equations; van der Waals attraction is computed for an inter-particle separation of 30 Å. The skeletal force is shown for $\sigma' = 10$ kPa and $\sigma' = 1$ MPa). These compressive forces mobilize the electrical repulsion forces and bring particles together until compression and repulsion are balanced. Changing the pore fluid can alter the inter-particle distance at equilibrium; the upper part of the figure shows the strain caused by changing the pore fluid from fresh-water to seawater concentrations (axis on the right - Equation 9 combined with Equations 1 and 8). The following observations can be made:

- Particle weight loses relevance with respect to capillary forces for particles smaller than $d \approx 3$ mm (Point 1 in the figure), and with respect to van der Waals attraction for particles smaller than $d \approx 30 \mu\text{m}$ (Point 2 in the figure).
- Capillary forces can exceed the contribution of $\sigma' = 10$ kPa confinement for particles smaller than $d \approx 20 \mu\text{m}$ (Point 3) and the contribution of $\sigma' = 1$ MPa for $d < 0.2 \mu\text{m}$ (Point 4).
- Judging by the strain level, chemical-mechanical coupling gains relevance for micron and sub-micron particles: the smaller the particles or the lower the effective confinement, the greater the effect of changes in pore fluid chemistry.
- Particles are considered "coarse" when skeletal forces due to boundary loads prevail. This is the case for particles larger than $d \approx 20 \mu\text{m}$ (Point 3).
- Particles are "fine" when contact-level capillary and electrical forces gain relevance. This is the case when particles are smaller than $d \approx 1-10 \mu\text{m}$.

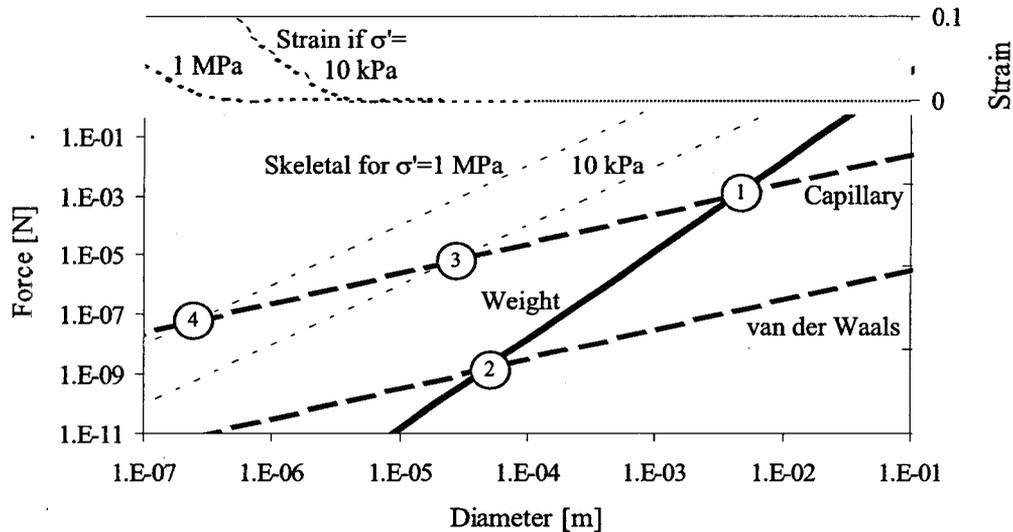


Figure 8. Skeletal vs. contact-level capillary and electrical forces. The upper part of the figure shows the strain (axis on right) caused by changing the pore fluid ionic concentration from fresh-water to seawater conditions. Note slopes: skeletal 2:1 (Equation 1), weight 3:1 (Equation 2), capillary and van der Waals 1:1 (Equations 6, 7 and 8).

The contribution of the van der Waals attraction force remains much smaller than the skeletal force under normal engineering conditions, even for very small particles. In the absence of cementation (or edge-to-face coordination), the classification of fine soils as "cohesive soils" is misleading and physically unjustified (two related views are presented in Santamarina 1997, and in Schofield 1998). Still, electrical forces determine fabric formation (Figure 7), which in turn affects soil behavior. A soil that has formed within a high ionic concentration fluid develops a characteristic fabric; if it is then leached with fresh water while confined, it may preserve the salient features of its fabric, yet, for a different inter-particle force condition. Therefore, it is not at its minimum energy configuration (as shown in Figure 7) and it is unstable. This is the case of sensitive marine clays.

The opposite case is equally important: when the soil fabric is already formed within a certain pore fluid, leaching a soil with a fluid with higher ionic concentration and/or valence produces a reduction in the inter-particle repulsion force (Equation 9), causing: a decrease in volume under a given effective stress condition, an increase in shear wave velocity, and an increase in hydraulic conductivity. Note that there are documented exceptions to these trends (an extensive compilation of experimental studies can be found in Santamarina et al. 2001-b). The response depends on the history of the test and whether stress or strain-controlled boundary conditions are imposed. Similar observations also apply to clay swelling (Ladd 1959).

Skeletal Forces -vs- Cementation - Sampling Effects

The stress-strain behavior, the strength and the volume change tendency of soils can be drastically affected by the degree of cementation (Clough et al. 1981; Lade

and Overton 1989; Airey and Fahey 1991; Reddy and Saxena 1993; Cuccovillo and Coop 1997). Two regions can be identified: the "cementation-controlled" region at low confinement, and the "stress-controlled" region at high confinement. In the cementation-controlled region, the small-strain shear stiffness can increase by an order of magnitude, the strength is cementation controlled, the buckling of chains is hindered (lower initial volume contraction), and the soil tends to brake in blocks (immediately after breaking, the inter block porosity is null, hence shear tends to cause high dilation, even if the cemented soil within the blocks has high void ratio).

The relative relevance of cementation and confinement can be identified by comparing the shear wave velocity in situ V_{S0} with the velocity in a remolded specimen $V_{s-remold}$ that is subjected to the same state of stress. In general, one should suspect cementation if $V_{S0} > V_{s-remold}$ (Note: creep and viscous effects also increase the shear wave velocity; these effects may be altered during sampling as well).

In most cases, natural cementation occurs when the soil is under confinement. When the soil is sampled, the applied confinement is removed, the center-to-center distance between particles increases, and the cement at contacts is put into tension. If the change in contact force $\Delta N = \Delta \sigma' d^2$ due to the stress reduction $\Delta \sigma'$ exceeds the tensile capacity of the cement at contacts (Equation 10), debonding occurs and the soil is permanently damaged or destructured. The cementation thickness can be related to the shear modulus of the soil G_s by considering a modified Hertzian formulation (Fernandez and Santamarina, 2001). Then, the following expression can be derived to predict the magnitude of stress reduction that can cause debonding σ'_{debond} :

$$\Delta \sigma' > \sigma'_{debond} = 1.9 \sigma_{ten} \left(\frac{G_s}{G_g} \right)^2 \quad \text{condition for de-bonding} \quad (11)$$

where σ_{ten} is the tensile strength of the bonding agent and G_g is the shear modulus of the mineral that makes the particles. The shear modulus of the soil G_s can be determined from the shear wave velocity V_{S0} measured in situ as $G_s = \rho_s V_{S0}^2$, where ρ_s is the mass density of the soil mass. The probability of debonding increases if the stress reduction $\Delta \sigma'$ approaches or exceeds the soil capacity $\Delta \sigma'_{debond}$. The upper bound for stress reduction is the in situ state of stress $\Delta \sigma' \approx \sigma'_o$ (details and data in Ladd and Lambe 1963). While G_g and σ_{ten} cannot be readily evaluated, this equation highlights the importance of the in situ shear wave velocity V_{S0} in determining whether a soil will experience sampling effects due to stress reduction. Figure 9 summarizes these observations into four regions, and provides a framework for organizing available data and further studies on sampling effects (sampling disturbance is reviewed in Jamiolkowski et al. 1985). Note that it is premature to predict the potential impact of sampling on the bases of the in-situ shear wave velocity V_{S0} alone.

De-bonding during sampling affects the behavior of both sands and clays, and it can cause important differences between the soil response measured in the laboratory

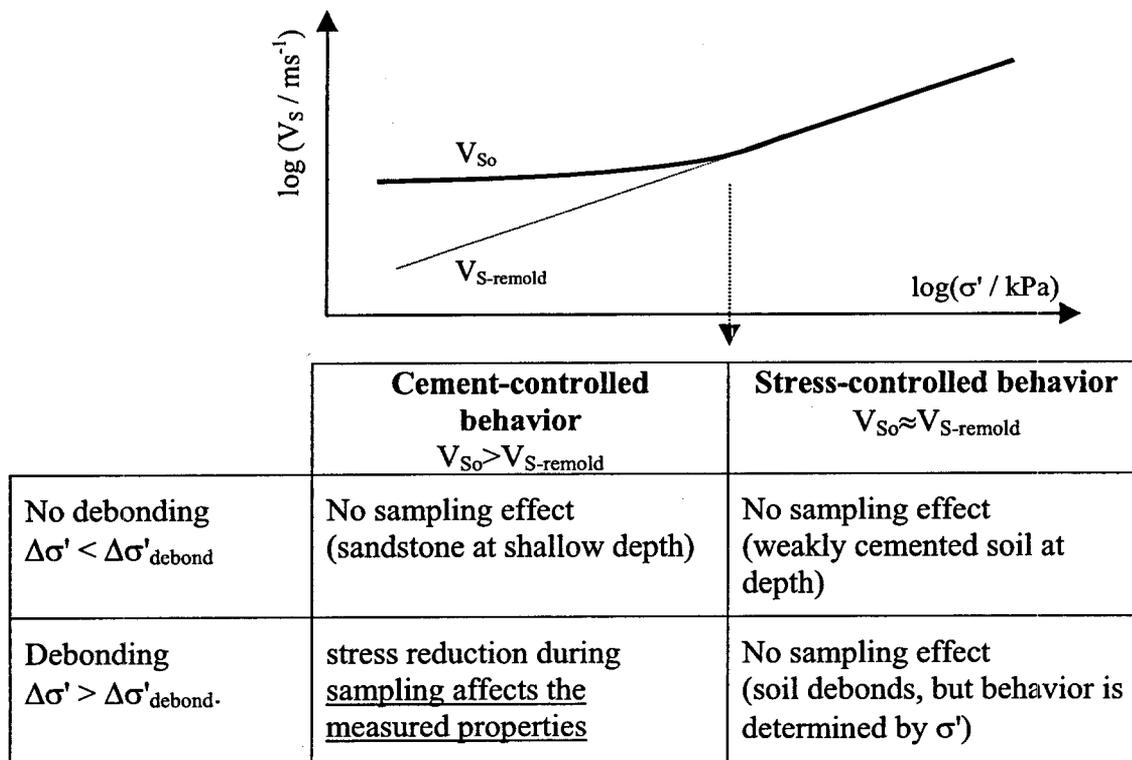


Figure 9. Skeletal forces vs. cementation strength – Sampling and debonding.

and in the field (Tatsuoka and Shibuya 1992; Leroueil, 2001; data by Stokoe published in Stokoe and Santamarina 2000).

In line with the main emphasis of this manuscript, the preceding analysis was done in terms of contact forces and stress reduction. Alternatively, the analysis can be generalized in terms of strains (or particle-level deformation) and compared against the linear and degradation threshold strain of the soil. Stress reduction is the prevailing mechanism in block sampling; with other samplers, insertion and removal of the specimens cause additional stresses, pore pressure changes and volume changes in the soil that must also be taken into consideration.

Drag Force, Weight and Electrical Forces - Fines Migration

The potential for fines migration during seepage depends on the balance between the drag force, the weight of the particle and other resisting contact forces. Figure 10 shows the drag force for different pore flow velocities in comparison with the sum of the weight and the van der Waals attraction (computed for a possible Rep-Att minimum at 30 Å and at 100 Å inter-particle distances). Notice that:

- The migration of particles greater than about 100 μm is unlikely (the required pore flow velocity would render a turbulent regime).
- The migration of particles less than ~10 μm is determined by the electrical forces. In this case, changing the pore fluid chemistry can alter the force balance. For

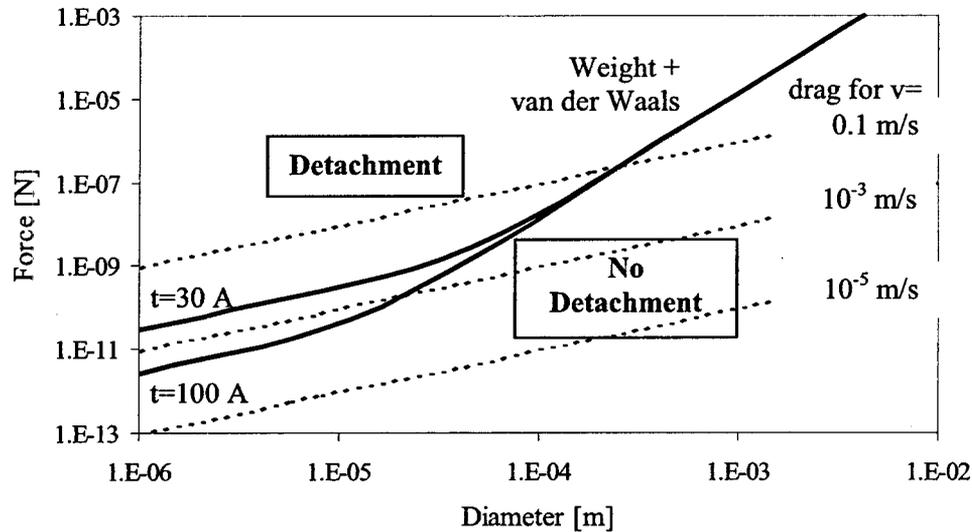


Figure 10. Drag vs. weight and net electrical attraction force.

example, a low concentration front may promote massive particle detachment allowing for particle migration.

