R

Radiofrequency

▶ Thermal Cancer Ablation Therapies Using Nanoparticles

Random Matrix Theory

▶ Harnessing Disorder at the Nanoscale

Rapid Atomic Processing (RAP)

▶ Ultraprecision Surfaces and Structures with Nanometer Accuracy by Ion Beam and Plasma Jet Technologies

Rapid Electrokinetic Patterning and Its Applications

Katherine Clayton[1], Jian-Wei Khor[1] and Steven T. Wereley[2]
[1]Birck Nanotechnology Center, School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA
[2]School of Mechanical Engineering, Purdue University, Birck Nanotechnology Center, West Lafayette, IN, USA

Synonyms

Electrokinetics; Hybrid opto-electric technique; Optoelectrokinetic Technique; Optofluidics

Definition

Rapid electrokinetic patterning (REP) is a hybrid optoelectric technique that concentrates and manipulates particles via the combination of a uniform alternating current (AC) electric field and laser light. Alone, the AC electric field induces electric double layer (EDL) polarization of the dielectric particles. Coupling the AC electric field with the laser light produces a toroidal electrothermal flow. This flow sweeps colloids suspended in the solution and brings them close to the electrode surface where particles are trapped by particle-electrode interactions.

Overview

The ability to concentrate and manipulate colloids on chip provides potential solutions in fields of micromanufacturing, disease diagnostics, biology, and plasmonics [1–3]. The ability to dynamically configure clusters of particles decreases time-to-assemble and manufacture structures, which is a continuing issue in many lab-on-a-chip (LOC)-based systems [4]. Many optical and

electrokinetic techniques have attempted to alleviate these issues however certain limitations still exist. Purely optical techniques, such as optical tweezers, allow for only small numbers of particles to be captured within the beam waist of a laser beam [5]. On the other hand, purely electric techniques, such as dielectrophoresis, require complex fabrication of microelectrode that cannot be changed dynamically [6].

Rapid electrokinetic patterning is a dynamic, high-throughput technique that allows for micro- and nano-based patterning of colloids. REP requires the application of a uniform AC electric field and optical illumination in order to concentrate particles within solution. Optical illumination induces an axisymmetric temperature gradient within the chip, with the highest temperature at the beam focus. This temperature variation causes physical property variation, most importantly the electrical conductivity and permittivity. The applied electric field interacts with these material property variations to create a body force within the fluid and consequently an electrothermal vortex. The vortex transports colloids to the center of the torus where particles are trapped by particle-electrode interactions. This flow is produced whether the laser is focused at the top electrode or bottom electrode. Figure 1a provides an example of a laser focused on the bottom electrode. When focusing the laser at the top electrode, it would look like Fig. 1a but vertically reflected in the horizontal midplane.

REP enables particle manipulation, translation, sorting, and patterning. This wide variety of processes that the technique enables make it useful in fields like micromanufacturing, plasmonics, biological detection and cell sorting, as well as chemical analysis. Using light patterning, particles can be concentrated into different patterns and structures on the electrode surface [8]. For example, detection of pathogenic bacteria requires cell sorting and concentration. The ability to sort such entities with ease could be potentially resolved with REP [9, 10]. These tasks are highly desirable in the biomedical community, potentially allowing full-scale laboratory analysis with small sample volumes and on platforms the size of a credit card [4].

REP can simultaneously collect particles of a specific size and manipulate the colloidal cluster. This technique requires a heat source, such as a laser or microheaters, to provide a temperature gradient along with an AC electric field. REP provides several benefits. First, it is a noncontact high-speed manipulation method, accumulating a cluster of particles in a matter of seconds. Second, REP is highly repeatable, as under the same settings, the same particle clusters will form each time. Third, REP chips require simple fabrication and manufacturing.

REP is a promising technique in its potential to contribute to a wide variety of fields in nano- and micro-technologies.

Physics

REP combines optics (using laser light) with electrokinetics (using an AC electric field). When coupled with one another, an electrothermal flow is produced in a toroidal fashion. This flow pattern "sweeps" particles toward the center of the torus' minor radius, where the temperature gradient generated by the nonuniform light source is maximum. The cluster of particles can be trapped onto the top or bottom electrode surface (depending on where the laser light is focused). An example of this is provided in Fig. 2b.

When the light source is removed from the apparatus, particles are no longer collected into the trap. Smaller-sized particles (i.e., <3 μm), post-light deactivation, are repelled from one another due to dipole-dipole repulsion (Fig. 2c). However, this repulsive effect is largely due to the AC electric field being maintained in chip. If the alternating current is removed, but the laser light is maintained, natural convection is seen within the apparatus [11] but usually at a speed an order of magnitude slower than that created by the electrical field. The coupling of the light source and AC electric field is vital for maintaining the concentration of particles.

In more detail, dielectric particles are polarized due to the introduction of the electric field. There is a dipole force that acts between two neighboring particles. Due to their repulsion and self-organization on the electrode surface when
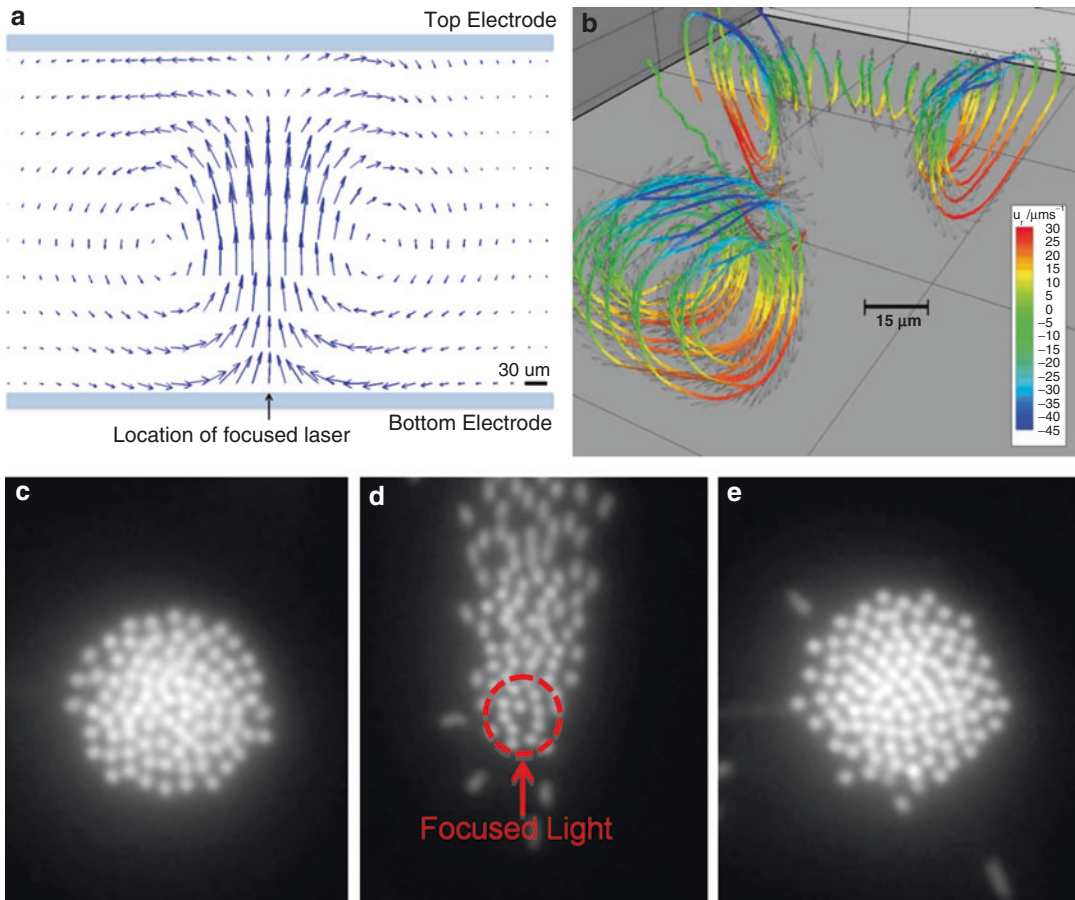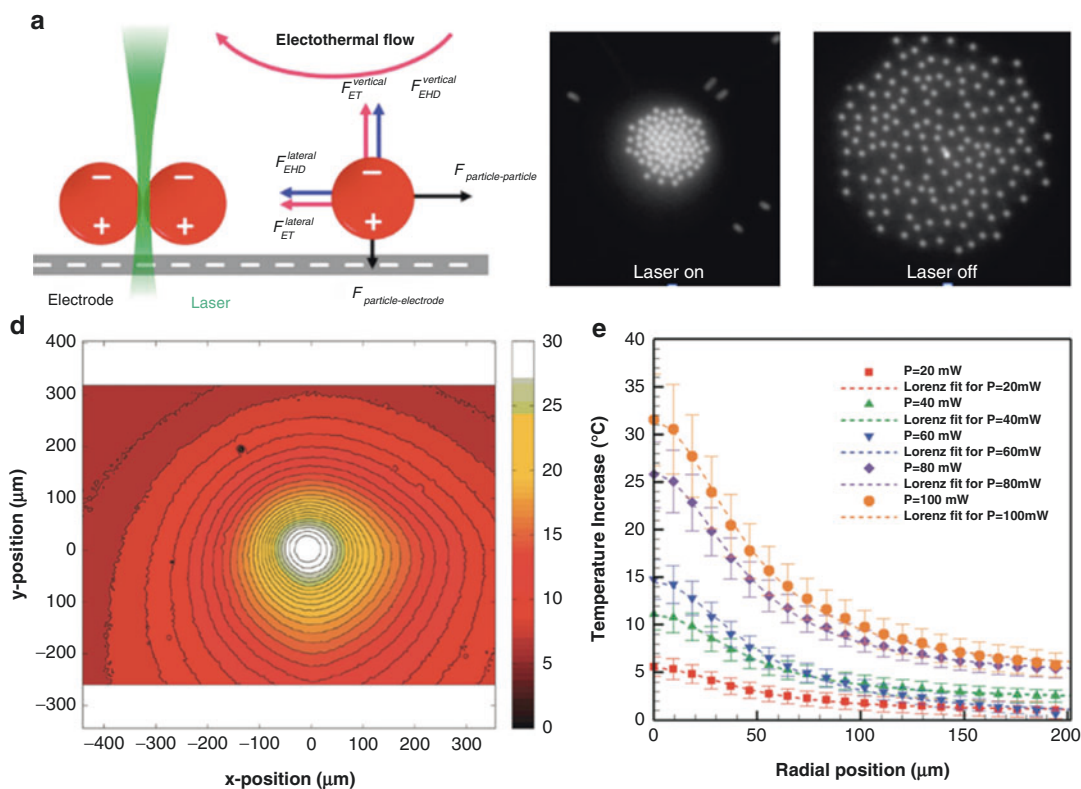
Rapid Electrokinetic Patterning and Its Applications, Fig. 1 **a** A 2D projection of electrothermal flow in REP. In this circumstance, a laser is focused on the bottom electrode of an REP chip, producing an upward flow on the inner radius of the toroidal fluid motion. **b** A 3D trajectory of the electrothermal flow at a frequency of 200 kHz. In this case the laser is focused at the top electrode surface (**b** Printed with permission from Kumar et al. Copyright 2010 Springer) [7]. **c** 1 μm particles are trapped in a monolayer using REP. These particles can be manipulated with ease, such as with **d** where the laser light is suddenly moved to an area below its original spot, and the particles are swept to the new location, until they are again in equilibrium in the trap, as seen in (**e**)

solely affected by the electric field, their lines of centers remain perpendicular to the electric field lines. As the laser light is combined with these electric field effects, the temperature gradient provided by the light induces a gradient for the conductivity and permittivity of the solution. This combination produces a toroidal electrothermal vortex, which is characterized by the time-averaged electrothermal body force equation:

$$\langle f_e \rangle = \frac{1}{2}\mathrm{Re}\left[\left(\frac{(\sigma\nabla\varepsilon - \varepsilon\nabla\sigma)\cdot E}{\sigma + i\omega\varepsilon} \right)E^* - \frac{1}{2}|E|^2\nabla\varepsilon \right] \quad (1)$$

In the time-averaged fluid body force equation, Re represents taking the real valued portion of the expression. The electric field is represented by E and its complex conjugate is E^*. The equation encompasses the gradient of conductivity, σ, and permittivity, ε, of the solution, which is created due to the temperature gradient produced by the optical source, and then finally the ω term represents the angular frequency. On the right-hand side of the equation, the left-hand term represents the Coulomb force, while the right-hand term is the dielectric force component [12].
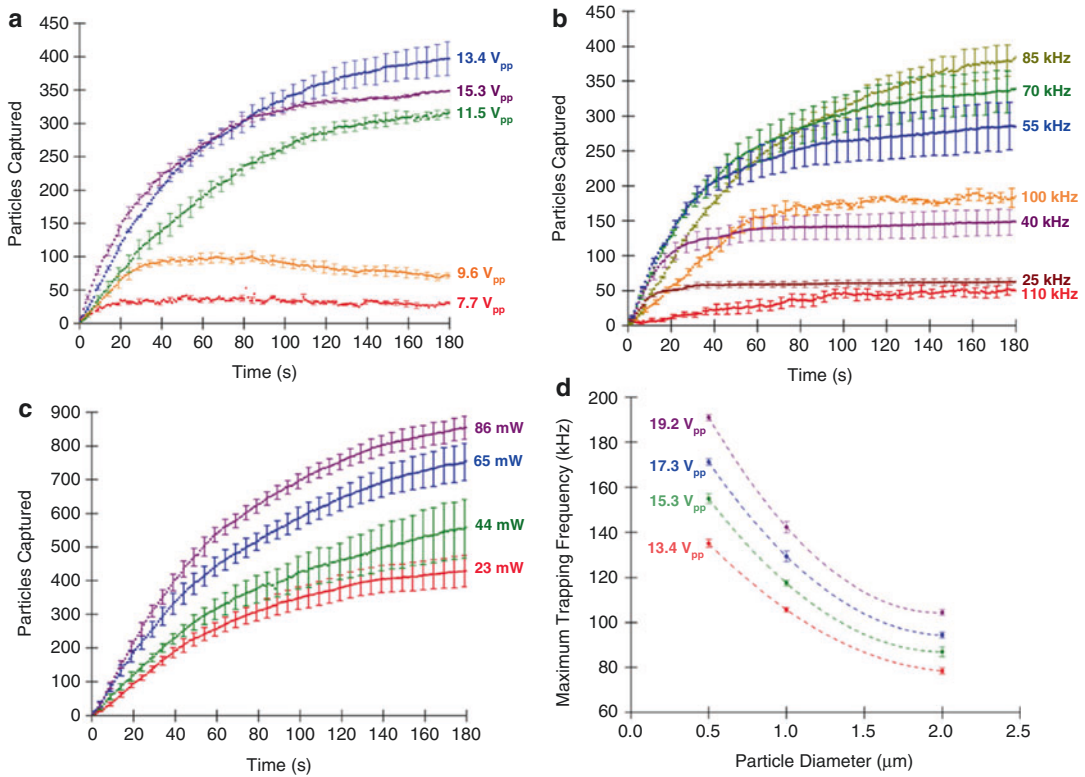
Rapid Electrokinetic Patterning and Its Applications, Fig. 2 **a** The force relationships on a particle undergoing REP – the bottom electrode is negatively charged, where the particle with an induced dipole interacts with its surface. Both lateral and vertical induced drag force, and electrohydrodynamic forces act on the particle as well. Additionally, the particle-particle interaction force is shown (Reprinted with permission from Kwon et al. Copyright 2012 Springer) [13]. **b** shows the congregation of particles during REP trap activation and **c** dipole-dipole repulsion of the particles after REP is "turned off." **d** The temperature increase on an ITO cover-slip introduced to a 150 mW powered laser. The laser is focused at the origin. **e** The effect of different laser powers (in mW) and its corresponding temperature increase, starting at the origin of the laser spot and propagating out radially (**d** and **e** Reprinted with permission from Kumar et al. Copyright 2010 American Chemical Society) [15]

From current understanding, four main electro-kinetic relationships can be analyzed. The first relationship is the force between particles. Denoted $F_{particle}$, this force is what particles experience between one another (i.e., $F_{particle}$- $F_{particle}$). Under the AC electric field, the dipole-induced particles experience repulsion from one another. The second is the force between the electrode and particle ($F_{electrode}$- $F_{particle}$). As the electrothermal flow sweeps the particle into a monolayer formation, the particles undergo a physical interaction with the electrode. This electrothermal flow aids in a drag force that can be broken into vertical and lateral component ($F^{lateral}_{ET}$, $F^{vertical}_{ET}$).

The fourth is an attractive particle-particle force ($F^{lateral}_{EHD}$, $F^{vertical}_{EHD}$). Polarized particles produce a local electrohydrodynamic flow due to the interaction between the electric field and an ionic double layer on the particle and electrode [13]. This force balance is demonstrated in Fig. 2a.

At a specific frequency, (1) the drag force due to electrohydrodynamic flow, (2) the force between the electrode and particle, (3) the force between particles, and (4) the drag force due to the electrothermal vortex balance one another at the trapping point. If the frequency is increased, the flow becomes so strong that particles are

Rapid Electrokinetic Patterning and Its Applications, Fig. 3 **a** Accumulation of 1.0 um particles at the ITO electrode over a period of 3 min for various V_{pp}. Laser power is 23 mW and AC frequency is 70 kHz. **b** Accumulation of 1.0 um particles at the ITO electrode over a period of 3 min at various frequencies. Laser power is 23 mW and AC voltage is 15.3 V_{pp}. **c** Accumulation of 1.0 um particles at the ITO electrode over a period of 3 min for various frequencies. AC peak-to-peak voltage is 15.3 V_{pp}. and AC frequency is 70 kHz. **d** Maximum trapping frequency for particles of various diameters to cluster at different applied AC peak-to-peak voltage. Laser power is 23 mW (Reprinted with permission from Williams et al. Copyright 2009 IOP) [12]

rapidly swept and unable to be maintained within their respective cluster. However, this relationship is still under investigation.

Although these relationships do remain under investigation, Fig. 3 represents trends seen experimentally in REP. In Fig. 3a the number of particles captured increases linearly and then plateaus as the particle cluster reaches steady state. Furthermore, it is observed that higher applied voltages cause larger numbers of particles trapped (except for the 15.3 V_{pp} case). This 15.3 V_{pp} outlier case may be due to an increased electrothermal body force. This increase may overcome particle-particle and particle-electrode forces, ripping particles out of their original trap space. In Fig. 3b it is observed that the particles captured for various frequencies increase linearly and then

plateau as the cluster reaches steady state. Figure 3c shows that the number of particles within the cluster increases linearly and then begins to reach steady state. The increase in laser power increases the number of particles captured. This may be due to the fact that as the laser power increases, the electrothermal flow velocity increases. This is due to a steeper temperature gradient, leading to faster particle circulation. This circulation brings particles closer to the colloid cluster, increasing the probability of particles to be pulled into the REP trap. Figure 3d shows the maximum trapping frequency for particles of different diameters before it fails to form a cluster. A particle of smaller diameter will remain trapped at a higher frequency when compared to a larger particle. Furthermore, the maximum trapping

frequency of the same particle increases as the applied AC peak-to-peak voltage increases [14].

Methodology

Fabrication of REP chips is relatively simple. The chip is made of a thin-film electrode coated onto a glass substrate. Indium tin oxide (ITO)-coated glass substrates are typically used to conduct REP. This coating is highly transparent allowing for observation within the REP chip under a microscope.

Two of these substrates are placed with the electrodes facing one another with a spacer in between. This spacer is designed to define microchannels and wells. The spacer material is adhered to both electrodes and is nonreactive with the sample. In some cases, a commercial double-sided tape can be used to conduct short-term REP studies. The typical setup for conducting REP requires a laser source and an AC function generator. The laser is applied to a desired spatial location on chip. A representation of this setup is presented in Fig. 4.

The channel is often illuminated using a near-infrared (NIR) laser of 1064 nm wavelength. However, alternative wavelengths may be used. Traditionally, the laser power used in REP is less than 200 mW. Furthermore, the laser frequency has to be considered so that the heat absorption of the laser onto the sample is minimal. Objective

lenses are typically chosen to be at a magnification of 20× or larger. This ensures that the laser spot size is focused enough to create a steep temperature gradient. A schematic of the REP setup can be found in Fig. 5.

For electrical effects, an AC electric field with a frequency less than 200 kHz must be used. This induces polarization of the dielectric material. In higher frequency regimes, the particles relax, hindering the collection of the particles under REP. Very low-frequency regimes often result in self-organization of particles on the electrode surface, but in turn induce strong particle-electrode interactions.

Particles must be suspended in low conductivity solutions (i.e., $\sigma \leq 10$ mS/m) [14, 15]. When particles are submerged in higher conductivity solutions, the electric double layer becomes thinner, increasing the particle-electrode adhesion.

Applications

In recent years, researchers have investigated potential applications of rapid electrokinetic patterning. This spawned a wide spectrum of uses for this opto-electrokinetic technique, from biology to plasmonics.

In Biology

Biology wet labs are looking for alternative systems that allow for appropriate sampling of
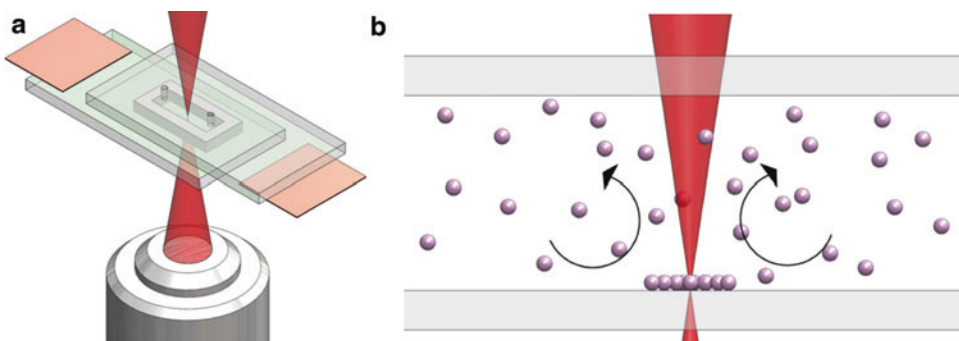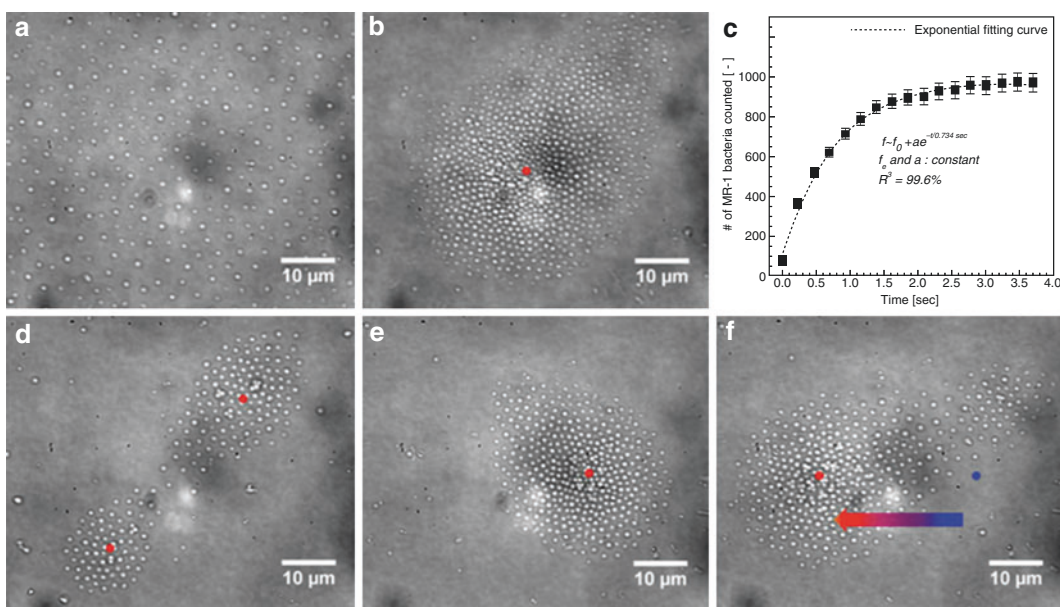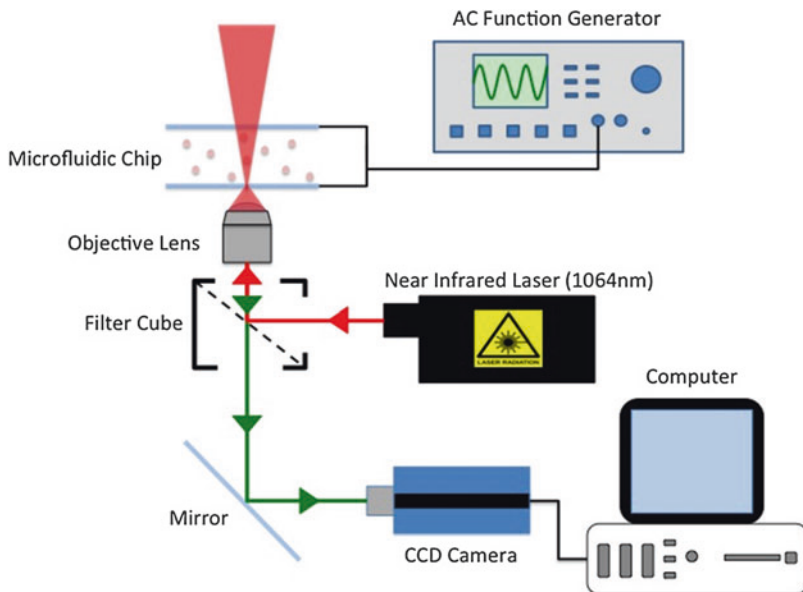

Rapid Electrokinetic Patterning and Its Applications, Fig. 4 In image **a**, a demonstration of an REP chip with laser light focused on a particular plane, where **b** shows how at the beam waist of the laser particles are trapped on the electrode surface; note the *black arrows* correspond with the flow direction of the electrothermal vortex

Rapid Electrokinetic Patterning and Its Applications, Fig. 5 A schematic of a standard REP setup. A fabricated REP chip undergoes an AC electric field from a function generator as well as heating from laser light. The 1064 nm laser is often directed to the substrate via an objective lens from the microscope. The CCD camera records images from the objective
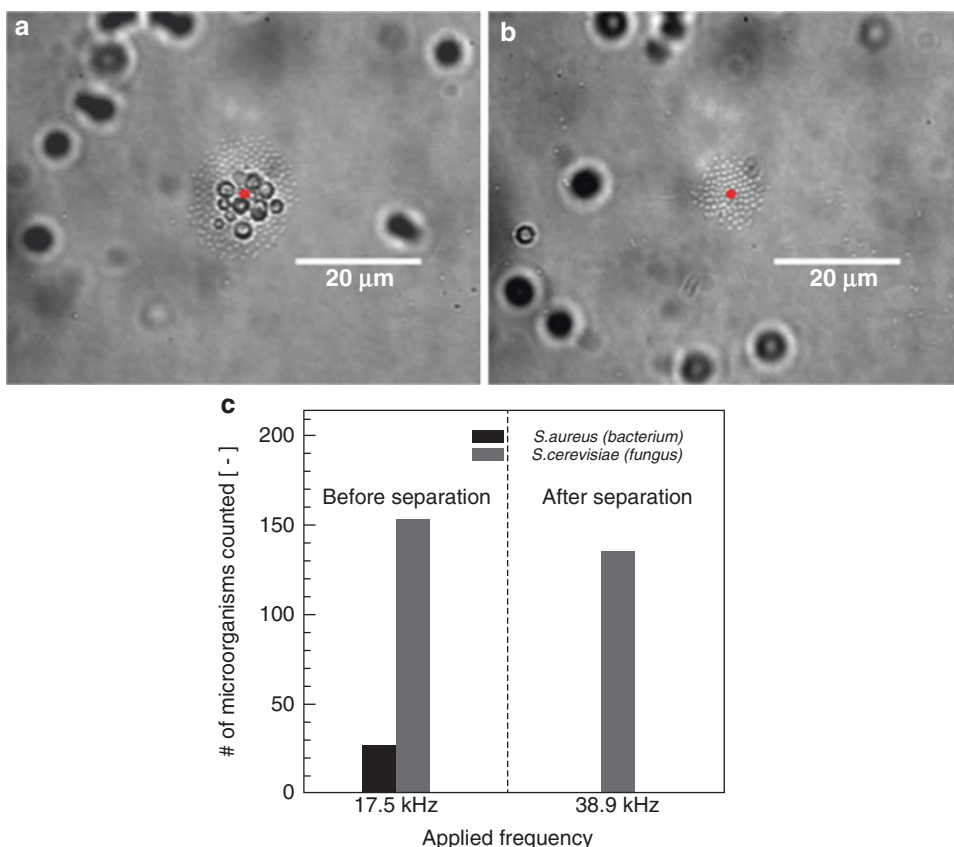

Rapid Electrokinetic Patterning and Its Applications, Fig. 6 From Kwon et al. image **a** shows *S. oneidensis* prior to REP trapping and **b** during trapping. In **c** it is shown how over time more bacteria are collected in the REP trap, eventually reaching a plateau, image **d** shows two laser spots and the particles trapped about the focused light, and **e** and **f** demonstrate how moving the light about the chip allows for dynamic manipulation of the particles on chip (Reprinted with permission from Kwon et al. Copyright 2012 Royal Society of Chemistry) [3]

solutions. Current techniques often require large amounts of consumables (both from reagents and materials), which are both costly and environmentally unfriendly. Many optical and electrokinetic-based methods wish to alleviate these issues. However, biological applications are difficult to implement in optoelectric systems due to issues with their charge properties, uniformity in sample,
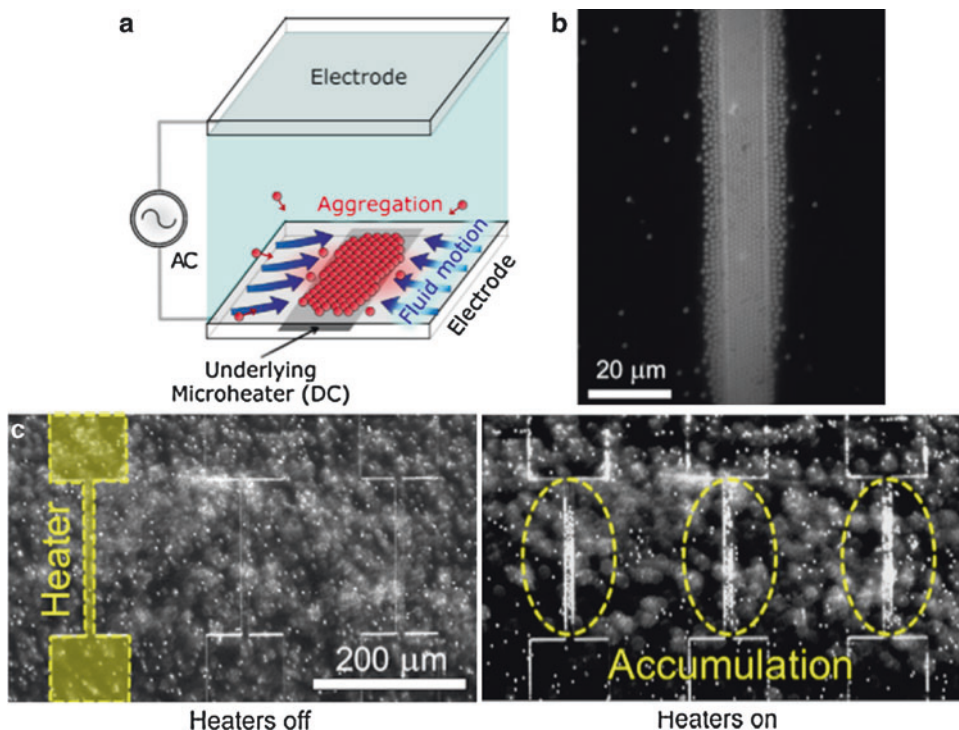
Rapid Electrokinetic Patterning and Its Applications, Fig. 7 **a** Showing how multiple species – *S. aureus* and *S. cerevisiae* – are collected with REP. **b** After changing frequency, the larger *S. cerevisiae* are swept out of the trap, and only the *S. aureus* remains. **c** Providing how in image (**a**) there are 150 bacteria and 8 fungi trapped and how after in (**b**) there are 125 bacteria and no fungi within the trap (Reprinted with permission from Kwon et al. Copyright 2012 Royal Society of Chemistry) [3]

and viability/compatibility with the system. Therefore, like with many optical and electrokinetic techniques, biological use with REP is currently being investigated. Kwon et al. published the first work, using REP as a bioassay platform. *Shewanella oneidensis*, a 1 μm rod-shaped bacteria, was examined with rapid electrokinetic patterning. This gram-positive MR-1 mutant was placed under an AC electric field, where the bacterium's major axis aligned with the electric field. After introducing the laser, the bacteria collected onto the electrode surface in their vertical orientation. Within 0.73 s, 700 organisms were maintained in the trap, saturating out at ~1,000. Figure 6a, b shows the *S. oneidensis* before and during trapping. This demonstrates the ability to

quickly and dynamically collect the species under REP. Likewise, multiple laser spots allowed for the collection of multiple *S. oneidensis* clusters, and through movement of the laser, these traps could translate about the chip, as seen in Fig. 6d [3].

Kwon et al. also demonstrated collecting multiple organisms within an REP trap, depending on its critical frequency effects (Fig. 7). *Saccharomyces cerevisiae* (a 5 μm spherical-shaped fungus) and *Staphylococcus aureus* (a ~1 μm spherical-shaped bacteria) were investigated. These two geometrically similar species were chosen due to their shape and similar zeta potentials. Under a 17.5 kHz frequency, 10.07 V_{pp}, and 20 mW laser power, both species clustered together within the

Rapid Electrokinetic Patterning and Its Applications, Fig. 8 **a** A schematic of a fabricated microheater to induce a temperature gradient on the bottom electrode surface, allowing for REP to take place without laser heating. **b** Accumulation of polystyrene particles along the microheater. **c** A display of an array of three microheaters.

Initially when the heaters are off, particles are freely dispersed in solution. After the heaters are turned on with the electric field, particles begin to accumulate on the surface of the electrode, in the location of the thin-film heaters (Reprinted with permission from Velasco et al. Copyright 2012 Elsevier) [17]

REP trap. However, by increasing the frequency to almost twice as much, the *S. cerevisiae* were swept away from the trap, whereas the *S. aureus* sample was maintained. This proved the capability of using REP as a method for separating small microbial samples on the basis of size [3].

Alternative Platforms

Producing a cost-effective, easily accessible, and minimally power-consuming REP system is a key requirement for developing REP-based portable devices. Currently, a 1,064 nm laser is used to illuminate the REP chip. This is easily controllable within the laboratory space but provides little-to-no mobility of the system. Williams et al. investigated thin-film heaters as an alternative for laser-induced electrothermal flow [16]. Velasco et al. took this theory further by experimentally

fabricated resistors onto a glass surface and tested 1–2 μm polystyrene spheres in the system [17]. A 10 μm wide wire trace was used which produced a 0.5 K temperature gradient. The system allowed for trapping, patterning, and sorting of colloidal particles. The electrothermal vortex was slower as compared to traditional REP [17]. Examples of this work are provided in Fig. 8a–c. This successful work showed the potential capabilities of microfabricating REP systems that can be used in the field. This opens up a potential opportunity for future use as a portable electrokinetic system.

Park et al. investigated twin opposing microvortices (TOMV). Fabricating two coplanar ITO electrodes onto glass surface (partly covered with SU-8), an AC electric field was applied between them while introducing laser light through the chip. Depending on where the
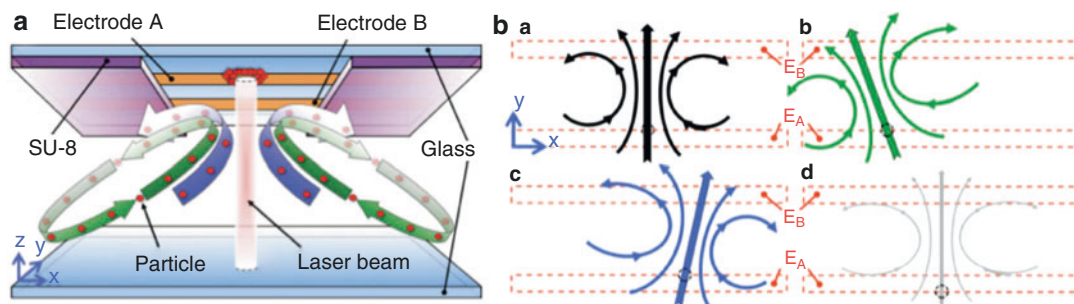
Rapid Electrokinetic Patterning and Its Applications, Fig. 9 In **a** is a schematic of microfabricated ITO electrodes, partially blocked by a layer of SU-8 on a glass substrate. These coplanar electrodes are alternative to the traditional parallel ITO coverslips used. This design creates twin opposing microvortices, collecting particles at the electrode surface. In **b**, where the laser light is focused (demonstrated by the point drawn on the *bold colored line*), changes the orientation of the opposing microvortices (Reprinted with permission from Park et al. Copyright 2012 Royal Society of Chemistry) [1]

light was focused, the TOMV changed their orientation within the chip (Fig. 9). These tunable microvortices are of interest due to the specific control of location and mixing of particles. Asymmetric flows and interesting mixing effects are possible with this technique [1].

Ndukaife et al. fabricated plasmonic nanostructures to increase heating effects in a traditional REP chip [2]. This nanostructure array absorbed light optimally at 1,064 nm. By exciting the structure at resonance, a rapid heat gradient was produced. Due to the increased heating effect, the velocity of the electrothermal vortex increased. This showed the potential of plasmonic structures as an REP platform. This could be used further in biosensors as well for surface-enhanced Raman spectroscopy (SERS) [2]. However, it is important to note that the fabrication of plasmonic structures is cost and labor intensive due to the use of e-beam lithography. A carefully selected ITO electrode can produce a sufficiently strong flow without the need of nanostructures.

Conclusions

Rapid electrokinetic patterning poses an array of current and future applications in a variety of fields. Now that the technique has been developed, work is being done to apply it. As this space is being pursued, REP will be used for novel applications within a large spectrum of nano- and microscale phenomena.

