G

Gallium Arsenide (GaAs)

▶ Physical Vapor Deposition

Gamma-butyrolactone (GBL)

▶ SU-8 Photoresist

Gammatone Filters

▶ Bio-inspired CMOS Cochlea

Gas Etching

▶ Dry Etching

Gas Phase Nanoparticle Formation

Yuehai Yang and Wenzhi Li
Department of Physics, Florida International
University, Miami, FL, USA

Synonyms

Nanoscale particle; Nano-sized particle

Definition

In the materials science community, the term "nanoparticle" is generally used to indicate particles with atomic domains spatially confined to less than 100 nm since the physical properties of materials start to change significantly due to a variety of confinement effects in this size range. Another commonly used concept, "cluster," indicates smaller particle with a diameter of only several nanometers, corresponding to 10^4 molecules or atoms. Sharing the same features with other types of nanostructured materials, nanoparticles also have significant atom fractions associated with interfacial environments, and interactions between their constituent domains. Gas phase particle formation refers to production of particles from individual atoms or molecules in the gas phase.

Overview

In the last several decades, intensive research has been performed on nanostructured materials, that is, materials with grain sizes less than 100 nm. Nano-sized particles are considered as the starting point for preparations of an extensive variety of nanostructured materials. Meanwhile, nanoparticles themselves have played a significant role in novel technology development [1]. As such, they have been prepared by a variety of methods and their synthesis is a crucial part of rapidly growing nanoscale research.

Gas phase synthesis technique is well known for producing a wide range of nanoparticles. Gas phase particle generation is captivating because it excludes the wet byproducts of liquid phase processes and the as-prepared product particles promptly separated from the gas stream are in high purity without post-growth

treatment. The gas phase synthesis has been extensively developed and widely applied in industries. Meanwhile, the commercialization of new products has experienced a considerable progress and the commercial nanoparticles produced by this method comprise the largest share of the market. In the synthesis of nanoparticles from atomic/molecular precursors, one wants to be able to control different aspects of this condensed ensemble, in which the most important one is the size and size distribution of the resultant particles. The reason, as indicated above, is that the material parameters of a solid experience a dramatic change when cluster size becomes less than a certain threshold value. A property will be altered if the entity is confined within a space smaller than the critical length of the entity responsible for this property. Therefore, the product quality and application characteristics of nanoparticles strongly depends on the distribution of particle size and on the particle morphology. Well-controlled gas phase routes are able to synthesize nearly spherical and nonporous particles with high purity. The resulting particles are small in size with narrow size distribution.

In gas phase synthesis of nanoparticles, conditions include usual situations of a supersaturated vapor, or chemical supersaturation in which it is thermodynamically favorable for the gas phase molecules to undergo chemical reactions and condensation. If the reaction and condensation kinetics permit, particles will nucleate homogeneously when the degree of supersaturation is sufficient. Once nucleation occurs, the reaction/condensation of the gas phase molecules will occur on the resulting particles and will keep relieving the remaining supersaturation, since then the nucleation process stops. With this gas phase conversion process, particles are built all the way up to the desired size.

Although the processes of gas phase nanoparticle synthesis vary from one to another, they all have in common the fundamental aspects of particle formation mechanisms. In gas phase reactors, three major formation mechanisms determine the particle formation within the first few milliseconds of the synthesis process, and will determine the final characteristics of particulate products. The influences of these mechanisms on particle formation, growth, and final morphology are shown in Fig. 1.

1. The precursor chemical reactions form product clusters by nucleation, and then the clusters grow into particles by surface growth [2].
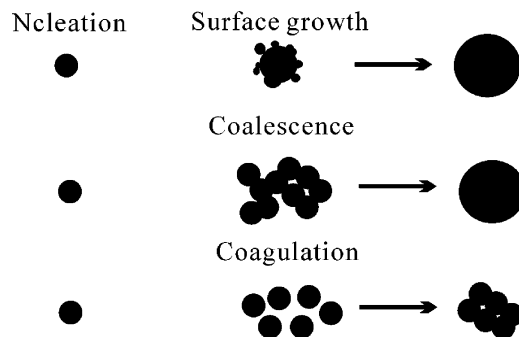
Gas Phase Nanoparticle Formation, Fig. 1 Gas phase particle formation mechanism

2. Because of the sintering processes, coalescence occurs in the high-temperature zones of the reactor to reduce aggregation and enhance the formation of spherical particles [3].

3. Also, coagulation occurs inevitably in the gas phase synthesis with high particle concentrations, after the formation of particles in a gas phase; their coagulation rate is weakly dependent on particle size and proportional to the square of particle number concentration.

If temperature is sufficiently high, the coalescence of particles is much faster than their coagulation to produce spherical particles. On the contrary, loose agglomerates are formed with open structures.

Most advances in this field closely relate to gaining better understanding and better control of particle aggregation and coalescence in order to produce nonporous particles with desired size and narrow size distribution. Gas phase synthesis is one of the best techniques in controlling nucleation-condensation growth and collecting/handling nanoparticles afterward. There are various parameters such as flame temperature, precursor residence time, ions, gas concentration, cooling rate, and additives which can affect the process of coagulation and sintering, and consequently, the characteristics of the resulting particles with respect to the size monodispersity. In addition to the particle size and size distribution, the composition of the constituent phases and the nature of the interfaces created between constituent phases are also of crucial importance if the particles are constituted by more than one material. The interplay among size, composition, and interfaces determines the properties of nanoparticles.

Specific Synthesis Methodologies

The physical and chemical processes for nanoparticle formation in gas phase are qualitatively understood. The process has been used in the formation of millions of tons of particles in nano-/submicroscale every year in different industries. Particle generation in the gas phase is carried out either by precursor gases reaction or by evaporation. To generate the high temperature for these processes, high-energy sources such as Joule heating, hydrocarbon flames, plasmas, ions, electrons, or laser beams are needed. Therefore, different kinds of furnace, flame, plasma, and laser reactors have been used to facilitate gas phase reactions and produce the desired materials.

Gas Phase Condensation (for Metal Nanoparticles)

To achieve supersaturation, the most straightforward method is to heat certain solids and evaporate them into background gases, and then reduce the temperature by mixing the vapors with cold gases. Formed by homogeneous nucleation in the vicinity of the gas source, the clusters grow in this process by gas phase coalescence of the source atoms. It has been found that the average particle size of the resulted nanoparticles increases with the increase of gas pressure and the applied inert gas mass. Experimentally, it has also been found that the particle size distribution is lognormal. This has been explained theoretically by the particle's growth mechanisms [4]. This method is also well suited for producing metal nanoparticles as for most metals, the evaporating temperature is attainable and evaporation rates are reasonable. Furthermore, compounds of the evaporated metal can be prepared by including a reactive gas in the cold gas stream [5].

In the gas-condensation process, it has been studied [6] that the formation of the atom clusters is essentially controlled by three fundamental rates:

1. The evaporation rate
2. The rate that the hot atoms lose their energy in the condensation region
3. The removal rate of the nucleated clusters from the supersaturation region

The evaporation-condensation process functions as a form of distillation. Therefore, with volatile impurities removed in this process, the purity of the material can be further improved. However, there are a few limitations for this synthesis method: (1) it is difficult to control composition in the case of synthesizing more complex materials by this method and (2) this process is restricted to materials which have low melting points or high vapor pressures.

Nanoparticle Formation in Flames

So far, in the nanoparticle synthesis, the most commercially successful approach is carrying out the synthesis process within a flame, which provides the energy to evaporate the precursors so that the heat needed to induce reaction and particle nucleation is produced by the combustion reactions. Every year, this method alone produces millions of metric tons of nanostructured materials, such as carbon black and metal oxides. Since the flame provides oxidizing environment, it has been primarily used to make oxides materials.

High temperatures (up to $2,400°C$) can be realized in the flames and the precursor concentration in the flame can be quite high due to the high energy density. In the highest temperature region where the primary particles are formed, the residence time only ranges between 10 and 100 ms. After this region, the only things that can be influenced are the particle size and the aggregation morphology. The attainable particle sizes using flame synthesis is from a few nanometers up to 500 nm. The specific surface areas of these particles can go $400 \text{ m}^2/\text{g}$ and higher.

Temperature profile, reactor residence time, and reactant concentration are the three crucial parameters for generating the desired particles. Unfortunately, the coupling of the flame chemistry to the nanoparticle production makes this process very complex and difficult to control. Every adjustment to the feed flow changes all three parameters simultaneously. Therefore, it is impossible to change these three parameters independently in flame reactors.

Furnace Flow (Hot-Wall) Reactor

Furnace flow systems employ tubular furnace-heated reactors for initiating the synthesis reaction. Temperatures for reaction are often set below $1,700°C$. The gas composition is freely selectable with variable concentrations. Usually, the system pressure is atmospheric, but it can also be set at other values and used as a process parameter.

With high efficiency, the reactor has a relatively simple design. The process of this technique is precisely controllable and flexible with respect to gas composition and system pressure. Therefore, it allows the production of particles with specific

characteristics. It also allows different kinds of particle production in the range from atomic to micrometer dimensions. However, it often causes a high degree of aggregation at high aerosol concentrations. Moreover, compared to other methodologies, it has high energy requirements.

Plasma Reactor

Supersaturation and particle nucleation can also be achieved by injecting the precursors into thermal plasma to provide the energy needed for inducing reactions. It is quite common to deposit thin films and coatings by plasma spraying and thermal plasma vapor (chemical or physical) deposition. The most commonly used electrical methods to produce plasmas are high-intensity arcs and inductively coupled high-frequency discharge. Solid precursors are conveyed in the plasma where they are fully evaporated and decomposed into atoms in the plasma center by the high temperatures. The precursor vapors are then cooled rapidly by mixing with quenching gas at the end of the plasma. As a result of high supersaturation, ultrafine particles are obtained.

Ion Sputtering

The method of sputtering is to vaporize materials by bombarding a solid surface with inert gas ions shot in high velocities. The sputter source, for example, an ion gun, usually works in systems with of 10^{-3} mbar or even higher vacuum, since a higher pressure will block the sputtered material transportation. Electrons can be used to replace ions for the same purpose. The major advantage of sputtering is that the material being heated is mainly the target material, and thus the composition of the sputtered material is that of the target. However, this process needs to be carried out at relatively low pressures which can cause difficulties for further nanoparticles processing in gas phase.

Spark Discharge

High-current spark between two solid electrodes is used as another means of vaporizing metals. In the presence of an inert background gas, the electrode material is evaporated to produce nanoparticles until it reaches the breakdown voltage. This technique is very useful for materials with a high melting point (e.g., Si or C), which cannot be evaporated in a furnace. The spark (arc) formed across the electrodes

can only vaporize a small amount of metal to produce very small amounts of nanoparticles, but this process is relatively reproducible.

Laser Ablation

Instead of evaporating certain material to produce a supersaturated vapor, a pulsed laser can be utilized to evaporate a piece of spatially confined material. The laser beams are focused onto the center of the reaction chamber where the precursor (e.g., coarse metal oxide powder) is fluidized, resulting in a power supply of a few kilowatts in a volume of only a few cubic millimeters. The precursor powder is evaporated by the focused laser, and nanoparticles are formed by the vapor condensation. This method can be used to vaporize materials that cannot readily be evaporated even though laser ablation can generally only produce small amounts of nanoparticles. For solids having high vaporization temperatures, typically ceramics and metal oxides, from which nanoparticles cannot be synthesized by standard gas phase processes, the high-power intensity of the laser opens a gate for them to be used as precursors. The nanoparticles synthesized by laser have potentials for new applications since their morphologies are significantly different from typical pyrogenic oxides due to the high cooling rates.

Nanoparticle Collection

Among the collection devices, the simplest one is a small-size pore filter for separating nanoparticles from a gas stream. The disadvantage is that impurities and defects can result from the incorporation of the filter material parts into the nanocrystallined particles. Therefore, by employing thermophoretic forces, a collection device was designed to apply a permanent temperature gradient, which results in the separation of nanoparticles from the gas that has been used. In addition, collection of the nanoparticles in liquid suspension can be used to enhance the stability of the collected nanoparticles against compositional changes, sintering, and agglomeration.

Limitation of Gas Phase Process

1. Hazardous gaseous reactants and its byproducts are apparent disadvantages of gas phase nanoparticles synthesis.

2. Nanoparticles agglomeration is inevitable in the gas phase synthesis. However, by capping the particles with appropriate ligands in the liquid phase synthesis, the dispersion of nanoparticles can be indefinitely stabilized.

3. In gas phase synthesis, it is often difficult to produce composite materials with uniform chemical compositions because heterogeneous compositions within an individual particle or from particle to particle can be caused by differences in the vapor pressures, chemical reaction rates of the reactants, and nucleation/growth rates of the products.

Engineering Applications

Nanoparticles have been applied or been evaluated for use in many fields by employing their electrical, magnetic, thermal, optical, and mechanical properties. Nanoparticles synthesized in gas phase have also emerged into engineering fields such as heterogeneous catalysis [7], biomaterials, electroceramics manufacturing, dental materials, and fuel cell membrane construction. The advantage of nanoparticle application is that it incorporates many effects related to their sizes. These effects include "the quantum size effects" (electronic effects caused by delocalized valence electron confinement), "the many-body effect" (altered cooperative atom phenomena), for example, lattice melting or vibrations, and suppression of the lattice-defect mechanisms, such as dislocation generation and migration within the confined grain sizes. The capability of engineering a large variety of useful technological applications will be impacted by possibly size-selected nanoparticle synthesis and assembling the nanoparticles into novel materials with unique and improved properties including controlled mechanical, electronic, optical, and chemical properties. Microelectronics and even medical applications [8] will profit from developing this nanoparticle formation process.

