Chapter 2

Literature Review

2.1 Theoretical modeling of porous AAO formation

Nanoporous anodic aluminum oxide (AAO) formed by anodization has been widely studied during the last few decades. This nanoscale pattern formation phenomenon by electrochemical reactions has also attracted many theoretical studies to decipher its instability mechanism but without much success. The basic mechanism behind pore formation phenomenon still awaits explanation and a self-consistent theory.

There are two forms of AAO exist, the barrier-type film (blanket oxide) and the pore-type film (porous oxide)[Li 1998-2470]. When aluminum is anodized in a neutral or basic solution, a flat and featureless insulating barrier oxide with few tiny pores forms. When aluminum is anodized in an acid solution, such as sulfuric (H_2SO_4), oxalic ($H_2C_2O_4$), and phosphoric (H_3PO_4) acids, deep nanopores can continuously grow, with diameters varying between ten and several hundred nanometer and lengths up to several microns. Both the barrier-type and pore-type AAO usually start from fairly smooth aluminum surface, but with pits formed at lattice imperfections or by electropolishing[Li 1998-2470]. During the initial stage of the pore formation, the small pits suffer more electric field due to the curvature itself and the ionic current density at the pore bottoms more than between the pores. Thus, the pores start to grow from the pits. The AAO pores grow perpendicular to the surface with an equilibrium of field-enhanced aluminum oxide dissolution at the oxide/electrolyte interface and aluminum oxide growth at the metal/oxide interface[Jessensky 1998-1173]. However, in the barrier-type AAO, any local electric field increase does not encounter the acid-catalyzed oxide dissolution, and thus pores do not grow. A schematic of the electrochemical reactions and ionic paths involved during anodization of aluminum are presented in Figure 2-1[Thamida 2002-240]. At the metal/oxide interface, aluminum ions are formed due to charge transfer or electrochemical reaction:

$$Al_{(s)} \rightarrow Al^{3+} + 3e^{-} \qquad (2.1)$$

The metal/oxide interface is shifted inward in the sample due to interaction of moving

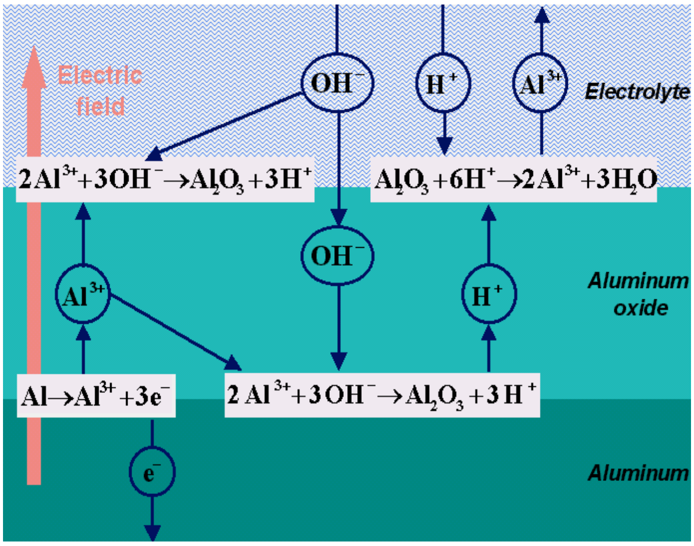

Figure 2-1 Schematic diagram showing the electrochemical reactions and ionic paths involved during anodization of aluminum[Thamida 2002-240].

oxygen species with aluminum:

$$2Al^{3+} + 3OH^- \rightarrow Al_2O_3 + 3H^+ \tag{2.2}$$

At the same time there are aluminum oxide growth at the oxide/electrolyte interface as a result of the outward migration of aluminum ions (mainly by a place-exchange mechanism) and their reaction with oxygen-containing electrolyte species:

$$2Al^{3+} + 3OH^- \rightarrow Al_2O_3 + 3H^+ \tag{2.3}$$

This aluminum oxide is also dissolved as a result of field-stimulated interaction of electrolyte species with the oxide surface:

$$Al_2O_3 + 6H^+ \rightarrow 2Al^{3+} + 3H_2O \tag{2.4}$$

The local electric field is the key variable along the interfaces that determines the reaction rates or moving speeds of the interfaces, and is dependent on the shape or topography of the interfaces. By multiplying the current density with the appropriate faradaic coefficients α_1 and β_1, the oxide/electrolyte interfacial speed (U_e) can be expressed as[Valand 1983-71]

$$U_e = \alpha_1 A_1 10^{-pH} \exp(k_d E_e) - \beta_1 B_1 \exp(k_0 E_e) \tag{2.5}$$

where A_1, B_1, k_d, and k_0 are kinetic constants and E_e is the electric field strength at the oxide/electrolyte interface. The first kinetic term corresponds to the dissolution reaction [Equation (2.4)] while the second term corresponds to the oxidation reaction [Equation (2.3)] at the oxide/electrolyte interface. Both the two ionic reactions are enhanced by the local electric field. On the other hand, the metal/oxide interfacial speed (U_m) is determined by the same dissolution and oxidation reactions as in Equation (2.5) but both contribute to the metal/oxide interface dissolution as[Parkhutik 1992-1258]

$$U_{\mathrm{m}} = \frac{E_{\mathrm{m}}}{E_{\mathrm{e}}}\left[\gamma_1 A_1 10^{-p\mathrm{H}} \exp(k_{\mathrm{d}} E_{\mathrm{e}}) + \gamma_1 B_1 \exp(k_0 E_{\mathrm{e}})\right] \tag{2.6}$$

where E_{m} is the electric field strength at the metal/oxide interface and γ_1 is the faradaic coefficient at the metal/oxide interface. By assuming during initial stages of aluminum oxide formation, the interfaces remain flat and lie perpendicular to the direction of propagation. The oxide thickness (D_{a}) would increase until both interfaces move with the same speed. Since the oxide layer is flat, the electric field is constant throughout the oxide and is given by

$$E_{\mathrm{e}} = E_{\mathrm{m}} = E_{\mathrm{b}} = \frac{V_0}{D_{\mathrm{a}}} \tag{2.7}$$

where E_{b} is the base state electric field and V_0 is the applied voltage. By substituting Equation (2.7) in Equation (2.5) and (2.6) and by equating the two interfacial speeds, base state electric field can be obtained[Thamida 2002-240]:

$$E_{\mathrm{b}} = \frac{1}{k_0 - k_{\mathrm{d}}} \ln\left[\frac{(1-\gamma_1/\alpha_1)\alpha_1 A_1 10^{-p\mathrm{H}}}{(1+\gamma_1/\beta_1)\beta_1 B_1}\right] \tag{2.8}$$

After inserting the kinetic and faradaic coefficients, these results give a critical value of $p\mathrm{H} = 1.77$[Thamida 2002-240]. As the pH value of the electrolyte is near or above the critical value of $p\mathrm{H} = 1.77$, the base state electric field tends to zero and the aluminum oxide barrier layer becomes very thick. The metal/oxide interface then continues to advance and oxidize while the oxide/electrolyte interface slows to a slow dissolution rate. As a result, the oxide layer increases in thickness indefinitely and no pore formation is expected (barrier-type film). Therefore, the porous AAO preparation is always performed in an acidic electrolyte in order to maintain the low pH condition and the electrochemical reaction kinetics. Experimental data from literatures[Woods 1970-1865; Keller 1953-411; Takahasi 1994-677] also support such a transition from blanket oxide layer to nanoporous oxide layer at a particular pH value.

Thompson *et al.*[Thompson 1983-205] assumed the following mechanistic sequence for porous AAO formation: (1) electropolishing or other pretreatment left a slightly scalloped surface of aluminum covered by oxide; (2) a scalloped native barrier oxide film grew over this scalloped surface; (3) as anodization started, pores would start at pits, cracks, and imperfections in the surface, leaving an electric field concentrated below the regions where the oxide film was thinner, thus (4) aiding the local dissolution of oxide; (5) this new pore bottom deepened, and a major pore formed, at the expense of the former shallow pores[Thompson 1983-205]. At the metal/oxide interface the average field across the barrier layer determined the barrier film growth rate. At the oxide/electrolyte interface the local field at the pore bottom, assisted by local heating, determined the oxide dissolution rate[Thompson 1983-205]. The film growth rate was approximately constant and independent of pore bottom curvature, while the dissolution rate increased as the pore bottom curvature decreases. As the pore radius of curvature decreased, the film dissolution rate increased leading to enlarge the pores. If the pore radius became too big, the dissolution slowed and the pores tended to shrink gradually. These two competing processes kept the pore radius constant under the constant anodization conditions.

Parkhutik *et al.*[Parkhutik 1992-1258] presented a kinetic theory for single-pore growth. By assuming a hemispherical pore bottom, they found that the pore radius (R) is a function of the applied voltage:

$$R = -\frac{V_0}{E_{e0}(1-k)\ln(1-k)} \tag{2.9}$$

where E_{e0} is a constant and k is a function of k_0, k_d, α_1, β_1, and γ_1:

$$k = 1 - \left[\frac{(\beta_1/\gamma_1)(k_0/k_d - 1)}{1 + \beta_1 k_0/\alpha_1 k_d}\right]^{1/2} \tag{2.10}$$

and that the pore radius is also a function of pH[Parkhutik 1992-1258]:

$$R = \frac{R_0}{1 - 2.3\,\eta\, p\mathrm{H}\left[\ln\left(\frac{\alpha_1 k_d A_1}{\beta_1 k_0 B_1}\right)\right]^{-1}} \tag{2.11}$$

where R_0 is the pore radius extrapolated to pH = 0 and $\eta = 1$ or 0[Parkhutik 1992-1258]. These equations for R reproduced the experimental linear dependence of the pore radius on applied voltage and also the experimental quasi-linear dependence of pore radius on pH. These models allows us to use them in direct numerical simulation of the real porous AAO formation, but further work is necessary to take into consideration a local increase in the electrolyte temperature and pH value at the pore bottom.

