ARTICLE

Early Mars: How Warm and How Wet?

Steven W. Squyres and James F. Kasting

Early in its history, Mars underwent fluvial erosion that has been interpreted as evidence for a warmer, wetter climate. However, no atmosphere composed of only CO_2 and H_2O appears capable of producing mean planetary temperatures even close to 0°C. Rather than by precipitation, aquifer recharge and ground water seepage may have been enabled by hydrothermal convection driven by geothermal heat and heat associated with impacts. Some climatic warming was probably necessary to allow water to flow for long distances across the surface. Modest warming could be provided by even a low-pressure CO_2 atmosphere if it was supplemented with small amounts of CH_4, NH_3, or SO_2. Episodic excursions to high obliquities may also have raised temperatures over some portions of the planet's surface.

Of all the planets, Mars appeals most strongly to the imagination because it is the one most like Earth. Mars has polar caps and seasons, shows evidence of having had liquid water on its surface, and might once have supported life. But present conditions at the surface of Mars are not Earth-like. Today, Mars is a cold desert with a surface atmospheric pressure of 7 mbar and a global mean temperature of 218 K. Liquid water cannot persist at the surface under these conditions, nor can life. Even if liquid water did exist, the highly oxidizing conditions found by Viking would be fatal to terrestrial organisms (*1*). The real fascination of Mars comes not from its present characteristics but from the most exciting discovery of the Mariner 9 mission: Conditions at the surface of Mars were once substantially warmer and wetter than they are today.

There is little doubt that conditions on Mars during the first half billion years of its history were dramatically different from those at present. The rate of meteoritic bombardment was certainly much higher. On the moon, radiometric dating of heavily and lightly cratered terrains has established that the average impactor flux during the first half billion years of lunar history was two orders of magnitude or more higher than the flux during the last 3 billion years (*2*). This abrupt drop-off in impactor flux coincided with the final gravitational sweep-up of the bodies from which the planets were formed. Dynamical considerations suggest that a qualitatively similar drop-off would have been experienced by all of the terrestrial planets (*3*).

Rates of erosion and deposition on early Mars also appear to have been higher than those during most of martian history (*4*). The best evidence for this comes from crater morphology in the ancient highlands of Mars (*5*). Flat-floored, rimless craters are common there, indicating that substantial topographic degradation has occurred. However, many fresh craters are superimposed on this degraded population, and in some younger regions, virtually all craters are fresh. The areal density of the fresh craters in both instances is so high that degradation cannot have taken place at a constant rate but was concentrated early in the planet's history. Crater densities indicate that degradation was most efficient during the earliest recorded geologic epoch on Mars, known as the Noachian. Fluvial activity has been advocated as a major agent of early crater degradation (*5*), but eolian mantling and erosion, volcanic infilling, and downslope movement of debris (mass wasting) could have played significant roles as well.

The most dramatic evidence that conditions early in martian history were different from those today comes from valleys common in the ancient heavily cratered terrain of Mars (Fig. 1). Most of these are small, with lengths of up to hundreds of kilometers and widths of one to a few kilometers. Channels (the actual conduits once occupied by fluid) have not been resolved in images of any valleys and may have been substantially narrower. Most valleys are U-shaped in cross section, with depths of 100 to 250 m (*6*). Dendritic patterns are typical, but tributary development and drainage density vary widely.

Most small valleys (more than 95%) are found in ancient terrains (*7*), suggesting that the valleys themselves are old. Where valley surface areas are sufficient to allow ages to be determined directly from the density of superimposed craters, this ancient age is confirmed (*8*). As with enhanced erosion-deposition rates, the epoch during which most small valleys formed appears to have been roughly the first half billion years of martian history. Because the valleys are so small and their discharges correspondingly modest, it has commonly been inferred that conditions warmer than those on Mars today were required for their formation (*9*).

If conditions early in martian history really were warmer, climate may be only part of the story. If the valleys formed by precipitation and runoff, then climate is indeed the controlling factor. If they formed by ground water outflow ("sapping"), however, then geothermal heat may also have been important. We argue below, as have others, that sapping best accounts for the morphology of most valley systems. Therefore, we consider not just climate but the broader issues of volatile abundance and temperature in the near-surface regions of Mars.

Environmental Conditions on Early Mars

Conditions on early Mars depend largely on the availability of volatiles, particularly CO_2 and H_2O. Estimates for the initial volatile abundance of Mars vary widely (*10*). Anders and Owen (*11*) predicted the planetary equivalents of 9.4 m of water and 0.14 to 0.53 bar of CO_2 on the basis of scaling from the measured abundance of ^{36}Ar. Like other nonradiogenic noble gases, ^{36}Ar is highly depleted on Mars compared to Earth. This estimate assumes that H_2O, CO_2, and Ar were incorporated into the planet in the same manner and that loss processes affect each equally. As discussed below, the latter assumption, at least, is probably incorrect. Simple mass scaling to Earth (that is, assuming that volatile abundance ratios are the same on the two planets) yields much higher values: ~1200 m of water and 10 bars of CO_2. Straightforward notions about planetary accretion suggest that the planetesimals available at Mars's orbit should have been, if anything, more volatile-rich than those that formed Earth because the solar nebula was cooler at that distance (*12*). Hence, mass scaling to Earth should provide a lower bound to the amount of volatiles delivered to Mars early in its history. This implies that we must seek an alternative explanation for the low abundance of noble gases in the martian atmosphere.