Future Research

Based on the newly arisen questions and challenges, research in this field underlines the need for synthesizing nanoscale particles (especially mixed oxides particles) with precisely controlled characteristics. Future research includes topics such as the scale-up synthesis of particles with controlled properties, verification of mesoscopic chemistry relationships that can be applied in research, and nano-thin coated nanoparticles made in large quantities. In this endeavor, scientists and engineers should accumulate their knowledge and contribute to small particles generation, granulation, flocculation, crystallization, and classification.

Cross-References

▶ Cellular Mechanisms of Nanoparticle's Toxicity
▶ Ecotoxicity of Inorganic Nanoparticles: From Unicellular Organisms to Invertebrates
▶ Electric Field–Directed Assembly of Bioderivatized Nanoparticles
▶ Electric-Field-Assisted Deterministic Nanowire Assembly
▶ Exposure and Toxicity of Metal and Oxide Nanoparticles to Earthworms
▶ Fate of Manufactured Nanoparticles in Aqueous Environment
▶ Genotoxicity of Nanoparticles
▶ In Vitro and In Vivo Toxicity of Silver Nanoparticles
▶ In Vivo Toxicity of Titanium Dioxide and Gold Nanoparticles
▶ Mechanical Properties of Nanocrystalline Metals
▶ Nanoparticles
▶ Optical Properties of Metal Nanoparticles
▶ Perfluorocarbon Nanoparticles
▶ Physicochemical Properties of Nanoparticles in Relation with Toxicity
▶ Self-assembly of Nanostructures
▶ Synthesis of Gold Nanoparticles
▶ Synthesis of Subnanometric Metal Nanoparticles

References

1. Afzaal, M., Ellwood, K., Pickett, N.L., O'Brien, P., Raftery, J., Waters, J.: Growth of lead chalcogenide thin films using single-source precursors. J. Mater. Chem. **14**, 1310–1315 (2004)
2. Pratsinis, S.E., Spicer, P.T.: Competition between gas phase and surface oxidation of $TiCl_4$ during synthesis of TiO_2 particles. Chem. Eng. Sci. **53**, 1861–1868 (1998)
3. Johannessen, T., Pratsinis, S.E., Livbjerg, H.: Computational fluid-particle dynamics for the flame synthesis of alumina particles. Chem. Eng. Sci. **55**, 177–191 (2000)
4. Granqvist, C.G., Buhrman, R.A.: Log-normal size distributions of ultrafine metal particles. Solid State Commun. **18**, 123–126 (1976)

5. Wegner, K., Walker, B., Tsantilis, S., Pratsinis, S.E.: Design of metal nanoparticle synthesis by vapor flow condensation. Chem. Eng. Sci. **57**, 1753–1762 (2002)
6. Siegel R.W. In: Cahn, R.W. (ed.): Materials Science and Technology **15**, 583 (VCH, Weinheim, 1991)
7. Chiang, W.H., Sankaran, R.M.: Synergistic effects in bimetallic nanoparticles for low temperature carbon nanotube growth. Adv. Mat. **20**, 4857–4861 (2008)
8. Boisselier, E., Astruc, D.: Gold nanoparticles in nanomedicine: preparations, imaging, diagnostics, therapies and toxicity. Chem. Soc. Rev. **38**, 1759–1782 (2009)

Gas-Phase Molecular Beam Epitaxy (Gas-Phase MBE)

▶ Physical Vapor Deposition

Gecko Adhesion

Elmar Kroner[1] and Eduard Arzt[2]
[1]INM – Leibniz Institute for New Materials, Saarbrücken, Germany
[2]INM – Leibniz Institute for New Materials and Saarland University, Saarbrücken, Germany

Synonyms

Dry adhesion; Fibrillar adhesion

Definition

The gecko adhesion system is a dry adhesive based on a fibrillar surface pattern and allows easy, repeatable, and residue-free detachment.

Conventional Adhesives and the Gecko Adhesion System

Adhesives are an important ingredient of modern technology. Their function is to reliably connect two different objects with each other. Although other techniques may be used to assemble two objects (e.g., screwing or welding), the application of adhesives has many benefits. For example, they provide a more uniform distribution of stresses along joints, can be applied at ambient (or moderately elevated) temperature, or can fulfill additional requirements such as acoustical damping or energy dissipation. Adhesives have also been key to new technologies, e.g., in the development of laminated fiber-reinforced materials. The need for adhesives in special applications has led to the development of more than 250,000 adhesives in use today.

Nature has also evolved adhesives for numerous purposes. A special case is a reusable adhesive in which attachment, detachment, and reattachment occurs without loss in adhesion strength: insects, spiders, and geckos can stick to walls and ceilings, attaching and detaching their feet rapidly and repeating this for thousands of times without losing the ability to stick. Geckos belong to the heaviest animals supporting such an adhesive system, which exhibits the following characteristics:

- *High adhesion to smooth and rough surfaces based on physical interactions*: Unlike conventional adhesives, which are usually based on chemical bonds or mechanical interlocking, gecko adhesives work without pretreatment of the surface, e.g., by cleaning or roughening.
- *Adhesion barely depends on the chemical properties of the surface*: Conventional adhesives often have to be tailored to different surfaces to achieve high adhesion.
- *Directional adhesion, allowing easy detachment*: Unlike conventional adhesives, gecko adhesives enable anisotropic adhesion properties. While sticking occurs in one direction, detachment happens in another.
- *High reversibility of adhesion*: The reversibility of conventional adhesives is poor and seldom lasts more than a few attachment cycles. Gecko adhesives can open and close hundreds or thousands of times.
- *Residue-free detachment*: Most conventional chemistry-based adhesives leave traces of residues on the surface during removal.
- *Self-cleaning properties*: Dust and particles, which stick to conventional adhesives, can be removed easily in gecko adhesives.

Mimicking the attachment system of geckos has attracted the attention of a worldwide research community. A successful development could lead to significant improvements in countless fields of application, or even lead to new technologies. After about a decade of research, first applications begin to appear on the horizon.

G
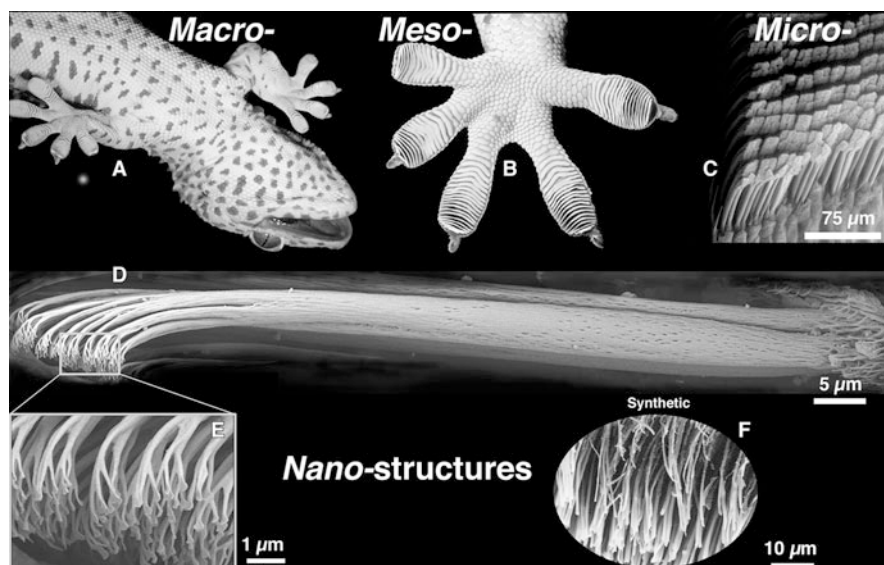
History of Gecko Adhesion Research

The ability of geckos to climb up walls and run along ceilings on different materials and various degrees of roughness has fascinated humans for thousands of years. Geckos were first mentioned in writing by the Greek philosopher *Aristotle* (350 B.C.), who stated in his work *Historia Animalum*: "The woodpecker [...] can run up and down a tree in any way, even with the head downwards, like the gecko-lizard." However, the reason behind this astonishing adhesion performance was discovered only with the development of electron microscopy in the twentieth century.

A brief study of the gecko toe structure using scanning electron microscopy was published by Hiller and Blaschke in 1967 [1]. They showed that gecko toes are covered with fine hairs (called *setae*) with tens of micrometers in length and several micrometers in thickness. These *setae* branch into even finer hairs (*spatulae*) and end in plate-like tips with diameters of about 200 nanometers (nm) (Fig. 1). The results of their observation led them to the now refuted conclusion that the surface structure of the gecko toes provides high adhesion and friction due to mechanical interlocking on micro and nanorough surfaces. In 1968, Hiller found a correlation between water contact angle and attachment forces of gecko toes and corrected his statement from the previous year by indicating that their function depend on "adhesion processes" rather than mechanical interlocking. In 1980, Stork [2] published an extensive experimental study on the beetle *Chrysolina polita*, which also suggests an adhesion system based on surface structures similar to those found in geckos. In his paper, he identified the main contribution to adhesion as molecular interactions, possibly with additional contribution from capillary forces.

Two decades later, research on the gecko adhesion system heated up to become a prime topic in biomimetics. In their 2000 publication, Autumn et al. [4] measured the adhesion force of single gecko *setae*. They extrapolated that, for all *setae* in perfect contact, a single gecko foot could adhere with a shear force of approximately 100 Newton (N). Triggered by this astonishing sticking ability, many key developments happened over recent years:

- Modern micro- and nanofabrication methods allowed bioinspired surface patterns to be produced with increasingly complex geometries.
- New methods for structural and mechanical characterization enabled extensive studies on the adhesion behavior of micropatterned surfaces.
- New micromechanical models helped explain the "gecko effect" and served as design guides for successful artificial surface patterns.
- The properties of gecko-inspired adhesives, i.e., high physical adhesion with high reversibility and repeatability, rapid and easy detachment, self-cleaning properties, provoked interest from several application sectors, e.g., for robot locomotion, microfabrication, biomedical use, or sports articles.
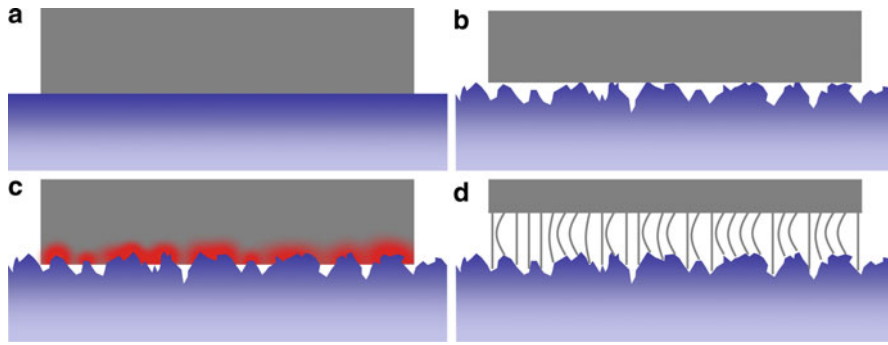
Gecko Adhesion, Fig. 2 Contact scenarios of two bodies: (**a**) two smooth surfaces with maximum contact area, (**b**) a smooth and a rough surface, both rigid, (**c**) as in (**b**) but with elastic deformation of the upper solid (high stresses are schematically indicated in *red*), and (**d**) a patterned surface which can adapt to a rough counterpart with small elastic energy penalty

While the gecko-inspired adhesion system offers the most spectacular characteristics, other adhesion systems in nature, e.g., of spiders and insects, are at least as interesting as the gecko system. An overview of adhesion in nature can be found in [5].

Contact Mechanics of Patterned Surfaces

To understand the gecko attachment system in its complexity, it is necessary to apply principles of contact mechanics to the special case of patterned surfaces. The adhesive interaction between two surfaces can be investigated in a static (no movement allowed) or dynamic (movement allowed, e.g., peeling) system.

Static system for adhesion – The interplay between attractive and repulsive forces determines the strength of the adhesive force between two touching objects. Attractive forces can arise, e.g., from molecular interactions due to van der Waals forces, electrostatic or magnetic forces, capillary forces, or gravity. For the gecko adhesion system, the main contribution is the presence of molecular interactions as demonstrated by Stork [2]. More recent studies have shown that humidity effects also contribute to adhesion; absorbed water may reduce surface roughness [6] or influence the material properties of the gecko structures [7], both leading to increased adhesion. Repulsive forces can result, e.g., from electrostatic and magnetic repulsion or – most important in the case treated here – the

energy stored in the elastically deformed bodies. If the attractive forces (positive per definition) and repulsive forces (negative) are now balanced, adhesion occurs, if the attractive forces are larger than the repulsive forces. If the surfaces adhere to each other, an *external* force is necessary to detach the two surfaces. This force is defined to have a negative sign and is called pull-off force P_c.

