**Supplementary information: Causes of ice-age intensification across the Mid-Pleistocene Transition**

Thomas B. Chalk1,2,\*, Mathis P. Hain1\* Gavin L. Foster1, Eelco J. Rohling3,1, Philip F. Sexton4,Marcus P.S. Badger4,5, Soraya G. Cherry1, Adam P. Hasenfratz6, Gerald H. Haug7, Samuel L. Jaccard8, Alfredo Martínez-García7, Heiko Pälike9,1, Richard D. Pancost5, Paul A. Wilson1

**1. Sample locations and age model**

**2. Methodology**

**2.1 Analytical techniques**

**2.2 Determination of pH from δ11B of *Globigerinoides ruber***

**2.3 Determination of CO2 from δ11B-derived pH**

**3. Forcing-to-SL relationships**

**4. Carbon cycle modeling**

**4.1 Model details**

**4.2 Extension of ODP 1094 Ba/Fe proxy forcing**

**4.3 Model inversion**

**5. Quantification of IGCO2, IGCO2 and IG-G∆CO2**

**6. Supplementary figure captions**

**1. Sample locations and age model**

We present a new highly resolved (1 sample per 3.5-4.5 kyr) δ11B-derived atmospheric CO2 record from ODP Site 999 (Caribbean Sea, 12o44.64’N and 78o44.36’W, see Fig. S1) that spans the 0-260 kyr (i.e., LP260) and 1080-1250 kyr (i.e., eMPT) time intervals. Sedimentation rates in these intervals are ~3 cm/kyr. Across large parts of the oceans (typically the oligotrophic regions of the low latitudes) the CO2 content of the gas phase of surface seawater approximately equates to the CO2 content of the overlying atmosphere (i.e., the surface water is in approximate CO2 equilibrium with the atmosphere). It is from these locations, such as ODP Site 999 (Fig S1), that atmospheric CO2 concentrations can be most reliably reconstructed using the δ11B method (34). Site 999 is today in near air-sea CO2 equilibrium (approximately +20 μatm; Fig S1, ref 35) and has likely maintained this near-equilibrium state through the Pleistocene and Pliocene (20,36). The foraminifera at ODP 999 are well preserved meaning diagenesis is not likely to negatively impact our reconstructions. It is also worth noting that even moderate to severe diagenesis has been shown to have little or no impact on the δ11B of planktic foraminifera (37).

For our age model we generated a benthic δ18O record for ODP Site 999 from ~600 to 1500 kyr at a resolution of 3 kyr using *Cibicidoides wuellerstorfi* that includes the same samples used to determine CO2 during our eMPT interval (Fig. S2). The samples were measured on a Finnigan MAT 253 gas isotope ratio mass spectrometer connected to a Kiel IV automated carbonate preparation device at the Zentrum für Marine Tropenökologie, Bremen. A detailed age model for Site 999 was generated by graphically aligning our benthic foraminiferal δ18O record to the LR04 benthic δ18O stack (26) (Fig. S2) using the Analyseries software (38). For the interval 0-500 kyr, we generated a detailed age model by aligning the published planktic foraminiferal (*G. ruber*) δ18O record from Site 999 (at ~0.5 to 2.0 kyr resolution, 39)6 to the LR04 benthic δ18O stack (26, Fig. S2). The resultant age model provides excellent agreement between our new lower-resolution benthic δ18O data and the LR04 benthic δ18O stack in the interval 110 to 260 kyr (Fig. S2).

**2. Methodology**

**2.1 Analytical techniques**

Between 140 and 220 individuals of *Globigerinoides ruber* (white, *sensu stricto*) *(G. ruber*, ~10 μg/shell) were picked from the 300-355 μm size fraction of the coarse (> 63 μm) fraction of washed sediments from ODP 999. Foraminiferal samples were cracked and cleaned in the boron isotope clean lab at the University of Southampton. The cracking was done between 2 glass slides under a microscope and care was taken to open all major chambers in order to allow for effective clay removal. The cleaning followed established methods for oxidative cleaning (40-42). Following cleaning, *G. ruber* samples were dissolved in weak (~0.15 M) Teflon-distilled nitric acid and separated into two fractions: an “isotope” fraction (90 %) and “trace element” fraction (8-10 % of sample volume). Boron was separated from the matrix of the isotope fraction with anion exchange resin (Amberlite IRA-743) in purpose-built columns (18,43).

Boron isotope ratios were measured at the University of Southampton on a Thermo Scientific Neptune multicollector inductively coupled plasma mass spectrometer (MC-ICPMS) following established methodology (18, 36 , 42, 43). External reproducibility was calculated based on the reproducibility of an in-house standard (Japanese coralline *Porites* – JCP-1 = 24.2‰ (see 42)) described by equation 1 below where [11B] = intensity in volts of 11B; for this study the 2σ uncertainty typically corresponds to ~0.2 ‰.

