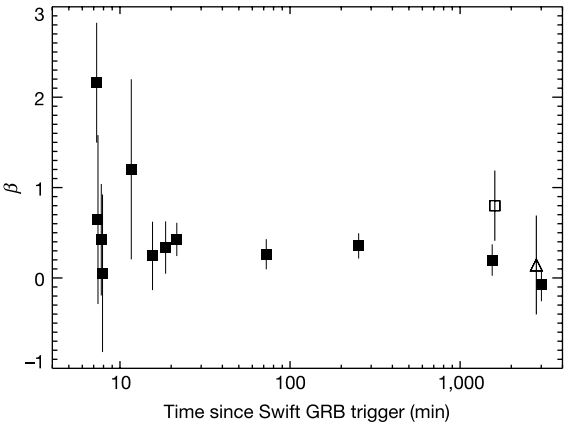

Figure 4 Colour evolution of the afterglow of GRB 041219a, corrected for extinction due to interstellar dust. The dust extinction[29] at the location of GRB 041219a is high: A(V band) = 5.9 (and highly uncertain). Using a common dust extinction law ($R_V = 3.1$), the extinction in magnitudes is A(J) = 1.6, A(H) = 1.04, and $A(K_s) = 0.66$. We fitted (filled squares) for a spectral index β with our J, H and K_s measurements at each epoch using the flux relation $F_\nu(t) = F_0(t)(\lambda_c/\lambda_H)^\beta$, where λ_c is the central wavelength of the filter bandpasses (c = [J, H, K]). Including the z-band detection[30] ($\lambda_c \approx 9{,}100$ Å) that was nearly coeval with our IR observations results in a steeper spectral slope of 0.80 ± 0.39 (open square) than found with the IR data alone. Our fit to the Apache Point JHK data[20] is also shown with an open triangle. With fits that yielded χ^2/d.f. greater than unity, we scale the resulting 1σ errors by $\sqrt{\chi^2}$/d.f. Though there is some evidence for secular trends, the resulting values for β are consistent with a single value of $\beta \approx 0.4$. We caution, however, that this common value of the spectral index is strongly dependent upon the assumed dust column. For different values of the extinction, A(V) = 4.0 and A(V) = 7.0, the fits to the spectral index are $\beta \approx 0.7$ and $\beta \approx 0.2$, respectively. Indeed, the standard forward-shock model[22] tends to favour a larger value ($\beta \approx 0.55$) than found with the IR data alone. For the first peak (during the GRB), there is marginal evidence that the source is intrinsically redder than at later times.

imaging and the rapid fall behaviour, the interpretation is largely unchanged by various analyses. □

Received 12 January; accepted 1 March 2005; doi:10.1038/nature03520.

