Supplemental Materials for "Robust A-type order and spin-flop transition on the surface of the antiferromagnetic topological insulator $MnBi_2Te_4$"

Paul M. Sass[1], Jinwoong Kim[1], David Vanderbilt[1], Jiaqiang Yan[2], and Weida Wu[1]*.

[1]*Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA.*

[2]*Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.*

*Correspondence to: wdwu@physics.rutgers.edu (WW).

Table of content

Methods

Sample preparation. Platelike single crystals of $MnBi_2Te_4$ were grown out of a Bi-Te flux and have been well characterized by measuring the magnetic and transport properties. They order magnetically below T_N = 24 K with ferromagnetic Mn-Te layers coupled antiferromagnetically. At 2 K, $MnBi_2Te_4$ shows a spin-flop transition at $\mu_0 H_{SF} = 3.5\,\mathrm{T}$ w followed by moment saturation at saturation transition at $\mu_0 H_S = 7.8\,\mathrm{T}$ with a magnetic field applied along the crystallographic *c*-axis.

MFM measurement. The MFM experiments were carried out in a homemade cryogenic magnetic force microscope using commercial piezoresistive cantilevers (spring constant $k \approx 3$ N/m, resonant frequency $f_0 \approx 42$ kHz). The homemade MFM is interfaced with a Nanonis SPM Controller (SPECS) and a commercial phase-lock loop (SPECS). MFM tips were prepared by depositing

nominally 150 nm Co film onto bare tips using e-beam evaporation. MFM images were taken in a constant height mode with the scanning plane nominally ~100 nm (except specified) above the sample surface. The MFM signal, the change of cantilever resonant frequency, is proportional to out-of-plane stray field gradient. Electrostatic interaction was minimized by nulling the tip-surface contact potential difference. Dark (bright) regions in MFM images represent attractive (repulsive) magnetization, where magnetizations are parallel (anti-parallel) with the positive external field.

Simulation. The numerical simulations were performed with the revised Mills model,

$$\begin{aligned} E = \sum_{i=1}^{N-1} J_i \mathbf{s}_i \cdot \mathbf{s}_{i+1} \quad + \quad \boldsymbol{H} \cdot \sum_{i=1}^{N} \mathbf{s}_i + \sum_{i=1}^{N} \frac{K_i}{2} (\mathbf{s}_i \cdot \hat{\mathbf{z}})^2 , \\ \mathbf{s}_i \quad \rightarrow \quad \mathbf{s}_i + \delta_{i=1,N}(\lambda_s - 1)\mathbf{s}_i, \\ J_i \quad = \quad J + \delta_{i=1,N-1}(\lambda_J - 1)J, \\ K_i \quad = \quad K + \delta_{i=1,N}(\lambda_K - 1)K, \end{aligned}$$

where $\lambda_A, (A = s, J, K)$ represents the reduction of surface magnetization, exchange coupling, and anisotropy energy, respectively. The reduced surface magnetization causes a pinning of the spin-flop state at the surface [34]. One end of the system described by a "clamped boundary condition" in which the spin constrained to that of the bulk,

$$\theta_{i=N} = \theta_{Bulk} = \begin{cases} 0, \wedge H_{FM} < H \\ \pm cos^{-1} \dfrac{H}{H_{FM}}, \wedge H_{BSF} < H \leq H_{FM}, \\ (\pm 1 - 1)\dfrac{\pi}{2}, \wedge H \leq H_{BSF} \end{cases}$$

where $H_{FM} = 4J - K$ and $H_{BSF} = \sqrt{K(4J - K)}$ are threshold fields for the bulk forced ferromagnetic and bulk spin-flop transitions respectively, the anisotropy $K < 2J$ is assumed to be small, and the sign indicates two types of surfaces: + for parallel and – for antiparallel.

The phase diagram of a semi-infinite system is obtained by using of revised Mills model with the one-side-clamped boundary condition for $N = 16$ layers with a parameter set of ($K/J = 0.4, \lambda_s = 0.6$, $\lambda_J = 0.8$, and $\lambda_K = 0.6$). The ground state at each sampling point is searched by comparing total energies of spin configurations relaxed from 200 initial random configurations. Two SSF states in the phase diagram are illustrated by comparing the net spin canting of two surfaces, which is defined as $C_{A,P} \equiv 1/N \sum_{i=1}^{N} sin^2\left(\theta_i^{A,P}\right)$ for antiparallel (A) and parallel (P) surfaces. The difference $C_A - C_P$ vanishes if both surfaces are in the bulk spin-flop (BSF) state or collinear states, and remains finite only when one surface is in the SSF state.

Simulation of the MFM contrast. In order to compare the simulation result of spin structure with experiment, the spatial gradient of magnetic force acting on the MFM tip is calculated with the simulated spin configurations with respect to the applied field for a given set of parameters. For simplicity, we approximate the magnetization of septuple layers with series of point magnetic dipole moments which are vertically aligned below the tip. Thus, the force gradient is calculated as $\partial F = \sum_{i=1}^{1000} cos(\theta_i)/\left(d + t(i-1)\right)^5$, where $d = 150$ nm is the tip-surface distance, $t = 13.6$ Å is the thickness of one septuple layer, and θ_i is a zenith angle of the *i*-th layer. For $i \geq N$, the angles are constrained or assumed to be that of bulk. The point dipoles approximation of septuple layers provides a qualitative description of the magnetic field dependence of MFM signal of domain contrast, though the oversimplified assumption prevents a quantitative account.

Note 1: Asymmetry in step edge contrast due to difference between forward and backward scan directions.