Van der Waals forces, being the result of dipole interactions, are short ranged and become significant only for distances below several nanometers. Surface chemistry has only a small influence on the magnitude of the attractive forces compared to the distance and is fixed for a contact pair. Thus, the only parameter, which can be modified to increase the attractive interactions, is the distance between the contacting surfaces.

For illustration, consider some special cases: when perfectly smooth surfaces, e.g., the flat sides of two silicon wafers, are brought into contact, the true contact area is large and elastic deformation is unnecessary; it will then be impossible to detach the wafers from each other without damaging them (this effect is exploited in a processing step called "wafer bonding"), see Fig. 2a. However, if a stiff solid is placed onto a rough surface, the true contact area will be negligible (Fig. 2b) and it will not stick to it; a small roughness is sufficient to significantly decrease the contact area and, due to the extremely short force range, the attractive forces. In order to increase the contact area, an external force compressing the two surfaces is necessary.

However, elastic deformation stores energy in the two objects and causes repulsive forces, resulting in vanishing adhesive forces.

To obtain significant adhesion, it is necessary to increase the true contact area while minimizing the elastic energy penalty. This can be realized by designing an adhesive with an *effective* Young's modulus as low as possible. The pioneering adhesives scientist *Carl A. Dahlquist* defined the limiting effective Young's modulus to achieve "tackiness" to be below 100 kilo Pascal (kPa). Such low values may be obtained by either choosing very soft adhesives, usually soft polymeric materials, or by introducing surface structures with high aspect ratios (Fig. 2d), where the effective Young's modulus is decreased due to additional bending of the structures. Although the contact area is decreased by patterning the surface, the reduction of the repulsive forces due to the gain in compliance can be higher, thus leading to an increased adhesion force.

The fabrication of low modulus materials seems to be tempting as a complicated microstructuring process can be avoided. However, such materials have several drawbacks for reversible and repeatable attachment. First, low modulus materials usually wear quickly, which would significantly decrease the repeatability of the adhesive performance. Second, soft materials tend to transfer material to the counter surface, e.g., free oligomers in soft polymers, disqualifying them as residue-free. Third, continuous flat surfaces will contaminate easily with dust particles and dirt, necessitating a periodic cleaning process when applied in dirty environments. Fourth, due to the isotropy within the surface plane, the adhesion will be isotropic as well. Thus, there will be no preferred detachment direction, which is a key property of the gecko adhesion system to allow locomotion. And finally, under dynamic attachment and detachment conditions, the soft non-patterned surface will be inferior to patterned surfaces, which will be explained in the following paragraph.

Dynamic system for adhesion – Patterned surfaces are superior in adhesion compared to flat ones, especially when the dynamics of detachment is considered. If detachment is initiated for a flat adhesive, the elastically stored strain energy freed in the process is available to drive the detachment front further.

For a disrupted surface, however, the strain energy stored in single fibrils will be lost on detachment and will not contribute to advance the detachment front (Fig. 3a). An adhesive contact can be described as a fracture problem, where the initiation of a detachment front is regarded as crack initiation and the advancing detachment corresponds to crack propagation. In fracture processes, the initiation of a crack is usually energetically more difficult than crack propagation. Once initiated, a crack will propagate easily. This analogy can be translated to flat and patterned adhesives. For a flat adhesive, the detachment front needs to be initiated only once and will subsequently propagate. However, for a patterned surface, new detachment cracks need to be initiated for each single contact, leading to an energetically more difficult detachment process.

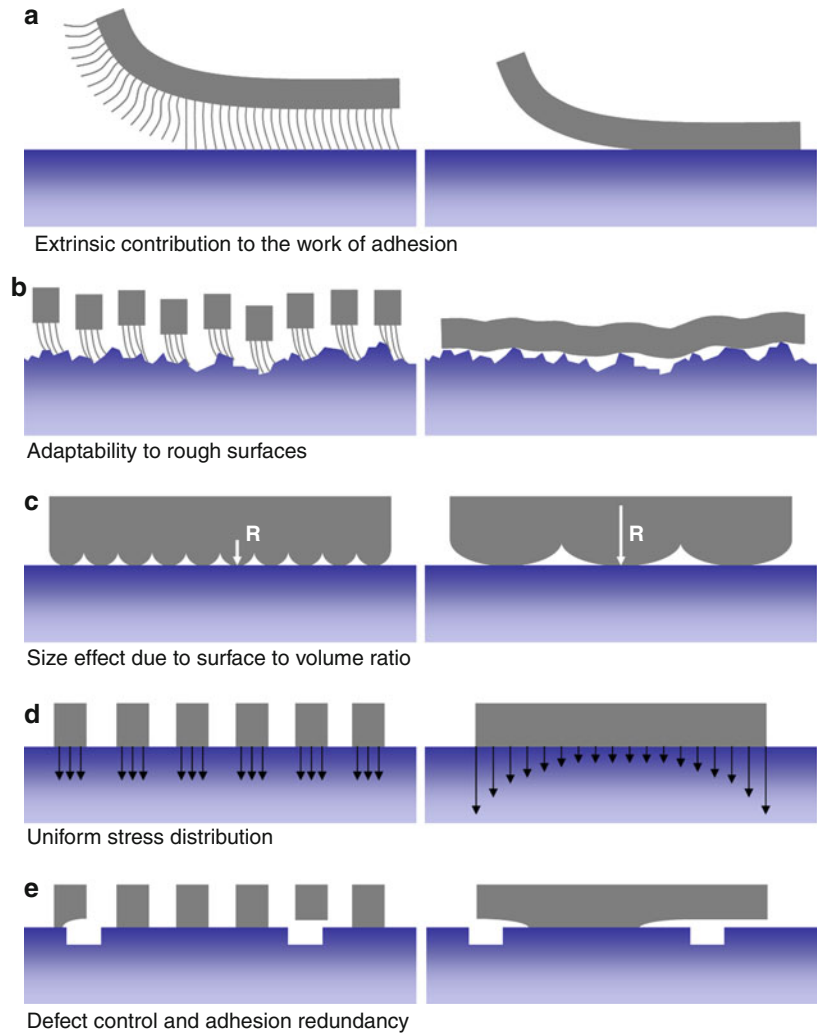## Size Matters – The Principle of "Contact Splitting"

In 1971, Johnson, Kendall, and Roberts [8] introduced a theory of contacts between soft elastic spheres, which soon became the most highly cited publication for contacts between soft materials, now known as the JKR theory. By balancing elastic, potential, and surface energy, they obtained a theoretical expression for the external pull-off force P_c required to separate two spherical bodies:

$$P_c = -\frac{3}{2}\pi\gamma R \qquad (1)$$

Here, R is the relative radius of curvature defined by $1/R = 1/R_1 + 1/R_2$, R_1 and R_2 being the radii of the two contacting spheres, and γ the "work of adhesion," defined as $\gamma = \gamma_1 + \gamma_2 - \gamma_{12}$, where γ_1 and γ_2 are the surface energies of the two contacting materials, respectively, and γ_{12} is the interfacial energy. The negative sign in Eq. 1 occurs because tensile forces are defined to be negative in the contact mechanics literature. A thorough derivation of the JKR theory, additional theories for soft elastic contacts, and adhesion problems of other contact geometries such as flat contacts can be found in the books by Johnson [9] and Maugis [10].

a

Extrinsic contribution to the work of adhesion

b

Adaptability to rough surfaces

c

Size effect due to surface to volume ratio

d

Uniform stress distribution

e

Defect control and adhesion redundancy

In 2003, Arzt et al. [11] showed that Eq. 1 results in a generic size effect for adhesion: The pull-off force for a large contact is smaller than the sum of the pull-off forces for many smaller contacts together, covering the same apparent contact area. They termed this concept the "principle of contact splitting." One consequence of this principle, e.g., is the following: The number of equally sized spherical contacts n is proportional to the contact area A, which itself is proportional to the square of a length. Considering Eq. 1, it follows that the pull-off force is proportional to the square root of the number of contacts:

$$P_c \sim R \sim \sqrt{A} \sim \sqrt{n} \qquad (2)$$

Thus, with increasing number of contacts, the pull-off force increases as well. Other contact geometries than spherical contacts lead to different exponents for the number of contacts but still result in an increasing pull-off force for increasing number of contacts. This simple model offered an explanation for the experimental observation that with increasing animal body weight, the size of their contact elements decreased. Their prediction fitted well to about 6 orders of magnitude in animal body weight, although the effect mainly reflected the differences between different species (flies, beetles, spiders, and geckos), but not within these groups. It was later pointed out by Federle [12] that the data included a transition from fluid-mediated to dry adhesion, which can give rise to additional effects.

The experimental finding that smaller contacts showed increased adhesion compared to larger ones has triggered numerous theoretical explanations (Fig. 3). They may be summarized as follows (classification after [13]):

- Extrinsic contribution to the work of adhesion (Fig. 3a): Patterned surfaces are more resistant to peeling because the stored strain energy in a pillar just before detachment is not available to drive detachment to the next fibril.
- Adaptability to rough surfaces (Fig. 3b): A patterned surface with long fibrils can adapt better to rough surfaces with less storage of elastic strain energy.
- Size effect due to surface-to-volume ratio (Fig. 3c): The volume, controlling the storage of elastic energy, decreases more rapidly than the gain in surface energy, favoring smaller contacts.
- Uniform stress distribution (Fig. 3d): Below a critical size of a contact, the stress distribution within the contact area becomes uniform, resulting in adhesion strengths near the theoretical values.
- Defect control and adhesion redundancy (Fig. 3e): If the adhesion is controlled by defects, smaller contacts will be more tolerant to defects than large contacts.

These considerations have led to the conclusion that a contact should be split into densely packed elements which should be as long and thin as possible to achieve maximum adhesion. However, if densely packed pillars exceed a critical aspect ratio (ratio between pillar length and diameter), they may become unstable with regard to "condensation" (also called "bundling" or "collapse"), as shown in Fig. 4. The reason for condensation of pillars is the low bending resistance combined with inter-pillar adhesion. This effect leads to a trade-off between pillar length, diameter, and elastic properties which can be conveniently displayed in "adhesion design maps" (Fig. 5). Illustrating the theoretical optimum in the parameter space, these maps show that, e.g., "condensation" is the overriding limitation for the refinement of a patterned adhesive. To adhere to surfaces with higher asperities, longer pillars would be needed, requiring large diameters to avoid condensation. However, there is a solution to this problem which can be found in the natural gecko adhesive: a hierarchical structure.
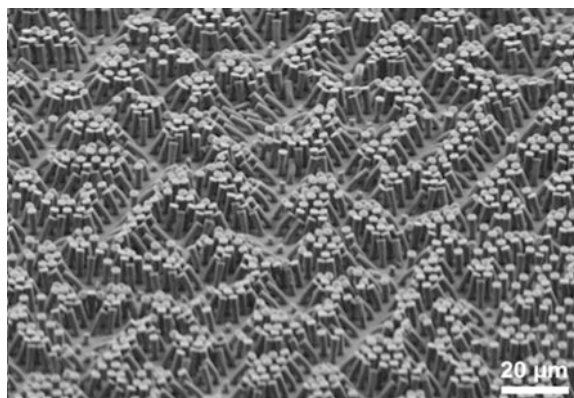
Gecko Adhesion, Fig. 4 "Condensation" of polydimethyl-siloxane pillars. At a certain aspect ratio of the pillars, the attractive forces between the single pillars overcome their resistance to bending, leading to a collapse of the structures (Picture reproduced with permission of C. Greiner)

## Role of Hierarchical Geometry for Gecko Adhesion Systems

On the macroscopic scale, a gecko has four legs with five toes each to adapt to large asperities in the range from centimeters to millimeters. On the microscopic scale, the *lamellae*, *setae*, and *spatulae* can adapt to asperities in the range of several tens of microns down to the sub-micrometer range. This hierarchical structure combines the advantages of small terminal contacts with stability against condensation. The influence of hierarchy on adhesion has been modeled repeatedly. Besides finite element method simulations and analytical calculations, models based on a hierarchical coupling of springs have been proposed.