1. Fenimore, E. E., Madras, C. D. & Nayakshin, S. Expanding relativistic shells and gamma-ray burst temporal structure. *Astrophys. J.* **473,** 998 (1996).
2. Mészáros, P. & Rees, M. J. Optical and long-wavelength afterglow from gamma-ray bursts. *Astrophys. J.* **476,** 232–237 (1997).
3. Sari, R. & Piran, T. Predictions for the very early afterglow and the optical flash. *Astrophys. J.* **520,** 641–649 (1999).
4. Beloborodov, A. M. Radiation front sweeping the ambient medium of gamma-ray bursts. *Astrophys. J.* **565,** 808–828 (2002).
5. van Paradijs, J., Kouveliotou, C. & Wijers, R. A. M. J. Gamma-ray burst afterglows. *Annu. Rev. Astron. Astrophys.* **38,** 379–425 (2000).
6. Baring, M. G. & Braby, M. L. A study of prompt emission mechanisms in gamma-ray bursts. *Astrophys. J.* **613,** 460–476 (2004).
7. Akerlof, C. *et al.* Observation of contemporaneous optical radiation from a gamma-ray burst. *Nature* **398,** 400–402 (1999).
8. Gotz, D., Mereghetti, S., Shaw, S., Beck, M. & Borkowski, J. GRB 041219—A long GRB detected by INTEGRAL. *GCN Circ.* 2866 (2004).
9. Fox, D. W. *et al.* Early optical emission from the γ-ray burst of 4 October 2002. *Nature* **422,** 284–286 (2003).
10. Fox, D. W. *et al.* Discovery of early optical emission from GRB 021211. *Astrophys. J.* **586,** L5–L8 (2003).
11. Li, W., Filippenko, A. V., Chornock, R. & Jha, S. The early light curve of the optical afterglow of GRB 021211. *Astrophys. J.* **586,** L9–L12 (2003).
12. Rykoff, E. S. *et al.* The early optical afterglow of GRB 030418 and progenitor mass loss. *Astrophys. J.* **601,** 1013–1018 (2004).
13. Klose, S. *et al.* The very faint K-band afterglow of GRB 020819 and the dust extinction hypothesis of the dark bursts. *Astrophys. J.* **592,** 1025–1034 (2003).
14. Zerbi, F. M. *et al. Proc. SPIE* **4841,** 737–748 (2003).
15. Winkler, C. *et al.* The INTEGRAL mission. *Astron. Astrophys.* **411,** L1–L6 (2003).
16. Gehrels, N. *et al.* The Swift gamma-ray burst mission. *Astrophys. J.* **611,** 1005–1020 (2004).
17. Blake, C. & Bloom, J. S. GRB 041219: Infrared afterglow candidate. *GCN Circ.*, 2870 (2004).
18. Vestrand, W. T. *et al.* A link between prompt optical and prompt γ-ray emission in γ-ray bursts. *Nature* doi:10.1038/nature03515 (this issue).
19. Matthews, K., Soifer, B. T. & McLean, I. in *Infrared Astronomy with Arrays, the Next Generation* (ed. McLean, I.) 239 (Kluwer, Dordrecht, 1994).
20. Hearty, F. *et al.* NIR observations of GRB 041219. *GCN Circ.* 2916 (2004).
21. Kobayashi, S. Light curves of gamma-ray burst optical flashes. *Astrophys. J.* **545,** 807–812 (2000).
22. Sari, R., Piran, T. & Narayan, R. Spectra and light curves of gamma-ray burst afterglows. *Astrophys. J.* **497,** L17–L20 (1998).
23. van der Horst, A., Rol, E. & Strom, R. GRB 041219: Second epoch WSRT radio observations. *GCN Circ.*, 2895 (2004).
24. Sari, R. Hydrodynamics of gamma-ray burst afterglow. *Astrophys. J.* **489,** L37–L40 (1997).
25. Zhang, B., Kobayashi, S. & Mészáros, P. Gamma-ray burst early optical afterglows: Implications for the initial Lorentz factor and the central engine. *Astrophys. J.* **595,** 950–954 (2003).
26. Ramirez-Ruiz, E. & Fenimore, E. E. Pulse width evolution in gamma-ray bursts: Evidence for internal shocks. *Astrophys. J.* **539,** 712–717 (2000).
27. Nakar, E. & Piran, T. Early afterglow emission from a reverse shock as a diagnostic tool for gamma-ray burst outflows. *Mon. Not. R. Astron. Soc.* **353,** 647–653 (2004).
28. Moon, D.-S., Cenko, B. & Adams, J. GRB041219: Continued NIR observations. *GCN Circ.* 2884 (2004).
29. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500,** 525–553 (1998).
30. Cenko, S. B. GRB 041219: Optical afterglow detection. *GCN Circ.* 2885 (2004).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements J.S.B. was supported by a Junior Fellowship from the Harvard Society of Fellows. PAIRITEL was made possible by a grant from the Harvard Milton Fund. Additional funding from the Smithsonian Institution for the PAIRITEL project is acknowledged. We thank the entire Mt Hopkins Ridge staff for support of PAIRITEL, especially W. Peters, R. Hutchins and T. Groner. J. Huchra is thanked for stewardship of the telescope under adverse conditions in the week leading up to GRB 041219a. This publication makes use of data products from the Two Micron All Sky Survey (2MASS), which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.B. (jbloom@astro.berkeley.edu).

Albedo of the south pole on Mars determined by topographic forcing of atmosphere dynamics

Anthony Colaprete[1], Jeffrey R. Barnes[2], Robert M. Haberle[1], Jeffery L. Hollingsworth[3], Hugh H. Kieffer[4] & Timothy N. Titus[4]

[1]*NASA Ames Research Center, Space Science Division, Moffett Field, Mountain View, California 94035, USA*
[2]*College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA*
[3]*San José State University Foundation, NASA Ames Research Center, Moffett Field, Mountain View, California 94035, USA*
[4]*US Geological Survey, Astrogeology Team, Flagstaff, Arizona 86001, USA*

The nature of the martian south polar cap has remained enigmatic since the first spacecraft observations[1–6]. In particular, the presence of a perennial carbon dioxide ice cap, the formation of a vast area of black 'slab ice' known as the Cryptic region and the asymmetric springtime retreat of the cap have eluded explanation. Here we present observations and climate modelling that indicate the south pole of Mars is characterized by two distinct regional climates that are the result of dynamical forcing by the largest southern impact basins, Argyre and Hellas. The style of surface frost deposition is controlled by these regional climates. In the cold and stormy conditions that exist poleward of 60° S and extend 180° in longitude west from the Mountains of Mitchel (~30° W), surface frost accumulation is dominated by precipitation. In the opposite hemisphere, the polar atmosphere is relatively warm and clear and frost accumulation is dominated by direct vapour deposition. It is the differences in these deposition styles that determine the cap albedo.