Equation 1:

The trace element fraction was diluted and Me/Ca ratios of Li, B, Na, Mg, Al, Mn, Fe, Sr, Cd, Ba, Nd, U were measured on a Thermo Scientific Element 2-XR ICPMS at the University of Southampton. Over the period of this study reproducibility of 3 internal standards with a range of Me/Ca ratios was 4 % for Mg/Ca and 5 % for Al/Ca. We use the trace element data to screen for diagenesis and other potential artifacts. Samples with Al/Ca > 100 μmol/mol were strictly removed to preserve data quality, and we find no anomalies in other element ratios (e.g. Mn/Ca, Ba/Ca, and Fe/Ca).

Mg/Ca temperatures were calculated following the approach of Evans and Müller (2012) ref. (44) and using a Mg/Ca of seawater (Mg/Casw) that was calculated from the modeled study of Fantle and DePaolo (2006, ref. 45). We use an H value of 0.41, though species-specific to *Trilobatus (formerly Globigerinoides) sacculifer* (46), as no calibrated H value for *G. ruber* is currently available. Note that that the accuracy of this reconstruction is not crucial to our findings, since SST has a relatively weak effect on calculated CO2 (see Supplement section 2.3), and also the potential change in Mg/Casw over the last million years is small due to the long residence times of the elements involved.

**2.2 Determination of pH from δ11B of *G. ruber***

Boron is present in two principal forms in seawater: boric acid (B(OH)3) and borate ion (B(OH)4−). The relative proportion of these two species depends on pH and the dissociation constant for boric acid, pK\*B:

Equation 2:

An isotopic fractionation between the two stable isotopes of boron-11 (~80 % abundance) and boron-10 (~20 %) is also associated with this equilibrium (above) as the bond strengths between boron and the hydroxyl ion in boric acid and the borate ion differ. 11B concentrates in the more strongly bonded boric acid giving it a higher δ11B than the borate ion by ~27.2 ‰ (47). The delta notation (δ11B) is used to express differences in boron isotope ratios:

Equation 3:

where 11B/10Bref is the isotopic ratio of NIST SRM 951 boric acid standard (11B/10B = 4.04367) ref. (48), and 11B/10Bsmp is the isotopic ratio of the sample. The basis of the boron isotope-pH proxy is that, because the borate ion is tetrahedral and charged, it is more readily substituted into CaCO3 (49, 50). Therefore, we can solve for pH based on the reconstructed isotopic composition of δ11Bborate:

Equation 4:

where pK\*B is the dissociation constant for boric acid at reconstructed in situ temperature, salinity and pressure (51)*,* δ11Bsw is the isotopic composition of seawater (39.61 ‰ ref. 52), δ11Bborate is the isotopic composition of borate ion, and αis the isotopic fractionation factor between the two aqueous species of boron in seawater (α=1.0272, equivalent to an equilibrium isotope fractionation of ~27.2 ‰ ref. (47)). Boron is well mixed in the oceans with a residence time of 10-20 Myr; to account for likely (small) changes in the boron isotopic composition of seawater (δ11Bsw), we use a simple linear extrapolation with a central value at modern δ11Bsw (39.61±0.1 ‰ ref. 52) with uncertainty increasing by ±0.12 ‰ by 1200 kyr (0.1 ‰ per million years). This central value is consistent with an independent constraint (54) and any possible changes (55) are well accounted for by the generous uncertainties we ascribe to all parameters.

As δ11Bruber (the δ11B of boron in *G. ruber* calcite)is offset from δ11Bborate it is necessary to first account for this relatively minor deviation in order to calculate pH from δ11Bruber, and a size-specific core-top, culture, and field measurement calibration (Equation 5) is applied for δ11Bborate reconstruction (33,39). This relationship has been shown in other studies to provide for accurate atmospheric CO2 reconstructions (11, 18, 19). Uncertainties in this calibration (*G. ruber* 300-355μm) are propagated through all relevant calculations (shown below at 2σ).