Anisotropic Adhesion

Another key feature of geckos is their ability to rapidly detach their feet from a surface, which is a necessary feature for swift locomotion. The key to high adhesion and rapid detachment lies in the combination of their biomechanics of motion and the anisotropy of their adhesive structures. A closer look into their adhesive structures (Fig. 1) reveals slightly bent features with plate-like tips. When these features form contact with a surface under normal loading, the plate-like

Gecko Adhesion,
Fig. 5 Adhesion design map to predict successful surface patterns. The map spans a space containing material properties (Young's modulus) and geometrical parameters (radius of the fibers). Limitations are drawn in, such as fiber fracture (the adhesive strength is stronger than the material strength, causing the fiber to break rather than detach), ideal contact strength (maximum adhesion force generated by van der Waals forces), fiber condensation, and adaptability of the surface pattern. The *light blue lines* indicate the aspect ratio λ. The theoretical optimum can be found at the lower apex of the *orange triangle* [14] (Reproduced with permission of E. Arzt and Elsevier)
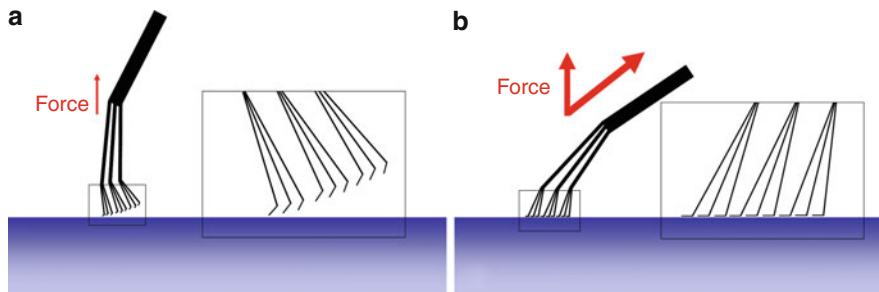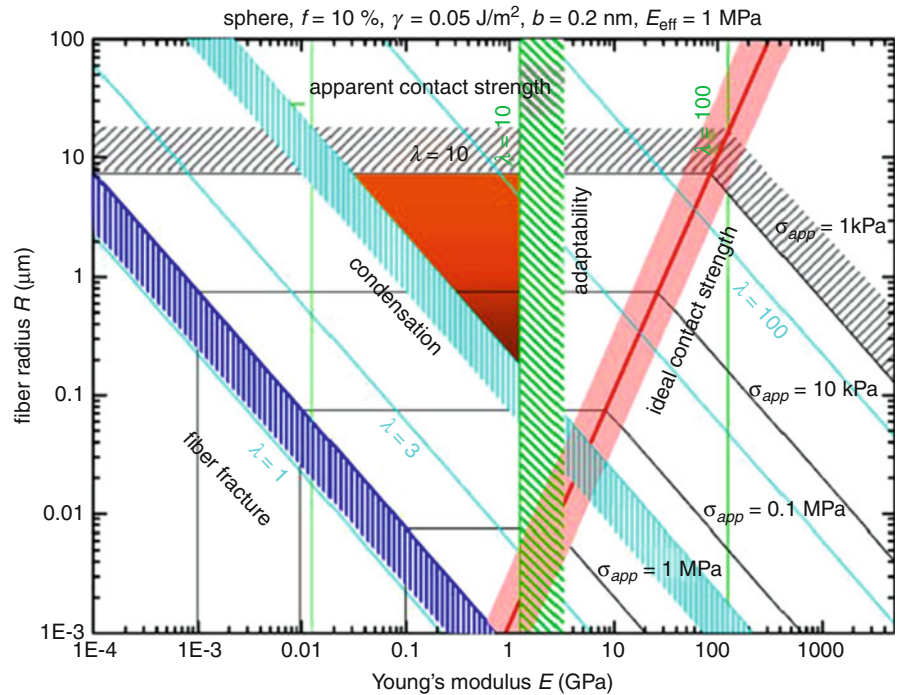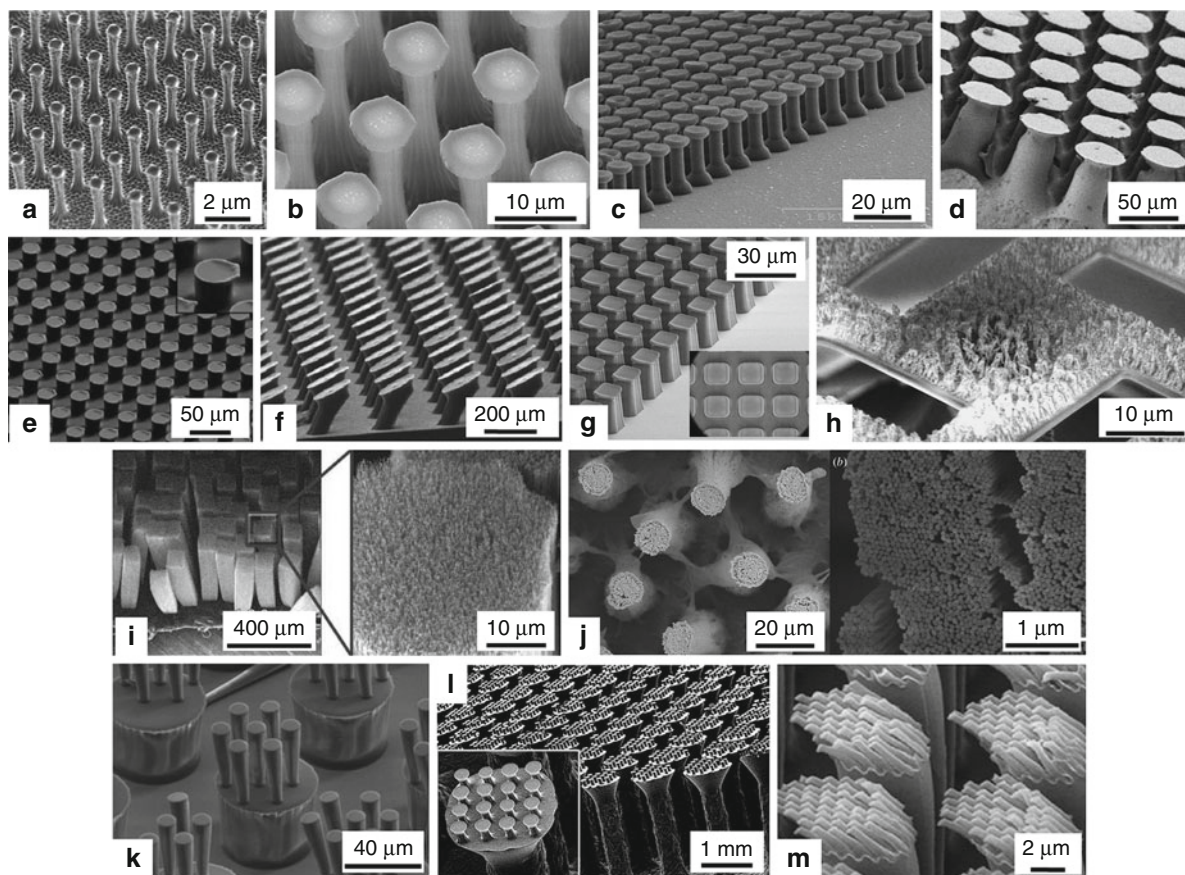

Gecko Adhesion, Fig. 6 Hierarchical gecko structures in contact with a smooth surface. If a normal force is applied, the tips of the hierarchical structures will be misaligned with the surface, leading to a low pull-off force (**a**). By applying a shear force, the structure tips become oriented, forming a large contact area and enhancing adhesion (**b**)

tips will be ill oriented to the surface and form insufficient contact (Fig. 6a) with low pull-off force. On application of a shear force, the structures will bend, allowing the plate-like tips to orient along the surface (Fig. 6b). The tips will then form a large contact area, enhancing adhesive forces. This so-called frictional adhesion allows the gecko to combine high sticking forces with fast detachment: Orientation of the tips by pulling induces adhesion, whereas pushing or a release of the pull causes rapid pull-off. Additionally, the special anatomy of the gecko toe allows hyperextension, i.e., rolling of the toes upward, which simplifies detachment further. This shear-induced adhesion also explains the paradox that a surface provides high adhesion and, at the same time, has self-cleaning properties, i.e., no dust or particles stick to the attachment pad; the attachment organs are only adhesive if a shear force is applied. In the absence of a shear force, the adhesion is low, hindering dust particles to adhere to the surface structure.

Gecko Adhesion, Fig. 7 Gecko-inspired adhesives, fabricated to achieve different structure geometry. (**a–d**): symmetric structures [15–18], (**e–g**): asymmetric structures [19–21], and (**h–m**) hierarchical structures [22–27] (Reprints with kind permission of: Nature Publishing Group (**a**), Elsevier (**b**, **h**), American Institute of Physics (**c**), The Royal Society Publishing (**d**), American Chemical Society (**e**, **l**), WILEY-VCH (**f**, **k**), IOP Publishing Ltd. (**g**, **j**), National Academy of Science, USA (**i**, **m**))

Artificial Gecko Adhesives

Over the last decade, scientists around the world have tried to mimic the gecko adhesion system with artificial structures, usually fabricated from polymeric material. The fabrication methods reflect the strategies employed in micro fabrication: bottom-up patterning (assembling structures from atomic or molecular materials), template-less top-down patterning (e.g., removing material from bulk) and top-down patterning using a template (e.g., molding of pre-patterned surfaces). Figure 7 shows examples of recently fabricated gecko-inspired adhesives using different processes.

An example for bottom-up patterning is the fabrication of oriented carbon nanotubes on a substrate. Although carbon nanotube-based adhesives can provide high shear forces, the compressive preload necessary to trigger adhesion sometimes even exceeds the resulting normal and shear force.

Template-less top-down methods are based on either irradiation techniques to cross-link or decompose a polymer, such as UV light, lasers, electron or ion beams, or on "mechanical" methods, i.e., cutting, writing, or melting of polymers. These techniques allow high precision in fabrication of surface structures, and features below 1 μm are accessible. Although some techniques allow the fabrication of larger areas in a short time (e.g., interference lithography), template-less top-down techniques are usually serial processes, where one structure is fabricated after another. This largely increases the fabrication time to achieve samples in feasible size.

The most widely used fabrication techniques are based on top-down structuring using a template. A template is usually fabricated by photolithography and is filled by a liquid polymer ("molding"). Although this process is usually more complex than the template-less techniques, it has a large benefit: the template may be used many times. If desired, the fabricated samples may be processed further, especially to modify the tip geometry of the pillars. Del Campo et al. [28] showed systematically in 2007 that the tip geometry of bioinspired adhesive structures is crucial to obtain high adhesive forces. They found that pillars with spherical, flat punch, or suction-cup tips showed relatively low adhesive forces, while spatular or mushroom-shaped tips showed high adhesion. To mimic the gecko adhesion system to its full extent, several of the fabrication processes presented above may be repeated to obtain hierarchical structures with controlled tip geometry. A current overview of fabrication methods to obtain patterned surfaces is given in the recently published book by del Campo and Arzt [29]. A review of recently fabricated gecko-like structures with high adhesion was published by Samento and Menon in 2010 [30].

Although the fabrication of gecko-inspired adhesives is already well advanced in the laboratory, there is still plenty of room to improve the performance of such artificial dry adhesives. For example, experimental studies of the adhesion performance on rough surfaces – ideal territory for a gecko – are rare. In addition, surface specimens with high adhesion seldom exceeded several square centimeters in area and the methods used are not feasible for large-scale fabrication. To establish a product based on gecko-inspired adhesives, new processes are now being developed which allow fabrication of surfaces with highly controlled structures in a cost-efficient way.

Applications

Possible applications of gecko-inspired adhesives can be identified by exploiting their key properties; i.e., high reversible adhesion, easy detachment, residue-free contact, and self-cleaning properties. Such adhesives are being implemented into a locomotion system for robots enabling them to overcome obstacles. Such a robot would be able to access terrain which would be difficult or dangerous for humans. Another related

application would be the development of "gecko-grabbers," e.g., in fabrication plants. It is likely that gecko adhesives could find a large field of application in household or sports articles, e.g., gecko-fastening strips or goalkeeper gloves.

A true breakthrough for gecko-inspired adhesives, however, would be in the class of "intelligent materials." These new kind of adhesives would not only have the function of high reversible (and switchable) adhesion, but would combine these properties with other functions. For example, gecko-inspired adhesives could be designed in such a way that they adhere *selectively* to different materials, allow a reversible adhesion to glass and serve as an optical coupling at the same time, or stick to skin and decompose with time, which could replace conventional plasters. Research over the last decade has laid the foundations for such ideas; now technological development must follow to make them come true.

