Supplemental information

Section 1 Normalization of the photocurrent

To make the decay of the photocurrent at different temperature could be mutual compared, we normalize them by Eq. (S1)

$$I_N = \frac{Current - I_D}{I_{P_0}} \times 100\% \quad \text{(S1)}$$

Where $Current - I_D$ (=I_P) refers to the (net) photocurrent obtained from Fig. 1c at a certain temperature, and I_{P_0} is corresponding photocurrent at onset of the decay (=150 s) as shown in Fig. 1c. From Eq. S1, we can found that I_D at all the temperature is normalized to 0, while the normalized photocurrent under all temperature is always normalized to 1 when at onset of the decay.

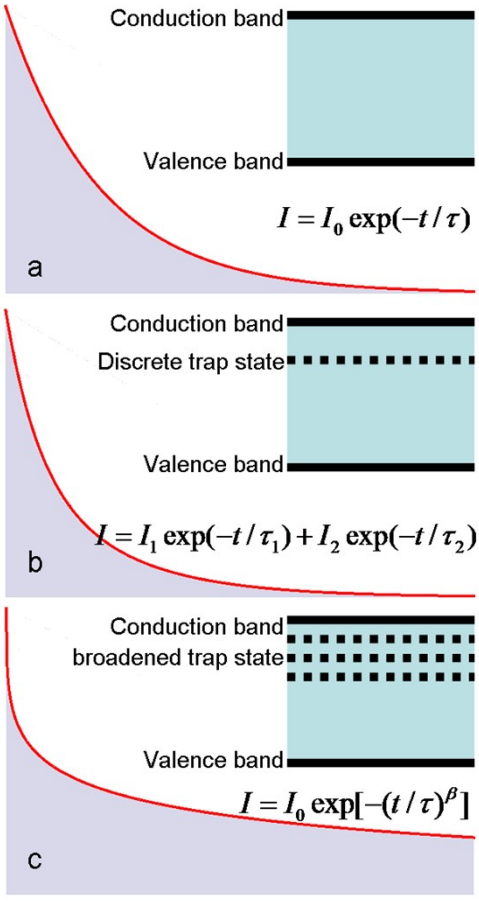

Fig.S1. the decay of the photocurrent of the semiconductors (a) without traps, (b) with discrete trap state and (c) with broadened trap state.

Section 2 Photocurrent decay depending on the trap distribution

In general, depending on the energy distribution of the traps, the decay of the photocurrent Ip in the semiconductors can be cataloged as three types shown in Fig. S1[2]: (1) for crystalline semiconductors with little traps, the decay follows exponential-law, i.e $Ip=I_0\exp(-t/\tau)$, where I_0 is the generated photocurrent under the illumination, the decay feature of the current is governed by τ, the life time of photo-exited carriers. (2) In the case of a few traps existed in the material, the decay follows double-exponential-law, $Ip=I_1\exp(-t/\tau_1)+I_2\exp(-t/\tau_2)$, and the two feature times τ_1 and τ_2 correspond to the direct recombination of electron-hole pair and traps-participated recombination, respectively. (3) For the traps with a broad-range energy-level distribution, the decay will follow the stretched-exponential-law, $Ip=I_0\exp[(-t/\tau)^{\beta}]$ [1, 2], where the decay not only depends on the feature time τ but also the dispersive parameter β ($0<\beta<1$). Briefly speaking, a smaller β ($\beta\rightarrow 0$) implies a relative broad energy range of the traps (trap band). In this case, the photo-exited carriers might be captured by some traps with deep energy level, and then become hard to escape and thus recombine with the holes. In this process the current cannot decay to the initial level (dark current level) but still keeps it in a metastable high level, this is called persistent-photocurrent (PPC)[2-4].