Equation 5:

**2.3 Determination of CO2 from δ11B-derived pH**

In order to calculate aqueous CO2 from pH a second carbonate-system parameter is required (34). Here we briefly outline two approaches to estimate total alkalinity (ALK) in order to derive CO2 from our pH data: (a) ALK remained within a specified range of its modern value, and (b) ALK change correlates with reconstructed pH change. Below we describe why both approaches yield very similar results, and we describe a sensitivity test to demonstrate that our main conclusions are not affected by the choice of how to estimate ALK. In the main text and our tabulated dataset, we assume ALK has remained within a generous range around its modern value (~2330±175 µmol/kg) over the investigated period. The outcome of our own model inversion of CO2 change of the last 1,500 kyr (see Supplement section 4.3 for details on the model) suggests a total ALK range of only 130 µmol/kg, much less than the 350 µmol/kg uncertainty range propagated here. This means we fully explore the likely ALK range based on modeling of whole ocean ALK variability on G-IG timescales and over the last 1,500 kyr (11, 22). This treatment of the second carbonate system parameter essentially assumes that any change in ALK will drive the majority of its impact on CO2 by changing pH (which we reconstruct directly) rather than by changing the abundance of carbonate and bicarbonate ions. This assertion is based on first-principle carbon chemistry and can be verified by comparing reconstructed pH with the ice-core CO2 record (Fig. S3a). When we cross-plot δ11B-based pH against ice-core CO2, we find that the data are well described by the theoretical pH-to-CO2 relationships, assuming either constant DIC or constant ALK (Fig. S3a), in support of our assertion. The alternative view is that much of the reconstructed pH change was caused by changes in ALK, such that ALK and pH are correlated (i.e., the ~0.2 pH increase during the Last Glacial Maximum corresponds to +100 µmol/kg ALK increase; e.g., see refs. 22 and 56). We carry out a sensitivity test to estimate how much difference in terms of CO2 it makes to account for this correlation (Fig. S3b). We find that including the ALK-to-pH correlation reduces our estimate for additional glacial-stage CO2 reduction by about 5 µatm, and it reduces our estimated increase of the G-IG CO2 range by about 5 µatm (Fig. S3c). As described above, this relatively modest change is due to the fact that ALK change causes most of its impact on CO2 by changing surface pH, the parameter that we reconstruct (i.e., Fig. S3b). We argue that the modest difference between the two ALK assumptions does not change the main conclusions of this study, especially given the agreement between our new data, previously published fully independent CO2 reconstructions and our carbon cycle modeling results (Fig. 4, Fig. S4). Since the ALK-to-pH correlation is itself uncertain we prefer the calculation using a large and constant ALK uncertainty range that more than covers existing estimates of G-IG ALK change (e.g., refs. 22 and 56). With this approach, the small deviation between the ALK assumptions is effectively covered by our uncertainty propagation and there is little independent evidence to suggest that this assertion is invalid (57, 58).

Using δ11B-derived pH and the prescribed ALK and modern aqueous CO2 disequilibrium, uncertainties are propagated via a Monte Carlo simulation (n = 10,000) in the statistical analysis program R (59)*,* and 95 % confidence intervals are calculated from the variation within these simulations. Two standard deviation uncertainties on the individual input variables are typically: δ11B (± analytical uncertainty defined by equation 1, normal distribution), Mg/Ca-derived temperature (± 3 ºC, normal distribution), salinity (± 3 psu, normal distribution), ALK (± 175 μmol/kg, uniform distribution), δ11Bsw (39.61 ± 0.10-0.22 ‰, normal distribution ref. (52)).Atmospheric CO2 was then calculated from aqueous CO2 using Henry’s Law, subtracting the modern extent of disequilibria with respect to CO2atm at the site (given a nominal uncertainty at 2σ of ±50 % (i.e., ±10 μatm)). Uncertainties in CO2 also include propagation of the uncertainties in the δ11B calibration of *G. ruber* (Equation 5). All subsequent carbonate system calculations in R were performed using the seacarb package (59, 60)*.*

**3. Forcing-to-SL relationship**

To extract the underlying relationship between two empirical datasets requires careful consideration of uncertainty in both X and Y coordinates (61). In our case the X coordinate is CO2 climate forcing, ∆RCO2, which is calculated directly from reconstructed CO2 following ref. 33:

Equation 6:

We calculate mean and standard deviation of ∆RCO2 as the half-point and quarter-width of the ±2σ interval of our CO2 reconstruction. As for the Y coordinate, sea-level (SL) in our case, we point out that there are three sources of uncertainty: (1) the inherent uncertainty of the SL estimate documented alongside each SL record, (2) the range of SL within the age window defined by the relative age uncertainty of the SL and the CO2 age models, and (3) the systematic discrepancies between different SL records. The latter point we address by carrying out our analyses independently for a number of recent SL records (3, 4, 20, 62), and we find only modest differences suggesting that our conclusions are robust to these systematic SL discrepancies. To account for both (a) and (b), for each CO2 data point we estimate SL and its uncertainty by constructing the cumulative distribution function (CDF) of all SL data that fall within the relative age uncertainty window (which we take to be 1.5-2 kyr for our LP260 data, an uncertainty ramps from 2 kyr to 4 kyr over the course of the 800 kyr compilation of ice core CO2 reconstructions, and 6 kyr for our eMPT data). For each individual CDF we estimate the appropriate SL mean and standard deviation as the half-point and half-width of the 16%-to-84% interval (corresponding to ±1σ). This procedure yields typical individual data point SL uncertainties about 20 meters (±2σ), and significantly larger uncertainties for a few points that happen to correspond to times of rapid SL change.