The explanation that seems most satis-

S. W. Squyres is with the Center for Radiophysics and Space Research, Cornell University, Ithaca, NY 14853, USA. J. F. Kasting is with the Department of Geosciences, Pennsylvania State University, State College, PA 16802, USA.

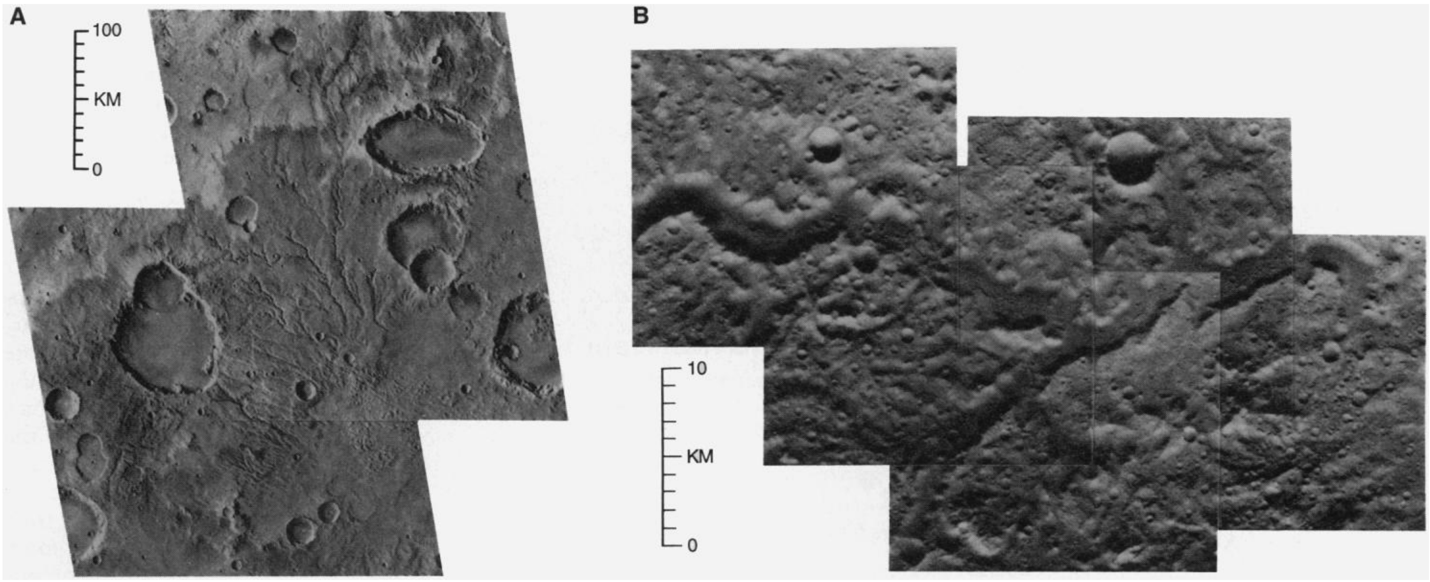

Fig. 1. Regional (**A**) and high-resolution (**B**) views of typical ancient valley systems on Mars. (A) Viking orbiter images 651A91 and 651A94 near latitude −20°, longitude 10°; (B) Viking orbiter images 130S17–21 near latitude 0°, longitude 258°.

factory is that Mars was supplied with abundant volatiles but that those confined primarily to the atmosphere (noble gases and molecular nitrogen) were depleted by impact erosion (the loss of atmospheric gases to space during large impacts) or hydrodynamic escape, or both (*13*). Because of its low mass and low surface gravity, Mars is much more susceptible to impact erosion than is Venus or Earth. Melosh and Vickery (*14*) showed that plausible early impact fluxes could have stripped Mars of most of its original atmosphere (*15*). Hydrodynamic escape of hydrogen may also have stripped off heavier gases by dragging them off into space. Evidence that this process occurred is provided by the relative abundances of noble gases, which are mass-fractionated compared to the terrestrial pattern (*13, 16*). Hydrodynamic escape could have occurred from an accretionary steam atmosphere (*17*) or from a subsequent hydrogen-rich paleoatmosphere.

Both impact erosion and hydrodynamic escape would have affected only atmospheric volatiles. H_2O, which would have existed primarily as liquid water or ice, need not have been lost as efficiently. Substantial CO_2 may likewise have been retained if carbon was partitioned into dissolved carbonate species and carbonate rocks. Taken together, available information supports the view that equivalents of up to several bars of CO_2 and several hundred meters of water could have been outgassed during the earliest stages of martian history and been present when the valley systems were forming.

The geothermal heat flow of Mars must also have been much higher early in the planet's history. Dynamical models show that the time scale for the accretion of Mars and the other terrestrial planets was on the order of 10^8 years (*3*). The impactors that struck Mars during accretion deposited tens of percent of their kinetic energy below the surface as heat, and large impactors buried their heat deeply. Because accretion is rapid relative to conductive heat loss, substantial melting of the interior of Mars should have occurred during accretion (*18*).