Section 3 Physics meaning of the active energy extracted from the thermally activated behavior of the feature time τ in PPC

Before the discussion to the general condition with arbitrary dispersive parameter β and feature time τ in the stretched-exponential law model, let's consider a degraded condition with $\beta=1$, which means the traps band would shrink to the discrete trap state, and the weighting center of the traps (WCT) is just the energy level of the discrete trap state itself. In this case, the decay of the photocurrent follows double-exponential model representing the combination of direct recombination of electron-hole pair and traps-participated recombination. The energy difference ΔE between the discrete trap state (WCT in this degraded case) and the conduction band bottom is estimated from the feature time τ_1 and τ_2 *via* $\tau_1/\tau_2=\exp(-\Delta E/kT)$[5-7], where k and T is the Boltzmann constant and temperature. Hence it is rational to deduct that when discrete trap state is broaden to the traps band, the active energy extracted from the feature time τ could be also deemed as the energy difference between the WCT and the conduction band bottom.

Section 4 The estimation of E_{CB}

The value of E_{CB} can be estimated from the E_a^{τ} and $E_a^{I_D}$ (extracted from temperature dependent PPC), as well as the band gap of Bi_2S_3 (1.3 eV) and its n-type nature[S1,S4]. Firstly, since Bi_2S_3 is n-type, E_{CB} should be lower than the half of the band gap (=0.65 eV). On the other hand, the E_{CB} (= $E_a^{\tau}+E_{WA}$) should be higher than $E_a^{\tau}+E_a^{I_D}$ (=0.44 eV), because $E_{WA}>E_a^{I_D}$ (=$E_{DT,}$), *i.e.* $0.44\text{eV}<E_{CB}<0.65$ eV. Therefore, a rough but rational estimation of E_{CB} could be the average of its upper-bound and lower-bound as $(0.44+0.65)/2\approx 0.55$eV.

Section 5 Double-exponential-decay of the photocurrent in Bi_2S_3 thin film (BSTF)

The Bi_2S_3 thin film (BSTF) used in the photoconductivity study is grown by chemical bath

deposition[S3]: a FTO glass is immersed in the mixture solution containing Triethanolamine (TEA,8 ml), Thioacetamide (TAA, 0.3 g), $Bi(NO_3)_3 \cdot 5H_2O$ (1.22 g) and H_2O (42 ml) at 40 °C for 2 h. As a result, BSTF with a thickness of ~150 nm is deposited on the FTO substrate, which is used as the bottom electrode. And the top electrode is silver paste (~0.8 mm) with ohmic contact to BSTF. It is noteworthy that there is the surrounded SiO_2 layer to prevent the potential short circuit brought by the porous structure in BSNN, but it is not necessary for BSTF with thin film structure. The morphology of the BSTF is presented in the Fig. S2a.

The decay of the photocurrent of BSTF follows double-exponential-law (Fig. S2b)[S2]. The two feature times (τ_1 and τ_2) are 2.43s and 6.60s, respectively. It can be found that τ_1 and τ_2, corresponding to the direct recombination of electron-hole pair and traps-participated recombination, respectively, are rather close to each other. This is because here the (discrete) traps are just slightly below the conduction band bottom (shadow level traps) and cannot embar the electrons effectively[7]. The small energy-level difference ΔE between the discrete shadow level trap and the conduction band bottom is just 0.025 eV according to the estimation from the formula $\tau_1/\tau_2=\exp(-\Delta E/kT)$. Consequently, such shadow level traps in the BSTF is hard to induce the RS behavior produced by the electron trapping/detrapping.