- While individual small particles may not be detached, flocks of particles may. For the conditions considered in the figure, a minimum pore flow velocity $v \approx 10^{-3}$ m/s is needed to cause detachment of any particle. Such pore flow velocity can be attained in sands or in coarse silts at high gradients. Therefore movable particles or "fines migration" is only relevant in the coarser formations and at high gradients, such as near a well.

Particles that are dragged may be flushed out of the soil or may form bridges at pore throats clogging the soil. Flushing and clogging depend on the relative size of the pore throats between skeleton-forming particles d_{large} , the size of the smaller migrating particles d_{small} , their ability to form bridges, and the volumetric concentration of fines in the permeant (Valdes 2002). In general, the required condition for flushing to occur is $d_{large}/d_{small} > 15$ -to-30. These microscale considerations provide insight into filter criteria. Whether flushing or clogging develops, the movement of the movable particles renders fluid flow non-linear, causing pressure jumps and changes in effective stress.

Skeletal Force Distribution: Effective Stress Strength (Friction Angle)

Micromechanical analyses and simulations show the relevance of particle coordination, rotational frustration and the buckling of chains on the ability of a soil to mobilize internal shear strength (Figures 1, 2 and 3). Such analyses predict that: (1) the friction angle is highest in plane strain, then in axial extension, and least in axial compression, (2) the difference in peak friction between plane strain and AC increases with inter-particle friction and density due to the enhanced rotational frustration, and (3) the difference among critical state friction angles determined at different b -values is smaller than among peak friction values. All these

micromechanical-based predictions are matched with experimental data gathered with sands (Figure 11-a and results in references such as Cornforth 1964; Bolton 1986; Schanz and Vermeer 1996). The evidence for any variation in critical state friction angle with b is less conclusive, yet minor differences may exist. Differences in friction angle in AE and AC are observed in clays as well, as shown in Figure 11.

The effect of the intermediate stress is not captured in Coulomb's failure criterion which predicts equal frictional resistance for all b -values, and it was first considered in the model by Lade and Duncan (1975).

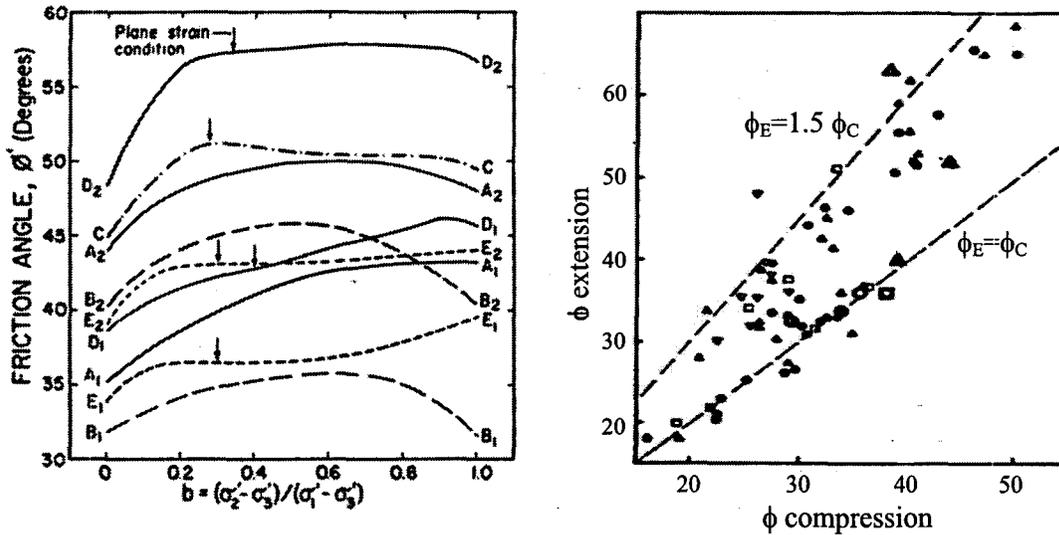


Figure 11. The effect of the intermediate stress on friction angle. (a) Sands - data from different authors compiled by Ladd et al. 1977 - Compare with Figure 3. (b) Clays: friction angle measured in AE $b=0$ and AC $b=1$ loading paths - from Mayne and Holtz (1985 - Most specimens are normally consolidated under K_0 conditions).

Skeletal Force Distribution: Undrained Strength (D_r and PI effects)

Ladd (1967) disclosed differences in undrained shear strength measured in different loading paths. As the undrained strength is controlled by the generation of pore pressure, the following microscale mechanisms should be considered:

1. The buckling of particle chains and the consequent transfer of confinement onto the pore fluid pressure. Buckling vulnerability increases in soils that have been anisotropically consolidated (Figure 3), and when subsequent loading reverses the direction of deformation and the lateral stability of columnar chains is altered. Displacement reversal also faces lower skeleton stiffness (Figure 1c). As mentioned earlier, a higher tendency to early volume contraction in AE than AC is observed in micromechanical simulations.
2. Spatial variability in void ratio and the increased instability of chains around large pores (Figure 1b).

3. Cementation (even slight), long-range electrical forces in small particles, and menisci at particle contacts in a mixed fluid-phase condition (e.g., soils with oil-water mixtures) provide contact stability and hinder buckling.
4. Inherent fabric anisotropy and its effects on skeletal stability and compressibility.

It follows from the first observation that AE loading should be more damaging than AC loading. The supportive experimental evidence is overwhelming, in sands (Hanzawa 1980, Vaid and Sivathayalan 1996, Yoshimine et al. 1999, Robertson et al. 2000, Vaid and Sivathayalan, 2000), silts (Zdravkovic and Jardine 1997), and clays (Bjerrum 1972, Ladd et al. 1977, Mayne and Holtz 1985, Jamiolkowski et al. 1985). Differences between AC and lateral extension LE, or between AE and lateral compression LC are less conclusive (e.g., Campanella and Vaid 1972; Parry 1960). Figure 12 shows data for sands and clays.

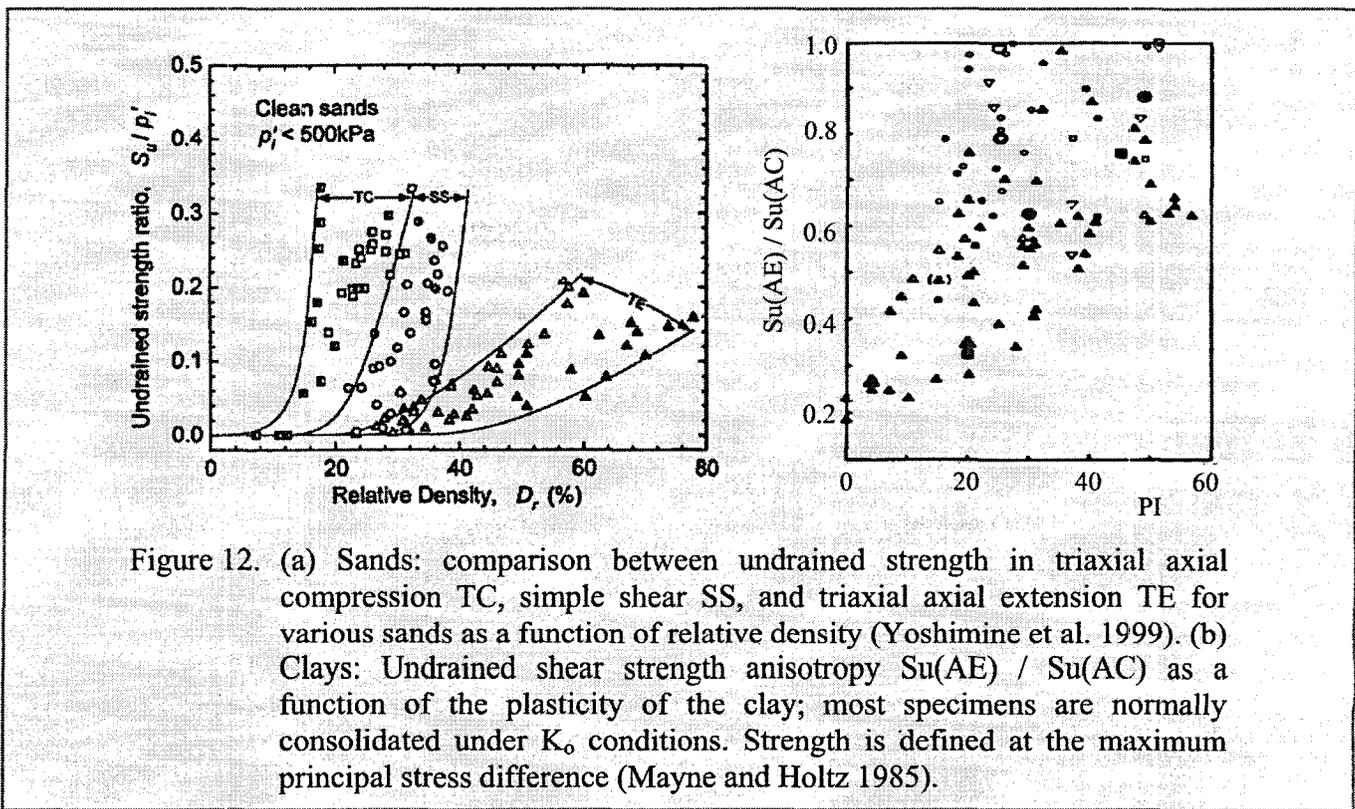


Figure 12. (a) Sands: comparison between undrained strength in triaxial axial compression TC, simple shear SS, and triaxial axial extension TE for various sands as a function of relative density (Yoshimine et al. 1999). (b) Clays: Undrained shear strength anisotropy $S_u(AE) / S_u(AC)$ as a function of the plasticity of the clay; most specimens are normally consolidated under K_0 conditions. Strength is defined at the maximum principal stress difference (Mayne and Holtz 1985).

Clearly, the collapse vulnerability of chains increases with decreasing relative density of sands, as particle coordination decreases (Figure 12-a). A related factor not captured in this figure is the effect of spatial variability of void ratio: arches around large pores are most vulnerable during shear (Figure 1b), hence, large pores tend to close first, as experimentally observed in soils (Sridharan et al. 1971; Delage and Lefebvre 1984). The spatial void ratio variability is related to specimen preparation methods (Castro 1969; Jang and Frost 1998). Hence, the undrained strength in soils is not only affected by the mean porosity, but by the pore size distribution as well (soil response for different specimen preparation methods is reviewed in Ishihara 1996).

The effect of plasticity in clays, shown in Figure 12-b, deserves special consideration. The typical role of electrical forces extends from about $h=50$ to 100 \AA . Figure 13 compares a Hertzian-type analysis of the degradation threshold strain γ_{dt} applicable to coarse elastic particles, with an analysis applicable to small rigid particles where the threshold strain is calculated for a limiting thickness h that keeps particles "in touch". The corresponding relations are:

$$\gamma_{dt} = 1.3 \left(\frac{\sigma'}{G_g} \right)^{2/3} \quad \text{Coarse grains} \quad (12)$$

$$\gamma_{dt} = 1.2 \frac{h}{d} \quad \text{Fine grains} \quad (13)$$

where d is the particle diameter, G_g is the shear modulus of the mineral that makes the grains and σ' is the applied effective confining stress.



Figure 13. Contact forces and contact deformability - Degradation threshold strain. (a) Large particles: Hertzian deformation. (b) Small particles: electrical interaction.

The distance h for relevant inter-particle electrical interaction can be related to the thickness of the double layer ϑ . The value of h is in the range of 20 \AA to 70 \AA . The liquid limit and the plastic index of a soil are proportional to h and $1/d$ (Muhunthan 1991) therefore, Equation 13 confirms the link between PI and the degradation threshold strain observed by Vucetic and Dobry (1991). Then, the higher the plasticity of the clay, the higher the degradation threshold strain, the less vulnerable force chains are to buckling, and the lower the undrained strength anisotropy in axial extension vs. axial compression (Figure 12) The stabilizing effect of electrical forces can be readily confirmed by saturating soils specimens with non-polar fluids (S. Burns, personal communication).