Cross-References

▶ AC Electrokinetics
▶ Dielectrophoresis
▶ Lab-on-a-Chip

References

1. Park, C., Wereley, S.T.: Rapid generation and manipulation of microfluidic vortex flows induced by AC electrokinetics with optical illumination. Lab Chip **13**, 1289–1294 (2013)
2. Ndukaife, J.C., Mishra, A., Guler, U., Nnanna, A.G. A., Wereley, S.T., Boltasseva, A.: Photothermal heating enabled by plasmonic nanostructures for electrokinetic manipulation and sorting of particles. ACS Nano **8**, 9035–9043 (2014)
3. Kwon, J.-S., Ravindranath, S.P., Kumar, A., Irudayaraj, J., Wereley, S.T.: Opto-electrokinetic manipulation for high-performance on-chip bioassays. Lab Chip **12**, 4955–4959 (2012)
4. Franke, T.A., Wixforth, A.: Microfluidics for miniaturized laboratories on a chip. Chem. Phys. Chem. **9**, 2140–2156 (2008)
5. Arai, F., Ng, C., Maruyama, H., Ichikawa, A., El-Shimy, H., Fukuda, T.: On chip single-cell separation and immobilization using optical tweezers and thermosensitive hydrogel. Lab Chip **5**, 1399–1403 (2005)

6. Pethig, R.: Review article-dielectrophoresis: Status of the theory, technology, and applications. Biomicrofluidics **4**, 1–35 (2010)
7. Kumar, A., Cierpka, C., Williams, S.J., Kähler, C.J., Wereley, S.T.: 3D3C velocimetry measurements of an electrothermal microvortex using wavefront deformation PTV and a single camera. Microfluid. Nanofluid. **10**, 355–365 (2010)
8. Williams, S.J., Kumar, A., Wereley, S.T.: Electrokinetic patterning of colloidal particles with optical landscapes. Lab Chip **8**, 1879–1882 (2008)
9. Bridle, H., Miller, B., Desmulliez, M.P.Y.: Application of microfluidics in waterborne pathogen monitoring: A review. Water Res. **55**, 256–271 (2014)
10. Yoon, J.-Y., Kim, B.: Lab-on-a-chip pathogen sensors for food safety. Sensors (Basel) **12**, 10713–10741 (2012)
11. Ramos, A., Morgan, H., Green, N.G., Castellanos, A.: Ac electrokinetics: A review of forces in microelectrode structures. J. Phys. D. Appl. Phys. **31**, 2338–2353 (1998)
12. Mishra, A., Kwon, J.-S., Thakur, R., Wereley, S.: Optoelectrical microfluidics as a promising tool in biology. Trends Biotechnol. **32**, 414–421 (2014)
13. Kwon, J., Thakur, R., Wereley, S.T.: Encyclopedia of Nanotechnology. Springer Netherlands, Dordrecht (2012)
14. Williams, S.J., Kumar, A., Green, N.G., Wereley, S.T.: Optically induced electrokinetic concentration and sorting of colloids. J. Micromech. Microeng. **20**, 015022 (2010)
15. Kumar, A., Kwon, J.-S., Williams, S.J., Green, N.G., Yip, N.K., Wereley, S.T.: Optically modulated electrokinetic manipulation and concentration of colloidal particles near an electrode surface. Langmuir **26**, 5262–5272 (2010)
16. Williams, S.J.: Enhanced electrothermal pumping with thin film resistive heaters. Electrophoresis **34**, 1400–1408 (2013)
17. Velasco, V., Williams, S.J.: Electrokinetic concentration, patterning, and sorting of colloids with thin film heaters. J. Colloid Interface Sci. **394**, 598–603 (2013)

Rate Sensors

▶ Gyroscopes

Reactive Current Clamp

▶ Dynamic Clamp

Reactive Empirical Bond-Order Potentials

J. David Schall[1], Paul T. Mikulski[2], Kathleen E. Ryan[3], Pamela L. Keating[3], M. Todd Knippenberg[4] and Judith A. Harrison[3]
[1]Department of Mechanical Engineering, Oakland University, Rochester, MI, USA
[2]Department of Physics, United States Naval Academy, Annapolis, MD, USA
[3]Department of Chemistry, United States Naval Academy, Annapolis, MD, USA
[4]Department of Chemistry, High Point University, High Point, NC, USA

Synonyms

Bond-order potential; REBO

Definition

Reactive empirical bond-order potentials are interatomic energy functions used in molecular dynamics simulation and modeling of nanosystems where an accurate but efficient description of chemical reactivity is required.

Genesis of Reactive Bond-Order Potentials

Atomistic simulation of a large number of atoms using molecular dynamics (MD) is a powerful tool for understanding the fundamental mechanisms present in nanomaterials systems. The ability to accurately model chemistry including reactivity, bonding, charge transfer, polarizability, mixing, etc., is central to this understanding. Underpinning these calculations is the atomic interaction potential. Ideally, the atomic interactions would be obtained directly from first-principles, i.e., through solution of Schrodinger's equation. However, such calculations are orders of magnitude too slow for the large number of energy evaluations required to study nanosystems

R

of reasonable size and practical interest. To obtain useful information in a reasonable amount of time, empirical, and semiempirical approximations to the atomic potentials have been developed. In this entry, a survey of reactive empirical bond-order and related interatomic potentials is presented.

An effective analytic potential should have a relatively simple functional form, which captures the essential essence of the underlying quantum mechanical bonding, and will be able to mimic experimental quantities by utilizing empirically derived functions and parameters. Empirical potentials are likely to include a wide range of fitting parameters, which are selected to reproduce cohesive energies, elastic constants, lattice constants, and surface energies. Ideally, the potential will also have some degree of transferability, having the ability to describe structures not included in the fitting database, at least in a qualitative sense. Finally, the resulting function should be relatively efficient computationally. The process of developing a potential, sometimes referred to as an art as well as a science, requires a combination of chemical insight, trial and error, and tenacity on the part of its developer.

Covalent materials, such as silicon and carbon, form strong directional bonds. This poses a challenge for potential development for this important class of materials. Many standard potential functions, such as the Lennard-Jones (LJ) potential or the embedded atom method (EAM), do not include any bond directionality. The Stillinger–Weber (SW) potential for solid and liquid phases of silicon was one of the first attempts to use a classic potential to overcome this challenge [1]. Stillinger and Weber based their potential model on a many-body approach. In this approach, the total energy is given as a linear superposition of terms representing different types of interatomic interactions, e.g., stretching, bending, rotation, and torsion. For the SW potential, the total energy includes only two of these terms, a pair- and triplet-term based on the geometric factors of bond length and bond angle, respectively. The total potential energy, E_{tot}, is given by

$$
\begin{aligned}
E_{tot} = {} & \frac{1}{2} \sum_{\substack{ij \\ (i \neq j)}} \phi_{ij}(R_{ij}) \\
& + \sum_{ijk} g(R_{ij}) g(R_{ik}) \left(\cos \theta_{ijk} + \tfrac{1}{3} \right)^2 .
\end{aligned}
\tag{1}
$$

Here ϕ_{ij} is the pair-term (bond stretching) representing electrostatic interactions between atoms i and j. The second summation is a three-body term, which represents bond bending between atoms i, j, and k. The term $g(r_{ij})$ is a decaying function with a cutoff between the first- and second-neighbor shell, and θ_{ijk} is the bond angle described by two neighbors j and k of atom i. The inclusion of the three-body term allows the potential to achieve an acceptable description of short-range order and of atom-exchanging diffusive motion in the liquid phase. While this potential is reasonably accurate when used for modeling solid silicon in the diamond cubic phase, it is biased toward the tetrahedral bond angle through the explicit inclusion of the factor $\left(\cos \theta_{ijk} + 1/3 \right)$ in the three-body term. The inclusion of this trigonometric factor discriminates in favor of pairs of bonds with the tetrahedral geometry, i.e., $\left(\cos \theta_{ijk} = -1/3 \right)$, and limits the transferability of the potential. For instance, it cannot accurately predict the energies for various non-tetrahedrally bonded, high-pressure phases of silicon, it does not correctly predict relative energies of surface structures, and the coordination of liquid-phase silicon is too low.

The family of potentials, CHARMM, AMBER, OPLS, etc., derived from the many-body approach described above for the Stillinger–Weber potential is often used in simulation of organic systems. These so-called force-field models are able to model structural and dynamical properties of very large molecules with a high degree of accuracy. However, it is important to note that the force-field methods generally do not allow for bond-forming or breaking to occur during simulation. Connectivity must be determined a priori. For a review of these potentials see for example Mackerell [2].

Abell presented a very general description of bonding based on the observation of a universal relationship between binding energy and bond

length in 1985 [3]. This relationship placed the bonding in crystalline solids and molecules on the same footing. The only criterion for bonding preference was the optimization of the binding energy with respect to the local coordination. The local coordination is, in turn, the dominant topological variable in the determination of binding energy. Soon after the publication of Abell, Tersoff developed a potential constructed to guarantee that this universal behavior was obtained [4]. The Tersoff potential was the first to attempt to incorporate the structural chemistry of covalently bonded systems into empirical potential energy function. The general form of the potential is

$$E = \sum_i E_i = \frac{1}{2} \sum_{i \neq j} V_{ij},$$
$$V_{ij} = f_c \left[V_R(R_{ij}) - B_{ij} V_A(R_{ij}) \right], \quad (2)$$

where E is the total energy of the system, E_i is the site energy for site i, V_{ij} is the interaction energy between atoms i and j, and R_{ij} is the distance between these atoms. The sum is over the j nearest neighbors of i; $V_R(R)$ and $V_A(R)$ are pair-additive repulsive and attractive interactions, respectively; f_c is a cutoff function to limit the range of the potential. For the Tersoff potential, the repulsive and attractive terms are represented by the Morse-type functions

$$\begin{aligned} V_R &= A \exp(-\lambda_1 R_{ij}) \\ V_A &= B \exp(-\lambda_2 R_{ij}), \end{aligned} \quad (3)$$

where A, B, λ_1, and λ_2 are all positive constants with $\lambda_1 > \lambda_2$. All deviations from the simple-pair potential are ascribed to the dependence of the bond-order B_{ij} on the local bonding environment or coordination. More specifically, the bonding strength for a bonded pair should be a monotonically decreasing function of the number of competing bonds, the strength of the competing bonds, and the cosines of the angles of the competing bonds. The innovative feature of the Tersoff potential is that it does not assume different forms for the angular functions for different hybridizations. In his work, B_{ij} has the form

$$\begin{aligned} B_{ij} &= \left(1 + \beta^n \xi_{ij}^n \right)^{-1/2n}, \\ \xi_{ij}^n &= \sum_{k \neq i, j} f_c(r_{ijk}) g(\theta_{jik}) \exp\left[\lambda_3^3 (r_{ij} - r_{ik})^3 \right], \\ g(\theta) &= 1 + c^2/d^2 - c^2/\left[d^2 + (h - \cos\theta)^2 \right], \end{aligned} \quad (4)$$

where θ_{jik} is the bond angle between bonds ij and ik. The angular function $g(\theta)$ is determined globally by a fit to solid structures with different coordination. This gives the function a high degree of transferability. Tersoff extended his original silicon potential to include germanium, carbon, and combinations thereof. This potential form has also been extended to a variety of III-V semiconductors.

An empirical bond-order expression that described both hydrocarbon molecules and solid-state carbon was developed by Brenner [5] to model chemical vapor deposition of diamond films. Taking inspiration from the Tersoff potential, the form of the reactive empirical bond-order potential (REBO) potential allows for bonds to form and break with changes in hybridization. This first-generation Brenner potential is very similar to the Tersoff potential. Pair repulsive and attractive terms are again represented by Morse-type potentials. The primary difference is the way in which the bond order is handled for hydrocarbon molecules. The expression for the total bond order is

$$\overline{B}_{ij} = \frac{B_{ij} + B_{ji}}{2} + F_{ij}\left(N_i^t, N_j^t, N_{ij}^{\mathrm{conj}} \right) \quad (5)$$

where

$$B_{ij} = \left[1 + \sum G_i(\theta_{jik}) f_c(r_{ik}) e^{\alpha\left[(r_{ij} - R_{ij}^E) - (r_{ik} - R_{ik}^E) \right]} + H_{ij}\left(N_i^H, N_i^C \right) \right]^{-\delta}, \quad (6)$$

The quantities Ni^C and Nj^H are the number of carbon and hydrogen atoms bonded to atom i. The total number of neighbors, Ni^t of atom i is $(Ni^C + Nj^H)$, Nij^{conj} depends on whether a bond between carbon atoms i and j is part of a conjugated system, $G(\theta)$ is a function of the angle between bonds i-j and i-k and has the same form

as the Tersoff potential. The two- and three-dimensional cubic splines, H_{ij} and F_{ij}, have discrete nodal values that are modified so that the potential reproduces the energies of various hydrocarbon molecules. The resulting potential successfully describes the different bonding characteristics of hydrogen and carbon radicals and nonconjugated double and triple carbon-carbon bonds.

Extensions of the First-Generation REBO Potential

Because the Tersoff-Brenner formalism has been used successfully to simulate a wide range of processes such as chemical vapor deposition, diamond growth, carbon nanotubes, graphene, diamond-like carbon, etc., it was extended to other systems by various groups. In what follows, some examples of ways in which the first-generation REBO potential has been extended are provided. This analysis is not meant to include all such examples, however.

Parameterization of Additional Atom Types

Because bond-order potentials were designed to model chemical reactivity in covalent systems, it was logical to adopt the first-generation REBO formalism to model Si-H systems [6]. Parameterizations also exist for Si-F and Si-Cl. The first attempts to construct bond-order potentials capable of modeling three different atom types were two independently developed C-Si-H potentials. The C-Si-H potential of Beardmore and Smith is a hybrid potential which makes use of the C-H, Si-H, and Si-C parameters developed by Brenner [5], Murty and Atwater [6], and Tersoff, respectively. Because it is a hybrid potential, it inherits the strengths and weaknesses of the potentials on which it is based. The C-Si-H potential of Dyson and Smith, known as the extended Brenner (XB) potential, adopts the C-H parameters of Brenner and develops three new sets of parameters for Si-Si, Si-C, and Si-H [7] while maintaining the same functional form.

When chemical reactions occur during a simulation, the bond order of atoms must change in a

continuous manner from reactant to product values. This is accomplished in the first-generation REBO hydrocarbon potential through the use of a bicubic spline $H_{ij}(N_i^H, N_i^C)$ (Eq. 6), which is a function of the number of neighbors surrounding atom i. To extend a bond-order potential to three atom types, such as Si-C-H, the spline function can be extended into ways. First, the cubic-spline function can be extended to three dimensions, such that it has the form $H_{ij}(N_i^H, N_i^C, N_i^{Si})$. This method provides the most flexibility in adjusting the bond order but can be difficult to code. Alternatively, because C and Si have similar electronegativities and possess similar bonding characteristics, a two-dimensional cubic spline that considers the number of neighbors of C and Si together, i.e., $H_{ij}(N_i^H, (N_i^C + N_i^{Si}))$, could also be used. This approach was adopted in both the Si-C-H potentials discussed above. This simplification has the advantage of simplifying the MD code. However, it is at the expense of a considerable loss in flexibility in fitting and, therefore, accuracy in reproducing a wide range of properties. For example, the errors in the energies of organosilane molecules calculated with the XB potential systematically increase with increasing molecule size (Fig. 1).

The performance of the XB potential was evaluated by Sbraccia et al. in 2002 [8]. Several
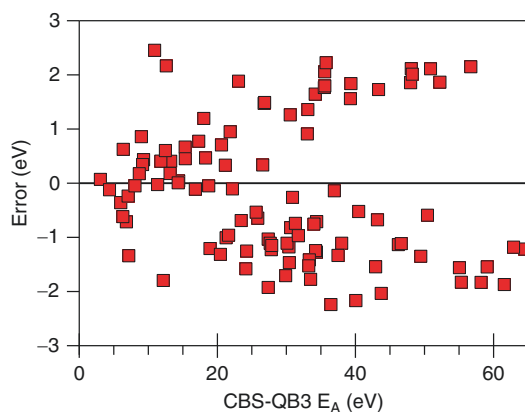
Reactive Empirical Bond-Order Potentials, Fig. 1 Error in atomization energies of $Si_xC_yH_z$ molecules predicted using the mXB potential relative to the atomization energy calculated using the CBS-QB3 method in the ab initio code Gaussian03 illustrating increasing error with atomization energy E_A (increasing molecule size)

problems, such as spurious minima in some potential energy curves, the incorrect energetic ordering for chemisorption of hydrocarbon molecules on Si surfaces, and over-coordinated carbon and silicon atoms were identified. These problems were corrected by altering a minimal set of parameters and the C–C interaction cutoff values. This parameterization is known as the modified extended Brenner (mXB) potential. It should be noted, however, that these changes do not improve the elastic properties of the solid phases of carbon or silicon [9] or the energies of organosilane molecules shown in Fig. 1.

Extension to Metals

Recently, a classical many-body potential for transition metal (Fe, Co, Ni) carbide clusters based on the bond-order type potential was developed [10]. The covalent interactions are based on the first-generation REBO potential. Deviations from the REBO potential formalism occur in the bond-order terms for carbon–metal interactions. The bond order for carbon–metal interactions has the form where b and δ are fitting parameters. For metal–metal interactions, instead of $B^* = \left(1 + b\left(N^C - 1\right)\right)^{\delta}$, multiplying V^A by B^*, the binding energy D_e and equilibrium bond length R_e are expressed as functions of the metal coordination number N^M and are given by

$$D_e = D_{e1} + D_{e2}exp\left\{-C_D\left(N^M - 1\right)\right\} \qquad (7)$$

$$R_e = R_{e1} + R_{e2}exp\left\{-C_R\left(N^M - 1\right)\right\}. \qquad (8)$$

The reader is referred to Shibuta and Maruyama [10] for all parameter values appearing in Eqs. 7 and 8.

Limitations

Before undertaking MD simulations with empirical potentials, careful thought should be given to the selected potential's inherent strengths and weaknesses and how these may impact the phenomena to be simulated. For example, multiple potential energy functions exist for modeling Si, including three versions of Tersoff's bond-order potential [4,11], SW [1], and the environment-

dependent interatomic potential (EDIP) [12]. The majority of empirical potentials for Si do not accurately describe the three elastic constants for diamond cubic silicon [13]. In contrast, the EDIP and the 2B-Si [9] potentials do a good job describing elastic constants. However, the bond lengths for many of the crystalline silicon structures are overestimated when using the EDIP potential. For a more complete comparison of the strengths and weaknesses of Si potentials the reader is referred to Balamane et al. [13] and Schall et al. [9].

In general, four limitations have been identified with the first-generation REBO potential. First, the functional form of the pair potentials was not flexible enough to allow a parameter set that could simultaneously reproduce structural energies and force constants. In his original work, Brenner provided two parameter sets, one that provided an excellent fit to bond energies and a second that provided a to fit force constants [5]. Second, because the zero-Kelvin elastic constants C_{11}, C_{12}, and C_{44} for diamond were not included in the fitting database, they are not accurately reproduced by the potential [14]. Third, both the attractive and repulsive pair terms have finite values at zero separation. Thus, in highly energetic atomic collisions, it may be possible for an atom pass through another atom without being repulsed. Finally, the derivatives of the cutoff function $f_{ij}(r)$ are nonrealistic and may lead to spurious minima in energies for certain structures, particularly amorphous carbon [5].

R

Second-Generation REBO Potentials

To correct the shortcomings of the original potential, Brenner and coworkers developed a second generation of the REBO potential [15]. The repulsive and attractive pair terms have the forms

$$V^R(r) = f^c(r)(1 + Q/r)Ae^{-\alpha r} \qquad (9)$$

and

$$V^A(r) = f^c(r) \sum_{n=1,3} B_n e^{-\beta_n r} \qquad (10)$$

where A, Q, α, B_n and β_n are fitting parameters. The screened Coulomb function used for the repulsive pair interaction goes to infinity as interatomic distances approach zero, and the attractive term has sufficient flexibility to simultaneously fit the bond properties that could not be fit with Morse-type terms. The bond-order B_{ij} term is also significantly different from either the Tersoff or first-generation REBO potential. Separate terms are included that depend on local coordination and bond angles, radical character and conjugation, and dihedral angle for C–C double bonds.

By altering the functional form and expanding the fitting database, the second-generation REBO potential is able to provide significantly better descriptions of bond energies, lengths, and force constants for hydrocarbon molecules relative to the first-generation REBO potential. In addition, the zero-Kelvin elastic properties, interstitial defect energies, and surface energies for diamond are well reproduced. This potential also qualitatively reproduces the correct behavior of the elastic constants C_{11} and C_{44} of diamond with temperature. Coincidentally, the relatively poor agreement of the calculated C_{12} with the experimentally determined data combined with a fast softening of C_{11} with temperature results in bulk moduli that agree fairly well with the experimental values (Fig. 2) [14].

As a first step toward developing a Si–C–H bond-order potential, Schall et al. recently published parameters for silicon potential (2B-Si) based on the second-generation REBO formalism [9]. This potential does a reasonable job producing a broad range of properties of diamond cubic and amorphous silicon. In principle, the second-generation REBO formalism could be extended to a wide range of interatomic interactions to model chemical reactivity in nanoscale systems that contain multiple atom types. For example, Sinnott and coworkers used density functional theory to develop parameters within the second-generation REBO formalism for C–O, O–O, and O–H [16]. Because this potential is based on the second-energy REBO potential, it possesses all of its inherent strengths and weaknesses, such as a lack of intermolecular forces.
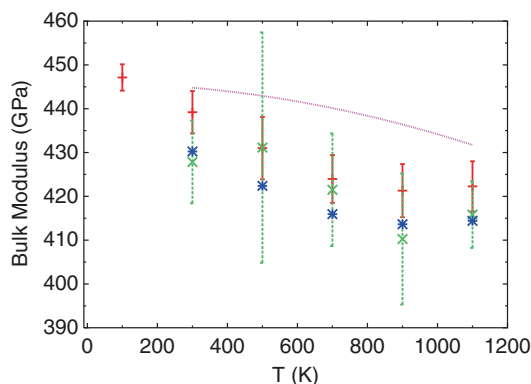
Reactive Empirical Bond-Order Potentials, Fig. 2 Bulk modulus of diamond as a function of temperature. Bulk values were calculated from the elastic constants using the relationship $B = \frac{1}{3}(C_{11} + 2C_{12})$. (+ signs), (*crosses*), and (*asterisks*) symbols represent values calculated with the direct, strain fluctuation, and stress fluctuation methods, respectively

The addition of a third atom type complicates the functional form of the bond order as discussed above. This potential suffers from the limitation discussed above by making use of a two-dimensional cubic spline instead of a three-dimensional cubic spline in the bond-order function. In addition, O is treated in a similar manner as C, despite the inherent differences in electronegativities of the two atoms. Additional parameter sets have also been developed recently by this group for Mo–S and C–F.

Covalent + Intermolecular Forces

To model chemical reactions in condensed phases, such as liquids, graphite, and self-assembled monolayers (SAMs), intermolecular forces must be added to the REBO potential. Goddard and coworkers added intermolecular forces to the first-generation REBO potential for hydrocarbons. Distance-based switching functions were used to turn the intermolecular forces "on" and "off." Intermolecular forces and torsional interactions have also been added to the second-generation REBO potential [17]. This potential, known as the adaptive intermolecular reactive empirical bond-order (AIREBO) potential, uses a unique algorithm to determine the conditions under which the intermolecular forces should be "on" and "off."

The AIREBO potential includes intermolecular interactions between nonbonded atoms and torsional interactions associated with a connected sequence of three bonds. This extension makes the AIREBO potential a particularly attractive tool for studying interfacial/tribological systems and liquids. The discussion below focuses on how intermolecular interactions are introduced without compromising the reactivity of the potential. For a discussion of how torsional interactions are modeled, the reader is referred to the original publication [17].

Intermolecular interactions are modeled through a Lennard-Jones (LJ) potential,

$$V^{LJ} = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right). \quad (11)$$

Only four parameters are taken as independent: σ_{CC}, σ_{HH}, ε_{CC} and ε_{HH}. The heterogeneous parameters are fixed by Lorentz-Berthelot combining rules,

$$\sigma_{CH} = \frac{1}{2}(\sigma_{CC} + \sigma_{HH}) \quad (12)$$

and

$$\varepsilon_{CH} = \sqrt{\varepsilon_{CC}\varepsilon_{HH}}. \quad (13)$$

To merge this with a reactive potential, the LJ potential may be either completely or partially "turned off" in response to the chemical environment of the interacting pair. The key feature of the AIREBO potential is that it provides a means of smoothly interpolating between purely bonded and nonbonded interactions through a set of three switching functions S_{distance}, S_{bond}, and $S_{\text{connectivity}}$. Each one of these switching functions may turn off the LJ interaction partially or entirely,

$$E^{LJ} = (1 - S_{\text{distance}}S_{\text{bond}})(1 - S_{\text{connectivity}})V^{LJ}. \quad (14)$$

For each switching function, a value of 1 is associated with turning off the LJ interaction completely, 0 is associated with a full LJ interaction, and values in between are associated with partial LJ interactions.

The distance-based switch, S_{distance}, is 1 for distances below r^{LJmin} and 0 for distances above r^{LJmax}. A cubic-spline function is used to interpolate at distances falling between the setpoints r^{LJmin} and r^{LJmax} to give a value for the switch between 0 and 1. The setpoints are fixed by requiring that the LJ potential minimum remain unchanged and that no artificial repulsive barrier is present as the LJ potential is switched off. These constraints require that $r^{\text{LJmin}} = \sigma_{ij}$ and $r^{\text{LJmax}} = 2^{1/6}\sigma_{ij}$.

The hypothetical bond-order-based switch S_{bond} interpolates to give values between 0 and 1 in the region $b^{\text{min}} < b^* < b^{\text{max}}$ using a cubic spline, 0 for a bond-order below this range and 1 for a bond-order above this range. The bond-order b^* is evaluated at r^{LJmin} for pairs separated by intermolecular distances (distances larger than r^{LJmin}), and thus at these distances b^* is a *hypothetical* bond-order that assesses the potential for these atoms to bond they approach one another. In the calculation of b^*, the distances of each atom in the pair to its neighbors remain unchanged.

The distance-based switch $S_{\text{connectivity}}$ takes into account the set of all one-, two-, and three-bond sequences that connect the pair of atoms under consideration. For each atom pair in each sequence of bonds, a weight is calculated that is 1 for distances below r^{LJmin} and 0 for distances above r^{LJmax}. These setpoints are the same as in S_{distance}; however, for compatibility with the original REBO potential, a shifted half-period cosine function is used to interpolate between r^{LJmin} and r^{LJmax}. A weight for a sequence of bonds is calculated as the product of the sequence's individual bond weights. The $S_{\text{connectivity}}$ switch is taken to be the largest sequence-weight found in the total set of sequences. In the adopted form for E^{LJ}, the switches S_{distance} and S_{bond} appear together as a product while $S_{\text{connectivity}}$ appears on its own. Consequently, a full LJ interaction will be included for atom pairs that are not (1,2), (1,3), or (1,4) neighbors and are *either* beyond the cutoff distance r^{LJmax} *or* have a bond-order below b^{min}.

The treatment of the van der Waals interactions can be improved in the AIREBO potential by

treating the long-range forces so that the chemical environment is taken into account [18]. In the spirit of the REBO potential, the LJ parameters for carbon are functions of the number of N^C and N, or $\sigma_C(N^C, N^H)$ and $\varepsilon_C(N^C, N^H)$. Two-dimensional cubic-spline functions are used to interpolate between values with integer numbers of neighbors.

Modeling Reactivity in Systems Containing Charge

The electronic properties and nature of bonding in molecules and solids stems from the distribution of charges within the system of interest. The simplest way to add charges to MD simulations is to assign a fixed charge to each atom in the simulation. Fixed-charge schemes have several limitations; however, the most important of which is their inability to redistribute charges in a physically realistic manner. Charge clouds can be distorted when molecules are brought into close proximity, without reacting. When chemical reactions occur, charge needs to be transferred. For instance, the charge states of hydrogen and chlorine change dramatically when they approach each other to form an HCl molecule.

The first atomic-level, charge transfer model was developed by Rappé and Goddard [19] and is known as the electronegativity equalization (QEq) method. In this method, the charges in the system distribute themselves so that the electronegativity at each atomic site is equalized for a given nuclear configuration. This is equivalent to saying that the driving force to shift charge from one atom to another is zero, or the derivatives of the potential with respect to charge are zero. The self-energy $E_i^s(qi)$ on particle i is a function of the charge of the ion and is expressed in terms of the charge, electronegativity χ, and atomic hardness (a term related to the change in energy as a function of charge). Obtaining the charge on each atom for each nuclear configuration of N atoms requires taking the inverse of an N by N matrix. This method was utilized by Streitz and Mintmire in their examination of the aluminum/alumina interface using their ES + EAM potential [20].

Inversion of an N by N matrix at each MD time step usually scales as $O(N^2)$ and becomes computationally intractable as the size of the system increases. Several alternative numerical methods have been developed to solve the N linear equations in charge associated with the N by N matrix. To date, these methods have not found widespread application.

To avoid the matrix inversion, Rick et al. developed an approximate method to equalize the electronegativity based on an extended Lagrangian approach. In this method, charges are treated as dynamic variables that evolve explicitly with time [21]. Equations of motion for the positions and charges of the atoms can be shown to be:

$$m_i\ddot{r}_i = \frac{\partial}{\partial r_i}E_T(\{r_i\}\{q_i\}) \qquad (15)$$

and

$$s_i\ddot{q}_i = \frac{\partial}{\partial q_{r_i}}E_T(\{r_i\}\{q_i\}) \qquad (16)$$

where m_i, r_i, and q_{r_i} are the mass, position, and charge of the ith atom. The term s_i is a fictitious mass of the charge, which is chosen to provide numerical stability. The individual charges respond to deviations of the electronegativity equilibration by moving to a new charge which more closely satisfies the equilibrium condition. One advantage of this approach is that both equations of motion Eqs. 15 and 16 can be integrated using standard numerical integration algorithms. In addition, because the matrix-based solution for determining charge is not used, the code remains $O(N)$, a crucial requirement for large-scale simulations.

The Lagrangian-based dynamic-charge method, which amounts to an approximate implementation of the electronegativity equalization condition, suffers from one significant drawback, however. The dynamic-charge method is closely related to a single iteration, steepest-descent search for the electrostatic energy minimum. The magnitude of the charge transfer is controlled by a

constant factor for all atoms and simulation steps. Thus, the charge "lags" behind the atomic positions used in minimizing the energy and, as a result, the electronegativity equalization condition is not satisfied. To remedy this situation, Ma and Garofalini have introduced the iterative fluctuation charge (IFC) model [22]. The IFC model is able to satisfy the electronegativity equalization condition by performing multiple iterations for electrons (charge) while keeping the atomic positions fixed. The exact numerical criterion, or convergence factor Δ, for electronegativity equalization is $\Delta\left(\frac{eV}{atom}\right) = \frac{\sqrt{\left(\chi_i - \chi_{eq}\right)^2}}{N} = 0$, where χ_i and χ_{eq} are the electronegativities on the ith atom and the equilibrium electronegativity. In practice, a small convergence factor, such as $\Delta = 1.0 \times 10^{-8}$ is used in MD simulations. In addition, charges will oscillate about their fully converged, equilibrium values unless a damping factor is used. The selection of the damping factor and the fictitious charge-mass values can be optimized to achieve optimal damping while not preventing thermal coupling between atomic and electronic degrees of freedom.

Recently, the extended Lagrangian approach with the damping factor discussed by Ma and Garofalini [22] was used to examine the interface between Si and SiO$_2$ surfaces using a charge optimized many-body (COMB) potential [23]. The potential has the form:

$$E_T = \sum_i \left[E_i^S + \frac{1}{2} \sum_{j \neq 1} V_{ij}\left(r_{ij}, q_i, q_j\right) + E_i^{BB} \right], \tag{17}$$

where E_T is the total potential energy of the system, E_i^S is the self-energy term of atom i, V_{ij} is the interatomic potential between the ith and jth atoms, r_{ij} is the distance of the atoms i and j, and q_i and q_j are charges of the atoms, and E_i^{BB} is the bond-bending term of atom i. The interatomic potential energy V_{ij} consists of four components: short-range repulsion, U_{ij}^R, short-range attraction, U_{ij}^A, long-range Coulombic interaction, U_{ij}^I, and long-range van der Waals energy, U_{ij}^V, which are defined as

$$V_{ij}\left(r_{ij}, q_i, q_j\right) = U_{ij}^R\left(r_{ij}\right) + U_{ij}^A\left(r_{ij}, q_i, q_j\right) \\ + U_{ij}^I\left(r_{ij}, q_i, q_j\right) + U_{ij}^V\left(r_{ij}\right), \tag{18}$$

The repulsive and attractive terms in Eq. 18 are of the Tersoff form:

$$U_{ij}^R = f(r)Ae^{-\lambda r} \text{ and } U_{ij}^A = f(r)b_{ij}Be^{-\alpha r} \tag{19}$$

In Tersoff's original work, A, B, λ, and α are constants and $f(r)$ is a distance-based cutoff function, which limits the range of the potential. In shorthand notation, the energy contribution due the repulsive and attractive terms (i.e., the short-range pair potential) can be written as

$$E_{pair} = \sum_{j \neq i} f(r)Ae^{-\lambda r} + \sum_{j \neq i} f(r)b_{ij}Be^{-\alpha r} = V^R \\ + bV^A \tag{20}$$

where b_{ij} is the bond-order function.

Modeling Charge Transfer During Chemical Reactions

Model fluctuating charges using the extended Lagrangian approach seems ideally suited to examine chemical reactions using MD. However, the use of this approach requires the definition of a charge-neutral entity, such as a molecule or the simulation cell. For example, Yu et al. recognized the problem with defining a charge-neutral simulation cell when simulating Si/SiO$_2$ interfaces. To circumvent this problem, the self-energy terms of the COMB potential were modified [23]. Rick et al. defined each water molecule as a charge-neutral entity in their examination of the structural properties of water [21]. Charge equilibration (QE) methods have been applied successfully to relatively homogeneous small-molecule systems near equilibrium. However, application of these methods to large molecules yields atomic charges and molecular dipole moments that are too large. This problem has been traditionally addressed by constraining sub-entities of large molecules to be charge neutral. This is problematic in a reactive environment where it would not be possible to

meaningfully adhere to a fixed set of molecule-based charge-neutral entities.

Recently, Mikulski et al. developed a method to integrate charge equilibration with bond-order potentials that does not require the assignment of charge neutrality to molecules [24]. The bond-order potential/charge equilibration (BOP/SQE) method is a bond-centered approach, where the charge of atom i is the sum of all charges transferred to it across each of its bonds. This has been referred to as a split-charge equilibration (SQE). The problem of the overestimation of dipole moments of large molecules is not solved by merely adopting the bond-centered (SQE) approach. A novel method is needed whereby each bond partly decouples from its bond network and settles into its own equilibrium. The BOP/SQE method connects the bond order of each bond to an amount of shared charge in each bond and interprets the split charge as an imbalance in where the shared charge is concentrated. Split charges cannot grow in size beyond this shared-charge limit, and if the nature of the bond is covalent, the size of the split charges should not get close to this limit. Thus, a fractional split charge is defined as the ratio of the split charge $\overline{f}_{ij}$ on atom i transferred from atom j to its shared charge $\overline{q}_{ij}^{\max}$, charge on atom i is expressed as

$$Q_i = \sum_j \overline{f}_{ij}\,\overline{q}_{ij}^{\max}. \qquad (21)$$

The shared-charge limit $\overline{q}_{ij}^{\max}$ is calculated directly from the bond order bij assigned to the bond from the BOP, $\overline{q}_{ij}^{\max} = \overline{q}_{ij}^{\max}\big(b_{ij}\big)$. Focus is now shifted from equilibrating split charges to equilibrating fractional split charges.

The equilibration process for bond ij is controlled by three factors: (1) The constant pull to increase charge separation associated with the traditional constant electronegativity difference $|\chi_i - \chi_j|$; (2) Coulomb interactions within the bond and with the surrounding environment; (3) A restoring force that approaches and infinite wall as the shared-charge limit is approached. The first two factors are analogous to the SQE condition but are applied to $\overline{f}_{ij}$ rather than $\overline{q}_{ij}$. The third, unique, element is implemented through inverse hyperbolic tangent function added to the charge-dependent electronegativity of each fractional split charge,

$$\overline{\chi}_{ij} = \left[\chi_i + J_{ii}^o Q_i + \sum_{k \neq i} J_{ij} Q_k\right] + \overline{\in}_{ij} \tan h^{-1}\big(\overline{f}_{ij}\big). \qquad (22)$$

The scale parameter $\overline{\in}_{ij}$ can be tailored to each pair of atom types. Large values of $\overline{\in}_{ij}$ cause the wall to gradually rise as soon as the split charges grow. The last term in this equation, or the charge-asymmetry penalty, serves to physically constrain the growth of the split charges and partially decouples each bond from its bond network. Therefore, the BOP/SQE formalism becomes a locally driven process that does not map back to the global equilibration of an entire molecule. With the selection of the scale parameter $\overline{\in}_{ij}$ to reproduce the dipole moment of water, this method was shown recently to reproduce the correct trends in molecular dipole moments of long-chain alcohols (Fig. 3).

It should be noted, that other methods for handling chemical reactions where charge is included are being developed [25]. Because these methods are not bond-order potentials, they are not included in this overview.

Improved Description of Alloys

In a recent study of the thermodynamics an empirical potential description of Fe–Cu alloys, Caro et al. concluded that to study alloys in an empirical framework, heteronuclear interactions require additional modifications to the empirical formalism to incorporate concentration dependent interactions that reproduce the magnitude of excess enthalpy of mixing and asymmetry around the equi-atomic composition [26]. To illustrate this point, data for the heat of mixing in the Fe-Cr system as determined in an ab initio study conducted by Olsson has been replotted in Fig. 4. This figure clearly illustrates the deviation from ideality with both positive and negative heat of mixing in evidence.
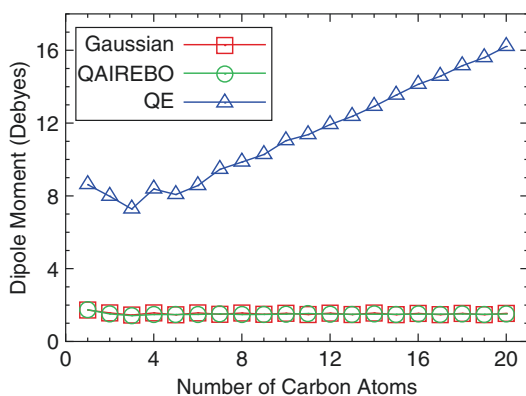
Reactive Empirical Bond-Order Potentials, Fig. 3 Molecular dipole moments of straight chain alcohols calculated using Gaussian, charge equilibration (QE), and a fluctuating-charge MD code that uses the BOP/QE method of charge equilibration and qAIREBO potential
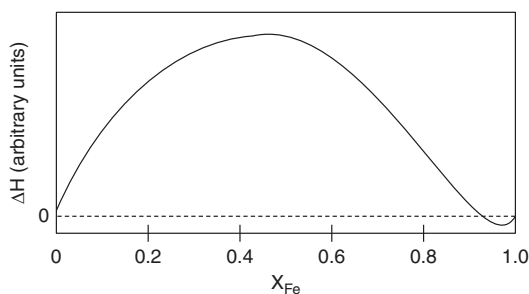
Reactive Empirical Bond-Order Potentials, Fig. 4 The enthalpy of mixing ΔH for the Fe–Cr system as a function of Fe fraction. (replotted from the work of Olsson) The solid line was calculated via first-principles. The dashed line indicates a linear interpolation corresponding to an ideal solution

Traditionally, empirical models for alloys use the Lorentz-Berthelot mixing rules (Eqs. 12 and 13). These rules reproduce the ideal heat of mixing, which is, strictly speaking, only accurate for dilute solutions of solute atoms in the host matrix. Caro et al. has proposed a method for reproducing the thermodynamic properties of concentrated alloys through a modification of the embedded atom method (EAM) formalism. The basic EAM description gives the total energy as a the sum of a pair potential, V, and an embedding function, F, which in turn depends on the electron density, ρ

$$E_{EAM} = \frac{1}{2}\sum V(r) + \sum F\left(\sum \rho(r)\right) \quad (23)$$

Consideration of first-principles calculations gives important information about the behavior of these functions. First, the embedding energy relative to the free atom energy goes to zero for zero electron density and should have a negative slope and positive curvature for the background electron density in metals. Second, the pair interaction term $V(r)$ is purely repulsive.

Caro et al. eschewed the commonly used Lorentz-Berthelot mixing rules and developed a new functional form for the cross potential for random alloy mixtures:

$$E_{RAND} = x_A^2 V_{AA} + x_B^2 V_{BB} + 2x_A x_B V_{AB} \\ + x_A F_A(\widetilde{\rho}) + x_B F_B(\widetilde{\rho}). \quad (24)$$

Here x_A and x_B are the atomic fractions of alloy element A and B, respectively. The pair potential for each element denoted as V_{AA} or V_{BB}. The model adopts the assumption that atoms are embedded in the same average environment, through $F(\widetilde{\rho})$ where $\widetilde{\rho}$ is the average electron density seen by either atom type A or B in the host material. Caro et al. found that the contributions to the formation energy from the embedding term were negligible, leaving only the pair potential as the sole contributor to the formation energy for alloys. By assuming V_{AB} is a function of both composition x and distance r and can be separated into a product $h(x)$ $u_{AB}(r)$, an appropriate choice of the cross term is

$$V_{AB}(x,r) = h(x)\frac{1}{2}[V_A(r) + V_B(r)] \quad (25)$$

Where $h(x)$ is a fourth-order polynomial. This selection of the cross term allows for the description of any type of formation energy curves, with an ideal solution of $h(x) = 1$, for a regular solution. Through appropriate fitting of the polynomial coefficients, a positive or negative heat of mixing can be achieved. Stukowski et al. recently published a version of this potential function.