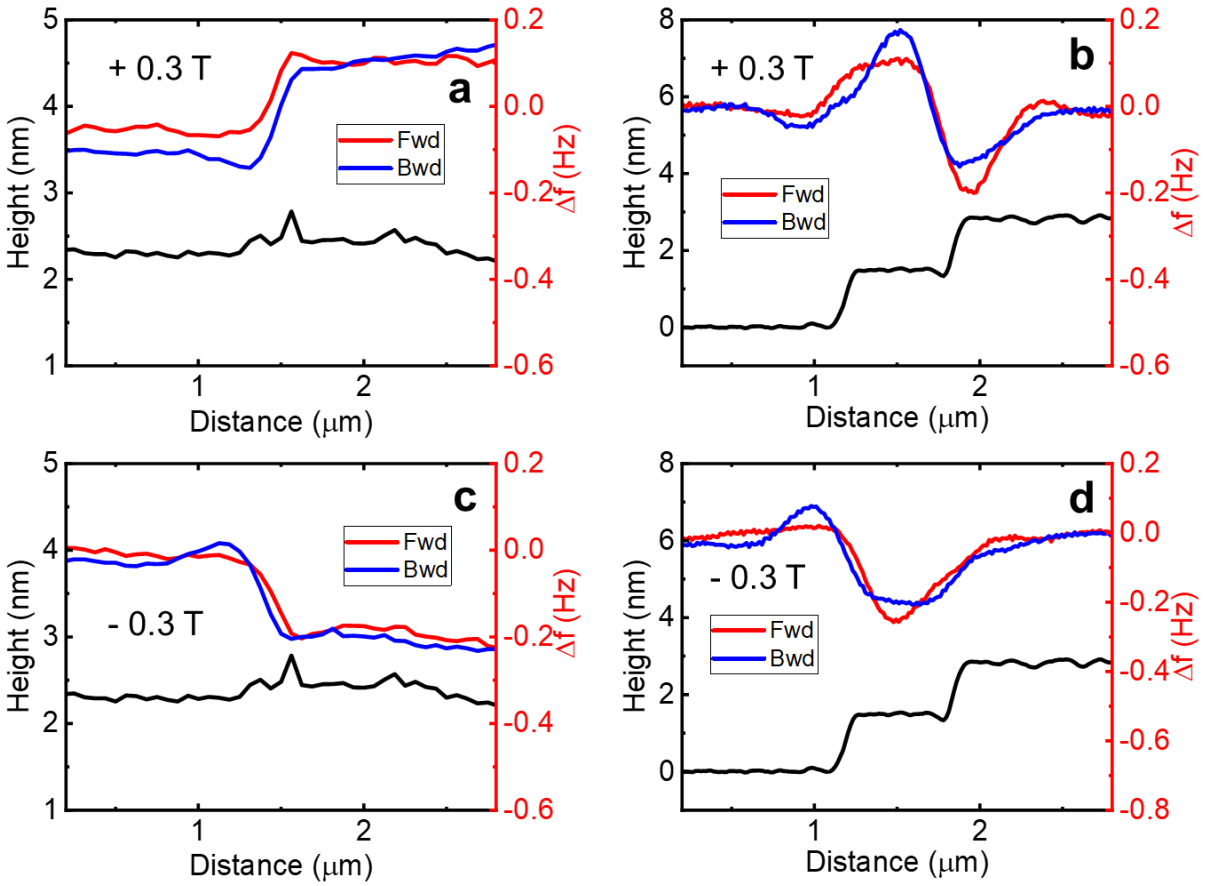

Fig. S1 | Comparison of forward and backward line profiles of topographic and MFM images. a, b, Line profiles of topographic and MFM images from Fig. 1 showing both forward and backward scanning directions at + 0.3 T. **c, d,** Line profiles in the same location at - 0.3 T. The forward and backward lines in **b** and **d** are different. The signal dips more on the right (left) side for the forward (backward) lines. This is due to the feedback of tip reacting to the abrupt change. The forward and backward lines in **a** and **c** do not differ very much.

Note 2: Average height comparison between 2 SL and 3 QL of MBT and impurity phase, respectivly.

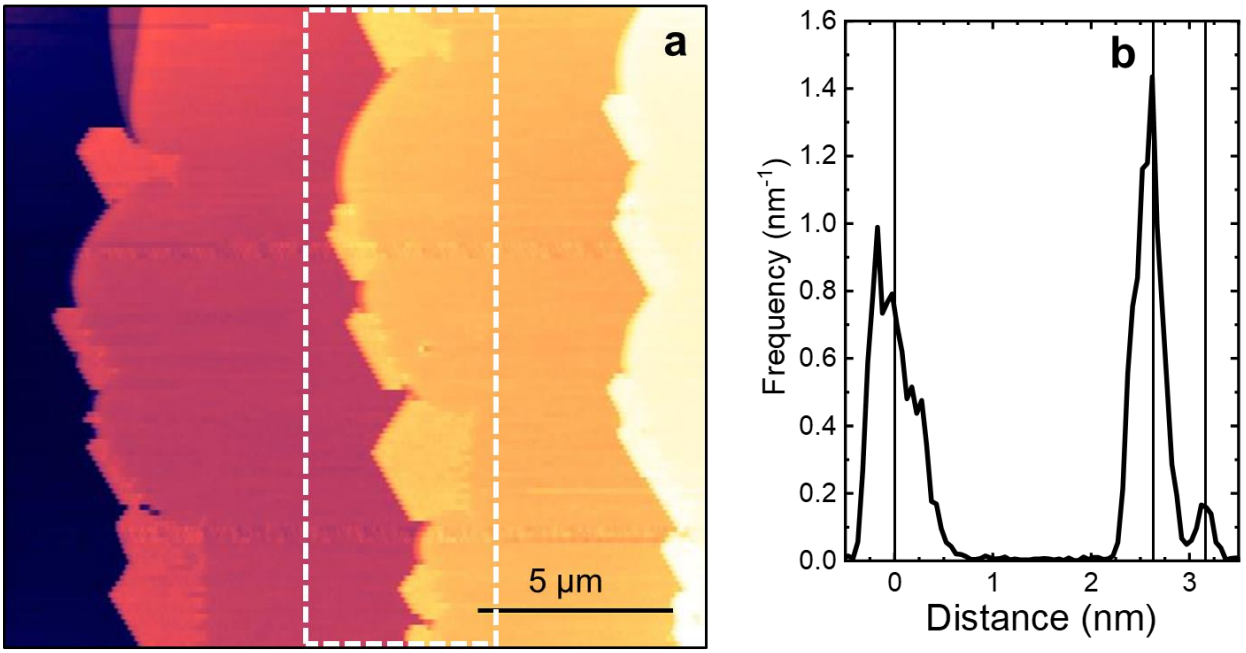

Fig. S2 | Topographic image and height distribution of as-grown $MnBi_2Te_4$ and $Bi_{2-x}Mn_xTe_3$ impurity phase. a, Topographic image of as-grown $MnBi_2Te_4$ surface from Fig. 2. **b**, Height distribution of dotted square area in **a**. The two large peaks show the difference in height between the two septuple layers. The smaller third peak shows the height of the impurity phase. The average peak values are 2.63 and 3.15 nm, for the two SLs and impurity phase, respectively. The value of 3.15 nm is close to the height of three quintuple layers (QL) of $Bi_{2-x}Mn_xTe_3$, and 2.63 nm is close to the height of two SLs. The theoretical values are 3.06 and 2.72 nm. The color scale for the topographic image is 7 nm.