Strength anisotropy data reflect the combined consequences of inherent fabric anisotropy and stress-induced anisotropy. Inherent fabric anisotropy results from either particle eccentricity and/or the biasing effects of deposition in a gravitational field. Its effect on undrained strength anisotropy can be explicitly studied by rotating the direction of the specimen an angle α with respect to the deposition direction (see data for sands and clays in Ladd et al. 1977; Jamiolkowski et al. 1985; Vaid and Sivathayalan 2000; for drained response in sand: Vaid and Sayao 1995). To facilitate

the visualization of depositional anisotropy effects, consider the extreme anisotropic packing of platy particles illustrated in the sketches on the left side of Figure 14: all contact normals and normal contact forces are in the vertical direction, all particle axes and contact shear forces are in the horizontal direction, and all pores are of equal size and geometry. Clearly, particle slenderness enhances contact and force anisotropy. The following responses are expected in the mind experiments proposed in the sketches:

- AC ($b=0, \alpha=0$) causes minimum pore pressure generation.
- AE or LC ($b=1, \alpha=90$) causes high initial pore pressure, followed by the development of kinematic constraints and possible dilatancy after "phase transformation".
- Simple shear SS ($b>0, \alpha>0$) may not only produce excess pore pressure but also a failure that is aligned with particle orientation so minimum dilatancy may be mobilized.

The structure sketched in Figure 14 could form during the slow deposition of large platy particles, such as mica platelets, so that gravity prevails over electrical forces. However, particles with slenderness as low as $\sim 1.1/1.0$ can confer large fabric anisotropy effects to the soil (Rothenburg 1993). This is the case not only in clays but also in most sands; data for Fraser River sand is presented in Figure 14³.

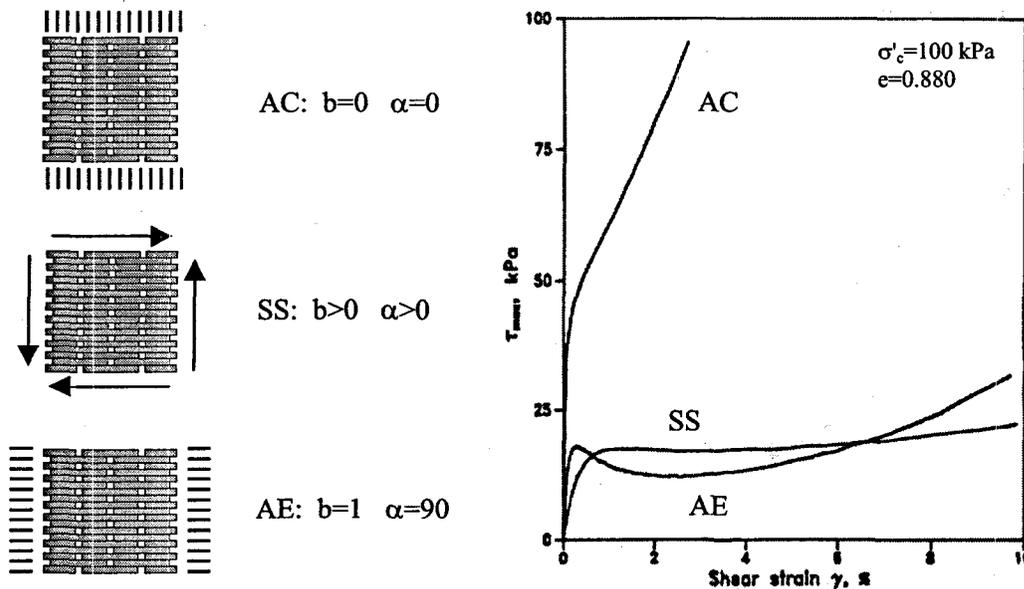


Figure 14. Inherent anisotropy effects on undrained strength. Left: conceptual models. Right: data for sands from Vaid and Sivathayalan (1996).

³ Note: most numerical micromechanical simulations do not include gravity and depositional anisotropy (e.g., simulation in Figure 3). Slender particles accentuate the effects of stress induced anisotropy.

Spatial and Temporal Scales in Particle Forces

The variation in sedimentation volume with concentration shown in Figure 7 and the distribution of inter-particle skeletal forces forming particle chains shown in Figure 1 highlight the presence of multiple internal spatial scales in the medium. These scales add scale-dependent phenomena. For example, the distribution of skeletal forces (which reflects the interplay between chain buckling and rotational frustration) causes an uneven displacement field related to the mobilization of normal and shear forces at particle contacts and the threshold for frictional slippage, $T_{ult} = N\mu$. Individual particles move together by forming wedges that displace relative to each other along inter-wedge planes where the deformation localizes; eventually, the displacement becomes kinematically restricted, columns buckle, wedges break and new inter-wedge planes form. This behavior can be readily verified by assembling a 2D random packing of coins on a flat surface and enforcing the displacement of one boundary (Figure 15 – see Drescher and de Josselin de Jong 1972). Note that domains made of fine particles form conglomerates that can move in wedges as coarse grains; this observation can facilitate explaining the similarities between fine and coarse grained soil response, such as in Figure 11.

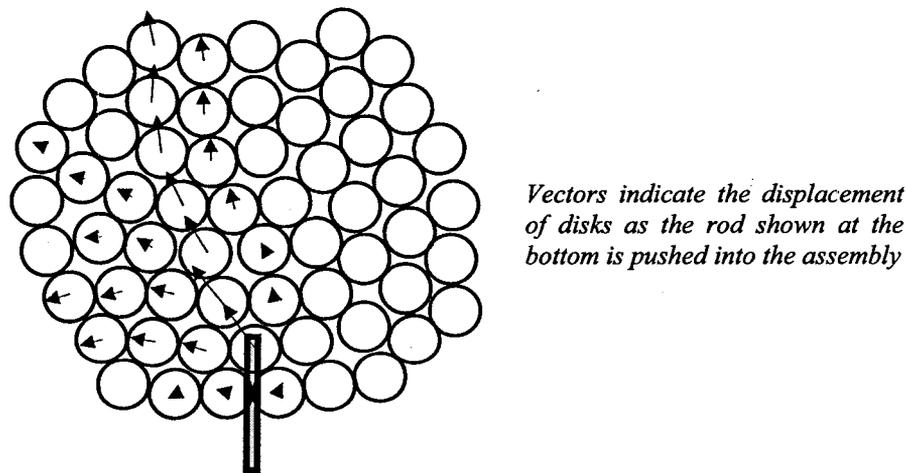


Figure 15. Localization at the particle level – Wedges. At the particle level, the deformation in granular media is inherently uneven. (Test procedure: pennies on a scanner).

At larger scales, the localization of deformation leads to the development of shear bands, where the buckling of particle columns tends to concentrate (Oda and Kazama 1998). Therefore, the localization of deformation is an inherent characteristic of particulate materials, and it has been observed in dense-dilative soils under drained loading and under undrained loading (if cavitation is reached), loose-contractive soils under undrained loading, lightly cemented soils, unsaturated soils, soils with platy or rod-like particles, and heterogeneous soils (e.g., Vardoulakis 1996; Finno et al. 1997;

Saada et al. 1999; Mokni and Desrues 1998; Cho 2001). The development of global localization affects the interpretation of laboratory data, including results presented in Figures 11 and 12: the strain level at the formation of a shear band is a function of b (Lanier et al. 1997, Wang and Lade, 2001), peak strength becomes specimen-size dependent as a function of soil rigidity and brittleness due to the associated "progressive failure", the measured global void ratio at large strain deviates from the critical state void ratio (Desrues et al. 1996), and the interpretation of the critical state friction angle is affected by the inclination of the shear band.

Particle forces also experience time-effects. The following examples apply to the different particle forces addressed above:

- *Normal and shear skeletal forces - Creep.* Two mechanisms are identified: First, material creep within particles near the contact (Kuhn and Mitchell 1993; Rothenburg 1993). Second, frictional slip; in this case the time scale at the particle level is determined by the inertial effects and the stress drop σ , $t = \sqrt{d^2 \rho / \sigma}$. Supporting evidence is obtained with acoustic emission measurements.
- *Pore fluid pressure - Transient.* The most common time dependent effect in soils is the diffusion of excess pore pressure. While the boundary-level effect is considered in standard practice (this is the typical case in Terzaghi's consolidation), pressure differences also develop at the level of pores, for example, in dual porosity soils. The time scale for the dissipation of these local gradients or pressure diffusion is related to the internal length scale L_{int} and the internal coefficient of consolidation c_{v-int} which depends on the skeletal stiffness and permeability, $t = L_{int}^2 / c_{v-int}$.
- *Capillary force.* When an unsaturated soil is subjected to shear, equilibrium in the water-air potentials is not regained immediately. When the water phase is continuous (funicular regime), the time scale is determined by the hydraulic conductivity of the medium and tends to be short. When the water phase is discontinuous and remains only at contacts (pendular regime), pressure homogenization occurs through the vapor pressure and is very slow. Evidence gathered with shear wave velocity is presented in Cho and Santamarina (2001).
- *Electrical forces.* Forcing a relative displacement between particles alters the statistical equilibrium position of counterions around particles, hence, the inter-particle forces. The time for ionic stabilization can be estimated using the ionic diffusion coefficient D , as $t = d^2 / D$. Time varying changes in electrical conductivity after remolding a soil specimen was observed by Rinaldi (1998). Additionally, viscous effects take place during the transient.
- *Cementation.* Diagenesis is clearly time-dependent and depends on the rate of chemical reactions and diffusion.

Particle-level time scales and macro-scale time scales in soils can be very different. The steady-state distribution of forces within a particulate medium presumes that equilibrium has been reached at all particles. When equilibrium is not attained at a given particle, the particle displaces and alters the equilibrium of its neighbors. Therefore, while it would appear that the particle-level time scales listed above tend to be short in general, their manifestation at the level of the soil can extend for a long period of time due to the large number of particles that are recursively involved, even if only the particles along main chains are considered.

Time-dependency in particle forces is related to macroscale phenomena observed in soils such as creep, strain rate effects on strength, secondary consolidation, thixotropy, aging, pile capacity, and changes in penetration resistance (Mitchell 1960, Kulhawy and Mayne 1990, Mesri et al. 1990, Schmertmann 1991, Díaz-Rodríguez and Santamarina 1999).

The coexistence of multiple internal spatial and temporal scales in particle forces hints to important potential effects when trying to relate laboratory measurements to field parameters (strength, stiffness, diffusion and conduction parameters). Thus, parameters obtained in the lab must be carefully interpreted in order to select design parameters.

Reassessing Customary Concepts

Many commonly used concepts and accepted soil phenomena gain new clarity when they are re-interpreted at the level of particle forces. The principle of effective stress and the phenomenon of hydraulic fracture in soils are briefly addressed next.

Effective Stress and Modified Effective Stress Expressions

The concept of effective stress plays a pivotal role in understanding and characterizing soil behavior. Several earlier observations are relevant to the concept of effective stress, in particular:

- The hydrostatic pore pressure around a particle provides buoyancy. The intensity of the pore pressure around a particle does not affect the skeletal force transmitted between particles (Figure 4).
- Hydraulic gradients i cause fluid flow. The ensuing drag and velocity gradient forces act on particles and alter the effective stress transmitted by the skeleton.
- Changes in electrical and pendular capillary forces produce changes in volume, stiffness and strength, particularly in the finer soils and at low confinement (Figure 8).

Skeletal forces are defined at the boundaries (e.g., membrane in a triaxial specimen or equipotential lines in seepage) while other forces are determined at the particle or contact levels. Therefore the impact of these forces on soil behavior is different, and mixing both types of forces in a single algebraic expression can lead to incorrect predictions (as observed in Bishop and Blight 1963, for unsaturated soils). For example, some soils collapse upon wetting even though suction decreases and expansion should be expected; others experience a decrease in stiffness with an increase in ionic concentration even though a lower repulsion force is expected.

It follows from this discussion, that the use of modified effective stress expressions to accommodate suction or electrical repulsion and attraction forces should be discontinued (modified effective stress expressions are tabulated in Santamarina et al. 2001-a). Instead, behavior should be re-interpreted taking into consideration the separate and independent contributions of the skeletal force due to boundary loads and the other contact-level forces. This has been recognized in unsaturated media (Fredlund and Morgenstern 1977; Alonso et al. 1990), but it still

requires further attention in the context of contact-level electrical forces (developments in constitutive modeling can be found Gajo et al. 2001; Guimaraes et al. 2001; Frijns et al. 1997).

Hydraulic Fracture in Soils

Hydraulic fracturing is intentionally or unintentionally produced in soil masses in a wide range of situations: thermal changes (both in the laboratory and in in situ, such as in thermal ground improvement techniques - Figure 16-a), as an experimental tool to determine the state of stress (similar to applications in intact rocks – reviewed in Ladd et al. 1977), it has been hypothesized as a failure mechanism in the failure of large dams (including Teton dam – exacerbated by arching and stress redistribution), it has been used in the context of deformation control (Mair and Hight 1994; Jafari et al. 2001), it is routinely utilized to increase the hydraulic conductivity (in view of enhanced oil recovery or even in decontamination strategies), and it occurs during grouting (even when compaction grouting is intended).