Strong evidence that the early martian interior was hot comes from the shergottite-nakhlite-chassignite (SNC) meteorites, for which a martian origin is generally accepted (*19*). The U-Pb data from SNC meteorites intercept the concordia curve at both a young age, representing crystallization or ejection, and at an age of 4.5 billion years ago (Ga) (*20*). This observation, combined with whole-rock Rb-Sr ages of about 4.5 to 4.6 Ga for SNCs, indicate that Mars differentiated very early (*21*). It is likely, then, that much of the martian mantle was at or near the solidus at the end of accretion. Because radiogenic heating is modest compared to the heat of accretion and of core formation, the planet would subsequently have cooled monotonically.

The geothermal gradient on Mars depends on both heat flow and the thermal conductivity of the regolith. Regolith conductivity is poorly known, but the conductivity of terrestrial permafrost may provide a guide. Typical permafrost conductivities are 0.2 to 2 W m^{-1} K^{-1}, with most in the range of 0.5 to 1.0 W m^{-1} K^{-1} (*22*). Taking this narrower range and the thermal history model of Schubert *et al.* (*23*), which has a hot start of the sort described above, one can estimate the evolution of the martian geothermal gradient (Fig. 2). The predicted gradient is 0.03 to 0.06 K m^{-1} at present but 0.18 to 0.36 K m^{-1} at 4.2 Ga. These latter values are large; for the present climate, they would allow liquid to exist just 150 to 300 m below the surface, a depth comparable to that of most valley systems. If ground water were saline, as is likely, the melting isotherm would be still shallower. Most importantly, local heat flow transients produced by concentrated magmatic activity and large impacts would have been substantially higher, allowing liquid water to exist very close to the surface and producing water temperatures well above 0°C.

Climatic Considerations

Until recently, it was believed that early Mars could have been warmed substantially by the greenhouse effect of a dense CO_2 atmosphere (*24*). Such an atmosphere on early Earth could explain why our planet

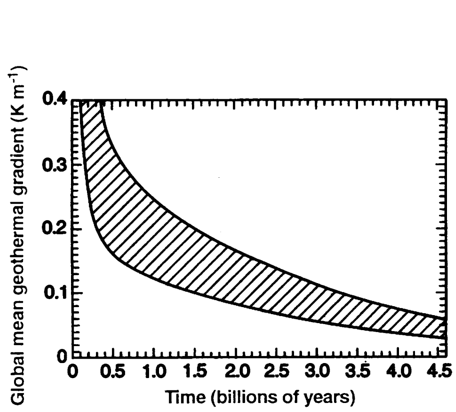

Fig. 2. A model of the evolution of the global mean geothermal gradient versus time on Mars that uses the mantle evolution model of (*23*) and a regolith thermal conductivity of 0.5 to 1.0 W m^{-1} K^{-1}.

remained habitable despite a 25 to 30% lower solar luminosity at that time (25). Unfortunately, this idea fails for Mars. Mars receives only 43% as much sunlight as does Earth today and would have received even less in the past. Its atmosphere should therefore have been filled with clouds of CO_2 ice (26). These clouds would have cooled the planet's surface by increasing the planetary albedo and reducing the tropospheric lapse rate. The first phenomenon is hard to simulate, but the second can be included in a one-dimensional, radiative-convective climate model (RCM) by requiring that the lapse rate follow a moist adiabat.

Figure 3 (26) shows a series of RCM calculations at different solar luminosities, S, performed with a model that includes the lapse rate effect. As S decreases below its present value, S_0, CO_2 condensation becomes increasingly important. For $S/S_0 = 0.75$ (the expected value at 3.8 Ga), the surface temperature T_s could not have exceeded 230 K. Furthermore, CO_2 partial pressures exceeding ~3 bars are ruled out because any additional CO_2 would have condensed on the planet's surface. A corresponding inverse calculation indicates that surface temperatures exceeding 0°C are possible only for $S/S_0 > 0.86$ (26), which should not have been reached until ~1.9 Ga (27). This is too late in the history of Mars to explain the age of most valley networks. Therefore, the conventional "warm early Mars" hypothesis has an apparently fatal flaw.

There are several ways out of this dilemma. One that we explore further below is geothermal heat. A second is greenhouse gases (or particles) in the martian paleoatmosphere beyond just CO_2 and H_2O. Potential candidates include CH_4, NH_3, and SO_2 (28, 29). The most promising is CH_4 because it is the least soluble and has the longest photochemical lifetime, particularly if shielded from solar ultraviolet radiation by hydrocarbon particles produced from its photolysis (30). Potential methane sources include volcanism and, if life evolved, methanogenic bacteria (31). A third possibility is that solar evolution models are incorrect and the young sun was brighter than assumed. A more massive and hence brighter young sun has been suggested as a means of explaining the observed Li depletion in the solar atmosphere (32). This hypothesis would imply, however, that young, solar-type stars should be losing mass at $\sim 10^4$ times the current mass flux in the solar wind, and there is no observational evidence for this.

Another way to promote valley formation in the presence of a weak CO_2 greenhouse would be to invoke spatial inhomogeneities in climate. For moderate obliquities (like the current value of 25°), the equatorial regions might be expected to be substantially warmer than the poles. However, solar flux at the martian equator at 3.8 Ga was still only about 50% of the current terrestrial average. At this low solar insolation, a large greenhouse effect, equivalent to 1 bar or more of CO_2, would be needed to produce equatorial temperatures above 0°C. But a CO_2 atmosphere of this density has such a long radiative time constant that the equator-to-pole temperature gradient should be very small (33), contradicting the initial assumption of spatial inhomogeneity. Therefore, it is difficult or impossible to produce a warm equatorial climate with CO_2 and H_2O as greenhouse gases.