Note 3: Topographic, MFM, differential, and height distribution of septuple layer and $Bi_{2-x}Mn_xTe_3$ impurity phase.

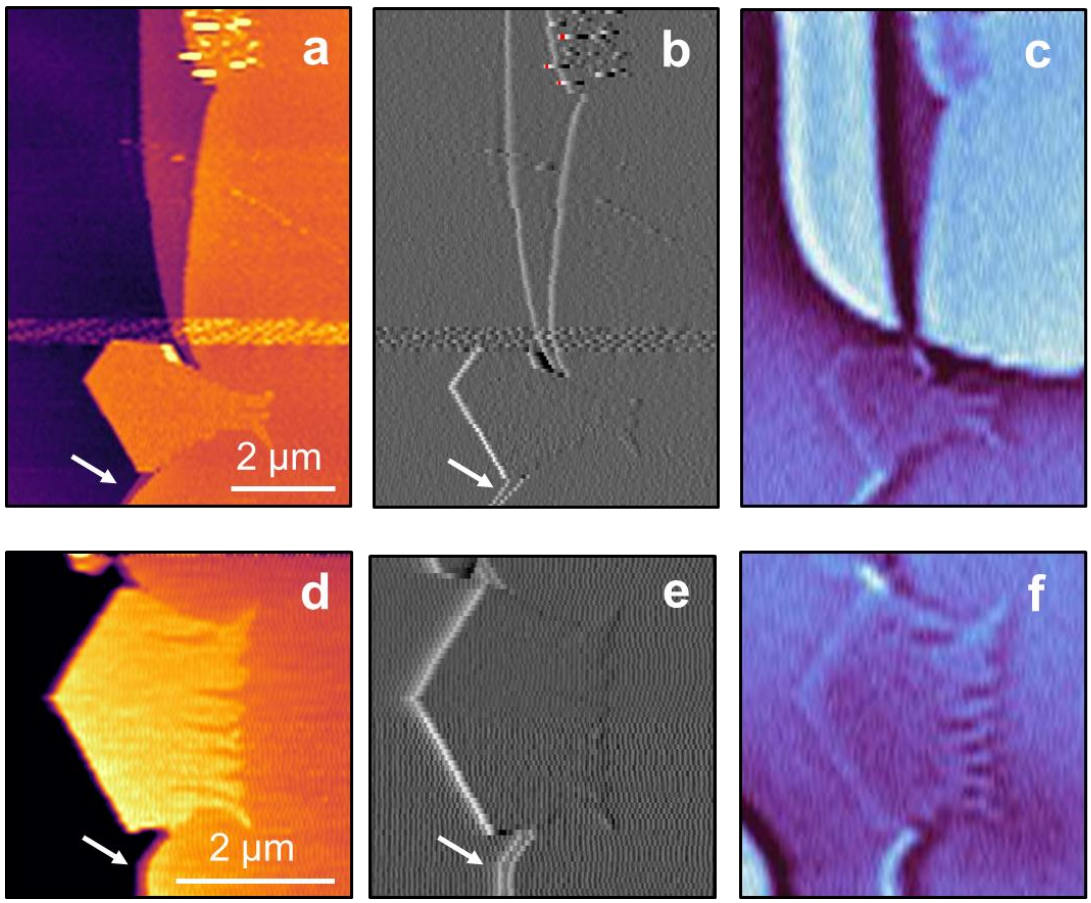

Fig. S3 | Topographic, MFM, and differential images of step edges. a-c, Topographic, differential topographic, and MFM images, respectively, of scan in Fig. 1. **d-f**, Topographic, differential topographic and MFM images of impurity phase at another location from Fig. 1. The step edges of two septuple layers (SL) are easily seen above and below the $Bi_{2-x}Mn_xTe_3$ impurity phase in **a** and **d**, indicated by white arrows. However, it is not clear from the topography if there is a step edge on the impurity. A differential map of the topography shows clearly the separation of the step edges, but shows no such separation on the impurity edge. Thus, the bright contrast in **c** and **f** comes from a single SL. The color scales for the images are 6 nm and 0.2 Hz (**a-c**) and 3 nm, 0.5 nm and 0.2 Hz (**d-f**).

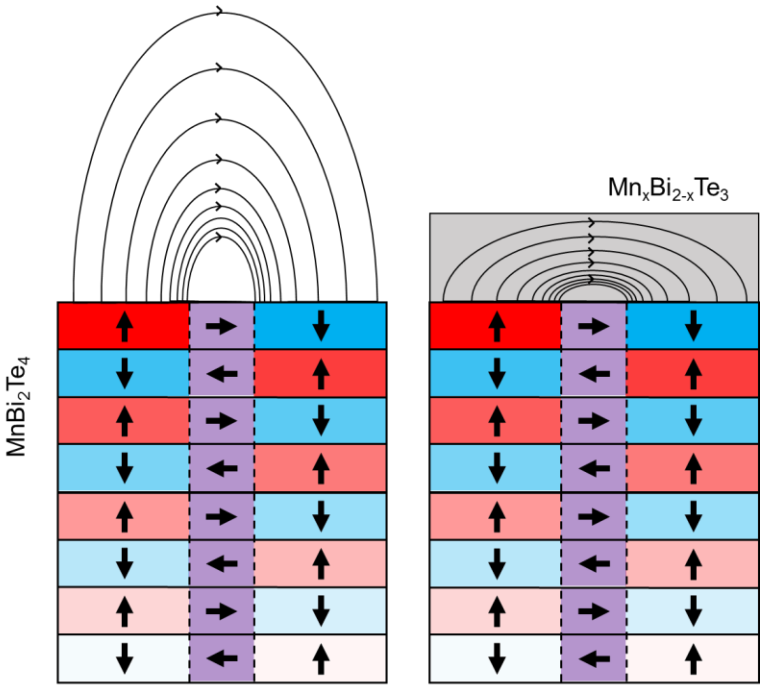

Fig. S4. Cartoon illustration of the screening effect of the impurity phase ($Bi_{2-x}Mn_xTe_3$). The magnetic flux lines are effectively trapped inside the impurity phase because of higher permeability.