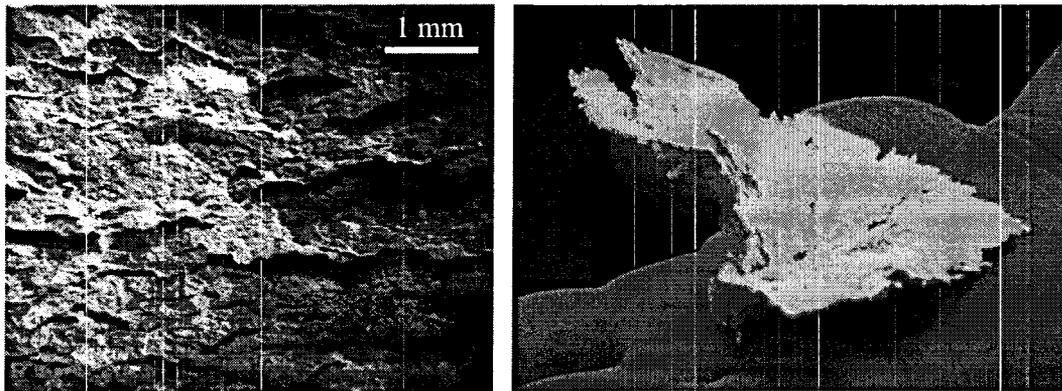


Figure 16. Hydraulic fracture in soils. It can take place in both fine grained soils (electrical attraction greater than particle weight) and in uncemented coarse grained soils (electrical attraction is irrelevant). (a) Kaolinite specimen mixed with water at the LL, and subjected to microwave radiation for fast heating. (b) Grouting in Ottawa sand: the picture shows the fracture that formed by grouting gypsum after the gypsum hardened and was retrieved - obtained in cooperation with L. Germanovich.

How can hydraulic fracture take place in soils? Current fracture mechanics theories apply to media with tensile strength and fracture propagation involves tensile failure at the tip. Yet, there is virtually no tensile capacity in uncemented soils. For clarity, consider uncemented dry sands where the adhesion force between particles is much smaller than the weight of particles (Figure 16-b; refer to Figure 8). The stress anisotropy in soils cannot exceed the limiting anisotropy determined by friction, say $\sigma_1/\sigma_3 = \tan^2(45 + \phi/2)$. Then, it appears that hydraulic fracturing in soils is the result of yield at the tip (stress path similar to AE), probably combined with other phenomena

such as hydrodynamic drag and/or cavitation of the pore fluid in the process zone. From this perspective, the application of standard fracture mechanics theory invoking the undrained shear strength of the soil as a form of cohesion violates the fundamental behavior of soils.

Closing remarks

Summary of Main Observations. Discerning soil behavior at the microscale brings enhanced physical meaning and understanding that can be applied to comprehend both available results and new measurements. Such insight guides research as well as the understanding of soil behavior in view of engineering applications. The main observations follow:

- Soils are particulate materials. Therefore, particle forces determine soil response. Particle forces are interrelated with particle characteristics (including size and distribution, slenderness, mineralogy), contact behavior and fabric .
- The behavior of coarse-grained soils is controlled by skeletal forces related to boundary stresses. On the other hand, the finer the particle and the lower the effective confinement, the higher the relevance of contact-level electrical and capillary forces. The transition size from coarse to fine for standard engineering applications is around $d \approx 10 \mu\text{m}$.
- Hydrodynamic forces alter the skeletal forces and can displace movable particles. Either fines migration or clogging renders non-linear fluid flow, and affects the pore pressure distribution and the effective stress.
- Cementation, even if small, alters stiffness, strength and volume change tendencies. Two regions can be identified: the low-confinement cementation-controlled region and the high-confinement stress-controlled region.
- While previously suggested modified effective stress expressions can incorporate the various contact-level forces, they are physically incorrect and can lead to inadequate predictions. Therefore, their use should be discontinued.
- The drained strength of a soil reflects the balance between two competing micro-processes: the decrease in inter-particle coordination to reduce rotational frustration (in order to minimize friction), and the buckling of chains (that increases coordination).
- The undrained strength is determined by the (tendency to) volume compressibility of the skeleton, which depends on the vulnerability of load carrying chains.
- The degradation threshold strain in coarse soils increases with the applied load and decreases with the stiffness of the particles. In fine grain soils, it increases with the effective distance of electrical forces and with decreasing particle size, hence, the higher the plastic index the higher the degradation threshold strain.
- There are multiple internal scales inherent to particle forces in soils (both spatial and temporal).
- The re-interpretation of common concepts at the level of particle forces, such as "cohesive soil", "effective stress" and "hydraulic fracture", provides enhanced insight into soil behavior.

Current Capabilities. The soil researcher today has exceptional experimental and numerical tools to study soil behavior, including particle forces. Likewise, today's geotechnical engineer can use robust, versatile and inexpensive numerical capabilities, standard testing methods, and a new generation of testing principles and procedures. These tools must be implemented within the framework of a clear understanding of soil behavior.

Challenging future. There are abundant fascinating research questions in soil behavior, including: the implications of omnipresent strain-localization and multiple internal scales on soil properties and engineering design; the reinterpretation of friction in light of a large number of new studies (coarse and fine soils) and harnessing friction through inherent noise-friction interaction; new characterization-rehabilitation methodologies based on dynamic energy coupling; bio-geo phenomena; methane hydrates (characterization and production); and subsurface imaging (similar to medical diagnosis) combined with the determination of engineering design parameters or within the framework of an advanced observational approach.

Today's availability and easy access to information were unimaginable only 20 years ago. Yet, paradoxically, this great facility to access almost unlimited information appears to enhance the risk of forgetting knowledge. Indeed, we are challenged not only to address new fascinating questions, but also to preserve the great insight and understanding developed by the preceding generations. This should be our commitment today, as we celebrate Prof. C. Ladd's leading example.

Acknowledgements

Prof. Ladd's contributions have profoundly enhanced our understanding of soil behavior and its implementation in engineering design. Many colleagues provided valuable comments and suggestions, including: J. Alvarellos, S. Burns, G. Cascante, G.C. Cho, D. Fratta, D. Frost, L. Germanovich, M.S. Guimaraes, I. Herle, K.A. Klein, J.S. Lee, P. Mayne (his knowledge of the literature is enviable), A.M. Palomino, G. Rix, J.R. Valdes, V. Uchil (for the many runs to the library), Y.H. Wang, and exceptional anonymous reviewers. References acknowledge the insight gained from other authors. Funding was provided by the National Science Foundation.

References

- Airey, D.W. and Fahey, M. (1991), Cyclic Response of Calcareous Soil from the North-West Shelf of Australia, *Géotechnique*, vol. 41, no. 1, pp. 101-121.
- Alonso, E.E. Gens A. and Josa, A. (1990), A constitutive Model for Partially Saturated Soils, *Géotechnique*, vol. 40, no. 3, pp. 405-430.
- Bear, J. (1972), *Dynamics of Fluids in Porous Media*, Dover, New York, 764 pages.
- Bennett, R.H. and Hulbert, M.H. (1986), *Clay Microstructure*, International Human Resources Development Corporation, Boston, 161 pages.
- Bhushan, B., Israelachvili, J. N. and Landman, U. (1995), Nanotribology: Friction, wear and lubrication at the atomic scale, *Nature*, vol. 374, pp. 607-616.

- Bishop, A. W. and Blight, G. E. (1963), Some aspects of effective stress in saturated and partly saturated soils, *Géotechnique*, vol. 13, pp. 177-196.
- Bishop, A.W. and Garga, V.K. (1969), Drained Tension Tests on London Clay, *Géotechnique*, vol. 19, pp. 309-313.
- Bjerrum, L. (1972), Embankments on Soft ground, *Proc. Earth & Earth-Supported Structures (Purdue Conference)*, vol. 2, ASCE, Reston, Virginia, pp 1-54.
- Bolton, M. D. (1986), The strength and dilatancy of sands, *Géotechnique*, vol. 36, pp. 65-78.
- Campanella R.G. and Vaid, Y.P. (1972), A simple Ko Triaxial Cell, *Canadian Geotechnical Journal*, vol. 9, pp.249-260.
- Castro G. (1969), Liquefaction of Sands, Ph.D. Dissertation, Harvard University, Massachusetts, 112 pages.
- Chantawarangul, K. (1993), Numerical Simulation of Three-Dimensional Granular Assemblies, Ph.D. Dissertation, University of Waterloo, Ontario, 219 pages.
- Cho, G.C. and Santamarina, J.C. (2001), Unsaturated Particulate Materials - Particle Level Studies, *ASCE Geotechnical Journal*, vol. 127, no. 1, pp. 84-96.
- Cho, G.C. (2001), Unsaturated Soil Stiffness and Post Liquefaction Shear Strength (Chapter III: Omnipresent Localizations), PhD Dissertation, Georgia Institute of Technology, 288 pages.
- Clough, G.W., Sitar, N., Bachus, R.C. and Rad, N.S. (1981), Cemented Sands Under Static Loading. *Journal of Geotechnical Engineering, ASCE*, vol. 107, no. 6, pp. 799-817.
- Cornforth, D.H. (1964), Some Experiments on the Influence of Strain Conditions on the Strength of Sand, *Géotechnique*, vol. 14, no. 2, pp. 143-167.
- Cuccovillo, T. and Coop, M.R. (1997), Yielding and Pre-Failure Deformation of Structure Sands, *Géotechnique*, vol. 47, no. 3, pp. 491-508.
- Cundall, P.A. and Strack, O.D.L. (1979), A discrete Numerical Model for Granular Assemblies, *Geotechnique*, vol. 29, no. 1, pp. 47-65.
- Dantu, P. (1968), Etude Statistique des Forces Intergranulaires dans un Milieu Pulverulent, *Géotechnique*, vol. 18, pp. 50-55.
- Delage, P. and Lefebvre, G. (1984), Study of the structure of a sensitive Champlain clay and of its evolution during consolidation, *Canadian Geotechnical Journal*, vol. 21, pp. 21-35.
- Delville, A. (1995), Monte Carlo Simulations of Surface Hydration: An Application to Clay Wetting, *J. Phys. Chem.* vol. 99, pp. 2033-2037.
- Delville, A. (2001), The Influence of Electrostatic Forces on the Stability and the Mechanical Properties of Clay Suspensions, *Proc. Workshop on Clay Behaviour: Chemo-Mechanical Coupling*, Edts., C. Di Maio, T. Hueckel, and B. Loret, Maratea, Italy. 20 pages.
- Deresiewicz, H. (1973), Bodies in contact with applications to granular media, *Mindlin and Applied Mechanics*, Edited by B. Herrman, Pergamon Press, New York.
- Desrues, J., Chambon, R., Mokni, M. and Mazerolle, F. (1996), Void Ratio Evolution Inside Shear Bands in Triaxial Sand Specimens Studied by Computed Tomography, *Geotechnique*, vol. 46, no. 3, pp. 529-546.

- Díaz-Rodríguez, J.A. and Santamarina, J.C. (1999), Thixotropy: The Case of Mexico City Soils, XI Panamerican Conf. on Soil Mech. and Geotech. Eng., Iguazu Falls, Brazil, Vol. 1 pp. 441-448.
- Drescher, A. and de Josselin de Jong, G. (1972), Photoelastic Verification of a Mechanical Model for the Flow of a Granular Material, *J. Mechanics and Physics of Solids*, vol. 20, pp. 337-351.
- Duffy, J. and Mindlin, R. D. (1957), Stress-strain relations of a granular medium, *Journal of Applied Mechanics*, vol. 24, pp. 585-593.
- Durelli, A. J. and Wu, D. (1984), Loads between disks in a system of discrete elements, *Experimental Mechanics*, December, pp. 337-341.
- Emerson, W.W. (1959), The structure of Soil Crumbs, *J. Soil Science*, vol. 10, no. 2, pp. 234.
- Fernandez, A. and Santamarina, J.C. (2001), The Effect of Cementation on the Small Strain Parameters of Sands, *Canadian Geotechnical Journal*, vol. 38, no. 1, pp. 191-199.
- Finno, R.J., Harris W.W., Mooney, M.A. and Viggiani, G. (1997), Shear Bands in Plane Strain Compression of Loose Sand, *Géotechnique*, vol. 47, no. 1, pp. 149-165.
- Fredlund, D. G. and Morgenstern, N. R. (1977), Stress state variables for unsaturated soils, *Journal of Geotechnical Engineering, ASCE*, vol. 103, pp. 447-466.
- Fredlund, D. G., and Rahardjo, H. (1993). *Soil Mechanics for Saturated Soils*. John Wiley and Sons, New York.
- Frijns, A.J.H., Huyghe, J.M. and Janssen, J.D. (1997), A Validation of the Quadriphasic Mixture Theory for Intervertebral Disc Tissue, *Int. Journal Engineering Science*, vol. 35, pp. 1419-1429.
- Gajo, A., Loret, B. and Hueckel, T. (2001), Electro-Chemo-Mechanical Coupling in Homoionic and Heteroionic Elastic-Plastic Expansive Clays, *Proc. Workshop on Clay Behaviour: Chemo-Mechanical Coupling*, Edts., C. Di Maio, T. Hueckel, and B. Loret, Maratea, Italy. 26 pages.
- Gherbi, M. Gourves, R. and Oudjehane, F. (1993), Distribution of the Contact Forces Inside a Granular Material, *Powders and Grains 93*, *Prod. 2nd International Conference on Micromechanics of Granular Media*, Brimingham, Elsevier, Amsterdam, pp. 167-171.
- Graf, W.H. (1984), *Hydraulics of Sediment Transport*, Water Resources Publications, Highlands Ranch, Colorado, 513 pages.
- Guimaraes, L., Gens, A, Sanchez, M. and Olivella, S. (2001), A Chemo-Mechanical Model for Unsaturated Expansive Clays, *Proc. Workshop on Clay Behaviour: Chemo-Mechanical Coupling*, Edts., C. Di Maio, T. Hueckel, and B. Loret, Maratea, Italy. 28 pages.
- Hanzawa, H. (1980), Undrained Strength and Stability Analysis for a Quick Sand, *Soils and Foundations*, vol. 20, no. 2, pp. 17-29.
- Horn, R.G. and Israelachvili, J. (1981), Direct Measurement of Structural Forces Between Two Surfaces in a Non-Polar Liquid, *J. Chemical Physics*, vol. 75, no. 3, pp. 1400-1411.
- Ingles, O.G. (1962), Bonding Forces in Soils - Part 3, *Proc. Of the First Conference of the Australian Road Research Board*, vol. 1, pp. 1025-1047.

