Mechanism and observation of Mott transition in VO$_2$-based two- and three-terminal devices

Hyun-Tak Kim$^1$, Byung-Gyu Chae$^1$, Doo-Hyeb Youn$^1$, Sung-Lyul Maeng$^1$, Gyungock Kim$^1$, Kwang-Yong Kang$^1$ and Yong-Sik Lim$^2$

$^1$ Telecom. Basic Research Lab., ETRI, Daejeon 305-350, Republic of Korea
$^2$ Department of Applied Physics, Konkuk University, Chungju, Chungbuk 380-701, Republic of Korea
E-mail: htkim@etri.re.kr and kimht45@hotmail.com

New Journal of Physics 6 (2004) 52
Received 24 September 2003
Published 17 May 2004
Online at http://www.njp.org/
DOI: 10.1088/1367-2630/6/1/052

Abstract. When holes of about 0.018% are induced into a conduction band (breakdown of critical on-site Coulomb energy), an abrupt first-order Mott metal–insulator transition (MIT) rather than a continuous Hubbard MIT near a critical on-site Coulomb energy $U/U_c = 1$, where $U$ is on-site Coulomb energy between electrons, is observed on an inhomogeneous VO$_2$ film, a strongly correlated Mott insulator. As a result, discontinuous jumps of the density of states on the Fermi surface are observed and inhomogeneity inevitably occurs. The off-current and temperature dependences of the abrupt MIT in a two-terminal device and the gate effect in a three-terminal device are clear evidence that the abrupt Mott MIT was induced by the excitation of holes. Raman spectra measured by a micro-Raman system show an MIT without the structural phase transition. Moreover, the magnitude of the observed jumps $\Delta J_{\text{observed}}$ at the abrupt MIT is an average over an inhomogeneous measurement region of the maximum true jump, $\Delta J_{\text{true}}$, deduced from the Brinkman–Rice picture. A brief discussion of whether VO$_2$ is a Mott insulator or a Peierls insulator is presented.
In a strongly correlated system, a metal–insulator transition (MIT) near a critical on-site Coulomb energy $U_c$, $U/U_c = 1$, where $U$ is the on-site Coulomb energy between electrons, has long been controversial as to whether the transition is abrupt or continuous in experiments [1]–[4], although a first-order MIT with temperature was observed by Morin [5]. An abrupt MIT indicates a Mott transition (first order) [1] and a continuous MIT a Hubbard transition (second order) [6]. The MIT breaks down an energy gap formed by the strongly correlated on-site Coulomb energy $U$, between sub-bands in a main band. Mott first predicted that the abrupt MIT occurs when a lattice constant is larger than a critical value [1]. Brinkman and Rice theoretically demonstrated an abrupt MIT near $U/U_c = 1$ for a strongly correlated metal with an electronic structure of one electron per atom [7]; this is called the Brinkman–Rice (BR) picture. The abrupt MIT with band filling was also developed by extension of the BR picture [8]; this is called the extended BR picture. This was based on a fractional charge justified by means of measurement in an inhomogeneous metallic system [8]. Meanwhile, Hubbard first derived that, when sub-bands overlap just below $U_c$, there is a finite minimum density of states (DOS) at the Fermi level, the DOS increases with decreasing $U$, and the system is metallic [6]; this is Hubbard’s continuous MIT. Later, the continuous MIT was confirmed in the infinite-dimensional Hubbard model [2].

Experimentally, applying an electric field to a two-terminal structure, Ohmic behaviour was measured in an organic Mott insulator in a regime where conduction from non-conduction (insulating behaviour) occurs [9]. Through a theoretical consideration based on the Hubbard model, the Ohmic behaviour was described in terms of a universal Landau–Zener quantum tunnelling [10]. A similar metallic behaviour for VO$_2$ was also observed [11]. Thus, on the basis of the metallic behaviours and the Oka et al analysis [10], the MIT just below $U_c$ seems to follow Hubbard’s continuous model. However, considering the abrupt MIT near $T_r \approx 340$ K ($67^\circ$C) observed in resistance measurement (figure 3), whether an MIT near $U/U_c = 1$ is abrupt or continuous in the experiments is not clear.

Another problem to consider is whether VO$_2$ is a Mott insulator or a simple band (Peierls) insulator. For VO$_2$ at $T_r \approx 340$ K, it is known that the electrical MITs are accompanied by a structural phase transition from monoclinic below $T_r$ to tetragonal above $T_r$ [5, 12, 13, 14]. VO$_2$ has also been classified as an ordinary simple band (Peierls) insulator for the structural phase
transition [15]–[17]. However, Rice et al [18] suggested that VO₂ has an M₂ phase undergoing an MIT and is a Mott insulator. Thus the problem remains controversial.

Generally, strongly correlated oxide materials are compounds composed of more than two elements including oxygen. Most of the materials are spatially inhomogeneous [19]–[21]. In particular, multiphases in VO₂ thin films were observed [22, 23]. To find an abrupt MIT in a VO₂ film, an MIT model (or theory) describing an inhomogeneous system is needed.