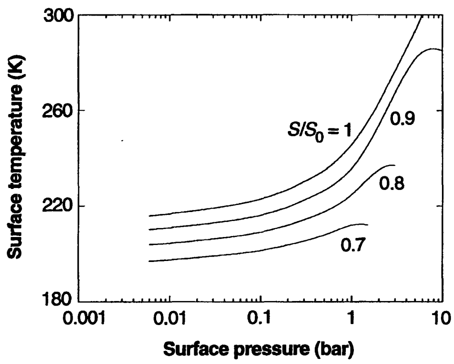

Fig. 3. Surface temperature versus surface pressure for a pure CO_2 atmosphere on early Mars. The effect of CO_2 condensation on the tropospheric lapse rate is included in the climate model. The curves are for various solar luminosities S, scaled to the present value S_0. [Figure from (26)]

The situation changes dramatically at high obliquity. At obliquities exceeding 54°, the average insolation at the poles exceeds that at the equator (34). According to recent numerical simulations (35), the obliquity of Mars varies chaotically and may reach values as high as 60° every few tens of millions of years. For a few months each year, the summer pole could receive up to 3.5 times the average planetary insolation; even at 3.8 Ga, it would have been heated more strongly than Earth is today. At times of high orbital eccentricity, the perihelion flux could have been more than 50% higher than the current mean terrestrial value. With this much insolation, little or no greenhouse effect would have been needed to produce local surface temperatures above 0°C (36) at high latitudes.

High obliquities cannot by themselves account for valley formation. The valley systems are widespread, whereas the warmest regions should have been localized near the poles. Substantial polar wander (37) would be required to distribute the valleys over the surface. Moreover, this mechanism alone cannot explain why the valley systems are mostly ancient. High planetary obliquities could have been achieved as recently as 10 or 20 million years ago. Some additional secular change, such as a monotonic decrease in atmospheric temperature or geothermal heat flow, must still be invoked to explain why most valleys formed in the distant past.

Geologic Considerations

Although valley systems are often cited as evidence for warmer conditions on Mars, their actual environmental implications are difficult to infer. This is in part because dissimilar geologic processes can produce similar landforms and in part because of the great age of the valleys and the significant modification they have suffered. The key environmental question is whether their formation was dominated by surface runoff, which requires precipitation, or by subsurface sapping, which may not.

The arguments for sapping being responsible for the formation of most ancient martian valley systems are strong (8, 38). Drainage densities for such valleys are typically far lower than those for terrestrial drainage systems. In contrast to the space-filling drainage patterns formed by precipitation, martian drainage patterns typically have broad, undissected interfluves. These occur most readily when valley formation is initiated by localized sapping rather than by widespread rainfall or snowfall. First-order tributaries are usually short and stubby, with steep-walled alcoves (or "theater" terminations) at their upstream terminations. Tributary valleys are comparable in depth and width to main trunk valleys, and tributaries join higher order valleys at angles that are greater than those typically found in terrestrial drainage systems. Structural control of martian valleys is strong. All of these characteristics are typical of terrestrial sapping valleys but atypical of drainage systems developed by runoff (8, 38). Some martian valleys deviate from these characteristics, most notably on the flanks of several volcanoes where local runoff, perhaps associated with volcanic outgassing, may have occurred (39). There are also some valley systems in the highlands that have morphologies that are sufficiently unclear or degraded that precipitation cannot be ruled out. However, we conclude that sapping accounts most readily for the morphology of the vast majority of ancient martian valleys.

Many apparent sapping valleys are sufficiently developed that they could not have been formed by a one-time discharge of the aquifer that fed them. Measurements of valley topography (6) provide an estimate of the amount of material moved during valley system formation. Drainage system boundaries may be mapped on the basis of both drainage patterns and regional topography. Such maps, coupled with models of regolith porosity (40), provide estimates of the maximum amount of water available from aquifer discharge. Even with very liberal assumptions about the ease with which the regolith is eroded, the amount of water that could be contained in most local aquifers is inadequate to carve the valleys with which they are associated (41). So, some form of hydrologic cycle is required in order to replenish the water in these aquifers and to allow erosion to occur.

There are at least two possible mechanisms for replenishing the water in martian aquifers. One, of course, is precipitation. Another is hydrothermal convection. Hydrothermal convection occurs when an aquifer is heated locally, setting up density contrasts in the water and initiating buoyancy-driven flow. On Earth today, the primary driver of hydrothermal convection is magmatic activity; on early Mars, both magmatic activity (42) and impact heating (43) would have been drivers. Gulick *et al.* (42) found that magmatic intrusions 100 km^3 in size could have circulated fluids continuously for $\sim 10^5$ years. Aquifers drawn down by seepage could therefore have been replenished for long periods as hydrothermal convection circulated water into them. This finding alone does not prove that valley formation would take place as a consequence of hydrothermal convection because aquifer recharge is only part of the problem; seepage and erosion must also operate in order for valleys to form. However, it does suggest that hydrothermal convection is an adequate alternative to precipitation for aquifer recharge if magmatism and cratering were sufficiently vigorous during the Noachian.