- Ishihara, K. (1996), *Soil Behaviour in Earthquake Geotechnics*, Calderon Press, Oxford, 350 pages.
- Israelachvili, J. (1992), *Intermolecular and Surface Forces*, Academic Press, New York, 450 pages.
- Jafari, M.R., Au, S.K., Soga, K., Bolton, M.D. and Komiya, K. (2001), *Fundamental Laboratory Investigation of Compensation Grouting in Clay*, Cambridge University, 15 pages.
- Jaeger, H. M., Nagel, S. R. and Behringer, R. (1996), The physics of granular materials, *Physics Today*, April, pp. 32-38.
- Jamiolkowski, M., Ladd, C.C., Germaine, J.T. and Lancellotta, R. (1985), New Developments in Field and Laboratory Testing of Soils, *Proc. 11th ICSMFE*, San Francisco, vol. 1, pp 67-153.
- Jang, D.J. and Frost, J.D. (1998), Sand Structure Differences Resulting from Specimen Preparation Procedures, *Prod. ASCE Conf. Geotechnical Earthquake Engineering and Soil Dynamics*, Seattle, vol. 1, pp. 234-245.
- Kulhawy, F.H. and Mayne, P.W. (1990), *Manual on Estimating Soil Properties for Foundation Design*, EPRI EL-6800, Electric Power Research Institute, Palo Alto.
- Kuhn, M. R. and Mitchell, J.K. (1993), New Perspectives on Soil Creep, *ASCE J. Geotechnical Engineering*, 119(3): 507-524.
- Ladd, C.C. (1959), Mechanisms of Swelling by Compacted Clays, *Highway Research Board Bulletin* 245, pp. 10-26.
- Ladd, C.C. and Lambe, T.W. (1963), The strength of Undisturbed Clay Determined from Undrained Tests, *NRC-ASTM Symposium on Laboratory Shear Testing of Soils*, Ottawa, ASTM STP 361, pp. 342-371.
- Ladd, C.C. (1967), Discussion, *Proc. Geotechnical Conference on Shear Strength Properties of Natural Soils and Rocks*, NGI, Oslo, vol. 2, pp. 112-115.
- Ladd, C.C., Foott, R., Ishihara, K., Schlosser, F. and Poulos, H.G. (1977), Stress-Deformation and Strength, *Proc. 9th ICSMFE*, vol. 2, Tokyo, pp. 421-494.
- Ladd, C.C. (1991), Stability Evaluation during Staged Construction, *J. Geotechnical Engineering*, vol. 117, no. 4, pp. 540-615.
- Lade, P.V. and Duncan, J.M. (1975), Elasto-plastic Stress-Strain Theory for Cohesionless Soil, *ASCE Journal of Geotechnical Engineering*, vol. 101, no. 10, pp. 1037-1053.
- Lade, P.V. and Overton, D. D. (1989), Cementation effects in Frictional Materials, *ASCE Journal of Geotechnical Engineering*. 115, pp. 1373-1387.
- Landman, U., Luedtke, W.D. and Gao, J.P. (1996), Atomic-Scale Issues in Tribology: Interfacial Junctions and Nano-Elastohydrodynamics. *Langmuir*, vol. 12, pp. 4514-4528.
- Lanier, J., Viggiani, G. and Desrues, J. (1997), Discussion to the paper by J. Chu, R. Lo and I.K.Lee entitled Strain Softening and Shear band Formation of Sand in Multi-axial Testing, *Geotechnique*, vol. 47, no. 5 pp. 1073-1077.
- Larsen, G. and Chilingar, G.V. (1979), *Diagenesis in Sediments and Sedimentary Rocks*, Elsevier Scientific Publishing Company, New York, N.Y.
- Leroueil, M.S. (2001) Natural Slopes and Cuts: Movement and Failure Mechanisms, *Géotechnique*, vol. 51, no. 3, pp. 195-244

- Mair, R.J. and Hight, D.W. (1994), Compensation Grouting, *World Tunneling*, November, pp. 361-367.
- Mayne, P.W. and Holtz, R.D. (1985), Effect of principal stress rotation on clay strength, *Proc. 11th ICSMFE, San Francisco*, vol. 2, pp 579-582.
- Mesri, G., Feng, T.W. and Benak, J.M. (1990), Postdensification Penetration Resistance of Clean Sands, vol. 116, no. 7, pp. 1095-1115.
- Mitchell, J.K. (1993), *Fundamentals of Soil Behavior*, J.Wiley, New York, 437 pages.
- Mitchell, J.K. (1960), Fundamentals Aspects of Thixotropy in Soils, *Journal of the Soil Mechanics and Foundations Division, ASCE*, vol. 86, no. 3, pp. 19-52.
- Mokni, M. and Desrues, J. (1998), Strain Localization Measurements in Undrained Plane-Strain Biaxial Tests on Hostun RF Sand, *Mechanics of Cohesive-Frictional Materials*, vol. 4, pp. 419-441.
- Muhunthan, B. (1991), Liquid Limit and Surface Area of Clays, *Géotechnique*, vol. 41, pp. 135-138.
- Öberg, A.L. (1997), *Matrix Suction in Silt and Sand Slopes*, Ph.D. Dissertation, Chalmers University of Technology, Sweden, 160 pages.
- Oda, M., Nemat-Nasser, S. and Konishi, J. (1985), Stress-Induced Anisotropy in Granular Masses, *Soil and Foundation*, vol. 25, pp. 85-97.
- Oda, M. (1972), The mechanism of fabric changes during compressional deformation of sand, *Soils and Foundations*, vol. 12, no. 2, pp. 1-18.
- Oda, M. and Kazama, H. (1998), Microstructure of Shear Bands and its Relation to the Mechanism of Dilatancy and Failure of Dense Granular Soils, *Géotechnique*, vol. 48, no. 4, pp. 465-481.
- Parry, R.H.G. (1960), Triaxial Compression and Extension Tests on Remoulded Saturated Clay, *Géotechnique*, vol. 10, pp. 166-180.
- Persson, B. N. J. (1998), *Sliding Friction - Physical Principles and Applications*, Springer, New York, 462p.
- Radjai, F., Wolf, D.E., Jean, M. and Moreau, J.J. (1998), Bimodal Character of Stress Transmission in Granular Packings, *Physical review Letters*, vol. 80., no. 1, pp. 61-64.
- Reddy, K.R. and Saxena, S.K. (1993), Effects of Cementation on Stress-Strain and Strength Characteristics of Sands, *Soils and Foundations*, vol. 33, no. 4, pp. 121-134.
- Rinaldi, V.A. (1998), Changes in electrical conductivity after remolding a clayey soil, personal communication. A similar effect was observed in Mexico City soils and is documented in Diaz-Rodriguez and Santamarina, 1999.
- Robertson, P.K. et al. (2000), The CANLEX Project: Summary and Conclusions, *Can. Geotechnical J.*, vol. 37, pp. 563-591.
- Rothenburg, L. (1993), Effects of particle shape and creep in contacts on micromechanical behavior of simulated sintered granular media, *Mechanics of Granular Materials and Powder Systems*, vol. 37, pp. 133-142.
- Rothenburg, L. and Bathurst, R. J. (1989), Analytical study of induced anisotropy in idealized granular material, *Géotechnique*, vol. 49, pp. 601-614.
- Saada, A.S., Liang, L., Figueroa, J.L. and Cope, C.T. (1999), Bifurcation and Shar Band Propagation in Sands, *Géotechnique*, vol. 49, no. 3, pp. 367-385.
- Santamarina, J.C., Klein, K. and Fam, M. (2001-a), *Soils and Waves*, J. Wiley, Chichester, UK, 488 pages.

- Santamarina, J.C, Klein, K. Palomino, A. and Guimaraes, M (2001-b), Micro-Scale Aspects Of Chemical-Mechanical Coupling - Interparticle Forces And Fabric; Edts., C. Di Maio, T. Hueckel, and B. Loret, Maratea, Italy. 26 pages.
- Santamarina, J.C. (1997), 'Cohesive Soil.' A Dangerous Oxymoron, The Electronic J. Geotechnical Eng. - Magazine, August (URL: under Electronic Publications in <http://www.ce.gatech.edu/~carlos>).
- Schanz, T. and Vermeer, P.A. (1996), Angles of friction and dilatancy in sand, *Géotechnique*, vol. 46, no. 1 , pp. 145-151.
- Schmertmann, J.H. (1991), The Mechanical Aging of Soils, *ASCE J. Geotechnical Engineering*, vol. 117, no. 9, pp.1288-1330.
- Schofield, A. (1998), Don't Use the C-Word, *Ground Engineering*, August, pp. 29-32.
- Skipper, N.T. (2001), Influence of Pore-Liquid Composition on Clay Behaviour: Molecular Dynamics Simulations of Nano-Structure, Proc. Workshop on Clay Behaviour: Chemo-Mechanical Coupling, Edts., C. Di Maio, T. Hueckel, and B. Loret, Maratea, Italy. 18 pages.
- Skipper, N.T., Refson, K. and McConnell, J.D.C. (1991), Computer Simulation of Interlayer Water in 2:1 Clays, *J. Chemical Physics* 94, no. 11, pp. 7434-7445.
- Sridharan, A., Altschaeffle, A. G. and Diamond, S. (1971), Pore size distribution studies, *Journal of the Soil Mechanics and Foundations Division, Proceedings of the ASCE*, vol. 97, SM 5, pp. 771-787.
- Stokoe, K.H. and J.C. Santamarina (2000), Seismic-Wave-Based Testing in Geotechnical Engineering, *GeoEng 2000*, Melbourne, Australia, November, pp. 1490-1536
- Tatsuoka, F. and Shibuya, S. (1992), Deformation Characteristics of Soils and Rocks from Field and laboratory Tests, *The University of Tokyo*, vol. 37, no. 1, 136 pages.
- Terzaghi, K (1920), Old Earth Pressure Theories and New Test Results, *Engineering News-Record*, vol. 85, n. 14, pp. 632-637.
- Thornton, C. (2000), Numerical Simulations of Deviatoric Shear Deformation of Granular Media, *Geotechnique*, vol. 50, pp. 43-53.
- Vaid, Y.P. and Sayao, A. (1995), Proportional Loading Behavior in Sand Under Multiaxial Stresses, *Soils and Foundations*, vol. 35, no.3, pp. 23-29.
- Vaid, Y.P. and Sivathayalan, S. (1996), Static and Cyclic Liquefaction Potential of Fraser Delta Sand in Simple Shear and Triaxial Tests, *Canadian Geotechnical J.*, vol. 33, pp. 281-289.
- Vaid, Y.P. and Sivathayalan, S. (2000), Fundamental Factors Affecting Liquefaction Susceptibility of Sands, *Can. Geotechnical J.*, vol. 37, pp. 592-606.
- Valdes, J.R. (2002), Fines Migration and Formation Damage, PhD Dissertation, Georgia Institute of Technology.
- van Olphen, H. (1991), An Introduction to Clay Colloid Chemistry, Krieger Publishing Cy, Florida, pp. 318.
- Vardoulakis, I. (1996), Deformation of Water-Saturated Sand: I. Uniform Undrained Deformation and Shear Banding, *Géotechnique*, vol. 46, no. 3, pp. 441-456.
- Vucetic, M. and Dobry, R. (1991), Effect of soil plasticity on cyclic response, *Journal of Geotechnical Engineering*, ASCE, vol. 117, pp. 89-107.