In this paper, we introduce an extended BR picture [8] that describes an inhomogeneous system and also includes a first-order MIT for an inhomogeneous system. We first observe an abrupt jump of current (or DOS) without a structural phase transition at a given electric field in a two-terminal structure and measure the gate effect of the jump in a three-terminal device (switching transistor), for VO₂ films below \( T_c \approx 340 \) K (assuming that VO₂ is a Mott insulator). The abrupt jump is analysed in terms of an abrupt Mott MIT without a structural phase transition. Inducing internal hole charges of about 0.018% in hole levels into conduction band (or the valence band) with a source–drain field or a gate field of a transistor [24] is an effective method to reveal the MIT mechanism. Furthermore, a brief discussion on whether VO₂ is a Mott insulator or a Peierls insulator is presented.

2. A new method to observe the Mott transition

In an inhomogeneous strongly correlated metallic system with two phases of a metal region and an insulating region, when measured e.g. by photoemission spectroscopy, a spectral-weight value in \( k \)-space is observed; however, the inhomogeneous phases cannot be deduced from the observed spectral weight. In other words, a reverse transformation from \( k \)-space into real-space is not defined (figure 1(a)). This indicates that two real- and \( k \)-spaces are not mathematically equivalent. The inhomogeneous system is different from the metal in having both an electronic structure of one electron per atom and mathematical equivalence between the two spaces. To overcome this problem, we consider measured data to be averaged data. When the inhomogeneous system is measured, carriers in the metal region should be averaged over lattices (or atoms) in the entire measurement region. Then, the inhomogeneous system is changed to a homogeneous one with an electronic structure of one effective charge per atom (figure 1(b)). The observed effective charge becomes \( e' = \rho e \), where \( 0 < \rho = n/L \leq 1 \) is band filling (or local density), \( n \) is the number of carriers in the metal region and \( L \) the number of total lattices in the measurement region. The fractional effective charge is justified only when the inhomogeneous system is measured. Otherwise, it becomes a true charge in the metal region.

On the basis of the fractional charge, the fractional Coulomb energy is defined as

\[
U \equiv \langle e'e'/r \rangle = \rho^2 U_{true},
\]

where \( U \) is the effective Coulomb energy and \( U_{true} \) is the true Coulomb energy. For a given \( U_c, U_{true} = \kappa U_c \) is given, where \( \kappa \) is the correlation strength used in the BR picture [7]. Thus \( U \) is defined as \( U = \kappa \rho^2 U_c \).

For an averaged system with one effective charge per atom, the effective mass of a quasiparticle \( m^* \), calculated by the Gutzwiller variational theory [8], is given by

\[
\frac{m^*}{m} = \frac{1}{1 - (U/U_c)^2} = \frac{1}{1 - \kappa^2 \rho^4}.
\]
where $m$ is the bare electron mass, $U/U_c$ is $\kappa \rho^2 \neq 1$, $\kappa$ is the strength of Coulomb energy between carriers when $\rho = 1$, and $0 < \rho \leq 1$ is band filling (see figure 2). When $\rho \neq 1$ for an inhomogeneous system, equation (1) is the effect of measurement and an average of the true effective mass in the BR picture [8]; electric conductivity $\sigma \propto (m^*/m)^2$ [1]. This defines the extended BR picture (hole-driven MIT theory) [8].

The material at $\kappa \rho^2 = 1$ in equation (1) can be assumed to be a paramagnetic insulator (or Mott insulator). The metal at a critical $\rho$ value ($= \rho'$) just below unity shows the best metallic characteristics [25]. The MIT from a metal at both $\rho'$ and $\kappa = 1$ ($\kappa \rho^2 \neq 1$) to the insulator at both $\rho = 1$ and $\kappa = 1$ ($\kappa \rho^2 = 1$) is abrupt (or a jump); this is the reason for observing the Mott transition near $U/U_c = 1$. Holes corresponding to the difference (critical hole content, $\Delta \rho' = 1 - \rho'$) between $\rho'$ and $\rho = 1$ are induced into a conduction band (the valence band) with free holes by a gate electric field. Note that the valence band for hole carriers and the conduction band for electron carriers are regarded as conduction band (a potential with charges $\geq 0$) in view of conduction. An occurrence of free holes indicates the disappearance of electrons in a band; this causes the decrease in the Coulomb energy from $U_c$ to $U$ (breakdown of Coulomb energy) [8]. Then, the energy gap breaks down and the metallic system becomes inhomogeneous due to the induced holes [8]. The number of induced holes may be $n_c \approx 3 \times 10^{18} \text{cm}^{-3}$ [26], as predicted by Mott from the Mott criterion $n_c^{1/3} a_0 \approx 0.25$ [1]. Here, $a_0$ is the Bohr radius and $n_c$ corresponds to about 0.018% of the number of carriers in the half-filled band, when one electron
in the cell volume, 59.22 \times 10^{-24} \text{ cm}^3, of VO_2 is assumed; \Delta \rho' = 0.018\%. The idea is shown in figure 2.