Cross-References

▶ Adhesion in Wet Environments: Frogs
▶ Bioadhesion
▶ Bioadhesives
▶ Biomimetics
▶ Biomimetics of Marine Adhesives
▶ Biopatterning
▶ Disjoining Pressure and Capillary Adhesion
▶ Gecko Effect

References

1. Hiller, U., Blaschke, R.: Zum Haftproblem der Gecko-Füsse. Naturwissenschaften **54**, 344–345 (1967)
2. Stork, N.E.: Experimental analysis of adhesion of *Chrysolina polita* (Chrysomelidae: Coleoptera) on a variety of surfaces. J. Exp. Biol. **88**, 91–107 (1980)
3. Autumn, K., Gravish, N.: Gecko adhesion: evolutionary nanotechnology. Phil. Trans. R. Soc. A **366**, 1575–1590 (2008)
4. Autumn, K., Liang, Y.A., Hsieh, S.T., Zesch, W., Chan, W.P., Kenny, T.W., Fearing, R., Full, R.J.: Adhesive force of a single gecko foot-hair. Nature **405**, 681–685 (2000)
5. Gorb, S.N.: Functional Surfaces in Biology 2: Adhesion Related Phenomena, vol. 2. Springer, Heidelberg (2010)
6. Huber, G., Mantz, H., Spolenak, R., Mecke, K., Jacobs, K., Gorb, S.N., Arzt, E.: Evidence for capillarity contributions to gecko adhesion from single spatula nanomechanical

measurements. Proc. Natl. Acad. Sci. USA **102**, 16293–16296 (2005)

7. Puthoff, J.B., Prowse, M.S., Wilkinson, M., Autumn, K.: Changes in materials properties explain the effects of humidity on gecko adhesion. J. Exp. Biol. **213**, 3699–3704 (2010)

8. Johnson, K.L., Kendall, K., Roberts, A.D.: Surface energy and the contact of elastic solids. Proc. R. Soc. Lond. A. **324**, 301–313 (1971)

9. Johnson, K.L.: Contact Mechanics. Cambridge University Press, Cambridge (1987)

10. Maugis, D.: Contact, Adhesion and Rupture of Elastic Solids. Springer, Heidelberg (2000)

11. Arzt, E., Gorb, S., Spolenak, R.: From micro to nano contacts in biological attachment devices. Proc. Natl. Acad. Sci. USA **100**, 10603–10606 (2003)

12. Federle, W.: Why are so many adhesive pads hairy? J. Exp. Biol. **219**, 2611–2621 (2006)

13. Kamperman, M., Kroner, E., del Campo, A., McMeeking, R.M., Arzt, E.: Functional adhesive surfaces with "Gecko Effect": the concept of contact splitting. Adv. Eng. Mater. **12**, 335–348 (2010)

14. Spolenak, R., Gorb, S., Arzt, E.: Adhesion design maps for bio-inspired attachment systems. Acta Biomater. **1**, 5–13 (2005)

15. Geim, A., Dubonos, S., Grigorieva, I., Novoselov, K., Zhukov, A., Shapoval, S.: Microfabricated adhesive mimicking gecko foot-hair. Nat. Mater. **2**(7), 461–463 (2003)

16. Davies, J., Haq, S., Hawke, T., Sargent, J.P.: A practical approach to the development of a synthetic Gecko tape. Int. J. Adhes. Adhes. **29**, 380–390 (2009)

17. Kim, S., Sitti, M.: Biologically inspired polymer microfibers with spatulate tips as repeatable fibrillar adhesives. Appl. Phys. Lett. **89**, 261911 (2006)

18. Varenberg, M., Gorb, S.: Close-up of mushroom-shaped fibrillar adhesive microstructure: contact element behaviour. J. R. Soc. Interface **5**, 785–789 (2008)

19. Del Campo, A., Greiner, C., Arzt, E.: Contact shape controls adhesion of bioinspired fibrillar surfaces. Langmuir **23**, 10235–10243 (2007)

20. Murphy, M.P., Aksak, B., Sitti, M.: Gecko-inspired directional and controllable adhesion. Small **5**(2), 170–175 (2009)

21. Sameoto, D., Menon, C.: Direct molding of dry adhesives with anisotropic peel strength using an offset lift-off photoresist mold. J. Micromech. Microeng. **19**, 115026 (2009)

22. Northen, M.T., Turner, K.L.: Meso-scale adhesion testing of integrated micro- and nano-scale structures. Sensor. Actuator. A **130–131**, 583–587 (2006)

23. Ge, L., Sethi, S., Ci, L., Ajayan, P.M., Dhinojwala, A.: Carbon nanotube-based synthetic gecko tapes. Proc. Natl. Acad. Sci. USA **104**(26), 10792–10795 (2007)

24. Kustadi, T.S., Samper, V.D., Ng, W.S., Chong, A.S., Gao, H.: Fabrication of a gecko-like hierarchical fibril array using a bonded porous alumina template. J. Micromech. Microeng. **17**, N75–N81 (2007)

25. Del Campo, A., Arzt, E.: Design parameters and current fabrication approaches for developing bioinspired dry adhesives. Macromol. Biosci. **7**, 118–127 (2007)

26. Murphy, M.P., Kim, S., Sitti, M.: Enhanced adhesion by Gecko-inspired hierarchical fibrillar adhesives. Appl. Mater. Interface **1**(4), 849–855 (2009)

27. Jeong, H.E., Lee, J.-K., Kim, H.N., Moon, S.H., Suh, K.Y.: A nontransferring dry adhesive with hierarchical polymer nanohairs. Proc. Natl. Acad. Sci. USA **106**(14), 5639–5644 (2009)

28. del Campo, A., Greiner, C., Arzt, E.: Contact shape controls adhesion of bioinspired fibrillar surfaces. Langmuir **23**, 10235–10243 (2007)

29. del Campo, A., Arzt, E.: Generating Micro and Nano Patterns on Polymeric Materials. Wiley-VCH, Weinheim (2011)

30. Samento, D., Menon, C.: Recent advances in the fabrication and adhesion testing of biomimetic dry adhesives. Smart. Mater. Struct. **19**, 103001 (2010)

Gecko Effect

Bharat Bhushan
Nanoprobe Laboratory for Bio- & Nanotechnology and Biomimetics, The Ohio State University, Columbus, OH, USA

Synonyms

Gecko feet; Reversible adhesion

Definition

Geckos have the largest mass and have developed the most complex hairy attachment structures among climbing animals that are capable of smart adhesion – the ability to cling to different smooth and rough surfaces and detach at will. These animals make use of about three million microscale hairs (setae) (about 14,000/mm^2) that branch off into nanoscale spatulae, about three billion spatulae on two feet. The so-called division of contacts provides high dry adhesion. Multiple-level hierarchically structured surface construction provides the gecko with the compliance and adaptability to create a large real area of contact with a variety of surfaces.

Overview

The leg attachment pads of several animals, including many insects, spiders, and lizards, are capable of attaching to and detaching from a variety of surfaces

Fig. 1 Terminal elements of
the hairy attachment pads of
a beetle, fly, spider, and gecko
shown at two different
scales [13]
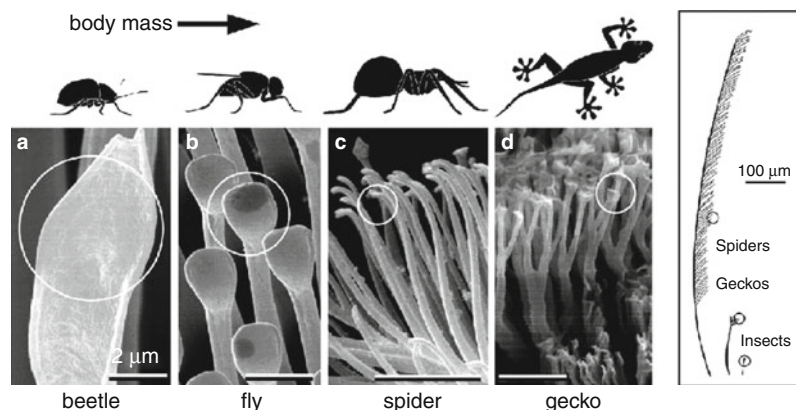
and are used for locomotion, even on vertical walls or across the ceiling [1–4]. Biological evolution over a long period of time has led to the optimization of their leg attachment systems. This dynamic attachment ability is referred to as reversible adhesion or smart adhesion [5]. Many insects (e.g., beetles and flies) and spiders have been the subject of investigation. However, the attachment pads of geckos have been the most widely studied due to the fact that they have the highest body mass and exhibit the most versatile and effective adhesive known in nature.

Although there are over 1,000 species of geckos [6, 7] that have attachment pads of varying morphology [8], the Tokay gecko (*Gekko gecko*) which is native to southeast Asia, has been the main focus of scientific research [9–11]. The Tokay gecko is the second largest gecko species, attaining lengths of approximately 0.3–0.4 m and 0.2–0.3 m for males and females, respectively. They have a distinctive blue or gray body with orange or red spots and can weigh up to 300 g [12]. These have been the most widely investigated species of gecko due to the availability and size of these creatures.

There is great interest among the scientific community to further study the characteristics of gecko feet in the hope that this information could be applied to the production of micro/nanosurfaces capable of recreating the adhesion forces generated by these lizards [2–4]. Common man-made adhesives such as tape or glue involve the use of wet adhesives that permanently attach two surfaces. However, replication of the characteristics of gecko feet would enable the development of a superadhesive tape capable of clean, dry adhesion. These structures can bind components in microelectronics without the high
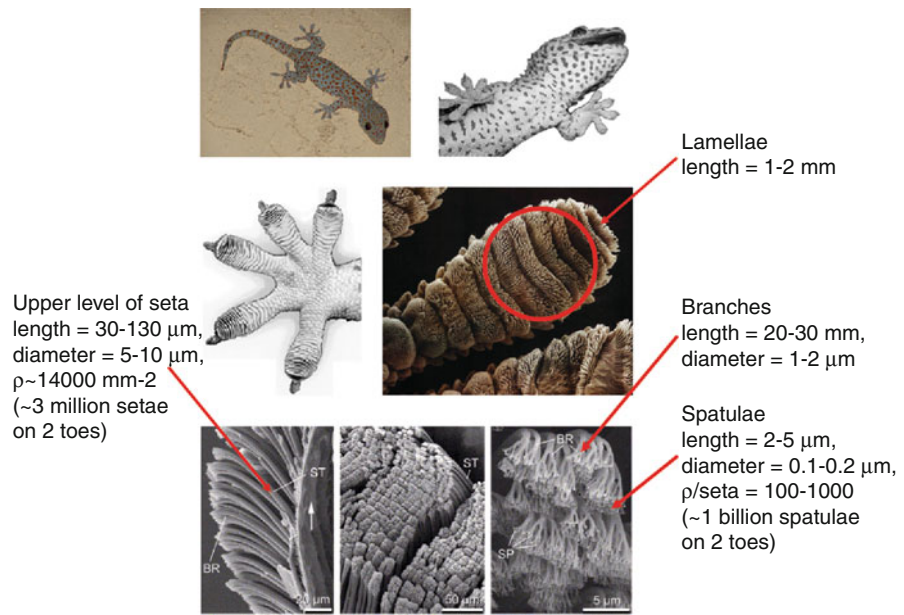
heat associated with various soldering processes. These structures will never dry out in a vacuum – a common problem in aerospace applications. They have the potential for use in everyday objects such as adhesive tapes, fasteners, and toys and in high technology such as microelectronic and space applications. Replication of the dynamic climbing and peeling ability of geckos could find use in the treads of wall-climbing robots.

Tokay Gecko

There are two kinds of attachment pads – relatively smooth and hairy. Relatively smooth pads, so-called arolia and euplantulae, are soft and deformable and are found in tree frogs, cockroaches, grasshoppers, and bugs. The hairy types consist of long deformable setae and are found in many insects (e.g., beetles, flies), spiders, and lizards. The microstructures utilized by beetles, flies, spiders, and geckos have similar structures and can be seen in Fig. 1. As the size (mass) of the creature increases, the radius of the terminal attachment elements decreases. This allows a greater number of setae to be packed into an area, hence increasing the linear dimension of contact and the adhesion strength [13, 14].

The explanation for the adhesion properties of gecko feet can be found in the surface morphology of the skin on the toes of the gecko. The skin is comprised of a complex hierarchical structure of lamellae, setae, branches, and spatulae [8]. Figure 2 shows various SEM micrographs of a gecko foot, showing the hierarchical structure down to the nanoscale. Figure 3 shows a schematic of the structure, and Table 1 summarizes the surface characteristics.

Gecko Effect, Fig. 2 The hierarchical structures of a Tokay gecko foot: a gecko foot [35] and a gecko toe [11]. Two feet contain about three million setae on their toes that branch off into about three billion spatulae. Scanning electron microscope (SEM) micrographs of the setae and the spatulae in the bottom row [54]. *ST* seta, *SP* spatula, *BR* branch

Lamellae
length = 1-2 mm

Upper level of seta
length = 30-130 μm,
diameter = 5-10 μm,
ρ~14000 mm-2
(~3 million setae
on 2 toes)

Branches
length = 20-30 mm,
diameter = 1-2 μm

Spatulae
length = 2-5 μm,
diameter = 0.1-0.2 μm,
ρ/seta = 100-1000
(~1 billion spatulae
on 2 toes)