- Wang, Q. and Lade, P.V. (2001), Shear Banding in True Triaxial Tests and Its Effect on Failure of Sand, in print.
- Yoshimine, M., Robertson, P.K. and Wride, C.E. (1999), Undrained Shear Strength of Clean Sands to Trigger Flow Liquefaction, Canadian Geotechnical Journal, vol. 36, pp. 891-906.
- Zdravkovic, L. and Jardine, R.J. (1997), Some Anisotropic Stiffness Characteristics of a Silt Under General Stress Conditions, Géotechnique, vol. 47, no. 3, pp. 407-437.

ENGINEERING GEOLOGICAL PROBLEMS ASSOCIATED WITH KARST TERRAINS: THEIR INVESTIGATION, MONITORING, AND MITIGATION AND DESIGN OF ENGINEERING STRUCTURES ON KARST TERRAINS

G.S. Xeidakis¹, A.Torok², S.Skias¹ and B. Kleb²

¹*Department of Civil Engineering, Democritus University of Thrace, 67100, Xanthi, Greece.*

²*Department of Engineering Geology and Engineering Materials, Budapest University of Technology and Economics, H-1521, Budapest, Hungary.*

ABSTRACT

The design and construction of civil engineering structures in karst regions confronts many problems due to unpredictable location, dimensions and geometry of the karst structure and voids. Karst terrain is one of the most intricate grounds to be assessed for civil engineering purposes. Conventional methods of site exploration like desk studies, site reconnaissance, borings, test pits, geophysical techniques, have their advantages and disadvantages; none of them are 100% accurate; therefore they should be used in concert, adapted to each project, the available budget and the undertaken risk. As not two sides are identical in karst, site investigation should be tailored to each site. Factors that should be considered when designing site investigation in karst are: maturity of karst landforms, depth of the karst features, overburden thickness, lateral extent of the karst features, hydrogeology of the area, loading, etc.

The main problems confronted by engineers designing structures on or in karst terrain are: difficulties in excavation and grading the ground over pinnacled rockheads; collapse of the roof over subsurface voids, subsidence of cover soil over sinkhole, difficulties in founding a structure over an irregular or pinnacled rockhead, loss of water from dam reservoirs, pollution of groundwater, etc. A number of solutions have been practiced by engineers to solve these problems like: relocating the structure on a safer site, filling the voids and the fractures with concrete, improving the foundation ground with grouting and/or geogrids, replacing foundation soil, bridging the voids with rigid mats or beams, using deep foundations (piling, drilled shafts, etc.), minimizing future sinkhole development by controlling surface and ground water, etc.

1 SITE INVESTIGATION ON KARST TERRAIN

A Karst terrain consists one of the most difficult ground conditions that have to be evaluated for civil engineering purpose. Conventional methods like: thorough data review, to obtain as much information as possible, including topographic and geologic maps, air photos, sinkholes maps, hydrogeology reports, water well records, and previous test boring information, are useful but not enough. These sources should be used to provide an indication of existence of caves, sinkholes and disappearing streams, faulting, rock quality, depth of overburden, and well yields which might foretell the degree of dissolution or fracturing of the rock. Investigation should be followed by site reconnaissance, by experienced personnel, to verify and extend the findings of the preliminary (desk) site evaluation. This stage should also include interviews with persons familiar with the site. The subsurface investigation program, that follows, should maximize the data obtained with a reasonable effort and cost, using various exploration techniques such as test boring, test pits, air track probes, georadar and geophysical surveys (resistivity, microgravity, electromagnetic, seismic refraction, etc).

1.1 Reliability of methods

Conventional practices show that the aforementioned methods may be adequate for investigation in karst sites of classes kI and kII, i.e. for simpler cases. Sites with more mature karst (classes kIII-kV) demand more rigorous ground investigations, managed by a multidisciplinary team that fully appreciates the complex characteristics of karst. No two karst sites are exactly alike in topography and geology, and no method is 100% accurate. Each site may require the use of different combination of investigation methods. A major difficulty in karst is locating subsurface voids.

Table 1 shows the estimated reliability of the various investigation methods from a sinkhole investigation project, carried out in karst region in eastern Pennsylvania and northern New Jersey, USA. In that project 12 site investigation companies took part, applying the 12 most commonly used site investigation methods. The findings were statistically processed and the results of their reliability in detecting existing sinkholes and subsurface voids (caves) are presented in Table 1.

Table 1. Estimated reliability of the various site investigation methods for detecting sinkholes and caves in karst terrain (in Thomas and Roth, 1999)

| Method | Reliability associated with finding existing sinkholes | Reliability associated with finding subsurface voids |
|--|--|--|
| Borings/air track drilling/cone penetrometer | NA (not applicable) | Poor to very good |
| Area reconnaissance | Good to very good | Poor |
| Review of existing mapping | Fair to good | Poor to fair |
| Review of aerial photographs | Fair to very good | Poor |
| Resistivity survey | NA | Poor to fair |
| Seismic refraction survey | NA | Poor |
| Electromagnetic survey | NA | Poor to fair |
| Ground penetrating radar | NA | Poor to fair |
| Trenching | NA | Fair to good |
| Microgravity survey | NA | Poor to fair |
| Video televiewer | NA | Fair to good |
| Borehole to borehole tests/tomography | NA | Fair to very good |

As it is shown in Table 1 none of the methods employed has a good applicability for detecting both sinkholes and subsurface voids. Some are better for the first and some for the second. In the following paragraph the pro and con of each method is discussed shortly.

1.2 Brief description of the various methods

1.2.1 Probing (boring, air track drilling, cone penetration, etc.)

The reliability of these methods to locate an existing subsurface void is related directly to the number of probes made and the size of the void. If each probe is completed to the bedrock surface, the probability of locating a void in the ground was estimated by Beacher, et al., 1980, (in Thomas and Roth 1999) to be:

$$P_r = 1 - (1 - A_v/A_s)^n$$

where P_r = probability of detection of a void, n = number of probes, uniformly distributed over an area A_s , and A_v = horizontal cross sectional area of the void itself.

It has been estimated statistically that a density of 2500 probes per hectare is needed to have 90% chance to locate one void 2.5 m in diameter in the ground. For example, in Belgium for the construction of a viaduct on class kIII karst, 31 boreholes had opened for five pier sites finding no subsurface void; during the excavation for foundation of the piers two caves were located. A second phase of 308 probes for investigation of the broader area found no more caves (Waltham, et al 1988, in Waltham and Fookes, 2003). This example shows the difficulty of locating voids in the ground by probing. It also shows that the true ground conditions in karst are discovered only after foundations are excavated. In general, 3-5 probes beneath every pile foot and column base are probably the best option in karst classes kI-kIII and are essential at pinnacle rockhead in classes kIV-kV karst.

Depth of probing: The depth of probing is a function of void size and its depth. Waltham and Fookes 2003 recommended that in karst of classes kI to kIII, in their karst classification, caves more than 5 m wide are unusual, and probing of 3.5 m in depth should be satisfied. Concerning the depth of probing, inside bedrock beneath pile tips, engineering practice varies considerably from 2 m in North Carolina karst, up to 5 m in cavernous Florida karst and 4 m, under foundations, in South Africa (in Waltham and Fookes 2003); some of the boreholes should be inclined up to 15°.

Tunneling: Probing is also necessary during tunnel excavation in karstified limestone; 3-5 boreholes, 10-12 m deep, should be drilled to the rock, from the face of the tunnel, in advance of the excavation cycle to check for voids or excess groundwater (Marinos 2001, in Beck et al, 2001).

1.2.2 *Review of existing maps and aerial photographs*

Review of existing maps (topographic, geologic, sinkhole inventory, hydrogeology maps, and aerial photographs), provide a fair to good means to detect surface karst features like sinkholes; however their use for locating subsurface voids are quite limited and need much experience from the user.

1.2.3 *Field reconnaissance*

Visual inspection of the site, on the ground or from the air, is a good practice in locating existing sinkholes and previously filled sinkholes. Heavy vegetation may limit the effectiveness of the method. It cannot be used to locate existing subsurface voids which have no surface features.

1.2.4 *Geophysical methods*

Geophysical methods (resistivity, seismic refraction, microgravity, magnetic, GPR, etc.) on karst have not produced consistently reliable results, so far. However technology is advancing and there are geophysical methods that can produce useful results in certain situations. For example, *resistivity surveys* have shown to perform better in detecting subsurface voids than electromagnetic surveys, if abrupt changes in topography and man made subsurface objects are present. However their reliability is fair to poor in areas with highly irregular soil-bedrock surfaces. In general, the reliability of resistivity surveys is fair to good for rockhead profiling in classes kI-kIII, in Waltham and Fookes 2003 classification, but its reliability in pinnacled rockheads of classes kIV-kV is poor.

Resistivity tomography combined with microgravity can be used to identify rockhead and distinguish buried sinkholes from caves, but it is still quite expensive. *Microgravity* can recognize missing mass within the ground and produce good data with increasing sophistication of their analysis. For example, Fourier analysis of microgravity data from a grid with spacing of 2 m can locate caves up to one meter across at specific depths (in Waltham and Fookes 2003). Wider grids cover larger areas and can distinguish low density fills in buried sinkholes, but the analysis is still quite cumbersome.

Seismic methods (refraction, reflection, etc), measure the velocity of compression waves, traveling through the ground. Wave velocity decreases in more fissured and more cavernous ground and thus can be correlated with engineering classification of rock mass. In the future, they may be used to characterize karst classes. So far, seismic methods are most effective in locating boundaries between strata and the interface of soil-rockhead. The methods seldom can locate, even large, cavities in the overburden from the ground surface, because the shock waves through the cover soil travel faster than through the cavities. However, the use of cross-hole seismic methods is possible to locate subsurface voids. In this case, data can be analyzed by computer using tomography to construct two and three dimensional representations of the anomalies. The technique is usually restricted to critical location at the site, due to the high expense of the closely spaced boreholes, the multiple shocks needed, and the volume of data to be processed by the computer (Sower, 1996, p.110).

Ground penetrating radar (GPR) emits high-frequency electromagnetic waves traveling through the ground and producing high resolution profiles of subsurface strata. Due to high attenuation of the electromagnetic energy, as it penetrates the ground, its use is limited to shallow depths (<5m). Specifically, the depth of penetration and resolution quality is decreasing as the percentage of clay increases in the soil. *Static Penetration Test (SPT)* has been used sometimes to detect potential sinkhole failure with limited success.

In conclusion geophysical surveys are good for reducing the investigation cost by identifying drilling sites but all geophysical anomalies require verification by drilling.

1.2.5 Trenching

Test pits or trenches are highly reliable for local studies at shallow depths, but they are impractical for greater depths (>5m) or below water table; they are expensive for large areas, as well.

2 DESIGN OF ENGINEERING STRUCTURES' FOUNDATIONS ON KARST TERRAIN

Once the subsurface exploration has been completed and the earthwork and design parameters of the structure determined, a risk assessment of the site should be made as an integral part of the design of the structure. This should include an overall qualitative assessment of the probability (low, moderate, high) of future subsidence occurrence in the site. The client should also be informed about the risk of his property from future subsidence (Hu, et al 2001, Destephen and Wargo 1992). The design professionals must be prepared to provide foundation alternatives so as to reduce or eliminate risk. These alternatives, in general, are: regular shallow spread footings with or without soil improvement, rigid mats and grade beams, and deep foundations (piles and piers).

2.1 Foundations over karst rockhead

In karst class kI and kII, the rockhead is usually sound, except for some unpredictable isolated fissures and shallow caves. They, usually, create only minor problems to foundations of structures. Regular *spread footings* are, principally, sufficient for foundation of ordinary buildings. Installation of piles may be required in some parts of a site and reinforced (grated) beams can be designed to span small new voids (Fisher and Canace, 1989, Destephen and Wargo, 1992, Beck and Herring 2001, Waltham and Fookes 2003, Wagener 1985, etc.).

In class kIII karst, rigid rafts and/or *grade beams* may bridge cavities in the ground (Sowers 1996, Beck, et al., 2001, 1999, 1997, 1995, 1989, 1986, Green, et al, 1995). *Mats or preparatory grouting* are preferred in Florida, USA, over new sinkholes. *Heavy geogrid* can be used, as well, to reinforce the soil over voids and reduce the impact of any future catastrophic void collapse. *Grouting* may be employed to fill fractures and small voids in rockhead, before founding on spread footings within the soil profile; this may be more economical than piling to rockhead.

Compaction grouting through boreholes has been used, as well, in Pennsylvania, USA, for soil improvement (Reith, et al 1999). *Grade beams* are sometimes applied to bridge any small soil subsidence that could occur with time. These beams are often extended beyond the structure's ends to prevent settlement from subsidence at the building corner (Destephen, et al 1992).