2.2 Synthesis methods of CNTs

There are many methods being developed to synthesize carbon nanotubes

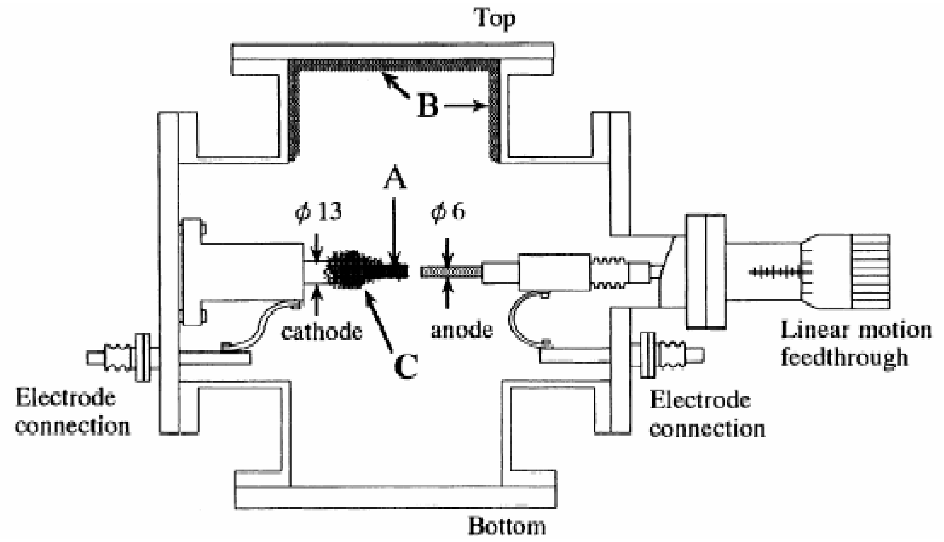

Figure 2-2 Schematic diagram of an arc-discharge system[Saito 2000-169]. (A) Carbonaceous hard deposit grown on the end of the cathode is called "cylindrical hard deposit"; (B) soot grown on surfaces of the cathode is "chamber soot"; (C) soot deposited on the ceiling of the evaporator is "cathode soot".

(CNTs), where arc-discharge, laser ablation, and chemical vapor deposition (CVD) are the most popular methods. In those methods, carbon sources can be in gas or solid phases. The morphology and properties of CNTs are often controlled by manipulating the following process parameters: substrate temperature, precursor gases and gas ratio, catalyst species, pre-treatment conditions, applied bias, etc. However, the proposed methods still suffer the following problems: low yielding, low uniformities in structure and property, low purity, etc.

(a) Arc-discharge:

Arc-discharge method is believed to be the earliest way to synthesize CNTs. When CNTs were found for the first time by Iijima[Iijima 1991-56], it was produced by this method. Figure 2-2 shows the schematic diagram of the arc-discharge system[Saito 2000-169]. There are two graphitic rods as anode and cathode. Between these two electrodes, arcing occurs when dc voltage is applied. In the situation of anode containing small amount of catalyst, such as iron, cobalt, nickel, and yttrium, the single-walled CNTs (SWCNTs) can be synthesized[Bethune 1993-605]. Multi-walled CNTs (MWCNTs) are able to be synthesized in the case of using pure graphite as two electrodes. Generally, the discharge is operated at a voltage range between 20 and 40 V with current from 40 to 100 A under helium or argon atmosphere of 10-500 Torr. Carbon clusters collided out from the anodic graphite rod by electron bombardment are deposited on the cathode surface. The production on the cathode may include amorphous carbon, fullerenes (C_{60}),

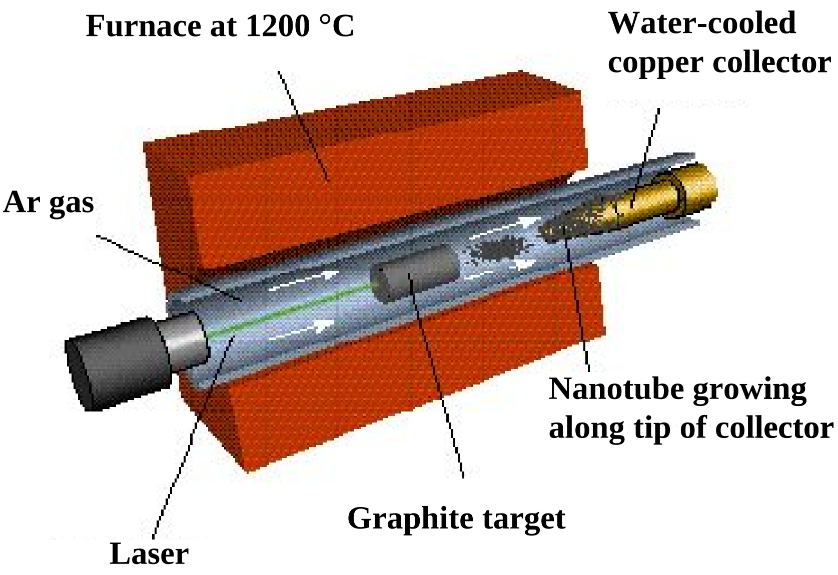

Figure 2-3 Schematic diagram of a laser ablation system[Guo 1995-49].

carbon cluster, CNTs, and varieties of other carbon structures. Therefore, purification of the nanostructures is generally an important issue for further applications. Another drawback of this process is its low yielding in producing CNTs.

(b) Laser ablation:

Laser ablation method was first reported by Guo's group in 1995[Guo 1995-49], as shown schematically in Figure 2-3. There is an incident laser beam for vaporizing graphite target under helium or argon gas atmosphere at pressure of 500 Torr. The productions are swept out by the flowing gas and to be deposited on the water cooled collector. Therefore, it is also called laser vaporization method. The graphite target containing cobalt, nickel, iron, or yttrium is a more favor condition to form SWCNTs.

(c) Chemical vapor deposition:

CVD method is a mature technique in thin film processes. Most of films can be fabricated by CVD method, including metals, semiconductors, and insulators. By using the CVD method, arrays of CNTs can be produced by decomposing of carbon-containing gaseous species (such as CH_4, C_2H_2, C_2H_4, C_6H_6, CO, etc.) on various substrates that contain transition metal catalysts. It is essential for the CVD process to introduce some forms of the energy to decompose precursor gases and deposit the reaction product on the substrate surface. The introduced energy may include thermal, microwave, rf power, or others. Thus, it gives rise to different process names, depending

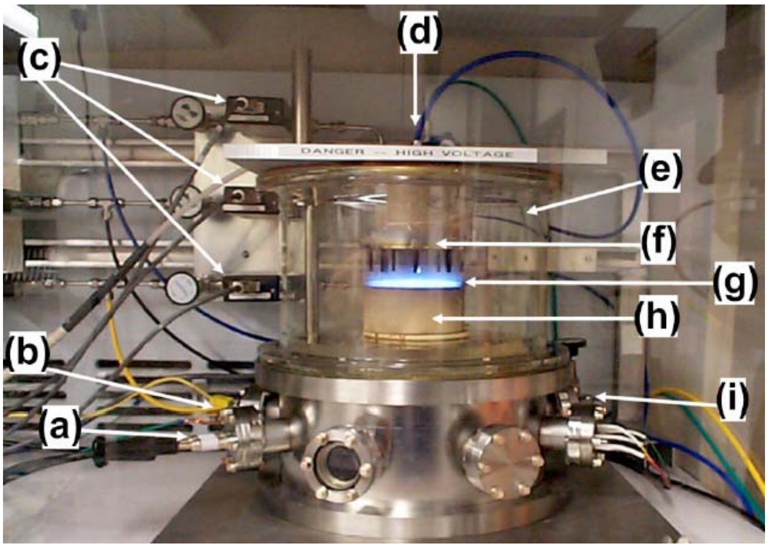

Figure 2-4 A photograph of a dc-PE-CVD reactor for the growth of carbon nanotubes or nanofibres [Melechko 2005- 041301]. (a) High current heater wiring; (b) thermocouple wiring; (c) mass flow controllers for precursor gases; (d) gas inlet; (e) glass cylinder vacuum chamber; (f) gas showerhead and anode; (g) cathode glow of plasma above a silicon wafer; (h) substrate heater and cathode; and (i) pressure transducer.

on the source of the applied energy. Various CVD processes to synthesize CNTs have been proposed by many researchers, including microwave plasma enhanced CVD (MPE-CVD)[Tsai 1999-3462], electron cyclotron resonance CVD (ECR-CVD)[Lin 2002-922], inductively coupled plasma CVD (ICP-CVD)[Delzeit 2002-6027], rf plasma enhanced CVD (rf-PE-CVD)[Kato 2004-2], dc plasma enhanced CVD (dc-PE-CVD)[Hofmann 2004-1171], thermal CVD[Lee 2000-3397], hot filament CVD (HF-CVD)[Yang 2004-433], etc. Figure 2-4 shows a color photograph of a dc-PE-CVD system for CNTs growth[Melechko 2005-041301]. Compared with arc discharge and laser ablation methods, CVD enhanced by plasma is a cheaper way (lower reaction temperature) and the reaction processes can be easily controlled. Moreover, it is superior to other methods in respect to purity, yield, and controlled alignment, thus current attention has focused on developing new techniques for the preparation of vertically aligned CNTs by using the CVD methods.