3. Experiments

Thin films of a Mott insulator VO_2, with a sub-energy gap of about 0.6 eV in the d band [18, 27], have been deposited on Al_2O_3 and Si substrates at the substrate temperature of 400 °C.
in the narrow range of 55–60 mTorr in an Ar + 10% O$_2$ ambient atmosphere by laser ablation [22, 23]. The thickness of each VO$_2$ film is about 900 Å. The resistance of the film decreases with increasing temperature and shows an abrupt MIT near a transition temperature $T_r \approx 340$ K (figure 3(a)). This corresponds to that measured by De Natale et al [28] and Borek et al [29]. It was proposed that the abrupt MIT is the structural phase transition from monoclinic below $T_r$ to tetragonal above $T_r$ [5, 12, 13, 14, 30]. The decrease in the resistance up to 340 K indicates an increase in hole carriers, and two kinds of electron and hole carriers coexist near $T_r = 340$ K (figure 3(b)). From 332 to 340 K, the number of carriers is not discernable because of mixing of carriers from holes to electrons as measured by the Hall effect. A change of carriers from holes to electrons is seen at 332 K. The minus sign indicates that the carriers are holes. (c) Temperature dependence of resistance above 340 K in (a). The resistance is not proportional to either $T$ (evidence of the Peierls picture) or $T^2$ (evidence of the Fermi liquid) with respect to temperature.

**Figure 3.** (a) Temperature dependence of the resistance of a VO$_2$ film deposited on an Al$_2$O$_3$ substrate. Hysteresis is shown. (b) Number of carriers as measured by the Hall effect. A change of carriers from holes to electrons is seen at 332 K. The minus sign indicates that the carriers are holes. (c) Temperature dependence of resistance above 340 K in (a). The resistance is not proportional to either $T$ (evidence of the Peierls picture) or $T^2$ (evidence of the Fermi liquid) with respect to temperature.
Figure 4. (a) Schematic representation of devices 1, 2, 7 and 8 with two terminals. (b) Schematic representation of three-terminal devices 3–6 with thermally treated amorphous SiO\textsubscript{2} of gate insulator. The dotted line between VO\textsubscript{2} and insulator is a channel (or path) where current flows. (c) Electrical circuit scheme of a three-terminal device. \(I_s, I_d\) and \(I_g\) represent the source, drain and gate current respectively. By Kirchhoff’s law, \(I_d + I_g + I_s = 0\) at a node.

We fabricated two- and three-terminal devices (or transistors) to observe the Mott transition. Schematic representations of the devices are shown in figure 4. Devices 1, 7 and 8 with a channel length \(L_{ch} = 5 \, \mu m\) and a channel width \(L_w = 50 \, \mu m\) and device 2 with \(L_{ch} = 3 \, \mu m\) and \(L_w = 50 \, \mu m\) were fabricated on Al\textsubscript{2}O\textsubscript{3} substrates by lithography (figure 4(a)). Devices 3–6 with \(L_{ch} = 5 \, \mu m\) and \(L_w = 25 \, \mu m\) were manufactured on Si substrates. Their structure is shown in figure 4(b). SiO\textsubscript{2} as the gate insulator was thermally treated. It can be seen that an interface between the polycrystal VO\textsubscript{2} film and the amorphous SiO\textsubscript{2} film is not sharp and complicated, and the VO\textsubscript{2} films are inhomogeneous [22]. However, the SiO\textsubscript{2} insulator is strong with respect to a high field for electronic application. Au/Cr electrodes were prepared for Ohmic contact. WSi was used as gate electrode. Electrical characteristics of the devices were measured by a precision semiconductor parameter analyser (HP4156B). In measurements, a short-pulse mode in HP4156B was used to prevent a Joule heat occurring for a long pulse; the pulse width is

electrons and holes. We speculate that the number of hole carriers may be \(n_c \approx 3 \times 10^{18} \, cm^{-3}\), according to the Mott criterion at \(T_{tr} \approx 340\,K\) and on the general basis that an exponential decrease in the resistance with temperature in semiconductor physics indicates an exponential increase in carriers. Generally, in oxide materials, there are holes of about \(5.5 \times 10^{18} \, cm^{-3}\), which corresponds to 0.034\% for d-band charges [31]–[33]. The holes are coupled with optical phonons [31]. In the metal regime above 340\,K, the carriers are electrons (figure 3).
about 64 $\mu$s. To protect the devices from excess current, the maximum current was limited to a compliance (or restricted) current. The temperature dependence in figures 9 and 10 was measured in a cryostat.

A Raman experiment was performed with a micro-Raman system equipped with a CCD detector. A 17 mW Ar ion laser with a 514.5 nm line in a micro-Raman system (Renishaw 2000) having a spectral resolution of $<2$ cm$^{-1}$ was employed. The Raman system was also equipped with an integral microscope (Olympus BH2-UMA). When the Raman spectra were measured, the current after the abrupt MIT was limited to compliance (or restricted) current to protect the device from possible damage due to excess current. Even though the compliance current was extended from 2 to 100 mA (the measurable maximum current of our system), an abrupt current jump was always observed at an MIT voltage $V_{MIT} = 10 \sim 11$ V.

4. Results and discussion

Figure 5(a) shows a plot of the drain–source current density $J_{DS}$ versus the drain–source voltage $V_{DS}$ for Si-based device 3, measured at gate open (two-terminal structure) at room temperature. An abrupt jump and Ohmic behaviour as a characteristic of metal above $J_{observed} \approx 1 \times 10^{5}$ A cm$^{-2}$ are exhibited. Figure 5(b) shows hysteresis loops of five times, which were continuously measured. These loops may be due to Joule heating and are characteristic of the first-order MIT and are reproducible. The loop measurements were carried out with Si-based device 4.