As for the regression of the underlying CO2 forcing-to-SL relationship we make use of the approach laid out by York et al. (61), a generalized form of “reduced major axis” regression method that takes full account of normal error in X and Y. We carry out this regression technique using four independent SL reconstructions (3, 4, 20, 62), both for our LP260 CO2 dataset (Fig.S7a) and our eMPT dataset (Fig.7d). In all cases we display the regression confidence interval. The confidence intervals of the regressed slopes for LP260 (Fig.7b1) and eMPT (Fig.7b3) are of particular interest. We consistently find the CO2 forcing-to-SL slope to be significantly steeper during LP260 (about 40-50 meters per Wm-2) than during eMPT (less than 25 meters per Wm-2), supporting the main line of argument in our study.

To build confidence in our regression method we offer three tests: (i) Are the regressions based on our LP260 dataset consistent with regressions based on the compilation of the continuous ice core CO2 record of the last 800 kyr? (ii) Does preindustrial SL fall within the prediction interval of our regressed CO2 forcing-to-SL relationships? and (iii) Is the York et al. regression method skillful in extracting the correct CO2 forcing-to-SL relationships when put to the test with synthetic data that mimics our eMPT data and its uncertainties?

To address (i) we carry out the regression using the ice core CO2 record against all SL records based on a 200-kyr sliding window of data. Because the sampling interval of the ice core CO2 compilation changes dramatically over its 800 kyr range (and in particular over the course of the last deglaciation) we regularize the CO2 data by binning it into ±1 kyr increments and estimate bin mean and standard deviation analogous to the SL data. We find that throughout the last 800 kyr the CO2 forcing-to-SL slopes regressed in that way generally fall within a band of 35 to 55 meters per Wm-2 (grey shading in Fig.S7b2), in good agreement without LP260 data and significantly greater than for our eMPT data. We note that the slope regressed from the SL record that is based on Mg/Ca-δ18O deconvolution (3) exhibits a long-term oscillation not seen in any of the other SL regressions, but even at its lowest point it is still significantly larger than that regressed during the eMPT. Likewise, while the Mediterranean SL reconstruction (4) falls into a constant narrow range for most of the 800 kyr, its regressed CO2 forcing-to-SL slope yields anomalous variability during a brief interval around 500 kyr where the SL record is known to be overprinted by a series of sapropel layers. The regression based on the Red Sea SL record (21), arguably the best constrained SL reconstruction, is remarkably constant over the entire 500 kyr period it spans, as is the case for the entire last 800 kyr for the regression based on simulated SL by de Boer et al. (62). Based on that analysis we conclude that there is no evidence for substantial changes in the CO2 forcing-to-SL slope over the entire 800 kyr duration of the ice core CO2 record, and that the significantly lower regressed slopes for our eMPT dataset therefore speak to a change in the CO2 forcing-to-SL relationship associated with the MPT. Further, we note that in Fig. 2 of the main manuscript we only present the regression results of the Mediterranean SL record, which this supplementary analysis suggests to be both robust and the most conservative record covering both the LP260 and eMPT intervals. We have further confidence in our conclusion of a MPT reduction in the CO2 forcing-to-SL slope because we find an equivalent change when considering the LR04 benthic foraminifera δ18O stack instead of SL (Fig.S8), which is observationally well constrained and carries a significant component SL signal.

To address (ii), the test for SL predicted for preindustrial CO2 forcing, we replicate our analysis of the CO2 forcing-to-SL slope also for the intercept regression parameter, shown in analogous way in Fig.S7c. When using the regressed CO2 forcing-to-SL relationship, however, it is the prediction interval (dashed in Fig.S7c) and not the regression confidence interval that is relevant because the CO2 forcing-to-SL regression does not embody residual SL change caused by factors such as orbital change. Hence the prediction interval at the intercept is wider than the intercept confidence interval, and it comfortably includes preindustrial SL at preindustrial CO2 forcing during LP260.