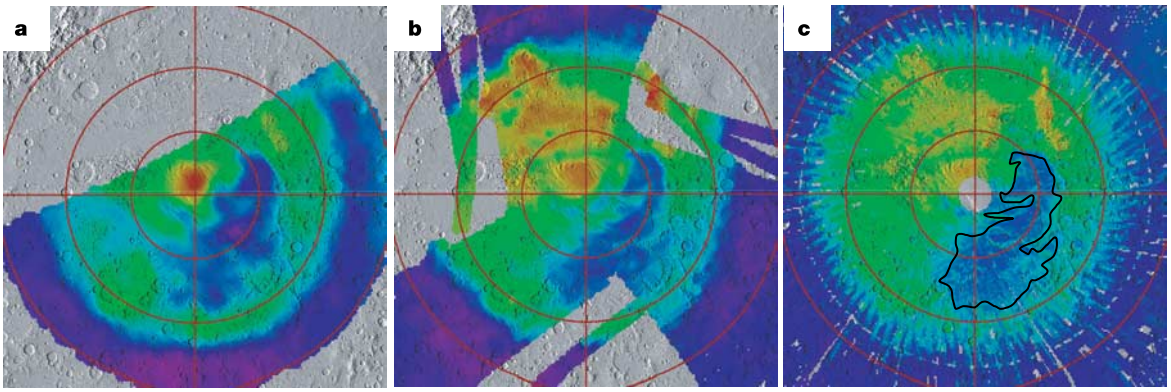

Figure 1 South pole springtime albedo maps from the Viking and Mars Global Surveyor spacecraft. Three albedo maps of the martian south springtime pole overlain on MOLA topography. In each panel, high albedos are indicated by warm colours (red shading) and low albedos are indicated by cool colours (blue shading). The red lines indicate every 10° of latitude and 90° of longitude. At the time of these observations the seasonal polar cap, composed primarily of CO_2 ice, extends from the pole to approximately 65° S. **a**, Albedo measured by the Infrared Thermal Mapper on board the Viking spacecraft in 1976. **b**, **c**, Albedo derived from observations made by the Thermal Emission Spectrometer on board the Mars Global Surveyor (MGS) spacecraft in 1999 (**b**) and 2001 (**c**). The black curve in **c** indicates the approximate boundary of the Cryptic region. Comparing a with **b** and **c**, there appear to be only minor differences in albedo between the Viking and MGS eras, with the largest differences occurring in parts of the Cryptic region.

The observed springtime (solar longitude, $L_s \approx 200$) surface albedo in the martian southern polar region is shown in Fig. 1. The Hellas impact basin is centred at −25° S and 65° W. In general, the hemisphere west of Hellas ('the western hemisphere') is marked by relatively high values of surface albedo. In contrast, the hemisphere east of Hellas ('the eastern hemisphere') contains extensive regions of very low surface albedo. One of the brightest features within the western hemisphere is the south pole residual cap (SPRC). The dark region that dominates the eastern hemisphere is the Cryptic region[5].

The nature of the SPRC has been the source of considerable debate since its identification as CO_2 ice by the Viking spacecraft. Two fundamental questions still exist regarding the SPRC's formation, location and stability. First, why is the SPRC offset from the geographic pole? There are no local topographic features or surface properties that can account for the offset in the SPRC. Second, does the SPRC represent a large or a small reservoir of CO_2? If the former, then it could possibly buffer the surface pressure[1]. If the latter, then the SPRC may not survive every year.

Just as mysterious as the SPRC is the region of 'black' CO_2 ice that dominates the eastern hemisphere (Fig. 1). Known as the Cryptic region[5], this area, first observed by Viking and later by Mars Global Surveyor (MGS), becomes visible during the recession of the seasonal cap. The Cryptic region is characterized by its very low visible albedo (<0.2) and high 25-μm relative emissivity (>0.95) and is believed to be composed of very clean CO_2 'slab' ice. It is the Cryptic region that first sublimes in spring, leading to the observed asymmetric retreat of the southern seasonal cap.

The difference in albedo between the SPRC and the Cryptic region suggests two styles of frost emplacement, and yet, to date, no obvious correlation between topography and surface properties that would favour a SPRC or Cryptic region have been identified[5]. However, independent observations from MGS indicate that the atmospheric circulation within the polar night produces two distinct climates that encompass the SPRC and the Cryptic region. These observations include cloud echoes from the Mars Orbiter Laser Altimeter (MOLA), ice grain sizes from the Thermal Emission Spectrometer (TES) and temperatures from TES and MGS Radio Science (RS).