A Picture of Early Mars

The case against widespread precipitation early in recorded martian history appears strong. The morphology of most valley systems does not require it, and some aspects of this morphology (for example, low drainage density) argue against it. Moreover, greenhouse warming sufficient to allow precipitation seems difficult to achieve. We therefore consider the climatic and geothermal conditions that could allow valley systems to form without precipitation. This is a three-part problem. The first requirement is a mechanism for aquifer recharge, and we argue, because of the prevalence of volcanism and cratering then, that hydrothermal convection is likely to have been sufficiently widespread in the Noachian for it to have done the job. The second requirement is that water must be able to emerge from a seepage face and cause headward erosion without the face freezing and ceasing the flow. The third is that the released liquid must be able to flow for up to hundreds of kilometers across the surface.

Recent calculations (44) address the conditions required for seepage. Results are calculated by solving coupled equations of heat diffusion and fluid flow in a hydrothermally charged aquifer behind an eroding seepage face (45). The face moves headward as erosion occurs, with a rate determined by fluid discharge and the grain size and cohesion of the regolith. If the atmosphere is warm, erosion can proceed indefinitely. If it is cold, whether or not erosion can proceed is determined by the relative rates of erosion and propagation of the surface thermal wave. If erosion is rapid, material can be removed from the face faster than it is cooled and the erosion can persist. If it is slow, however, the surface thermal wave propagates inward, allowing pore ice to form. This reduces the hydraulic conductivity, discharge, and erosion further and chokes off the flow.

The model begins with no ice present and liquid flowing freely. One calculation that has been done is for the low-temperature greenhouse model of Kasting (26), 1.5 bars of CO_2, and a local geothermal heat flux of 0.65 W m^{-2}, which for the assumed regolith thermal conductivity and sapping depth gives a water temperature of 0°C. Flow stops in a matter of hours for any plausible wind conditions. Significantly, flow stops more rapidly than it does for lower atmospheric pressure; the slight rise in temperature is more than offset by the increased atmospheric density and improved ability of the wind to remove heat. Calculations have also been performed for the same greenhouse model but a local geothermal heat flux of 1.19 W m^{-2}. In this case, the water in the aquifer is 50 K above the melting temperature, and seepage can proceed indefinitely for some plausible wind speeds. This local gradient, which could be generated by magmatic activity or cratering, is about three times the calculated global average at this point in martian history.

The model shows other interesting behavior. Hydraulic conductivity of the regolith influences seepage rate, and values typical of compact terrestrial soils yield slow seepage and cessation of erosion under both sets of conditions described above. However, the hydraulic conductivities that have actually been calculated for martian regolith from outflow channel dimensions (46) readily yield continuous flow and erosion for a high local heat flow. More important than hydraulic conductivity is atmospheric pressure. At high atmospheric pressure, heat loss to the atmosphere is rapid, and it is difficult to get continuous erosion unless both water temperature and hydraulic conductivity are high. Flow and erosion of course are promoted by higher atmospheric temperatures. However, the pressure effect is sufficiently important that if CO_2 is the only greenhouse gas, sapping is aided by a reduction of atmospheric pressure, not an increase. For example, if the water is hot (50°C) and even if the hydraulic conductivity is typical of compact terrestrial soil, seepage can proceed indefinitely under the present conditions of 218 K and 7 mbar.

The above calculation considers the seepage problem only up to the point where the water emerges from the sapping face; it does not deal with the further flow of the water across the surface. In fact, it is surface flow that appears to be the part of the problem that most requires climatic warming. The only influence that geothermal heat has on surface flow is on the water's temperature as it leaves the sapping face. This can be high, but the water will cool quite quickly once exposed to air.

We have performed calculations to examine the cooling of a martian stream. Energy terms include insolation, radiation from and to the water, conduction to the atmosphere, and evaporative cooling. Atmospheric heat and water vapor transport are calculated following Brutsaert (47). Streamflow velocity is calculated with the Manning equation (41) and a channel gradient of 0.012 (48). Evaporative cooling is the dominant energy term at high temperatures. This is especially true for a low-pressure atmosphere because hot water will boil vigorously when it first comes into contact with such an atmosphere. Because the heat of vaporization of water is large compared with the product of water's heat capacity and the temperature change a stream undergoes, the volume of water evaporated is a small fraction of the total. However, the cooling produced by the removal of the heat of vaporization is very rapid, and even for the steep channel gradient used, the distance flowed before the water begins to freeze can be small under a wide range of pressure conditions. For example, for a CO_2-only greenhouse with a pressure of 1.5 bars, a stream a meter deep and initially at 80°C will cool to 0°C in a few hours, flowing just a few tens of kilometers before freezing begins. Substantial floods (water depths of several meters, requiring very high hydraulic conductivities) are needed for water to flow hundreds of kilometers under these climatic conditions before beginning to freeze. Water flow distances can be increased by reducing atmospheric pressure or by raising atmospheric temperature with minimal pressure increase.

Once freezing begins, substantial flow still might continue. In fact, calculations have suggested that flow could proceed for hundreds of kilometers under a protective ice cover even under the present climatic (*49*). However, terrestrial experience suggests that such models are incorrect (*50*). In arctic streams, structures called icings can form (*51*). As an ice cover begins to develop, hydraulic pressure builds and the ice fractures, allowing water to seep to the surface and spread, often overflowing the channel banks to freeze in thin sheets. Icings halt stream flow much more effectively than calculations predict and could do so on Mars. Icings are common only in shallow streams, so from this standpoint also, deep streams and high hydraulic conductivities promote valley growth under cold climates.