The gecko attachment system consists of an intricate hierarchy of structures beginning with lamellae, soft ridges 1–2 mm in length [8] that are located on the attachment pads (toes) that compress easily so that contact can be made with rough, bumpy surfaces. Tiny curved hairs known as setae extend from the lamellae with a density of approximately 14,000 per square millimeter [15]. These setae are typically 30–130 μm in length, 5–10 μm in diameter [8, 9, 16, 17], and are composed primarily of β-keratin [18, 19] with some α-keratin component [20]. At the end of each seta, 100–1000 spatulae [8, 9] with typically 2–5 μm in length and a diameter of 0.1–0.2 μm [8] branch out and form the points of contact with the surface. The tips of the spatulae are approximately 0.2–0.3 μm in width [8], 0.5 μm in length, and 0.01 μm in thickness [21] and garner their name from their resemblance to a spatula.
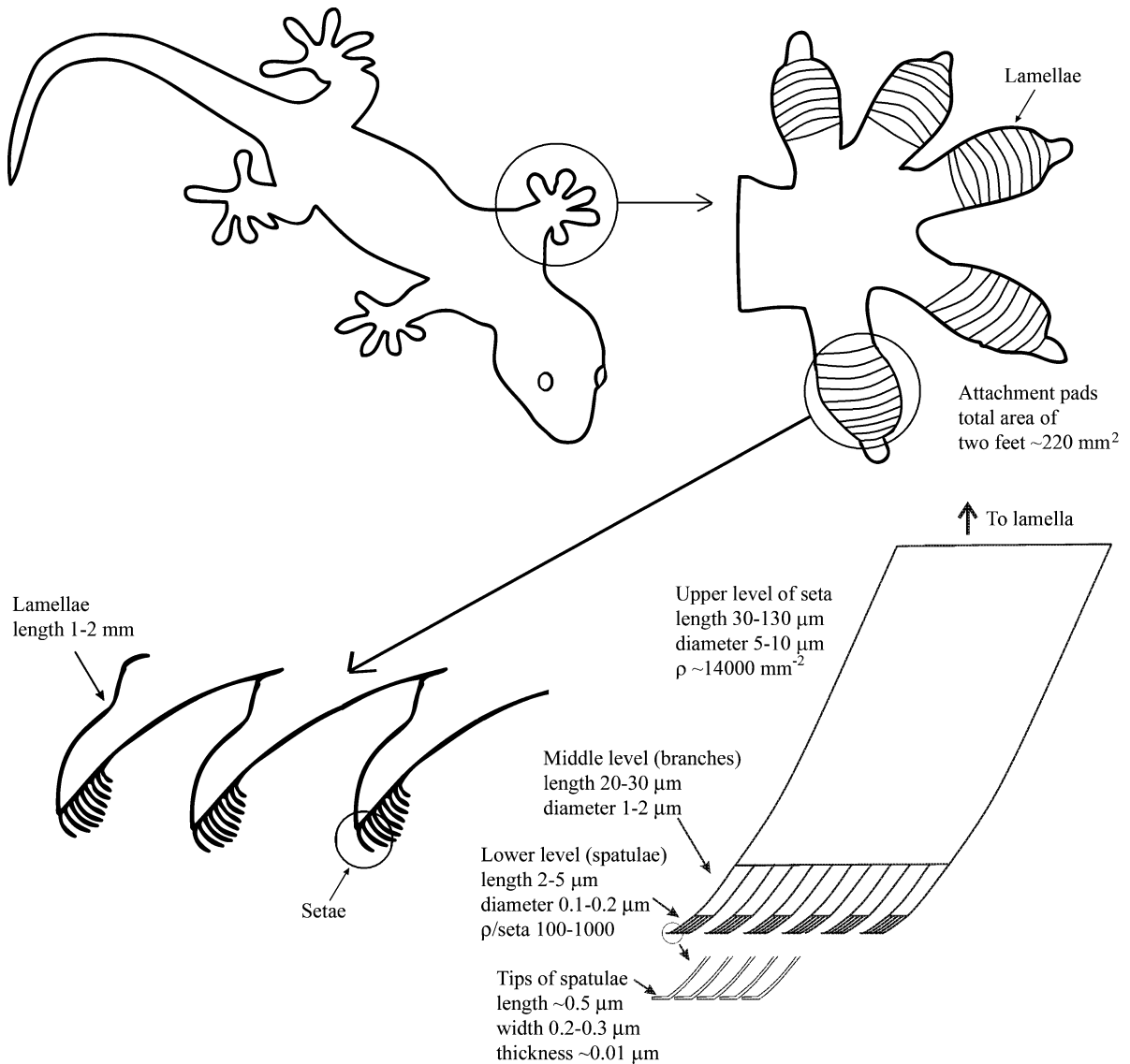
The attachment pads on two feet of the Tokay gecko have an area of about 220 mm^2. About three million setae on their toes that branch off into about three billion spatulae on two feet can produce a clinging ability of about 20N required to pull a lizard down a nearly vertical (85°) surface [10] and allow them to climb vertical surfaces at speeds over 1 m/s with the capability to attach and detach their toes in milliseconds. In isolated setae, a 2.5 μN preload yielded adhesion of 20–40 μN, and thus the adhesion

coefficient, which represents the strength of adhesion as a function of preload, ranges from 8 to 16 [22].

Typical rough, rigid surfaces are able to make intimate contact with a mating surface only over a very small portion of the perceived apparent area of contact. In fact, the real area of contact is typically two to six orders of magnitude less than the apparent area of contact [4, 23, 24]. Autumn et al. [22] proposed that divided contacts serve as a means for increasing adhesion. The surface energy approach can be used to calculate adhesion force in the dry environment in order to calculate the effect of division of contacts. Adhesion force of a single contact is proportional to a linear dimension of the contact. For a constant area divided into a large number of contacts or setae, the adhesion force increases linearly with the square root of the number of contacts (self-similar scaling) [13].

Attachment Mechanisms

When asperities of two solid surfaces are brought into contact with each other, chemical and/or physical attractions occur. The force developed that holds the two surfaces together is known as adhesion. In a broad sense, adhesion is considered to be either physical or chemical in nature [3, 4, 23–29]. Chemical interactions such as electrostatic attraction charges [30] as well as intermolecular forces [9] including van der Waals and capillary forces have all been proposed as potential

Lamellae

Attachment pads
total area of
two feet ~220 mm^2

To lamella

Upper level of seta
length 30-130 µm
diameter 5-10 µm
ρ ~14000 mm^{-2}

Lamellae
length 1-2 mm

Middle level (branches)
length 20-30 µm
diameter 1-2 µm

Lower level (spatulae)
length 2-5 µm
diameter 0.1-0.2 µm
ρ/seta 100-1000

Setae

Tips of spatulae
length ~0.5 µm
width 0.2-0.3 µm
thickness ~0.01 µm

Gecko Effect, Fig. 3 Schematic of a Tokay gecko including the overall body, one foot, a cross-sectional view of the lamellae, and an individual seta. ρ represents number of spatulae

adhesion mechanisms in gecko feet. Others have hypothesized that geckos adhere to surfaces through the secretion of sticky fluids [31, 32], suction [32], increased frictional force [33], and microinterlocking [34].

Through experimental testing and observations conducted over the last century and a half many potential adhesive mechanisms have been eliminated. Observation has shown that geckos lack any glands capable of producing sticky fluids [31, 32], thus ruling out the secretion of sticky fluids as a potential adhesive

mechanism. Furthermore, geckos are able to create large adhesive forces normal to a surface. Since friction only acts parallel to a surface, the attachment mechanism of increased frictional force has been ruled out. Dellit [34] experimentally ruled out suction and electrostatic attraction as potential adhesive mechanisms. Experiments carried out in vacuum did not show a difference between the adhesive forces at low pressures compared to ambient conditions. Since adhesive forces generated during suction are

Gecko Effect, Table 1 Surface characteristics of Tokay gecko feet (Young's modulus of surface material, keratin = 1–20 GPa[a,b])

Component	Size	Density	Adhesive force
Seta	30–130[c–f]/5–10[c–f] length/diameter (μm)	~14,000[h,i] setae/mm^2	194 μN[j] (in shear) ~20 μN[j] (normal)
Branch	20–30[c]/1–2[c] length/diameter (μm)	–	–
Spatula	2–5[c]/0.1–0.2[c,g] length/diameter (μm)	100–1000[c,d] spatulae per seta	–
Tip of spatula	~0.5[c,g]/0.2–0.3[c,f]/~0.01[g] length/width/ thickness (μm)	–	11 nN[k] (normal)

[a]Russell [19]
[b]Bertram and Gosline [51]
[c]Ruibal and Ernst [8]
[d]Hiller [9]
[e]Russell [16]
[f]Williams and Peterson [17]
[g]Persson and Gorb [21]
[h]Schleich and Kästle [15]
[i]Autumn and Peattie [52]
[j]Autumn et al. [35]
[k]Huber et al. [53]

based on pressure differentials, which are insignificant under vacuum, suction was rejected as an adhesive mechanism [34]. Additional testing utilized X-ray bombardment to create ionized air in which electrostatic attraction charges would be eliminated. It was determined that geckos were still able to adhere to surfaces in these conditions, and therefore, electrostatic charges could not be the sole cause of attraction [34]. Autumn et al. [35] demonstrated the ability of a gecko to generate large adhesive forces when in contact with a molecularly smooth SiO_2 microelectromechanical system (MEMS) semiconductor. Since surface roughness is necessary for microinterlocking to occur, it has been ruled out as a mechanism of adhesion. Two mechanisms, van der Waals forces and capillary forces, remain as the potential sources of gecko adhesion.

## Multilevel Hierarchy for Adaptability to Rough Surfaces

In order to study the effect of the number of hierarchical levels in the attachment system on attachment ability, models with one [5, 36, 37], two [5, 36, 37], and three [36, 37] levels of hierarchy were simulated (Fig. 4). The random rough surfaces used in the simulations were generated by a computer program [23, 24]. Figure 5 shows the calculated spring force–distance curves for the one-, two- and three-level hierarchical models in contact with rough surfaces of different values of root mean square (RMS) amplitude σ ranging from $\sigma = 0.01$ μm to

$\sigma = 10$ μm at an applied load of 1.6 μN which was derived from the gecko's weight. When the spring model is pressed against the rough surface, contact between the spring and the rough surface occurs at point A; as the spring tip presses into the contacting surface, the force increases up to point B, B′, or B″. During pull off, the spring relaxes, and the spring force passes an equilibrium state (0 N); tips break free of adhesion forces at point C, C′, or C″ as the spring moves away from the surface. The perpendicular distance from C, C′, or C″ to zero is the adhesion force. Adhesion energy stored during contact can be obtained by calculating the area of the triangle during the unloading part of the curves. The data in Fig. 5 show that adhesion forces for various levels of hierarchy are comparable when in contact with the smooth surface. However, the adhesion force for the three-level hierarchy is the largest when in contact with rough surface. Thus, multilevel hierarchy facilitates adaptability to a variety of rough surfaces.

Peeling

Although geckos are capable of producing large adhesion forces, they retain the ability to remove their feet from an attachment surface at will by peeling action. The orientation of the spatulae facilitates peeling. Autumn et al. [35] were the first to experimentally show that the adhesion force of gecko setae is dependent on the three-dimensional orientation as well as the preload applied during attachment. Due to this fact, geckos have developed a complex foot

One-level hierarchy

Two-level hierarchy

Three-level hierarchy

Gecko Effect, Fig. 4 One-, two-, and three-level hierarchical spring models for simulating the effect of hierarchical morphology on interaction of a seta with a rough surface. In this figure, $l_{I, II, III}$ are lengths of structures, s_I is space between spatulae, $k_{I, II, III}$ are stiffnesses of structures, *I, II*, and *III* are level indexes, *R* is radius of tip, and *h* is distance between upper spring base of each model and mean line of the rough profile [36]
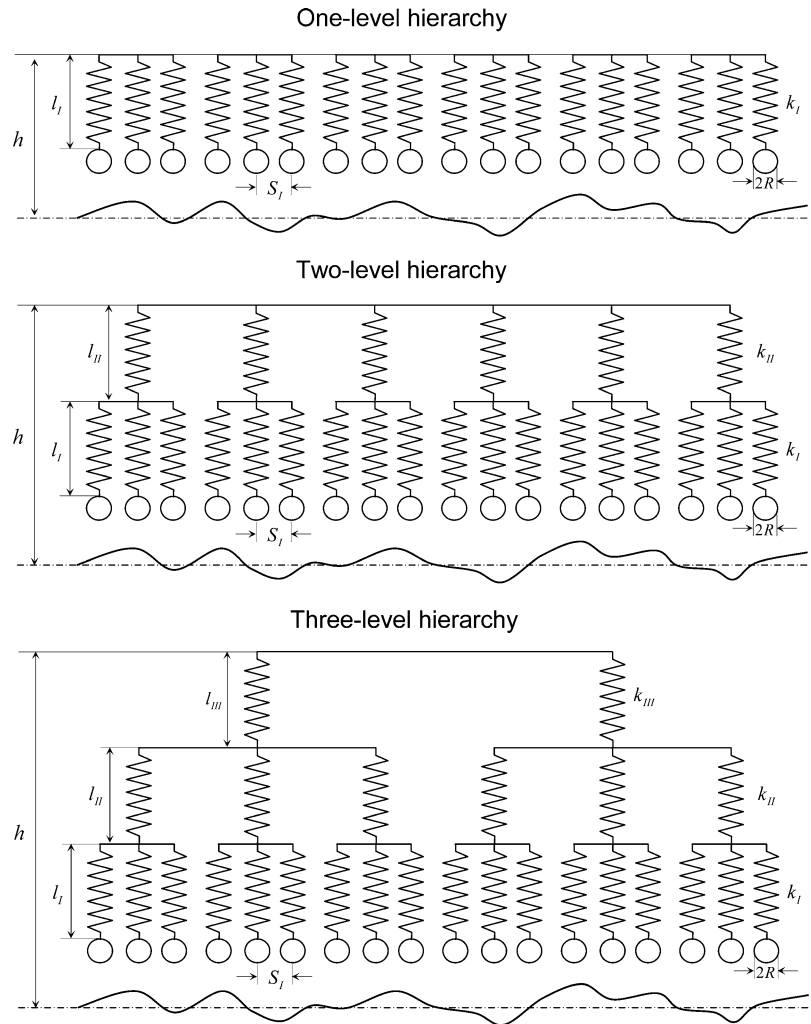
motion during walking. First, the toes are carefully uncurled during detachment. The maximum adhesion occurs at an attachment angle of 30º – the angle between a seta and mating surface. The gecko is then able to peel its foot from surfaces one row of setae at a time by changing the angle at which its setae contact a surface. At an attachment angle greater than 30º, the gecko will detach from the surface.