The methodology described above can also be extended to Tersoff-type potentials. The foundation of this approach was laid out by Brenner who

recognized that the EAM potential and Tersoff many-body potential are formally identical in the appropriate limits. The essential result is that it is reasonable to assume that embedding function $F(\rho(r))$ in the EAM formalism is equal to the attractive part of the Tersoff pair potential, bV^A. Similarly, it is a logical to assume that the repulsive terms in both EAM and Tersoff are equivalent. With these points in mind, a combination of the Tersoff potential with Caro's compositional dependence method can be expressed in shorthand notation as

$$E_{RAND} = x_A^2 V_{AA}^R + x_B^2 V_{BB}^R + 2x_A x_B V_{AB}^R + x_A \tilde{b} V_{AA}^A$$
$$+ x_B \tilde{b} V_{BB}^A. \tag{26}$$

where V_{AA} and V_{BB} are the repulsive and attractive pair terms, denoted by superscripts ^{R} or ^{A}, respectively, for the pure elements A or B as given in Eq. 3. The repulsive cross term V_{AB} is given in Eq. 26. Schall et al. have fit a multicomponent bond-order potential for Si–C–H systems, and it was found that the average bond-order $\tilde{b}$ must be normalized using a scaling factor to ensure compatibility between bond orders for derived for different pure element systems. The average bond order is then given as

$$\tilde{b} = \chi_{ij} \frac{b_{ij} + b_{ji}}{2}. \tag{27}$$

The normalizing factor χ_{ij} is equal to one for homonuclear atom pairs. For heteronuclear atom pairs, χ_{ij} is determined by fixing the average bond order to a value of one for a specified crystal structure. For example, in the Si–C system, χ_{ij} was determined by setting the average bond order for the β-Zn phase of SiC solving Eq. 27 for χ_{ij}. A similar scheme was utilized by Tersoff for multicomponent systems.

Summary and Future Directions

In the foregoing sections, a discussion of reactive empirical bond-order potentials was given.

Such potentials were developed to describe the universal relationship between binding energy and bond length and have been extended from simple descriptions of single elements to multicomponent systems and then to a combination of solids and molecules. The motivation for the development of these potentials was to provide a computationally efficient means to model chemical reactivity, including an accurate description of bond-breaking and forming processes, while still using sound quantum mechanical principles. The original formalism has since been expanded to include descriptions of long-range interactions, torsion, charge transfer, and descriptions of alloys. This family of potentials has since been used to describe a wide variety of processes in nanomaterials systems. One serious limitation of the family of reactive empirical bond-order potentials is that only a relative few materials systems have been parameterized. As mentioned previously, parameterization of a potential involves quite a bit of art as well as science. In the future, semi-automated methods such as force matching that combine ab initio force calculations systematically with parameter fitting could help overcome this hurtle. In addition, the incorporation of more than three elements presents a significant challenge due to the complexity of the spline functions used to interpolate the bond order correction terms. To overcome this barrier, either compromises in the accuracy of the potential or new developments in multidimensional interpolation functions must be made.

Cross-References

▶ Computational Study of Nanomaterials: From Large-Scale Atomistic Simulations to Mesoscopic Modeling

References

1. Stillinger, F.H., Weber, T.A.: Computer simulation of local order in condensed phases of silicon. Phys. Rev. B **31**, 5262 (1985)

2. Mackerell, A.D.: Empirical force fields for biological macromolecules: overview and issues. J. Comput. Chem. **25**, 1584 (2004)

3. Abell, G.C.: Empirical chemical pseudopotential theory of molecular and metallic bonding. Phys. Rev. B **31**, 6184 (1985)

4. Tersoff, J.: New empirical model for the structural properties of silicon. Phys. Rev. Lett. **56**, 632 (1986)

5. Brenner, D.W.: Empirical potential for hydrocarbons for use in simulating chemical vapor deposition of diamond films. Phys. Rev. B **42**, 9458 (1990)

6. Murty, M.V.R., Atwater, H.A.: Empirical interatomic potential for Si-H interactions. Phys. Rev. B **51**, 4889 (1995)

7. Dyson, A.J., Smith, P.V.: Extension of the brenner empirical interatomic potential to C-Si-H systems. Surf. Sci. **355**, 140 (1996)

8. Sbraccia, C., Silvestrelli, P.L., Ancilotto, F.: Modified XB potential for simulating interactions of organic molecules with Si surfaces. Surf. Sci. **516**, 147 (2002)

9. Schall, J.D., Gao, G., Harrison, J.A.: Elastic constants of silicon materials calculated as a function of temperature using a parametrization of the second-generation reactive empirical bond-order potential. Phys. Rev. B **77**, 115209 (2008)

10. Shibuta, Y., Maruyama, S.: Bond-order potential for transition metal carbide cluster for the growth simulation of a single-walled carbon nanotube. Comput. Mater. Sci. **39**, 842 (2007)

11. Tersoff, J.: Empirical interatomic potential for silicon with improved elastic properties. Phys. Rev. B **38**, 9902 (1988)

12. Justo, J.F., Bazant, M.Z., Kaxiras, E., Bulatov, V.V., Yip, S.: Interatomic potential for silicon defects and disordered phases. Phys. Rev. B **58**, 2539 (1998)

13. Balamane, H., Halicioglu, T., Tiller, W.A.: Comparative study of silicon empirical potentials. Phys. Rev. B **46**, 2250 (1992)

14. Gao, G.T., Van Workum, K., Schall, J.D., Harrison, J. A.: Elastic constants of diamond from molecular dynamics simulations. J. Phys. Condens. Matter **18**, S1737 (2006)

15. Brenner, D.W., Shenderova, O.A., Harrison, J.A., Stuart, S.J., Ni, B., Sinnott, S.B.: Second generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J. Phys. Condens. Matter. **14**, 783 (2002)

16. Ni, B., Lee, K.H., Sinnott, S.B.: A reactive empirical bond order (REBO) potential for hydrocarbon-oxygen interactions. J. Phys. Condens. Matter **16**, 7261 (2004)

17. Stuart, S.J., Tutein, A.B.., Harrison, J.A.: A reactive potential for hydrocarbons with intermolecular interactions. J. Chem. Phys. **112**, 6472 (2000)

18. Liu, A., Stuart, S.J.: Empirical bond-order potential for hydrocarbons: adaptive treatment of van der Waals interactions. J. Comput. Chem. **29**, 601 (2008)

19. Rappe, A.K., Goddard, W.A.: Charge equilibration for molecular-dynamics simulations. J. Phys. Chem. **95**, 3358 (1991)

20. Streitz, F.H., Mintmire, J.W.: Electrostatic potentials for metal-oxide surfaces and interfaces. Phys. Rev. B **50**, 11996 (1994)

21. Rick, S.W., Stuart, S.J., Berne, B.J.: Dynamical fluctuating charge force-fields: applicatoin to liquid water. J. Chem. Phys. **1001**, 6141 (1994)

22. Ma, Y., Garofalini, S.H.: Iterative fluctuation charge model: a new variable charge molecular dynamics method. J. Chem. Phys. **124**, 234102 (2007)

23. Yu, J.G., Sinnott, S.B., Phillpot, S.R.: Charge optimized many-body potential for the Si/SiO2 system. Phys. Rev. B **75**, 085311 (2007)

24. Mikulski, P.T., Knippenberg, M.T., Harrison, J.A.: Merging bond-order potentials with charge equilibration. J. Chem. Phys. **131**, 241105 (2009)

25. van Duin, A.C.T., Dasgupta, S., Lorant, F., Goddard, W.A.: ReaxFF: a reactive force field for hydrocarbons. J. Phys. Chem. A **105**, 9396 (2001)

26. Caro, A., Turchi, P.E.A., Caro, M., Lopasso, E.M.: Thermodynamics of an empirical potential description of Fe-Cu alloys. J. Nucl. Mater. **336**, 233 (2005)

Reactive Ion Beam Etching (RIBE)

▶ Ultraprecision Surfaces and Structures with Nanometer Accuracy by Ion Beam and Plasma Jet Technologies

Realtime Green's Function Approach

▶ Nanowire FET Simulations Based on the Nonequilibrium Green's Function Formalism

REBO

▶ Reactive Empirical Bond-Order Potentials

Reduced Graphene Oxide

▶ Plasmonic Photothermal Therapy with Gold Nanorods/Reduced Graphene Oxide Core/Shell Nanocomposites

R

Refractometric Sensing Using Plasmonic Nanoparticles

Erik Martinsson and Daniel Aili
Division of Molecular Physics, Department of
Physics, Chemistry and Biology, Linköping
University, Linköping, Sweden

Synonyms

LSPR-based sensing; Nanoplasmonic sensing

Definition

Refractometric sensing using plasmonic nanoparticles exploits the excitation of localized surface plasmons in metal nanostructures to monitor changes in the dielectric properties in the close vicinity of the nanoparticle surface, caused by, for instance, the binding of a biomolecule to immobilized receptors. The sensitivity to small changes in refractive index enables real-time and label-free detection of molecular interactions at the metal nanoparticle surface.

Overview

Surface plasmons are collective oscillations of the free electron gas at the surface of metals that can be induced by an external electric field, including electromagnetic radiation. When the size of the metal structures is reduced to nanoscale dimensions, the electron oscillations are confined in the nanostructures creating a phenomenon known as localized surface plasmon resonance (LSPR). The excitation of LSPR gives rise to a strong enhancement of electromagnetic fields in the near-field region, and the ability to concentrate electromagnetic radiation into particles of nanoscale dimensions has made plasmonic materials attractive for a large variety of applications, including molecular sensing. For most types of silver and gold nanostructures, the resonance frequency is in the visible region of the electromagnetic spectrum, which gives rise to distinct colors for nanoparticle ensembles. The colors emanate from the extensive absorption and scattering (extinction) of light by the nanoparticles at the LSPR frequency. The resonance conditions are highly dependent on a number of parameters, including (i) the nature of the nanostructures, i.e., their size, shape, and composition; (ii) their interaction with other, closely spaced metal nanostructures; and (iii) the dielectric properties of the surrounding medium. The latter enables metal nanostructures to be used as sensitive probes for detecting changes in the refractive index (RI) in their close proximity caused by for instance the adsorption of biomolecules to the metal surface. The RI dependency of the LSPR is the basis of nanoplasmonic refractometric sensing (Fig. 1). The highly localized sensing volume around metal nanostructures makes these materials attractive signal transducers in miniaturized sensors with high sensitivity to local RI changes.

Metal nanoparticles of different sizes and geometries, such as spherical, rod shaped, and triangular, can be prepared by wet-chemical synthesis and display plasmon resonance frequencies over the entire visible and near-infrared region of the electromagnetic spectrum. Moreover, metal nanoparticles can be immobilized on solid substrates by self-assembly for cost-efficient production of large nanoplasmonic arrays (>1 cm^2), which facilitates development of inexpensive LSPR-based biosensing devices.

Historical Background

The vibrant colors of metal nanoparticles have fascinated mankind for centuries and have been extensively used to decorate glass and ceramics. However, an explanation of the origin of these colors was not presented until the mid-nineteenth century by the pioneering work of Michael Faraday. Faraday synthesized gold nanoparticles by reducing gold salt with phosphorus, and after conducting a series of experiments, he concluded that the colors were caused by the presence of small particles of metallic gold that were suspended in the aqueous media [1]. Some decades
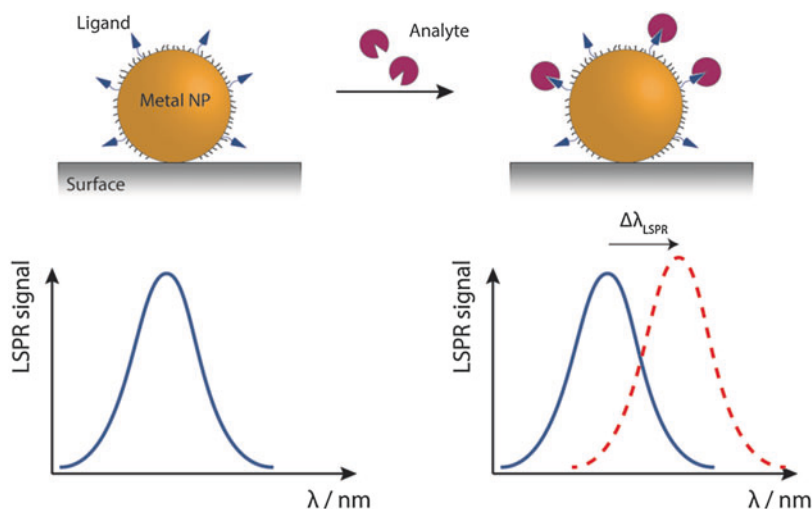
Refractometric Sensing Using Plasmonic Nanoparticles, Fig. 1 Schematic illustration of the basic concept behind refractometric sensing using plasmonic nanoparticles. Molecular interactions that take place at the surface of a metal nanoparticle change the conditions for the LSPR excitation that result in measurable changes in the extinction spectra, which enables real-time and label-free detection of the binding event

later, Richard Zsigmondy could verify the nanoscale dimensions of these particles by investigating them using an ultramicroscope. The techniques for fabrication and characterization of nanomaterials have advanced considerably since then, and it is now possible to produce well-defined nanostructures with respect to physical and chemical properties and to investigate their properties in great detail.

Refractometric sensing using surface plasmons was first described by Liedberg et al. in 1983 [2]. In this seminal work, propagating surface plasmons at the interface between a planar silver film and a dielectric were exploited to monitor antigen-antibody interactions. It took an additional 15 years before Englebienne showed that the same sensing principle also could be applied to localized surface plasmons in metal nanoparticles [3]. Both sensing schemes offer label-free and real-time detection of molecular interactions taking place at (or close to) the metal surface. Biosensing based on surface plasmon resonance (SPR) is today a commercially established technique that is frequently used to characterize a wide variety of biomolecular interactions. LSPR-based techniques are just about to find their way into the commercial market. In contrast to SPR, LSPR enables cheap, miniaturized, and multiplexed sensing using very simple optical setups.
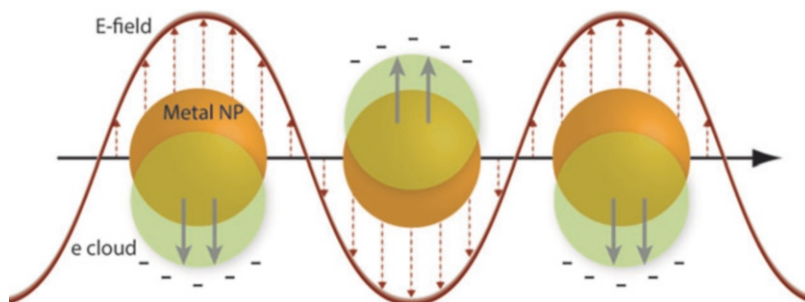
Theory

Metals have partially filled conduction bands, which give rise to "free" electrons that can move within the material. When metal nanostructures are exposed to an external electric field, these electrons are displaced relative to the positive core of the atoms. The resulting induced dipole will oscillate with the same frequency as the external electric field (e.g., electromagnetic radiation). For small nanostructures the electrons will oscillate coherently at a specific frequency (Fig. 2). This creates a distinct peak in the extinction spectrum at the resonance frequency, known as the LSPR peak.

In 1908, Gustav Mie described the interaction between light and small metal nanoparticles in a homogenous medium by analytically solving Maxwell's equations [4]. He derived an expression for the optical extinction (sum of absorption and scattering) cross section (σ_{ext}) for metal nanoparticles much smaller than the wavelength of light ($2r << \lambda$):

Refractometric Sensing Using Plasmonic Nanoparticles, Fig. 2 Schematic illustration of localized surface plasmon resonance in spherical gold nanoparticles induced by an external electric field

$$\sigma_{ext}(\lambda) = \frac{24\pi^2 r^3 \varepsilon_d^{3/2} N}{\lambda \ln(10)} \frac{\varepsilon_i(\lambda)}{\left(\varepsilon_r(\lambda) + 2\varepsilon_d\right)^2 + \varepsilon_i(\lambda)^2}$$

(1)

where r is the radius of the nanoparticle, ε_d is the dielectric constant of the surrounding medium, N is the electron density, and ε_r and ε_i are the real and imaginary part of the complex dielectric function of the bulk metal, respectively. The plasmon resonance condition is fulfilled when $\varepsilon_r(\lambda) = -2\varepsilon_d(\lambda)$ if the imaginary part of the dielectric function is small (low losses). For small gold and silver nanoparticles, this resonance frequency falls in the visible region of the electromagnetic spectrum (gold ~ 520 nm and silver ~380 nm) which explains their distinct colors. However, the resonance frequency can be changed by varying the size and shape of the nanoparticles; thus, the plasmon resonance peak can be tuned over the entire visible region. For larger nanoparticles, the dipolar approximation is no longer valid since the electrons will not experience a homogeneous electric field (phase differences) which means that the expression in Eq. 1 cannot be used to accurately determine their optical properties. Instead, complex numerical methods like discrete dipolar approximation (DDA) and finite-difference time-domain (FDTD) simulations are frequently employed to describe the optical response for larger, spherical, or anisotropic nanoparticles.

Equation 1 shows that the extinction depends on the size and the optical properties of the metal nanoparticles as well as the dielectric properties of the surrounding medium. Consequently, the LSPR excitation will change when the refractive

index of the surrounding medium changes. Biomolecules have a higher refractive index (~1.46) than water (1.33). A binding event of a molecular analyte to a ligand immobilized on the nanoparticle surface displaces water molecules and replaces them with the biomolecule, resulting in an effective increase in the refractive index of the medium surrounding the nanoparticle and a concomitant change in the extinction cross section.

Synthesis of Metal Nanoparticles

Procedures for metal nanoparticle synthesis have improved significantly during the last decades, and it is now possible to produce metal nanoparticles with a range of different structures and morphologies and that are composed of either a single or several different materials, resulting in nanostructures with unique optical properties. One of the most commonly used wet chemistry techniques to produce metal nanoparticles is based on metal ion nucleation using citrate reduction. This method was developed by Turkevich et al. in 1951, where gold chloride ($HAuCl_4$) was reduced by sodium citrate to generate fairly monodisperse gold nanoparticles in the size range of approximately 10–50 nm in diameter (Fig. 3) [5]. The sodium citrate acts both as a reducing agent and electrostatic stabilizer of the resulting nanoparticles. The size of the nanoparticles can be tuned by varying the ratio between the concentrations of gold salt and sodium citrate in water. A common synthesis method to obtain anisotropic gold nanoparticles is the seed-mediated growth method, which
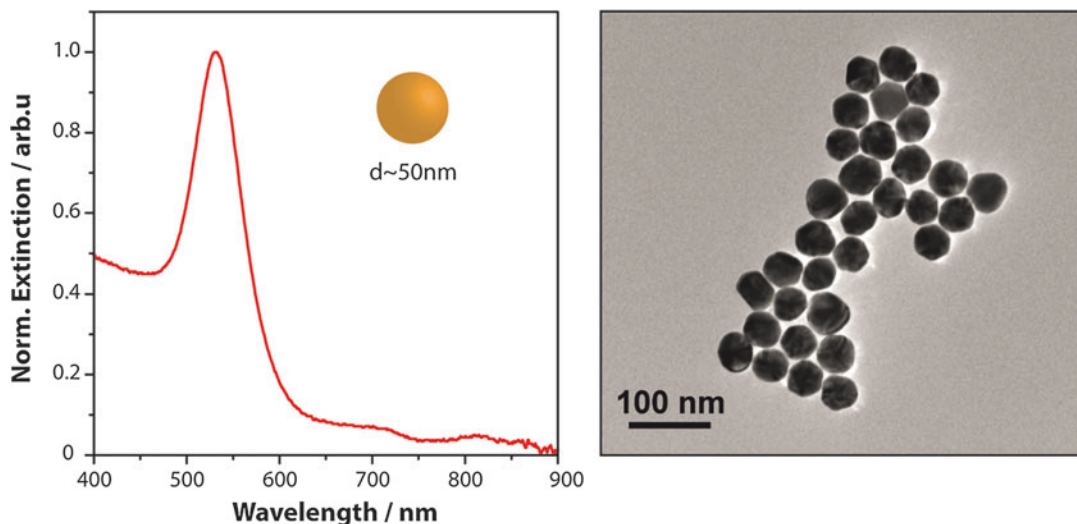
Refractometric Sensing Using Plasmonic Nanoparticles, Fig. 3 Extinction spectra (*left*) and transmission electron microscopy (*TEM*) image (*right*) of spherical gold particles with an average diameter of 50 nm synthesized by citrate reduction of gold salt

utilizes reagents, typically surfactants in order to restrict the growth at certain crystal facets. This method has been used to realize a wide variety of nanostructures including nanorods, nanocubes, stare-shaped nanoparticles, and nanoprisms, which all show different LSPR frequencies.

Refractive Index Sensitivity

Plasmonic metal nanoparticles can be utilized as transducer elements in refractometric sensors because of the sensitivity of LSPR to small changes in refractive index in the immediate vicinity of their surface. The sensitivity (η_B) is defined as the magnitude of the spectral response caused by a change in refractive index (Δn):

$$R = \eta_B \Delta n \qquad (2)$$

where the spectral response (R) can be measured as a change in the plasmon peak wavelength position (nm) or in the signal intensity (extinction). This linear relationship refers to RI changes in the entire sensing volume around the nanoparticles, and η_B is therefore normally referred to as the bulk RI sensitivity, usually expressed in nm/RIU or ext/RIU. In order to achieve a large response for a specific RI change, the bulk RI sensitivity should obviously be as high as possible. Consequently, there are numerous reports on the synthesis of metal nanoparticles exhibiting different RI sensitivities, ranging from ~50 nm/RIU for small, spherical gold nanoparticles to >1000 nm/RIU for more complex, anisotropic nanoparticles [6]. However, Miller et al. have shown that the bulk RI sensitivity for metal nanoparticles is strongly related to their plasmon resonance frequency, and a linear relationship between the two parameters was obtained for particles with a plasmon resonance below 800 nm [7]. Accordingly, nanoparticles that exhibit a plasmon resonance at longer wavelength will display higher bulk RI sensitivity, which means that the aim of creating nanoparticles with high bulk RI sensitivity has merely been quests for synthesizing nanoparticles with longer plasmon resonance wavelength, i.e., lower energy. Silver nanoparticles generally exhibit higher sensitivity as compared to gold nanoparticles because of the different dielectric properties of the two metals. Silver nanoparticles can, however, easily undergo oxidation, which makes them less suitable for most sensing applications.

Metal nanoparticles that exhibit a high bulk RI will not necessarily exhibit a high sensing

R

performance since also the resolution must be taken into account. Large nanoparticles that display a plasmon resonance at longer wavelengths (in the near-infrared region) usually have a broad plasmon band due to dephasing and increased radiative damping. This makes it harder to accurately determine the plasmon peak maxima, which has a negative effect on the resolution. The figure of merit (FOM) is a measure that takes both the sensitivity and the resolution into account and is expressed as:

$$\text{FOM} = \frac{\eta_B}{\text{fwhm}} \quad (3)$$

where η_B is the bulk RI sensitivity and fwhm is the full width at half maximum, i.e., the plasmon peak width (in nm) at half its maximum value. Otte et al. have shown that the FOM value for gold nanorods is maximized for a plasmon resonance close to 700 nm, i.e., where an optimum relationship between sensitivity and resolution can be observed [8]. Also, it is advantageous to stay in this region since cheaper optical components can be used for plasmon resonance excitation and detection.

Sensing Volume

The sensing volume around a metal nanoparticle depends on the spatial distribution of the electromagnetic near field created as a result of the excitation of localized surface plasmons. The intensity of the near field decays exponentially from the metal surface, and the decay length is defined as the distance where the magnitude has decreased by 1/e (~37 %). Both the decay length (~10–20 nm) and the intensity of the near field depend on the physical properties of the metal nanoparticle and the dielectric properties of the surrounding medium. The sensing distance, i.e., the average distance from the metal surface where a change in RI can be detected, is strongly correlated to the decay length of the electric field. Numerous studies have been carried out in order to determine the sensing distance for various metal nanoparticle structures using different layer-by-layer techniques. In general, the reported
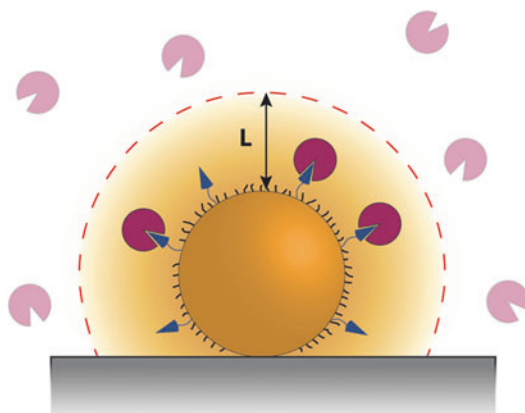
Refractometric Sensing Using Plasmonic Nanoparticles, Fig. 4 Schematic illustration of the highly localized sensing volume around a gold nanoparticle. The sensing volume depends strongly on the decay length (L) of the electromagnetic near field

sensing distances range between 20 and 30 nm. Hence, the ability of metal nanoparticles to detect changes in their surroundings is strictly limited to a small volume around the nanoparticles, which makes them suitable to probe highly localized RI changes, caused by for instance a molecular binding event to an immobilized receptor (Fig. 4). As the sensing response obtained as a result of adsorption of molecules on the nanoparticle surface is more relevant than the bulk RI sensitivity, an expression for the plasmon shift that can be expected when a molecular layer is formed around a metal nanostructure has been derived [9]:

$$\Delta\lambda_{\max} = \eta_B \left(n_{\text{layer}} - n_{\text{medium}} \right) \left(1 - e^{-\frac{2d}{L}} \right) \quad (4)$$

where n_{layer} and n_{medium} are the refractive indices of the adsorbed layer and the surrounding medium, respectively, d is the thickness of the adsorbed molecular layer, and L is the decay length of the electric field. Although, this expression is based on empirical findings, it gives a rather good prediction of the sensing response that can be obtained when a thin film of molecules are adsorbed at the surface of a metal nanoparticle.

Conventional SPR, based on propagating surface plasmons, exhibits a bulk RI sensitivity that is roughly one order of magnitude higher than LSPR. Despite this fact, the two sensing schemes

show a similar sensing performance when considering molecular adsorption close to the metal interface [10]. As the sensing volume of metal nanoparticles is small, the background signal, also known as the bulk signal, can be significantly lower than in conventional SPR-based sensors, which is advantageous in the analysis of low molecular compounds that will only induce small RI changes.

However, the strictly limited sensing volume associated with LSPR-based sensing limits the usage of bulkier surface chemistries and large recognition molecules (e.g., antibodies), since the molecular interactions of interest then will be located outside the actual sensing volume. Thin surface coatings and small recognition elements are thus appropriate to employ in LSPR-based sensing, ideally keeping the molecular interactions in the direct vicinity of the nanoparticle surface. However, it is a great challenge to realize a thin, functional surface coating that exhibits a high resistance against unspecific molecular adsorption.

Single Nanoparticle Detection

The highly localized sensing volume around metal nanoparticles makes them suitable transducer elements for miniaturized sensing. Single nanoparticle detection has been an area of intense research since this technique enables a significant reduction in the absolute detection limit, i.e., fewer analytes per nanoparticle, which is important both for miniaturization and multiplexed sensing. Single nanoparticles also show narrower plasmon bands as compared to an ensemble of particles, which improve the spectral resolution since smaller LSPR shifts can be detected. Dark-field microscopy techniques enable simultaneous single particle spectral plasmon mapping of multiple nanoparticles for single nanoparticle analysis. Although metal nanoparticles have a large extinction cross section, the scattered light intensity from a single nanoparticle is still rather weak, and the signal-to-noise ratio in the obtained extinction spectra is generally relatively low. The overall sensing performance is thus usually better when an ensemble of nanoparticles is used for detection since this gives a higher signal intensity and an improved signal-to-noise ratio.

Substrate Effect

Nanoplasmonic sensing devices where the metal nanoparticles are immobilized on a solid support have several practical advantages as compared to using dispersed nanoparticles since immobilization facilitates handling, washing, storage, and regeneration of the sensor surface. Immobilization of nanoparticles can be achieved by self-assembly, using, for instance, thiol chemistry, or by physisorption using, e.g., electrostatic interactions between charged species adsorbed on the nanoparticle surface and the substrate. The self-assembly process can be controlled by an appropriate and defined surface chemistry at both the substrate and the metal nanoparticles and tuned by changing, e.g., the pH, ionic strength, nanoparticle concentration, or incubation time. However, tethering nanoparticles to a solid support will influence their RI sensitivity since the substrate will occupy a certain fraction of the sensing volume around the nanoparticles. This is known as the substrate effect and results in a reduction in the RI sensitivity. The magnitude of this reduction depends strongly on the fractional surface area of the nanoparticle that is in direct contact with the underlying substrate [11]. Thus, a flat structure like a nanodisk or nanoprism will be affected to a larger extent by the substrate effect than, for example, a nanosphere. Consequently, in order to reduce the influence of a substrate immobilization on the RI sensitivity, the contact area between the nanoparticle and the substrate should be minimized.

In addition, if the substrate has a higher RI value than the surrounding medium, the electric field created around the nanoparticles upon excitation can get redistributed, shifting a large part of the field toward the substrate. This will increase the substrate effect additionally since the RI detection is associated with the electric field. The redistribution of the electric field can, however, be reduced when using substrates with a low RI value, like Teflon ($n \approx 1.32$).

Plasmonic Coupling

Another way to tune the plasmonic response of metal nanoparticles is to utilize dipolar interactions between closely spaced nanoparticles. This is known as plasmonic coupling and results in appearance of hybridized plasmon modes where the plasmon frequency is highly dependent on the separation between the nanoparticles. Plasmonic coupling between closely spaced nanoparticles gives rise to regions where the electromagnetic field is strongly enhanced as compared to the incident electric field that is used for the excitation (Fig. 5) [12, 13]. These so-called hot spots are of large interest for sensing applications and have been extensively utilized in surface-enhanced Raman spectroscopy (SERS) where the signals from Raman-active molecules can be significantly enhanced. In addition, these regions have also been shown to exhibit a higher RI sensitivity, and an improved sensing performance can be achieved if the molecular interactions can be guided into these areas [14]. Plasmonic coupling is also utilized in colorimetric sensing schemes based on analyte recognition-induced changes in colloidal stability. This method was pioneered by Mirkin et al. who used DNA to facilitate an aggregation of gold nanoparticles [15].
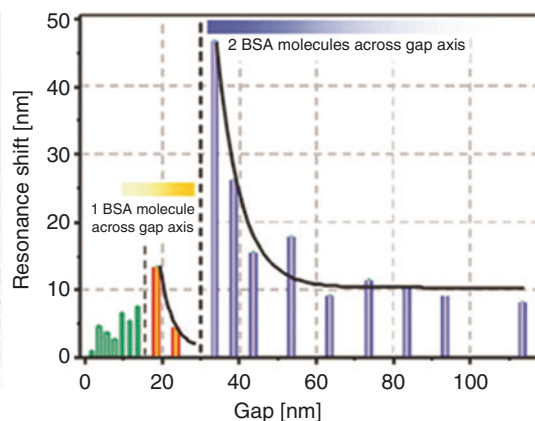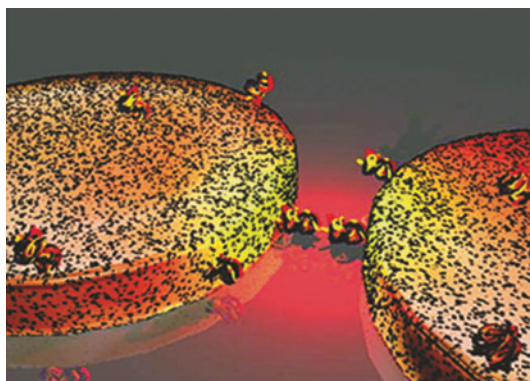
Nanoplasmonic Biosensing

Numerous examples of LSPR-based biosensing schemes have been proposed for detection of a wide variety of analytes including DNA, proteins, and small organic molecules using different recognition elements. Due to the limited sensing volume around metal nanoparticles, LSPR-based biosensing is a suitable method for detection of small analytes. The smaller the analytes are, the smaller the RI change will be upon binding to the ligand, which consequently requires a high sensitivity in order to obtain an acceptable limit of detection. For example, Kabashin et al. showed that small molecular weight compounds can be detected by localized surface plasmons created in gold nanopillars, and they were able to detect biotin (244 Da) instead of streptavidin (52.8 kDa), which is normally detected [16]. In addition, Hall et al. showed that LSPR can be used to monitor

Refractometric Sensing Using Plasmonic Nanoparticles,
Fig. 5 Electric field distribution plots for (**a**) a single gold nanoparticle and (**b**) a nanoparticle dimer. (**c**) Field enhancement factor for a nanoparticle dimer at different gap distances with respect to a single nanoparticle (Reproduced with permission from Ref. [12])
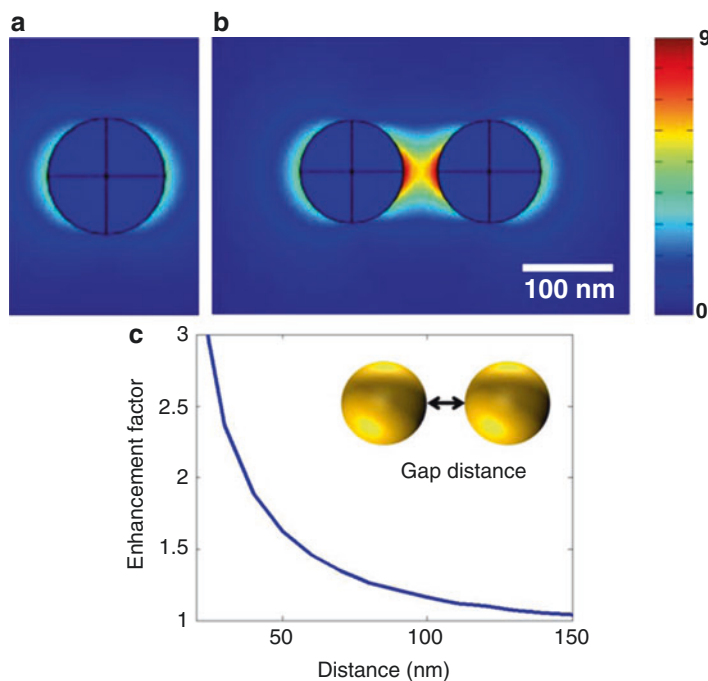
Refractometric Sensing Using Plasmonic Nanoparticles, Fig. 6 Binding of BSA in the gap of a gold nanodimer. By varying the gap distance, the binding of one or two BSA molecules could be identified and discriminated (Reprinted with permission from Ref. [18]. Copyright 2009 American Chemical Society)

protein conformational changes using an array of silver nanoprisms fabricated by nanosphere lithography [17]. Aćimović et al. were able to discriminate between the bindings of one or two BSA (bovine serum albumin) molecules by varying the gap distance between gold nanodisks (Fig. 6) [18]. Recently, a LSPR biochip based on gold nanorods was realized that enabled a multiplexed detection of cancer biomarkers (human alpha-fetoprotein and prostate-specific antigen) down to concentrations of 500 pg/mL [19]. These examples demonstrate the great potential of LSPR in the development of high-performance sensing platforms that can be produced at a low cost.

Future Perspective

Refractometric sensing using plasmonic nanoparticles is a promising strategy for detecting small, local changes in the surrounding dielectric environment. This real-time, label-free sensing technique enables miniaturization and multiplexed analysis, which is difficult to achieve using conventional SPR-based sensing. However, many challenges remain before cost-effective refractometric sensing platforms based on LSPR that exhibit high sensitivity, specificity, and robustness can be practically implemented and can compete with existing sensing technologies. Some of these challenges are discussed below:

Biofunctionalization: The highly localized sensing volume around metal nanoparticles is advantageous in many aspects but also put constraints on the surface chemistry that can be used since the distance between the recognition moiety and the nanoparticle surface must be kept to a minimum. A reliable and biofunctional surface chemistry, which both exhibit selectivity and low unspecific binding, is critical in the development of high-performing sensing devices. In order to fully exploit the advantages of LSPR-based sensing devices, the surface coating must hence be very thin while still displaying low nonspecific adsorption and selective recognition of target analytes.

Reproducibility: Although the techniques for synthesis of metal nanoparticles has advanced significantly during the last decade, there is still room for improvements with respect to batch-to-batch reproducibility and minimizing size and shape polydispersity that give rise to broad plasmon bands and lower the overall sensing performance. Lithographic techniques can be used to make nanostructures and patterns with high reproducibility, but generally at a higher production cost and facing problems with up-scaling as compared to colloidal self-assembly. Self-assembly of metal nanoparticles thus offers a simple and cost-effective way to produce nanoplasmonic substrates, but the assembly conditions needs to

be tightly controlled in order to produce nanopatterns with high reproducibility and to facilitate multiplexed sensing.

Localization: Plasmonic coupling between adjacent nanoparticles gives rise to areas with enhanced electromagnetic fields. If the molecular sensing can be guided into these regions, a significantly higher sensitivity can be obtained. Thus, it would likely be more beneficial to be able to control the orientation and interparticle coupling between closely spaced nanoparticles rather than optimizing the structure of individual nanoparticles. However, it is a great challenge to both obtain well-defined plasmon coupling and to guide the immobilization of recognition elements into these hot-spot regions to exploit the enhanced sensitivity.

Cross-References

▶ Active Plasmonic Devices
▶ Biosensors
▶ Gold Nanorods
▶ Light Localization for Nano-optical Devices
▶ Nanoparticles
▶ Optical Properties of Metal Nanoparticles
▶ Plasmon Resonance Energy Transfer Nanospectroscopy
▶ Plasmonic Structures for Solar Energy Harvesting
▶ Surface Plasmon-Polariton-Based Detectors
▶ Synthesis of Gold Nanoparticles

References

1. Faraday, M.: The bakerian lecture: experimental relations of gold (and other metals) to light. Philos. T. Roy. Soc. **147**, 145–181 (1857)
2. Liedberg, B., Nylander, C., Lundstrom, I.: Surface-plasmon resonance for gas-detection and biosensing. Sensor Actuator. **4**, 299–304 (1983)
3. Englebienne, P.: Use of colloidal gold surface plasmon resonance peak shift to infer affinity constants from the interactions between protein antigens and antibodies specific for single or multiple epitopes. Analyst **123**, 1599–1603 (1998)
4. Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. Ann. Phys. **3**, 377–445 (1908)
5. Turkevich, J., Stevenson, P.C., Hillier, J.: A study of the nucleation and growth processes in the synthesis of colloidal gold. Discuss. Faraday Soc. **11**, 55–75 (1951)
6. Charles, D.E., Aherne, D., Gara, M., Ledwith, D.M., Gun'ko, Y.K., Kelly, J.M., Blau, W.J., Brennan-Fournet, M.E.: Versatile solution phase triangular silver nanoplates for highly sensitive plasmon resonance sensing. ACS Nano **4**, 55–64 (2010)
7. Miller, M.M., Lazarides, A.A.: Sensitivity of metal nanoparticle surface plasmon resonance to the dielectric environment. J. Phys. Chem. B **109**, 21556–21565 (2005)
8. Otte, M.A., Sepulveda, B., Ni, W.H., Juste, J.P., Liz-Marzan, L.M., Lechuga, L.M.: Identification of the optimal spectral region for plasmonic and nanoplasmonic sensing. ACS Nano **4**, 349–357 (2010)
9. Willets, K.A., Van Duyne, R.P.: Localized surface plasmon resonance spectroscopy and sensing. Annu. Rev. Phys. Chem. **58**, 267–297 (2007)
10. Svedendahl, M., Chen, S., Dmitriev, A., Kall, M.: Refractometric sensing using propagating versus localized surface plasmons: a direct comparison. Nano Lett. **9**, 4428–4433 (2009)
11. Martinsson, E., Otte, M.A., Shahjamali, M.M., Sepulveda, B., Aili, D.: Substrate effect on the refractive index sensitivity of silver nanoparticles. J. Phys. Chem. C **118**, 24680–24687 (2014)
12. Chung, T., Lee, S.Y., Song, E.Y., Chun, H., Lee, B.: Plasmonic nanostructures for nano-scale bio-sensing. Sensors **11**, 10907–10929 (2011)
13. Jain, P.K., El-Sayed, M.A.: Plasmonic coupling in noble metal nanostructures. Chem Phys Lett **487**, 153–164 (2010)
14. Feuz, L., Jonsson, P., Jonsson, M.P., Hook, F.: Improving the limit of detection of nanoscale sensors by directed binding to high-sensitivity areas. ACS Nano **4**, 2167–2177 (2010)
15. Mirkin, C.A., Letsinger, R.L., Mucic, R.C., Storhoff, J.J.: A DNA-based method for rationally assembling nanoparticles into macroscopic materials. Nature **382**, 607–609 (1996)
16. Kabashin, A.V., Evans, P., Pastkovsky, S., Hendren, W., Wurtz, G.A., Atkinson, R., Pollard, R., Podolskiy, V.A., Zayats, A.V.: Plasmonic nanorod metamaterials for biosensing. Nat. Mater. **8**, 867–871 (2009)
17. Hall, W.P., Anker, J.N., Lin, Y., Modica, J., Mrksich, M., Van Duyne, R.P.: A calcium-modulated plasmonic switch. J. Am. Chem. Soc. **130**, 5836–5837 (2008)
18. Acimovic, S.S., Kreuzer, M.P., Gonzalez, M.U., Quidant, R.: Plasmon near-field coupling in metal dimers as a step toward single-molecule sensing. ACS Nano **3**, 1231–1237 (2009)
19. Acimovic, S.S., Ortega, M.A., Sanz, V., Berthelot, J., Garcia-Cordero, J.L., Renger, J., Maerkl, S.J., Kreuzer, M.P., Quidant, R.: LSPR chip for parallel, rapid, and sensitive detection of cancer markers in serum. Nano Lett. **14**, 2636–2641 (2014)

Relaxation

▶ Nanomechanical Properties of Nanostructures

Relaxation Calorimetry

▶ Nanocalorimetry

Reliability of Nanostructures

Tuncay Alan
Mechanical and Aerospace Engineering
Department, Monash University, Victoria,
Australia

Synonyms

Mechanical characterization of nanostructures;
Nanomechanics

Definition

The continuous reduction in the size of structural device components promises improved efficiency and, often, new functionalities, which would otherwise be impossible with conventional macroscale devices. As a result, in the past decades numerous microsystem devices with nanoscale mechanical parts have been demonstrated and used in a variety of engineering applications, ranging from electrical circuit components to sensors, actuators, and biomedical applications [1]. This ongoing drive to create smaller new devices to perform critical functions requires significant attention to the mechanical reliability of materials used at nanoscales. To make use of the full mechanical potential of materials at decreasing sizes, fundamental properties, such as the elastic modulus, yield, fracture and fatigue properties, adhesion and intrinsic stresses, should be determined by testing relevant sized samples that have been fabricated or processed using the same procedures as the structure of interest and subjected to the same environmental conditions.