Note 4: Additional MFM images from positive and negative field sweeps, additional sample location, thermal cycle, and MFM images of cleaved surface.

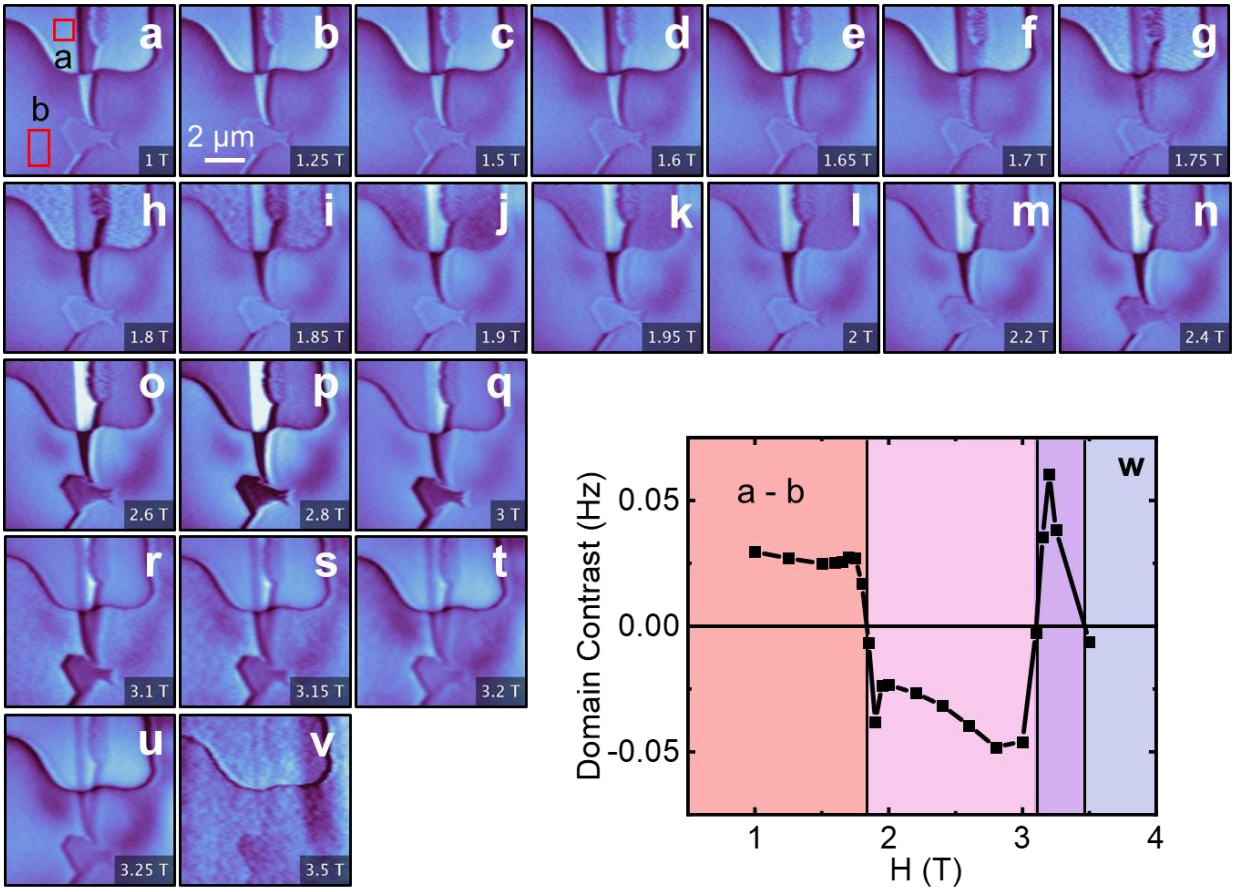

Fig. S5 | MFM images and positive magnetic field dependence of domain contrast on as-grown $MnBi_2Te_4$. a-v, MFM images taken at 5 K with increasing positive field labelled in lower right corners. **w**, Domain contrast between red squares in **a** versus applied field. Below 1.75 T, the domain contrast is constant. As the applied field is increased, the domain contrast quickly reverses around 1.85 and 3.1 T. Above 3.5 T, the system enters the canted AFM (CAFM) phase. Near 1.85 T, the bright domain starts to appear rougher and darker, i.e. the antiparallel surface domain spin-flops. At 3.1 T, the next lower SL spin-flops due to its interaction with the spin-flopped surface layer, and thus, the bright domain again begins to appear rougher and darker. The color scale for MFM images is 0.3 (**a-q**) and 0.8 (**r-v**) Hz.