Self-cleaning

Natural contaminants (dirt and dust) as well as man-made pollutants are unavoidable and have the potential to interfere with the clinging ability of geckos. Particles found in the air consist of particulates that are typically less than 10 µm in diameter while

those found on the ground can often be larger [38, 39]. Intuitively, it seems that the great adhesion strength of gecko feet would cause dust and other particles to become trapped in the spatulae and that they would have no way of being removed without some sort of manual cleaning action on behalf of the gecko. However, geckos are not known to groom their feet like beetles [40], nor do they secrete sticky fluids to remove adhering particles like ants [41] and tree frogs [42], yet they retain adhesive properties. One potential source of cleaning is during the time when the lizards undergo molting, or the shedding of the superficial layer of epidermal cells. However, this process only occurs approximately once per month [43]. If molting were the sole source of cleaning, the gecko would

The effect of multi-level hierarchical structure

$F_n = 1.6$ μN, $\beta^* = 200$ μm
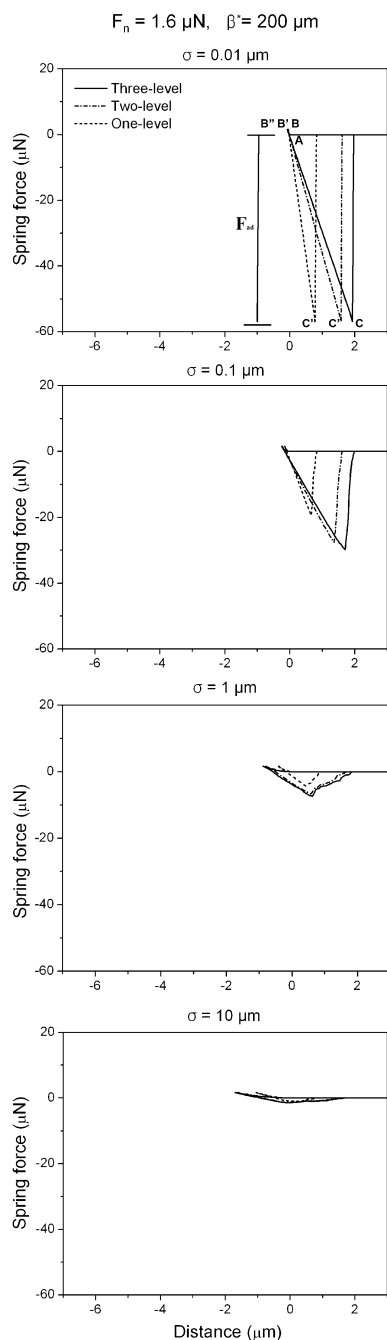
Gecko Effect, Fig. 5 Force–distance curves of one-, two-, and three-level models in contact with rough surfaces with different σ values for an applied load of 1.6 μN

rapidly lose its adhesive properties as it was exposed to contaminants in nature [44]. They reported that gecko setae become cleaner with repeated use – a phenomena known as self-cleaning.

Fabrication of Biomimetic Gecko Skin

Based on the studies reported in the literature, the dominant adhesion mechanism utilized by gecko and some spider attachment systems appears to be van der Waals forces. The hierarchical structure involving complex divisions of the gecko skin (lamellae-setae-branches-spatulae) enable a large number of contacts between the gecko skin and mating surface. These hierarchical fibrillar microstructured surfaces would be capable of reusable dry adhesion and would have uses in a wide range of applications from everyday objects such as adhesive tapes, fasteners, toys, microelectronic and space applications, and treads of wall-climbing robots. The development of nanofabricated surfaces capable of replicating this adhesion force developed in nature is limited by current fabrication methods. Many different techniques have been used in an attempt to create and characterize bio-inspired adhesive tapes. Attempts are being made to develop climbing robots using gecko-inspired structures [45–48].

As an example, Geim et al. [49] created arrays of polyimide nanofibers using electron-beam lithography and dry etching in oxygen plasma (Fig. 6a (left)). By using electron-beam lithography, thermal evaporation of an aluminum film, and lift-off, an array of nanoscale aluminum disks were prepared. These patterns were then transferred in the polyimide film by dry etching in oxygen plasma. A 1 cm^2 sample was able to create 3 N of adhesive force under the new arrangement. This is approximately one-third the adhesive strength of a gecko. They fabricated a Spiderman toy (about 0.4 N) with a hand covered with molded polymer nanohairs, which could support its weight Fig. 6b.

As another example, Bhushan and Sayer [50] characterized the surface roughness, friction force, and contact angle of a structured sample and compared the results to an unstructured control. The macroscale coefficient of kinetic friction of the structured sample was found to be almost four times greater than the unstructured sample. This increase was determined to be a result of the structured roughness of the sample and not the random nanoroughness. It is also noteworthy that the static and kinetic coefficients of friction were approximately equal for the structured sample. It is believed that the divided contacts allow the broken contacts of the structured sample to constantly recreate contact.

Gecko Effect, Fig. 6 (**a**) An array of polyimide nanohairs and (**b**) a Spiderman toy (about 0.4 N) with a hand covered with the molded polymer nanohairs, clinging to a glass plate [49]
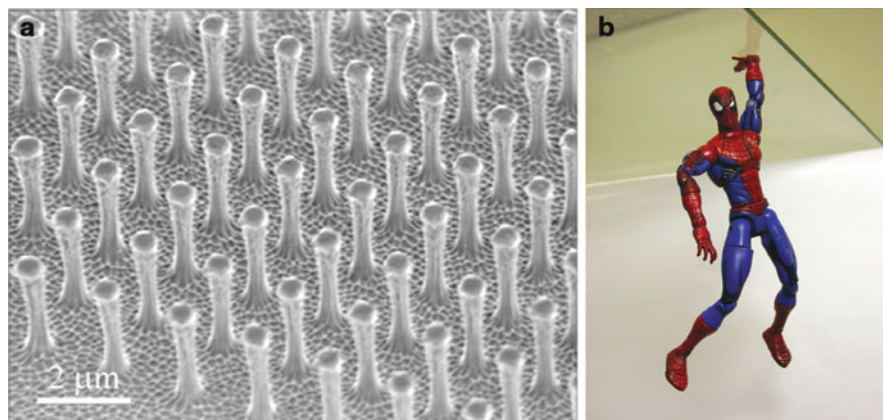
Closure

The adhesive properties of geckos and other creatures such as flies, beetles, and spiders are due to the hierarchical structures present on each creature's hairy attachment pads. Geckos have developed the most intricate adhesive structures of any of the aforementioned creatures. The attachment system consists of ridges called lamellae that are covered in microscale setae that branch off into nanoscale spatulae, about three billion spatulae on two feet. The so-called division of contacts provides high dry adhesion. Multiple level hierarchically structured surface construction plays an important role in adapting to surface roughness bringing the spatulae in close proximity with the mating surface. These structures, as well as material properties, allow the gecko to obtain a much larger real area of contact between its feet and a mating surface than is possible with a non-fibrillar material. Two feet of a Tokay gecko have about 220 mm^2 of attachment pad area on which the gecko is able to generate approximately 20 N of adhesion force. Although capable of generating high adhesion forces, a gecko is able to detach from a surface at will – an ability known as smart adhesion. Detachment is achieved by a peeling motion of the gecko's feet from a surface.

There is a great interest among the scientific community to create surfaces that replicate the adhesion strength of gecko feet. These hierarchical fibrillar microstructured surfaces would be capable of reusable dry adhesion and would have uses in a wide range of applications from everyday objects such as adhesive tapes, fasteners, toys, microelectronic and space applications, and treads of wall-climbing robots.

References

1. Gorb, S.: Attachment Devices of Insect Cuticles. Kluwer, Dordrecht (2001)
2. Bhushan, B.: Adhesion of multi-level hierarchical attachment systems in gecko feet. J. Adhes. Sci. Technol. **21**, 1213–1258 (2007)
3. Bhushan, B. (ed.): Springer Handbook of Nanotechnology, 3rd edn. Springer, Heidelberg (2010)
4. Bhushan, B. (ed.): Nanotribology and Nanomechanics I – Measurement Techniques and Nanomechanics, II – Nanotribology, Biomimetics, and Industrial Applications, 3rd edn. Springer, Heidelberg (2011)
5. Bhushan, B., Peressadko, A.G., Kim, T.W.: Adhesion analysis of two-level hierarchical morphology in natural attachment systems for "smart adhesion". J. Adhes. Sci. Technol. **20**, 1475–1491 (2006)
6. Kluge, A.G.: Gekkotan lizard taxonomy. Hamadryad **26**, 1–209 (2001)
7. Han, D., Zhou, K., Bauer, A.M.: Phylogenetic relationships among gekkotan lizards inferred from C-mos nuclear DNA sequences and a new classification of the Gekkota. Biol. J. Linn. Soc. **83**, 353–368 (2004)
8. Ruibal, R., Ernst, V.: The structure of the digital setae of lizards. J. Morphol. **117**, 271–294 (1965)
9. Hiller, U.: Untersuchungen zum Feinbau und zur Funktion der Haftborsten von Reptilien. Z. Morphol. Tiere **62**, 307–362 (1968)
10. Irschick, D.J., Austin, C.C., Petren, K., Fisher, R.N., Losos, J.B., Ellers, O.: A comparative analysis of clinging ability among pad-bearing lizards. Biol. J. Linn. Soc. **59**, 21–35 (1996)
11. Autumn, K.: How Gecko toes stick. Am. J. Scient. **94**, 124–132 (2006)
12. Tinkle, D.W.: Gecko. In: Cummings, M. (ed.) Encylcopedia Americana, vol 12, 359 p. Grolier, Norwich (1992)
13. Arzt, E., Gorb, S., Spolenak, R.: From micro to nano contacts in biological attachment devices. Proc. Natl. Acad. Sci. USA **100**, 10603–10606 (2003)
14. Federle, W.: Why are so many adhesive Pads Hairy? J. Exp. Biol. **209**, 2611–2621 (2006)
15. Schleich, H.H., Kästle, W.: Ultrastrukturen an Gecko-Zehen. Amphibia Reptilia **7**, 141–166 (1986)

16. Russell, A.P.: A contribution to the functional morphology of the foot of the tokay, Gekko gecko. J. Zool. Lond. **176**, 437–476 (1975)

17. Williams, E.E., Peterson, J.A.: Convergent and alternative designs in the digital adhesive pads of scincid lizards. Science **215**, 1509–1511 (1982)

18. Maderson, P.F.A.: Keratinized epidermal derivatives as an aid to climbing in gekkonid lizards. Nature **2003**, 780–781 (1964)

19. Russell, A.P.: The morphological basis of weight-bearing in the scansors of the tokay gecko. Can. J. Zool. **64**, 948–955 (1986)

20. Rizzo, N., Gardner, K., Walls, D., Keiper-Hrynko, N., Hallahan, D.: Characterization of the structure and composition of gecko adhesive setae. J. Royal Soc. Interf. **3**, 441–451 (2006)

21. Persson, B.N.J., Gorb, S.: The effect of surface roughness on the adhesion of elastic plates with application to biological systems. J. Chem. Phys. **119**, 11437–11444 (2003)

22. Autumn, K., Sitti, M., Liang, Y.A., Peattie, A.M., Hansen, W.R., Sponberg, S., Kenny, T.W., Fearing, R., Israelachvili, J.N., Full, R.J.: Evidence for van der Waals adhesion in gecko setae. Proc. Natl. Acad. Sci. USA **99**, 12252–12256 (2002)

23. Bhushan, B.: Principles and Applications of Tribology. Wiley, New York (1999)

24. Bhushan, B.: Introduction to Tribology. Wiley, New York (2002)

25. Bikerman, J.J.: The Science of Adhesive Joints. Academic, New York (1961)

26. Zisman, W.A.: Influence of constitution on adhesion. Ind. Eng. Chem. **55**(10), 18–38 (1963)

27. Houwink, R., Salomon, G.: Effect of contamination on the adhesion of metallic couples in ultra high vacuum. J. Appl. Phys. **38**, 1896–1904 (1967)

28. Israelachvili, J.N.: Intermolecular and Surface Forces, 2nd edn. Academic, San Diego (1992)

29. Bhushan, B.: Tribology and Mechanics of Magnetic Storage Devices, 2nd edn. Springer, New York (1996)

30. Schmidt, H.R.: Zur anatomie und physiologie der geckopfote. Jena. Z. Naturwissen **39**, 551 (1904)

31. Wagler, J.: Naturliches System der Amphibien. J. G. Cotta'schen Buchhandlung, Munich (1830)

32. Simmermacher, G.: Untersuchungen uber haftapparate an tarsalgliedern von insekten. Zeitschr. Wissen Zool. **40**, 481–556 (1884)

33. Hora, S.L.: The adhesive apparatus on the toes of certain geckos and tree frogs. J. Asiat. Soc. Beng. **9**, 137–145 (1923)

34. Dellit, W.D.: Zur anatomie und physiologie der Geckozehe. Jena. Z. Naturwissen **68**, 613–658 (1934)

35. Autumn, K., Liang, Y.A., Hsieh, S.T., Zesch, W., Chan, W.P., Kenny, T.W., Fearing, R., Full, R.J.: Adhesive force of a single gecko foot-hair. Nature **405**, 681–685 (2000)