This work, using simulations with the NASA Ames Mars General Circulation Model (MGCM), provides a consistent explanation for the various observations and offers a simple interpretation—the topography associated with the Tharsis rise and the Hellas and Argyre impact basins modulates the south polar circulation, producing two distinct climates. Ultimately, the two climates generated by the circulation result in very different surface properties for the CO_2

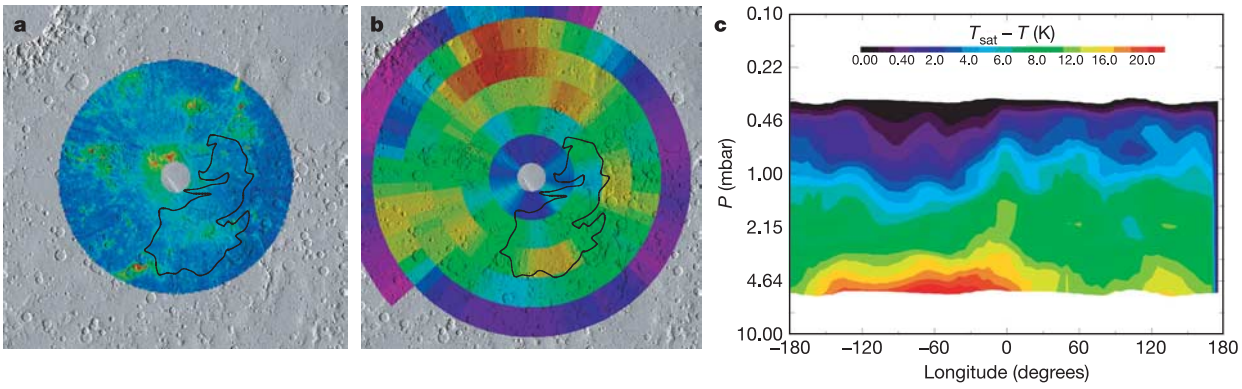

Figure 2 Observations from MGS showing south pole climate asymmetry in grain size and $\delta T_{sat} = T_{sat} - T$. **a**, Average grain size index from $L_s = 87$–110. High values of the grain index (smaller grains) are indicated by warm colours (red shading), and low values of the grain index (larger grains) are indicated by cool colours (blue shading). **b**, Column-integrated δT_{sat} derived from TES nadir atmospheric temperatures for the period $L_s = 90$–110. The black contour shows the approximate boundary of the Cryptic region (Fig. 1). **c**, Longitude cross-section at 75° S showing δT_{sat} derived from TES as a function of pressure.

ice cap. Within the hemisphere that includes the SPRC, atmospheric precipitation dominates direct surface deposition, resulting in high albedo frosts; in the Cryptic region, cap ice is accumulated by direct deposition, forming dark CO_2 slab ice.

The albedo and emissivity of the polar cap has been associated with the size of the ice grains that make up the surface ice deposits[5,7,8]. Small ice grains result in brighter surfaces, while larger grains result in darker surfaces. A measure of the ice grain size can be derived from differences in the 18-μm and the 25-μm brightness temperatures observed by the TES. The 18-μm brightness temperature is a good measure of the true kinetic temperature, while the 25-μm band is very sensitive to grain size. The 'grain size index' (Fig. 2a) is the difference between the two bands ($T_{18} - T_{25}$). Regions with high values of the grain size index include the SPRC and the Mountains of Mitchel. It has been suggested that high grain size indexes are the result of small ice grains in the atmosphere or freshly precipitated to the surface[8]. MOLA detected cloud echoes within the polar night that have been attributed to the presence of CO_2 clouds[9–11]. The greatest rates of occurrence are in the western hemisphere. In the eastern hemisphere there are fewer cloud echoes, with a minimum poleward of Hellas basin[12].

Also correlating with surface albedos are temperatures derived from TES and RS. Frequently the atmosphere within the polar night chills to below the condensation temperature of CO_2 ice and becomes supersaturated. The greatest and most persistent supersaturation, $\delta T_{sat} = T_{sat} - T$, occurs in the western hemisphere and in the longitude corridor of the SPRC (Fig. 2b). Atmospheric temperatures are significantly colder in the western hemisphere compared with the eastern hemisphere. The greatest values of δT_{sat} correlate well with surface brightness and grain size.