No channels have been resolved in images of ancient martian valleys, so we cannot estimate the depths of the streams that occupied them. For them to have formed at atmospheric temperatures well below the maximum possible with a CO_2-only greenhouse would have required vigorous hydrothermal convection and high hydraulic conductivity in the regolith. This is not impossible, but explaining the flow of water for hundreds of kilometers across the surface is certainly more straightforward with some climatic warming. Mean annual temperatures in excess of 273 K clearly are not necessary, but a few tens of degrees of warming above the nongreenhouse temperature would make the conditions for valley formation much more favorable.

Modest climatic warming could have been produced by supplementing a CO_2 atmosphere with small concentrations of CH_4 and NH_3. A low overall atmospheric pressure is most favorable. If significant climatic warming is required to explain the lengths of the observed valleys (that is, if hydraulic conductivities and stream depths were modest), then a low-pressure CO_2 atmosphere supplemented by CH_4 or NH_3 is the explanation that we favor. The warmest conditions would have occurred near the summer pole during times of maximum obliquity, although for this to have contributed significantly to valley system formation would probably require polar wander, for which good evidence has not been found.

Early Mars was certainly wet. We believe it was warm enough beneath the ground to recharge aquifers by hydrothermal convection and to allow continuous seepage from them. It was also warm enough in the atmosphere to allow the water to flow for long distances across the surface. How warm the atmosphere must have been cannot be determined from current data, but temperatures a few tens of degrees below freezing would have sufficed, and it could have been colder if hydrothermal convection was vigorous, hydraulic conductivities in the regolith were high, and atmospheric pressure was low. Precipitation does not appear to have been necessary, although available data do not rule it out conclusively. Small-scale fluvial activity decreased with time as hydrothermal activity diminished and the atmosphere cooled, although local hydrothermal systems could have accounted for the continuation of isolated fluvial events well into post-Noachian times.

It is clear that near-surface aqueous environments existed on Mars early in its history and could have provided conditions much more favorable for life than those that exist today. Moreover, a speculative suggestion concerning life on Earth is that it originated in submarine hydrothermal vents (*52*), and hydrothermal activity on Mars could have produced similar environments there. Whatever conditions were on early Mars, the best evidence for them may be preserved in materials associated with the ancient valleys. One very important goal for future exploration should be high-resolution imaging of valley floors to search for channel landforms, although this search is likely to be frustrated by the extreme age and degradation state of the valleys. Another goal should be to retrieve samples of the sediments deposited in association with these valleys, so that we can examine the record of environmental and perhaps biological conditions that they could contain.