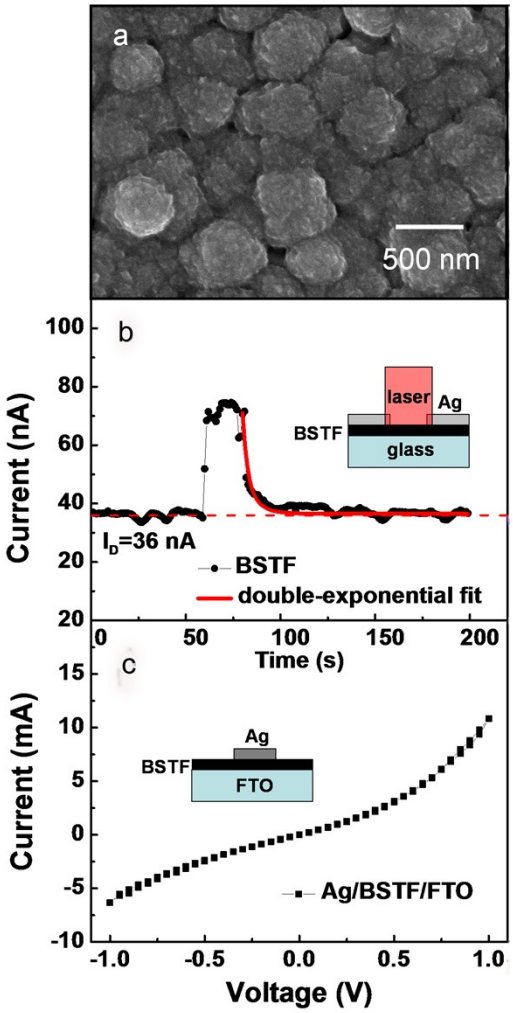

Fig. S2. (a) The morphology, (b) decay of the photocurrent and (c) switching features of the Bi_2S_3 thin film

Section 6 Switching saturation of R_{LR}

In our previous work, the RS behavior of Ag/BSNN/FTO had been observed in the voltage sweep (0 V → 1V→ -1V → 0 V) shown in Fig. S3， with a typical interval of 1 ms and step of 50 mV, corresponding to an equivalent sweep speed of 50V/s [S4]. The sweep-speed-dependence of the RS behavior can be investigated by tuning the interval. It can be found that as the interval increases to 50 ms (i.e. sweep speed slows down to 1V/s), resistance at LR state (R_{LR}) becomes saturated, implying the traps are almost fully filled.

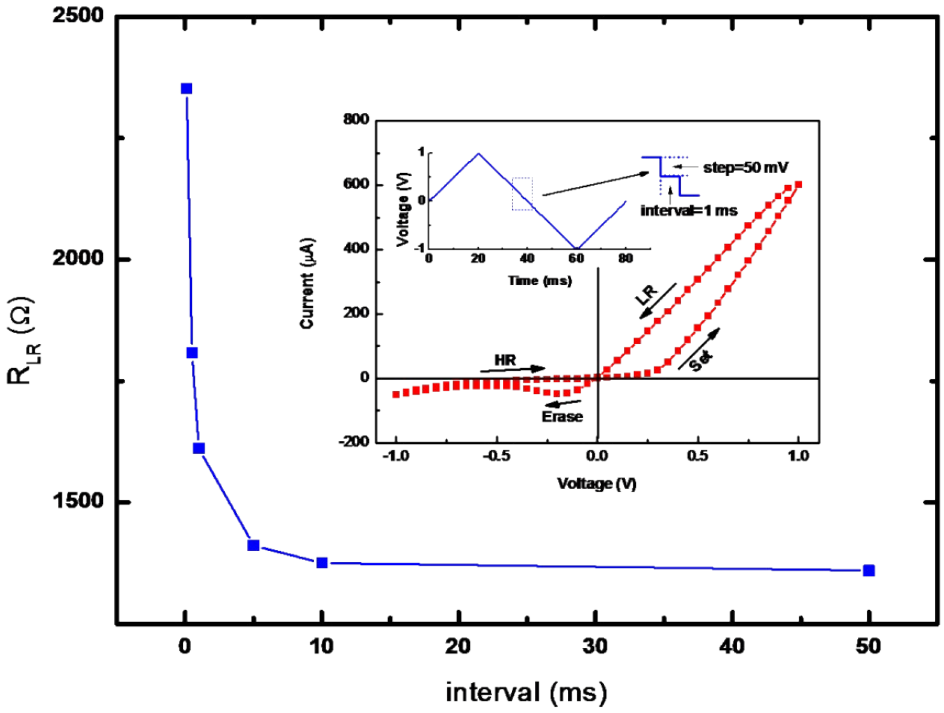

Fig. S3 Switching saturation of R_{LR}

Section 7 Positive voltage dependence of low resistive (LR) state

In our previous work[S4], the positive voltage dependence of low resistive (LR) state of Ag/BSNN/FTO had been observed in the voltage sweep (0 V → x V→ -1V → 0 V, x=0.5, 0.6, 0.7, 0.8, 0.9, 1 V) shown in Fig. S4. It can be found that at all positive voltage x from 0.5 to 1 V, the conductance G at LR state obtained via ohmic fit can be approached as a function of x:

$$G(x)=I/V=(ax + b)/x \quad (S2)$$

where a and b is slope and intercept of the inclination (blue line in Fig. S4).