Figure 6 shows Raman spectra measured after the current jump of $V_{MIT} = 10 \sim 11$ V at room temperature with a two-terminal device of Al$_2$O$_3$-based device 1 with $L_{DS} = 5 \mu$m (figure 4(a)). Peaks without current dependence near 400 cm$^{-1}$ are peaks of the substrate Al$_2$O$_3$. Raman peaks below $I_{DS} = 20$ mA in the monoclinic structure of VO$_2$ are not exhibited above a compliance (or restricted) current $I_{DS} = 20$ mA due to screening of free carriers. From $I_{DS} = 0$ to 20 mA, Raman-peak positions of $A_\gamma$ modes do not shift to a lower direction and their bandwidths do not also broaden; this is evidence for the MIT. When a structural phase transition occurs, the most dominant $A_\gamma$ mode in the monoclinic structure near 622 cm$^{-1}$ should shift to a broad $A_{1g}$ of 576 cm$^{-1}$ in the tetragonal structure [12, 13, 14]. The $A_{1g}$ mode is seen slightly at $I_{DS} = 40$ mA and is more clear at $I_{DS} = 50$ mA, as shown in figure 6. This result is explained by the following reasons: firstly, excess current flows after a current jump as shown in figure 5(a) occurs; secondly, Joule heating takes place; thirdly, the temperature of VO$_2$ in the device increases; fourthly, a structural phase transition occurs due to Joule heating; and, finally, the $A_{1g}$ mode of the tetragonal structure is observed. This indicates that the structural phase transition is a secondary effect, and the current jump occurs primarily in the monoclinic structure. Thus we observed a current jump in the monoclinic structure of $I_{DS} = 2$ or 10 mA, which suggests that the abrupt jump in current with electric field is not accompanied by the structural phase transition.

Figure 7 shows a plot of $J_{DS}$ versus the drain–source electric field $E_{DS}$ for Al$_2$O$_3$-based device 2, measured at room temperature. $J_{DS}$ and $E_{DS}$ were obtained from the drain–source current and drain–source voltage, respectively; $I_{DS} = J_{DS}S_{DS}$ (where $S_{DS}$ is the cross section) and $V_{DS} = E_{DS}L_{ch}$. The $J_{DS}$ behaviour below point A (figure 7, inset), as observed by Kumai et al [9] and Boriskov et al [11], is linear from $E_{DS} \approx 0.8$ to 4 MV m$^{-1}$, but is non-linear in the total regime. It was suggested that the linear Ohmic behaviour occurs due to an applied field [11] and an induced current [10], not scattering between hole carriers by an increase in sample
temperature by Joule heating arising from leakage current ($I_{DS}$ near $V_{DS} \approx 0$ V). Oka et al [10] have described well the Ohmic behaviour by a theoretical consideration in terms of a universal Landau–Zener quantum tunnelling based on the Hubbard model. These authors indicated that a transition point from non-conduction to conduction, such as near $E_{DS} \approx 0.8$ MV m$^{-1}$ of the first arrow in the inset, is the Mott transition point [10]. Note that the Ohmic behaviour observed in VO$_2$ is similar to that observed in the Oka et al study [10]. However, since holes were observed below 334 K (figure 3(b)), it is asserted that the carriers for Ohmic behaviour are holes and the number of holes is very small. Thus the Ohmic behaviour is not an intrinsic property of metals and may be due to scattering of only a few carriers existing at room temperature by a weak electric field because the DOS does not depend on the electric field (line 1 in figure 7). Here, the differential

---

**Figure 5.** (a) $J_{DS}$ versus $V_{DS}$ measured with Si-based device 3 at gate open. Ohmic behaviour is seen above $J_{DS} \approx 1 \times 10^5$ A cm$^{-2}$. (b) Hysteresis loops measured with device 4. Measurements were performed five times continuously.
Figure 6. Raman spectra of VO$_2$ as a function of a compliance (or restricted) current of source–drain. Metallic behaviour is shown above $I_{DS} = 20$ mA. A used device with $L_{ch} = 5 \mu$m has the two-terminal structure of the Al$_2$O$_3$-based device 1. The spectra were measured at room temperature with a micro-Raman system at $V_{MIT} = 10 \sim 11$ V in which the current jump occurs. Peaks without the current dependence near 400 cm$^{-1}$ and above are peaks of an Al$_2$O$_3$ substrate.

Conductance $dI/dV$ with unit $\Omega^{-1}$ at negative (positive) voltage reflects the DOS below (above) the Fermi level $E_F$; $dI/dV \approx \int_{-\infty}^{+\infty} N_F(eV) [-\partial f(E + eV, T)/\partial(eV)] dE \approx N_F(eV)$ and $N_F(eV)$ with unit $(eV^{-1} \text{ cm}^{-3})$ is the DOS on the Fermi level and $f(E + eV, T)$ is the Fermi–Dirac function [35]. For a metal, $N_F(eV)$ is constant. Moreover, the non-linear behaviour is semiconduction due to the increase in hole carriers by Zener’s impact ionization and is a doping process wherein $n_c$ of holes are induced by the electric field. This is supported by the fact that the DOS increases exponentially with an increasing electric field (figure 7, line 2). Thus the VO$_2$ film below point A is regarded as a semiconductor.