In pinnacled rockheads of karst classes kIV and kV deep foundations are used, ending on sound rock. *Drilled shafts (caissons)* are preferred to piles because piles (both driven and auger cast) may be doglegged (deflected) on pinnacles or deviate on sloped rock and it is difficult to discern their actual bearing capacity (Figure 1). Each pile or shaft tip is probed by 2-3 boreholes to ensure lack of voids beneath its end; some of the boreholes must be inclined (splayed) to 15° from the vertical to examine the ground surrounding the pile tip. As a guide for planning, the mean final length of end bearing piles should be about 30% greater of the mean rockhead depth, into the sound bedrock (see Sowers 1996, Waltham and Fookes 2003, p. 114).

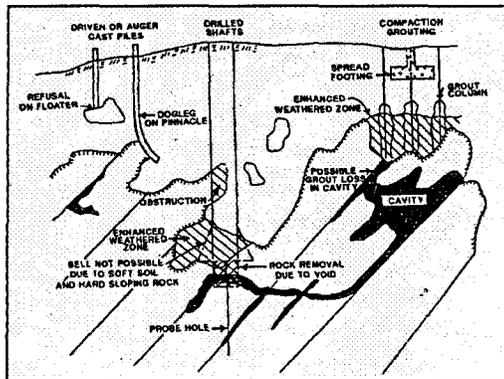


Figure 1. Various types of foundations in karst terrain. (after Destephen and Wargo 1992)

Although *drilled shaft or pile foundations* can be designed with virtually no risk of subsidence problems, they are generally very expensive and prohibited for relatively small buildings. Some of the expense will be for reinforced (grade) beams, if the floor slab is structurally supported to span between sound pinnacle tops. The most cost-effective bored piles or shafts are the *belled drilled shafts* in which the load can be carried by an expanded bearing area formed by a bellling tool. This minimizes the shaft diameter, reducing concrete volume. Unfortunately drilled shafts or piles cannot always be successfully belled in karst areas due to perched groundwater, soft overburden soil prone to caving or sloping bedrock (Destephen and Wargo 1992).

Rockhead pinnacles, some 50 m high, encountered in some tropical karsts, consist a nightmare for founding heavy structures that demand foundations on bedrock. Each pile location requires its own ground investigation and design considerations.

Gypsum has low strength, thus it cannot support high loads of rockhead pinnacles nor heavily loaded end bearing piles.

Roads and light structures pose not much difficulty in founding on soil over a deeply pinnacled rockhead of class KIV-KV.

2.2 Foundations over caves

2.2.1 Generalities.

Subsurface voids in karst are unpredictable. Every site in such terrain has to be assessed individually in the context of its geology and geomorphology. The design of engineering structures in karst with caves must respond to the local conditions. Local maps and records may indicate typical and maximum possible cave size; local people may also know the larger caves in the area. The maximum size of a cave defines the minimum thickness of sound rock cover needed and the possible depth of probing (Figure 1). The more mature the karst, the larger its caves. There are of course, exceptions to this rule. Fortunately, the karst processes are very slow, compared to effective life of human structures (50-100 years). Therefore if there is not immediate collapse of a void during construction or loading, the danger of future collapse is limited. For example in Slovenia karst voids discoveries and collapses are common during road construction, but subsequent collapses under road operation are very rare (Sebela, et al 1999, Jennings, 1966).

Caves 10 m across are typically found in karst class kIV but can occur in less mature karst of class kIII or even in class kII, as well. In more mature karst (class kV) larger caves are quite common. Many large caves at shallow depths, have entrances open to ground surface and can be visited and assessed by direct observation (e.g. Piges cave in Drama, Alistrati cave in Serres, Diross cave in Mani, and many others in Greece and elsewhere).

2.2.2 Methods of foundations design on caves

If caves are critical to planned foundations they are *filled with concrete or bridged with graded beams or slabs*. Considerable loss of ground may occur during grouting a cave at shallow depth due to flowage into neighboring karst voids or water flow inside the cave; this will increase the foundation cost. To avoid this situation, perimeter grout curtains with thicker grout can be used. *Relocation of footings* or of the whole structure, when possible, may be proved to be more economical in some cases.

Drilled piles or shafts should be preformed or cast in geotextile sleeves and founded on solid rock at the cave bottom. The cost of such piles may be greater than filling the cave with grout when it is small enough; this is due to inability of dewatering, additional boring needed, depth to rock where troughs occur and rock excavation performed, owing to unsuitable bearing surfaces such as sloping rock, mud seams, voids and weathered zone.

Grout filling in *gypsum caves* is inappropriate because the greater dissolution rates of gypsum allows the excavation of a new cavity around the concrete plug, within the life time of an overlying engineered structure (Figure 2).

2.3 Foundation over sinkholes

A number of viable foundation support solutions exists that help to confront the problem of placing structures over soil with sinkholes and karst voids. The two most commonly used methods are *excavation and back filling of the sinkhole, and load transfer to sound rock by bridging or piling*. The simplest but not necessarily the most economical solution is to excavate the sinkhole soil to rock and backfill it with a sand and gravel mix and cement grout. If the cavity at the bottom of the

sinkhole, is large, the throat can be blocked (choked) with a graded rock fill which can be sealed with thick grout and/or layers of impermeable soil fill (clay). The area is then returned to grade with a compacted, impervious material and sealed with geomembrane to prevent future water infiltration. The other solution is the *installation of drilled piles or shafts to sound rock strata*, as described before (figure 1, 3).

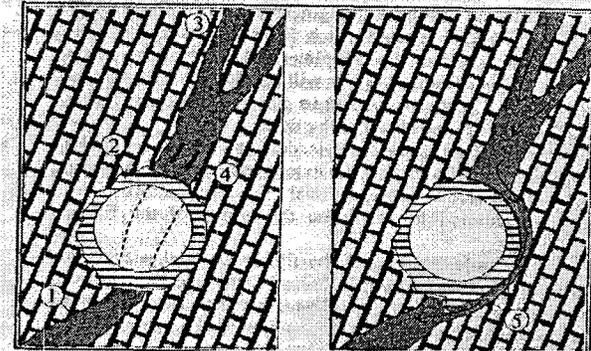


Figure 2. When blocking a big karstic underground water conduit by tunneling or concrete, water may open a new pipe around the block. This is more common practice in gypsum. (after Milanivic, 2000, in Marinos 2001).

In order to prevent the sinkhole collapse, the key factor is the proper *control of water flow at the ground surface and through soil*. Downwards percolation of water should be minimized or eliminated, if possible. Sources of potentially dangerous water flows include water from parking and roadway areas, roof down-spouts, catch basins, flow along utility lines, backfills, cultivated areas and irrigated gardens, runoff from impervious surfaces, etc. (Fisher and Canace 1989, Destephen and Wargo 1992, Sowers 1996, Waltham and Fookes 2003, Kannan 1999, Beck et al 1999, ch. 4). *Control of water abstraction and lowering the water tables* in the area is also critical, especially where the water table is close above rockhead; the water table should always be kept above rockhead.

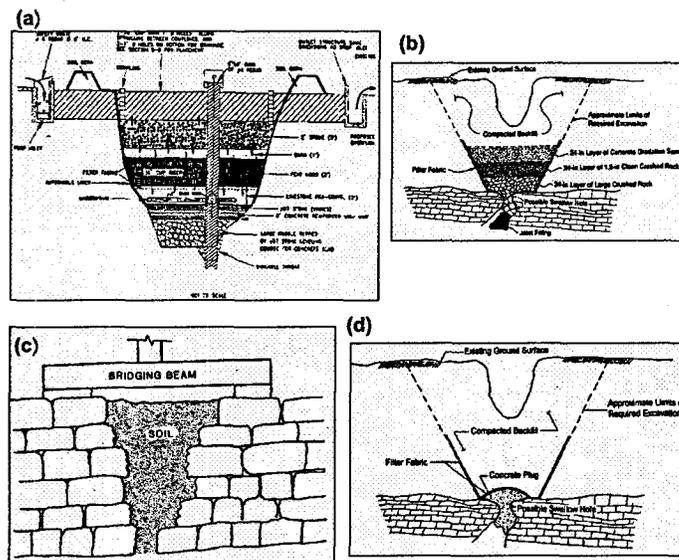


Figure 3. Foundations over sinkholes (a) Excavation and filling with graded materials plus graded beam and drainage, in a highway pavement in Tennessee, USA (Stephenson et al 1997), (b) Excavation and filling with graded drain material (in Beck et al 1997, p180) (d) Structural plug of the sinkhole throat (Ripp and Baker 1997, in Beck et al 1997, p296) and (c) Bridging with a graded beam (Sowers 1996)

A good example is Florida's Disney World which stands on 20-30 m of soil over soft, karstified limestone of class kIII, with many sinkholes. They managed to avoid sinkhole subsidence by closely monitoring pumping from water wells and keeping the water table more or less at constant elevation. The pumping is switched off where local water table decline is detected (in Waltham and Fookes 2003). Dewatering by quarrying or tunneling should be banned in karst areas (Marinos 2001, in Beck et al, 2001).

Other measures for remediation and prevention of sinkhole failures are *grout sealing at rockhead fissures*. This may be applied to karst class kII and kIII, but it is problematic in class kIV and kV with large pinnacles. In the last case "cap grouting" may be required with cement slurries, after plugging open fissures below.

Compaction grouting, using a low slump grout (slump <25mm) under pressure, to displace and consolidate soft zones, can be used above groundwater table. The grout forms within the soil a solid block that bridges over fissures and pinnacles. As grout flows in uncontrollable way its placement may not be so effective. *Grout columns* are sometimes used to improve soil conditions at soil-rock contact (Figure 1,3). *Pressure grouting* is not recommended "to fill all voids" since the flow and the losses of the grout are uncontrollable.

Deep dynamic compaction for densifying soil, by dropping big weights from a height, is controversial. The method may reduce the permeability of the soil and collapse soil-rock cavity roofs, but if it fails to do so, it may increase the potential of future void roof collapse (Destephen et al 1992).

In karst of class kII-kIV, sinkhole hazard can be reduced by *laying geogrid* into the soil combined with proper drainage control (Villard, et al 2000).

3 TUNNELING IN KARST

Tunneling in karstic environment is confronted with two main problems: *inundation by ground water and collapse of karstic voids* with inrush of mud and water into the excavation. The first problem is faced by lowering groundwater either through controlled drainage inside the tunnel or dewatering the ground by pumping at the ground surface, and *grouting and reducing the rock mass permeability*. Dewatering may have adverse effects, like: development of sinkholes, settlement at the ground surface due to soil consolidation, depletion of groundwater reservoirs which in turn may affect agriculture in the broader area, sea water intrusion at coastal zone, contamination of ground water, etc. Grouting, on the other hand is difficult in large voids and under high water head (Fig.1, 2).

In mines, dewatering is usually managed by massive pumping through deep wells to maintain huge cone of depression around each well, for, as long as the mine is under operation.

Confronting a karstic void during tunneling is really a challenge: it can be overcome by bridging or filling the void, if empty, by stabilizing with grouting its soft filling material, and tunneling through it afterwards, or by controlling mud and groundwater inrush by fore-boring, etc., before the tunnel face approaches the void (Marinos 2001, in Beck, et al, 2001).

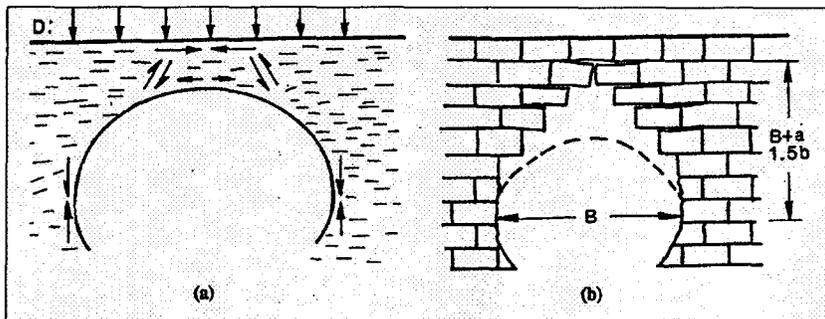


Figure 4. Stress and displacements in the rock surrounding a cavity, (a) Homogenous rock with thin cover-rock, and (b) Stratified hard rock (Sowers 1996)

4 PLANNING TO REDUCE RISK IN KARST SITES DEVELOPMENT

The most economical approach in developing a site, underlain by karstified carbonate rock is to define the problem sites by site investigation and to plan location, size, and type of significant constructions out of the problem areas of the site. That means, if the aerial extent of the problem sites are known, it may be possible to place the most important structures in the safest part(s) of the construction site and leave the problem areas for non critical facilities such as golf courses, hiking trails, parking lots, parks, roadways, grass fields, etc.