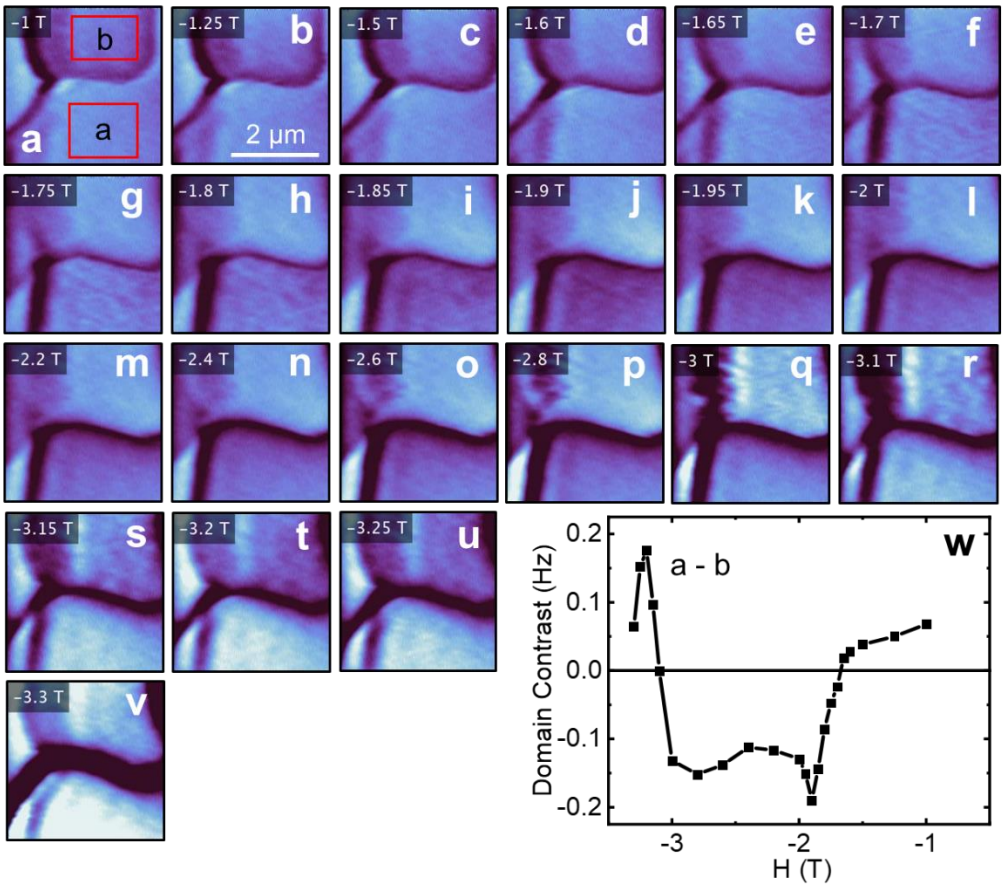

Fig. S6 | MFM images and negative magnetic field dependence of domain contrast on as-grown $MnBi_2Te_4$. a-v, MFM images of upper right domain in Fig. S4 with increasing negative field labelled in lower right corners. **w**, Domain contrast between red squares in **a** versus applied field. Below -1.65 T, the domain contrast is decreases slowly, until around -1.75 T, the domain contrast quickly reverses, and again at -3.1 T. Above -3.5 T, the system enters the canted AFM (CAFM) phase. Near -1.7 T (**f**), the bright domain starts to appear rougher and darker, i.e. the antiparallel surface domain spin-flops. At -3.1 T (**r**), the next lower SL spin-flops due to its interaction with the spin-flopped surface layer, and thus, the bright domain again begins to appear rougher and darker. The domain contrast shows similar qualitative behavior to the positive field run in Fig. S4, thus, the effect is symmetric in positive and negative field. The color scale for MFM images is 0.3 Hz.

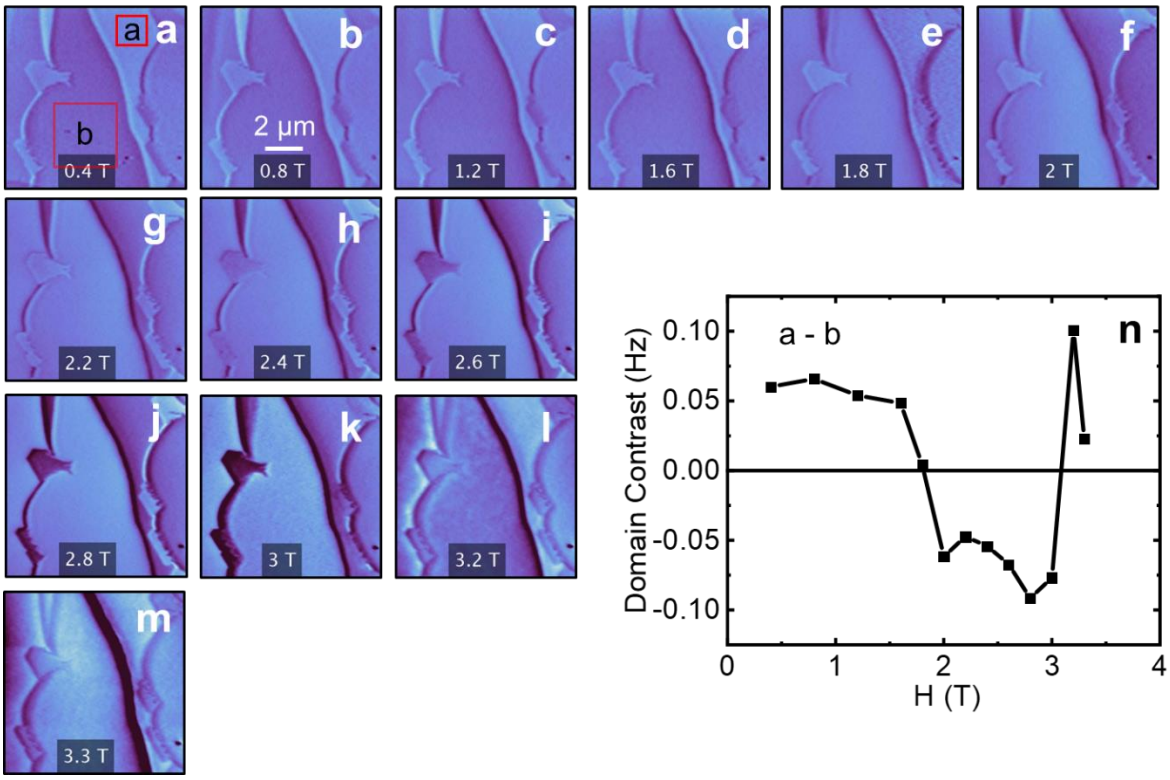

Fig. S7 | MFM images and magnetic field dependence of domain contrast on as-grown $MnBi_2Te_4$ at different location and domain configuration. a-m, MFM images taken at 5 K of zoomed out location of Fig. S4 after thermal cycling. A new domain wall was observed running down the middle of **a**. **n**, Domain contrast between red squares in **a** versus applied field. Again, similar qualitative behavior in the domain contrast vs applied field was observed. The bright antiparallel domain in **e** begins to reverse around 1.8 T and then reverses again around 3.1 T (**I**). Thus, the domain reversal is reproducible even after thermal cycling to a different domain configuration. The color scale for the MFM images is 0.5 Hz.