2.3 Growth mechanisms for catalyst assisted CNTs

Two different growth modes of catalyst assisted CNTs have been observed: tip-growth mode, in which the catalyst particle is detached from the substrate and remains at the tip of the growing tube, and base-growth mode, in which the catalyst remains on the substrate. Regarding the use of CVD methods for CNT growth, many growth mechanisms have been proposed. Most of the mechanisms are based on the

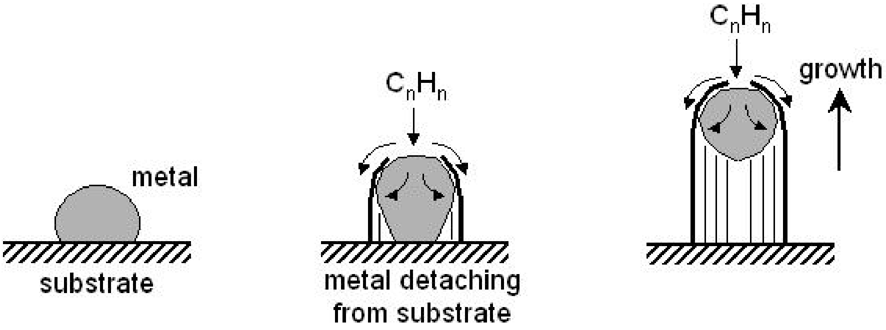

Figure 2-5 Schematic representation proposed for CNT growth from the pyrolysis of hydrocarbons on a metal particle[Ebbesen 1997-56].

original model of carbon nanofibres proposed by Baker[Baker 1989-315]. However, most models were proposed without sufficient and systematic supporting experimental evidence, and they often lacked details about the physical mechanisms and the effects of various process parameters. Thus, the kinetics of CNT nucleation and growth are not well known yet.

(a) Tip-growth CNTs:

When metal catalyst exists, CNTs are formed by pyrolysis of carbon-containing precursors on the isolated metal nanoparticles. Figure 2-5 is schematic representation proposed for a tip-growth CNT from the pyrolysis of hydrocarbons on a metal particle[Ebbesen 1997-56]. The precursors decomposed on the front exposed surfaces of the metal particles to release hydrogen as gas and carbon atoms, which dissolve in the particles. The dissolved carbon atoms diffuse through the particle and are precipitated at the trailing end from the body of the tube. This mechanism depends on the temperature gradient across the catalyst particle due to the exothermic decomposition of precursors. Since the solubility of carbon in a metal is temperature dependent, precipitation of excess carbon will occur at the colder zone behind the particle and initiate the CNT growth. Such a process will continue until the leading tip of the catalyst particle is poisoned or covered by carbon completely. Support for this diffusion model comes from experiments on the kinetics of growth of carbon filaments from acetylene at 1000 ^{o}C which yield an activation energy of 140 kJ/mol. The concentration of carbon in the iron catalyst particle is probably close to that at the eutectic, and the activation energy for diffusion is 133 kJ/mol[Tibbetts 1983-666].

Recently, Helveg *et al.*[Helveg 2004-426] presented *in situ* high-resolution transmission electron microscope (HRTEM) observations of the formation of carbon nanofibres from

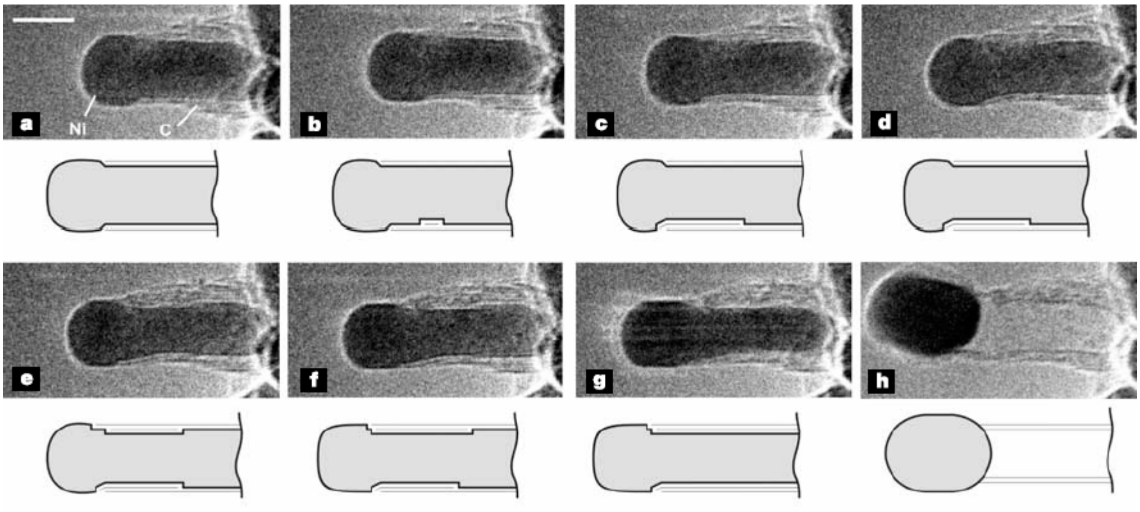

Figure 2-6 Image sequence of a growing carbon nanofibre[Helveg 2004-426]. The images are acquired *in situ* with CH_4:H_2 = 1:1 at a total pressure of 2.1 mbar with the sample heated to 536 °C.

CH_4 decomposition over supported nickel catalyst. Figure 2-6 shows image sequence of a growing carbon nanofibre[Helveg 2004-426]. In these pictures, the dissociative CH_4 adsorption is facilitated at the step edges and carbon atoms adsorb preferentially at the step sites. The graphene layer forms at the terrace between the steps on the curved and dynamically changing catalyst nanoparticle surface. The process involves surface diffusion of carbon and nickel atoms from the step edge, and includes the breaking of the carbon bond to the nickel step on the free surface, incorporation under the graphene sheet, and diffusion at the graphene-nickel interface. Moreover, their observations and calculations suggest that it is not necessary to include the bulk diffusion of carbon through the nickel particle, however, they do not eliminate such possibility.

(b) Base-growth CNTs:

Lee *et al.*[Lee 2000-3397] have proposed a growth model for the base-growth CNTs with bamboo-like shape. The schematic diagrams of this model are shown in Figure 2-7[Lee 2000-3397]. First, carbons produced from the decomposition of C_2H_2 molecules adsorb on a catalytic particle. The carbons diffuse via the surface and/or bulk of the metal particles, forming graphite sheets as a cap on the catalytic particles, as shown in Figure 2-7(a). Once the formation of the graphite sheet starts, the diffusion of carbons probably accelerates into the reaction zone of the catalytic particle, thus the carbons are continuously added to the edge of the cap. As the cap lifts off the catalytic particle, a closed tip with the inside hollow is formed [see Figure 2-7(b)]. The motive force departing from the catalytic particle may be the stress accumulated under the graphite

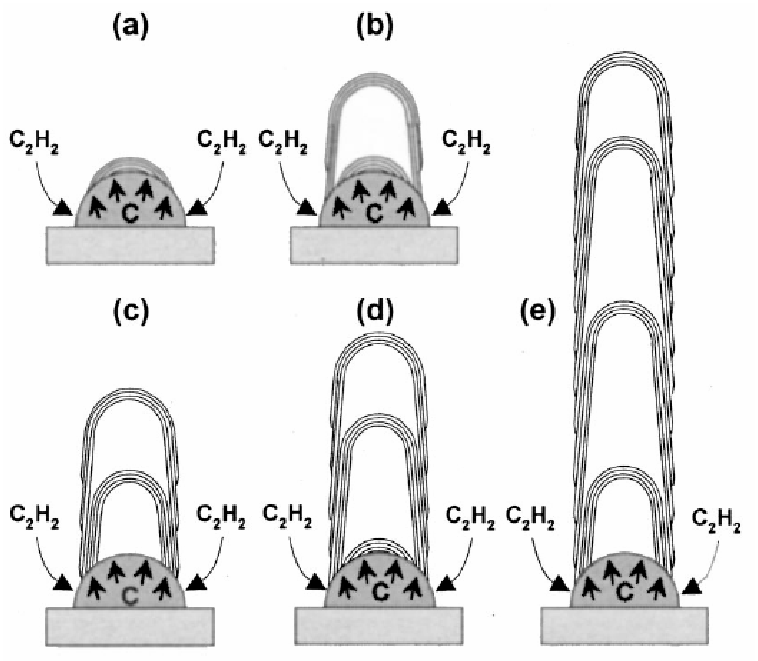

Figure 2-7 Schematic diagrams of the base-growth model[Lee 2000-3397].

cap. The wall would grow toward the vertical direction with a certain angle and the outer graphite sheets disappear. The accumulation of carbons at the inner surface of the catalytic particle occurs mainly via bulk diffusion, which produces the compartment graphite sheets. The compartment graphite sheets grow by forming the joint with the wall. As the wall grows further, the compartment layer departs from the catalytic particle due to stress [see Figure 2-7(c)]. Next, the compartment layer grown on the catalytic particle is again connected with the wall [see Figure 2-7(d)]. The shape of the compartment layers is probably influenced by the surface geometry of the catalytic particle. When the carbons are supplied continuously under steady-state condition, the compartment layers can appear periodically [see Figure 2-7(e)]. Wetting ability or adhesion of the catalyst with the substrate and temperature gradients across the catalyst particle are considered to be the most important factors that determine the CNT growth mechanisms (either tip- or base-growth)[Baker 1989-315].