Overview

Even though bulk elastic properties are generally expected to remain constant down to a length scale of several nanometers, the mechanical reliability of materials depends significantly on the specimen size [2]. Despite their high theoretical strengths, most materials used in practical applications cannot sustain very high stresses due to intrinsic defects. An estimate for the minimum defect size that will cause failure can be obtained by an order of magnitude calculation considering Griffith fracture criterion (which suggests that an infinitesimal crack will propagate when the available strain energy is larger than that needed to create new surfaces) [3]. Accordingly, the plane-stress fracture strength of the material is given by

$$\sigma_f = \frac{1}{1.12} \sqrt{\frac{2E\gamma}{\pi l_c}}, \qquad (1)$$

where l_c is the characteristic size of an ideal defect, γ is the surface energy per unit area, and E is the elastic modulus. Considering the theoretical strength of Si (calculated by atomistic simulations), $\sigma_f = 23$ GPa [4], $\gamma = 1{:}25$ J/m^2 [5] and $E = 130$ GPa, Eq. 1 suggests that *a single 1-nm-long edge crack will reduce the strength of an otherwise defect-free Si component by 60 % to 9 GPa!*

Key Research Findings

As the dimensions of a material decrease down to nanometers, approaching the typical size of critical defects, the strength is expected to increase significantly with a potential to reach the theoretical limit. Recent experiments demonstrated that near-theoretical strength values that are more than 25 times higher than typical macroscale strengths could indeed be achieved for 190-nm-thick single crystal silicon structures [6, 7] and 15-nm-thick

R

low-pressure chemical vapor deposited low-stress SiN [8]. These ultrahigh strengths are very promising as they enable novel MEMS devices, which can reliably operate under very high stresses [8]. Yet, at decreasing size scales, the number of intrinsic volume defects decreases and surface-to-volume ratio increases. Hence, surfaces and process parameters gain importance. The above experiments on silicon [6] as well as stochastic simulations [7] show that atomic scale surface features, which would otherwise be negligible in larger structures start acting as defects (Eq. 1) and that they may significantly compromise the reliability of the nano-material. For instance, a slight increase in process induced root-mean-square (rms) surface roughness (from 0.4 to 1.5 nm) decreases the strength of Si nanobeams by approximately 20 % (Fig. 1). A similar effect is observed when the samples are exposed to air for an extended period. When surfaces are not effectively protected against oxidation, 1-nm-thick native oxide islands that form randomly on the surface have been shown to reduce the strength of structures by 30 % [9].

The above examples emphasize that at decreasing size scales, mechanical performance of structures is often dominated by a competition between size, grain structure, surface defects, and environmental conditions. How would failure statistics change as the average defect size approaches the size of the device? At what size scales would the environmental factors start dominating failure? Could the mechanical service life of structures be improved by modifying process parameters? Answering these three questions is crucial to effectively characterize mechanical reliability of any material used at the nanoscale.

Methodology

So far, various nanomechanical test procedures have been proposed, and a few commercial devices have appeared in the market as ultimate nanoscale test platforms. Initial studies have focused on commonly used MEMS materials, and more recently, nanomechanical experiments were also adapted to characterize biological
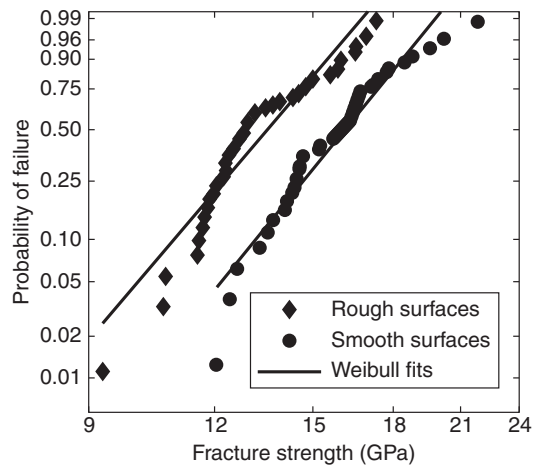
Reliability of Nanostructures, Fig. 1 Fracture strength statistics of nanobeams with different surface roughnesses. The estimated strength of the beams increases from 14.2 to 16.8 GPa as the surface roughness decreases from 1.5 to 0.4 nm

nanostructures such as cells and proteins. Due to differences in experimental techniques, geometric properties of the samples used and fabrication procedures, there is a significant variation in the failure statistics reported by different groups.

Because of the number of parameters affecting mechanical response, there are still no standard techniques and it is difficult to come up with a universal method. The available experiments generally involve variations to well-established macroscale tensile, bulge, and bending tests. The choice of procedure depends on several factors among which, the most important are: size (a micron and nano-sized sample cannot always be tested by the same instrument), required statistical accuracy (due to their complexity some methods do not allow easy repetition), tested material (not all materials are compatible with the fabrication procedures necessary for the preparation of the samples).

Tensile Tests

The main advantage of a tensile test is its analytical simplicity. Since the gauge length of the specimen is under uniform stress and strain fields, load–displacement curves can easily be used to determine elastic modulus and failure properties

directly. Required displacement and load resolutions can generally be accommodated by commercially available piezoelectric or capacitive actuators and high-precision load cells [2]. Yet, handling tensile test specimens remains a major challenge: It is very difficult to grip an individual micron-sized sample from both ends to apply a uniform load. Sharpe and collaborators got around this problem by electrostatic gripping. They used an external probe that was temporarily attached to hundreds of micrometer long test samples by electrostatic attraction forces. The probe, which was connected to a load cell was then manipulated to load the samples to failure, while displacements were monitored via image processing. There have been many variations of this procedure as reviewed in detail elsewhere [2]. Using a different approach, Gaspar and coworkers [10] developed a homebuilt setup which uses a commercial load sensor and an automated stage to test micron-scale tensile samples that were supported by MEMS springs. The biggest advantage of the method, despite the relatively large length scales, is that it enables a very large number of tests to be performed at a short time.

The experiments that focused on smaller devices have generally employed specifically designed MEMS-based test setups, where the sample is integrated with the test device. For instance, Haque and coworkers [11] developed a tensile tester that allows in situ analysis of deformation inside an electron microscope while the sample is being loaded by MEMS beams. Similarly, some recent commercial devices by Hysitron [12] and others have promised similar functionalities. A comprehensive review of the test methods can be found in [2, 13, 14].

Bulge Tests
As another approach, bulge tests can effectively be used to test membranes with thicknesses approaching several nanometers while keeping the lateral dimensions in the order of hundreds of micrometers. This well-established method consists of measuring the out-of-plane deflection of a thin membrane under the influence of increasing pressure, applied from its backside [14]. The central deflection of the membrane is monitored continuously and the maximum stress at failure can be inferred from the corresponding pressure–deflection curves. Shrinking only the thickness (while keeping the other dimensions at μm/mm scales relaxes the experimental constraints on detection considerably. Recently, a homebuilt bulge test setup was successfully used to statistically characterize membranes that are as thin as 10 nm [8], while displacements were measured optically. Bulge tests also provide a very good alternative to study 2D materials such as graphene. However, the major disadvantage of the technique is the complicated analysis that is required to interpret the experimental data accurately.

Nanoscale Bending Tests with AFM
The third approach, which is reviewed in more detail here, is the nanoscale bending test. Bending tests offer many benefits over the above methods. Most importantly, gripping of the test samples is no longer a problem and the experimental procedures are generally simpler than in tensile tests. Since the test samples and test setup are independent from each other, there are fewer process restrictions and the effects of different parameters on reliability can be more easily studied. Finally, in a bending test, the area of fracture initiation can be effectively controlled without needing external notches. For an un-notched tensile test sample, fracture load can be reached anywhere along the beam: Sidewall surfaces and beam top or undersides have equal stresses and hence the defect distribution on all sides should be carefully considered. On the other hand, bending samples can be designed so that maximum stresses occur on the critical sites, hence, simplifying the corresponding statistical analysis. This point proves to be very useful when studying the effects of specific defects on mechanical performance [7]. Yet, the major drawback of the method is that the stress–strain distributions within the sample and its failure characteristics must be acquired indirectly, by processing the experimental load–displacement measurements.

Overview of the Method
Among all of the available methods, atomic force microscopy (AFM) is arguably the most popular

one for nanoscale bending tests, being extensively employed to characterize both mechanical device components and biological samples. AFM is a powerful analytical tool used mainly to determine the surface topography of structures with the help of a (usually) compliant cantilever, which is attached to a piezo scanner. For instance, in contact mode operation, the cantilever scans the surface of the material while it is in contact with the sample. A feedback control loop moves the piezo scanner vertically at each scan point to maintain a fixed force between the tip and the sample. This vertical piezo displacement can then be transformed into a topographic image of the surface. A nanomechanical AFM-based test consists of using an AFM cantilever, with known stiffness, to deflect a test sample until it fails. The applied force, F, and the sample deflection, δ_{sample}, are simultaneously measured during loading and the fracture strength of the tested structure is obtained from either a simple theoretical calculation or finite element analysis (FEA).

The first step in the experiment is to determine the position of the AFM cantilever tip relative to the test sample accurately. For this purpose, the sample of interest is imaged by a contact mode scan and the load application point is picked from the AFM image. Using the piezoelectric controls of the instruments, the cantilever tip is offset to the loading point with an accuracy of a few nanometers. Next, the sample is loaded. The schematic in Fig. 2 shows a cantilever loading a micron-sized test beam. When the piezo scanner is extended by δ_{piezo}, the AFM cantilever pushes against the sample, which is deflected by δ_{sample}, while the cantilever undergoes an upward displacement of $\delta_{\text{cantilever}}$. Since the total piezo extension, δ_{piezo} equals the sum of the cantilever and sample displacements, deflection of the test structure can be expressed as

$$\delta_{\text{sample}} = \delta_{\text{piezo}} - \delta_{\text{cantilever}} \qquad (2)$$

Deflection Measurement To obtain load–deflection curves for individual samples, $\delta_{\text{cantilever}}$ is continuously recorded throughout the experiment by monitoring the position of the laser light, which is reflected off the back of the AFM
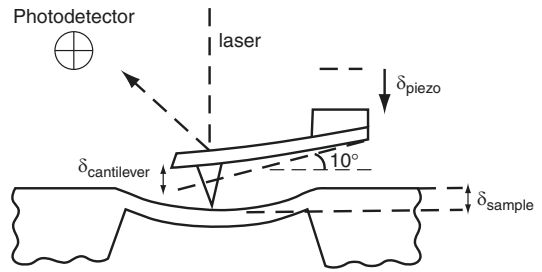
Reliability of Nanostructures, Fig. 2 Schematic of a nanoscale bending test with an AFM cantilever loading a double clamped test beam

cantilever and collected by a position-sensitive photodetector. The upward cantilever deflection results in a change in the photodetector output signal (S_p). Hence, every time a cantilever is attached to the piezo scanner, the relation between $\delta_{\text{cantilever}}$ and the S_p (in Volts) must be carefully calibrated. To relate S_p (V) to $\delta_{\text{cantilever}}$ (m), the cantilever is brought into contact with a noncompliant surface, such as a rigid Si piece. The piezo is extended by a known distance, δ_{piezo}, pushing the cantilever against the surface and the corresponding change in S_p is monitored. Throughout the calibration, since the surface remains rigid (neglecting elastic contact deformations), $\delta_{\text{cantilever}} = \delta_{\text{piezo}}$. Considering the nonlinearity often observed at higher deflections, this relation is best characterized by a third-order polynomial ($\delta_{\text{cantilever}} = c_1 S_p^3 + c_2 S_p^2 + c_3 S_p$), the constants of which are determined by a least-squares fit.

Force Measurement The force, F, applied by the cantilever on the test structure is equal in magnitude and opposite in direction to the load that causes the cantilever to deflect by $\delta_{\text{cantilever}}$. As seen in Fig. 2, the AFM cantilever makes a 10° angle with the horizontal. Hence, any load, F, applied on the test structure will have a vertical component, $Fv = F\cos(10°)$ and a much smaller, negligible, horizontal component, $Fh = F\sin(10°)$. The vertical force applied on the beam is given by

$$F_v = k_{\text{cantilever}}\delta_{\text{cantilever,}} \qquad (3)$$

where $k_{\text{cantilever}}$ is the stiffness of the AFM cantilever.

Accuracy of the force measurements strongly depends on $k_{cantilever}$. Using commercially available cantilevers may save lengthy fabrication steps; however, since the specifications are not precisely known, a careful calibration process is needed to accurately determine $k_{cantilever}$.

AFM cantilever stiffness calibration. When uncoated single-crystal Si cantilevers are used, the stiffness can be obtained from the cantilever dimensions and known elastic properties of Si. An SEM can be used to measure the lateral dimensions and the offset of the tip, which defines the effective length of the cantilever. Similarly, the tip height can be obtained using an optical surface profilometer. Measuring the cantilever thickness accurately is often the most important, yet, most difficult step, due to the inclined sidewalls resulting from the fabrication process.

An iterative method based on resonant measurements was recently proposed to accurately determine h and calculate $k_{cantilever}$ [6]. The resonant frequency of the cantilever in air, f_{air}, is measured with the AFM and the vacuum frequency, f_{vac}, is estimated by correcting for the effects of the air damping.

$$f_{vac} = \left[1 + \frac{\pi b \rho_{air}}{4h\rho_{cantilever}}\right]^{0.5} f_{air}, \qquad (4)$$

where ρ_{air} and $\rho_{cantilever}$ are the densities of air and the cantilever material respectively, b is the average width, and h is the thickness of the cantilever. (For typical cantilevers used in the experiments in [6], $f_{vac} \sim 1.002\, f_{air}$.) The vibration of the cantilever is then simulated by FEA while the thickness, h, is iterated until simulated and measured frequencies match. The stiffness, $k_{cantilever}$, is then calculated from a static FEA, which incorporates known elastic properties and the exact cantilever geometry.

Determining the Fracture Strength Finally, the experimental load–deflection curves are processed to obtain the Young's modulus, E, of the samples and the deformation of each sample is modeled with a nonlinear FEA simulation. The FEA model takes into account the maximum load

at fracture, load application point, geometric properties, and calculated E values to determine the fracture stress, *maximum tensile stress along the length of the sample*, for each sample. Results from repeat tests can then be analyzed statistically.

Accuracy of the Measurements
The precision of the strength measurements strongly depends on the accuracy of the force measurement, since the fracture load is used to extract fracture strength from FEA. An approximate expression for the error in strength measurements is given by:

$$\frac{\Delta\sigma}{\sigma} ; \frac{\Delta F_v}{F_v} ; \frac{\Delta k_{cantilever}}{k_{cantilever}} + \frac{\Delta_{cantilever}}{\delta_{cantilever}} \qquad (5)$$

For this analysis, the cantilever stiffness can be expressed as $k \approx 3EI/L^3$, where $I = bh^3/12$, with a cantilever width, b, thickness, h, and length, L. Similarly, the thickness $h \approx cf_{vac}L^2$, where c is a constant and f_{vac} is the resonant frequency in vacuum. Hence, Eq. 5 becomes:

$$\frac{\Delta\sigma}{\sigma} = \frac{\Delta b}{b} + \frac{3\Delta f_{vac}}{f_{vac}} + \frac{6\Delta L}{L} - \frac{3\Delta L}{L}$$
$$+ \frac{\Delta\delta_{cantilever}}{\delta_{cantilever}} \qquad (6)$$

The relative errors in the width and length measurements are related to the resolution of the SEM images, while $\frac{\Delta f_{vac}}{f_{vac}}$ depends on the sensitivity of the frequency measurement. All of these are much smaller than 1 %. On the other hand, the precision of the AFM cantilever displacement is directly related to the precision of the AFM piezo scanner and the photodetector readout and will dominate the measurement error, depending on the setup being used.

Future Directions in Research

AFM has great potential to become the standard testing method for a large class of materials. No doubt, the biggest advantage of the method is the use of commonly available standard instruments,

R

hence allowing similar structures to be tested at different locations without needing specialized test apparatus. Being highly flexible, it can effectively characterize the effects of process defects and environmental degradation on mechanical reliability, hence addressing the questions postulated in the "Key Research Findings" section. Moreover, there are no material restrictions, and, upon correct choice of cantilevers, samples ranging from hundreds of nanometers to a few micrometers can repetitively be tested with very good-to-excellent accuracy. However, further studies are still needed to better understand the interaction between the tip and the sample during loading, determine the optimum tip and cantilever geometry to test different samples, and decide upon standard calibration procedures. One of the current challenges is also to develop new generation of AFM cantilevers in order to adapt the existing methodology to perform tensile tests at the same resolutions.

Cross-References

▶ Ab Initio DFT Simulations of Nanostructures
▶ AFM
▶ AFM Probes
▶ Atomic Force Microscopy
▶ Finite Element Methods for Computational Nano-optics
▶ Mechanical Properties of Hierarchical Protein Materials
▶ Nanomechanical Properties of Nanostructures

References

1. Senturia, S.D.: Microsystem Design. Springer, New York (2004)
2. Hemker, K.J., Sharpe Jr., W.N.: Microscale characterization of mechanical properties. Annu. Rev. Mater. Res. **37**, 93–126 (2007)
3. Anderson, T.J.: Fracture Mechanics: Fundamentals and Applications, 3rd edn. CRC Press, Boca Raton (2004)
4. Roundy, M., Cohen, M.L.: Ideal strength of diamond, Si, and Ge. Phys. Rev. B **64**, 212103 (2001)
5. McCarty, A., Chasiotis, I.: Description of brittle failure of non-uniform MEMS geometries. Thin Solid Films **515**(6), 3267–3276 (2006)
6. Alan, T., Hines, M.A., Zehnder, A.T.: Effect of surface morphology on the fracture strength of silicon nanobeams. Appl. Phys. Lett. **89**, 091901 (2006)
7. Alan, T., Zehnder, A.T.: A Monte-Carlo simulation of the effect of surface morphology on the fracture of nanobeams. Int. J. Fract. **148**, 2 (2007)
8. Alan, T., Yokosawa, T., Gaspar, J., Pandraud, G., Paul, O., Creemer, F., Sarro, P.M., Zandbergen, H.W.: Microfabricated channel with ultra-thin yet ultra-strong windows enables electron microscopy under 4-bar pressure. Appl. Phys. Lett. **100**, 081903 (2012)
9. Alan, T., Zehnder, A.T., Sengupta, D., Hines, M.A.: Methyl monolayers improve the fracture strength and durability of silicon nanobeams. Appl. Phys. Lett. **89**, 231905 (2006)
10. Gaspar, J., Schmidt, M.E., Held, J., Paul, O.: Wafer-scale microtensile testing of thin films. J. Microelectromech. Syst. **18**(5), 1062–1076 (2009)
11. Haque, M.A., Espinosa, H.D., Lee, H.J.: MEMS for in situ testing-handling, actuation, loading, and displacement measurements. MRS Bull. **35**(5), 375–381 (2010)
12. Kiener, D.A., Minor, A.M.: Source truncation and exhaustion: insights from quantitative insitu TEM tensile testing. Nano Lett. **11**(9), 3816–3820 (2011)
13. Haque, M.A., Saif, T.: Mechanical testing at the micro/nanoscale. In: Sharpe, W.N. (ed.) Springer Handbook of Experimental Solid Mechanics. Springer, New York (2008)
14. Tabata, O., Tsuchiya, T.: Reliability of MEMS: Testing of Materials and Devices. Wiley-VCH, Weinheim (2007)
15. Vlassak, J.J., Nix, W.D.: A new bulge test technique for the determination of Young's Modulus and Poisson's ratio of thin films. J. Mater. Res. **7**(12), 3242–3249 (1992)

Remotely Powered Propulsion of Helical Nanobelts

Gilgueng Hwang[1] and Stéphane Régnier[2]
[1]Laboratoire de Photonique et de Nanostructures (LPN-CNRS), Site Alcatel de Marcoussis, Paris, France
[2]Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, Paris, France

Synonyms

Wireless propulsion of helical nanobelt swimmers

Definition

Remotely powered propulsion of helical nanobelts is the swimming propulsion of nanoswimmers created by helical nanobelts.

Overview

Since swimming mobile micro-agents at low Reynolds number experience severe viscous drag, their swimming performances are much limited. Considering the most of in vivo environments are dynamic system, highly dynamic microswimmers would have a big technological impact in many fields. Engineering artificial flagella is inspired from nature's bacteria having their excellent motility. They are created from self-scrolled helical nanobelts. These swimmers are designed with a head and a tail, similar to nature's micro-organisms such as bacteria and their flagella. In this entry, the highly dynamic propulsion of biologically inspired artificial flagella which are remotely powered by electroosmotic force is introduced. They can swim as fast as 24 times of their body lengths per second and apply pressure higher than 300 Pa from their high energy conversion efficiency. Their swimming performance outran the other inorganic microscale swimmers and even nature's bacteria like *Escherichia coli*. These electroosmotic propelled helical nanobelt swimmers could be the potential biomedical carriers, wireless manipulators, and as local probes for rheological measurements.

Introduction

Rapid advancement in semiconductor electronics and microelectromechanical systems (MEMS) has accelerated the size reduction of functional systems like sensors or actuators. Considering much increased needs in less invasive biomedical or clinical tools to prolong human life, micro- or nanoscale systems have many promising features. One of the most important features is to reduce their size and disconnect them from wires. In this context, recent interest toward developing untethered biomedical micro- and nanorobots could be an alternative way to achieve less invasive and targeted medical therapy. Although they seem to increase the complexity of conventional medications, they are still promising due to their controllability, local therapy, and repeatability. Recent developments of in vivo biomedical nanorobots draw much interest in discovering new and efficient wireless power transfer methods and in controlled locomotion mechanisms. This kind of novel tools would have many potential applications, such as targeting, diagnosing, and treating blood clots or cancer cells, for neural cell probing and so on.

Although they are promising to such applications, they are still remained in laboratory experiments due to several reasons. The micro- or nanoscale mobile agents experience much more viscous drag due to the decreased Reynolds number which results in severely limited swimming performance. Considering the capillary of human's circulatory system has the average flow speed which can go up to 1 mm/s, mobile microagents should swim faster than this flow speed for their survival and task continuity. It should be noted that nature's bacteria have well adapted in such a harsh environment.

Many approaches aim to mimic natural bacteria for the locomotion of artificial nanostructures because of their excellent motility, around 5–10 times of body sizes per second [1]. The Nature has adopted several different propulsion techniques. The swimming of real bacteria is mainly divided into corkscrew-type rotating propulsion and the oscillation of flagella tail [2]. Several studies exploring similar propulsion techniques exist in the literature, with different efficiencies. Purcell uses a definition of swimming efficiency which compares the power used to propel a body at a given velocity to the power required to simply pull the body through the fluid at the same velocity (ε = power required to simply pull body/power consumed during propulsion) [2]. Under this definition, direct pulling with field gradient is 100 % efficient (i.e., $\varepsilon = 1$), and other methods would always perform less. Recently, direct pulling of nanowires, nanotubes and microtubes with excellent swimming performances (2 mm/s^{-1} which is

approximately 50–100 times per body size) was demonstrated by catalytic decomposition of hydrogen peroxide (H_2O_2). However, this technique only works in a solution like H_2O_2, allowing a catalytic reaction with Pt or Ag [3].

Among different propulsion methods, the use of magnetic effect is the most widely encountered. A rotating magnetic field was used to drive the macroscale flagella-like structure [4]. Microbeads were attached to natural bacteria and moved by flagella propulsion [5], 30-µm-scale bacterial flagella–like propulsion was demonstrated by an attached ferromagnetic metal pad under an external rotating magnetic field [6]. A linear chain of colloidal magnetic particles linked by DNA and attached to a red blood cell can act as a flexible artificial flagellum [7]. Further size reduction and applying same principle to drive artificial bacteria was achieved [8]. Lately, helical propellers are shown as an efficient solution at low Reynolds numbers [9].

Concerning the driving power source, there are several limitations of using an external field. The magnetic field gradient rapidly decreases with the distance from the source. Therefore, although the helical morphology is advantageous at low Reynolds numbers by reducing the viscous drag, the swimming performance of artificial structures is still much lower than the natural bacteria. To improve the swimming performance, increasing field intensity would be necessary.

However, the major challenge toward potential in vivo biomedical applications is the lack of closed loop motion control either by teleoperation or autonomous navigation. This lack is related to position-tracking issues. Widely used noninvasive biomedical sensors or imaging devices include radiography, computed tomography (CT), ultrasound echography, and magnetic resonance imaging (MRI). Considering the real-time navigation and high spatial resolution requirements, MRI is one of the most promising solutions. Ferromagnetic objects, or even magnetotactic bacteria, were tracked and navigated under real-time MRI [10]. In case of smaller objects at low Reynolds numbers, an additional propulsion-dedicated field gradient generator would probably be required. However, magnetic field gradient–based propulsion would be largely limited when it has to be used with an observation device such as MRI because of the conflict between the imaging and propelling magnetic fields. This problem can be avoided by using nonmagnetic propulsion.

As nonmagnetic approaches, the flapping or undulating motion of flagella was achieved in mesoscale using a commercial piezoelectric bimorph [11]. Microrobots that harness natural bacteria have also been demonstrated [5]. There are very few works on electric field–based wireless micro/nanorobots. For example, diodes were actuated using electroosmotic pumping [12]. It showed that electroosmotic field can be a good candidate for actuating micro/nano robots. However, the work was only demonstrated in macro scale and not scaled down to micro or nano.

In this entry, remote propulsion of artificial micro swimmers with high dynamic mobility (5–30 body lengths per second) at low Reynolds numbers by electroosmotic effect is described. These swimmers are based on helical nanobelt (HNB) structures. As different types of HNBs showed different swimming performances, choices of surface material and geometry is discussed along with the fabrication processes. The electroosmotic propulsion principle and experimental setups are presented. Finally, the experimental results on electroosmotic propulsion of HNBs in comparison with different locomotion techniques are described.

Self-Scrolled Helical Nanobelts as Artificial Flagella

Advantages of HNBs as Artificial Flagella

There are several different types of bacterial locomotion, such as rotating and undulating motions. Their major features, especially in terms of surface condition, geometry, and motility have evolved to swim efficiently at low Reynolds numbers. For example, *E. coli* bacteria consist of a highly elastic body formed of protein. Therefore, they can self-transform their morphology to increase their swimming velocity in a viscous liquid environment. It is considered that such

ability of natural bacteria is mainly attributed to the nonlinear mechanics of their flagella. To mimic similar performances of natural bacteria, the ultraflexible nanostructures with large nonconstant mechanics are considered to be necessary. The mechanical durability is also an important issue especially in harsh environments.

HNBs are one of the most flexible nanostructures that can be fabricated using standard microfabrication techniques with controlled geometry. Helical morphology appears as an adapted solution for electroosmotic propulsion because of the large surface-to-volume ratio to maximize the external energy reception. Another advantage is their elasticity to passively self-adapt their morphology in thick liquid environments by reducing viscous drag. Different types of nanobelts were synthesized and mechanically characterized [13]. Recently, HNBs revealed very large nonlinear mechanical properties along with giant piezoelectricity [14]. In case of electroosmotic propulsion, different surface coatings with thin films can also be achieved by standard microfabrication techniques to improve electrokinetic energy. The high surface-to-volume ratio of HNBs can disadvantageously increase friction at the liquid/solid interface. HNBs with hydrophobic surface coatings can reduce this friction to improve the swimming performance.

Additionally, harvesting energy from the environment would be an important advantage for mobile micro/nanoagents. Piezoelectric nanogenerators based on zinc oxide nanowire arrays were firstly proposed [15]. Nanowire-based, self-powered nanosensor has also been demonstrated [16]. Passive motions of HNBs during their swimming at low Reynolds number can self-harvest electrokinetic energy due to their giant piezoelectricity [14].

Therefore, considering the limitations of conventional propulsion mechanisms, remotely powered gradient pulling of HNBs should be efficient to swim in low Reynolds number liquid [9].

Fabrication Process

Four different types of HNB-based nanoswimmers, varying the materials and geometries, summarized in Table 1 are produced for testing purposes. Their propulsion efficiency is then compared in electro-osmotic propulsion experiments.

Figure 1 illustrates the fabrication process. The initial layers were grown on semiinsulating GaAs (AXT wafers) using a molecular beam epitaxy system (VEECO, Gen II MBE) equipped with a valved cracker for As and solid sources for Ga and In (Fig. 1a). For n-type doping, a Si source was used. Substrate temperature was measured by a pyrometer. After a thin GaAs buffer layer, the 500-nm thick sacrificial AlGaAs layer is deposited. The layer contains 20 % Ga in order to prevent oxidation. The sacrificial layer in previous designs was made of AlAs. The oxidation of these layers after exposure damaged the InGaAs/GaAs bilayer. On top of the sacrificial layer the InGaAs/GaAs layer is deposited which later self-forms into a nanospring. An In content of 15 % in the InGaAs layer was determined by X-ray diffraction (XRD) measurements. The thickness of this layer must be smaller than the critical thickness to maintain elastic strain. The layer properties along with other specifications of the structures are summarized in Table 1.

During the deposition of the InGaAs/GaAs bilayer on this wafer for HNB 1, it was attempted to get a slightly lower doping concentration than on the wafer that was used for HNB 2, 3, and 4. From the results, it seems that the doping concentration was too high which resulted in a partial self-compensation and, therefore, in a decrease in the effective doping concentration. The doping concentrations of the structures are 4.4×10^{18} and 8.5×10^{18} cm^{-3}. The initial pattern can be created through photolithography. Reversible photoresist AZ5214 was used as a resist. After the development of the resist, reactive ion etching (RIE) with a SiCl$_4$ gas was used to transfer the pattern to the InGaAs/GaAs bilayers. For HNB4, the thickness of photoresist was reduced around 50 nm with O$_2$ plasmas. For metallic HNB3, the Cr/Ni layers on the surface and on the heads to which the structures are fixed at the end were created before RIE through a lift-off process with negative photoresist AZ5214.

The Cr layer is 10 nm thick and serves as an adhesion layer. The 10-nm thick Ni layer is used for metal HNB3. Finally, a 2 % HF aqueous

R

Remotely Powered Propulsion of Helical Nanobelts, Table 1 Specifications of different types of artificial bacteria (HNB)

	HNB 1 low	HNB 2 high	HNB 3 Cr/Ni	HNB 4 photoresist
Thickness $In_{0.15}Ga_{0.85}As/GaAs$ (nm)	11.6/15.6	8~10/15.6	11.6/15.6	
Thickness Cr/Ni or photoresist (nm)	NA		Cr/Ni 10/10	Photoresist ~50
Diameter (tail = head), pitch and width (tail) (μm)	2.1, 14 and 2.5			
Number of turns (tail)	4.5 turns			
Length (head/total) (μm)	12/74			
InGaAs/GaAs doping (ND)	4.4×10^{18} cm^{-3}	8.5×10^{18} cm^{-3}		

Remotely Powered Propulsion of Helical Nanobelts, Fig. 1 Basic fabrication process sequence of an HNB. (**a–c**) Initial planar bilayer, patterned through conventional microfabrication techniques, folds itself into a 3-D nanostructure during wet etch release. (**d**) Produced HNB with a tail and a head (SEM photo)
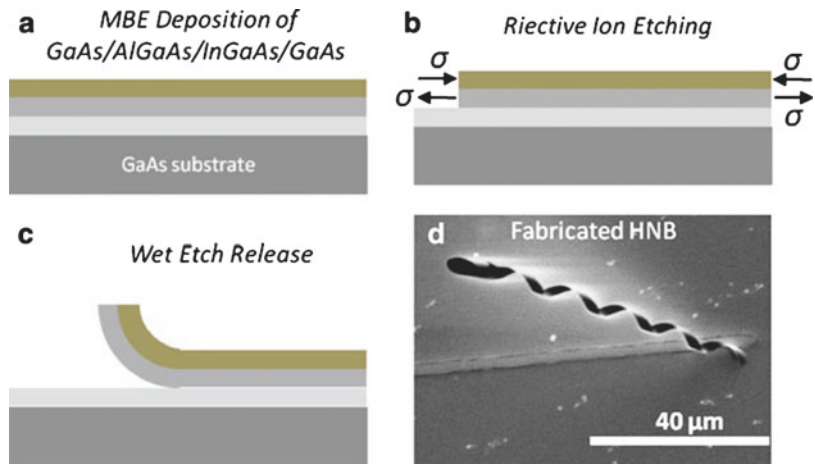
solution at 4 °C was used to selectively etch the AlGaAs sacrificial layer under the InGaAs/GaAs heterostructures for the self-forming of the nanostructures (Fig. 1a). During this wet etch, the patterned bilayer curled up along a <100> direction, releasing the internal strain and forming 3-D structures. The direction of the scrolling is determined by the anisotropy in stiffness of the InGaAs/GaAs bilayer. After the wet etch release, the chips were rinsed in deionized water and subsequently in isopropyl alcohol [14]. Samples are also conserved in isopropyl alcohol in order to prevent oxidation by water of heads to which the structures are fixed.

Electromechanical Property

To understand the electrokinetic function of the applied strain by liquid environment, the piezoresistivity of HNBs is characterized. For the intrinsic property characterizations, the HNBs with metal pads attached on both sides were fabricated using microfabrication techniques [14]. With the metal connectors, good electrical contact can be achieved. Besides the electromechanical characterization, such connectors also allow for the integration of these structures into more complex assemblies. Nanomanipulation inside an SEM was used for their electromechanical property characterization. The experimental results showed that the structures showed unusually high piezoresistive response. Moreover, electrostatic actuation was used to excite the structures at their resonance frequency and investigate their resistance to fatigue. With their low stiffness, high strain capability, and good fatigue resistance, the HNBs can be used as high-resolution and large-range force sensors. By variation of design parameters, such as the number of turns, thickness, diameter, or pitch, an HNB with the required stiffness can be designed through simulation.

For the quantitative comparison with other piezoresistors, the piezoresistance coefficient can be obtained as,

$$\pi_l^\sigma = \frac{1}{X} \frac{\Delta\sigma}{\sigma_0} \qquad (1)$$

where σ_0 is the conductivity under zero stress and X is the stress. The piezoresistance coefficient was defined with resistivity as $\pi_l^\sigma = \Delta\rho/\rho_0$. The conversion is $\pi_l^\sigma = -\pi_l^\rho$ for small σ. Uniaxial stresses were applied on HNB along their lengths. Since the linearity of each cycle from the measured data is almost constant, the widest measured region of the cycles was chosen to calculate the piezoresistance coefficient. Geometry information from Table 1 was used as the parameters.

These obtained piezoresistance coefficients of HNBs based on Eq. 1 are compared with other elements of the piezoresistors. Each reported piezoresistance coefficients of Bulk Si, boron-doped Si, and Si-NW, and CNT were summarized in Fig. 2 in comparison with HNB [14]. Especially, Si-NW was recently reported as showing giant piezoresistive effect. The revealed piezoresistance coefficients of the HNBs are even much higher than the one of Si-NW. These results are straightforward from the benefit of HNB's flexibility considering the Eq. 1 since the tiny input stress (X) can cause the resistance change ($\Delta\sigma/\sigma_0$). At least the torsional effect seems to affect much to the high piezoresistive response of the HNBs. As was discussed on the difficulty in exact modeling of the piezoresistive HNBs, mechanism behind such a high piezoresistive response was not understood yet. However, there is the fact that the HNBs are the promising nanostructures to be used as the elements of ultraflexible electromechanical devices such as force sensors and bioinspired systems.

Mechanical Property

In addition to piezoresistivity of HNBs, they can perform like a spring to save energy from large-range mechanical force based on their nonconstant mechanics. Tuning fork force calibration tool was utilized to characterize the mechanics of HNBs in full-range tensile elongations. The tuning fork as force interaction tools have been proposed mainly to replace the AFM optical cantilevers for imaging [17]. The mechanical property (spring constant) of HNBs was measured by attaching them to the tuning fork tip. The force is obtained from the stiffness of HNBs measured by tuning fork gradient force sensor and by the displacement measurement of HNBs with SEM imaging. Figure 3 summarizes the measured mechanical property while HNBs are under tensile elongation. The nonconstant stiffness behavior of HNBs during their controlled tensile elongation was clearly revealed in their full-range elongation. The obtained stiffness ranges from 0.009 to 0.297 N/m with a resolution of 0.0031 N/m during full elongation and 0.011 to 0.378 N/m with a resolution of 0.0006 N/m for the nondestructive method. It was transformed into full elongation tensile forces as high as 2.95 μN for the first experiment and 1.56 μN for the second.
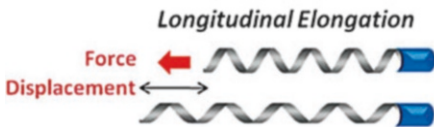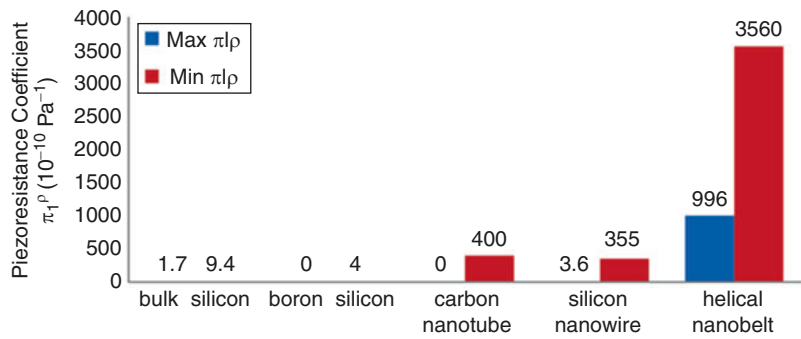
Wireless Propulsion of Helical Nanobelts

Principle of HNB Electroosmotic Propulsion

The propulsion of HNBs involves two conversions: transforming energy from an external source to a force and transforming this force into a motion. The proposed method to supply power remotely to HNBs is an electrokinetic process. Its main advantages for in vivo applications are the good tolerance of living organisms to electric fields and its compatibility with medical MRI imaging.