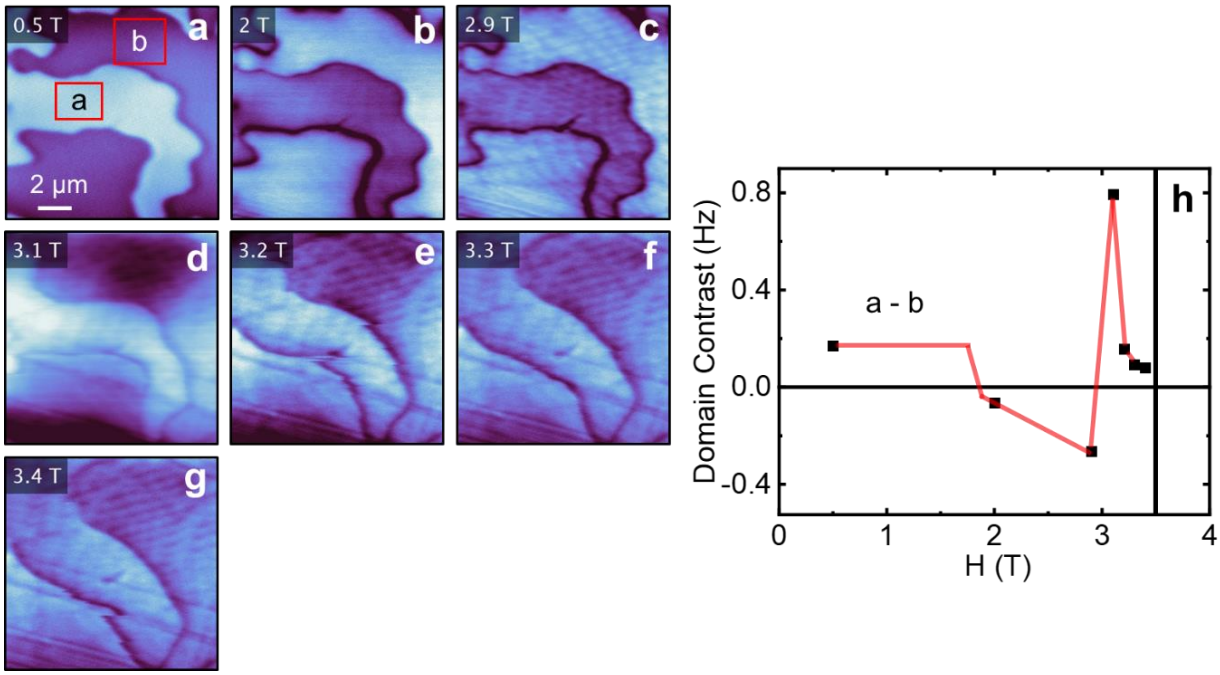

Fig. S8 | MFM images and magnetic field dependence of domain contrast on cleaved surface of $MnBi_2Te_4$. **a-g**, MFM images taken at 15 K with increasing magnetic field on cleaved MBT. **h**, Domain contrast between red squares in **a** versus applied field. The red line is a guide to show contrast reversals. The domain reversal behavior on the cleaved surface is qualitatively similar to that on the as-grown surface. This shows that the surface flop is independent of a particular type of surface termination. The color scale for the MFM images is 0.5 Hz.

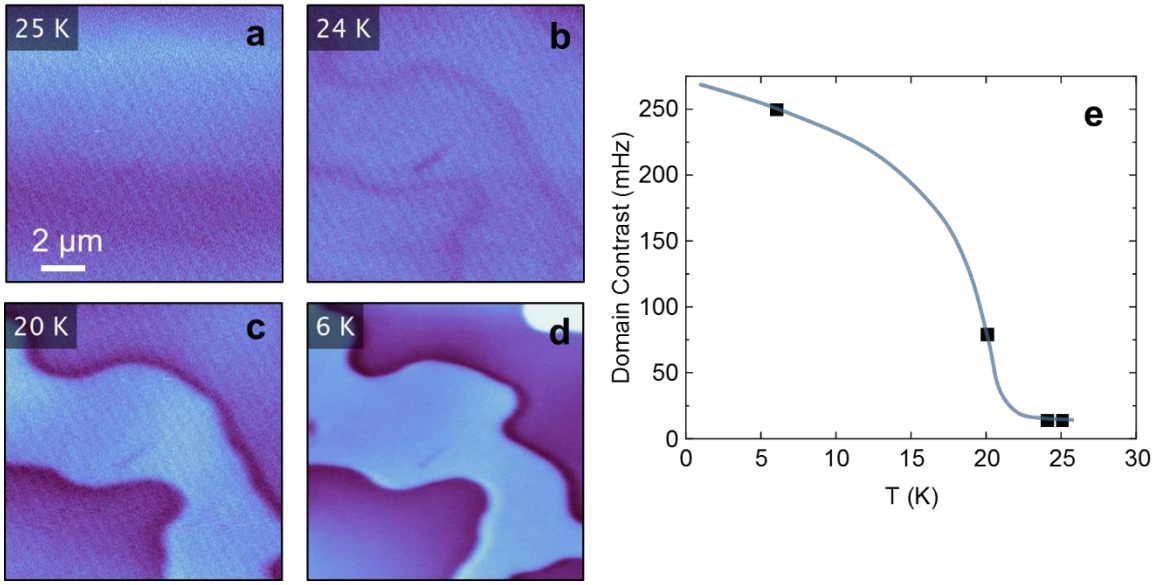

Fig. S9 | MFM images and temperature dependence of domain contrast on cleaved surface of $MnBi_2Te_4$. **a-c**, MFM images taken at zero field with decreasing temperature (25, 24 and 20 K) on cleaved surface of $MnBi_2Te_4$. **d**, MFM image of the same location taken at 5 K in 0.5 T. **e**, The temperature dependence of domain contrast. The blue line is a guide to eye. The color scale for the MFM images is 0.15 Hz for **a-c**, and 0.5 Hz for **d**.