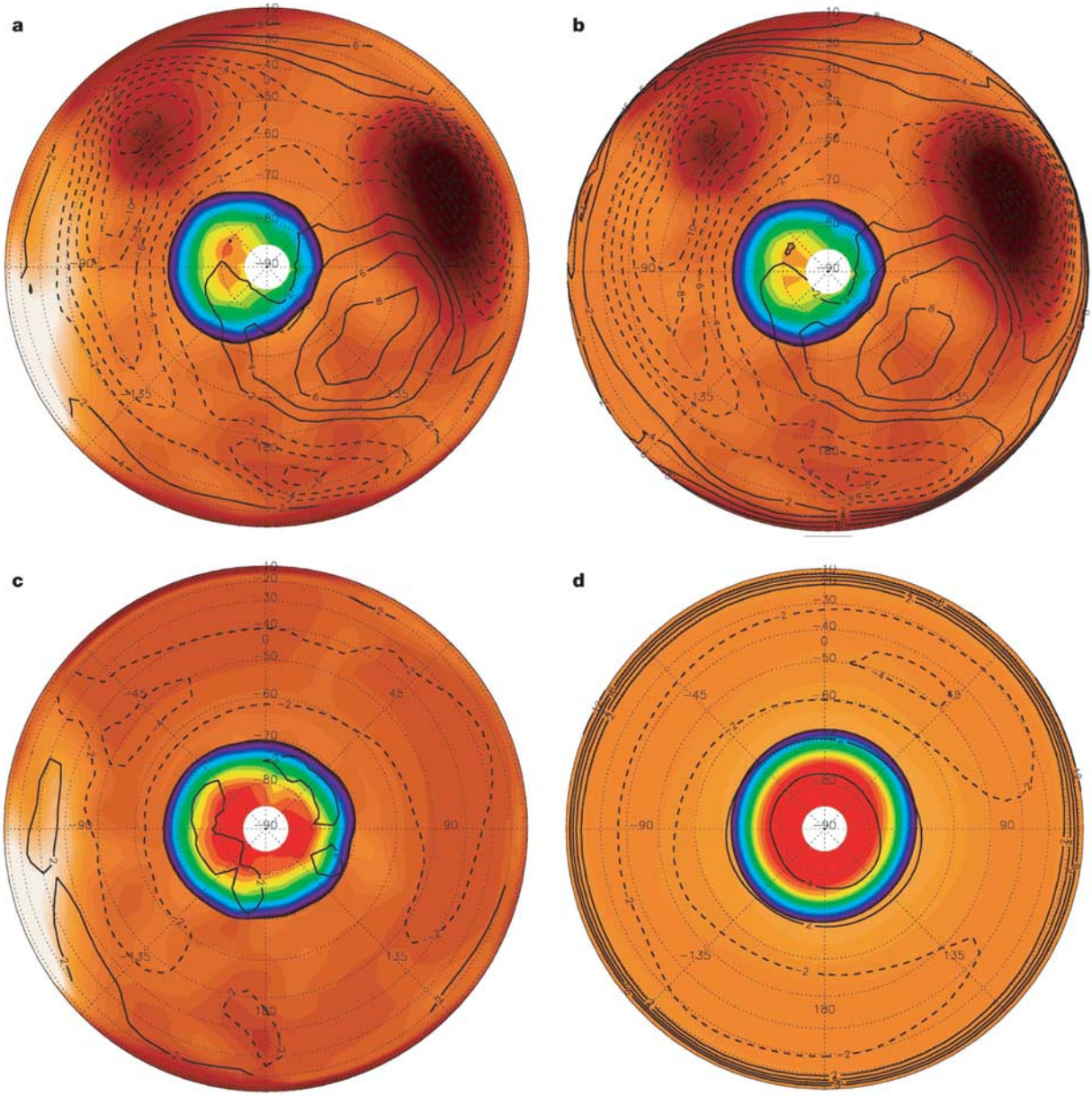

Figure 3 GCM results showing the dependence of the climate asymmetry on mid-latitude topography. All panels show NASA Ames MGCM simulation results for column-integrated δT_{sat} (colour shading with values corresponding to the colour bar shown in Fig. 2) at $L_s = 90$. Black contours indicate temperature deviations from the time-averaged (20 solar days) zonal mean at the 3.5-mbar pressure level, with solid contours indicating temperatures higher than the zonal mean (high-pressure regions) and dashed contours indicating temperatures lower than the zonal mean (low-pressure regions). Model topography is indicated with the dark red–white background shading: **a**, Current true topography; **b**, Tharsis removed; **c**, Hellas and Argyre removed; **d**, Flat topography.

It is the atmospheric temperatures that provide insight into what is controlling these two climate regimes. The δT_{sat} for a longitude cut at constant latitude of 75° S is shown in Fig. 2c. Evident in the temperature deviations (Fig. 2b, c) is a low-level planetary wave with a very strong stationary zonal wavenumber-one component. Hinson *et al.*[13] had previously identified this stationary wave in RS and TES observations. The wavenumber-one pattern establishes a low-pressure zone over much of the western hemisphere and a high-pressure zone over much of the eastern hemisphere. This wavenumber-one pattern dominates most of the southern winter, and extends into southern spring ($L_s \approx 230$). The strong wavenumber-one pattern results in a polar vortex that is offset from the geographic pole and sits over the SPRC.

Furthermore, the amount of variance in atmospheric temperatures is also greatest in the western hemisphere. A distinct 'storm track' exists, starting west of Hellas and extending west past Argyre[14]. Within this storm track temperatures can vary by as much as 15–20 K per solar day. The rapid decreases in temperature drive the formation of thick clouds and increased precipitation.

To further understand the southern winter dynamics and their association with the SPRC and Cryptic region, several simulations were conducted with the NASA MGCM[15]. The version of the MGCM used here includes a sophisticated CO_2 cloud scheme that predicts cloud formation and precipitation[16]. Simulations were conducted for a variety of topographic permutations in which the southern extent of Tharsis, Hellas and Argyre were included or removed (Fig. 3).