In general, the surface charge that develops at the solid–liquid interfaces plays an important role in a number of electrokinetic processes. Due to the presence of a thin interfacial layer named the Stern layer, it results in a nonzero electric potential at the liquid interface, also known as the zeta potential ζ. This potential is screened by mobile counterions diffusing in the liquid in a layer generally named Electrical Diffuse Layer (EDL), which has a typical thickness of a few tens of nanometers depending on the solution ionic strength. Under an electric potential difference, the resulting electric field puts in motion the excess counterions in the EDL. This moving layer drags the whole

Remotely Powered Propulsion of Helical Nanobelts, Fig. 2 Comparison of longitudinal piezoresistance coefficients in different nanostructures (bulk silicon, boron silicon, carbon nanotube, silicon nanowire, and helical nanobelt) [14]

Remotely Powered Propulsion of Helical Nanobelts, Fig. 3 Mechanical property of helical nanobelt with nonconstant stiffness for large-range force sensing

liquid inside the fluidic medium, resulting in "pluglike" flows also referred to as electroosmotic flows (EOF). In addition, one should also consider the surface charge developing at the HNB surface. In the case of a HNB floating in liquid, its own Stern layer generates a flow which applies a hydrodynamic pressure on the surface of the robot, propelling it in the opposite direction.

Figure 4 depicts the electroosmotic propulsion mechanism through the interface between HNB's thin membrane surface and liquid medium solution. Since the propulsion force occurs through the whole surface, the large surface-to-volume ratio of HNBs is very advantageous.

HNBs' motility results from a local electroosmotic flux powered by an external field. The specific direction of HNB's propulsion along the cathode and anode probes indicates that a DC field gradient along the HNB is responsible for this propulsion. The equivalent circuit can predict that the electroosmotic propulsion force is affected by the difference of electrical resistivity between medium solution and the HNB (Fig. 4). The electric voltage applied to the HNB by the external field can be estimated from a serially connected resistors model describing the ionic conductance through the liquid medium and the

capacitors for the ionic layers. In a DC electric field, the resistance is the leading contribution. The resulting DC voltage of magnitude V_d induced in the HNB is modeled as

$$V_d = \frac{R_2}{R_1 + R_2 + R_3} V_{ext} \tag{2}$$

where V_{ext} is the DC input voltage applied to the probes in the petri dish, and R_1 (resistance between left electrode and HNB), R_2 (resistance of HNB), and R_3 (resistance between HNB and right electrode) are depicted in Fig. 4. Assuming that the resistance of the liquid is linearly proportional to the distance between the probes, Eq. 2 can be simplified as $V_d = E_{ext} l_d$, where l_d is the length of the HNB body and E_{ext} is the external DC field. The external DC field between the electrodes makes electroosmotic fluid flow along the HNB body.

Considering a liquid electrolyte, consisting of positive and negative particles in liquid solution, the interface between the electrolyte and the container wall generally forms a double layer in equilibrium, where a nonzero surface charge is screened by a very thin diffuse layer of excess ionic charge of width λ the Debye screening
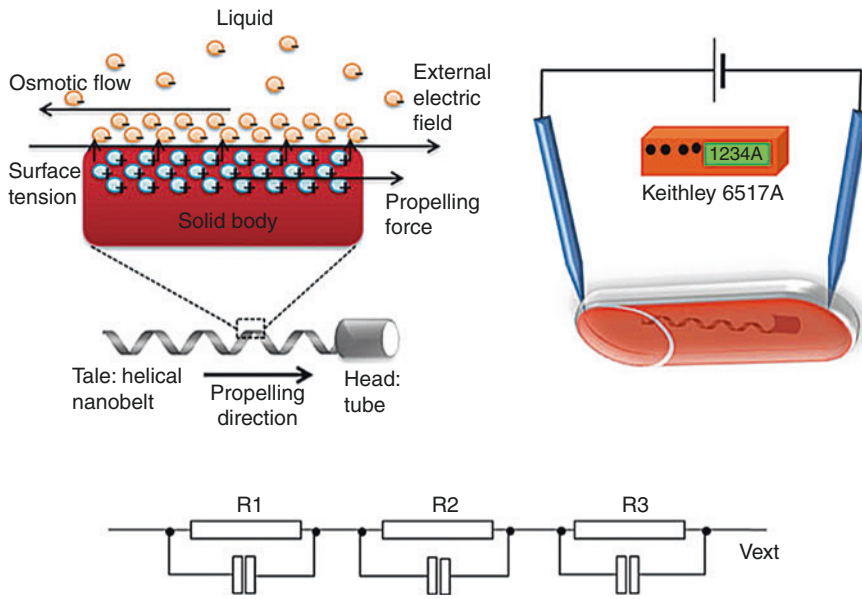
Remotely Powered Propulsion of Helical Nanobelts, Fig. 4 Schematic diagrams of the experiments for HNB swimming. A DC bias was applied between two electrodes located through axial direction of HNB body. Thus, the propulsion direction can be controlled and predicted by the configuration of electrodes. For controlling the direction of electric field, micromanipulators with tungsten probes were utilized to form electrodes

length (typically 1–100 nm). The double layer is effectively a capacitor skin at the interface, which has a small voltage ζ across called the zeta potential. The effect on the HNB and also the environment under electric field is further detailed in the following model which decomposes the part of intrinsic electrophoretic force directly generated to HNB and the electroosmotic flow from the surface interaction. Considering a tangential electric field E_{ii} applied in parallel to a flat surface, the electric field acts on ions in the diffuse part of the double layer, which drags the fluid to produce an effective slip velocity outside the double layer by the Helmholtz-Smoluchowski formula (Eq. 3) [18],

$$u_{11} = \left(\frac{2\varepsilon_0 \varepsilon_r \zeta_{HNB} H}{3\eta} - \frac{\varepsilon_0 \varepsilon_r \zeta}{\eta} \right) \vec{E} \qquad (3)$$

where $\varepsilon_0 \varepsilon_r$ and η are the electrolyte permittivity and fluid viscosity, which are constants (environment parameters). Increasing the speed of the HNB requires increasing ζ_{HNB} and H which are the zeta potential around the surface of the HNB and the Henri function, which depends on the geometry

and hydrodynamic properties and the external field (E). On the negative term of the equation, the Zeta potential ζ on the borders of the surface is inversely proportional to the resulting speed, thus it should be minimized and controlled by applying experiments under closed microfluidic channel.

An important application of electroosmotic flow is capillary electroosmosis, where an electric field is applied down a capillary tube to generate a uniform plug flow (Eq. 3), driven by the slip at the surface. Typical flow speeds of 100 μm/s are produced by a field of 10 V/mm. This value will be used to compare the swimming performance of HNBs. It should be noted here that the electroosmotic propulsion does not depend on the size of the microagents. Thus, microscopic HNB could move theoretically about as fast as a macrosized one.

Electroosmotic Propulsion Experimental Setup

HNBs are detached from the substrate by micromanipulation. At the start of the experiments, the fixed end of produced HNBs are cut and released

from chips using a setup with two probes with high positioning accuracy under an optical microscope. Same probes are also used to generate the electric field through the liquid medium and HNBs. Total experimental setup is shown in Fig. 5. Two nanorobotic manipulators (Kleindiek, MM3A) are installed under optical microscope; each has three degrees of freedom and, respectively, 5, 3.5, and 0.25 nm resolutions in X, Y, and Z directions. For the application of the electrical field and high resistance measurement, an electrometer with a DC power supply (Keithley 6517A) is used. Two probes attached to each side of manipulators are positioned in less than 1 mm distance. To avoid the optical reflection from the meniscus between probes and liquid medium interface, longer probes (~5 mm) with sharp tips (diameter around 100 nm) are used.

Controlled Swimming Experiments

Figure 6 demonstrates the forward and rotating propulsion of HNBs. Experiments are depicted in Fig. 6b, e. The swimming propulsions o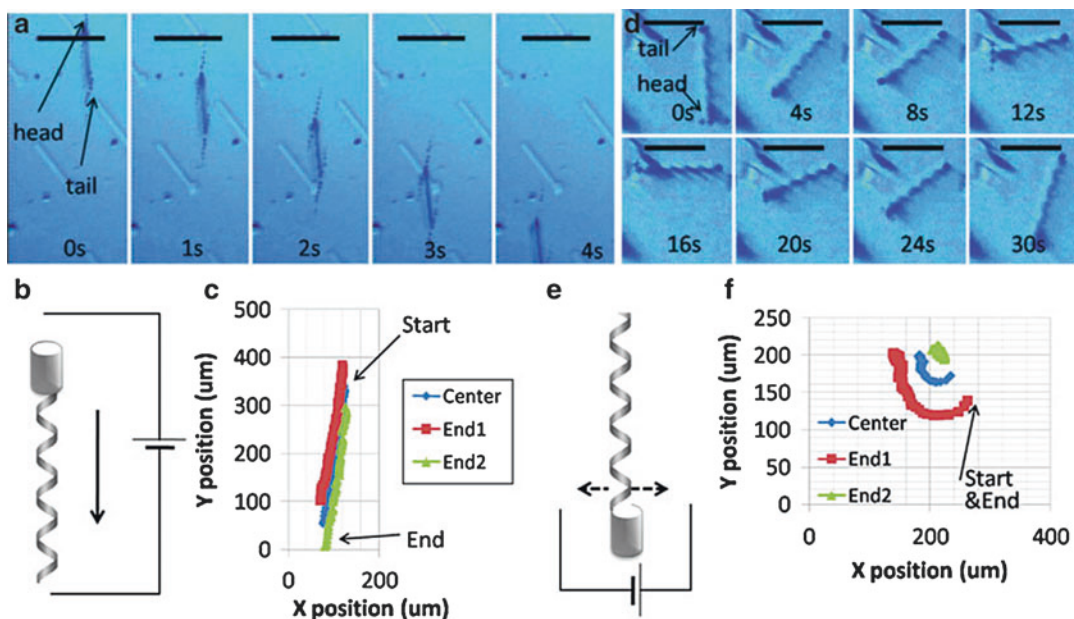f HNBs are achieved by modifying the electrodes' configuration. A series of snap shots during forward (Fig. 6a) and rotating (Fig. 6d) propulsions were taken during the propulsion experiments. The trajectories of each swimming experiment are displayed in Fig. 6c, f. For trajectories, video analysis software and a custom-made C++ program are utilized. Rotation of HNBs was achieved by their aligning to the applied field direction. The probes attached to manipulators can change the configuration to generate the desired electric field gradient. Note that in this case the rotation was achieved using an HNB fixed in one extremity, while the other is free. The electrode configuration was fixed to make the HNB rotate. Furthermore, the forward and backward swimming direction can be converted without the complicated tumbling motion of *E. coli* bacteria.

It is confirmed that the swimming directions of HNBs could be easily controlled by differentiating the field gradient. For the steering control, the field gradient needs to be aligned to the rotational axes of the HNBs. In case of an untethered HNB, the rotation is achieved by controlling the field gradient. The high surface-to-volume ratio of
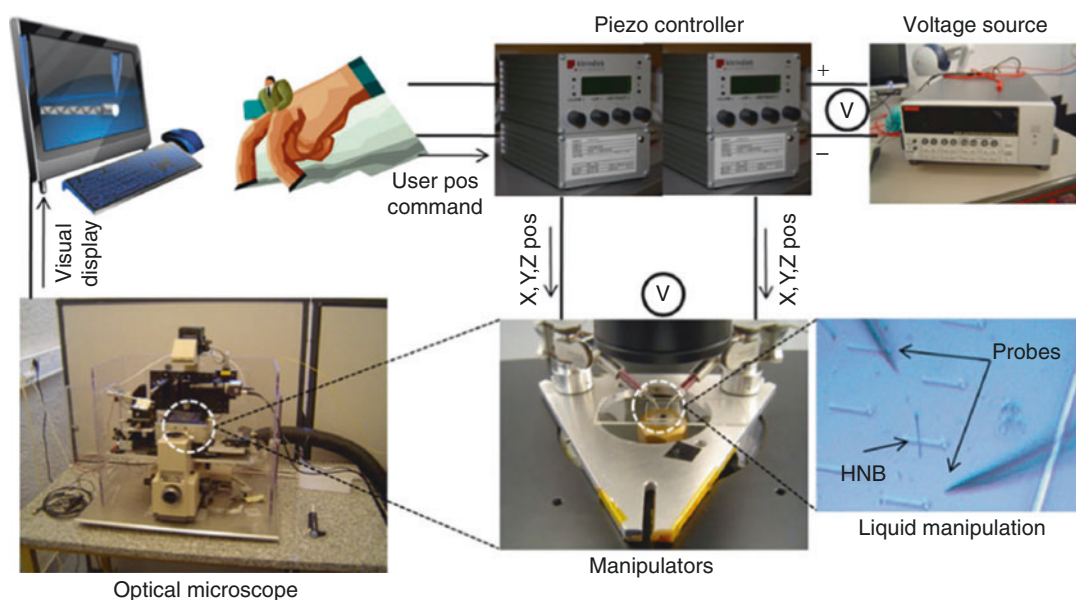
Remotely Powered Propulsion of Helical Nanobelts, Fig. 5 System setup for electroosmotic propulsion experiments: optical microscope with a camera is used to give user visual feedback and thus proper manipulation in liquid. User operates the piezo controller to manipulate tungsten probes, which are used to manipulate HNBs and to apply the electric field generated by an external voltage source

Remotely Powered Propulsion of Helical Nanobelts, Fig. 6 Different swimming propulsions of HNBs are achieved by modifying electrode configuration. Series of snap shots during backward (**a**) and rotating, (**d**) propulsions are taken during the experiments (**b** and **e**). The trajectories of each swimming experiment are displayed in (**c**) and (**f**). The scale bars in (**a**) and (**d**) are, respectively, 80 and 60 μm

HNBs increases the swimming efficiency by increasing the charged area in a given volume and mass.

The maneuverability of the currently demonstrated electroosmotic HNBs is still limited to the basic motions compared to the magnetic field based approaches. However, this issue can be addressed using an array of embedded electrodes.

Comparison of Swimming Performance

Electrokinetic Effect

To further understand the repeatable swimming behaviors due to an electroosmotic force and to find the optimized design parameters, especially the surface profile, the swimming performances of four different types of HNB were characterized. High- and low-doping HNB, a metallic (Cr/Ni) one and a dielectric counterpart were analyzed. Figure 7 shows the average and maximum propulsion velocity and acceleration normalized by the applied electric field. HNB performance was investigated in several steps (6–11 steps) ranging from 24 to 401 V/m. It was difficult to compare four HNBs in identical single field intensity because the experimental configuration of probes and HNB is difficult to maintain between each experiment. Therefore, a statistical average velocity was chosen to describe the general behavior of different HNB. It should also be noted that experiments on different types of HNBs were conducted in different days which might slightly change electrical resistivity of the liquid medium. Even though the exact resistance values of the utilized HNBs were not verified because of the difficulty in collecting them after experiments, the intrinsic resistance of HNB can be predicted as $R_{HNB4} > R_{HNB1} > R_{HNB2} > R_{HNB3}$. Hence, the average and maximum velocities and accelerations depend on the resistance of HNB as was expected from the Eq. 3. Highly doped HNB2 showed better swimming performance than low-doped HNB1. This can be explained by the fact that the InGaAs layer thickness (estimated as 8–10 nm) of HNB2 is slightly thinner than the one of HNB1 (11 nm). It results in better performance in HNB2 than HNB1 from the reduced gravity
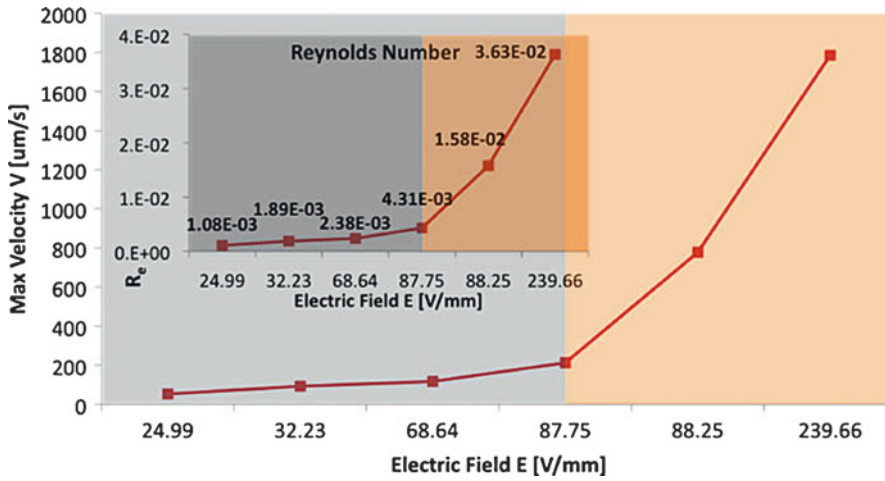
Remotely Powered Propulsion of Helical Nanobelts, Fig. 7 Dependence of HNB velocity on applied electric field intensity. The plot was obtained during the swimming experiments of a 74-μm-long HNB in isopropyl alcohol medium. The nonlinearity of the swimming performance is demonstrated at the region with higher electric field intensity than 87.75 V/mm. The *inset* figure shows the Reynolds numbers in different velocities versus applied electric field

effect and hydrodynamic friction, regardless of the higher conductivity. It reveals that the doping concentration does not play much role in the swimming performance compared to the geometry or volume effect. It should also be noted that the HNB3 and HNB4 are slightly thicker than HNB1 and HNB2 because of additional metallic or dielectric layers (Cr/Ni 10/10 nm, AZ5214 photoresist around 50 nm Table 1).

However, the HNB3 had lower performance, while HNB4 was better than the others. In the case of HNB3, this can be explained by the increase in hydrodynamic friction during swimming which slow down HNB. Furthermore, in case of metallic HNB3, the gravitational force can also increase the surface interaction force by dragging it down to the substrate. The detailed calculation and comparison of buoyancy force and gravitational force are not described here; however, the resulting nonfluidic forces (as the difference between the gravitational force and the buoyancy force) on HNB3 is 0.94 pN, which is two times higher than on HNB1 (0.43 pN). These negative effects on HNB3 reduce its swimming performance compared to the others (Fig. 8). The result of HNB4 is contrary to the HNB3 while additional layer was deposited onto the surface. It can be explained by the fact that the nonfluidic force of HNB4 was

decreased by the increased buoyancy force from the photoresist, which has much lower density than both gold and nickel. The result clearly shows the high dependency on the surface charge and thus zeta potential in dielectric surface coating to HNB compared to conductive or semiconductor materials (Fig. 8).

Swimming Performance Comparison with Other Swimmers

The force generated by the HNB's swimming is estimated from the drag it overcomes, considering its shape approximately as an ellipsoid [8].

$$F = \frac{4\eta\pi a v}{\left[\ln\left(\frac{2a}{b}\right) - \frac{1}{2}\right]} \qquad (4)$$

where a and b are the dimensions along the longitudinal and lateral axes respectively, η is the viscosity of the liquid medium; and v is speed of the HNB. For a 74-μm-long HNB with a diameter of 2.1 μm moving at 1.8 mm/s in isopropyl alcohol medium, the generated force of 1.3 nN (Eq. 4) is an order of magnitude higher than the conventional optical trap force which is less than 200 pN. Furthermore, HNBs can apply a large pressure
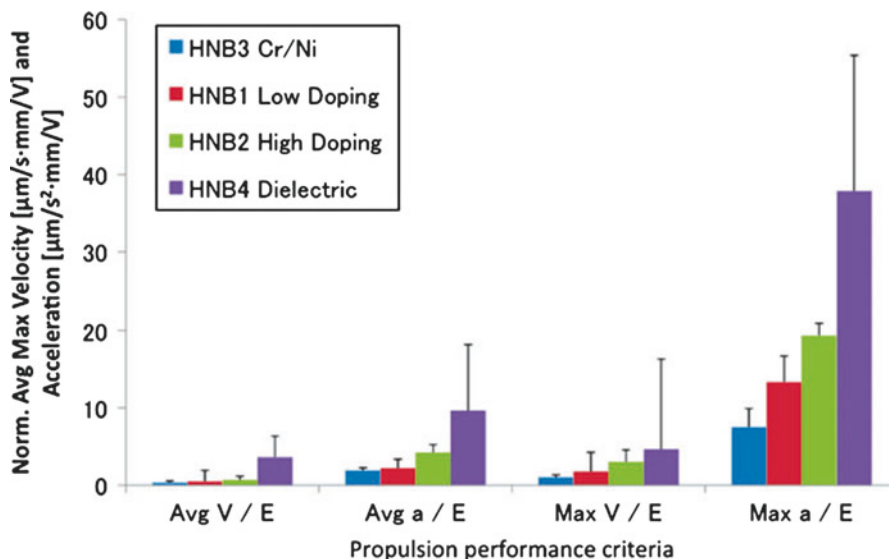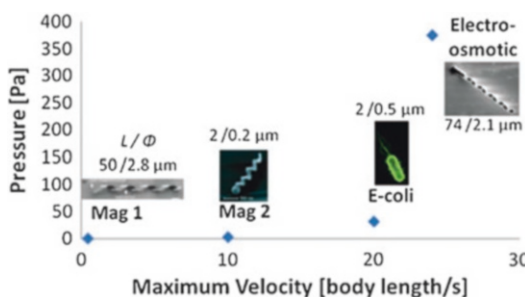
Remotely Powered Propulsion of Helical Nanobelts, Fig. 8 Dependence of normalized velocity and acceleration of propulsion on four types of HNBs. High- and low-doped semiconductor HNB showed higher swimming performance than metal (Cr/Ni)-coated HNB, and best performances are achieved with type 4 (dielectric)

up to 375.5 Pa, which can be advantageous for local manipulation, such as on biological membranes.

Compared to two rotating magnetic propulsions of helical nanostructures, the demonstrated electroosmotic propulsion of HNBs showed much higher swimming performance in terms of maximum swimming velocity and manipulation force. The demonstrated velocity (1,785 μm/s) can reach 90 times faster and the manipulation force and pressure reaching at around 1.3 nN and 375.5 Pa which are much higher than others respectively (Fig. 9).

Remotely Powered Propulsion of Helical Nanobelts, Fig. 9 Comparison of swimming performances of swimmers: Magnetic 1 [6], Magnetic 2 [8], *E. coli* [1, 19] and Electroosmotic HNB

Conclusion

This entry introduced biologically inspired microswimmers based on helical nanobelts. Helical nanobelts microswimmers have excellent electromechanical and mechanical properties which can mimic their original model in nature. The demonstrated direct pulling of HNBs by electro-osmotic pumping was proven to be an efficient energy conversion mechanism for microscopic artificial swimming objects and overcomes easily the viscous drag and gravitational forces in an aqueous environment at low Reynolds numbers. Swimming velocities as fast as 1.8 mm/s (24 times the body length per second) and pressure as high as 375.5 Pa are achieved by HNBs propelled by an electro-osmotic force under an external electric field. These performances are higher than other state-of-the-art microscale artificial swimmers with physical energy conversion. Note that the velocity of 24 body lengths per second is even faster than natural bacterial propulsion which is

their original model. Although helical morphology with a thin film is beneficial to such a fast swimming at low Reynolds numbers, the swimming performance appears to depend highly on the surface electrokinetic effects. Among the tested four different types, dielectric polymer coated HNB showed the best swimming performance compared to others with metallic or semiconductor surfaces. These HNBs can be used as wireless liquid manipulators, assemblers and nanorobots for biomedical or MEMS/NEMS applications or remote physical or chemical detection by functionalizing with proper readouts.

Cross-References

▶ Biomimetics
▶ Dielectrophoresis
▶ Microfabricated Probe Technology

References

1. Berg, H.C.: *E. coli* in Motion. Springer, New York (2004)
2. Purcell, E.M.: Lift at low Reynolds number. Am. J. Phys. **45**, 3–11 (1977)
3. Sundararajan, S., Lammert, P.E., Zudans, A.W., Crespi, V.H., Sen, A.: Catalytic motors for transport of colloidal cargo. Nano Lett. **8**, 1271–1276 (2008)
4. Honda, T., Arai, K.I., Ishiyama, K.: Micro swimming mechanisms propelled by external magnetic fields. IEEE Trans. Magn. **32**, 5085–5087 (1996)
5. Behkam, B., Sitti, M.: Bacterial flagella-based propulsion and on/off motion control of microscale objects. Appl. Phys. Lett. **90**, 023902(1–3) (2006)
6. Zhang, L., Abbott, J.J., Dong, L.X., Peyer, K.E., Kratochvil, B.E., Zhang, H., Bergeles, C., Nelson, B.J.: Characterizing the swimming properties of artificial bacterial flagella. Nano Lett. **9**, 3663–3667 (2009)
7. Dreyfus, R., Baudry, J., Roper, M.L., Fermigier, M., Stone, H.A., Bibette, J.: Microscopic artificial swimmers. Nature **437**, 862–865 (2005)
8. Ghosh, A., Fischer, P.: Controlled propulsion of artificial magnetic nanostructured propellers. Nano Lett. **9**, 2243–2245 (2009)
9. Abbott, J.J., Peyer, K.E., Lagomarsino, M.C., Zhang, L., Dong, L.X., Kaliakatsos, I.K., Nelson, B.J.: How should microrobots swim? Intl. J. Rob. Res. **28**, 1434–1447 (2009)
10. Martel, S., Tremblay, C., Ngakeng, S., Langlois, G.: Flagellated magnetotactic bacteria as controlled MRI-trackable propulsion and steering systems for medical nanorobots operating in the human microvasculature. Intl. J. Rob. Res. **28**, 571–582 (2009)
11. Kosa, G., Jakab, P., Hata, N., Jolesz, F., Neubach, Z., Shoham, M., Zaaroor, M., Szekely, G.: Flagella swimming for medical micro robots: theory, experiments and application. Proceeding of the 2nd IEEE RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 258–263. Scottsdale (2008)
12. Chang, S.T., Paunov, V.N., Petsev, D.N., Velev, O.D.: Remotely powered self-propelling particles and micropumps based on miniature diodes. Nat. Mater. **6**, 235–240 (2007)
13. Gao, P.X., Mai, W., Wang, Z.L.: Superelasticity and nanofracture mechanics of ZnO nanohelices. Nano Lett. **6**, 2536–2543 (2006)
14. Hwang, G., Hashimoto, H., Bell, D.J., Dong, L.X., Nelson, B.J., Schon, S.: Piezoresistive InGaAs/GaAs nanosprings with metal connectors. Nano Lett. **2**, 554–561 (2009)
15. Wang, Z.L., Song, J.H.: Piezoelectric nanogenerators based on zinc oxide nanowire arrays. Science **312**, 242–246 (2006)
16. Xu, S., Qin, Y., Xu, C., Wei, Y., Yang, R., Wang, Z.L.: Self-powered nanowire devices. Nat. Nanotechnol. **5**, 366–373 (2010)
17. Karrai, K., Grober, R.D.: Piezoelectric tip-sample distance control for near field optical microscopes. Appl. Phys. Lett. **66**, 1842 (1995)
18. Hunter, R.J.: Foundations of Colloid Science. Oxford University Press, New York (2010)
19. Berg, H.C., Brown, D.A.: Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. Nature **239**, 500–504 (1972)

Replica Molding

▶ Microcontact Printing
▶ Nanoscale Printing

Resistive Switching

▶ Graphene Nanostructures for Memristive Devices

Resonant Evanescent Wave Biosensor

▶ Whispering Gallery Mode Resonator Biosensors

Retina Implant

Retina Silicon Chip

Retinal Prosthesis

Reversible Adhesion

RF Filters

RF MEMS Reconfigurable Filter

RF MEMS Switches

RF MEMS/NEMS

RF-MEMS/NEMS

Rheology

Riblets

Rigorous Maxwell Solver

RNA Interference (RNAi)

RNAi in Biomedicine and Drug Delivery

Michèle Riesen and Nuria Vergara-Irigaray
Department of Genetics Evolution and
Environment, Institute of Healthy Ageing,
University College London, London, UK

R

Synonyms

RNA interference (RNAi); Short-interfering RNA (siRNA); siRNA delivery

Definitions

RNA interference (RNAi) is the targeted down regulation of gene expression, most commonly

on the translational level by introducing synthetic double-stranded or short interference RNA (dsRNA, siRNA) into the cell. Gene expression is then reduced by employing an existing cellular mechanism for processing the dsRNA into the active, single-stranded form incorporated in a complex called RISC, which targets and leads to cleavage of the specific messenger RNA (mRNA) substrate.

Biomedicine is the field of medicine that develops concepts or mechanisms discovered or used in basic research into clinical applications. This process is called *From Bench to Bedside*. In the context of RNAi or drugs that are delivered to the organism or target tissue by using nanoparticles, biomedicine is sometimes referred to as nanomedicine. The size of the particles ranges between 1 and 300 nm.

Overview

The recent advances in the understanding of the molecular mechanisms of RNA interference (RNAi) have led to the concept that dsRNA can be delivered into cells to modulate pathogenic functions of genes. This mechanism opens new possibilities to cure diseases since the underlying cause of nearly every pathological process is the dysregulation of one or several genes at different steps in a physiological process. This therapeutic application of RNAi as a new type of drug has been quickly embraced by many researchers as a promising strategy for diseases which were hitherto considered to be difficult or nigh impossible to treat. However, the delivery of dsRNA into cells is not a trivial task due to their large size of approximately 14 kDa and negative surface charge. Additionally, RNA has a rather short lifespan under physiological conditions due to the efficient removal by nucleases.

In order to successfully employ RNAi to target pathological processes, the stability of siRNA has to be addressed. This is generally achieved by introducing chemical modifications to the nucleotides without inducing toxicity or through the selected drug vesicle. Since RNAi is highly sequence-specific, an appropriate sequence in the

gene of interest has to be found and preferentially tested using different siRNAs against the same target. Further important factors which should be controlled for are off-target effects (OTEs), for example, cleavage of other mRNA than the sequence-defined target, which can arise despite the precision of the mechanism, and activation of the immune response which can be induced both sequence dependently and independently. The major challenge is still the delivery of the siRNA to the target organ for the treatment. This then expands the list of considerations to the choice of delivery vehicles and routes of administrations. For an excellent and comprehensive introduction into the field please see [11, 17].

Some of these issues, in particular the delivery of the siRNA, are solved by combining the expertise of material, chemical, and physical science to devise efficient vectors for delivery of the dsRNA into the target tissue. The development of carrier molecules in the nano-range, such as polymers, liposomes, peptides, and aptamers, goes hand in hand with the design of new therapeutics for various pathologies, such as cancer, genetic, and viral diseases. Notably, in less than 10 years since the discovery of RNAi, several drugs have been successfully developed and are currently tested in Phase I–III clinical trials in the United States (www.clinicaltrials.gov).

It is worth mentioning that RNAi can originate from either synthetic siRNA, which has to reach the cytosol for final processing and activation, or as plasmid-borne short hairpin RNA (shRNA). These plasmids must be delivered to the nucleus where the RNA is transcribed and undergoes a first step of processing before it is transported to the cytosol. This type of RNAi is generally carried by viral vectors and has been subject of extensive investigation. However, the focus of this review is nonviral delivery of siRNA; hence, the reader is referred to elsewhere for a further discussion of this area. Finally, bacteria have recently been employed as delivery vehicles for shRNA or siRNA in a process termed transkingdom RNAi (tkRNAi) or bacteria-mediated RNAi (bmRNAi).

In this entry, the basic cellular mechanism of RNAi will be introduced. Important criteria for the design of siRNA such as sequence design and

chemical modifications will be discussed together with observed off-target effects (OTEs). This will be followed by an overview over physiological barriers and approaches chosen to deal with these challenges. Further, different types of delivery strategies will be presented which are currently employed in clinical applications according. Finally, how nanotechnology is integrated with RNAi will be discussed with an overview and discussion of a selection of particles presently utilized for in vitro, in vivo, and clinical trials.

Basic Mechanisms

RNA Interference: A Shared to Defend the Genome

In 2006, Andrew Fire and Craig Mello were awarded the Nobel Prize in Medicine or Physiology for their discovery of gene silencing through double-stranded RNA (dsRNA) in *Caenorhabditis elegans (C. elegans)*. Notably, their work describing this phenomenon had only been published for 8 years before they received the honors [5]. The observation that gene silencing could be achieved by the introduction of plasmid DNA into plants was first described some 20 years ago, but the importance of this curious finding was not recognized until it was reported by Andrew Fire and colleagues in one of the most widely used model organisms, the nematode worm *C. elegans*.

Although the exact mechanisms of RNA interference (RNAi) were unknown when it was first reported, not much time passed before it became clear that the strategies and functions of RNAi are conserved in plants, invertebrates and vertebrates, as well as many of the components of the pathway. Once long dsRNA has entered the cell, it is trimmed into small duplexes of approximately 20–25 nucleotides in the cytosol by an RNase III enzyme called DICER1. This step is followed by integration into a complex called pre-RNA induced silencing complex (pre-RISC), also thought to be mediated by DICER1, where further processing into single-stranded RNA (ssRNA) takes place. RISC contains DICER1 and several proteins from the Argonaute family, AGO1–4. The passenger (sense) strand is then degraded

and the mature RISC is formed with the guide (antisense) strand integrated into RISC. The guide strand is complementary to mRNA and determines the specificity of RISC. When RISC encounters its target mRNA and perfect Watson–Crick base-pairing occurs, the mRNA is cleaved by the RISC-member AGO2. If the base-pairing is imperfect, the protein translation is disabled, but the mRNA is not cleaved by AGO2. This complex can be assembled stably over an extended period of time and catalytically regulate gene expression on the translational level (Fig. 1).

Since the discovery of RNAi, the knowledge of the field has progressed tremendously. Species possess a vast arsenal of their own small RNAs, for example, microRNAs (miRNA) or PIWI-interacting RNAs (piRNAs). The different classes of RNAs fulfill different functions such as regulating gene expression at the translational level, or they are expressed in specific tissues, for example, the germ line or neurons. Furthermore, RNA was found in the nucleus where it was found to regulate gene silencing at the transcriptional level by modulating histone modifications or DNA methylation levels. Further evidence proposes RNAi as some kind of nuclear pest control to defend the genome against insertion of mobile elements such as L1 transposons and viral DNA. Despite the rapid succession of discoveries of RNAi-governed processes, the detailed mechanisms by which these regulations are executed remain yet to be elucidated. Nevertheless, the impact the study of RNAi has had on the understanding of gene regulatory processes over the past decade must not be diminished. For comprehensive and recent reviews over the different pathways, processes, and the components of RNAi please refer to [12, 19].

Design of Synthetic Short RNA for Clinical Applications

Sequence Design

Intensive research into sequence design resulted in a set of criteria that greatly facilitates the development of suitable RNA duplexes for any given gene. There are a number of web tools available from either commercial or academic sources in order to identify appropriate RNAi target
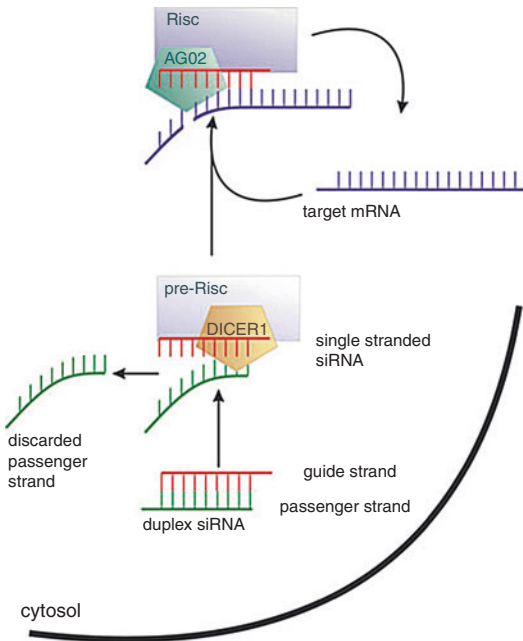
RNAi in Biomedicine and Drug Delivery, Fig. 1 siRNA-induced silencing. Double-stranded siRNA is delivered to the cytosol, where it is recognized by DICER1, and incorporated into a pre-RISC complex upon which the duplex is separated and the passenger strand discarded. This processing step is also mediated by DICER1. The mature RISC complex is then formed and guided to target mRNA. Perfect match interactions between the guide (antisense) strand and mRNA lead to cleavage of the substrate mRNA by AGO2. A fully assembled RISC complex is stable and can catalytically inactivate a large number of mRNA. The model is a simplified depiction of the siRNA pathway only and omits transport of siRNA into the cell as well as other small RNA processing pathways. The nomenclature refers to human proteins

sequences [15]. The basic structure of the siRNA generally contains a central domain of 19 basepairs with an overhang of two bases at the 3′-end and a 5′-phosphate group. The reason for this type of synthetic siRNA to be common is because it is identical with the end product of the cellular RNAi pathway and DICER1. However, other siRNA varying in length between 19 and 27 basepairs and with either blunt or asymmetric ends have been shown to be efficient in vitro and to be processed by DICER1. Secondary structures of the siRNA as well as the target sequence should be considered together with the

overall thermodynamic properties (T_m), which influence loading into RISC. As a rule, siRNA is more efficient when the 5′-end of the guide strand has a lower melting temperature than the 5′-end of the passenger strand by adjusting the A/U and G/C content, respectively. The positioning of individual bases has also been found to have an effect on efficiency of siRNA [2, 15].

Chemical Modifications to the Sugar or Phosphodiester Backbone

Chemical modifications of the dsRNA in order to increase the stability (T_m) under physiological conditions or reduce the susceptibility to RNase attacks include the introduction or replacement of various reactive groups on the base, sugar or phosphodiester backbone of the RNA. In general, the sugar backbone is modified at the 2′O- position, for example, to 2′-Fluoro. This modification does not interfere with loading into RISC or cleavage by DICER1. Another commonly used modification is 2′-O-Methyl. This nucleotide is occurring naturally in tRNA or ribosomal components and insertion of these modified nucleotides has been shown to abolish the induction of the innate immune system in mice. Locked nucleic acids (LNA), where the sugar ring is locked in the 3′-endo conformation due to the presence of a methylene bridge between the 2′-O and the 4′-C of the ribose, have been reported to increase stability successfully.

Modifications to the phosphodiester backbone include, for example, phosphorothioate, which has been shown to increase T_m and resistance to nuclease-mediated degradation. This modification can be easily integrated at defined positions, but it has been related to toxicity, loss of function, and nonspecific protein binding if large numbers of nucleotides in the siRNA are replaced with this modification. The replacement of the phosphodiester bond with a boranophosphate greatly increases stability of the dsRNA, but synthesis of this type of backbone requires in vitro transcription and the position of the insertion of boranophosphate can therefore not be defined. For excellent and detailed reviews of the design of siRNA and modifications, please refer to [1, 2, 8].

Off-Target Effects

A serious concern when developing siRNA as therapeutics are OTEs. They arise most commonly for the following reasons: activation of the innate immune system, sequence homology with other mRNAs additionally to the specific target, and when siRNA enters the cellular miRNA pathway. Initially, it was assumed that the short RNA duplexes could not induce an immune response. However, with the increasing number of work employing siRNA in vitro and in vivo, siRNAs were reported to stimulate the innate immune system and induce a strong interferon response. This activation has been shown to be both sequence dependent and independent, as well as cell-type specific. Intracellular receptors from the Toll-like family, in particular TLR3, 7, and 8 recognize RNA, either dsRNA (TLR3) or ssRNA (TLR7 and 8), and mediate the induction of IFN-α or IFN-β. These receptors are found either on intracellular organelles such as the endosome or lysosomes, as well as cells that constitute the innate immune system such as B cells or monocytes. Under given circumstances, siRNA can therefore induce the innate immune system [22].

In order to avoid OTEs based on sequence homology, it is recommended to carefully screen the transcriptome for similarities with the required siRNA. It is important to note that although single mismatches have been reported to abolish the efficiency of siRNA, short similarities of five to eight bases can direct RISC to unintentional targets. If siRNA engage in the miRNA pathway, the specific effect of siRNA on one gene is replaced by effects on many genes due to the nature of miRNA-mediated gene regulation. Generally, OTEs are dose-dependent and can therefore be circumvented by establishing dose–response curves and choosing the lowest efficient dose. Since it has been shown that downregulations of the targeted gene transcripts were in fact due to OTEs and not to the direct mediation of the administered siRNA, it is strongly recommended to test at least two or three different siRNAs against the same gene. This should result in reproducible effects and confirm that the expected and observed downregulation of the transcripts is due to the direct effect of siRNA [8, 15].