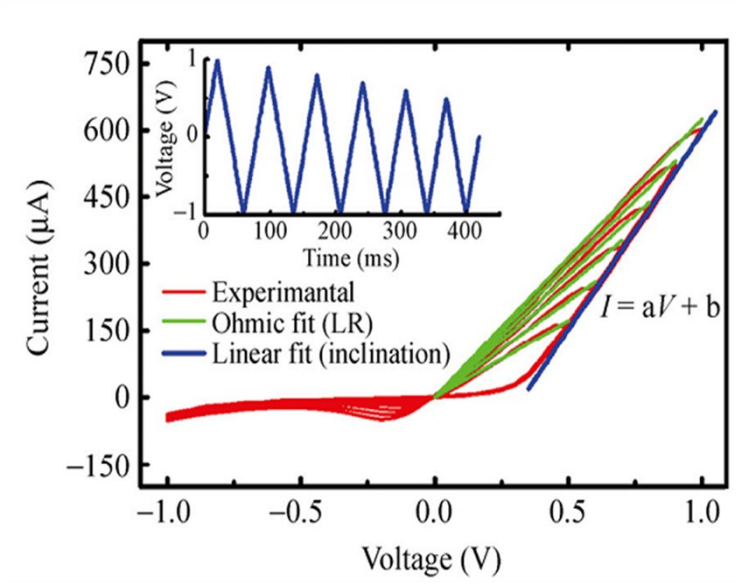

Fig. S4 the positive voltage dependence of low resistive (LR) state

Section 8 Theoretical model of the switching

For the switchable interface with a trap-density of N_t per unit area, partial electrons will fill into the interface traps with the energy lower than -eV and open high conductive paths. If the sweep speed is quite slow (e.g. 1V/s in our case, section 6), the traps whose energy level is lower than -eV could be almost fully filled, resulting in a low resistive (LR) state depending on the voltage (section 7). Accordingly, if assuming an energy distribution of f(E), we could get the interface conductance G contributed from the traping/detraping process as

$$G = G_H N_t \int_{-\infty}^{-eV} f(E)dE + G_L N_t (1 - \int_{-\infty}^{-eV} f(E)dE) \quad \text{(S3)}$$

where G_H is the conductance of the high conductive paths, and G_L is the low conductance of unfilled traps. Since $G_H >> G_L$, Eq(S3) can be simplified as $G = G_H N_t \int_{-\infty}^{-eV} f(E)dE$, and then $dG/dV = -eN_t G_H f(-eV) \propto f(-eV)$. Considering G=I/V, here dG/dV can be directly obtained from the experimental I-V characteristic and represents the profile of f(E)

References

1. C. Nebel, *Semiconductor science and technology*, 2003, **18**, S1.
2. K. Ghaffarzadeh, A. Nathan, J. Robertson, S. Kim, S. Jeon, C. Kim, U.-I. Chung and J.-H. Lee, *Applied Physics Letters*, 2010, **97**, 143510-143510-143513.
3. C. H. Qiu and J. Pankove, *Applied physics letters*, 1997, **70**, 1983-1985.
4. C. Johnson, J. Lin, H. Jiang, M. A. Khan and C. Sun, *Applied physics letters*, 1996, **68**, 1808-1810.
5. D. Berben, K. Buse, S. Wevering, P. Herth, M. Imlau and T. Woike, *Journal of Applied Physics*, 2000, **87**, 1034-1041.
6. C. Lee, G. Yu, D. Moses, K. Pakbaz, C. Zhang, N. Sariciftci, A. Heeger and F. Wudl, *Physical Review B*, 1993, **48**, 15425.
7. J. Hao, S. Studenikin and M. Cocivera, *Journal of Applied Physics*, 2001, **90**, 5064-5069.