2.4 Nanofabrications of well aligned CNT arrays

Alignment of CNTs is particularly important to enable both fundamental studies and field emission applications. Vertical alignment generally means that CNTs are oriented perpendicular to the substrate. Various methods for production of aligned arrays of CNTs have been demonstrated. Most frequently found in the literature are

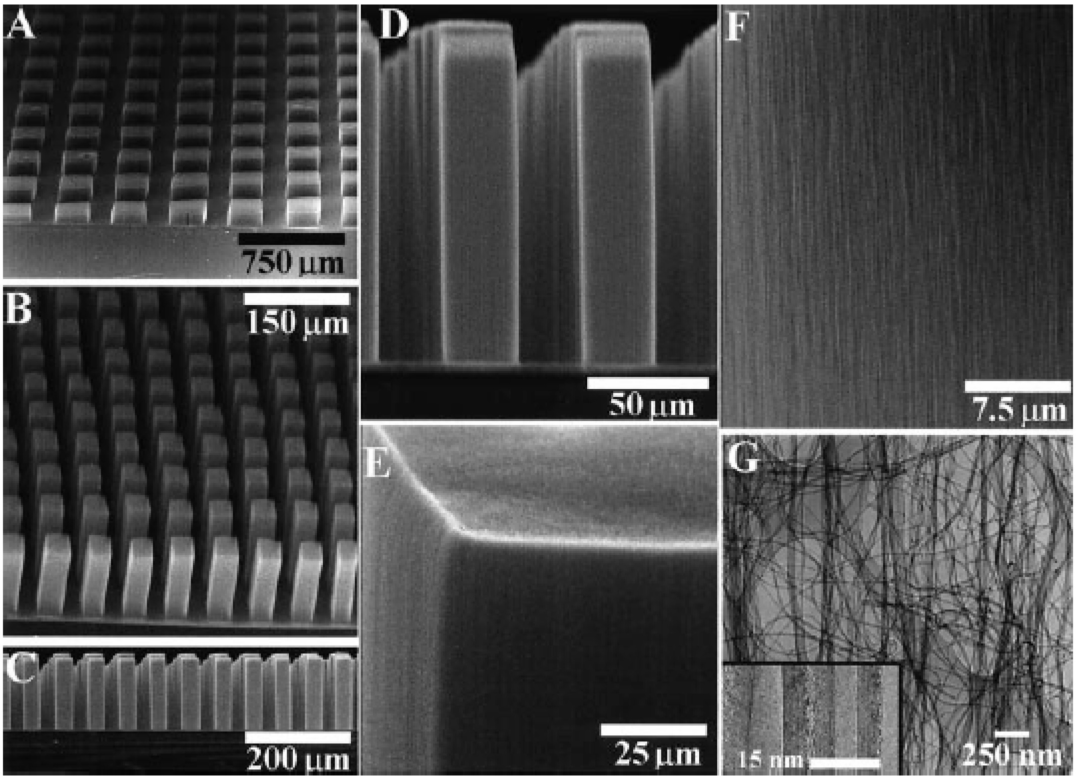

Figure 2-8 Electron micrographs of self-oriented nanotubes synthesized on n^+-type porous silicon substrates[Fan 1999-512].

nanotubes grown in a very dense organization with vertical alignment due to the crowding effect. Fan *et al.*[Fan 1999-512] used porous silicon substrates with evaporated iron catalyst pattern to produce nanotube blocks by thermal CVD that grew perpendicular to the substrate. As shown in Figure 2-8, the high density of nanotubes within each block confined the nearest neighbors and attracted the outermost nanotubes to their neighbors via van der Walls forces, resulting in the oriented growth.

Bower *et al.*[Bower 2000-830] showed that nanotubes can be grown on contoured surfaces and aligned in a direction always perpendicular to the local substrate surface by using the MPE-CVD. The alignment is primarily induced by the electrical self-bias field imposed on the substrate surface from the plasma environment. It is found that switching the plasma source off effectively turns the alignment mechanism off, leading to a smooth transition between the plasma-grown straight nanotubes and the thermally grown “curly” nanotubes, as shown in Figure 2-9[Bower 2000-830].

Merkulov *et al.*[Merkulov 2001-2970] proposed a model to describe the alignment mechanism of CNT growth in a PE-CVD process. They suggested the alignment as the result of a feedback mechanism associated with a non-uniform stress created across the interface of the catalyst particle with the nanotube by the plasma-induced electrostatic

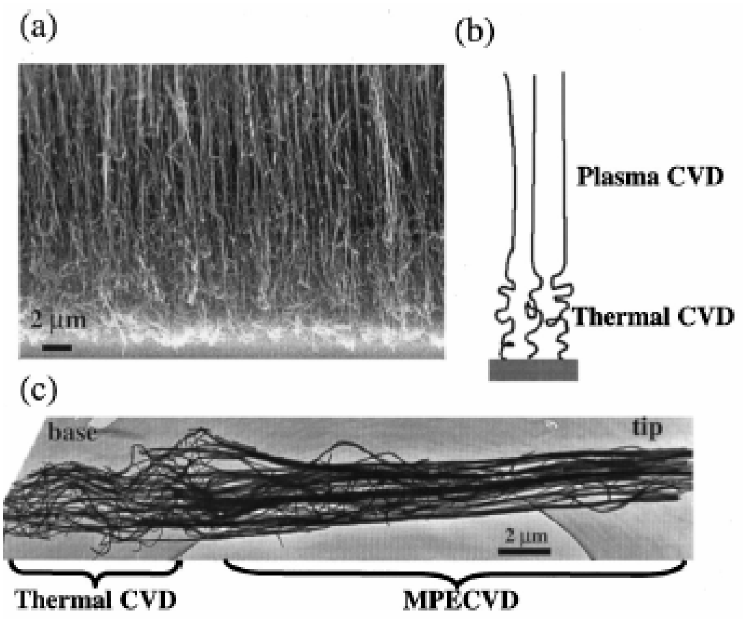

Figure 2-9 (a) An electron micrograph and (b) a schematic showing the straight/curled nanotube structure produced by an alternating plasma and thermal process. (c) is an electron micrograph showing a bundle of nanotubes with the upper portion straight and the lower portion curled[Bower 2000-830].

forces (see Figure 2-10)[Merkulov 2001-2970]. The axis of a nanotube growing perpendicular to the substrate corresponds with the direction of the applied electrostatic force leading to a uniform tensile stress across the entire nanotube/catalyst interface, as shown in Figures 2-10(a) and (b). Accordingly, carbon precipitates uniformly across the interface and the tube continues to grow vertically. On the other hand, if there were a spatial fluctuation in the carbon precipitation at the interface, nanotube growth would deviate from vertical alignment, as shown in Figure 2-10(c) and (d). In the case of nanotubes growing from the tip (tip-growth mechanism), the electrostatic force produces a compressive stress at the part of the catalyst/nanotube interface where the greater rate of the growth is seen [Figure 2-10(c)]. Similarly, a tensile stress is produced at the part of

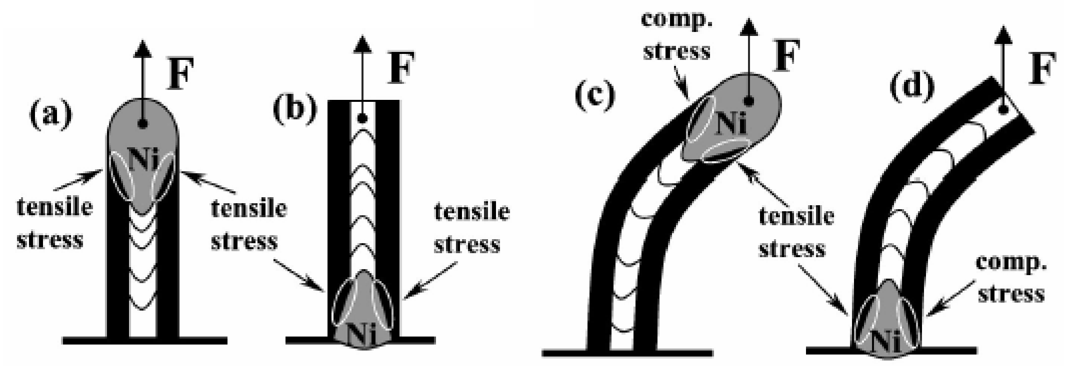

Figure 2-10 Alignment mechanism based on stress-dependent growth rate and stress distribution caused by interaction of nanotube with electric field[Merkulov 2001-2970].