Finally, to address (iii) the question as to the skill of York regression to extract the correct CO2 forcing-to-SL slope even in the face of substantial X and Y uncertainty of individual points and an overall low signal-to-noise ratio, we construct a synthetic test that mimics the signal-to-noise ratio of our eMPT data (Fig.S9). That is, we presuppose a known CO2 forcing-to-SL relationship and generate randomized data points based on the typical uncertainties of CO2 forcing and SL. To carry out this analysis we change the number of synthetic data points used for the regression and find that: (a) the true slope of the known synthetic relationship falls within the regressed slope confidence interval, (b) the width of the regressed slope confidence interval systematically narrows as the number of data points increases, and (c) at 50 data points the slope confidence interval is narrow enough to detect the change in slope we reconstruct between LP260 and eMPT.

**4. Carbon cycle modeling**

**4.1 Model details**

To simulate the global carbon cycle and atmospheric CO2 levels across the MPT, we make use of the recently updated CYCLOPS model (23). The model represents the oceanic carbon cycle and physical circulation as the exchange between 18 separate surface, mid-depth and deep-water reservoirs, as well as carbon fluxes between surface-water reservoirs and an atmospheric reservoir. Following previous work (63,64), the model simulates the open system CaCO3 cycle through explicit representation of undersaturation-driven seafloor dissolution of biogenic CaCO3 rain that originates from the surface.

To simulate global carbon-cycle variations in the Pleistocene, we apply three fully separate forcings to the model, all of which represent well-established modes of carbon-cycle and circulation change previously investigated using CYCLOPS: (a) major nutrient drawdown driven by glacial-stage iron-fertilization of the Subantarctic Zone of the Southern Ocean (8, 65); (b) coupled glacial-stage reduction in vertical exchange, export production and residual surface nutrient status of the Polar Antarctic Zone of the Southern Ocean (56,66-68); and (c) glacial-stage shoaling of the Atlantic Meridional Overturning circulation (69-72). The rationale and evidence base for these changes have been reviewed elsewhere (73) and the model’s atmospheric CO2 sensitivity to these mechanisms has been evaluated in detail (23). To represent the time-evolution in model forcing, we use ODP 1090 FeMAR (1), ODP 1094 Ba/Fe (25) and ODP 982/U1313 (26, 27) observational records to adjust model conditions in the Subantarctic Zone, Polar Antarctic Zone and the global deep-water circulation pattern, respectively. Other drivers of CO2 change (e.g., silicate-weathering feedback, temperature) are not considered here but will be investigated in the future. Similarly, by implementing Atlantic circulation changes based on North Atlantic carbon isotope gradients our model does not reflect exceptional circulation weakening during MIS 23 inferred from neodymium isotopes (74) or any other modes of Atlantic meridional overturning circulation reorganization.

**4.2 Extension of ODP 1094 Ba/Fe proxy forcing**

The previously published ODP 1094 Ba/Fe record from 0–1 million years ago was extended to allow us to start the forced simulation as early as 1.5 Myr. The composite section for the deeper part of ODP core 1094 (>121 mcd; 1–1.5 Myr) was mainly reconstructed using magnetic susceptibility (75), and Fe counts derived from XRF scanning on sediments of Hole A and D, and the δ18O of the planktonic foraminiferal species *Neogloboquadrina pachyderma*. The age model of the core is based on tuning the δ18O of the benthic foraminifer *Cibicidoides wuellerstorfi* to the LR04 benthic stack (26), as will be documented elsewhere. Oxygen isotope analyses were performed with a Thermo GasBench II coupled to a Thermo Delta V Plus mass spectrometer at the Geological Institute, ETH Zurich (1σ error of 0.07 ‰). The relative sedimentary concentrations of Ba and Fe were acquired with an AVAATECH profiling X-Ray Fluorescence (XRF) core scanner at the MARUM, University of Bremen, using the same settings and procedures as in ref. 25, but with a different detector (Canberra X-PIPS Silicon Drift Detector, Model SXD 15C-150-500). To assure that the XRF data are consistent throughout the entire record, four sections between 0 and 1 Myr have been rescanned. The elemental XRF data of the rescanned sediments are correlated to each other, and their linear fit is used to scale newly acquired data accordingly (r = 0.97 for Fe, r = 0.77 for Ba).

**4.3 Model inversion**

The relationship between the observational records and the forcing applied to the model (Subantarctic nutrient status, Polar Antarctic/Deep Southern Ocean exchange, Polar Antarctic surface nutrient status, and Atlantic overturning) is represented as four equations with a total of seven free parameters (i.e., intercept and slope, threshold in the case of Atlantic overturning). The Atlantic circulation is simulated as ‘shallow’ only if the reconstructed vertical stable carbon isotope gradient in the deep North Atlantic increases above a threshold value determined as part of the inversion; all other forcing functions are continuous and monotonic. For the Antarctic changes, vertical exchange scales with Ba/Fe while surface nutrient status scales with the square root of Ba/Fe, thereby in effect assuming a sub-linear scaling between Ba/Fe and export production (i.e., the property arguably recorded by Ba/Fe at ODP 1094 (25)). The seven forcing-function parameters are initially set to arbitrary (but sensible) values.