In the simulation with true (that is, current MOLA observed) topography, temperatures predicted by the model closely match those observed and, in particular, the coldest temperatures are observed in the western hemisphere and encompass the SPRC. Predicted CO_2 cloud cover, precipitation and δT_{sat} agree well with the observations described so far (compare Figs 3a and 2b). During the formation of the seasonal cap, cloud cover and precipitation are greatest in a longitude corridor that aligns with the SPRC and are almost completely absent over the Cryptic region. A strong wavenumber-one system dominates from late autumn to mid-winter and establishes a pattern of low and high surface pressures that are consistent with the observations.

When the topography is altered, the regional circulation and, hence, atmospheric temperatures are changed. Figure 3b shows the effect of removing the southernmost extent of the Tharsis rise. Without the southern extent of Tharsis, the hemispherical asymmetry in δT_{sat}, while still evident, is rotated west by about 45° and has somewhat smaller values compared with the case with true topography. If the Hellas and Argyre impact basins are removed (Fig. 3c), the wavenumber-one circulation pattern no longer dominates and cold temperatures occur with much greater symmetry about the pole. A very symmetrical pattern of circulation and cold temperatures results if all topography is removed (Fig. 3d). The simulations indicate that the south polar circulation is controlled by a quasi-stationary planetary (that is, Rossby) wave[17] that produces longitudinal variation in the time-mean (for example, seasonal-mean) circulation and climate. The longitudinal asymmetries of the order of 20–30 K seen in the thermal environment between the western and eastern hemispheres in southern high latitudes are a result of this planetary wave that is excited primarily by the planet's middle-latitude surface topography distribution[18,19].

The two climates established by the topographically forced circulation produce differences in the nature of the surface ice deposits. From the MGCM simulations, the relative amounts of CO_2 ice that is deposited by precipitation and by direct deposition can be compared. The deposition ratio is defined as the amount of atmospheric precipitation ($kg\,m^{-2}\,s^{-1}$) divided by the amount of direct surface deposition of CO_2 ice. The deposition ratio is area-weighted so that longitude corridors of differing extent can be compared. On average, across the entire polar cap, the deposition ratio is approximately 0.1. However, in some regions the deposition ratio can be as high as 0.9. The greatest deposition ratios occur in the western hemisphere. In the 'true topography' case of the MGCM, the deposition ratio between the longitudes of −135° W and 45° W is 1.56 times as large as the deposition ratio in the opposite hemisphere. If the western edge of this corridor is extended to −180° W, so that the western corridor now extends from −180° W to 45° W, and the eastern hemisphere corridor is confined to just the longitude extent of the Cryptic region, the deposition ratio is 22.4 times as large.

Jakosky and Haberle[20] suggested that the SPRC may be unstable from year to year, and that small perturbations in climate may result in the SPRC disappearing some years and reappearing in others. In spring, the recession of the seasonal cap is largely controlled by the surface brightness. Although albedos in local regions on the seasonal cap can change rapidly, the persistent asymmetric circulation creates conditions in which relatively larger grain sizes are stable in the Cryptic region through the sublimation season. Any changes to the circulation can result in changes to the surface brightness and poleward heat flow, and thus modify the recession of the cap. The current SPRC albedo is very close to the minimum required for year-round stability[20]. If these albedos are only slightly modified by changes in the location or quantity of precipitation, then the SPRC will undergo significant changes and may partially or completely disappear. The observed changes in polar ice features reported by Malin *et al.*[21] and Byrne and Ingersoll[22,23] are consistent with the notion of a SPRC that is only quasi-stable and may undergo significant changes from one year to the next.

The albedo of the south polar ice cap is primarily controlled by the frost grain size. Frost grain size is determined by the style of frost deposition. Smaller grains form when there is atmospheric precipitation to the surface and larger grains form when frost is directly deposited to the surface. Locations that favour one style of frost emplacement over the other are primarily determined by the regional circulation and resulting thermal state. Thus topography, through its influence on the circulation, is ultimately responsible for the nature and location of the SPRC and Cryptic regions on Mars. □

Methods

The NASA Ames MGCM simulations presented here were run at a resolution of 5° latitude × 6° longitude. A constant atmospheric dust optical depth of 0.3 was assumed in all runs. Included in these simulations is a sophisticated carbon dioxide cloud scheme that models the microphysical cloud processes of nucleation, condensation/sublimation, and sedimentation[16]. There are several significant improvements in this MGCM CO_2 cloud model over previous models. Most relevant to the work presented here is the treatment of the nucleation and growth processes. The formation of new cloud particles occurs through the heterogeneous nucleation on dust or water-ice particles. A condensational growth rate is calculated at each time step (that is, it is not assumed to be instantaneous) and cloud particles are treated as tracers with their sedimentation rate calculated across the distribution of particle sizes. To accurately treat the growth of both cloud particles and direct deposition of surface ice, CO_2 vapour is treated as a tracer, independent of the total pressure. These features of the cloud model allow for supersaturation, as seen in the observations, and provide a much better representation of actual polar processes including precipitation and direct surface deposition.