Abrupt jumps of $J_{DS}$ at point A and the DOS at point dA are shown in figure 7. The measured maximum current density at point B is $J_{\text{observed}} \approx 2 \times 10^5$ A cm$^{-2}$, which is of the order of the current density observed in a dirty metal. Note that the true maximum current density is much higher than $J_{\text{observed}}$, because the measurement was limited to 20 mA. It is believed that Ohmic behaviour exists above 20 mA. The jump of $J_{DS}$ between points A and B.
corresponds to a jump of conductivity $\sigma$, because $J_{DS} = \sigma E$ and $E$ is constant between points A and B; $\sigma \propto n_{\text{electron}}$, where $n_{\text{electron}}$ is the number of electron carriers (above point A) on the Fermi surface. This indicates that the number of carriers discontinuously increases, and the DOS on the Fermi surface jumps from points dA to dB (figure 7). The observed jump of the DOS (or $\Delta J_{\text{observed}}$) is an average over an inhomogeneous measurement region of the maximum true jump $\Delta J_{\text{true}}$, which can be deduced by using the effective mass in the BR picture (see section 5). The jump of the DOS is a typical behaviour of a first-order MIT, which was first theoretically derived by Brinkman and Rice [7]. Note that carriers for current measured in semiconductors and metals are on the Fermi surface. Derivatives $dI/dV$ at respective fields correspond to the DOSs on the Fermi surface (figure 7). Moreover, the abrupt jump was also measured more than 1500 times without breakdown using a two-terminal device. Thus we consider the jumps in figures 5 and 7 to be abrupt MIT (or Mott transition) and not the Peierls transition accompanied by a structural phase transition by an electron–phonon interaction, because a structure at the transition point (point A) is still monoclinic as indicated in figure 6. If Hubbard’s continuous MIT exists, the jump will not be observed and the non-linear behaviour with an increasing field

Figure 7. Left axis: $J_{DS}$ versus $E_{DS}$ for Al$_2$O$_3$-based device 2. Right axis: density of states (DOS), $dI/dV$, on the Fermi surface obtained by derivation with respect to $V_{DS}$. Inset: ohmic behaviour from $E_{DS} = 0.8$ up to 4 MV m$^{-1}$ is shown. The maximum current was limited to compliance current (20 mA) to protect the device.
Figure 8. $I_{DS}$ versus $V_{DS}$ measured with Al$_2$O$_3$-based devices and $L_{ch} = 3 \mu$m at room temperature. As $I_{off-current}$ at points C, D and E increases, MIT-$V_{DS}$ at points F, G and H decreases. To protect the devices, a compliance current was applied.

in the inset of figure 7 will be continuously exhibited from point A to point B in the electron system.

Figure 8 shows $I_{DS}$ versus $V_{DS}$ curves measured at room temperature with Al$_2$O$_3$-based devices and $L_{ch} = 3 \mu$m. The off-current is defined as $I_{DS}$ near $V_{DS} = 0$ V; other researchers refer to off-current as leakage current. With increasing off-current (points C–E), MIT-$V_{DS}$ decreases and MIT-$I_{DS}$ increases (points F–H). The off-current is caused by the excitation of holes in impurity levels such as oxygen deficiency.

Figure 9 shows the temperature dependence of an $I_{DS}$ versus $V_{DS}$ curve determined with Si-based device 5 at gate open. With increasing temperature, MIT-$V_{DS}$ decreases and MIT-$I_{DS}$ increases, as indicated by the transition line. The right panel shows an increase in the excitation of holes with increasing temperature at $V_{DS} = 1$ V, as indicated on the left panel. Thus the increases in the off-current (figure 8) and excitation (figure 9) indicate that the holes were excited.

Figure 10 shows the temperature dependence of the abrupt MIT measured with Al$_2$O$_3$-based device 7. The transition voltage of the abrupt MIT decreases with increasing temperature.
This arises from excitation of hole charges by temperature change. At 338 K, near the transition temperature of the abrupt MIT, and above 338 K, $I_{DS}$ values show Ohmic behaviour without any current jump, in contrast with the MITs with a current jump below 338 K; this is a very important observation. This indicates that the current jump occurs in a nonmetallic phase (or insulator) with the monoclinic structure, as shown below $I_{DS} = 10 \text{ mA}$ in figure 6. Note that the device was protected by a compliance (or restricted) current of 3 mA and the measurement was carried out without external resistance. The temperature dependence in figures 9 and 10 was measured in a cryostat.