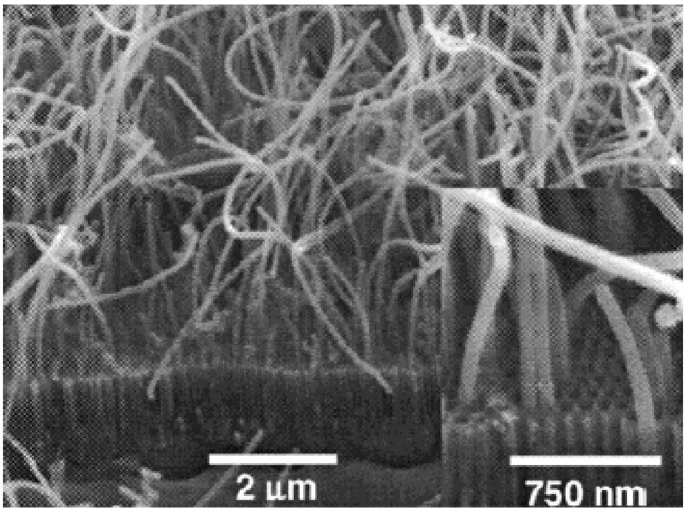

Figure 2-11 Electron micrographs of the cobalt-catalyzed CNTs grown out of the AAO template pores[Jeong 2002-1859].

the catalyst/nanotube interface where the lesser rate of growth is seen. It is proposed that these opposing stresses favor subsequent carbon precipitation at the interface experiencing tensile stress. The net result is a stable, negative feedback that acts to equalize the growth rate around the entire periphery of the catalyst/nanotube interface, leading to vertically aligned CNT growth[Merkulov 2001-2970]. The difference in the growth rates at the two boundaries may be caused by stress-induced diffusion[Laudon 2001-201] due to the stress gradient in the catalyst particle and possibly by the variation in the stress-dependent sticking of diffusing carbon atoms to the carbon side of the nickel-carbon interface. When the catalyst particle is located at the base of the CNT (base growth mechanism), however, the situation with the preferred location of carbon precipitation is different. Since the nanotube base is attached to the substrate, the stress created at the catalyst/nanotube interface with the greater growth rate is tensile [Figure 2-10(d)] and acts to continue the increased growth rate, thus causing the CNT to bend even further[Merkulov 2001-2970].

The most direct method of aligned growth of CNTs is to use a nanoporous AAO template. AAO template assisted growth of vertically aligned CNT arrays has been accomplished by several research groups[Kyotani 1996-2109; Li 1999-367; Iwasaki 1999-2044; Suh 1999-2047; Yao 2001-11395; Bae 2002-277]. To make a comprehensive survey, the AAO templated CNT growth can be divided into two methods. One is to pre-deposit metal catalyst into the nanopores, and the other is to directly grow CNTs inside the nanopores without metal catalyst. For the former approach, the obtained CNTs possess crystallinity as good as the CNTs grown on a silicon wafer by a conventional CVD process, however, precise

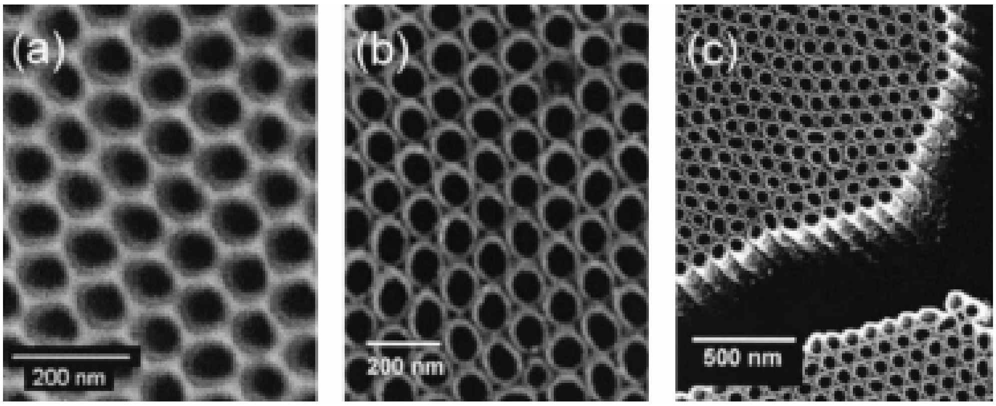

Figure 2-12 (a) Electron micrographs of a highly order CNT array grown on AAO after ion milling, (b) top view, and (c) side view of the CNT array after the AAO template was heavily etched in HF[Gao 2003-5602].

control of the length of aligned CNTs has met with little success[Lee 2001-2387; Jeong 2002-1859]. As shown in Figure 2-11[Jeong 2002-1859], the relatively high tube growth rate leads to overgrowth and entanglement of the cobalt-catalyzed CNTs without vertical alignment. When the AAO template is used alone for the fabrication of CNTs, the obtained nanotubes have very uniform diameter and length, which are the same dimensions of the nanopores in the AAO template, as shown in Figure 2-12[Gao 2003-5602]. However, these aligned CNTs have very poor crystallinity[Kyotani 1996-2109; Sui 2001-1523]. More efforts are necessary for preparing highly graphitized CNTs in the AAO template with suitable tube length.

2.5 Nanofabrication technologies of nanodot arrays

The formation of arrays of quasi zero-dimensional (0-D) nanostructure, i.e., nanodots, nanoislands, or quantum dots, on solid substrates has attracted considerable attention because of their unique physical and chemical properties and potential applications in optoelectronic, magnetic, and quantum devices. However, fabrication of the ordered nanostructures at surfaces still remains a difficult challenge due to the limit of lithographic and semiconductor processing techniques. As a consequence, many technologies have been proposed to provide alternative means for the fabrication of the 0-D nanostructure arrays that are mainly based on either the principle of self-organization or the template methods.

(a) Stranski-Krastanow growth:

One of the particularly promising approaches for fabricating large numbers of

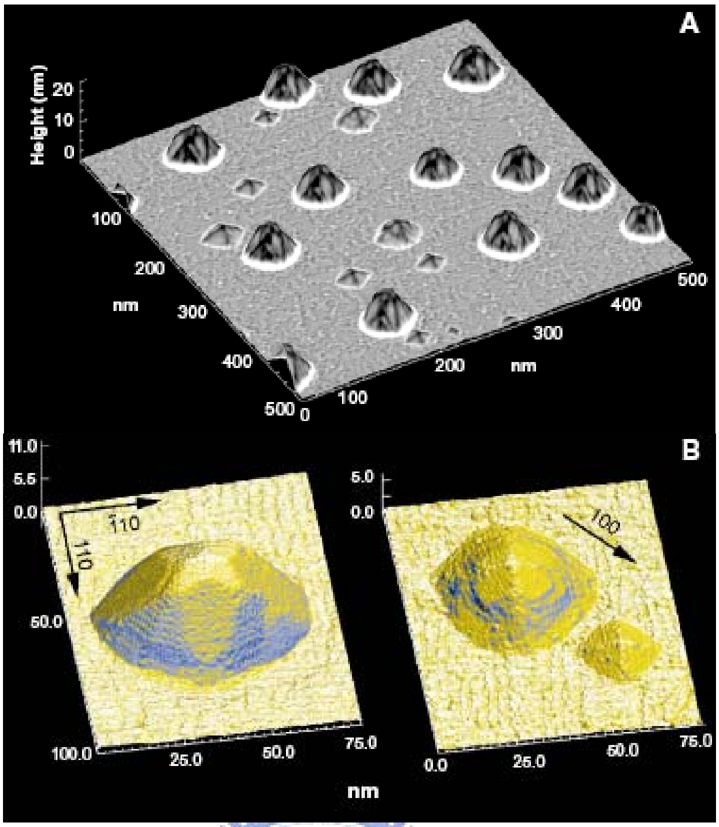

Figure 2-13 (A) STM topograph of strained germanium nanocrystals on silicon (001) substrate. (B) Higher magnification images of the nanocrystals: (left) a mature dome and (right) a nanocrystal entering the transition stage and a small pyramid[Gilberto 1998-353].

nanodots for device applications is via the spontaneous self-organization of nanoislands which occurs during Stranski-Krastanow (S-K) growth. During S-K growth, a growing epitaxial film becomes morphologically unstable as the thickness of the film exceeds a critical value (several monolayers), which results in the formation of nanoislands or nanocrystals on a substrate. The instability of a planar interface is driven by the elastic strain energy casing by the lattice mismatch or misfit between the epitaxial film and substrate[Eaglesham 1990-1943]. Therefore, S-K growth is constrained to a few material systems and size ranges in which the critical strain-mismatch conditions can be met (e.g., Ge/Si, InAs/GaAs, InGaAs/GaAs, InAs/InP, PbSe/PbTe, etc.)[Eaglesham 1990-1943; Ross 1999-1931; Springholz 1998-734]. Figure 2-13 shows the scanning tunneling microscopy (STM) images of strained germanium nanocrystals on silicon (001)[Gilberto 1998-353]. The nanoisland arrays can then be covered with another material to create the optoelectronic devices of interest. However, manipulation of spatial arrangement and size distribution of the nanoislands remains a major problem, which is generally required for practical applications. In order to control the spatial distribution, many efforts have been developed using a variety of techniques, such as selective growth on pre-patterned

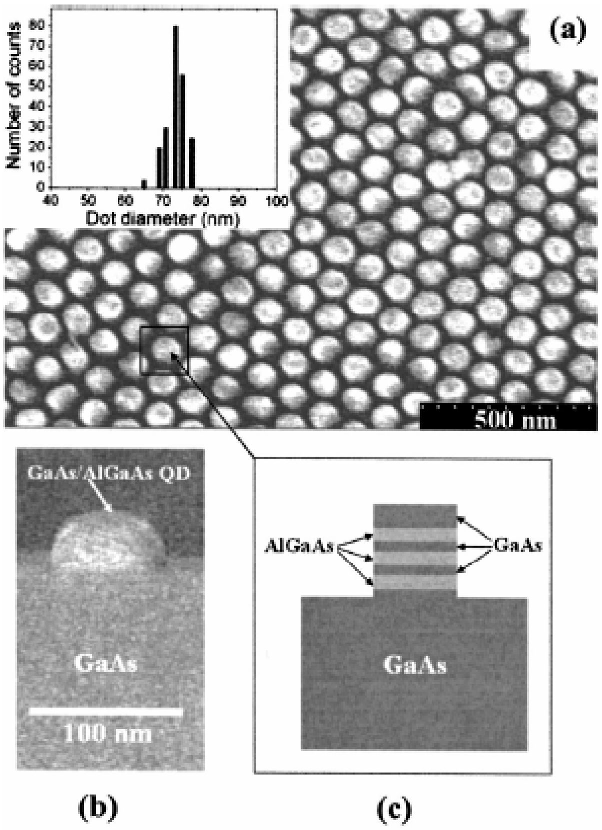

Figure 2-14 (a) Top-view SEM image of a double-well GaAs/AlGaAs quantum dot array grown on GaAs (001) surface after removing the AAO template; (b) cross-sectional SEM image of a typical dot; and (c) schematic illustration of layer structure in each dot[Mei 2003-967].

substrates with stripe mesas[Jin 1999-2752; Kitajima 2002-497] or on relaxed templates with dislocation networks[Leroy 2002-3078]. However, it is still far from the desired control of the width of the mesas or the periodicity of the dislocations.