Delivery of Synthetic Short Interference RNA for Clinical Applications

Physiological Barriers

After the successful design, synthesis, and in vitro tests of a suitable siRNA compound, the next and likely more challenging obstacle to be overcome is the delivery of siRNA to its site of action. There is a consensus in the field that the vast potential of RNAi cannot be fully exploited yet because of the lack of efficient delivery vehicles to the target tissue. The body has various highly efficient strategies for clearance of foreign particles from the system, which are briefly described below in order to understand the difficulties in delivery and how they are addressed by the choice of appropriate vesicles or siRNA modification or both combined. Tremendous efforts are directed toward the development of a range of particles with different physicochemical properties which are equipped to reach the target organ and pass the physiological barriers [7].

If siRNAs are not immediately degraded by free nucleases then the next barrier is renal clearance, which is thought to determine the half-life of modified or naked siRNA in vivo. The small size of around 14 kDa, which is far below the cutoff of 40 kDa for glomerular filtration, and the hydrophilic nature of siRNA target these molecules for removal by the kidneys. The next trial is the clearance via the reticuloendothelial system (RES). The RES consists of mobile, circulating phagocytic cells, such as monocytes, or resident tissue macrophages that remove pathogens and cellular debris which remains from apoptosis, wound healing, or tissue remodeling. Large numbers of RES-associated cells reside in the liver, the Kupffer cells, and in the spleen or generally in other tissues with high blood flow and fenestrated vasculature.

Should carriers have successfully dodged clearance by any of the above-described pathways, they will then encounter the endothelial barrier lining the vascular lumen. Endothelial cells are tightly connected with the underlying extracellular matrix through integrins and they are joined together via a number of strictly regulated junctions. Endothelial cells are also

perforated with a large number of small pores of about 45 Å and rare large pores of 250 Å. Most of the transport of vesicles or siRNA is thought to be mediated via leaks in junctions rather than the pores themselves; however, there is some controversy regarding this matter. In many tissues, particles larger than 5 nm will not be able to extravasate through the endothelial lining, whereas organs such as the liver or spleen are easily accessible because of the extensive vasculature of these organs.

One specific type of endothelial barrier is the blood brain barrier (BBB) at the boundary to the astrocytes and pericytes where the endothelium forms an exceptionally tight lining to protect the brain. As progress in research into neurodegenerative diseases such as Alzheimer's reveals an ever-increasing number of genes which are potentially "drugable," the brain becomes an attractive target for siRNA. Under normal physiological conditions, this barrier is intact. However, the integrity of the BBB can be compromised during disease or through administration of certain drugs. Also, specific endothelial reporters have been described which could be employed to aid passage of particles across the BBB to the brain. Therefore, development of carriers that can deliver drugs across the BBB will become increasingly important to treat this kind of diseases.

The final step for successful delivery is the cellular uptake and, importantly, endosomal release into the cytosol. In general, particles enter a cell via receptor-mediated endocytosis. Recognition and binding to the receptor induces the formation of a vesicle which leaves the cell membrane on the cytosolic site. From there, the vesicles undergo a series of intracellular trafficking steps and maturation until, finally, late endosomes or lysosomes have been formed. The last challenge the particles face is the release from the endosome into the cell. The *proton sponge effect* has been described as a mechanism by which the trapped drug can exit the endosome. Briefly, acidification of the endosome leads to protonation of amine groups on the vesicle. This leads to an influx of protons and chloride ions, followed by water to equilibrate the changes in osmolarity, eventually causing rupture of the vesicle and release of its contents into the cytosol. For further reading, please see [7, 20].

Types of Delivery

In vivo application of siRNA can be classified, for example, into active or passive systemic and local/cellular delivery. Passive systemic delivery does not require any specific targeting sequences or molecules. It is based on the property of nanoparticles to accumulate in certain tissues under certain conditions. For example, nanoparticles or naked siRNA have been shown to accumulate in liver or spleen because of the specific structure of the endothelium, with large fenestrations of the microvasculature of these organs. Delivery of vesicles to inflamed tissue is thought to be mediated by the RES. Accumulation of particles in tumors has been attributed to the *enhanced permeability and retention effect* (EPR) in tumors. This is thought to be caused by reduced lymphatic drainage, leaky microvasculature, and enhanced angiogenesis of cancerous tissue.

Active or targeted systemic delivery is more complex because it requires the carrier to be equipped with components such as antibodies, ligands against cell-surface receptors, or peptides from viral proteins targeting the nanoparticle to the desired cell type or tissue. Antibodies are used as whole entities, or fragments thereof, and they are generally directed against receptors which are increasingly expressed on tumor cells. Receptors that have been targeted include, for example, the epidermal growth factor receptors (EGFR) or human epidermal growth factor receptor-2 (HER-2), the iron-binding protein transferrin (Tf), which targets the transferrin receptor (TfR) responsible for iron uptake, as well as Toll-like receptors (e.g., TLR3). Successfully employed peptides are the rabies virus glycoprotein (RVG) or arginine-glycine-asparagine (RGD). Sugars, i.e., glucose or mannose, are recognized by a number of cell-type specific receptors and have also been successfully employed in delivery to specific cell types or across the BBB.

Naked siRNA has been injected using the hydrodynamic delivery method into mouse tail veins and was subsequently found to accumulate

in the liver and to a lesser extent in other organs such as the lungs. This method requires large volumes to be rapidly injected, and although successfully employed in mice and pigs, it is doubtful whether this way of administering systemic siRNA would be of therapeutic use in humans. The volume for injection needed is around 10 % of the body weight, and this would most likely lead to heart failure and death. Direct injection of siRNA is feasible if the target is easily accessible, such as is used, for example, for treatment of wet age-related macular disease or other diseases of the eyes where siRNA is injected directly into the eye. Several Phase I–III clinical trials have been approved by the US FDA. Other means of delivery that have been tested include topical administration which is useful for exposed sites, such as the skin. Inhalation is the means of choice for treatment of respiratory tract diseases. Electroporation has been used for delivery into muscle tissue. For more detailed information, please see [14, 18, 21, 22].

Delivery Vesicles and Nanoparticles

Considerations for the Design of an siRNA Carrier Investigation into suitable carriers over the past decade has resulted in a vast amount of materials and combinations of several components which all have advantages and disadvantages for delivery of specific cargo, the route of administration, and the target tissues. Criteria to be considered are the size and size distribution of particles, which is, for example, inversely related to cellular uptake and delivery to tumor tissue. Since the smallest microvessels have a diameter of 5–6 μm, particles must be considerably below this diameter in order to avoid accumulation and clogging which could lead to embolism. The surface properties of nanoparticles dictate whether they will be recognized and removed by the host immune system; non-modified hydrophobic particles are generally swiftly removed, whereas vesicles decorated with biodegradable materials which are hydrophilic such as polyethylene glycol (PEG) or polysorbate 80 (Tween 80) are better tolerated.

Drug loading capacities and the release characteristics are further important points. Loading can be achieved by two methods. The drug can be incorporated at the time of nanoparticle assembly or it can be absorbed after the formation of the carrier by incubation in a concentrated drug-solution. Highly efficient loading has been shown for siRNA due to the negative charge of the oligonucleotide and positively charged vesicles, such as cationic lipids or polymers. Drug release depends, for example, on the degradation of the carrier or how fast the drug can diffuse across the polymer membrane, as well as the solubility of the drug. A concise selection of the most frequently and successfully employed carriers is briefly described below. For a detailed description of these design considerations, please see [4, 18]. The reader is also referred to cross-referenced entries which elaborate in more detail the properties and design of nanoparticles specifically for drug delivery.

Aptamers Among biological macromolecules RNA has a unique place because of the ease of manipulation and synthesis similar to DNA, which is combined with a structural versatility found in proteins. These useful properties predestine RNA to be used itself as a building block for the assembly of nanoparticles called RNA-aptamers (Fig. 2a). Generally, aptamers are highly structured oligonucleotides that are obtained by SELEX (systematic evolution of ligands by exponential enrichment) for the required purpose. This is a protocol whereby a library of RNA-aptamers is subjected to iterative rounds of counter-selection and selection to yield aptamers which recognize, for example, specific receptors on target cells, proteins, as well as organic compounds. The prime features which make aptamers attractive carriers are their small size of a few nanometers compared to cationic polymers or liposomal carriers (upto 300 nm), that modified nucleotides can easily be integrated to increase stability in physiological conditions, they have low toxicity and are biodegradable. Instances where aptamers have been employed as carriers for siRNA include murine xenografts of prostate cancer where tumor growth and regression was mediated by aptamer-mediated siRNA delivery, or treatment against viral infection by

R

either targeting HIV gp120, a glycoprotein which recognizes and utilizes the CD4 receptors on T lymphocytes for entry, or the CD4 receptor itself [6, 10].

Cationic Polymers and Dendrimers The cationic polymer cyclodextrin (CDP) has been widely and successfully used as a carrier directed in particular against solid tumors. Advantages of this material are the low toxicity; it is not metabolized in humans and does not induce an immune response. CDPs are commonly of a tri-composite structure, where CDP is employed as the carrier for siRNA due to strong ionic interactions between the cationic polymer and the anionic nucleic acids. The particle is stabilized with the hydrophilic cycloalkane adamantane conjugated to PEG (AD-PEG). Targeting of the vesicle to tumors is achieved by adding an AD-PEG-Transferrin (AD-PEG-Tf) moiety because Tf is expressed abundantly in a number of tumors. This formulation carrying siRNA against the ribonucleotide reductase subunit 2 (RRM2), called CALAA-01 by Calanda Pharmaceuticals, was approved for Phase I clinical trials by the FDA in 2008. The reader is referred to [3, 9] for a comprehensive review on the development of CALAA-01.

Another cationic species that is frequently used for similar properties is polyethylene-imine (PEI), which can be synthesized to be of linear or branched structure and is then classified as a dendrimer. PEI efficiently forms stable nanoparticles with oligonucleotides and promotes endosomal release via the *proton sponge effect*. Because of its branched structure, it can carry a large payload. Despite the supreme quality of PEI as a siRNA delivery vesicle, its identified toxicity toward most cells is not a desirable side effect. Three approaches have been taken to address the toxicity issue of PEI. First, the dendrimer is equipped with targeting moieties such as RGD to reduce the toxicity to nontarget tissue. Second, maintaining the particle size below 10 kDa substantially reduced toxicity compared to larger particles of 25 kDa. Finally, ketalized carriers have been synthesized with the added advantage that these linkages are sensitive to acidic pH, and hence readily dissolve in endosome to release the cargo [13].
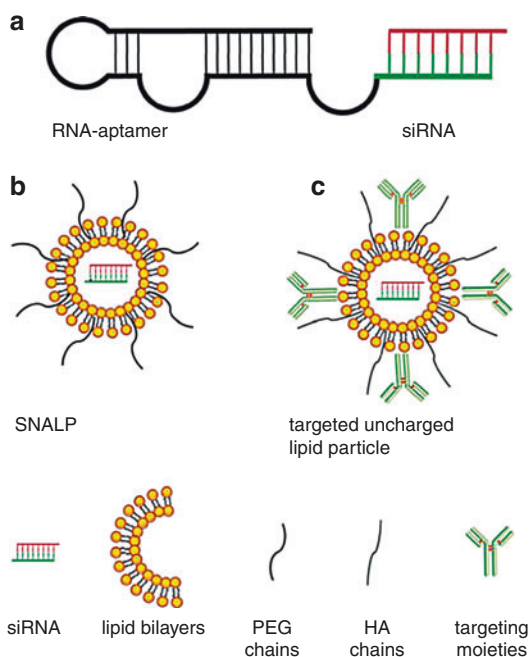
RNAi in Biomedicine and Drug Delivery, Fig. 2 Delivery vehicles for siRNA administration. Three types of vesicles are shown for delivery of siRNA. (**a**) RNA-aptamer with siRNA ligated. (**b**) Stable nucleic acid particle decorated with PEG (SNALP). (**c**) Uncharged lipid particle decorated with targeting moieties, i.e., single chain antibody fragment directed against Tf and HA chains

Liposomal Carriers Lipids as gene delivery vehicles have been used since the early 1980s. Cationic lipids are often used as vesicle due to the facile handling and preparation since they readily associate with the anionic nucleic acids. Moreover, routinely used reagents for in vitro transfection of cell culture are based on cationic liposomes, such as *Lipofectamine 2000* by Invitrogen or other equivalent products. The most commonly used cationic lipid is 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) because it efficiently associates with nucleic acids. However, DOTAP was again shown to be cytotoxic; hence, this carrier was equipped with modifications which reduce this side effect, such as lactosylation. Another strategy to inhibit the growth of pancreatic xenografts on a murine model involved DOTAP decorated with histidine-lysine peptides. This facilitated endosomal escape and single-chain antibody

fragment directed against Tf, and this vesicle did not induce an IFN-response and could be administered at low doses due to the target specificity.

Another promising type of vesicle based on cationic lipids is the stable nucleic acid-lipid particle (SNALP) (Fig. 2b). The siRNA is embedded within the lipid envelope and the composition of SNALPs contains reduced cationic lipid content and is PEG-ylated. SNALPs have been subjected to extensive study as siRNA carrier against in vivo viral models with considerable success of full protection against infection of guinea pigs with the deadly Zaire strain of the Ebola virus. Another trial employed SNALPs with siRNA against apolipoprotein B levels in nonhuman primates, which was well tolerated and showed reduction of Apolipoprotein B (ApoB) and serum cholesterol among other physiological improvements. Following on from these promising results, several therapeutic siRNA-SNALP formulations against tumors or metabolic diseases are now in preclinical or have been admitted to Phase I clinical trials, and targets include VEGF and ApoB among others [8]. Further information on these trials can be found on www.clinicaltrials.gov or the web sites listed.

Since cationic liposomes have been associated with significant toxicities, development of vesicles composed of uncharged lipids has been resumed with considerable success. One lipid in particular has emerged as a promising carrier: 1,2-dioleoyl-sn-glycero-3-phosphatidylcholine (DOPC). These vesicles successfully reduced tumor xenografts of different origins. Targeted vesicles against the thrombin receptor (TR) expressed on some tumors greatly reduced not only TR but also angiogenic and invasive tumor markers, such as vascular enodothelial growth factor (VEGF) and matrix metalloproteinase 2 (MMP2). Techniques developed in order to maximize the payload of these nanoparticles utilize lyophilization followed by rehydration with siRNA. The half-life in circulation under physiological conditions of these particles was increased by surface coating with hyaluronic acid (HA), a naturally occurring glucosaminoglycan with properties equivalent to PEG (Fig. 2c). Further information can be found in [17].

Other Types of Carriers: Atellocollagen, Cationic Cell Penetrating Peptides (CPPs), Polylysine There are a number of other options for building siRNA-nanocarriers. Atellocollagen is a preparation of digested cationic calf dermal collagen successfully employed and well tolerated in animal models of inflammatory disease or tumor xenografts. Despite the fact that the studied atellogen-siRNA nanoparticles were not decorated with a specific targeting moiety, they were found to accumulate in tumors and inflamed areas presumably due to the EPR effect in these tissues. CPPs are short peptides with a high content of positively charged amino acids, in particular arginine, which can be synthesized or are derived from naturally occurring peptides, e.g., TAT from HIV-1 to facilitate cellular uptake. Initially thought to be fully soluble, it appears that their efficiency is due to aggregation into nanoparticles. Successful trials in vivo used polyarginine fused to cholesterol and siRNA against VEGF in a tumor xenograft. Another polyarginine-based CPP coupled with RVG was found to cross the BBB and reduce green fluorescent protein expression in a mouse model. Another type of carrier is based on polylysine (K) stretches. Initial problems with toxicity such as complement activation were overcome by PEG-modifications and the addition of histidine (H), which facilitates endosomal escape. Further, it was found that the efficiency as siRNA carriers was improved by increasing the histidine ratio compared to lysine [9].

R

Further Reading

Currently, 20 studies are listed on www.clinicaltrials.gov when searched for the keyword siRNA. This clearly shows that the challenges of delivery posed by the physiological barriers or physicochemical properties of the particles can be overcome. The marriage between advances in material sciences and the developing understanding of the physiological and pathophysiological processes culminates in the development of novel RNAi-based therapeutics for the treatment of diseases ranging from viral to inherited and age-related pathologies.

Due to space restraints, a number of additional excellent reviews could not be referenced in this entry; however, they can be found when searching Pubmed or Web of Science with keywords such as "siRNA AND nanotechnology." Further information on research and development of siRNA-based therapeutics can be found on the web sites of the leading companies.

http://www.alnylam.com/
http://www.opko.com/
http://www.quarkpharma.com/
http://www.silence-therapeutics.com/
http://www.sirna.com/
http://www.tekmirapharm.com/Home.asp

An insightful and curated web site on the recent development of RNAi therapeutics, career opportunities, consulting, and overview over the market and enterprise portfolios can be found here:

http://rnaitherapeutics.blogspot.com/

Cross-References

► Genotoxicity of Nanoparticles
► Lab-on-a-Chip for Studies in *C. elegans*
► Liposomes
► Nanomedicine
► Nanoparticle Cytotoxicity
► Nanoparticles
► Nanotechnology
► Nanotechnology in Cardiovascular Diseases

References

1. Behlke, M.A.: Progress towards in vivo use of siRNAs. Mol. Ther. **13**, 644–670 (2006)
2. Behlke, M.A.: Chemical modification of siRNAs for in vivo use. Oligonucleotides **18**, 305–319 (2008)
3. Davis, M.E.: The first targeted delivery of siRNA in humans via a self-assembling, cyclodextrin polymer-based nanoparticle: from concept to clinic. Mol. Pharm. **6**, 659–668 (2009)
4. Fattal, E., Barratt, G.: Nanotechnologies and controlled release systems for the delivery of antisense oligonucleotides and small interfering RNA. Br. J. Pharmacol. **157**, 179–194 (2009)
5. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C.: Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391**, 806–811 (1998)
6. Guo, P.: RNA nanotechnology: engineering, assembly and applications in detection, gene delivery and therapy. J. Nanosci. Nanotechnol. **5**, 1964–1982 (2005)
7. Juliano, R., Bauman, J., Kang, H., Ming, X.: Biological barriers to therapy with antisense and siRNA oligonucleotides. Mol. Pharm. **6**, 686–695 (2009)
8. Lares, M.R., Rossi, J.J., Ouellet, D.L.: RNAi and small interfering RNAs in human disease therapeutic applications. Trends Biotechnol. **28**, 570–579 (2010)
9. Leng, Q., Woodle, M.C., Lu, P.Y., Mixson, A.J.: Advances in systemic siRNA delivery. Drugs Future **34**, 721 (2009)
10. Levy-Nissenbaum, E., Radovic-Moreno, A.F., Wang, A.Z., Langer, R., Farokhzad, O.C.: Nanotechnology and aptamers: applications in drug delivery. Trends Biotechnol. **26**, 442–449 (2008)
11. Martin, S.E., Caplen, N.J.: Applications of RNA interference in mammalian systems. Annu. Rev. Genomics Hum. Genet. **8**, 81–108 (2007)
12. Moazed, D.: Small RNAs in transcriptional gene silencing and genome defence. Nature **457**, 413–420 (2009)
13. Mok, H., Park, T.G.: Functional polymers for targeted delivery of nucleic acid drugs. Macromol. Biosci. **9**, 731–743 (2009)
14. Oliveira, S., Storm, G., Schiffelers, R.M.: Targeted delivery of siRNA. J. Biomed. Biotechnol. **2006**, 63675 (2006)
15. Peek, A.S., Behlke, M.A.: Design of active small interfering RNAs. Curr. Opin. Mol. Ther. **9**, 110–118 (2007)
16. Raemdonck, K., Vandenbroucke, R.E., Demeester, J., Sanders, N.N., De Smedt, S.C.: Maintaining the silence: reflections on long-term RNAi. Drug Discov. Today **13**, 917–931 (2008)
17. Schroeder, A., Levins, C.G., Cortez, C., Langer, R., Anderson, D.G.: Lipid-based nanotherapeutics for siRNA delivery. J. Intern. Med. **267**, 9–21 (2010)
18. Singh, R., Lillard Jr., J.W.: Nanoparticle-based targeted drug delivery. Exp. Mol. Pathol. **86**, 215–223 (2009)
19. Siomi, H., Siomi, M.C.: On the road to reading the RNA-interference code. Nature **457**, 396–404 (2009)
20. Tiemann, K., Rossi, J.J.: RNAi-based therapeutics-current status, challenges and prospects. EMBO Mol. Med. **1**, 142–151 (2009)
21. Tokatlian, T., Segura, T.: siRNA applications in nanomedicine. Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. **2**, 305–315 (2010)
22. Weinstein, S., Peer, D.: RNAi nanomedicines: challenges and opportunities within the immune system. Nanotechnology **21**, 232001 (2010)

Robot-Based Automation on the Nanoscale

Sergej Fatikow, Daniel Jasper, Christian Dahmen, Florian Krohs, Volkmar Eichhorn and Michael Weigel-Jech
Division Microrobotics and Control Engineering (AMiR), Department of Computing Science, University of Oldenburg, Oldenburg, Germany

Synonyms

AFM; Atomic force microscopy; Carbon-nanotubes; DNA manipulation; Manipulating; Nanorobotic assembly; Nanorobotic manipulation of biological cells; Nanorobotics; Nanorobotics for bioengineering; Nanorobotics for MEMS and NEMS; Scanning electron microscopy; Visual servoing for SEM

Definition

The handling of micro- and nanoscale objects is an important application field of robotic technology. Nanohandling includes the finding, grasping, tracking, cutting, etc., of objects as well as different characterization methods such as indenting or scratching on the nanoscale. Measurement of different features of the object, probe positioning with nanometer accuracy, structuring or shaping of nanostructures, and generally all kinds of changes to matter at the nanolevel could also be defined as nanohandling in the broadest sense. This entry addresses several approaches that can be automated with the help of nanohandling robots. The development of nanohandling robot systems working in constricted work spaces (in a vacuum chamber of an SEM or under an optical microscope) and with a nanometer resolution and accuracy is a big technological challenge for the robotics research community. Advanced actuator and sensor technologies suitable for nanohandling have to be investigated and implemented as well as vision-based feedback methods for tracking and data acquisition. Another crucial issue is the development of real-time robot control methods that meet the demands of automated nanomanipulation. This entry presents several promising solutions and applications of robot-based nanohandling.

Robot-Based Automation on the Nanoscale

Introduction to the Topic

The handling of micro- and nanoscale objects is an important application field of robotic technology. It is often referred to as nanohandling, having in mind the range of aspired positioning accuracy for the manipulation of micro- and nanoscale objects of different nature. The nanohandling of objects may include their finding, grasping, moving, tracking, releasing, positioning, pushing, pulling, cutting, bending, twisting, etc. Additionally, different characterization methods like indenting or scratching on the nanoscale, measurement of different features of the object, probe positioning with nanometer accuracy, structuring or shaping of nanostructures, and generally all kinds of changes to matter at the nanolevel could also be defined as nanohandling in the broadest sense. As in the field of industrial robotics, where humans leave hard, unacceptable work to robots, robots with nanohandling capabilities can help humans to handle extremely small objects with very high accuracy. The size of these robots also plays an important role in many applications [1, 2]. Highly miniaturized robots, often referred to as microrobots, are able to operate in constricted work spaces, for example, under a light microscope or in the vacuum chamber of a scanning electron microscope (SEM). Microsystem technology (MST) and nanotechnology require this kind of robot, since humans lack sensing capabilities, precision, and direct manipulation at those scales. Automated nanohandling by microrobots will have a great impact on both these technologies [3]. This entry will address several aspects of nanohandling automation.

The development of nanohandling robot systems is a big technological challenge for the

robotics research community. Advanced actuator and sensor technologies suitable for nanohandling have to be investigated and implemented. Another crucial task is the development of real-time robot control methods that meet the demands of automated nanomanipulation. The latter is challenging, especially due to the difficulty of getting real-time visual feedback and the lack of advanced control strategies able to deal with changing and uncertain physical parameters and disturbances.

Trends in Nanohandling

The following three approaches are being pursued in the nanohandling research community:

- Top-down approach utilizing serial nanohandling by microrobotic systems. The main goal is the miniaturization of robots, manipulators, and their tools as well as the adaptation of robotic technology (sensing, actuation, control, automation) to the demands of MST and nanotechnology. This approach is the major topic of this entry.
- Bottom-up approach or self-assembly, utilizing parallel nanohandling by autonomous organization of micro- and nanoobjects into patterns or structures.
- The use of a scanning probe microscope (SPM) as a nanohandling robot. In this approach, the (functionalized) tip of an atomic force microscope (AFM) probe or of a scanning tunneling microscope (STM) probe acts as a robot end-effector that is applied to a nanoscale part. This approach has been actively investigated in recent years [1, 2] and is discussed in section "AFM as a Nanohandling Robot."

Several other approaches like the use of optical tweezers or electrophoresis might also be adapted for automated nanohandling.

Automated Microrobot-Based Nanohandling

Microrobotics for handling micro- and nanoscale parts has been established as a self-contained research field for nearly 15 years and, in recent years, a trend toward the microrobot-based automation of nanohandling processes has emerged [4, 5].

Process feedback is the most crucial aspect of nanohandling automation. It is difficult to obtain reliable information when handling microscale and especially nanoscale parts. Vision feedback often is the only way to control a nanohandling process. For this reason, the vacuum chamber of a scanning electron microscope is, for many applications, the best place for a nanohandling robot. It provides an ample work space, very high resolution up to 1 nm, and a large depth of field. However, real-time visual feedback from changing work scenes in the SEM containing moving nanorobots is a challenging problem.

Automated Microrobot-Based Nanohandling Station (AMNS)

Figure 1 presents a generic concept of an automated microrobot-based nanohandling station (AMNS).

The microrobots are usually driven by piezoactuators that enable positioning resolutions down to sub-nanometer ranges. The microrobots of the station have a nanomanipulator integrated in their mobile platform, which makes them capable both of moving over longer distances and of manipulating with nanometer accuracy. This leads to more flexibility as the robots can be deployed anywhere inside the SEM vacuum chamber. Stationary robots can also be used in an AMNS. They are easier to control compared to mobile robots, which makes them more suitable for high-throughput automation. Depending on the application, different combinations of both robot types can be implemented.

The sensor system of the AMNS includes SEM, video cameras, force sensors, as well as position sensors integrated into the robots axes. The sensor data is sent to the station's control system for real-time signal processing. The control systems task is to calculate the positions of the robots and their tools as well as the positions of the parts to be handled or other objects of interest. The calculated positions serve as input data for the closed-loop robot control.

Teleoperation often is the first step on the way to automation. For this reason, a user interface is an important component of the AMNS.
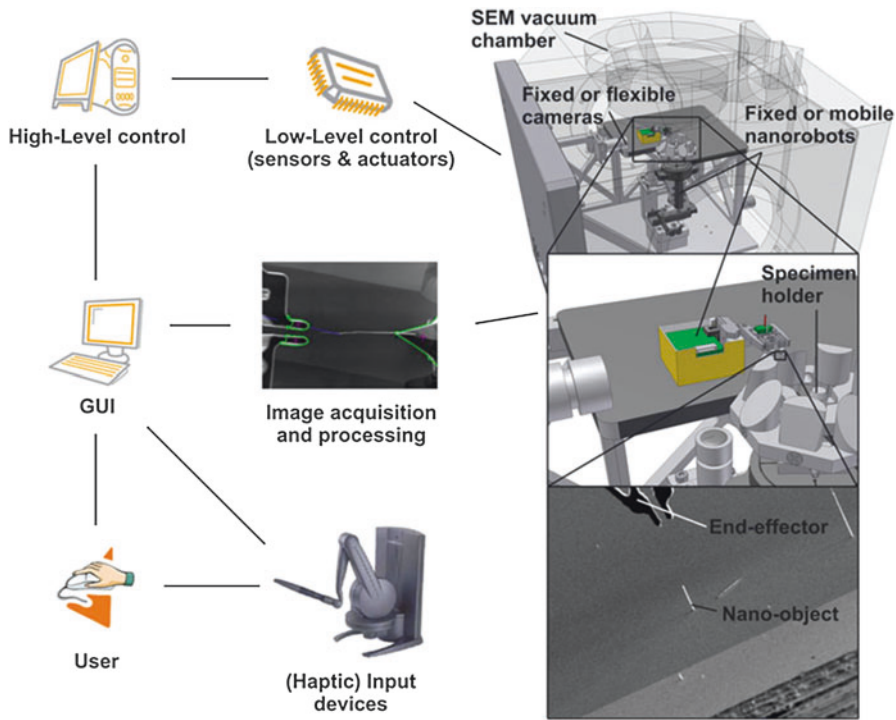
Robot-Based Automation on the Nanoscale, Fig. 1 Generic concept of the AMNS [2]

The user can influence the handling process by a haptic interface and/or by a graphical user interface (GUI). A good overview of state of the art teleoperation techniques and applications is given in [6].

The positioning accuracy of the microrobots during automated nanohandling is affected by several factors, so that a powerful robot control system is required. On the low level of the control system, driving voltages for the robot actuators are calculated in real time, which keep the robot and its end-effector on the desired path. The high-level control system is responsible among others for path planning, error handling, and the time-saving parallel execution of tasks. GUI and haptic interface are supported by the high-level control system as well.

The AMNS concept has been implemented in different application fields for (semi-) automated nanomanipulation [2, 3]. Relevant aspects of this work will be introduced in the following sections.

Structure of the Essay
The following is a brief summary of the topics covered in the following sections.

Section "Real-Time Vision Feedback" covers the vision feedback in nanohandling robot stations. The focus is on the use of SEM in combination with real-time image processing algorithms, which is proposed as the near-field sensor for the automation of nanohandling tasks.

Section "Position Control Inside SEM" introduces fundamental considerations for automated positioning inside the SEM. The section encompasses actuation types, motion principles, closed-loop control mechanisms for nanorobots, SEM-based visual servoing, as well as the difference between macro- and microscale control engineering.

Section "AFM as a Nanohandling Robot" introduces the atomic force microscope as a robotic system for manipulation at the nanoscale. The AFM tip acts as a nanohandling tool to

change the position or orientation of a nanoobject or to modify the surface of a substrate with nanometric precision. Different application fields are introduced and current challenges in this field of research are discussed.

Section "Application Example: Handling of CNT" describes nanorobotic strategies for the microgripper-based pick-and-place handling of individual carbon nanotubes. For a reliable alignment in z-direction, a shadow-based depth detection technique is presented. Furthermore, special handling strategies are introduced for placing and releasing the CNT on the target structure. As an example, the automated assembly of so-called CNT-enhanced AFM probes is demonstrated.

Section "Application Example: Handling of Biological Materials" deals with the characterization and manipulation of biological objects by an atomic force microscope. An overview of relevant parameters for AFM measurements of samples in liquids and of soft samples is given. After a short introduction of the biological background, the current state of the art in automated AFM-based handling of biomaterials is given, ranging from the design of DNA-based nanoelectric circuits to the structuring of biosensor chips for medical and forensic use.

Real-Time Vision Feedback

Manipulation and handling of objects on the nanoscale is subject to strong uncertainty on position and status of objects. This is due to the low mass of the objects of interest and the relatively high parasitic and interacting forces. Additionally, the choice of imaging modalities available at this scale is limited. Optical microscopes cannot deliver resolutions sufficient for imaging nanosized objects. Scanning probe microscopes like the AFM, though having the advantageous possibility to be used as a manipulator, do not exhibit the necessary speed for closed-loop feedback. The most flexible solution for imaging in this context is the SEM. It allows for observing processes at the nanometer range and still is fast enough to enable closed-loop control feedback. Drawbacks of the SEM are the sensitivity to noise and movement artifacts during fast scans, the possibility of distortions due to magnetic or

electric influences generated by the AMNS, and the lack of three-dimensional information. Various approaches have been developed to deal with these conditions. The important properties of the SEM imaging and the tasks and approaches to solve the problems associated with it will be described in the following sections.

Imaging Source Properties

Certain properties of the SEM and its imaging process are important for the applied computer vision approaches. The most important negative property is the high level of noise. Noise can be reduced by frame averaging or pixel averaging which, however, also reduces the frame rate. Figure 2 shows the SEM images with low speed and high speed and the corresponding noise level.

The noise in the SEM can be modeled as additive noise:

$$g_{\text{SEM}} = g + g_{\text{noise}}$$

With g the true gray level value, g_{noise} the noise component, and g_{SEM} the acquired gray level value.

Because the generation of secondary electrons is a statistic process and the number of electrons generated and detected can be assumed as Poisson distributed, the gray level fluctuation also has a Poisson characteristic.

Due to the scanning principle of the SEM, fast moving objects appear distorted in the images. This is due to the fact that the object has moved during the time needed to finish each line of the scan. This can significantly reduce the performance of computer vision algorithms. Therefore, the speed of robots has to be restricted if computer vision algorithms are required.

Typical Visual Feedback Tasks

For a successful nanohandling automation, visual feedback is needed for specific tasks. These can be described as object recognition, classification, and position determination. The latter can be executed in two dimensions and in three dimensions, respectively. Important especially for the position determination is speed in order to allow for closed-loop control. Still, even with fast imaging,
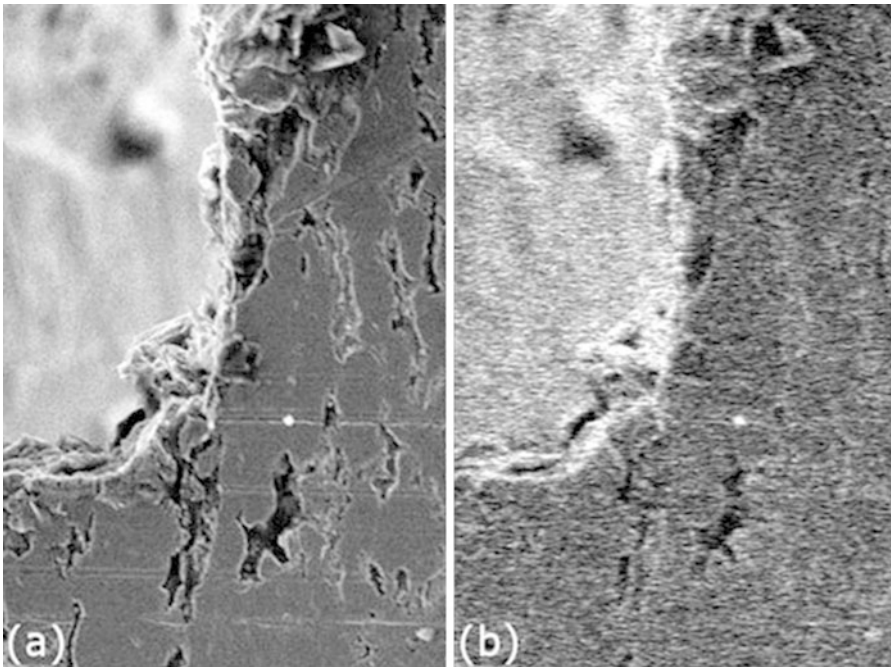
Robot-Based Automation on the Nanoscale, Fig. 2 SEM images at different scan speeds, acquired with a LEO 1,450 tungsten emitter SEM, (**a**) slow scanning, (**b**) fast scanning

control is often restricted to a look-then-move strategy, because the actuation is orders of magnitude faster than the imaging and subsequent visual feedback. When using SEM image acquisition with computer vision approaches as an image-based sensor, this is a limitation which cannot be overcome easily.

Object Recognition
Object recognition has the task of determining if an object of interest is in the view field or not. Many different approaches are known in traditional macro- or microscale imaging and computer vision. The approaches used at the nanoscale include template matching, statistical pattern recognition, and recognition based on local features. Object recognition algorithms used for nanohandling automation in many cases do not have special and elevated needs for noise robustness, because the recognition task is not time critical and low frame rates with high averaging can be used. Approaches used in nanohandling applications use support vector machines and various feature extraction

algorithms. Applications include CNT-detection and detection of TEM lamellae [7].

Two-Dimensional Position Determination and Tracking
The positions of objects and tools are not known precisely through any sensor except for the SEM. Object position determination and tracking allows image-based visual servoing approaches to work. Example methods which have been developed to work on SEM images rely on template matching [8], active contours [3], and rigid body models [9]. What is common in all these methods is that some measures have been included to diminish the effect of noise on the algorithm performance, either by use of inherent noise robustness like in the case of correlation-based template matching, the use of regions to cancel noise influence, or the usage of multiple or even manifold features and the selection and weighting of these features according to their expected significance and robustness.

Different algorithms may be suitable for different needs, depending on the degrees of freedom

and the boundary conditions imposed by the system setup and the object properties. If the objects of interest are rigid, their appearance constant and known, their expected movement restricted to translations, and no magnification changes are required, template matching using cross correlation has turned out to be a robust and powerful yet simple approach. The idea is to calculate the cross correlation of a template image calculate the cross correlation of a template image and an input image I:

$$C(x,y) = T(x,y) * I(x,y)$$

In the resulting matrix, each value corresponds to the similarity of the image part to the template. Assuming that the object of interest is in the image, the position is with very high probability at the position of the maximal value in this matrix:

$$C(x_0, y_0) = \max(C(x,y))$$

The algorithm delivers a single position. It may well determine the position with sub-pixel accuracy, using weighting or interpolating techniques. The speed can be improved by using the frequency domain for calculation. In Fig. 3, the result of the algorithm can be seen. The template is marked at the place the best match is found.

If the objects may rotate or change in size, active contours or rigid body models are more suitable, depending on whether an a priori model of the object is available or not. Also, the occurrence of occlusions is less problematic with these algorithms. The principle behind active contours is that a contour represented by a polygon or spline is combined with an energy function which will determine its shape and position. The energy function consists of two parts:

$$J = J_{int} + J_{ext}$$

J_{int} is the internal energy which purely depends on contour properties like length, compactness, or smoothness. For compactness, the internal energy J_{int} can be formulated as:
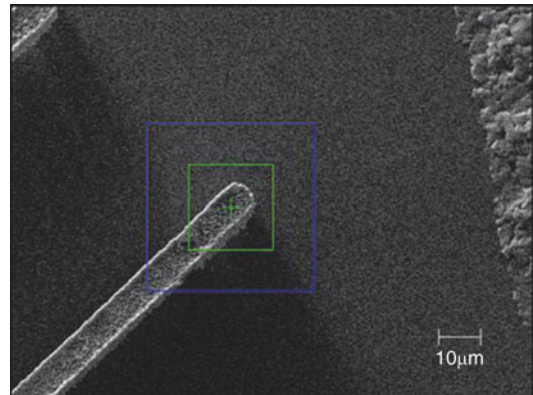
Robot-Based Automation on the Nanoscale, Fig. 3 Tip of a microgripper tracked with correlation tracking

$$J_{int} = \frac{L^2}{A}$$

The second part is the external energy, which depends on the image content or image features. Possible implementations of this evaluate the distance to edges or the region statistics. The energy function is minimized for each frame by applying transformations to the contour. These transformations can be either single point translations or Euclidean transformations of the whole contour.

Figure 4 shows an active contour tracking an object in SEM images.

Using single point translations, deformable and nonrigid objects can be tracked, but occlusion is problematic. Using Euclidean transformations of the entire contour, the algorithm is robust against occlusion at the expense of trackability of deformable objects. Using these kinds of algorithms, full automation of nanohandling applications has been performed.

Three-Dimensional Position Determination

A difficult issue with image-based position determination using SEM images is the extraction of the third dimension. Any setup used for nanohandling and nanomanipulation is essentially a three-dimensional one, which means that in many cases, a two-dimensional position extracted from an image is insufficient for reliable and flexible automation. Several approaches exist to solve this problem:
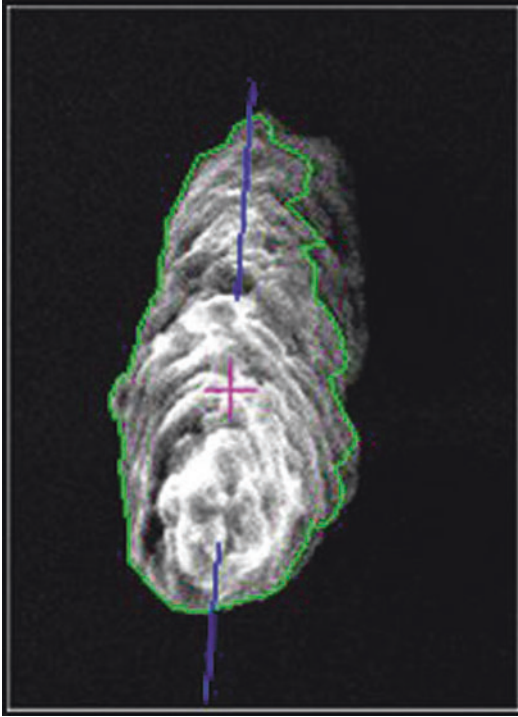
Robot-Based Automation on the Nanoscale, Fig. 4 Object tracked with active contours in SEM images

– Restrict the setup or degrees of freedom to a plane, which is an unpractical solution.
– Use different image sources to enable depth estimation from multiple views. This includes the usage of additional microscope cameras to roughly estimate the information missing in the SEM image, as well as the use of beam deflection mechanisms to acquire SEM images from different angles [3].
– Use models of the SEM image generation and the lens distortions to extract the third dimension. This has been successfully done with smaller magnifications by [9].
– Use additional information from the images, such as defocus [20] or shadow to estimate the missing information or at least determine if the object or tool is at the desired location in space.