Received 25 October 2004; accepted 10 March 2005; doi:10.1038/nature03561.

1. Leighton, R. L. & Murray, B. C. Behaviour of carbon dioxide and other volatiles on Mars. *Science* **153,** 136–144 (1966).
2. Paige, D. A. & Ingersoll, A. P. Annual heat balance of Martian polar caps Viking observations. *Science* **228,** 1160–1168 (1985).
3. Lindner, B. L. The hemispherical asymmetry in the Martian polar caps. *J. Geophys. Res.* **98,** 3339–3344 (1993).
4. James, P. B. *et al.* Seasonal recession of Martian South Polar Cap; 1992 HST observations. *Icarus* **123,** 87–100 (1996).
5. Kieffer, H. H., Titus, T. N., Mullins, K. F. & Christensen, P. R. Mars south polar spring and summer behaviour observed by TES: Seasonal cap evolution controlled by frost grain size. *J. Geophys. Res.* **105,** 9653–9700 (2000).
6. Thomas, P. C. *et al.* North–south geological differences between the residual polar caps on Mars. *Nature* **404,** 161–164 (2000).
7. Forget, F., Hourdin, F. & Talagrand, O. CO_2 snowfall on Mars: Simulation with a general circulation model. *Icarus* **131,** 302–316 (1998).
8. Titus, T. N., Kieffer, H. H., Mullins, K. F. & Christensen, P. R. TES premapping data: Slab ice and snow

flurries in the Martian north polar night. *J. Geophys. Res.* **106,** 23181–23196 (2001).
9. Pettengill, G. H. & Ford, P. G. Winter clouds over the north Martian polar cap. *Geophys. Res. Lett.* **27,** 609–613 (2001).
10. Ivanov, A. B. & Muhleman, D. O. Cloud reflection observations: results from the Mars Orbiter Laser Altimeter. *Icarus* **154,** 190–206 (2001).
11. Colaprete, A. & Toon, O. B. Carbon dioxide snow storms during the polar night on Mars. *J. Geophys. Res.* **107,** doi:10.1029/2001JE001758 (2002).
12. Colaprete, A., Haberle, R. M. & Toon, O. B. Formation of convective carbon dioxide clouds near the south pole of Mars. *J. Geophys. Res.* **108,** doi:10.1029/2003JE002053 (2003).
13. Hinson, D. P., Wilson, R. J., Smith, M. D. & Conrath, B. J. Stationary planetary waves in the atmosphere of Mars during southern winter. *J. Geophys. Res.* **108,** doi:10.1029/2002JE001949 (2003).
14. Hinson, D. P. & Wilson, R. J. Transient eddies in the Southern Hemisphere of Mars. *Geophys. Res. Lett.* **29,** doi:10.1029/2001GL014103 (2002).
15. Haberle, R. M. *et al.* General circulation model simulations of the Mars Pathfinder atmospheric structure investigation/meteorology data. *J. Geophys. Res.* **104,** 8957–8974 (1999).
16. Colaprete, A. & Toon, O. B. Carbon dioxide clouds in an early dense Martian atmosphere. *J. Geophys. Res.* **108,** doi:10.1029/2002JE001967 (2003).
17. Holton, J. R. *An Introduction to Dynamic Meteorology* 3rd edn (Academic, London, 1992).
18. Hollingsworth, J. L. & Barnes, J. R. Forced, stationary planetary waves in Mars' winter atmosphere. *J. Atmos. Sci.* **53,** 428–448 (1996).
19. Banfield, D., Conrath, B. J., Smith, M. D., Christensen, P. R. & Wilson, R. J. Forced waves in the Martian atmosphere from MGS TES nadir data. *Icarus* **161,** 319–345 (2003).
20. Jakosky, B. M. & Haberle, R. M. Year-to-year instability of the Mars south polar cap. *J. Geophys. Res.* **95,** 1359–1365 (1990).
21. Malin, M. C., Caplinger, M. A. & Davis, S. D. Observational evidence for an active surface reservoir of solid carbon dioxide on Mars. *Science* **294,** 2146–2148 (2001).
22. Byrne, S. & Ingersoll, A. P. A sublimation model for Martian south polar ice features. *Science* **299,** 1051–1053 (2003).
23. Byrne, S. & Ingersoll, A. P. Martian climate events on timescales of centuries: Evidence from feature morphology in the residual south polar ice cap. *Geophys. Res. Lett.* **30,** doi:10.1029/2003GL017597 (2003).