(b) AAO templated deposition:

Nanoporous AAO formed by anodization of aluminum in an acidic solution has been shown to be capable of yielding highly ordered nanochannels arrays with hexagonal lattice pattern. The nanoporous AAO membranes can be removed from their host aluminum substrates using a lift-off process, and then be bonded to desired substrates to work as template for generating well ordered nanostructures. By using the nanoporous AAO as a deposition mask, highly ordered arrays of nanosized dots of various materials can be readily fabricated, including Au, Ag[Masuda 2000-1031], Ni, Si[Liang 2002-2544], InAs[Liang 2004-5974], GaAs[Mei 2002-361], etc. Figure 2-14 shows the scanning electron microscopy (SEM) images of a double-well GaAs/AlGaAs quantum dot array grown on GaAs (001) surface by molecular-beam epitaxial (MBE) growth after removing the AAO template[Mei 2003-967]. As shown, the dot density is in excess of 10^{10}

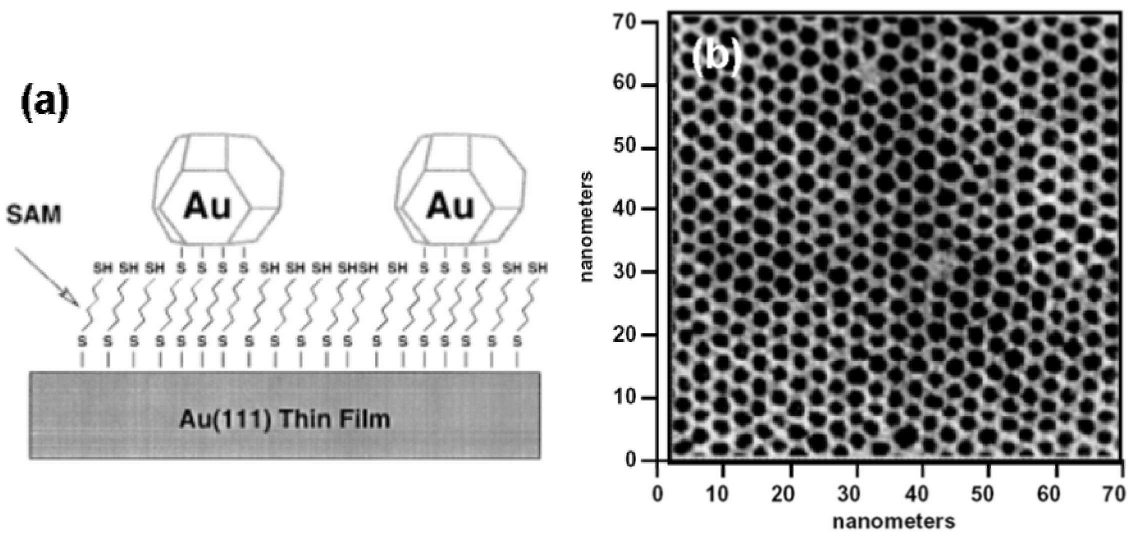

Figure 2-15 (a) Schematic side-view of gold nanoparticles attached to a SAM by molecular wires. (b) Top-view SEM image of a self-assembled array of gold nanoparticles. Each dark dot represents a gold particle[Derogi 1995-9071].

cm^{-2}, comparable to that for the S-K growth mode. This process involves several advantages: each dot has almost identical size and spacing, and the size and the spacing can be easily controlled by changing the anodization conditions of the AAO template. However, the lift-off process of AAO from aluminum host is difficult to control and incompatible with typical device processes. It is necessary to develop the techniques of direct formation of a regular AAO nanopore array on a solid substrate.

(c) Chemical self-assembly:

Chemical self-assembly is the spontaneous orientation of a number of molecules into an energetically favored supra-molecular structure without human intervention. Naturally occurring self-assembled systems of complex supra-molecular structures include deoxyribonucleic acid (DNA), cell membranes, chromophores, or even viruses[Kuhn 1994-33; Seeman 1991-149]. The autonomous components of organic molecules modified a substrate with monolayers via self-assembly process was thus named as self-assembled monolayers (SAMs). The principal driving force for formation of these monomolecular films is specific interactions between the molecule head group and the substrate surface. Depending on the structure of the molecule, these films can be disordered or well-packed, resembling the organization of crystal. Derogi *et al.*[Derogi 1995-9071] have adopted the SAM method to attach individual gold nanoparticles on top of the SAM by using self-assembling molecules with functional groups on both ends, as illustrated in Figure 2-15(a). As shown in Figure 2-15(b), this technique creates two-

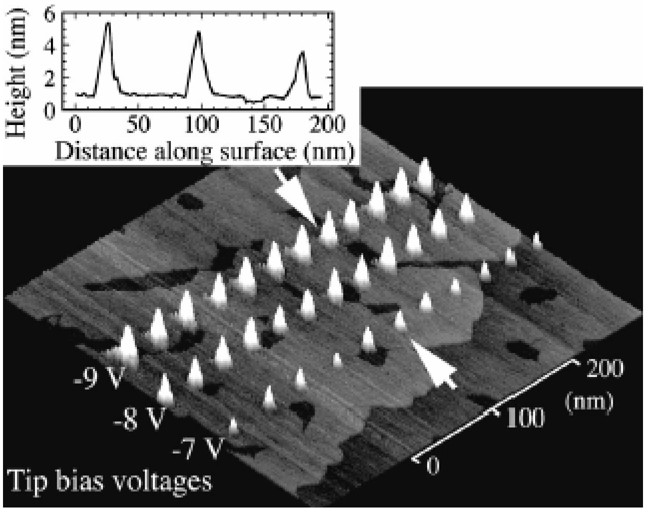

Figure 2-16 STM image of germanium islands on the silicon (111) substrate covered with 2.9 bilayer of germanium at 450 °C. Inset shows the height profile between white arrows[Shklyaev 2000-1397].

dimensional (2-D) arrays of regularly spaced gold nanoparticles. Molecular-scale circuits using such arrays of gold nanoparticles might be fabricated in this manner. The arrays are manufactured on the nanometer scale with minimum feature sizes around 2 nm[Derogi 1995-9071].

(d) SPM-induced growth or oxidation:

Scanning probe microscope (SPM) is continuously being developed for surface modification and for the investigation of different surface properties. The ability of SPM to initiate the transfer and accumulation of atoms for nanostructure fabrications is a topical aspect for further development. Nanosized islands have been grown on a sample surface by accumulating atoms from the surrounding area through directional surface diffusion initiated by the electric field of an atomic force microscope (AFM) or STM tip. Figure 2-16 shows the STM image of germanium islands on the silicon (111) substrate created with STM at a constant tunneling current of 0.3 nA and various negative tip bias voltages[Shklyaev 2000-1397]. As shown, the nanosized islands grew with a constant rate determined by the tip-sample bias voltage. Furthermore, SPM-induced local field oxidation is another feasible process for nanostructure fabrications. The local field oxidation process is similar to conventional electrochemical anodization except that an AFM/STM tip is used as the cathode and water from the ambient humidity is used as the electrolyte[Gwo 1999-1090]. García *et al.*[García 1999-1898] have reported the fabrication of high-areal-density oxide dot arrays on a hydrogen-passivated silicon substrate using non-contact AFM and Cooper *et al.*[Cooper 1999-3566] have reported the case of oxide array with an areal density of 1.6 Tbit/inch^2 on a titanium film using a SWCNT grown on an AFM cantilever. However, the SPM-based growth is poor efficiency for producing spatially dense nanodot arrays making it less attractive than the technique of self-organized nanodot fabrication.

2.6 Basic field emission theory

Electron field emission is a quantum-mechanical tunneling phenomenon of electrons extracted from a conductive solid surface, such as a metal or semiconductor, into vacuum in response to a very high electric field. Field emission has several advantages over thermionic emission. First, the emitters do not need to be heated. The energy spread of the emitted electrons is also far smaller, and the emitted current can be readily controlled with the applied voltage[Bonard 2002-1715]. It is therefore not surprising that thermoelectronic emission is replaced by field emission in various applications, e.g., displays and microwave tubes.