All these approaches have different advantages and disadvantages. Restricting the setup or degrees of freedom as mentioned has the advantage that no additional measures are needed to deal with the three-dimensional problem. This restricts the possibilities for handling and manipulation to an unacceptable extent. The use of different sources may deliver everything from rough estimates in case of microscope cameras to very precise (nm resolution) measurements in case of well-designed stereo imaging SEMs or SEM/FIB combinations. Drawbacks here are the reduction in speed when using stereo SEM and the complex correspondence problem in very dissimilar images when using a camera or the SEM/FIB combination. The use of models for the SEM imaging has produced quite good results, but depends strongly on the accuracy of the imaging model and on the accuracy of the rigid body model of the object which is tracked. This restricts the application to some extent to objects for which a model can be assumed, and to a certain parameter range of the SEM imaging. The use of additional information from the image mainly concentrates on defocus and shadow. Using defocus of objects can deliver an accuracy up to 5 μm using time-consuming focus sweeps, and it can also be used to estimate the out-of-focus displacement and therefore the position of any object for a given frame, but with reduced accuracy and reliability. This is partly due to the complex nature of the electron beam optics and the high number of factors influencing the image, especially image and object sharpness.

Position Control Inside SEM

A common challenge for the automation of handling and assembly tasks in the SEM is the precise alignment of two or more objects carried by different robots. Such objects can be wafers with micro- or nanostructured components as well as different tools including grippers, probes, and cantilevers. Dependent on the size of the involved objects, the accuracy of this alignment has to be in the micrometer or even nanometer range. Accuracy in this sense stands for the precision of the relative alignment, that is, a nanoobject needs to be brought precisely into the opening of a gripper. For automated positioning, a closed-loop controller is necessary.

Actuation Types and Motion Principles

To facilitate closed-loop motion control inside SEMs, actuators capable of generating the required resolution need to be designed. Although there are efforts to scale conventional actuators such as electromotors to enable such a resolution, for example, employing backlash-free gears, novel actuation principles are better suited. In general, an actuator capitalizes on a *physical effect* in order to create motion. Most of the currently employed actuators for nanohandling are based on the inverse piezoelectric effect, that is, the controlled deformation of a piezoelectric element using an electric field. Piezoelectric actuators are popular owing to their large forces, high precision, and fast response time. Such actuators can create forces in the kN-range and respond to an input signal within microseconds. Major downsides are the limited deformation of about 0.1 % as well as a strongly nonlinear behavior affected by hysteresis and creep. Other physical effects that are used include electrostatic, thermal, and magnetostrictive actuators as well as shape memory alloys [4].

Piezoelectric actuators can be further categorized into scanning and step-wise actuators. *Scanning actuators* use a backlash-free cinematic structure based on flexure hinges to amplify the relatively small displacements of the piezoelectric elements. Although commercial actuators can reach working ranges of up to 500 μm, the working range is a severe limitation and usually a second actuator is required for coarse positioning. Furthermore, the electric field used for actuation cannot be generated with arbitrary precision and without noise. Thus, there is a trade-off between working range and resolution. *Step-wise actuators* are a solution to overcome these downsides. Instead of amplifying the piezoelectric element's motion using a mechanical structure, the stroke is used to perform small steps. Step-wise actuators have a virtually unlimited working range. Most step-wise actuators are hybrid actuators, so that the piezoelectric element can also be used in a scanning mode to create movements smaller than single steps.

Scanning actuators have the advantage of generating a smooth motion over the entire working range. Furthermore, due to the high force that can be generated by a piezoelectric actuator, very high accelerations and speeds are possible. Another advantage is that a rough knowledge of the actuator's position can be directly derived from the applied voltage. Homing is not required for scanning actuators and even open-loop movements are reliable to a certain extent. A downside is that a scanning actuator always returns to its initial position on power loss. Thus, microrobotic systems have to be designed in a way that this movement cannot result in a severe damage to the employed robots, tools, or objects. Furthermore, the nonlinear characteristics including hysteresis and creep are scaled with respect to the full working range. Thus, these effects have a significant influence and make sensor feedback necessary.

In contrast, step-wise actuators have opposed properties. Their working range is only restricted by the mechanical guides, usually in the centimeter range. They require a sensor because the executed steps are not sufficiently repeatable. Furthermore, dependent on the sensor, homing is required to find the initial location of the actuators. In addition to the large working ranges, major advantages of step-wise actuators are that they do not perform a significant movement if the power is lost and that hysteresis and creep are scaled with the step length.

The most commonly applied step-wise motion principle based on piezoelectric actuators is the stick–slip actuation principle [4]. This principle makes use of the high movement speed and acceleration of a piezoelectric actuator. A mobile part is connected to a piezoelectric actuator mechanically using a contact with a specific friction. If the piezoelectric actuator is deformed slowly, the friction forces the mobile part to perform the same motion. If the actuator is deformed rapidly, however, the inertia of the mobile part keeps it in place and the friction force is overcome generating a net displacement between actuator and mobile part. Alternating slow and rapid deformations into opposite directions generate a step-wise motion over long travel ranges.

Vision-Based Closed-Loop Control

On the macroscale, the location of the manipulated objects in a world coordinate system is either

known from the process design or can be measured by a variety of available sensors. To perform closed-loop positioning, a robot can move to this location with sufficient accuracy using its internal sensors. On the nanoscale, the position of objects has to be derived by processing images of the employed SEM (see section "Real-Time Vision Feedback"). SEMs, however, are designed to create highly magnified images and not to create highly accurate measurements. Thus, the resulting position is often time variant and nonlinear, so that the object's location in a world coordinate system cannot be accurately derived. In addition, the positioning itself has limited accuracy due to the size ratio between the robotic system and the handled objects. Commonly, a robot with a size of a few centimeters is used to manipulate an object with a precision of a few nanometers, that is, the robot is about six orders of magnitude larger than the required accuracy. Effects such as thermally- induced drift and limited mechanical stiffness lead to a nonlinear and time-variant relation between the position measured by a robot's internal sensors and the real position of the robot's tool. Thus, internal sensors are insufficient for reliable nanopositioning.

As known from the macroworld, the limitations of positioning based on internal sensors can be overcome by using the information generated by a visual sensor. Using vision feedback as a sensor for robotic manipulation is promising as it is a noncontact measurement and mimics human behavior. The initial technique, which is still used for most SEM-based positioning, extracts a position deviation from an image and then moves the robot either open loop or based on internal sensors to decrease this deviation. When the robot stops, another image is taken, again extracting and compensating for the deviation. This is called the *look-then-move* approach [10]. Two advantages of the approach are that low throughput rates of image processing are acceptable and that the entire setup can be handled in a quasi-static fashion neglecting dynamic aspects. The effectiveness, however, entirely depends of the accuracy of the visual sensor and the robot's open-loop control.

An alternative is to use the visual feedback as a low-level element in the motion control loop of a robot. This is called *visual servoing* [10]. Two challenges need to be solved to achieve effective visual servoing. First, the image processing needs to be fast. As the available processing power has significantly increased over the recent years, several image processing algorithms can be used for visual servoing. More complex algorithms, however, still require computation times on the order of seconds and do not allow for effective servoing. Second, the dynamic aspects of visual servoing need to be considered. The robot's movement as well as resulting vibrations leads to movement during image acquisition. This movement creates image distortions rendering most of the current image processing algorithms useless. Enabling an effective visual servoing, however, is a necessary step on the way of high-speed micro- and nanohandling viable for industrial application.

Closed-Loop Control

Based either on an internal sensor or on vision feedback, a closed-loop controller can be implemented. Closed-loop control on the micro- and nanoscale is significantly different from macroscale control [1]. Macroscale control mostly solves challenges arising from the dynamic behavior of systems, for example, inertia. On the micro- and nanoscale, these effects become negligible. As an example, a microrobot might carry a mass m of 30 g and can create a force F of approximately 0.02 N. Thus, its maximum acceleration a is:

$$a = F/m = 2/3 \left[\text{m/s}^2\right]$$

If it needs to accelerate to a velocity v of 500 μm/s, which is high for a micro- or nanooperation, it requires:

$$t = v/a = 0.375 \ [\text{ms}]$$

This is significantly faster than the update rate of any visual sensor. Thus, acceleration effects can be neglected. Therefore, for many applications and actuator types, a simple proportional controller is sufficient to perform an effective alignment.

R

AFM as a Nanohandling Robot

Since its development by Binnig, Quate, and Gerber [11], AFM has been increasingly recognized not only as a tool for imaging purposes, but also for manipulating at the micro-, nano-, and atomic scale. The main advantages of the AFM are its high resolution in the sub-nanometer range as well as its flexibility for operating under physiological conditions and in vacuum or liquid environments, depending on the specific application. In addition to its visualization capabilities, the AFM can be used for force spectroscopy and, with functionalized tips, for measuring electric, magnetic, or chemical properties. Moreover, the AFM tip can be used to manipulate objects or materials with nanometric precision and thus the AFM can also be seen as a nanohandling robot. AFM-based manipulation benefits from the fact that the AFM directly measures forces that are exerted during the manipulation and thus gives important feedback on the progress of the manipulation. Potential applications of AFM-based manipulations are manifold, ranging from the fabrication of nanoscale devices, reparation or modification of nanostructures to the characterization and handling of biological samples. Different types of manipulation can be conducted with the AFM including pushing, pulling, cutting, indenting, and structuring of a wide range of materials (Fig. 5).

Starting with unintended manipulations of Au-Clusters, reported 1987 by Baró et al. in [12], the AFM as a tool for manipulating single nanoentities became subject of increasing interest in the research community. A controlled manipulation of nanoscale objects by an AFM has been actively investigated over the last decade. Many publications exist describing the utilization of the AFM as an instrument for manipulating spherical nanoobjects such as gold nanoparticles as well as longish nanoobjects such as carbon nanotubes (CNTs), nanorods, or also biological objects such as DNA [2]. Figure 6 illustrates a successful manipulation of multiwalled CNTs that was conducted by the AFM tip. A commercially available dynamic-mode cantilever was used for both – the image acquisition as well as the manipulation. For imaging, conventional intermittent mode was used, whereas for manipulation the tip was brought into proximity with the surface and then moved laterally to mobilize the CNTs.

To perform a controlled movement of a nanoscopic object, it has to be localized and identified on the substrate first. A major drawback of AFM-based manipulation is the lack of an additional, visual sensor. Therefore, before any manipulation is conducted, the AFM usually has to be utilized to obtain a detailed image of the scenery in order to plan the execution of the manipulation. This imaging scan is most often performed in dynamic mode (intermittent or noncontact) in order to avoid any damage or unintended manipulation of the sample. For the manipulation, the AFM tip is brought into proximity with the object and moved – either in dynamic or contact mode, but without any feedback to control the z-direction – toward a predefined position. As a result, the object is then pushed forward by the repulsive forces. After a manipulation step, reimaging of the area of interest often becomes necessary to reveal the result of the manipulation. The disadvantage of this "look and manipulate" scheme is that the image acquisition process usually needs several minutes depending on the required quality. Moreover, the manipulation itself cannot be monitored by means other than the measured forces exerted by the cantilever and thus have to be performed in a "blind" way. Unfortunately, due to a variety of uncertainties, the results of the manipulation are often not satisfying and require frequent trial and error experiments. Real-time imaging would greatly help to make the manipulation process faster, more robust, and accurate. However, optical microscopy cannot be used as an additional real-time capable imaging method because its resolution does not fit the requirements for the manipulation of objects with nanometric dimensions. An alternative is given by the integration of the AFM into the vacuum chamber of an SEM to allow for real-time observation of the manipulated object and the AFM probe. However, this entails the disadvantage of restriction to vacuum and to some extent conductive sample materials, which is not feasible for many applications.

Robot-Based Automation on the Nanoscale, Fig. 5 Schematic view of manipulation tasks that can be performed by AFM-based robotic systems [1, 2]

Moving/Pulling

Cutting

Structuring/Lithography

Scratching/Indenting
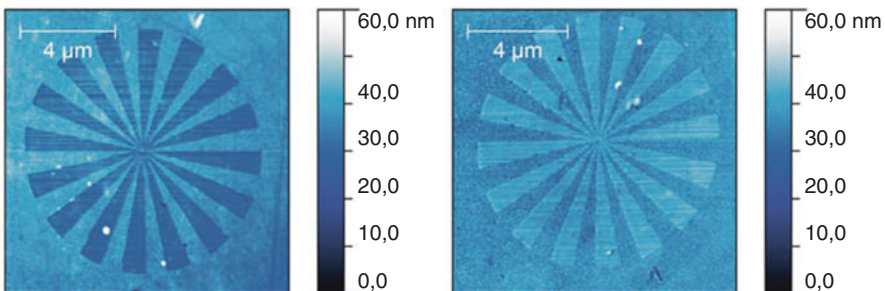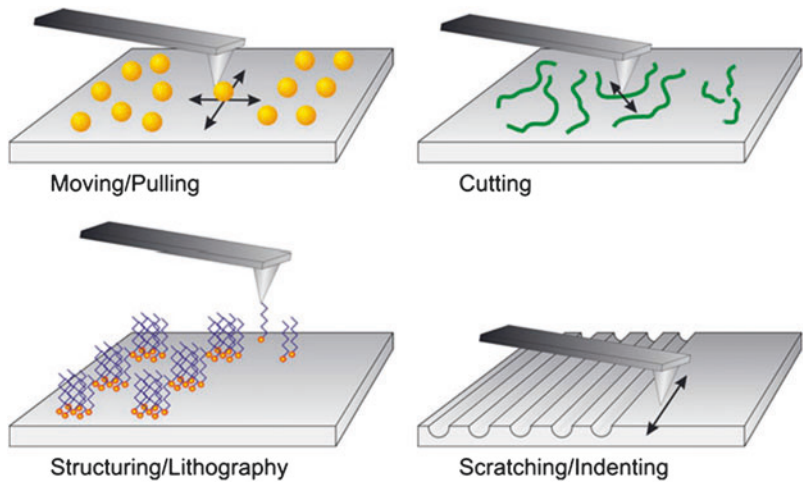
Robot-Based Automation on the Nanoscale, Fig. 6 AFM topography star with diameter of 10 μm, fabricated on a gold layer (*left*) using mechanical AFM-lithography and the replica of this structure fabricated on polypropylene (*PP*) using the microinjection molding technique (*right*)

Besides its ability to manipulate nanoscale objects, the AFM tip can also be used to modify surfaces with nanometric precision or to change an object's shape, for example, by scratching, indenting, cutting, and dissecting. For nanomachining purposes, the AFM tip can be exploited as a nanohandling tool in order to act as, for example, a milling cutter, nanoscalpel, or nanoindenter. However, the type of interaction between AFM tip and substrate is not only mechanical, but may also involve chemical, thermal, or electrical impact. Over the last two decades, a large number of promising nanolithographic technologies have evolved and are still emerging that are based on such AFM-based nanomanipulation of surfaces. These techniques can be used among others for mask-free lithography on the nanoscale and provide a promising alternative to conventional technologies. Using AFM-based surface modification, material can either be removed (e.g., through direct mechanical scratching), chemically changed (e.g., resist exposure techniques or oxidation), or deposited (e.g., dip-pen lithography). A detailed review of such SPM-based lithography can be found [13]. An example for a novel process chain for small series production of nanostructures that is based on AFM-based mechanical lithography, which is the most basic type of AFM lithography, was introduced [14]. Figure 7 depicts how the mechanical AFM-lithography is utilized to fabricate a nanostructure that was subsequently replicated several times by means of the microinjection molding technique.

Up to now, the AFM as a robot for the nanomanipulation can only be found in

R

Robot-Based Automation on the Nanoscale, Fig. 7 Multiwalled carbon nanotubes on HOPG substrate manipulated by the AFM. When moving the AFM tip laterally, the height control loop was turned off in order to avoid slipping of the AFM tip over the CNT. It could be observed that, depending on the parameters chosen for manipulation (tip velocity, damping amplitude), cutting as well as pushing and pulling of the CNTs can be accomplished

laboratories for prototypical issues. For industry, the serial AFM-based approach is still not reliable and fast enough to be successfully applied for productive use. To increase throughput and reliability of this sequential process, automation can be considered as a desirable goal. To reach this goal, however, a large number of challenging prerequisites have to be fulfilled.

Besides the lack of real-time imaging of the manipulation process discussed above, spatial uncertainties constitute an import error source. Although positioning inaccuracies caused by creep and hysteresis of piezoactuated scanning stages can nowadays most often be compensated by closed-loop operation, thermal drift of the AFM components lead to spatial positioning errors that are challenging to counteract. Solutions to overcome this problem have been addressed [1, 2, 15]. Moreover, sticking effects also play a crucial role in AFM-based nanohandling and therefore, substrate, cantilever, and tip material have to be carefully chosen in order to minimize these effects. A promising strategy that is pursued by several research groups aims at the precise modeling of the nanohandling task. To compensate for the uncertainty introduced by the inevitable "blind" manipulation, a model of the manipulation process is utilized enabling a manipulation in open-loop mode. Having a valid model of the nanomanipulation, including all relevant interactions between nanoobject, tip, and substrate, it is possible to calculate the expected position of the nanoobject during manipulation in real time. Using such a model for an augmented reality system, where a virtual reality is augmented by the calculated nanoworld situation, enables an additional visual feedback. However, this approach requires exact knowledge of the nanomanipulation phenomena, which is not yet available in the current state of nanosciences.

Application Example: Handling of CNT

Nanomaterials such as different kinds of nanowires and especially carbon nanotubes (CNTs) [16] have unique physical properties and can thus improve and enhance existing microsystems in a multitude of applications. Two main application areas have been identified where in particular the integration of CNTs into known microsystems is of high impact: CNT-enhanced AFM probes [17] for the analysis of high-aspect ratio structures and CNT-interconnects in microchip technology [18]. Even though parallel fabrication techniques will be required for a future bulk production of such devices, enabling technologies for the reliable manipulation of individual nanostructures and their prototypic integration into microsystems have to be developed. Automated robot-based assembly techniques are very promising to implement a fast

and systematic prototyping of CNT-based devices. In the following section, a microgripper-based strategy for the automated assembly of CNT-enhanced AFM probes is presented.

## Nanorobotic Strategies for Microgripper-Based Handling of CNTs

In prior experiments, electrothermally actuated microgrippers have already been used for teleoperated pick-and-place handling of individual CNTs [19]. The CNTs used for the experiments have a typical diameter of 200 nm and are grown vertically aligned on a substrate with a spacing of 50 μm between nanotubes. In order to automate the microgripper-based handling of CNTs, special nanorobotic strategies have been developed. Most important are methods for the z-approach of gripper and CNT and for the reliable gripping and placing of CNTs that lead to a well-defined system state.

Methods for z-Approach of Microgripper and CNT An AMNS has been developed that can be integrated into the SEM [2], so that visual feedback from the microscope is available during the nanohandling tasks in addition to the integrated sensors of the positioning systems. Image processing algorithms for object recognition and tracking have been developed (compare section "Real-Time Vision Feedback") providing the x, y-position of gripper and nanotube. This two-dimensional position information can be used to automate the approach in x,y-direction. However,

to receive information on the gripper's and CNT's relative z-position, additional contact sensors or three-dimensional systems are required. Such systems cause additional costs and reduce the available space within the SEM's vacuum chamber [20].

A novel so-called shadow-based depth detection method has been developed that, together with the focus information of the SEM, facilitates the automated z-approach of microgripper and CNT in any available SEM. First of all, the "depth-from-focus" technique [20] is used to coarsely align both objects in z-direction down to 50 μm. Then, for the fine approach of gripper and CNT, a shadow can be detected as soon as the CNT is located exactly between the gripper jaws. The gripper jaws block the secondary electrons from the detector. Figure 8 shows SEM images illustrating the shadow-based depth detection method. The clearly detectable shadow on the CNT in Fig. 8b is used as a stop criterion for the z-approach.

Gripping and Removing CNTs from the Growth Substrate Different gripping techniques for the reliable picking and removing of individual CNTs that are grown on a silicon substrate by chemical vapor deposition (CVD) have been discussed in [21]. The so-called shear gripping technique has been identified as the most reliable approach, requiring smaller gripping forces compared to tensile gripping. The microgripper is closed around the nanotube and
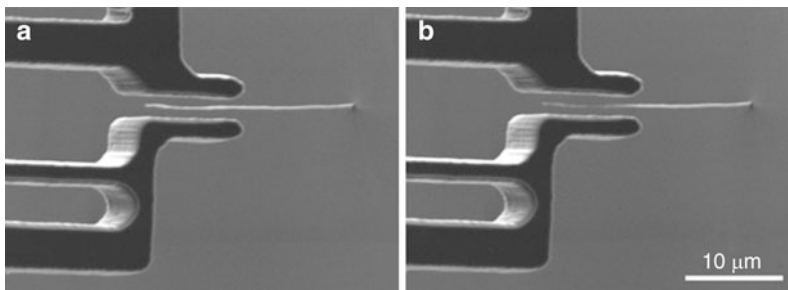
Robot-Based Automation on the Nanoscale, Fig. 8 Shadow-based depth detection for the alignment of microgripper and CNT in z-direction. (**a**) CNT is located

below the gripper jaws, no shadow is detectable. (**b**) CNT is between the gripper jaws, a shadow is clearly visible

moved sideways to break the CNT off the substrate. After that, the gripped CNT can be transferred to the target structure.

Placing and Releasing CNTs on the Target Structure A novel handling strategy for placing and releasing CNTs onto the target structure has been developed. This strategy relies on the adhesion forces between CNT and microgripper jaws and thus avoids the need of additional joining techniques. After reopening the microgripper, the nanotube can be oriented in four different configurations. Either the CNT is connected by a line contact to one gripper jaw (see Fig. 9a, b) or the CNT is connected by two-point contacts to the gripper jaws (see Fig. 9c, d).

In case of the two-point contacts, the adhesion forces are smaller compared to the line contacts. For this reason, the two-point contact is preferred in order to easily place and release the CNT on the target structure. In addition, obtaining a well-defined system state after reopening the gripper is required to allow full automation of the placing and releasing process. For the automated microgripper-based mounting of an individual CNT onto the tip of an AFM probe, the strategy shown in Fig. 10 has been developed.

The closed microgripper holding the CNT (Fig. 10a) is approached to the tip of an AFM probe until a bending of the nanotube can be detected (Fig. 10b). This bending is detected automatically and used as the stop criterion of the z-approach. At the same time, the exerted force assures that CNT and gripper jaws are forming a two-point contact with defined orientation after reopening the microgripper (Fig. 10c). After this, the CNT is moved toward the AFM tip until a line contact is achieved (Fig. 10d). Because the adhesion forces of the line contact between CNT and AFM tip are much higher than the forces of the two-point contacts between CNT and gripper jaws, the nanotube can be easily released from the gripper jaws by retracting the microgripper (Fig. 10e–f).

Automated Assembly of CNT-Enhanced AFM Probes

The position control techniques described in section "Position Control Inside SEM" have been
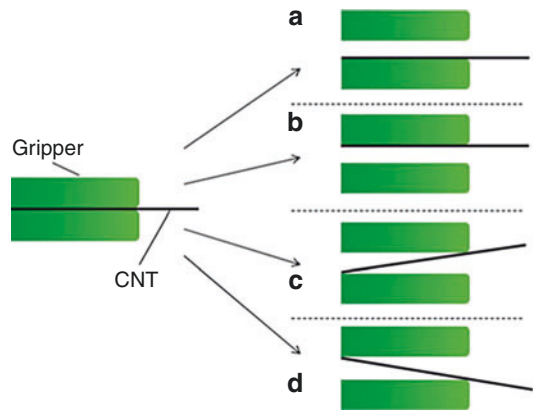
Robot-Based Automation on the Nanoscale, Fig. 9 The four possible configurations after reopening the microgripper. (**a**, **b**) Line contact between CNT and gripper jaw. (**c**, **d**) Two-point contact between CNT and gripper jaws
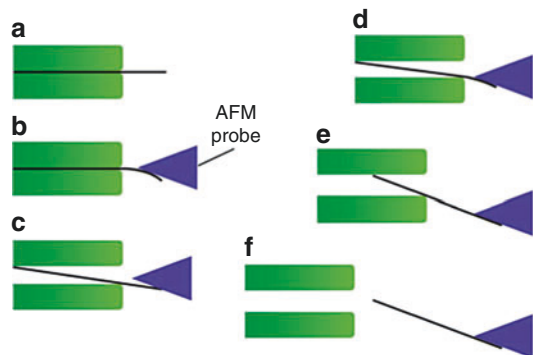
Robot-Based Automation on the Nanoscale, Fig. 10 Nanorobotic handling strategy for automated placing a CNT onto the tip of an AFM probe and releasing it from the microgripper

used to fully automate the nanohandling task of mounting a CNT onto the tip of an AFM probe. For this purpose, the whole automation sequence has been divided into three subsequences: The initialization process, the gripping and removing process, and the process of placing and releasing the CNT. In the following, these three subsequences are described in detail with respect to automation purposes.

Initialization Sequence The first step of an automated robotic micro-nano-integration is the

initialization. For this purpose, all actuators are moved to a well-known starting position and the coordinate system of the internal position sensors are mapped to the coordinates of the SEM image. Then, the electrothermal microgripper mounted to the mobile microrobots is calibrated. A pattern of the microgripper is recorded that allows for tracking the position of the gripper by applying cross-correlation algorithms to the SEM images (compare section "Real-Time Vision Feedback"). After this, the CNT substrate is scanned automatically to index all CNTs. For each CNT, length and orientation are checked as a quality criterion and only the positions of suitable CNTs are stored to the database. Finally, the AFM probes are indexed by saving the exact position of the probes' tips to the database.

Picking Sequence As described in section "Gripping and Removing CNTs from the Growth Substrate", the shear gripping technique is used to grip and remove an individual CNT from the sample substrate. Figure 11 shows SEM images of the gripping sequence illustrating the process of breaking the CNT off the substrate. First, so-called coarse and fine positioning steps are performed placing the CNT between the gripper jaws (see Fig. 11a). Second, the microgripper is actuated and the CNT is detached by moving the gripper sideways (see Fig. 11b). Third, the gripped CNT can be transferred to the target structure.

Placing Sequence For the automated placing of the CNT onto the tip of an AFM probe, the nanorobotic strategy described in section "Placing and Releasing CNTs on the Target Structure" is used. Figure 12 shows SEM images of the automated placing sequence. After the coarse and fine approach between CNT and AFM probe has been realized in x,y-direction (see Fig. 12a), the final z-approach is performed until a bending of the CNT is detected (see Fig. 12b). Then, the gripper is reopened and a two-point contact between CNT and gripper jaws as well as a line contact between CNT and AFM probe tip is established (see Fig. 12c). Finally, the CNT can be easily released from the gripper jaws and placed onto the tip of the AFM probe by retracting the microgripper (see Fig. 12d).

Application Example: Handling of Biological Materials

Current research areas in molecular and cell biology, medicine, and process sensor technology (sensor systems monitoring various processes in industry and science) often require advances in the nanoengineering technologies to look for molecular phenomena with the highest possible resolution in order to enlighten the "black-box" which still shades most metabolic reactions. The aim of present work, starting from current medical problems, is to propose novel studies especially in molecular and cellular microbiology as well as virology with direct applications in medicine and industry, closing the bottom-up cycle. Thanks to advances in micro- and nanofabrication and micro- and nanorobotics, robotic systems which offer the possibility of characterizing and
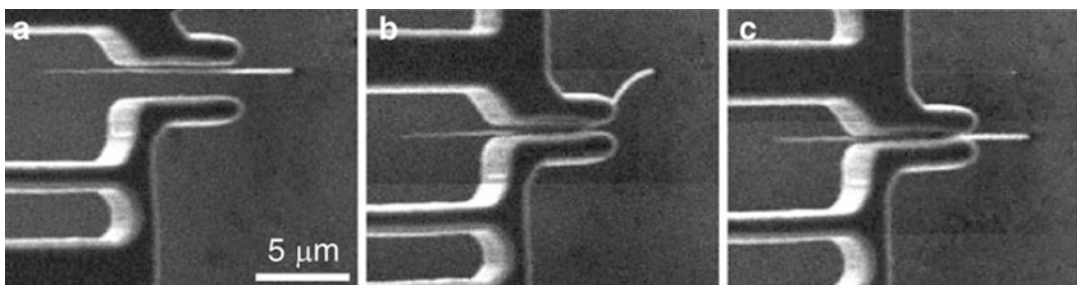

Robot-Based Automation on the Nanoscale, Fig. 11 SEM images of the automated gripping sequence. (**a**) Coarse and fine alignment of gripper of CNT. (**b**) Removing the CNT from the substrate by shear gripping. (**c**) Closed gripper holding the detached CNT

Robot-Based Automation on the Nanoscale, Fig. 12 SEM images of the automated assembly. (**a**) Coarse and fine alignment of gripped CNT and AFM tip. (**b**) Final z-approach until a bending of the CNT is detected. (**c**) Creating a two-point contact between CNT and gripper jaws and a line contact between CNT and AFM probe. (**d**) Releasing the CNT by retracting the gripper
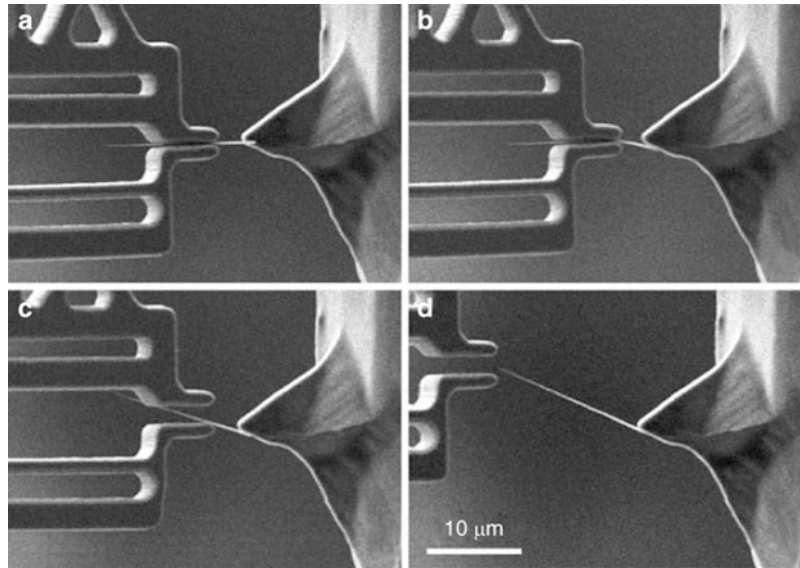
manipulating biological objects can be developed today. The use of such systems enable new studies of single cell phenomena with nanometer resolution (e.g., studying the local mechanical and electrical properties of single bacteria for characterization and evaluation of the resistance to antibiotics or to the spread of infections) up to complete cell compounds (e.g., studying the mechanical properties of bacterial biofilms) and deepen the appreciation of the processes at the nanoscale. The following section will give a brief overview of the handling of biological objects using automated AFM-based approaches.

Manipulation and Handling of Biomaterials Using Atomic Force Microscopy

Section "AFM as a Nanohandling Robot" has already shown the basics of the AFM for the automated handling of nanoscale materials. In this subsection, a brief overview of the differences for biomaterials is given.

For characterization or manipulation, samples have to be immobilized on an adequate substrate. A common method is the immobilization on freshly cleaved mica. Mica has a hydrophilic, charged surface that binds proteins and other biomolecules easily. Occasionally, mica is treated with buffer systems, which change the charging

of the surface. Gold substrates can also be used if coated with protein reactive monolayers.

The characteristics of the used cantilevers are determined by material, tip shape, and mechanical parameters such as spring constant, resonant frequency, and Q factor. For biological applications, cantilevers are usually made of silicon or silicon nitride. These materials are chemically inert and can be doped to induce electrical charges. Soft cantilevers ($C < 1$ N/m) are used for the contact mode and hard cantilevers ($C \approx 42$ N/m) are used for the tapping mode. If operated in liquids, the pH value and the presence of electrolytes influence the forces between tip and sample. Although a smaller tip radius results in more accurate images, it may be advantageous for soft biological samples, to use a duller tip reducing the risk of accidentally damaging the sample.

Compared to other methods (e.g., optical tweezers, dielectrophoresis) the AFM has some major advantages for the handling of biomaterials. The most important points are [1, 2, 22]:

- The AFM can be used in liquids, in vacuum, and at ambient conditions meaning that biological samples can be characterized and manipulated at their necessary physiological conditions. Furthermore, living samples and

their reaction to mechanical stimulation can be studied.

- The AFM can be used for imaging as well as for manipulation nearly at the same time.
- The possible resolution lies in the sub-nanometer range.
- Three-dimensional topological information can be obtained.
- No special preparation methods are necessary.

However, compared to the other methods AFM-based manipulation entails major drawbacks and limitations. Knowing these limitations is important to choose the right method or tool to handle or characterize a specific sample. The most important limitations are:

- A three-dimensional force measurement is not yet possible.
- The scanning speed is generally too low for real-time mapping.
- The scanning range for high-speed AFMs is still very low (about 3×3 μm).

Examples of Automated Handling of Biomaterials with an AFM

In the following sections, current work on the field of handling, characterization, and manipulation of biomaterials is described. Current research mainly deals with the development of fully automated solutions for handling and manipulation of DNA, for solving of packaging problems in the future of nanoelectronics, and for structuring of biosensors for medical and forensic use.

Handling and Manipulation of DNA In accordance with the progressing miniaturization of electric circuits, new ways to realize smaller channel widths are needed. To overcome the limitations of silicon-based processing, the use of objects like CNTs or biomolecules like DNA is a promising solution. Nanowires produced via the metallization of DNA are promising candidates for nanoelectronic devices of future generations. DNA can be considered a basic building block for nanostructure fabrication because it provides unique self-recognition properties.

DNA constitutes an ideal template for the organization of metallic and semiconductor particles into wire-like assemblies. Nanoscopic (thickness below 10 nm) and regular wire-like metallic structures can be efficiently produced because DNA molecules have a diameter of only two nanometers and a length in the micrometer range. Moreover, both the ends of DNA molecules and the surface of solid substrates can be functionalized through a variety of methods to create specific links between DNA and substrate, allowing for the integration of DNA molecules into specific, predefined sites of microstructured electronic circuits. By exploiting the specific recognition of complementary nucleotide sequences, complex structures made of DNA can be fabricated and metalized at a later stage. Packaging problems of next-generation nanoelectronics technologies, (e.g., single-electron transistors, quantum automata, molecular electronics, etc.) might also be addressed by using DNA strands.

Recently, many researchers have proven the feasibility of DNA manipulation by an AFM tip, primarily in liquid conditions. The next step toward high-throughput fabrication is automation of DNA handling. Figure 13 show two alternative immobilization procedures of DNA, which offer the potential of automated DNA manipulation in dry ambient conditions. Figure 13a shows a topographic scan of a DNA strand. The height of the DNA strand is about 1.5 nm. Figure 13b shows the same DNA strand after several manipulations by the AFM cantilever tip. The DNA was manipulated and successfully moved about 25–100 nm. Figure 13c, d shows an experiment to demonstrate the transport of DNA. Two separate DNA strands are shown. The upper DNA strand (Fig. 13c) was moved by several AFM strokes up to 300 nm. As a result, this DNA strand was removed from its location (Fig. 13d), without affecting the position of the lower DNA.

The presented methods for the handling of biomaterials show the potential for different usage scenarios in industry and scientific research. In case of the handling of DNA, the experiments demonstrate that a DNA manipulation and handling can be performed in dry ambient conditions.

R
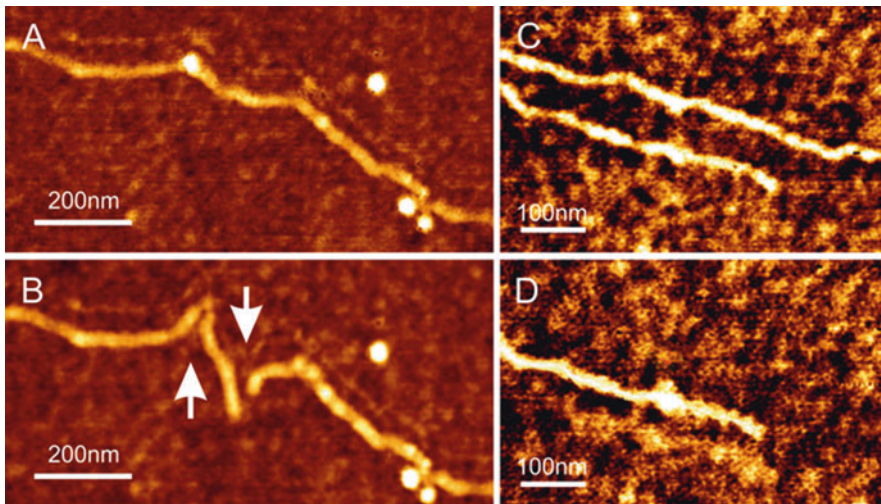
Robot-Based Automation on the Nanoscale, Fig. 13 (**a**, **b**) DNA immobilized on silicon without surface modifications (**a**) and the same DNA after the manipulation process (**b**). On the *left* side of the strand, two short distance movements from *bottom* to *top* were implemented. As a result, the corresponding DNA parts have been pushed about 25 nm. On the *right* side, two long distance movements were implemented, resulting in DNA displacement by 100 nm. (**c**, **d**) DNAs immobilized on silicon without surface modifications (**a**) and the same substrate with *upper* DNA strand pushed away on the surface by the AFM tip, without affecting the *lower* strand (**b**)

Automated Structuring of Biological Materials for Biosensors Biosensors consist of a sensitive biological component, a transducer, or detector element, and associated electronics or signal processors. An accurate and well-defined preparation of the sensitive components made, for example, of microorganisms, cell receptors, proteins, enzymes, anti-bodies, or nucleic acids is crucial part of the sensor design. Commonly used methods are microstamping and microprinting, both of which are unable to decrease the structures of the biological component down to the nanometer range. However, higher density is necessary to increase the measurement accuracy and the number of different, simultaneously useable biocomponents. In this section, current work on automated design of high-density sensitive components using AFM-based lithography is introduced.

The compensation of spatial uncertainties plays an important role in automation of AFM-based handling. Although the effects of hysteresis and creep of the PZT-based scanner can be reduced by closed-loop control, spatial uncertainties caused by thermal drift are crucial and less straightforward to deal with. Especially in life science applications, where the AFM is often operated in humid environments, the effect becomes strong and often prevents successful nanohandling, let alone automation of the handling task. To achieve reliable results, the recently developed method for real-time drift compensation described in section "AFM as a Nanohandling Robot" and in [1, 15] was applied, both for the presented manipulation of DNA as well as for the AFM-based nanotooling described below.

As a first step toward the automated fabrication of biosensor components, rectangular areas were successfully structured into an APTES monolayer on mica (Fig. 14). To further investigate the correlation between the forces applied during manipulation and the resulting machining depths, a series of experiments was performed to scratch multiple areas.