Acknowledgements We acknowledge discussions with F. Montmessin and F. Forget. This work was supported under NASA's Planetary Atmospheres Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (Anthony.Colaprete-1@nasa.gov).

Friction enhances elasticity in granular solids

C. Goldenberg[1]* **& I. Goldhirsch**[2]

[1]*School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences,* [2]*Department of Fluid Mechanics and Heat Transfer, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel*

* Present address: Physique et Mécanique des Milieux Hétérogènes, ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 5, France

For years, engineers have used elastic and plastic models to describe the properties of granular solids, such as sand piles and grains in silos[1–3]. However, there are theoretical[4–6] and experimental[7–14] results that challenge this approach. Specifically, it has been claimed[4–6] that stress in granular solids propagates in a manner described by wave-like (hyperbolic) equations, rather than the elliptic equations of static elasticity. Here we report numerical simulations of the response of a two-dimensional granular slab to an external load, revealing that both approaches are valid—albeit on different length scales. For small systems that can be considered mesoscopic on the scale of the grains, a hyperbolic-like, strongly anisotropic response is expected. However, in large systems (those typically considered by engineers), the response is closer to that predicted by traditional isotropic elasticity models. Static friction, often ignored in simple models, plays a key role: it increases the elastic range and renders the response more isotropic, even beyond this range.

Collections of macroscopic particles, known as granular materials, are commonplace in the energy, pharmaceutical, chemical and food industries, as well as in the environment. Under the influence of gravity, grains organize in static assemblies. Understanding the properties of these granular solids is of practical and scientific importance. Traditionally, engineers have used elastic theory[1,2] to describe granular bulk properties at low loads, and elasto-plastic models[1–3] to describe yield and quasi-static flow. Surprisingly, in some experiments the pressure on the floor under sand piles was found to possess a local minimum under the apex[13–14], whereas isotropic elastic theory (and common sense) would imply that the pressure is maximal there (we note that the presence of this minimum depends on the method of preparation of the pile[15]). Other experiments[7–9] revealed the existence of anisotropic chains of particle contacts ('force chains') along which the contact forces were stronger than average. These and other experiments[10–12] were interpreted as evidence against elastic theories of granular matter. Models[4–6] in which the forces propagate, like waves, along force chains, were proposed and some of their predictions were found to be in good agreement with experiment. The basic equations in these theories are hyperbolic, in contrast with the elliptic nature of static elasticity, implying a possible need to revisit important parts of engineering and environmental science. Here we show that this is not required, mostly because the new approaches (which often ignore interparticle friction[6]) apply mainly to mesoscopic-sized systems. An understanding of the crossover from the small- to the large-scale granular physics is essential for a better understanding of granular matter in general.

Consider a vertical force applied to the top of a rectangular layer ('slab') of a granular material, positioned on a horizontal surface. In a hyperbolic model[4–6], the forces propagate along a cone, or two lines in two dimensions (2D), so that the maximal pressure on the floor is not below the point of application of the force, but rather on a ring whose diameter increases with the depth of the layer (or two points, in 2D). The pressure distribution at the floor therefore exhibits two peaks (in 2D). In contrast, isotropic elasticity predicts a single peak, whose width is proportional to the depth of the layer[16,17], as observed in some experiments[16,18]. Other experiments appear to be compatible with a hyperbolic description (for example, for lattice configurations[8–10]). Experiments therefore seem to be providing mixed messages on the correct description of granular statics.

As two-dimensional simulations are easier to perform and visualize, we set out to study the microscopic basis of the above-mentioned findings in 2D. We studied two-dimensional slabs, composed of monodisperse or polydisperse disks, using a standard grain interaction model[19]: the normal and tangential forces, denoted by F_n and F_t respectively, are determined by spring-dashpot models (with spring constants k^n and k^t, respectively). A transition to sliding occurs when the Coulomb friction law is saturated: $F_t = \mu F_n$, where μ is the coefficient of static friction. Our numerical experiments proceeded as follows. First, an initial configuration of non-touching grains is prepared in an enclosure composed of rigid side walls and floor. The system is then relaxed under gravity to form a stationary slab (for details, see Fig. 1 legend). A downward vertical force (F_{ext}) is slowly applied to the top of the system until it reaches the desired value; the system is then further relaxed to a numerically static state.

Consider first a monodisperse system, with the disks arranged on a triangular lattice. The force profile on the floor for different values of F_{ext} is shown in Fig. 1a–c, for systems with different coefficients of friction. A crossover from a single-peaked to a two-peaked response occurs as the applied force is increased (except for $\mu = 1$), the crossover value of the force increasing with μ. The dependence of the value of the force on the floor below the point of application of F_{ext} ($x = 0$) on F_{ext} is shown in Fig. 1d. An elastic-like response (linear in F_{ext}) is observed for small F_{ext}. The extent of the range in which the response is linear in F_{ext} is significantly larger in frictional systems. As we have verified that superposition holds as