When a sufficient electric field is applied on the emitter surface, electrons will tunnel through the surface-potential barrier across the metal-vacuum interface and escape from the emitter surface, even under a very low temperature. This potential barrier is called the work function which corresponds to the potential difference between Fermi level (E_F, energy level of the free electrons) of the emitter and the field-free vacuum. Work function is a surface property of material and correlates with the electronic structure and orientation of the crystal plane (typically 4-5 eV for commonly used materials). For an atomically smooth molybdenum surface, the work function is 4.36 eV for the (112) plane and 4.95 eV for the (110) plane[Lin 1999-6]. Impurities and adsorbates on the vacuum/emitter interface will also change the electronic structure of the surface and thus modify the work function. In order to obtain high emission currents at low voltages, one desires to have an as low as possible work function.

The potential barrier is square when no electric field is present. If an external electric field is applied, the potential barrier at the surface becomes thinner and the vacuum energy level is reduced as shown schematically in Figure 2-17[Cheng 2003-1021]. The top of the barrier is rounded due to image force. The decrease in barrier width allows the electrons having energy less than vacuum level to tunnel through. The decrease in barrier height ($\Delta\varphi$) can be descried by the following relation:

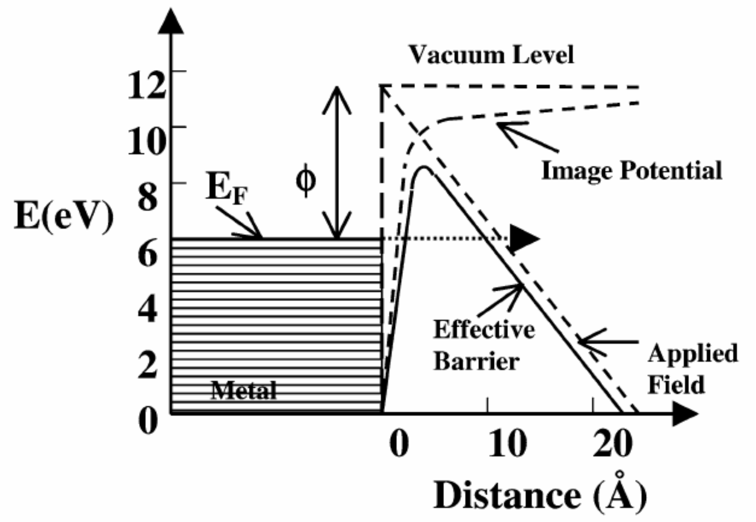

Figure 2-17 Potential-energy diagram illustrating the effect of an external electric field on the energy barrier for electrons at a metal surface[Cheng 2003-1021].

$$\Delta\varphi = \left(\frac{qE}{4\pi\varepsilon_0}\right)^{1/2} \tag{2.12}$$

where φ is the work function, E is the applied electric field, q is electron charge, and ε_0 is permittivity of free space. The applied electric field decisively affects the barrier height and thus the emitted current. Fowler and Nordheim[Fowler 1928-173] in 1928 derived a relationship between the field-emission current density (J) from a pure metal and the applied electric field. The widely quoted Fowler-Nordheim (F-N) equation can be expressed as[Fowler 1928-173]

$$J = \frac{AE^2}{\varphi t^2(\mathrm{y})}\exp\left[-\frac{B\varphi^{3/2}v(\mathrm{y})}{E}\right] \tag{2.13}$$

where $A = 1.54\times10^{-10}$ (A V^{-2} eV), $B = 6.83\times10^{9}$ (V $eV^{-3/2}$ m^{-1}), $\mathrm{y} = 3.7947\times10^{-4}E^{1/2}/\varphi$, $t^2(\mathrm{y}) \sim 1.1$, and $v(\mathrm{y})$ can be approximated as

$$v(\mathrm{y}) = \cos(0.5\pi\ \mathrm{y}) \quad \text{or} \quad v(\mathrm{y}) = 0.95 - \mathrm{y}^2 \tag{2.14}$$

Generally, the field emission current (I) is measured as a function of the applied voltage (V). Substituting relationships of $J = I/\alpha$ and $E = \beta V/d$ into Equation (2.13), where α is the emitting area, β is the local field enhancement factor at the emitting surface, and d is the distance between cathode and anode, following equation can be obtained:

$$I = \frac{A\alpha\beta^2V^2}{d^2\varphi t^2(\mathrm{y})}\exp\left[-\frac{Bd\varphi^{3/2}v(\mathrm{y})}{\beta V}\right] \tag{2.15}$$

In addition, the Napierian logarithm form of Equation (2.15) can be expressed as

$$\ln\left(\frac{I}{V^2}\right) = \ln\left[\frac{A\alpha\beta^2}{d^2\varphi t^2(\mathrm{y})}\right] - \frac{Bd\varphi^{3/2}v(\mathrm{y})}{\beta V} \tag{2.16}$$

using $v(\mathrm{y}) \sim 1$ in Equation (2.16), it gives

$$\ln\left(\frac{I}{V^2}\right) = \ln\left[\frac{A\alpha\beta^2}{d^2\varphi t^2(\mathrm{y})}\right] - \frac{Bd\varphi^{3/2}}{\beta V} \tag{2.17}$$

Here, a plot of $\ln(I/V^2)$ versus $1/V$ yields a linear curve. This plot is the well known F-N plot. From Equation (2.17), the slope (S) of the F-N plot is given by

$$S = -\frac{Bd\varphi^{3/2}}{\beta} \tag{2.18}$$

The value of field enhancement factor can be estimated from the slope of the measured F-N curve if the work function of the emitter is known. All the field emission sources rely on the field enhancement due to sharp tips or high aspect ratio of protrusions, so emitters tend to have smaller virtual source sizes because of the primary role of the β factor. The larger the β, the higher the field concentration, and therefore the lower the effective threshold voltage for emission. For a thin cylinder, the β value is roughly proportional to the height-diameter ratio (aspect ratio)[Utsumi 1991-2276]. For electron emission to occur, the electric field at the emitting surface must be on the order of 1-3 V/nm, which at reasonable applied voltages requires larger values of β.

It should be noted that the F-N model is valid only for flat metal surfaces at 0 K, and is in many case not satisfactory. However, the model is simple and widely used. One can only hope that efforts that aim to extend the model or to develop numerical approaches will soon supplement it.

2.7 Overview of cold cathode materials

The desirable field emitter materials should have a low work function, long life time, and compatibility with the processing technology. The first cone-shaped field emission cathode was developed by Spindt[Spindt 1968-3504] in 1968. This field emission cathode is formed by molybdenum evaporation deposition through an aperture on a substrate as depicted in Figure 2-18[Itoh 2002-514]. Molybdenum and tungsten emitters are commonly used for high current densities because of their high melting point property. Both molybdenum and tungsten emitters can be fabricated by thermal evaporation through a small aperture over the substrate.

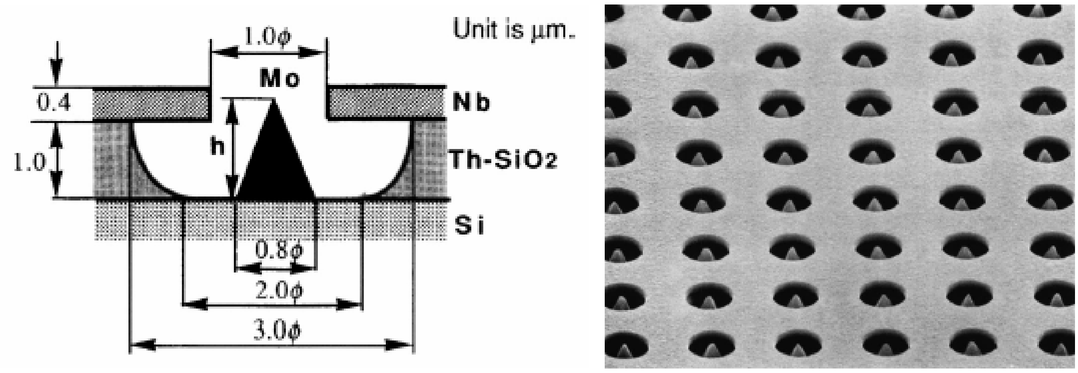

Figure 2-18 Structure and SEM image of a Spindt-type field emitter[Itoh 2002-514].

Hafnium and lanthanum hexaboride (LaB_6) are used as emitter materials because of their low work functions (about 3.6 eV)[Golubev 2000-67]. Since transconductance increases with decreasing work function, emitters with low work function are top contenders for vacuum microelectronics. However, in the case of LaB_6 the emission current is unstable due to atom migrations.

Silicon tips[Cade 1989-2704] formed by plasma or chemical etching techniques can also be used as emitters. By selecting a certain doping concentration, a degree of control (not available for metals) over electronic properties of the emitter can be achieved. Silicon tips coated with various refractory metals, such as molybdenum and tungsten, are also used as emitters[Busta 1989-2679].

Gallium arsenide (GaAs) multipoint cathodes have also been demonstrated by Bakhtizin *et al.*[Bakhtizin 1991-2398] GaAs is believed to be a promising material for the development of large area efficient sources of electrons and a new type of photo-cathodes due to its high mobility of electrons, short lifetime of the minority carriers, and greater energy band gap as compared to silicon and germanium.

Diamond is one of the main crystalline allotropes of carbon. Among the many unusual and fascinating characteristics that diamond possesses is the presence of a very small barrier to its electrons leaving the surface and emitting into vacuum. When diamond surfaces are terminated with hydrogen atoms, the electron affinity can become negative[Himpsel 1979-624]. This allows diamond to emit electrons efficiently at low applied fields, even from planar surfaces, without the need of fabricating sharp microtips to provide geometric field enhancement. Moreover, the excellent mechanical and chemical stability of diamond leads to highly durable and reliable emitters.