The temperature dependence also provides decisive information to understand the mechanism of the abrupt jump. If the total number of holes $n_{tot}$ in the hole levels is given by $n_{tot} = n_b + n_{free}(T, E)$, where $n_b$ is the number of bound holes in the levels and $n_{free}(T, E)$ is the number of holes freed by temperature $T$ and by electric field $E$ from the levels, then $n_b$ decreases with increasing $n_{free}(T, E)$. For the abrupt jump, $\Delta n \equiv n_c - n_{free} = 0$ should be satisfied, where $n_c \approx 3 \times 10^{18} \text{ cm}^{-3}$, as predicted by Mott [1, 26]. At $T_r \approx 340 \text{ K}$, it is
Figure 10. Temperature dependence of abrupt MIT observed with Al₂O₃-based device 7 above room temperature. Above $T = 338 \, \text{K} (65^\circ\text{C})$, apparent metallic behaviour appears. To protect the device, compliance current (3 mA) was applied.

suggested, as decisive evidence for the Mott transition, that the abrupt current jump will disappear, because $n_{\text{free}}(T \approx 340 \, \text{K}, E \approx 0) = n_c$ (i.e. $\Delta n = 0$) is excited only by temperature. Below $T_{tr} \approx 340 \, \text{K}$, the abrupt MIT voltage decreases with increasing temperature, because, from $n_c = n_{\text{free}}(T, E) = n_{\text{free}}(T) + n_{\text{free}}(E)$, the increase in $n_{\text{free}}(T)$ with increasing temperature decreases $n_{\text{free}}(E)$. Thus it is revealed that the mechanism of the abrupt MIT excited by temperature (figure 3(a)) is the same as that by an electric field.

Figure 11 shows $J_{DS}$ as a function of $E_{DS}$ near the abrupt jump of the Si-based device 6 with a gate thickness of 2500 Å. The electrical circuit scheme for a gate current is shown in figure 4(c). A gate current is about $I_{\text{gate}} \approx 3.78 \times 10^{-8} \, \text{A}$ at $V_{DS} = 0.4 \, \text{V}$ and $V_{\text{gate}} = 0 \, \text{V}$. Its characteristics are as follows. Firstly, the gate effect at $E_{DS} = 2.8 \, \text{MV m}^{-1}$ and $V_{\text{gate}} = -10 \, \text{V}$ is due to induced holes and occurs suddenly; this indicates attainment of $n_c$. Secondly, MIT-$E_{DS}$ increases with increasing negative gate field, which is due to a decrease in the conductivity $\sigma_{DS}$; $J_{DS}$ at MIT points may be regarded as constant (see figure 11) and $E_{DS} = J_{DS}/\sigma_{DS}$. This is due to an increase in hole carriers generated by the negative gate fields and indicates an increase in inhomogeneity (injection of holes to the electron system); this seems to follow a continuous curve of $\sigma \propto (m^*/m)^2$ below $\rho'$ in figure 2(a). It was also observed that MIT-$E_{DS}$ decreases with increasing positive gate voltage when the off-current is large. Thirdly, in the metal (electron system) regime over $J_{DS} = 0.9 \times 10^5 \, \text{mA cm}^{-2}$, Ohmic behaviour differs
Figure 11. $J_{DS}$ versus $E_{DS}$ near the abrupt jump of Si-based device 6. Electric field at the MIT increases with increasing negative gate voltage from $V_g = -10\, \text{V}$. Above $J_{DS} \approx 0.9 \times 10^5 \, \text{A cm}^{-2}$, Ohmic behaviour is seen.

from the Ohmic behaviour in the hole system in figure 7 and arises from a dirty interface between the polycrystal VO$_2$ film and the amorphous SiO$_2$. The dirty interface causes resistance [22] and is a channel where current flows (figure 4(b)). We suggest that the Ohmic behaviours in figures 5(a) and 11 are a true metallic characteristic. Finally, the gate-induced abrupt MIT reveals that the MIT depends on the hole carriers in the semiconductor regime, and that inhomogeneity inevitably arises from hole doping [8].

5. Effect of measurement

Compared with the magnitude of the jump $\Delta J \equiv J_{\text{electron}} = J_{\text{observed}}$ at MIT points, the observed current density $J_{\text{observed}} \approx 10^5 \, \text{A cm}^{-2}$, measured with Si-based devices, as shown in figures 5 and 11, is two or three orders less than $J \approx 10^7$–$10^8 \, \text{A cm}^{-2}$ in a pure metal and is much smaller than that observed with Al$_2$O$_3$-based devices. This can be explained by the following two reasons. Firstly, polycrystal VO$_2$ films (figure 4(b)), deposited on an amorphous SiO$_2$, are more inhomogeneous than crystallized films deposited on Al$_2$O$_3$ substrates (figure 4(a)); the
band filling $\rho$ for VO$_2$/SiO$_2$ films is smaller than $\rho$ for VO$_2$/Al$_2$O$_3$ films. The electronic properties of the film-substrate interfacial region affect the electrothermal switching of VO$_2$ thin-film device structures [34]. Secondly, on the basis of $J = \sigma E$ and $\sigma \propto (m^*/m)^2$, $J_{\text{observed}}$ with equation (1) is given as a function of $\rho$ (the extent of the metal region B); as $\rho$ decreases, $J_{\text{observed}}$ decreases. The physical meaning of $J_{\text{observed}}$ is an average over the measurement region (right panel of figure 2(b)) of a true value $J_{\text{true}}$, which is measured in region B of the right panel of figure 2(b) (see [8]). $J_{\text{true}}$ is the magnitude of the true jump, which can be obtained from equation (1) with $\rho = 1$ and $\kappa \neq 1$ [7]. Thus $J_{\text{observed}}$ is the effect of measurement.