By using the AFM, an automated structuring of biological materials for biosensors can be done.
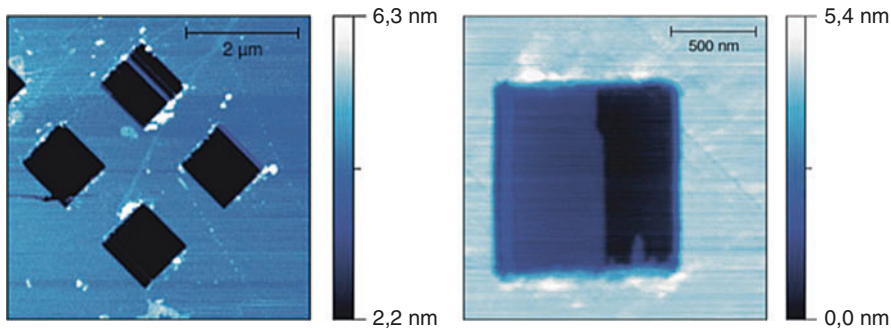
Robot-Based Automation on the Nanoscale, Fig. 14 Height image of 1×1 μm area scratched into APTES monolayer by AFM machining. The two visible machining depths arise from different force set points used during the processing (*left*) and AFM height image of a series of 1×1 μm sized scratched areas (*right*)

The automation of the necessary processes will also lead to a feasible industrial use, as well as for clinical and research laboratories. The size of the resulting structures correlates with the resolution of the commonly used read-out techniques such as fluorescence microscopy or surface plasmon resonance spectroscopy.

Future Trends

In the future, closed-loop motion control in SEMs needs to be changed from the quasi-static look-then-move-approach to an effective visual servoing in order to implement high-speed micro- and nanohandling viable for industrial application. Furthermore, to allow for a more effective avoidance of both collisions and vibrations, the commonly implemented position control needs to be extended to trajectory control so that the robot moves with a predetermined velocity, acceleration and jerk at all times. Considering visual feedback, an important future task is to enable more reliable and faster feedback especially in case of fast moving objects, because of the artifacts and distortions generated by the object movement. Also speed, accuracy, and resolution of the z-position estimation still is a topic where significant improvements can be made.

The AFM has emerged as an important tool for the fabrication of nanoscale structures, as well as the handling and manipulation of biological objects and for the manipulation of nanoscale objects. To broaden the applicability of those techniques for productive use, it will become necessary to increase throughput significantly. For the lithographic techniques used to structure surfaces, the usage of multiple AFM tips in parallel will greatly enhance throughput. To speed up the assembly of individual nanoentities, achieving a high level of automation is an important research goal for the future. Further miniaturization of gripping tools can facilitate the nanorobotic pick-and-place handling of CNTs with smaller diameters. The optimization of the presented automated assembly sequence of CNT-enhanced AFM probes will lead to a high-throughput prototyping of CNT-based devices, which is an important pre-stage toward a direct integration of CNTs into microstructures by CMOS-compatible CVD-based fabrication techniques.

Cross-References

▶ AFM
▶ AFM in Liquids
▶ AFM Probes
▶ Atomic Force Microscopy
▶ Basic MEMS Actuators
▶ BioPatterning
▶ Biosensors

R

References

1. Sattler, K.D.: Handbook of Nanophysics – Nanomedicine and Nanorobotics. Taylor & Francis/CRC Press, Boca Raton (2010)
2. Fatikow, S., Wich, T., Dahmen, C., Jasper, D., Stolle, C., Eichhorn, V., Hagemann, S., Weigel-Jech, M.: Nanohandling robot cells. In: Sattler, K.D. (ed.) Handbook of Nanophysics – Nanomedicine and Nanorobotics, pp. 47.1–47.31. Taylor & Francis/CRC Press, Boca Raton (2010)
3. Fatikow, S. (ed.): Automated Nanohandling by Microrobots. Springer, London (2008)
4. Fatikow, S., Rembold, U.: Microsystem Technology and Microrobotics. Springer, Berlin/Heidelberg/New York (1997)
5. Fatikow, S.: Mikroroboter und Mikromontage (in german). B.G. Teubner, Stuttgart/Leipzig (2000)
6. Ferreira, A., Mavroidis, C.: Virtual reality and haptics for nanorobotics. IEEE Robot. Autom. Mag. **13**, 78–92 (2006)
7. Fatikow, S., Dahmen, C., Wortmann, T., Tunnell, R.: Visual feedback methods for nanohandling automation. Int. J. Inf. Acquis. **6**, 159–169 (2009)
8. Sievers, T., Fatikow, S.: Real-time object tracking for the robot-based nanohandling in a scanning electron microscope. J. Micromechatron. **3**(3–4), 267–284 (2006). Special Issue on Micro/Nanohandling
9. Kratochvil, B.E., Dong, L.X., Nelson, B.J.: Real-time rigid-body visual tracking in a scanning electron microscope. Int. J. Robot. Res. **28**, 498–511 (2009)
10. Hutchinson, S., Hager, G.D., Corke, P.I.: A tutorial on visual servo control. IEEE Trans. Robot. Autom. **12**, 651–670 (1996)
11. Binnig, G., Quate, C.F., Gerber, C.: Atomic force microscope. Phys. Rev. Lett. **56**, 930–933 (1986)
12. Baró, A.M., Bartolome, A., Vazquez, L., García, N., Reifenberger, R., Choi, E., Andres, R.P.: Direct imaging of 13-Å-diam Au clusters using scanning tunneling microscopy. Appl. Phys. Lett. **51**, 1594–1596 (1987)
13. Tseng, A.A., Notargiacomo, A., Chen, T.P.: Nanofabrication by scanning probe microscope lithography: a review. J. Vac. Sci. Technol. B **23**, 877–894 (2005)
14. Brousseau, E., Krohs, F., Caillaud, E., Dimov, S., Gibaru, O., Fatikow, S.: Development of a novel process chain based on atomic force microscopy scratching for small and medium series production of polymer nanostructured components. ASME J. Manuf. Sci. Eng. **132**, 030901 (2010)
15. Krohs, F., Onal, C., Sitti, M., Fatikow, S.: Towards automated nanoassembly with the atomic force microscope: a versatile drift compensation procedure. ASME J. Dyn. Syst. Meas. Control **131**, 061106 (2009)
16. Iijima, S.: Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991)
17. Dai, H., Hafner, J.H., Rinzler, A.G., Colbert, D.T., Smalley, R.E.: Nanotubes as nanoprobes in scanning probe microscopy. Nature **384**, 147–150 (1996)
18. Srivastava, N., Li, H., Kreupl, F., Banerjee, K.: On the applicability of single-walled carbon nanotubes as VLSI interconnects. IEEE Trans. Nanotechnol. **8**, 542–559 (2009)
19. Sardan, O., Eichhorn, V., Petersen, D.H., Fatikow, S., Sigmund, O., Bøggild, P.: Rapid prototyping of nanotube-based devices using topology-optimized microgrippers. Nanotechnology **19**, 495503 (2008)

20. Eichhorn, V., Fatikow, S., Wich, T., Dahmen, C., Sievers, T., Andersen, K.N., Carlson, K., Bøggild, P.: Depth-detection methods for microgripper based CNT manipulation in a scanning electron microscope. J. Micro-Nano Mechatron. **4**, 27–36 (2008)
21. Andersen, K.N., Petersen, D.H., Carlson, K., Mølhave, K., Sardan, O., Horsewell, A., Eichhorn, V., Fatikow, S., Bøggild, P.: Multimodal electrothermal silicon microgrippers for nanotube manipulation. IEEE Trans. Nanotechnol. **8**, 76–85 (2009)
22. Castillo, J., Dimaki, M., Svendsen, W.E.: Manipulation of biological samples using micro and nano techniques. Integr. Biol. **1**, 30–42 (2009)

Robotic Insects

► Insect Flight and Micro Air Vehicles (MAVs)

Robotic Platforms for Nanomanipulations

► Nanorobotics for Bioengineering and Medical Interventions

Role of Cavitation Bubble Dynamics on Nanoparticle Generation

► Mechanisms of Nanoparticle Generation by Laser Ablation in Liquids

Role of Plasma Confinement on Nanoparticle Generation

► Mechanisms of Nanoparticle Generation by Laser Ablation in Liquids

Rolled-Up Nanostructure

► Nanorobotics for NEMS Using Helical Nanostructures

Rolled-Up Nanotechnology

► Electrical Impedance Tomography for Single-Cell Imaging

Rose Petal Effect

Bharat Bhushan[1] and Michael Nosonovsky[2]
[1]Nanoprobe Laboratory for Bio- and Nanotechnology and Biomimetics, The Ohio State University, Columbus, OH, USA
[2]Department of Mechanical Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Synonyms

Petal effect

Definition

Rose petal effect is the ability of certain rough surfaces to have a high contact angle with water simultaneously with high adhesion (large contact angle hysteresis) with water.

Theoretical

Recent experimental findings and theoretical analyses made it clear that the early Wenzel [1] and Cassie and Baxter [2] models do not explain the complexity of interactions during wetting of a rough surface which can follow several different scenarios. As a result, there are several modes of wetting of a rough surface, and therefore, wetting cannot be characterized by a single number such as the contact angle (CA). Contact angle hysteresis (the difference between the advancing and receding CA) is another important parameter which characterizes wetting (Fig. 1).

Wetting of a solid is dependent upon the adhesion of water molecules to the solid. On the one

R
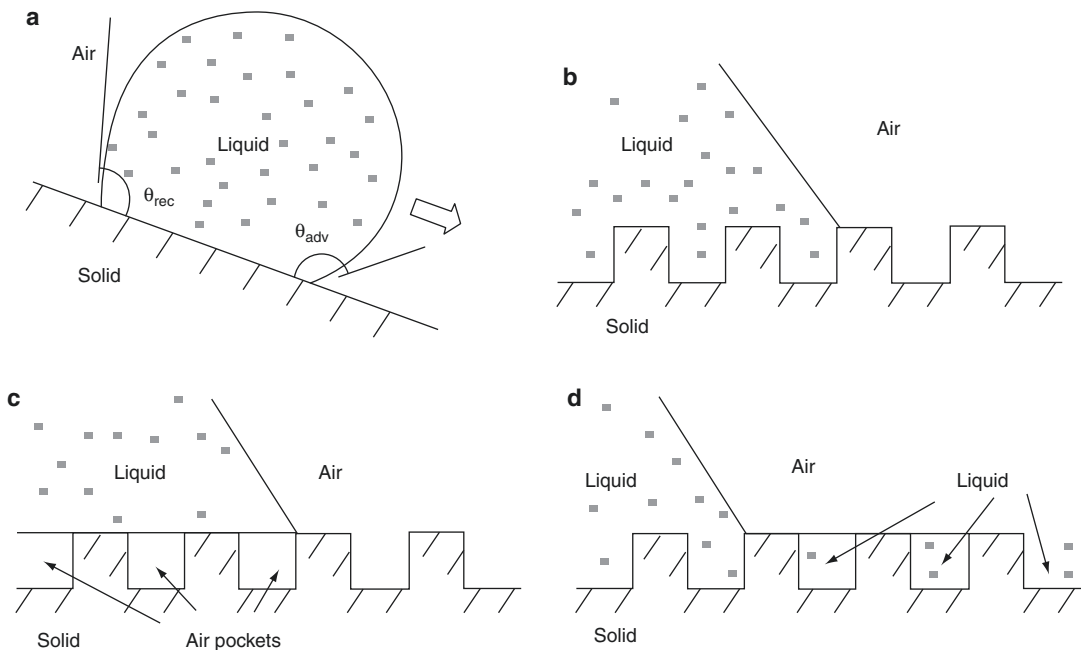
Rose Petal Effect, Fig. 1 (**a**) Schematics of a droplet on a tilted substrate showing advancing (θ_{adv}) and receding (θ_{rec}) contact angles. The difference between these angles constitutes the contact angle hysteresis. Configurations

described by (**b**) the Wenzel equation for the homogeneous interface, (**c**) Cassie–Baxter equation for the composite interface with air pockets, and (**d**) the Cassie equation for the homogeneous interface

hand, a high CA is a sign of low liquid–solid adhesion. On the other hand, low CA hysteresis is a sign of low liquid–solid adhesion as well. It is now widely believed that a surface can be superhydrophobic and at the same time strongly adhesive to water [3–9]. The so-called petal effect is exhibited by a surface that has a high CA, but also a large CA hysteresis and strong adhesion to water.
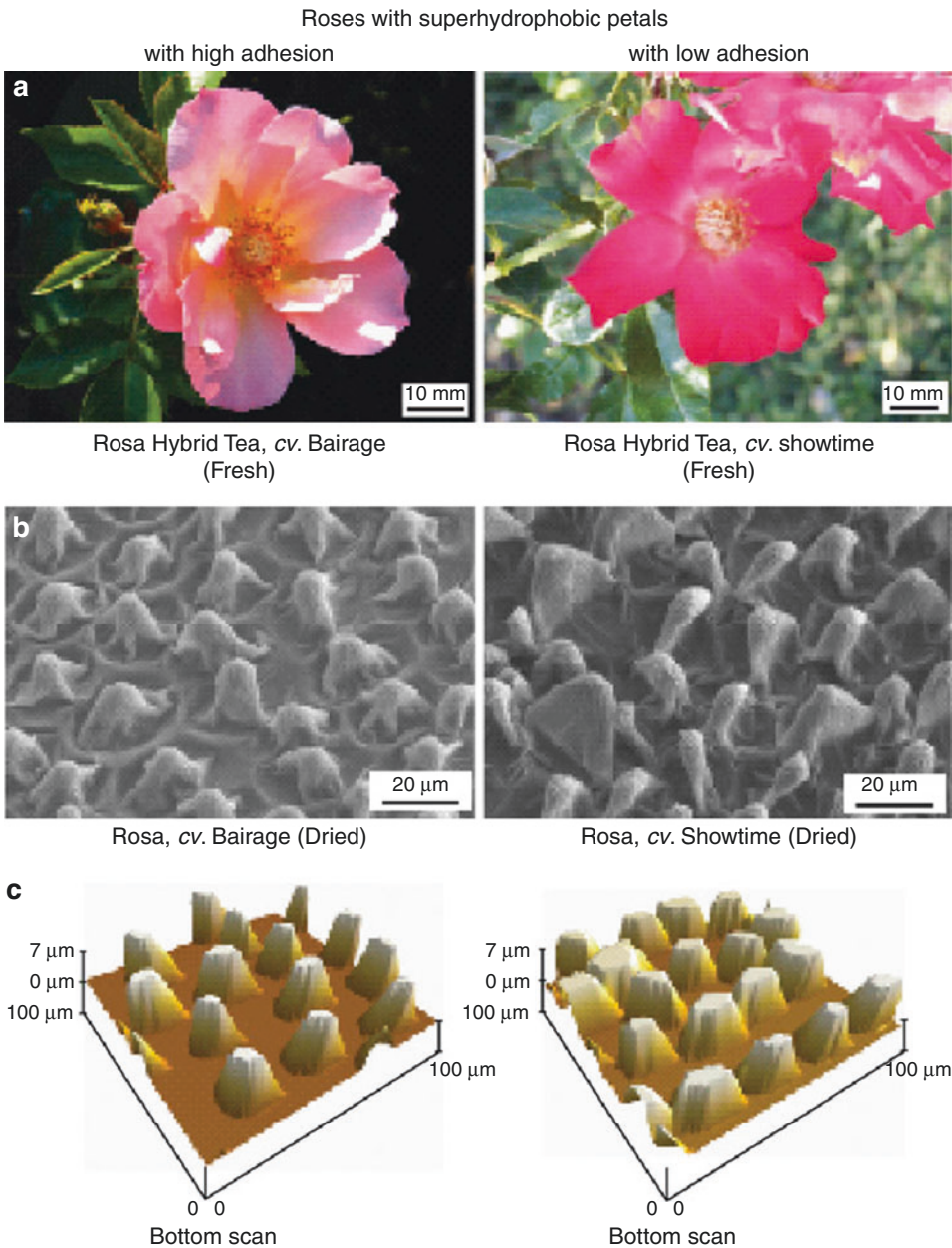
Main Findings

Plant leaves and petals provide an example of surfaces with high CA, and high and low CA hysteresis. Bhushan and Her [3] studied two kinds of superhydrophobic rose petals: (1) Rosa Hybrid Tea, cv. Bairage and (2) Rosa Hybrid Tea, cv. Showtime, referred to as Rosa, cv. Bairage and Rosa, cv. Showtime, respectively. Figure 2 shows optical micrographs, scanning electron microscopy (SEM) images, and atomic force microscope

(AFM) surface height maps of two rose petals. Figure 3 shows a sessile and a suspending water droplet on Rosa, cv. Bairage, demonstrating that it can simultaneously have high CA, high adhesion, and high CA hysteresis.

The surface roughness of the two rose petals was measured with the AFM, and the results for the peak-to-base height of bumps, the mid-width, peak radius, and bump density are summarized in Table 1. The data indicates that the low adhesion specimen (Rosa, cv. Showtime) has higher density and height of the bumps, indicating that the penetration of water between the microbumps is less likely. Wetting of a rough surface with a single level of hierarchy of roughness details can follow several scenarios (Table 2).

For a hierarchical structure with small bumps on top of larger bumps, a larger number of scenarios is available, and they are summarized in Table 3 and Fig. 4. Water can penetrate either into the micro- or nanostructure, or into both. In addition, the micro- or nanostructure can be

Roses with superhydrophobic petals

with high adhesion / with low adhesion

Rosa Hybrid Tea, *cv*. Bairage (Fresh) / Rosa Hybrid Tea, *cv*. showtime (Fresh)

Rosa, *cv*. Bairage (Dried) / Rosa, *cv*. Showtime (Dried)

Bottom scan / Bottom scan

Rose Petal Effect, Fig. 2 (**a**) Optical images, (**b**) scanning microscope micrographs, and (**c**) atomic force microscope roughness maps of petals of two roses (Rosa Hybrid Tea, cv. Bairage) (Rosa, cv. Bairage), and Rosa Hybrid Tea, cv. Showtime (Rosa, cv. Showtime) (Adapted from Bhushan and Her [3])

impregnated by water or air. The regimes with water penetrating into the microstructure can have high solid–water adhesion and therefore, high CA hysteresis.

Bhushan and Her [3] conducted a series of carefully designed experiments to decouple the effects of the micro- and nanostructures. They synthesized microstructured surfaces with pillars out of epoxy resin. The epoxy surfaces were reproduced from model Si templates and were created by a two-step molding process producing a dual replica (first a negative replica and then a

Droplet on Rosa, *cv.* Bairage

0° tilt angle   180° tilt angle
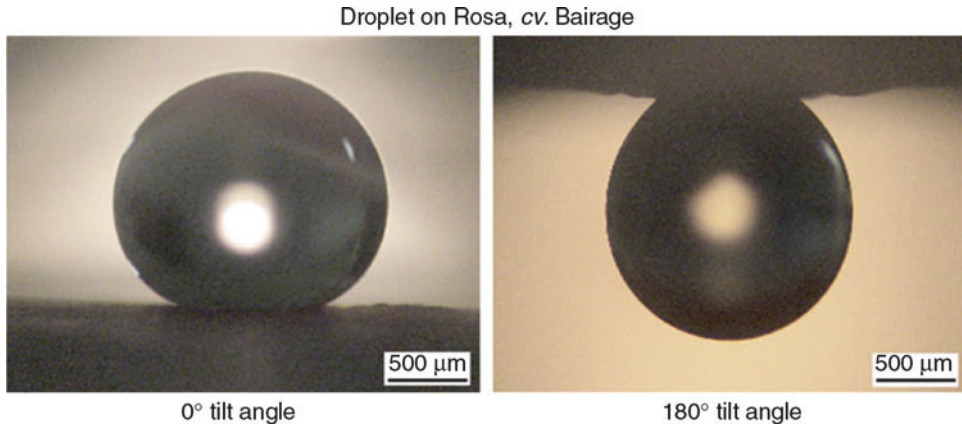
Rose Petal Effect, Fig. 3 Optical micrographs of water droplets on Rosa, cv. Bairage at 0° and 180° tilt angles. Droplet is still suspended when the petal is turned upside down [3]

**Rose Petal Effect, Table 1** Surface roughness statistics for the two rose petals [3]

	Peak-to-base height (µm)	Mid-width (µm)	Peak radius (µm)	Bump density (1/10,000 µm^2)
Rosa, cv. Bairage (high adhesion)	6.8	16.7	5.8	23
Rosa, cv. Showtime (low adhesion)	8.4	15.3	4.8	34

Rose Petal Effect, Table 2 Wetting regimes of a surface with a single level of hierarchy of roughness

State	Cassie–Baxter	Wenzel	Impregnating Cassie
Cavities	Air	Water under droplet	Water everywhere
CA	High	High	High
CA hysteresis	Low	Can be high	Low

Rose Petal Effect, Table 3 Different regimes of wetting of a surface with dual roughness

	Air in microstructure	Water under droplet in microstructure	Water impregnating microstructure
Air in nanostructure	Lotus, high CA, low CA hysteresis	Rose, high CA, high CA hysteresis	Rose-filled microstructure
Water under droplet in nanostructure	Cassie (air-filled microstructure, water in nanostructure), high CA, low CA hysteresis	Wenzel (water in micro- and nanostructure), high CA, high or low CA hysteresis	Wenzel-filled microstructure
Water impregnating nanostructure	Cassie-filled nanostructure	Wenzel-filled nanostructure	Wenzel-filled micro and nanostructure
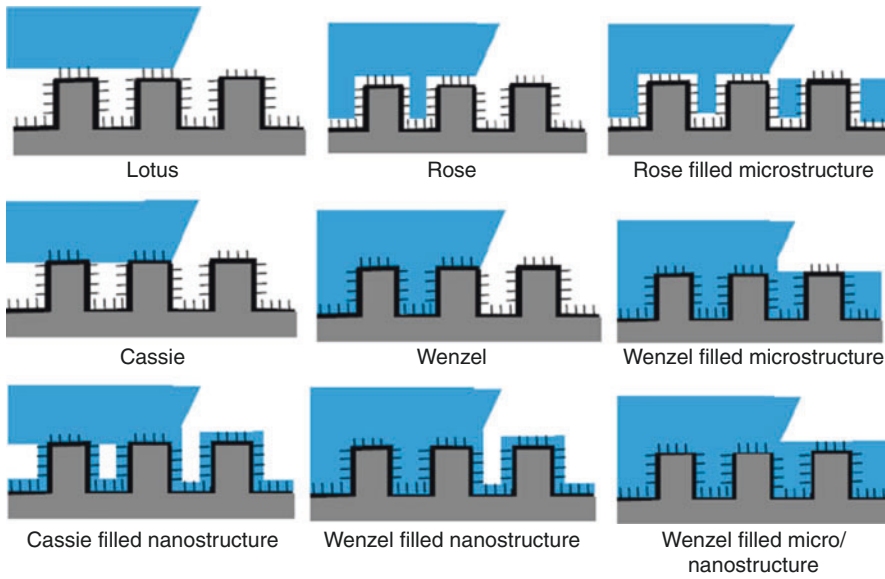
Rose Petal Effect, Fig. 4 Schematics of nine wetting scenarios for a surface with hierarchical roughness
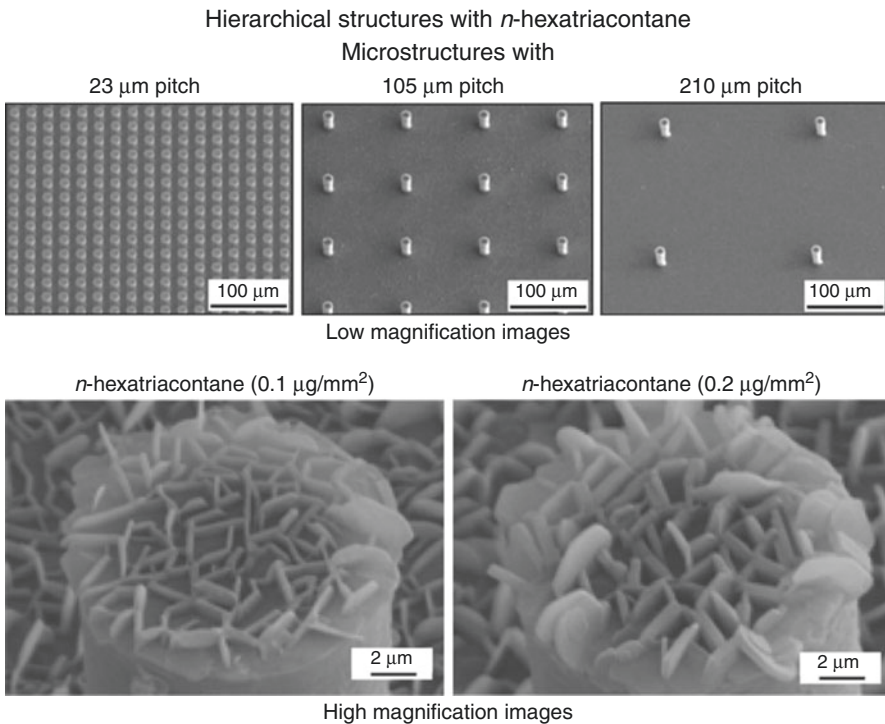
Rose Petal Effect, Fig. 5 SEM micrographs of the microstructures and nanostructures fabricated with two different masses of n-hexatriacontane for hierarchical structure. All images were taken at 45° tilt angle. All samples are positive replicas, obtained from negative replica with dental wax and Si micropatterned master template (14 μm diameter and 30 μm height) fabricated with epoxy resin coated with n-hexatriacontane [3]

positive replica of the original Si template). Surfaces with a pitch (the periodicity of the structure of the pillars) of 23, 105, and 210 μm and with the same diameter (14 μm) and height (30 μm) of the pillars were produced. After that, nanostructures were created on the microstructured sample by self-assembly of the alkane n-hexatriacontane $(CH_3(CH_2)_{34}CH_3)$ deposited by a thermal evaporation method. Alkanes of varying chain lengths are common hydrophobic compounds of plant waxes. On smooth surfaces, alkanes can cause a large contact angle and a small contact angle hysteresis for water droplets. To fabricate the nanostructure, various masses of n-hexatriacontane were coated on a microstructure. The nanostructure is formed by three-dimensional platelets of n-hexatriacontane. Platelets are flat crystals, grown perpendicular to the surface. They are randomly distributed on the surface, and their shapes and sizes show some

variation. Figure 5 shows selected images. When different masses of wax are applied, the density of the nanostructure is changed.
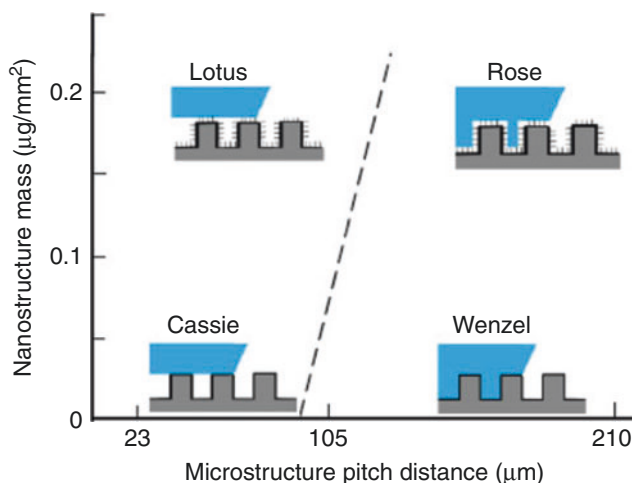
For surfaces with a small pitch of 23 μm, while the mass of n-hexatriacontane is changed, there are only small changes in the static contact angle and contact angle hysteresis values, which means that they are always in the "Lotus" wetting regime. On the surface with a 210 μm pitch value, as the mass of n-hexatriacontane is increased, the static contact angle is increased, and the reverse trend was found for the contact angle hysteresis. This was interpreted as evidence that the nanostructure is responsible for the CA hysteresis and low adhesion between water and the solid surface. The results are summarized in Table 4. The wetting regimes are shown schematically in Fig. 6 as a function of the pitch of the microstructure and the mass of n-hexatriacontane. Small mass of the nanostructure material

Rose Petal Effect, Table 4 CA and CA hysteresis for surfaces with various micro- and nanoroughness (Based on Bhushan and Her [3])

| Mass of n-hexatriacontane (μg/mm^2)	Pitch					
	23 μm		105 μm		210 μm	
	CA	CA hysteresis	CA	CA hysteresis	CA	CA hysteresis
0.1	164°	3°	152	87	135	45
0.12	165°	3°	153	20	135	42
0.16	166°	3°	160	5	150	12
0.2	167°	3°	168	4	166	3

Rose Petal Effect, Fig. 6 Schematic of a wetting regime map as a function of microstructure pitch and the mass of nanostructure material. The mass of nanostructure material equal to zero corresponds to microstructure only (with the Wenzel and Cassie regimes). Higher mass of the nanostructure material corresponds to higher values of pitch, at which the transition occurs

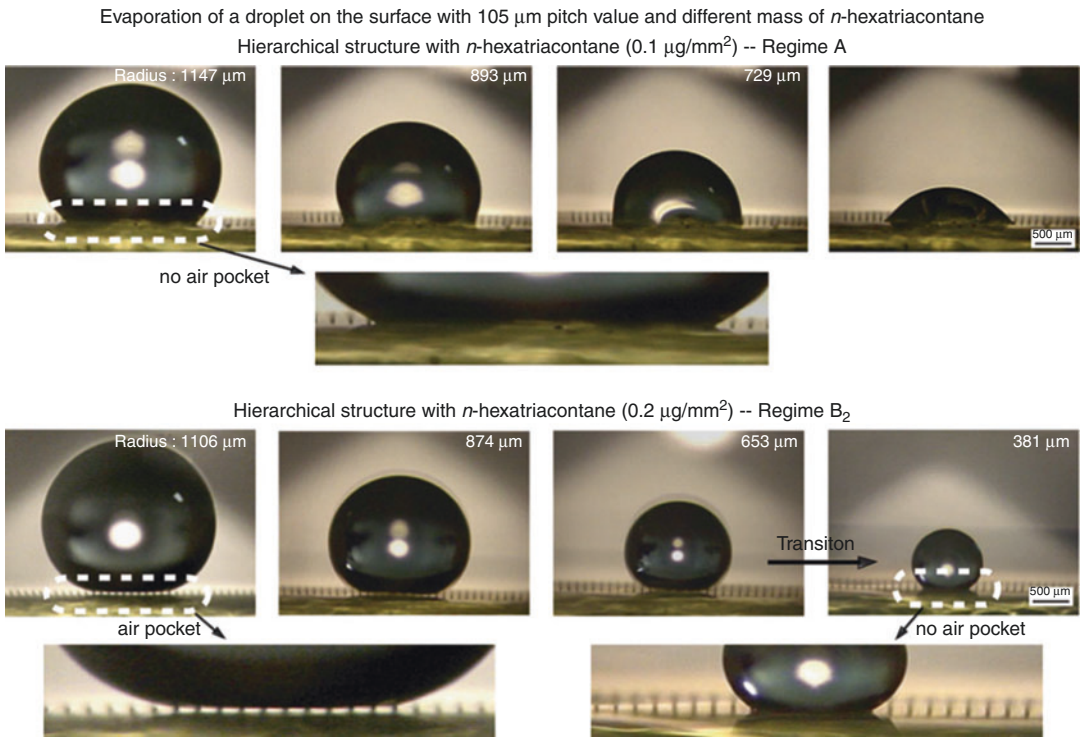**a** Shape of droplets on hierarchical structure with 23 μm pitch

Horizontal surface with *n*-hexatriacontane (0.1 μg/mm^2)

Regime B$_1$

b Shape of droplets on hierarchical structure with 105 μm pitch

Horizontal surface with different mass of *n*-hexatriacontane

n-hexatriacontane (0.1 μg/mm^2) · *n*-hexatriacontane (0.2 μg/mm^2)
Regime A · Regime B$_2$

Inclined surface with *n*-hexatriacontane (0.1 μg/mm^2)
Regime A
vertical · upside down

Rose Petal Effect, Fig. 7 (**a**) Droplet on a horizontal surface of hierarchical structure with 23 μm pitch and n-hexatriacontane (0.1 μg/mm^2) showing air pocket formation and (**b**) droplet on a hierarchical structure with 105 μm pitch and n-hexatriacontane (0.1 μg/mm^2) and 0.2 μg/mm^2 showing no air pocket and air pocket formation, respectively. Also shown is the image taken on the inclined surface with hierarchical structure with 0.1 μg/mm^2 showing that droplet is still suspended [3]
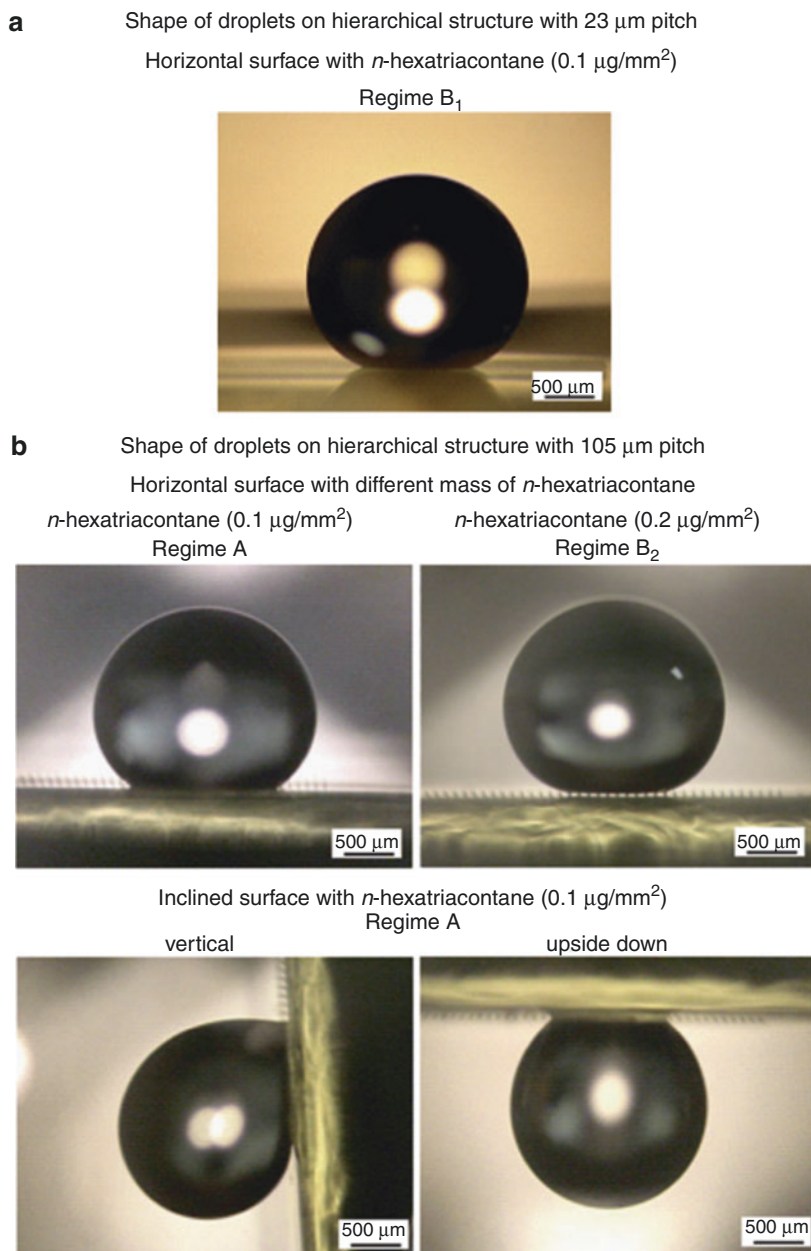
correspond to the Cassie and Wenzel regimes, whereas high mass of the nanostructure corresponds to the Lotus and Rose regimes. The Lotus regime is more likely for larger masses of the nanostructure material. Figure 7 shows a droplet on a horizontal surface of a hierarchical structure with 23 and 105 μm pitches and n-hexatriacontane (0.1 μg/mm^2). Air pockets are observed in the first case and not observed in the second case, indicating the difference between the two regimes [3].

To further verify the effect of wetting states on the surfaces, evaporation experiments with a droplet on a hierarchical structure coated with two different amounts of n-hexatriacontane were performed. Figure 8 shows the optical

Evaporation of a droplet on the surface with 105 µm pitch value and different mass of *n*-hexatriacontane

Hierarchical structure with *n*-hexatriacontane (0.1 µg/mm²) -- Regime A

Hierarchical structure with *n*-hexatriacontane (0.2 µg/mm²) -- Regime B$_2$

Rose Petal Effect, Fig. 8 Optical micrographs of droplet evaporation on the hierarchical structured surfaces with 105 µm pitch value. n-Hexatriacontane (0.1 µg/mm²)-coated sample has no air pocket formed between the pillars in the entire contact area until evaporation was completed. Hierarchical structure with n-hexatriacontane (0.2 µg/mm²) has air pocket, and then the transition from the "Lotus" regime to the "Rose petal" regime occurred [3]

micrographs of a droplet evaporating on two different hierarchical structured surfaces. On the n-hexatriacontane (0.1 µg/mm²)-coated surface, an air pocket was not visible at the bottom area of the droplet. However, the droplet on the surface has a high static contact angle (152°) since the droplet still cannot completely impregnate the nanostructure. The footprint size of the droplet on the surface has only small changes from 1820 to 1791 µm. During evaporation, the initial contact area between the droplet and hierarchical structured surface does not decrease until the droplet evaporates completely, which means complete wetting between droplet and microstructures. For the n-hexatriacontane (0.2 µg/mm²)-coated surface, the light passes below the droplet, and air pockets can be seen, so to start with the droplet is in the Cassie-Baxter regime. When the radius of the droplet decreased to 381 µm, the air pockets are

not visible anymore. The footprint size of the droplet on the surface is changed from 1177 to 641 µm, since droplet remained on only a few pillars until the end of the evaporation process [9].

The experimental observations of the two types of rose petals show that hierarchically structured plant surfaces can have both adhesive and non-adhesive properties at the same time with high CA. This is due to the existence of various modes of wetting of a hierarchical surface, so that water can penetrate either into macro- or nanoroughness, or into both. Water penetration into the microroughness tends to result in high adhesion with the solid surface, whereas the presence of the nanoroughness still provides high CA. As a result, two distinct modes of wetting are observed. One can be called the "Lotus" mode (with low adhesion) and another is the "rose" mode with high adhesion.

Cross-References

▶ Biomimetics
▶ Lotus Effect

References

1. Wenzel, R.N.: Resistance of solid surfaces to wetting by water. Ind. Eng. Chem. **28**, 988–994 (1936)
2. Cassie, A., Baxter, S.: Wettability of porous surfaces. Trans. Faraday Soc. **40**, 546–551 (1944)
3. Bhushan, B., Her, E.K.: Fabrication of superhydrophobic surfaces with high and low adhesion inspired from rose petal. Langmuir **26**, 8207–8217 (2010)
4. Bormashenko, E., Stein, T., Pogreb, R., Aurbach, D.: "Petal Effect" on surfaces based on lycopodium: high-stick surfaces demonstrating high apparent contact angles. J. Phys. Chem. C **113**, 5568–5572 (2009)
5. Chang, F.M., Hong, S.J., Sheng, Y.J., Tsao, H.K.: High contact angle hysteresis of superhydrophobic surfaces: hydrophobic defects. Appl. Phys. Lett. **95**, 064102 (2009)
6. Feng, L., Zhang, Y., Xi, J., Zhu, Y., Wang, N., Xia, F., Jiang, L.: Petal effect: a superhydrophobic state with high adhesive force. Langmuir **24**, 4114–4119 (2008)
7. Gao, L., McCarthy, T.J.: Teflon is hydrophilic. Comments on definitions of hydrophobic, shear versus tensile hydrophobicity, and wettability characterization. Langmuir **24**, 9184–9188 (2008)
8. Jin, M.H., Feng, X.L., Feng, L., Sun, T.L., Zhai, J., Li, T.J., Jiang, L.: Superhydrophobic aligned polystyrene nanotube films with high adhesive force. Adv. Mater. **17**, 1977–1981 (2005)
9. Bhushan, B., Nosonovsky, M.: The rose petal effect and the modes of superhydrophobicity. Philos. Trans. Royal. Soc. A. **368**, 4713–4728 (2010)

Rotary Nanoelectromechanical Systems (NEMS)

▶ Mechanism for Assembling Arrays of Rotary Nanoelectromechanical Devices

Rotary Nanomachines

▶ Mechanism for Assembling Arrays of Rotary Nanoelectromechanical Devices

Rotary Nanomotors

▶ Mechanism for Assembling Arrays of Rotary Nanoelectromechanical Devices

Rotation Sensors

▶ Gyroscopes

R