In recent years, one-dimensional (1-D) nanostructures, such as nanotubes, nanowires, nanotips, nanocones, nanorods, etc., of metals or semiconductors with high aspect ratios are considered to be ideal field emission electron sources that allow electrons to overcome the surface barrier and emit into vacuum at low electric fields. Among these materials, CNTs have received much attention, since field emission from a single MWCNT tip was first demonstrated in 1995[Rinzler 1995-1550]. Both theoretical predictions and experimental observations have shown that CNTs possess unique physical properties. In particular, the extraordinary geometric property of high aspect ratio leads to CNTs naturally attractive as field emitters. Although the large work function of about 5 eV associated with CNTs[Suzuki 2000-4007], high field concentrations can be effectively induced at the ultra-sharp CNT tips. Considerably large values of the field enhancement factors in the order of 10^2-10^3 can generally be achieved, which exceed those associated with metal microtips. As a result they are capable of emitting high currents (up to 1 A/cm^2) at low electric fields (about 5 V/μm). In addition, graphitic CNTs provide stable chemical and structural properties, whereas the conventional metal or silicon field emitters have shown increased tendency for oxidation during the operation, leading to the emission current fluctuation[Itoh 1993-647].

2.8 Potential applications of field emission

Field emission devices possesses with some unmatched advantages: electrons in vacuum can travel far faster with less energy loss than in any semiconductor. This enables faster modulation and higher electron energies than are possible with semiconductor structures, so field emission devices can operate at higher frequencies and higher power in a wider temperature range, as well as in high radiation environments. Due to the superior properties, possible applications of the field emission vacuum microelectronic devices include[Shieh 1992-22]

(a) Ultra-fast computers, memory, and signal processing with switching speed less than 1 pico-second.

(b) Small, lightweight, and inexpensive low-power to medium-power microwave and millimeter-wave amplifiers with very high efficiency and ultra-wide bandwidth.

(c) rf generators above 100 GHz, perhaps to 1 THz.

(d) Ultra-radiation hard analog and digital electronics.

(e) Temperature insensitive electronics for jet engine control without fuel cooling.

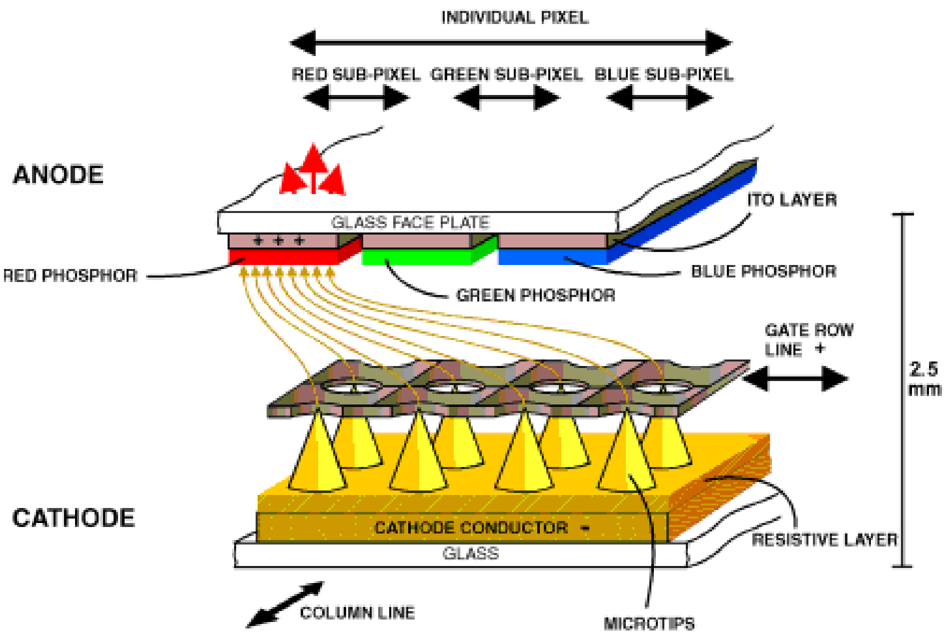

Figure 2-19 Perspective view of a Spindt-type FED in which one emitter array is shared by a red, green, and blue sub-pixel[Zhu 2001-293]. The drawing is not to scale.

(f) Electron sources for microscopes, ion guns, mass spectrometers, cathode-ray rubes, X-ray generators, high energy accelerators, and electron-beam lithography.

(g) Novel full-color, high brightness, flat panel field emission displays (FEDs).

Among these different applications, FEDs have moved beyond the research laboratory to actual prototypes and commercial products. FEDs are drawing attention as one of the most promising flat panel displays, which have following excellent features[Itoh 2004-1362; 2002-514]: thin panel thickness (about 2 mm), self-emissive distortion-free images, wide viewing angle (about 170°), quick response in the order of microseconds by controlling with analog or digital signals without active elements, tolerance to environment as high as that of receiving tubes, free from the terrestrial magnetic effect, free from the changes in the ambient magnetism, quick start of operation, and less dead space of image. FEDs are also expected as the low power consumption displays, because the emission current, in principle, flows only to the light emitting part in proportion to the light intensity.

The operation principle of a typical FED is depicted in Figure 2-19[Zhu 2001-293]. The field emission device sketched in Figure 2-19 is a Spindt-type triode device using metal microtips as emitters. The gate electrode serves as the electron extracting electrode and the glass plate which is coated with phosphors is used as a display screen. The electrons are extracted from emitters by a gate voltage of less than 100 volts and accelerated by the electric field toward the collector, which is biased at a higher

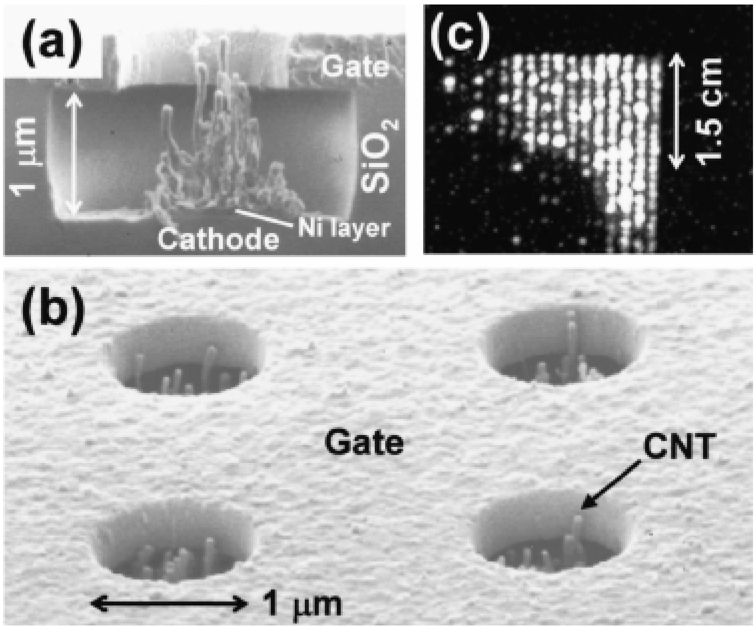

Figure 2-20 CNT emitters formed in gated holes. Cross-section view (a), top view (b), and photograph of the anode emitting at the gate voltage of 100 V (c)[Shiratori 2003-2485].

potential than the gate. The electrons excite the phosphor, which emits photons by cathodeluminescence (CL). The light is observed from the opposite side through the collector (anode) plate. The operation principle is similar to a conventional cathode-ray tube (CRT).

In the past three decades, Spindt-type FEDs consisted of microfabricated metal field emitter microtips were widely researched and developed[Spindt 1968-3504]. Recently, the construction of FEDs using various forms of nanostructures, especially CNTs, as emitters has become active. The most attraction of using CNT emitters is the hope of improving the field emission characteristics by utilizing their extremely sharp tube shape, and at the same time to reduce the manufacturing cost of the cold cathode. Two methods are generally used to fabricate gated CNT-FEDs. One is to squeeze the CNT mixtures (in most cases with organic vehicles) into the pre-patterned gated holes[Wang 2001-1294; Chung 2002-4045], and the other is to directly growth the CNTs inside the holes by CVD[Shiratori 2003-2485; Cheng 2001-H15]. The former approach is easy in processing but it is limited in making high-resolution displays. Therefore, in order to achieve high-resolution displays, it is necessary to direct growth of CNTs inside the gated holes and precise control of gate-to-CNT distances. Figure 2-20(a) and (b) are SEM images showing MWCNTs selectively grown inside the gated holes[Shiratori 2003-2485]. However, this approach is also limited by choosing substrate due to the high growth temperature (higher than 500 ^{o}C) of CNTs and is difficult to keep CNT uniformity over a large area.

Figure 2-21 Video running image of the CNT-based FED[Chung 2002-4045]. The size of active area is 5 inch diagonal and operated in the gate electrode of 100 V and anode biases of 1.5 kV, respectively, which was observed with a duty ratio of 1/120 and a frequency of 100 Hz.

Despite such problems, CNT-FEDs are still expected to be realized. Samsung Advanced Institute of Technology in 2002 had published a 5 inch CNT-FED prototype as shown in Figure 2-21[Chung 2002-4045]. The luminance as high as 510 cd/m^2 was achieved under an application of 100 V and 1.5 kV to gate electrode and anode, respectively.