Figure 12 shows the external resistance dependence of the magnitude of abrupt current jumps observed for the Al$_2$O$_3$-based device 8. With increasing external resistance, the magnitude of the abrupt current jump decreases and the MIT voltage increases. This can be explained by our model, as shown in figure 2(b), where, in the measurement region, the metal region decreases with increasing external resistance. If the measurement is confined to the metal region in figure 2(b), the magnitude of the current jump might be of the order of $\approx 10^7 \sim 10^8$ A cm$^{-2}$, the current density of a good metal. That is, $I_{\text{DS}}$ observed at 5 K$\Omega$ does not show the characteristic of a current jump, whereas $I_{\text{DS}}$ values measured at 1 K$\Omega$ or less display jumps. This indicates that the observed $I_{\text{DS}}$ varies with an external resistance, even though the intrinsic metal characteristic remains the same.
Thus the observed current density $J_{DS}$ of the order of $\approx 10^5 \text{ A cm}^{-2}$ is an average of the metal region over the measurement region. The average is the effect of measurement. Furthermore, since the VO$_2$ film has electrons and holes, as observed by Hall measurement in figure 3, it is believed that the VO$_2$ film is intrinsically inhomogeneous, although an external resistance effect is excluded. The inhomogeneity is an intrinsic characteristic of a material with an abrupt current jump and was confirmed by high-resolution cross-sectional transmission-electron microscopy [22]. Thus a true current jump cannot be measured in an inhomogeneous system, as has been suggested by the extended BR picture [8].

6. Is VO$_2$ a Peierls insulator?

The abrupt MITs, observed by pressure [36], doping [36] and temperature (figure 3), can arise from doping of hole charges and not by structural phase transitions. Chudnovski [37] suggested that the electrical MIT near 340 K seems to arise from instability of the lattice vibrations. This is different from the hole-driven MIT suggested in this paper. Using the micro-Raman experiment [30], we found that the structural phase transition is not abrupt (or continuous). Okazaki et al [38] revealed that, in the insulating phase, the V 3d band for VO$_2$ is strongly broadened as the temperature is increased, which is attributed to the electron–phonon interaction. Since temperature or pressure affects lattice vibrations, the abrupt MIT and the structural phase transition near 340 K are not distinguishable. This may prove that VO$_2$ is a Peierls insulator. However, on the basis of the Raman data shown in figure 6, the structural phase transition is secondary. Thus the abrupt MIT near 340 K occurs independent of the structural phase transition and is regarded as the hole-driven MIT.

The abrupt MIT occurs between 170 K and $T_n$, as shown in figures 9 and 10, and does not undergo the structural phase transition, which has not been explained by theories on the Peierls picture [15]–[17], [39, 40]. Since the abrupt MIT is a typical characteristic of a strong correlation for a Mott insulator, VO$_2$ is regarded as a Mott insulator and not a Peierls insulator. Moreover, Stefanovich et al [26] suggested that VO$_2$ is a Mott–Hubbard insulator. If we find evidence for the Peierls picture in the observations, it can be examined whether the temperature dependence of resistance is proportional to $T$ in the Peierls picture [39] or to $T^2$ in the Fermi liquid picture. Above 340 K in figure 3(c), the resistance does not follow either $T$ or $T^2$, although Ohmic behaviour is seen in figure 10. The temperature dependence in figure 3(c) may be attributed to inhomogeneity due to multiphases in the VO$_2$ film [22]. Therefore we could not find any evidence for the Peierls picture in this research.

For the structural phase transition and the MIT near 340 K for VO$_2$, there exists the Aronov–Kudinov (AK) model explaining the Mott transition, which can be regarded as the Peierls picture [37, 40]. The AK model is a composite one: it incorporates both the electron–phonon interaction between the lattice and the current carriers and the Coulomb interaction between current carriers at a given lattice site. This model suggests that the MIT is attributed to an instability that arises in the system of Frenkel’s excitons. The excitation energy of the excitons, corresponding to the exciton gap, is proportional to the exciton concentration. If the exciton concentration reaches $T_n/K_e$, where $T_n$ is the transition temperature and $K_e$ is the exciton–phonon interaction constant, then the exciton gap collapses; in other words, a Mott semiconductor–semiconductor phase transition occurs. This transition can stimulate the collapse of the main sub-gap (the entire current gap) and, thus, lead to a transition to a metallic state; this is the Mott transition.
compare the AK model with the hole-driven MIT model (extended BR picture, idea of this research). The sub-gap (or current gap) is formed by the electron–electron interaction in the strongly correlated system and the electron–phonon interaction in the AK model. Frank’s excitons are parametrized in the AK model, but holes in impurity levels are mediated in the extended BR picture. However, the extended BR picture, in contrast with the AK model, establishes exactly the abrupt MIT and does not require the structural phase transition. Results of Raman experiment in figure 6 support the MIT without the structural phase transition.

In addition, there is a negative differential resistance (NDR) observed by a current control method for VO₂ [26]. The NDR appears to be explained by stability criteria \( R_L + dV/dI < 0 \) or \( > 0 \) [41]. The NDR is also regarded as a semiconductor–metal transition resulting from temperature changes observed at \( T_{tr} = 340 \) K, based on measurement of large temperature spikes at the onset of switching in a growing filament [42].