We invert the model by finding the specific combination of the seven forcing-function parameters that minimize the root-mean-square (RMS) error of simulated atmospheric CO2, as evaluated against a composite reconstruction of atmospheric CO2 from Antarctic ice cores (14). That is, given the observational forcing records, the forward model and the CO2 record, we invert to obtain the model forcing parameters that minimize the RMS objective function. In order to be confident in the outcome of the minimization procedure across seven-dimensional parameter space, we incorporate into the model two fully independent methods and verify that they converge onto the same solution: (a) Powell’s conjugate direction method (76-78) and (b) the Fletcher-Reeves implementation of the Steepest Descent conjugate gradient method (79). Side conditions relating to physically possible parameter solutions are encoded directly into the model and both algorithms converge to the same solution at a RMS of 12.2 µatm.

**5. Quantification of GCO2 , IGCO2 and IG-G∆CO2**

Given our two δ11B-based CO2 reconstructions (eMPT, LP260), we wish to (a) quantify by how much glacial and interglacial CO2 levels (i.e., GCO2 and IGCO2 , respectively) are different across the MPT and by how much the magnitude of glacial-interglacial CO2 cycles (i.e., IG-G∆CO2) has changed, and (b) compare our new estimates to equivalent estimates derived from existing CO2 records (11, 17, 14) as well as from our model-derived predictions of CO2 (see supplement section 4). This section describes the details of the analysis and offers discussion of the rationale behind these calculations. Our analysis of the CO2 datasets consists of four steps: (a) define which individual data points represent glacial and which data points represent interglacial conditions, (b) estimate the average CO2 and its uncertainty from the two sets (i.e., GCO2 and IGCO2) of CO2 data, (c) estimate the glacial-interglacial CO2 difference (i.e., IG-G∆CO2) before and after the MPT, and (d) test whether the resulting changes are statistically significant.

In order to define which individual data points contribute, we sub-sample each record based on a percentile cut-off criterion, using for example only the 25 % “most interglacial” and the 25 % “most glacial” data points in the quantification of GCO2, IGCO2 and IG-G∆CO2. In the case of the δ11B-based CO2 reconstructions, we can rank the individual data points based on δ18O measured on the same sample to directly reflect climate state, but in the case of the direct ice-core measurements and our model-simulated CO2, we have to rank the individual data points based on their CO2 level. The value of 25 % for the cut-off criterion is arbitrary and therefore we carry out and present the analysis for the full 1 % to 50 % range of non-overlapping sub-samples (Fig S4). A low percentile cut-off implies that only the few most extreme data points are sub-sampled as glacial and interglacial, which maximizes the glacial/interglacial difference but also leads to a relatively large uncertainty because only few data contribute to the estimates. Conversely, a high percentile cut-off implies that more individual data contribute to the estimation of CO2, which reduces the uncertainty and increases robustness, but also progressively averages away the end-member changes we try to quantify. For brevity, we make the trade-off between robustness and sensitivity, and exclusively discuss the 25 % percentile cut-off in the main text, and we note that our qualitative results are independent of that choice (Fig. S4).

The second step of our analysis is to estimate GCO2 and IGCO2 based on their respective sub-samples (see above paragraphs) and to also propagate the uncertainty of the individual data points. To this end, we first sum up and normalize the probability density functions of the individual data points and their normally distributed uncertainty. We assign 1σ uncertainties of 20 µatm to our CO2 data, 14 µatm to the Hönisch et al data (ref. 11, as published; we note that ref. 11 does not include a number of sources of uncertainty that are included in the uncertainty calculations for our data), a nominal 10 µatm to ice-core data, and 12.2 µatm to simulated CO2 based on model-inversion residual RMS. From these cumulative sub-sample probability density functions we estimate the central tendency as the median probability and normal dispersion as half of the central 66 % cumulative probability interval. Thus, the dispersion of our estimates represents a combination of inherent measurement uncertainty associated with individual data points and of the true spread in CO2 levels within the sub-sample. To estimate IG-G∆CO2 we simply subtract the central estimate of GCO2 from IGCO2, whereby the normal dispersion of IG-G∆CO2 is propagated as the root-sum-of-squares of the normal dispersion estimates of GCO2 from IGCO2.