REFERENCES AND NOTES

1. K. Biemann *et al.*, *J. Geophys. Res.* **82**, 4641 (1977).
2. W. K. Hartmann, *Icarus* **12**, 131 (1970).
3. G. W. Wetherill, *Science* **228**, 877 (1985); in *Origin of the Moon*, W. K. Hartmann *et al.*, Eds. (Lunar and Planetary Institute, Houston, TX, 1986), pp. 519–550.
4. C. R. Chapman and K. L. Jones, *Annu. Rev. Earth Planet. Sci.* **5**, 515 (1977); J. J. Plaut, R. Kahn, E. A. Guiness, R. E. Arvidson, *Icarus* **75**, 357 (1988).
5. R. A. Craddock and T. A. Maxwell, *J. Geophys. Res.* **95**, 14265 (1990).
6. J. M. Goldspiel, S. W. Squyres, D. G. Jankowski, *Icarus* **105**, 479 (1993).
7. D. C. Pieri, *ibid.* **27**, 25 (1976); M. H. Carr and G. D. Clow, *ibid.* **48**, 91 (1981).
8. D. C. Pieri, *Science* **210**, 895 (1980).
9. C. Sagan, O. B. Toon, P. J. Gierasch, *ibid.* **181**, 1045 (1973); J. B. Pollack, *Icarus* **37**, 479 (1979); ______, J. F. Kasting, S. M. Richardson, K. Poliakoff, *ibid.* **71**, 203 (1987).
10. C. P. McKay, O. B. Toon, J. F. Kasting, *Nature* **352**, 489 (1991).
11. E. Anders and T. Owen, *Science* **198**, 453 (1977).
12. J. S. Lewis, *Icarus* **16**, 241 (1972).
13. D. M. Hunten, *Science* **259**, 915 (1993).
14. H. J. Melosh and A. M. Vickery, *Nature* **338**, 487 (1989).
15. An atmosphere should have accumulated during the main phase of the accretion of Mars because the planetesimals that arrived during this time would have come from nearby orbits and should therefore have collided with Mars at a relatively low velocity. Late-arriving impactors from farther out in the solar system would have collided at high velocities and would therefore have been more effective at stripping off atmospheric gases.
16. D. M. Hunten, R. O. Pepin, J. C. G. Walker, *Icarus* **69**, 532 (1987).
17. T. Matsui and Y. Abe, *Nature* **319**, 303 (1986).
18. A. Coradini, C. Federico, P. Lanciano, *Phys. Earth Planet. Inter.* **31**, 145 (1983).
19. J. T. Wasson and G. W. Wetherill, in *Asteroids*, T. Gehrels, Ed. (Univ. of Arizona Press, Tuscon, AZ, 1979), pp. 926–974; C. A. Wood and L. D. Ashwal, *Proc. Lunar Planet Sci. Conf.* **12**, 1359 (1981); D. D. Bogard and P. Johnson, *Science* **221**, 651 (1983); H. Y. McSween, *Rev. Geophys.* **23**, 391 (1985).
20. J. H. Chen and G. J. Wasserburg, *Geochim. Cosmochim. Acta* **50**, 955 (1986).
21. C.-Y. Shih *et al.*, *ibid.* **46**, 2323 (1982).
22. S. M. Clifford and F. P. Fanale, *J. Geophys Res.* **90** (suppl., Proc. Lunar Planet. Sci. Conf. 16), D144 (1985).
23. G. Schubert, S. C. Solomon, D. L. Turcotte, M. J. Drake, N. H. Sleep, in *Mars*, H. H. Kieffer *et al.*, Eds. (Univ. of Arizona Press, Tuscon, AZ, 1992), pp. 147–183.
24. J. B. Pollack, J. F. Kasting, S. M. Richardson, K. Poliakoff, *Icarus* **71**, 203 (1987), and references therein.
25. T. Owen, R. D. Cess, V. Ramanathan, *Nature* **277**, 640 (1979); J. C. G. Walker, P. B. Hays, J. F. Kasting, *J. Geophys. Res.* **86**, 9776 (1981); M. J. Newman and R. T. Rood, *Science* **198**, 1035 (1977).
26. J. F. Kasting, *Icarus* **94**, 1 (1991).
27. D. O. Gough, *Sol. Phys.* **74**, 21 (1981).
28. C. Sagan and G. Mullin, *Science* **177**, 52 (1972); S. E. Postawko and W. R. Kuhn, *J. Geophys. Res.* **91** (suppl., Proc. Lunar Planet. Sci. Conf. 16), D431 (1986).
29. L .L. Brown and J. F. Kasting, in *Workshop on Early Mars: How Warm and How Wet?*, S. Squyres and J. F. Kasting, Eds. (Lunar and Planetary Institute, Houston, TX, 1993), p. 3.
30. C. Sagan and C. F. Chyba, *Bull. Am. Astron. Soc.* **23**, 1211 (abstr.) (1991).
31. Volcanism is a plausible methane source because outgassing at depth shifts the chemical equilibrium from CO_2-H_2O toward CH_4. Deeply submerged hydrothermal vents could have existed if the initial water inventory of Mars was high. Mantle oxygen fugacities one or two log units below the current terrestrial value would be needed to allow significant methane outgassing to occur. Biological inhabitation of early Mars may or may not be considered plausible, but methanogenic bacteria would likely have been present if life had indeed evolved.
32. A. I. Boothroyd, I. J. Sackmann, W. A. Fowler, *Astrophys. J.* **377**, 318 (1991); T. E. Graedel, I. J. Sackmann, A. I. Boothroyd, *Geophys. Res. Lett.* **18**, 1881 (1991).
33. P. J. Gierasch and O. B. Toon, *J. Atmos. Sci.* **30**, 1502 (1973).
34. W. R. Ward, *J. Geophys. Res.* **79**, 3375 (1974). The equatorial/planetary average insolation ratio increases to 27% when Mars is at zero obliquity, which happens occasionally.
35. J. Laskar and P. Robutel, *Nature* **361**, 608 (1993); J. Touma and J. Wisdom, *Science* **259**, 1294 (1993).
36. We assume here that the ice caps would have migrated to the equator during times of high obliquity, allowing the polar surface albedo to be relatively low.
37. P. Goldreich and A. Toomre, *J. Geophys. Res.* **74**, 2555 (1969).
38. C. G. Higgins, *Geology* **10**, 147 (1982); J. E. Laity and M. C. Malin, *Geol. Soc. Am. Bull.* **96**, 203 (1985); R. C. Kochel and J. Piper, *J. Geophys Res.* **91** (suppl., Proc. Lunar Planet. Sci. Conf. 17), E175 (1985).
39. V. C. Gulick and V. R. Baker, *J. Geophys. Res.* **95**, 14325 (1990).
40. S. M. Clifford, in *Third International Colloquium on Mars* (contribution 441, Lunar and Planetary Institute, Houston, TX, 1981), pp. 46–48.
41. J. M. Goldspiel and S. W. Squyres, *Icarus* **89**, 392 (1991); V. C. Gulick and V. R. Baker, in (*29*), pp. 12–13.

42. V. C. Gulick, M. S. Marley, V. R. Baker, *Lunar Planet. Sci.* **XXII**, 509 (1992).
43. H. E. Newsom, *Icarus* **44**, 207 (1980); G. R. Brakenridge, H. E. Newsom, V. R. Baker, *Geology* **13**, 859 (1985).
44. J. M. Goldspiel, thesis, Cornell University, Ithaca, NY (1994).
45. In this model, the heat diffusion equation is solved for a porous regolith, with terms for advection by moving fluid and treatment of freezing or melting within pores. The equations of fluid flow are solved for spatially and temporally variable hydraulic conductivity, with variations caused by pore freezing or melting. The outer thermal boundary condition takes into account radiation and conduction at the seepage face, both to and from the atmosphere, and the effects of wind.
46. M. H. Carr, *J. Geophys. Res.* **84**, 2995 (1979).
47. W. H. Brutsaert, *Evaporation Into the Atmosphere: Theory, History, and Applications* (Reidel, Boston, 1982).
48. This gradient is determined from Earth-based radar altimetry of regional slopes in ancient martian drainage basins [J. M. Goldspiel, S. W. Squyres, M. A. Slade, R. F. Jurgens, S. H. Zisk, *Icarus* **106**, 346 (1993)] and hence probably overestimates actual channel gradients.
49. D. Wallace and C. Sagan, *ibid.* **39**, 385 (1979); M. H. Carr, *ibid.* **56**, 476 (1983).
50. V. R. Baker, M. H. Carr, V. C. Gulick, C. R. Williams, M. S. Marley, in (*23*), pp. 493–522.
51. C. E. Sloan, C. Zenone, I. R. Mayo, *U.S. Geol. Surv. Prof. Pap. 979* (1976).
52. J. A. Baross and S. E. Hoffman, *Origins Life* **15**, 327 (1985).