Note 5: Surface spin-flop theoretical model and simulation

Pinning of spin-flop state via surface magnetization reduction

In the original Mills' model, the spin-flop state first nucleates on the antiparallel surface and migrates into the bulk with increasing external field as reproduced in Fig. S10 (a) and (b). Figure S10 (a) shows the canting position $\langle d \rangle$ with respect to the external field, which is calculated as,

$$\langle d \rangle = \frac{1}{N} \frac{\sum_{i=1}^{N} i \sin^2(\theta_i)}{\sum_{i=1}^{N} \sin^2(\theta_i)}.$$

The migration occurs as a consequence of first order transitions that are clearly seen in the stepwise increase of the canting position. Two surfaces eventually become indistinguishable (at B_6) due to the migration. In the revised Mills' model, however, the flop state is found to be pinned on the surface if the surface magnetization is reduced more than 10% ($\lambda_s < 0.9$). Instead of the stepwise change, the flop state smoothly expands from the surface as shown in Fig. S10 (c). The more the surface magnetization is reduced, the more the flop state is localized on surface. An intermediate step appears for $\lambda_s = 0.4$ due to the spin-flop transition of the parallel surface as shown in Fig. S10(d).

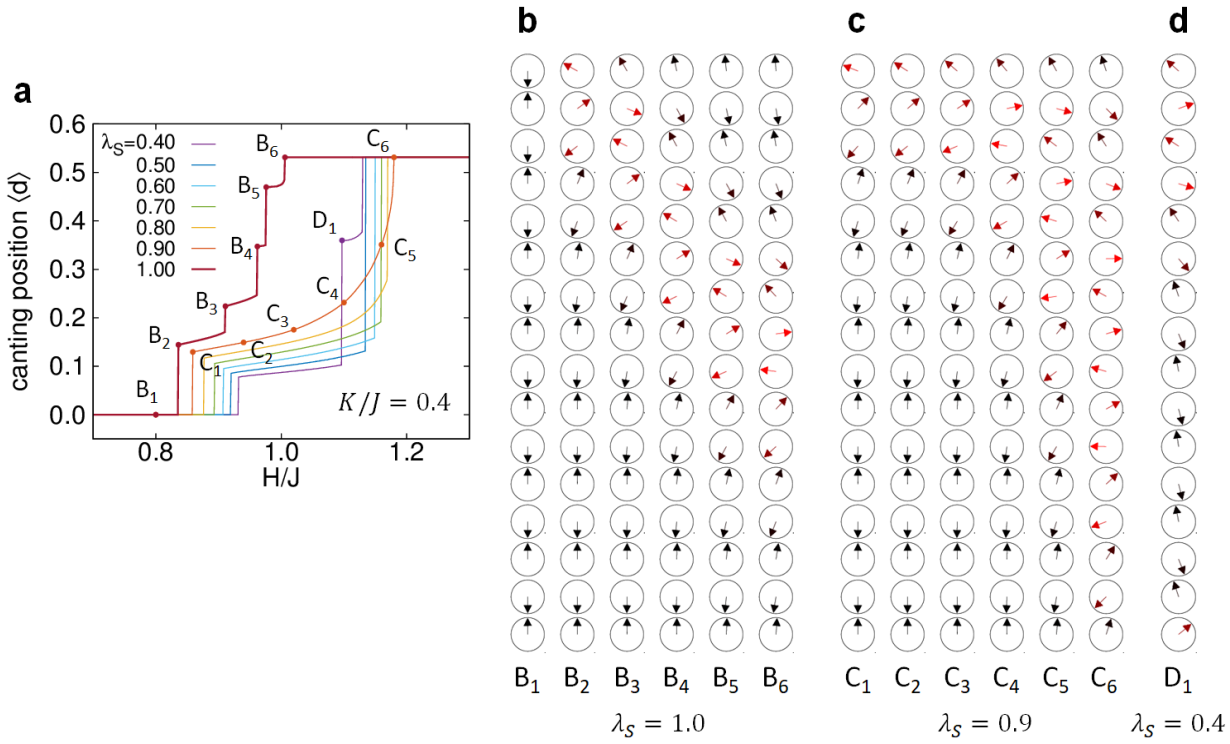

Fig. S10 | Numerical simulation of revised Mills' model for N = 16 layers with surface modification, $\lambda_S = S_{surf}/S_{bulk}$. **a**, Normalized depth of spin-flop state with respect to external field. Color of lines indicates surface magnetization. **b-d**, Spin configurations of original and revised Mills' model (λ_S = 1.0, 0.9, 0.4) at each point in (**a**). Color of arrows represents the amount of canting from layer normal.

Universality of the MFM contrast reversal

The reduction of surface magnetization induces SSF on the parallel surface as seen in Fig. S10 (d). In order to understand the individual effect of surface parameters (λ_s, λ_J, λ_K) on the SSF transition, the parameter space is swept at $K/J = 0.4$. Figure S11 (a) shows the parameter window where the relative threshold field, $r_{A,P} \equiv (H_{BSF} - H_{SSF}^{A,P})/H_{BSF}$ exhibits the values found in experiment ($r_A \approx 0.5$, $r_P \approx 0.1$) with 10% (dark shadow) and 20% (pale shadow) errors. The colored solid lines are contour plots satisfying $r_A = 0.5$ (blue) and $r_P = 0.1$ (red) as shown in Fig. S11 (b). It

clearly shows that SSF transition occurs on both surfaces in a broad parameter window and the MFM contrast is reproduced in good agreement with experiment (Fig. S11 (c)).