Other means to minimize risk of structures over karst are (Fisher and Canace, 1989, Destephen and Warge 1992, Beck et al, 2001, 1999, 1995, etc.):

- To avoid (unlined) detention basins/ponds in the area, unless one wishes to use such basins for ground water recharge. This however is risky, since it will increase water head and flow rate and may be pollution, in the soil and rock.
- To design measures to maintain ground water level consistent with that prior to development
- .To avoid ground grading, when possible, but if done it should reflect surface drainage away from the structures.
- To avoid placing utilities adjacent or beneath shallow foundations. If this is necessary place them in concrete duct bank.
- To provide water-tight storm drains and tie roof-drains directly into them.
- To seal pavement curbs and catch basins. Do not allow concentrated flows in unpaved or unlined ditches or swales (ponds).
- To use always lined retention basins and keep them away from the construction, if possible.
- To provide professional observation and inspection of site investigation, earthwork and foundation construction, so as to examine subsurface material and recommend changes in the initial design, aiming to reduce risk.

5 CONCLUSIONS

From the previous discussion it comes out that:

- Karst terrain is very complex to be grouped in classes and a difficult ground for construction of structures either on the ground surface or underground. It is a real challenge that can turn to a nightmare for engineers if the proper measures are not taken.
- Conventional methods of site investigation can be applied only to young karst with simple structure; for more mature karst none of these methods is 100% accurate. It demands more rigorous ground investigation by a multidisciplinary team that fully appreciates the complex characteristics of karst. Especially the location of underground voids is very difficult. No two karst sites are exactly alike in topography in geology and in structure.
- Risk assessment of the site should precede any design and the probability of future subsidence should be estimated.
- The design of engineering works on karst terrain, have to be adapted to the maturity of the karst.
- Various foundation alternatives such as spread footings on cover soil, with or without soil improvement, rigid mats and graded beams; deep foundations (piles or drilled piers); control of surface and ground waters and sinkhole development, etc., are practiced in karst. The kind of foundation applied is unique for each karst site and depends upon the maturity and the structure of the karst and the predicted foundation loading.
- The most dangerous sites for subsidence are those engaged to foundations over sinkholes. The controlling factors are: water percolating through the cover soil and overloading.
- The principal measures to be taken for avoiding subsidence are:
 - o To control of surface and ground water in the karst region.
 - o To avoid the ground overloading, and
 - o To place significant constructions out of the problem areas of the site.

- In tunneling, problem areas can be located by fore-boring from the face of the tunnel.
 In conclusions karst is a difficult ground for engineers and needs proper understanding for good engineering practice.

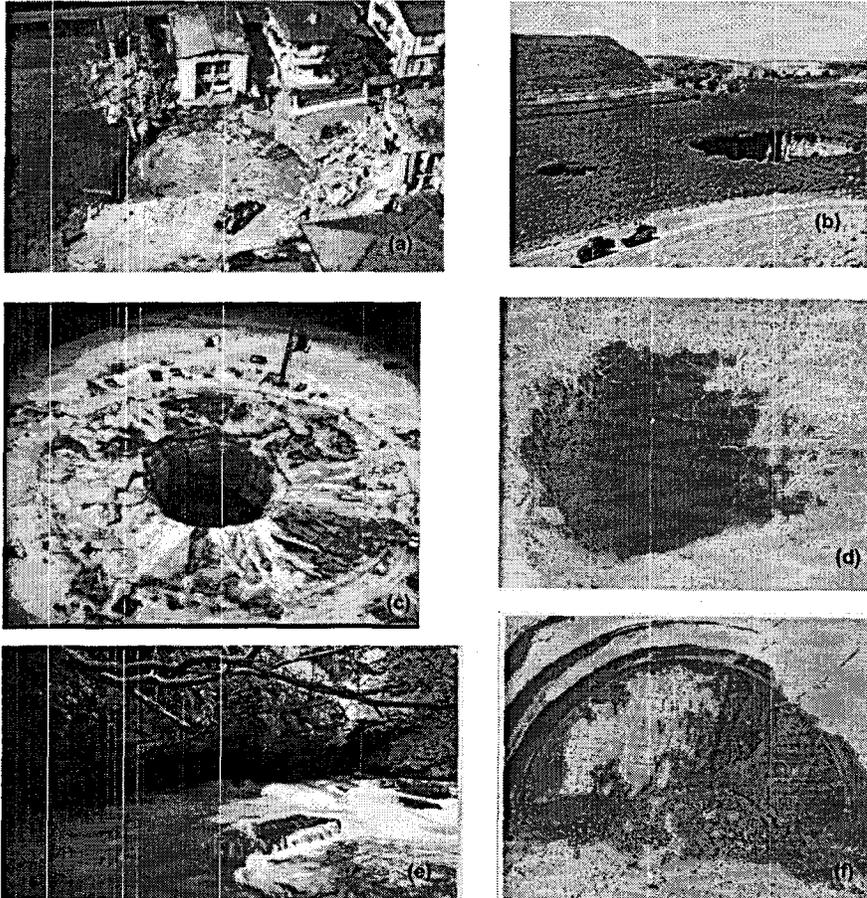


Figure 5. Examples of sinkholes collapse. (a) Sinkhole under repair in Camaione, Tuscany, Italy, 1995, diameter 30 m, depth 13 m; thickness of cover soil more than 100 m. (in Beck et al, 2001, cover photo), (b) Sinkholes on the bottom of May Dam's lake, they drained the lake Turkey (Ertunc, 2004), (c) Sinkhole collapse in Florida, USA, more than 130 m deep, (Fuleiham et al, 1997) (in Beck and Stephenson 1997, cover photo), (d) and (f) Sinkhole collapse over Dodoni tunnel, Epeiros, Greece, (Marinos 2001), (e) Cave and spring of Piges Aggiti river in marble, Drama, Greece. Behind the entrance there is a collapsed sinkhole.

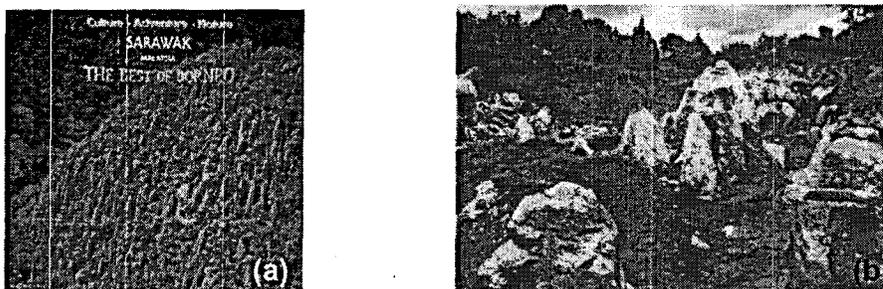


Figure 6 (a) and (b) Pinnaclated karst terrain in tropics

REFERENCES

- Adams, A.L. 2001. A geoelectrical investigation of the freshwater aquifer near the well fields of north Andros Island, Bahamas. Proc. Conf. on "Geotechnical and Environmental Applications of Karst Geology and Hydrology. Beck, B.F. & Herring, J.G. 2001 Balkema Publishers, pp. 341-346.
- Beck, B.F. & Herring, J.G. 2001. Geotechnical and Environmental Applications of Karst Geology and Hydrology. Balkema Publishers, 438pp
- Beck, B.F., Pettit, A.J. & Herring J.G. 1999 (eds). Hydrogeology and Engineering Geology of Sinkholes and Karst. *Conference proceedings*, Chapter 11, Special session on highway karst. Design, Construction and Repair. *Balkema*, Rotterdam, 478pp.
- Beck, B.F. & Stephenson J.B. 1997 (eds). The Engineering Geology and Hydrogeology of Karst Terraines. *Conference proceedings*. Balkema, Rotterdam. 516pp.
- Beck, B.F. & Pearson, F.M. 1995 (eds). Karst geohazards. Engineering and Environmental problems in Karst Terrane. *Conference proceedings*. Balkema, Rotterdam. 582pp
- Beck, B.F. 1989 (editor). Engineering and Environmental impacts of Sinkholes and Karst. *Conference proceedings*. Balkema, Rotterdam, 384pp.
- Beck, B.F. & Sinclair, W.C. 1986. Sinkholes in Florida: an introduction. Institute Report 85-86-4. Florida Sinkhole Research, USA.
- Bennett, D. 1997. Finding a foothold. *New Civil Engineer* (4), 24-25.
- Crawford, N. C. 2001. Field Trip Guide, Part I. Environmental problems associated with urban development upon karst, Bowling Green, Kentucky. Proc. Conf. on "Geotechnical and Environmental Applications of Karst Geology and Hydrology. Beck, B.F. & Herring, J.G. 2001 Balkema Publishers, pp. 397-424.
- Destephen, R.A. & Wargo, R.H. 1992. Foundation Design in Karst Terrain. *Bull. Assoc. Engng Geologists*, vol. XXIX (2), pp.165-173.
- Ertunc, A. 2004. Karstic limestone problems of some dams in Turkey. *Proc. 5th Intern. Symposium on Eastern Mediterranean Geology*. Thessaloniki, Greece, 14-20 April 2004.
- Fischer, J.A. & Canace, R. 1989. Foundation Engineering Construction in Karst Terrane. In "Foundation Engineering. Current Principles and Practices". *Proc. Conf., Evoston*, Illinois, ASCE, USA, 1989, pp. 29-42.
- Fookes, P.G. 1997. Geology for engineers: the geological model, prediction and performance. *Quarterly Journal of Engineering Geology*, 30, 293-424.
- Fookes, P.G. & Hawkins, A.B. 1988. Limestone weathering: its engineering significance and a proposed classification scheme. *Quarterly Journal of Engineering Geology*, 21, 7-31.
- Ford, D.C. & Williams, P.F. 1989. Karst Geomorphology and Hydrology. Unwin Hyman, London.
- Green, M.R., Forth, R.A. & Beaumont, D. 1995. Land subsidence on Magnesian Limestone terrain in County Durham, England. *International Association of Hydrological Sciences*, 234, 423-431.
- Hu, R.L., Yeung, M.R., Lee, C.F., Wang, S.S. & Xiang, J.X. 2001. Regional risk assessment of karst collapse in Tangshan, China. *Environmental Geology* 2001, 40, pp. 1377-1389.
- Jennings, J.E. 1966. Building on dolomites in the Transvaal. *The Civil Engineer in South Africa*, 8, 41-62
- Kannan, R. C. 1999. Designing foundations around sinkholes. *Engineering Geology*, 52, 75-82
- Lowe, D. & Waltham, T. 2002. Dictionary of karst and caves. *British Cave Research Association Cave Studies*, 10, 1-40.
- Marinos P.G.. 2001. Tunnelling and mining in karstic terrane. An engineering challenge. Keynote Address. Proc. Conf. on "Geotechnical and Environmental Applications of Karst Geology and Hydrology. Beck, B.F. & Herring, J.G. 2001 Balkema Publishers. Chapter 1., pp. 3-16.
- Merrit, A.M. 1995. Geotechnical aspects of the design and construction of dams and pressure tunnels in soluble rocks. In Beck, B.F. & Pearson, F.M. 1995 (eds). Karst geohazards. Engineering and Environmental problems in Karst Terrane. *Conference proceedings*. Balkema, Rotterdam.
- Reith, C.M., Cadden, A.W. & Naples, C.J. 1999. Engineers challenged by Mother Nature's twist of geology, In: Beck, B.F., Pettit, A.J. & Herring, J.G. (eds) *Hydrology and Engineering Geology of Sinkholes and Karst*. Balkema, Rotterdam, 149-155.
- Sebela, S., Mihevc, A. & Slabe, T. 1999. The vulnerability map of karst along highways in Slovenia. In: BECK, B.F., PETTIT, A.J. & HERRING, J.G. (eds) *Hydrology and Engineering Geology of Sinkholes and Karst*. Balkema, Rotterdam, 419-422.
- Sowers, G.F. 1996. Building on sinkholes. ASCE Press, USA, 202, pp.
- Tharp, T.M. 1999. Mechanics of upward propagation of cover-collapse sinkholes. *Engineering Geology*, 52, 23-33.
- Thomas, B. & Roth, M.J.S. 1999. Evaluation of site characterization methods for sinkholes in Pennsylvania and New Jersey. *Engineering Geology*, 52, 1999, pp. 147-152.
- Villard, P., Gourc, J.P. & Giraud, H. 2000. A geosynthetic reinforcement solution to prevent the formation of localized sinkholes. *Canadian Geotechnical Journal*, 37, 987-999.
- Wagener, F.v.M. 1985. Problems of soils in South Africa: dolomites. *The Civil Engineer in South Africa*, 27, 395-407.
- Waltham, A.C. & Fookes, P.G. 2003. Engineering classification of karst ground conditions. *Q.J Engng Geology and Hydrogeology*, 36, pp. 101-118.
- White, W.B. 2002. Karst Hydrology: recent development and open questions. *Engineering Geology*, 65, 2002, pp. 85-105.
- Yilmaz, I. 2001. Gypsum/anhydrite: some engineering problems. *Bull. Eng. Geol. Env.* 2001, 59, pp. 227-230.