53. This work was supported by the National Aeronautics and Space Administration Planetary Geosciences and Martian Surface and Atmosphere Through Time programs.

RESEARCH ARTICLE

Volume Holographic Storage and Retrieval of Digital Data

John F. Heanue, Matthew C. Bashaw, Lambertus Hesselink*

A multiple page fully digital holographic data storage system is demonstrated. This system is used to store and retrieve digital image and compressed video data with a photorefractive crystal. Architecture issues related to spatio-rotational multiplexing and novel error-correcting encoding techniques used to achieve low bit-error rates are discussed.

Holographic storage has long held promise for large digital storage capacity, fast data transfer rates, and short access times (*1, 2*). Current storage technologies are limited in that they do not simultaneously provide each of these three features. Recent developments in materials, spatial light modulators (SLMs), and charge-coupled-device (CCD) arrays have brought the promise closer to reality. Potential storage capacity of terabytes of data, transfer rates exceeding 1 gigabit per second, and random access times less than 100 μsec appear feasible, making holographic storage attractive for applications in the planned national information infrastructure, in conventional and parallel computing, and in the entertainment industry.

Holographic recording is accomplished by combining an image-bearing light beam and a reference beam in a recording medium. The variation in intensity in the resulting interference pattern causes the complex index of refraction to be modulated throughout the volume of the medium. In a photorefractive medium such as $LiNbO_3$, charges are excited from impurity centers in the presence of light and subsequently trapped (*3, 4*). The resulting space-charge field causes modulation in the index of refraction through the electro-optic effect. When the medium is exposed to a reference beam identical to one used in recording, the light will diffract in such a way as to reproduce the original image-bearing wavefront. In holographic data storage, data are converted to an optical signal by use of an SLM (Fig. 1). A hologram corresponding to the image (one data "page") on the SLM is then recorded in a photorefractive crystal or other suitable volume holographic recording medium. Multiple holograms, each corresponding to a page of data, are written in the medium using angular multiplexing, in which each hologram is written with a reference beam incident at a different angle. For a 1-cm-long $LiNbO_3$ crystal, holograms may be written with reference beams separated in angle by as little as 50 μrad. The angle can be changed either by steering the beam or by rotational multiplexing, a technique that involves mechanical rotation of the storage medium. The collection of stored data pages superimposed in a particular volume of the crystal is referred to as a stack. Spatial multiplexing is accomplished by dividing the crystal volume into a number of regions and recording one stack per region. Readout of a stored data page involves illuminating the crystal with the appropriate reference beam and imaging the diffracted optical signal onto a CCD array, which converts the optical signal back into an electronic signal. With parallel readout of the CCD array, fast data transfer rates can be achieved because all pixels in the stored data page are reconstructed simultaneously.

Performance criteria. The main criteria used in evaluating the performance of a holographic data storage system are capacity, data transfer rate, access time, and bit error rate. The transfer rate and access time are bounded by the speed of peripheral devices (electronic beam steerers, mechanical stages, CCD array), whereas capacity and bit error rate are determined by the noise level in the system. As the number N of holograms stored in a single stack increases, the diffraction efficiency falls as $1/N^2$. As the strength of the diffracted signal decreases, the signal-to-noise ratio (SNR) decreases because the strength of noise due to scatter is independent of N. The total number of holograms that can be stored is thus determined by the minimum acceptable SNR. A lower bound on the bit error rate (BER) can be determined using the SNR calculated assuming a $1/N^2$ falloff in diffraction efficiency; in practice, the

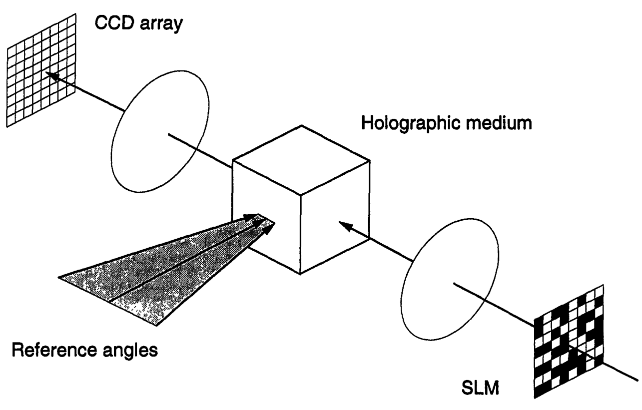

Fig. 1. General holographic recording scheme.

J. F. Heanue is in the Department of Applied Physics, Stanford University, Stanford, CA 94305–4090, USA. M. C. Bashaw and L. Hesselink are in the Department of Electrical Engineering, Stanford University, Stanford, CA 94305–4035, USA.

*To whom correspondence should be addressed.