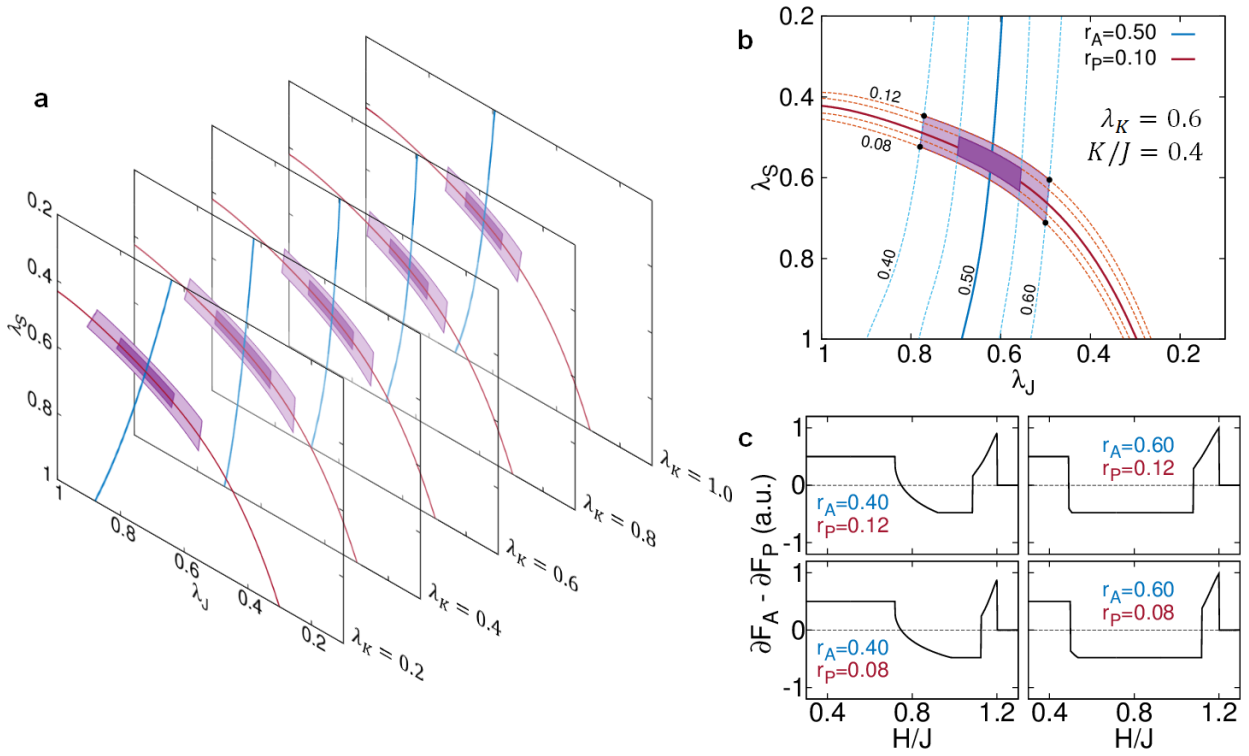

Fig. S11 | Numerical simulation of clamped model for N = 21 layers with surface modification, $\boldsymbol{\lambda_A = A_{surf}/A_{bulk}}$ ($\boldsymbol{A = J, K, s}$). a, Contour lines satisfying $r_A = 0.5$ (blue) and $r_P = 0.1$ (red) in the parameter space of (λ_J, λ_s). **b**, Contour lines for a specific case of $\lambda_K = 0.6$. Dark and pale colored area indicates 20% and 10% window from the crossing points, respectively. **c**, Simulated MFM contrast at the corners of 20% window in (**b**), which qualitatively agree with the experimental results in Fig. 3(i) of main text.

Note 6: MFM images of bubble domain at 18 K

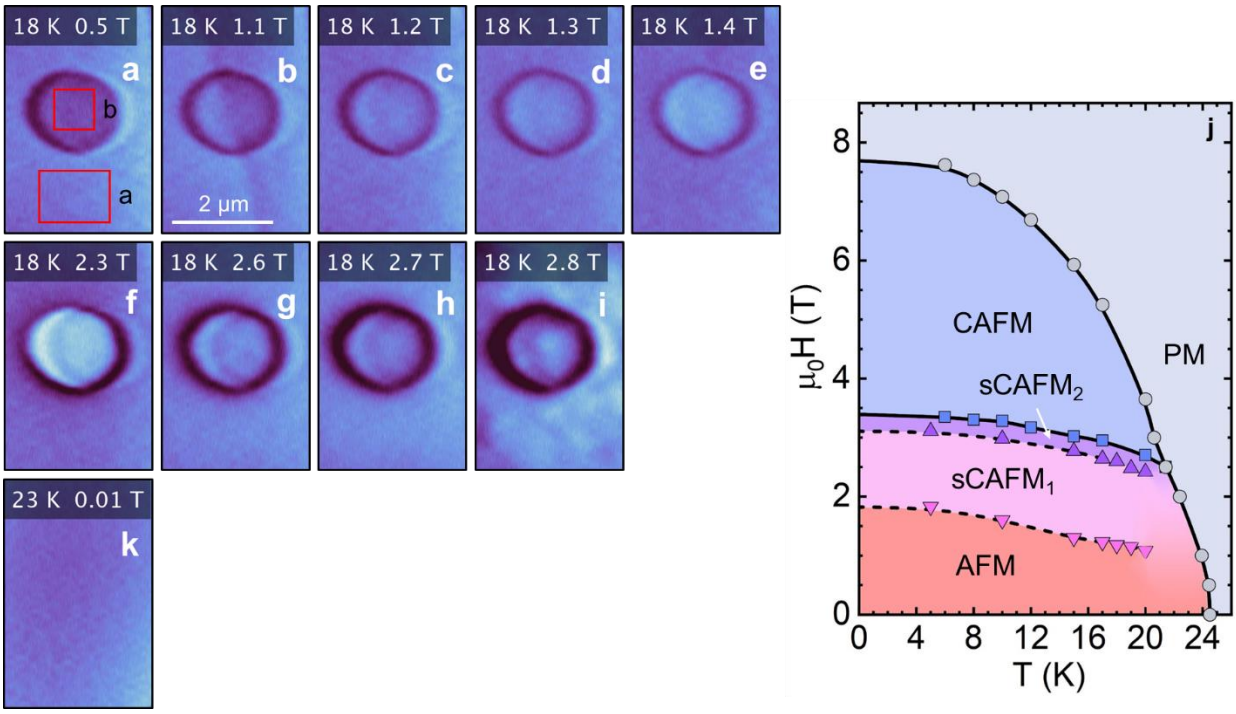

Fig. S12 | MFM images at high temperature used to construct surface-flop phase diagram. a-i, MFM images taken at 18 K on as-grown MBT with increasing field. The surface flop transitions appear around 1.2 T (**c**) and 2.6 T (**g**). Domain contrast between a and b in bubble domain was used to construct phase diagram of surface transitions, shown in **j**. **k**, MFM image at same location at 23 K and 0.01 T where bubble domain disappeared. The surface flop transition is likely suppressed somewhere in the range of 20-23 K because the spin-flop transition merges with saturation transition above the bicritical point (~22 K, 2.5 T). The color scale for the MFM images is 0.3 Hz.