The final step in our analysis is to evaluate the significance of the differences we find in glacial and interglacial CO2 and glacial/interglacial CO2 range before and after the MPT: i.e., ∆GCO2, ∆IGCO2 and ∆IG-G∆CO2 as also visualized in Figure 4 of the main text. To this end, we subtract eMPT GCO2, IGCO2 and IG-G∆CO2 from the corresponding LP260 estimates, again propagating the estimate dispersion as the root-sum-of-squares of the individual normal dispersion estimates. For all four CO2 datasets (i.e., two δ11B-based CO2 records, ice-core data, and model-inversion results) we find that estimated ∆IGCO2 falls onto zero within its 1σ dispersion (thick black whisker), and thus we cannot reject the null hypothesis that interglacial CO2 levels were identical during our intervals (eMPT and LP260). Conversely, we find that estimated ∆GCO2 is negative for all four CO2 datasets, with the deviation from zero exceeding the threshold of 1.64 σ for one-sided testing (thin black whisker in Figure 4). That is, we can reject at 95 % confidence level that glacial-stage CO2 prior to the MPT was as low or lower than after the MPT. Likewise, and largely driven by this significant glacial-stage CO2 decline across the MPT, the estimated MPT change in the G-IG CO2 range is significantly greater than zero for all four datasets (at 95 % confidence level for one-sided testing; thin black whisker in Figure 4). Thus, we can reject at 95 % confidence level the null hypothesis that the magnitude of glacial/interglacial CO2 change prior to the MPT was as large (or larger) as the glacial/interglacial CO2 change following the MPT.

Additional results of statistical analysis are tabulated in Table S1.

**6. Supplementary figure captions:**

**Figure S1 | Map of sediment core sample sites.** Colour scale represents CO2 disequilibrium between surface-water and air, with surface-water supersaturation indicated by red tones (35). The locations of ODP 999 and ODP 668 are indicated. Also indicated are the locations of cores used to inform our modelling results in green, ODP 982, IODP U1313, ODP 1090

and ODP 1093.

**Figure S2 | Foraminiferal Mg/Ca and oxygen isotope measurements from ODP 999.** a) Planktic Mg/Ca temperature estimated from Globigerinoides ruber (LP260 = blue, eMPT = red), temperatures calculated as stated in Supplement section 2.1, also shown is the previous record of Mg/Ca temperature from this Site (grey, 39). b) Planktic *Globigerinoides ruber* (blue) and benthic *Cibicides wuellerstorfi* (red) oxygen isotope results plotted *versus* the global benthic foraminiferal oxygen isotope stack (black) ref 26.

**Figure S3 | Assessment of alkalinity assumption.** Our treatment of the second carbonate-system parameter asserts that any change in alkalinity will cause most of its CO2 impact by changing pH, which we reconstruct directly, rather than by the change in the abundance of carbonate and bicarbonate ion. (a) To test this assertion we cross-plot reconstructed pH and ice-core CO2 and assess the relationship between these empirical data against our assertion. Blue and purple dots are our LP260 CO2 data and CO2 data from ref. 11, respectively, with the corresponding CO2 interpolated from the ice-core CO2  record (14). The three solid lines represent the theoretical pH-CO2 relationships under the following assumptions: (red) constant DIC, (black) constant alkalinity, and (green) DIC and alkalinity both change with a set ratio of 1:1. Either assumption of constant DIC or constant alkalinity gives faithful reproduction of ice-core CO2 from the Site 999 and 668 δ11B-pH records, in strong support of our assertion. The CO2 uncertainty in the empirical data relates to the age-model uncertainty (~2 kyrs) of the pH data-points and the resulting spread in interpolated ice-core CO2. (b) To illustrate our ALK assumption further, we show CO2 as a function of ALK and pH at 25˚C and salinity of 35, as predicted yielding nearly vertical contours that demonstrate a strong sensitivity of CO2 to pH and a weak sensitivity to ALK. As a point of reference we highlight pH of 8.2±0.15, 8.3±0.15 and 8.4±0.15 to represent interglacial pH both during eMPT and LP260, eMPT glacial pH and LP260 glacial pH, respectively. The large ALK uncertainty range that we propagate in our calculations (grey shading) more than covers estimates of ALK change over recent G-IG cycles (i.e., dashed line connecting IG with LGM). (c) From the sensitivity calculations we estimate the CO2 offset between two plausible treatments of ALK uncertainty: (solid lines) assuming ALK remained within constant range (i.e., ±175 µmol/kg, grey shading), and (dashed lines) assuming an ALK-to-pH correlation of +100 µmol/kg per 0.2 pH increase in addition to the ±175 µmol/kg ALK uncertainty. We find that the ALK-to-pH correlation causes a modest mean increase of ~5 ppm and ~8 ppm in the estimates for eMPT and LP260 glacial CO2 levels, respectively. The implied ~3ppm reduction in the difference between glacial stage eMPT and LP260 CO2 is too small to undermine the main conclusions of our study.