Transition can also occur in all compounds (or semiconductors) having a sub-energy gap (below \( E_{gap} \approx 2 \) eV) with hole levels. This phenomenon, similar to abrupt MIT, is known as a ‘dielectric breakdown’, which causes the breakdown of a device or degradation of the intrinsic characteristics of a material. This may be attributed to a large excess current produced by the abrupt MIT.

7. Conclusion

As predicted by the extended BR picture and Mott’s idea, in the Mott insulator of VO₂, an MIT near \( U/U_c = 1 \) occurs abruptly along with inhomogeneity by semiconduction as a doping process of internal holes of \( n_c \approx 0.018\% \). Hole doping of a low concentration to a Mott insulator causes a breakdown of on-site Coulomb energy from \( U_c \) to a large constant \( U (< U_c) \), as indicated in figure 2.

Acknowledgments

We thank Dr Soo-Hyeon Park at KBSI for Hall-effect measurement and Dr J H Park for fabrication of Si₃N₄ films with CVD. Y S Lim gratefully acknowledges Professors K Kim and S J Lee from the Laboratory of Intelligent Interface at SNU for their support with a micro-Raman system. H T Kim, the leader of this project, developed the concept and wrote the paper. B G Chae and D H Youn deposited VO₂ and BSTO films, performed the transistor fabrication process including Si-based transistors, and measured the \( I–V \) characteristics. K Y Kang prepared the laser-ablation and lithography equipments and generated this project with H T Kim. S L Maeng evaluated transistor characteristics, the shielding measurement system and the Si₃N₄ film fabrication. G Kim contributed a discussion on the difference between the Zener breakdown and the first-order MIT. Y S Lim measured the Raman spectra. Furthermore, this research was performed by a project of Core Researches in the Fields of Information and Telecommunication in ETRI, supported by the Ministry of Planning and Budget.

References

[1] Mott N F 1990 Metal–Insulator Transition (London: Taylor and Francis)
[2] Zhang X Y, Rozenberg M J and Kotliar G 1993 Phys. Rev. Lett. 70 1666
[3] Imada M, Fujimori A and Tokura Y 1998 Rev. Mod. Phys. 70 1039

New Journal of Physics 6 (2004) 52 (http://www.njp.org/)
[4] Fazekas P 1999 Lecture Notes on Electron Correlation and Magnetism (Singapore: World Scientific)
[5] Morin F J 1959 Phys. Rev. Lett. 3 34
[6] Hubbard J 1965 Proc. R. Soc. (Lond) 281 401
[7] Brinkman W F and Rice T M 1970 Phys. Rev. B 2 4302
[8] Kim H T 2002 New Trends in Superconductivity, NATO Science Series vol II/67 (Dordrecht: Kluwer) p 137; online at http://xxx.lanl.gov/abs/cond-mat/0110112
[9] Kumai R, Okimoto Y and Tokura Y 1999 Science 284 1645
[10] Oka T, Arita R and Aoki H 2003 Phys. Rev. Lett. 91 66406
[11] Boriskov P P, Velichko A A, Pergament A L, Stefanovich G B and Stefanovich D G 2002 Tech. Phys. Lett. 28 406
[12] Srivastava R and Chase L L 1971 Phys. Rev. Lett. 27 727
[13] Vikhnin V S, Gonchaurk I N, Davydov V Y, Chudnovskii F A and Shadrin E B 1995 Phys. Solid State 37 1971
[14] Adler D and Brooks H 1967 Phys. Rev. 155 826
[15] Goodenough J B 1969 Phys. Rev. 124 413
[16] Wentzcovich R M, Schulz W W and Allen P B 1994 Phys. Rev. Lett. 72 3389
[17] Rice T M, Launois H and Pouget J P 1994 Phys. Rev. Lett. 73 3042
[18] Pan S H et al 2001 Nature 413 282
[19] Lang K M, Madhavan V, Hoffman J E, Hudson E W, Eisaki H, Uchida S and Davis J C 2002 Nature 415 412
[20] Botto I L, Vassallo M B, Baran E J and Minelli G 1997 Mater. Chem. Phys. 50 267
[21] Remke R L, Walser R M and Bene R W 1982 Thin Solid Films 97 129
[22] Mourachkine A 2002 High-Temperature Superconductivity in Cuprates (Dordrecht: Kluwer) ch 2
[23] Oron A G and Kudinov E K 1969 Sov. Phys.—Tech. Phys. 20 999
[24] Okazaki K, Wadati H, Fujimori A, Onoda M, Murakha Y and Hiroi Z 2004 Phys. Rev. B 69 165104
[25] Wentzcovich R M, Schulz W W and Allen P B 1994 Phys. Rev. Lett. 73 3043
[26] McWhan D B, Rice T M and Remeika J P 1969 Phys. Rev. Lett. 23 1384
[27] Chudnovskii F A 1976 Sov. Phys.—Tech. Phys. 20 999
[28] Jackson J L and Shaw M P 1974 Appl. Phys. Lett. 25 666
[29] Duchene J 1975 J. Solid State Chem. 12 303

New Journal of Physics 6 (2004) 52 (http://www.njp.org/)