**Figure S4 | Quantification of CO2 change since the MPT.** As described in detail in Methods section 5, we quantify (top) interglacial CO2, (middle) glacial CO2, and (bottom) interglacial-minus-glacial CO2 for both data from the latest late Pleistocene (blue) and during or prior to the MPT (red). From left to right, quantification is carried out on four different datasets: boron isotope data from ODP999 (this study) and ODP668B (11), CO2 directly measured on air trapped in Antarctic ice using the compilation of continuous ice core CO2 reconstructions (14) and stratigraphically disturbed ~1-million-year-old “blue-ice” from the Allan Hills (17), and CYCLOPS model inversion (this study). To define glacial and interglacial subsets of the datasets we use on a cut-off criterion, sub-sampling the data with a given percentile of lowest/highest δ18O (marine records) or CO2 (ice core, model). For this figure we carry out this analysis while systematically changing the value of the critical percentiles (x-axis). The results discussed in the main manuscript and shown in Figure 4 correspond to a 25% cut-off percentile, but as shown here our conclusions are robust across a wide range of percentiles. The underlying data are compiled for percentile increments of 5% in Table S1.

**Figure S5| Overview of CYCLOPS model inversion results.** Top panel shows in black the composite ice-core CO2 record (14) used as the inversion target and in blue the model inversion result, with dark and light shading indicating ±1x and ±2x the root-mean-square error of the inversion (12.3 µatm). The middle three panels show the three palaeoceanographic proxy records used to drive the inversion, as described in detail in Supplement section 4. The bottom panel illustrates the contribution of dust-borne iron fertilization of the Subantarctic Zone of the Southern Ocean, highlighting the simulated CO2 change that can be attributed to that forcing.

**Figure S6 | Spectral analysis of various palaeoclimatological records.** Ice-core CO2 levels (14), benthic foraminiferal oxygen isotope stack (26), reconstructed sea-level change (4), our new boron-isotope-derived CO2 reconstruction, and CO2 calculated by inversion of the CYCLOPS carbon-cycle model (as described in Supplement section 4). Spectra on the left are derived from data down-sampled to the temporal resolution of the oxygen isotope stack, and spectra on the right are derived from data down-sampled to the temporal resolution of our boron-isotope-derived CO2 reconstruction. Evaluating the datasets at identical age points yields identical critical significance levels (horizontal dashed lines), shown here for 0.5, 0.9 and 0.99 for all panels. We use the Lomb-Scargle method for frequency analysis of unequally spaced data, as implemented in the “plomb” Matlab function.

**Figure S7 | Regression of relative sea-level change and CO2 radiative forcing.** Regression of our (a) LP260 and (d) eMPT data against different sea-level records (color coded). The error bars indicate ±2σ uncertainties for individual data points, and the light/dark envelope indicates the 1σ/2σ confidence interval of the regressions. The (b) slope and (c) intercept of the LP260 and eMPT regressions can be compared against equivalent regression results using a 200-kyr sliding window of the continuous ice core CO2 record (14), which yield relatively high slope values and intercepts around 0 meters. Slopes and intercepts based on ice core CO2 are consistent with the regression results for our LP260 data, whereas the slope regressed from our eMPT data is consistently lower. All regressions are based on the method by York et al. (61), which fully takes into account uncertainty in X and Y. For more discussion of methodology and uncertainty treatment see Supplementary Information section 3. Dashed lines in (c) indicate prediction interval.

**Figure S8 | Regression of δ18O and δ11B records.** Panels a) and b) display the regression results of our LP260 and eMPT boron isotope datasets against the LR04 benthic foraminifera oxygen isotope stack (see Figure S7 for details). Panels c) and d) compare the results of the regression slope and intercept, respectively, against equivalent analysis using a 200-kyr sliding window based on the ice core CO2 record (middle panels). The slopes mimic those seen in the equivalent sea level analyses.

**Figure S9 | Synthetic test of York regression method.** Test of regression robustness based on synthetic data that mimics our eMPT dataset. Noise is added to a known relationship of SL and CO₂ radiative forcing (Y=25X, shown as the fat black line) before the York regression is applied. We test the performance of the regression as a function of how many data points are contained in the dataset and find that the 51 samples in our eMPT dataset should be sufficient to recover the true CO₂-forcing-to-SL relationship with the required accuracy.