36. Kim, T.W., Bhushan, B.: The adhesion analysis of multi-level hierarchical attachment system contacting with a rough surface. J. Adhesion Sci. Technol. **21**, 1–20 (2007)

37. Kim, T.W., Bhushan, B.: Effect of stiffness of multi-level hierarchical attachment system on adhesion enhancement. Ultramicroscopy **107**, 902–912 (2007)

38. Hinds, W.C.: Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles. Wiley, New York (1982)

39. Jaenicke, R.: Atmospheric aerosol size distribution. In: Harrison, R.M., van Grieke, R. (eds.) Atmospheric Particles, pp. 1–29. Wiley, New York (1998)

40. Stork, N.E.: A comparison of the adhesive setae on the feet of lizards and arthropods. J. Nat. Hist. **17**, 829–835 (1983)

41. Federle, W., Riehle, M., Curtis, A.S.G., Full, R.J.: An integrative study of insect adhesion: mechanics of wet adhesion of pretarsal pads in ants. Integr. Comp. Biol. **42**, 1100–1106 (2002)

42. Hanna, G., Barnes, W.J.P.: Adhesion and detachment of the toe pads of tree frogs. J. Exper. Biol. **155**, 103–125 (1991)

43. Van der Kloot, W.G.: Molting. In: Cummings, M. (ed.) Encylcopedia Americana, vol 19, pp 336–337. Grolier, Norwich (1992)

44. Hansen, W.R., Autumn, K.: Evidence for self-cleaning in gecko setae. Proc. Natl. Acad. Sci. USA **102**, 385–389 (2005)

45. Autumn, K., Dittmore, A., Santos, D., Spenko, M., Cutkosky, M.: Frictional adhesion, a new angle on gecko attachment. J. Exp. Biol. **209**, 3569–3579 (2006)

46. Daltorio, K.A., Gorb, S., Peressadko, A., Horchler, A.D., Ritzmann, R.E., and Quinn, R.D.: A Robot that Climbs walls using micro-structured polymer adhesive. In: Proceedings of 30th annual meeting of the adhesion society, pp. 329–331 (2007)

47. Aksak, B., Murphy, M.P., and Sitti, M.: Gecko inspired micro-fibrillar adhesives for wall climbing robots on micro/nanoscale rough surfaces. In: Proceedings of 2008 IEEE conference on robotics and automation, Pasadena, CA, pp. 3058–3063 (2008)

48. Cutkosky, M.R., Kim, S.: Design and fabrication of multi-materials structures for bio-inspired robots. Phil. Trans. R. Soc. A **367**, 1799–1813 (2009)

49. Geim, A.K., Dubonos, S.V., Grigorieva, I.V., Novoselov, K.S., Zhukov, A.A., Shapoval, S.Y.: Microfabricated adhesive mimicking gecko foot-hair. Nat. Mater. **2**, 461–463 (2003)

50. Bhushan, B., Sayer, R.A.: Surface characterization and friction of a bio-inspired reversible adhesive tape. Microsyst. Technol. **13**, 71–78 (2007)

51. Bertram, J.E.A., Gosline, J.M.: Functional design of horse hoof keratin: the modulation of mechanical properties through hydration effects. J. Exp. Biol. **130**, 121–136 (1987)

52. Autumn, K., Peattie, A.M.: Mechanisms of adhesion in geckos. Integr. Comp. Biol. **42**, 1081–1090 (2002)

53. Huber, G., Gorb, S.N., Spolenak, R., Arzt, E.: Resolving the nanoscale adhesion of individual gecko spatulae by atomic force microscopy. Biol. Lett. **1**, 2–4 (2005)

54. Gao, H., Wang, X., Yao, H., Gorb, S., Arzt, E.: Mechanics of hierarchical adhesion structures of geckos. Mech. Mater. **37**, 275–285 (2005)

Gecko Feet

► Gecko Effect

Gel Chemical Synthesis

▶ Sol-Gel Method

Genotoxicity of Nanoparticles

Laila Benameur[1], Liu Wei[2] and Alain Botta[1]
[1]Laboratoire de Biogénotoxicologie et Mutagenèse Environnementale (EA 1784/FR CNRS 3098 ECCOREV), Aix-Marseille Université, Marseille Cedex 5, France
[2]CEREGE UMR 6635-CNRS, Aix-Marseille Université, Europôle de l'Arbois, Aix-en-Provence, France

Synonyms

Clastogenicity or/and aneugenicity

Definition

Genotoxicity literally is toxicity against genetic material (DNA damages). It may lead to mutagenicity and carcinogenicity in some circumstances.

Introduction

The unique behavior and physicochemical properties of engineered nanoparticles (NPs) confer novel capabilities to these particles, suitable for a plethora of industrial applications, but currently raise concerns about hazardous effects on biological systems [1]. Although NPs are increasingly used in wide variety of new products and widespread applications, there is insufficient knowledge regarding the human health and environmental implications of these NPs. In this respect, the detailed cellular and molecular mechanisms leading to the biological effects of NPs must be understood to develop safe nanoscale engineered materials and standardized assays of their early toxicity risk [2]. Moreover, genotoxicity is considered an important aspect of carcinogenesis, thus the interactions of NPs with DNA and the mechanisms of those genotoxic interactions must be understood to evaluate and predict mutation- and cancer-related risks of NPs [3–5].

The genotoxic properties of NPs are often closely linked to oxidative damage to DNA and proteins caused by oxidative stress resulting from hyperproduction of reactive oxygen species (ROS) and reactive nitrogen species (RNS). NPs-induced oxidative stress is perhaps the most broadly developed and accepted mechanism for the potential toxic activity of NPs [1]. NP-mediated ROS and RNS production mechanisms can be classified into three groups: intrinsic production, production by interaction with cell targets, and production mediated by the inflammatory reaction. The three groups share responsibility for most of the primary (direct or indirect) or secondary genotoxic effects so far observed with NPs. Primary genotoxicity is generally thought to be without threshold, while it seems that thresholds can be specified for secondary genotoxicity effects. The general mechanism may be clastogenic (i.e., direct or indirect damage to the genetic material during the interphase or during the mitotic process) or aneugenic (alteration of the mitotic apparatus proteins, particularly the spindle and nucleoli). Moreover an indirect consequence can be DNA adducts generated by electrophilic metabolites resulting from bioactivation of polycyclic aromatic hydrocarbons (PAH) adsorbed onto the NPs and thereby delivered to the cytosol.

Genotoxic Mechanisms
Direct Clastogenic Mechanisms
Direct clastogenic mechanisms underlie several DNA lesions, like base oxidization notably producing 8-hydroxy,2′deoxyguanosine (8-OHdG), base nitration by RNS, methylation, oxidative deamination, or depurination, generating abasic sites, ring opening, and finally, single strand breakages (SSB) and double-strand breakages (DSB) by deoxyribose ring opening. These mechanisms may have carcinogenic consequences, if the mutations resulting from oxidative DNA damage, for example, base pair mutations, deletions and insertions, affect and induce oncogene activation, and tumor suppressor gene inactivation.

Indirect Clastogenic Mechanism
The indirect clastogenic mechanism occurs through the mediation of preliminary lipid peroxidation by ROS, which generates electrophilic α,β-unsaturated aldehydes such as malondialdehyde (MDA) and

4-hydroxy-2-nonenal (4-HN), leading to the production of exocyclic DNA adducts (etheno and propano adducts).

Aneugenic Mechanism

The aneugenic mechanism involves protein oxidative lesions induced by ROS (e.g., cysteine oxidation) or nitration by RNS and their derivatives, especially dinitrogen trioxide N_2O_3, which notably results in nitrotyrosination. When lesions affect constituents of the mitotic apparatus (achromatic spindles, microtubule organizing centers, kinetochores), dysfunction can occur in chromosome segregation and migration during mitosis. The major consequence may be non-disjunction in the anaphase and chromosome loss.

Production of DNA Adducts

Environmentally, PAH can be adsorbed on NPs and secondarily delivered to the cytosol, where it may intracellularly concentrate genotoxic/mutagenic electrophilic metabolites produced by bioactivation of many PAH. These metabolites can generate DNA and protein bulky adducts. Certain NPs like diesel or carbon black can adsorb, transport, and deliver several PAH like benzo[a]pyrene (b[a]p) and nitrated/hydroxylated PAH derivatives that are much more genotoxic than their unsubstituted counterparts. Moreover, the ROS and RNS produced during PAH bioactivation might be enzymatic inducers of electrophilic PAH bioactivation and might also inhibit nucleotide excision repair (NER), a process demonstrated on the human alveolar epithelial cell line A549 [6].

Detection and Assessment of Genotoxicity

To detect and assess genotoxicity, short-term assays are implemented that are classified in three main families depending on the type of detected abnormality: primary DNA alterations, gene mutations, and chromosome mutations.

Detection of Primary DNA Alterations

The most often tests used are the unscheduled DNA synthesis test (UDS) and the comet assay.

The UDS test reveals genotoxic lesions by measuring the intensity of DNA synthesis required for their reparation. This technique is based on direct evaluation of primary DNA damages by means of measurement of cellular DNA repair capacity. After contact with the studied product cells are cultivated in presence of tritied thymidine. If DNA has been damaged by the studied product, the cellular repairing process (excision-resynthesis) is started up and the tritied thymidine is incorporated into the repairing enzymes. The measure of the incorporation level of radioactivity is straightly linked to the cellular repairing level. The UDS test can be applied in vitro on several cultured cells (e.g., hepatocytes) and *ex vivo* on exposed animals.

The comet assay or single cell gel electrophoresis (SCGE) assay is a rapid, sensitive and relatively simple method for detecting DNA damage in individual cells. It allows the detection of DNA single strand breaks (SSB, strand breaks, and incomplete excision repair sites), alkali-labile sites, and cross-linking. It was initially introduced under neutral conditions and then developed in alkalin conditions in order to measure low levels of strand breaks with high sensitivity. The comet assay is based on the ability of negatively charged loops/fragments of DNA to be drawn through an agarose gel in response to an electric field. The extent of DNA migration depends directly on the DNA damage present in the cells. In practice, a suspension of cells is mixed with low melting point agarose and spread onto a microscope glass slide. Following lysis of cells with detergent at high salt concentration, DNA unwinding and electrophoresis are carried out at a specific pH. Unwinding of the DNA and electrophoresis at neutral pH [7, 8] predominantly facilitates the detection of double-strand breaks and cross-links whereas unwinding and electrophoresis at pH 12.1–12.4 facilitates the detection of single and double-strand breaks, incomplete excision repair sites and cross-links. At a pH greater than 12.6, alkali labile sites are detected. DNA subjected to an electric field migrates out of the cell, in the direction of the anode, appearing like a 'comet' with a distinct head, comprising of intact DNA and a tail, consisting of damaged or broken pieces of DNA. The size and shape of the comet and the distribution of DNA within the comet correlate with the extent of DNA damage. In combination with certain bacterial enzymes (e.g., formamidopyrimidine glycosylase, endonuclease III, uracil-DNA glycosylases), which recognize oxidized purines and pyrimidine bases, the Comet Assay has been used to determine oxidative DNA damage. The main advantages of the Comet Assay include: (1) the collection of data at the level of the individual cell, allowing more robust statistical

analyses; (2) the need for a small number of cells per sample (<10,000); (3) sensitivity for detecting DNA damage; and (4) use of any eukaryote single cell population both in vitro and in vivo, including cells obtained from exposed human populations and aquatic organisms for ecogenotoxicological studies and environmental monitoring.

Single cell gel electrophoresis assay is a fast, reproducible, and sensitive microelectrophoretic technique for visualizing and evaluating the extent of DNA damage: SSB and DSB, alkali-labile sites, cross-links, and sites not yet fully repaired, during the excision phase of the BER (base excision repair) and NER systems, in single prokaryotic and eukaryotic cells. It can be used in vitro on several cell cultures, *ex vivo* on human lymphocytes or epithelial cells, and in vivo on the whole animal to pinpoint any organ specificity of the genotoxic substance. The comet assay can also be used to identify apoptotic cells and to test the capacity to repair DNA lesions. In addition, the comet assay can be carried out in the presence of bacterial endonucleases, such as formamidopyrimidine DNA glycosylase (Fpg) or endonuclease III, which handle specific excision repair of oxidative DNA lesions (with ensuing resynthesis of the eliminated strand and consequent SSB after the excision stage). This provides indisputable evidence of the oxidative DNA base lesions induced by ROS.

Detection of Gene Mutations

Gene mutations are either substitution of a single base pair (point mutations) or multiple alterations (addition or deletion) of several base pairs (frameshift mutations). The classical assays are the Ames test, the mouse lymphoma test, the hypoxanthine guanine phosphoribosyl transferase (HPRT) test, and the use of transgenic mouse strains.

The Ames test detects gene mutations in various strains of *Salmonella typhimurium* carrying a mutation in the operon governing essential amino-acid histidine synthesis, making them unsuited to develop in a histidine-deficient culture medium (auxotrophy). In the presence of a genotoxic substance, a reverse mutation gives the bacteria the ability to synthesize histidine once again (prototrophy), whereupon they can then develop in a medium deficient in this amino-acid. Various strains have been developed with different sensitivities to genotoxic substances but for oxidative lesions strain TA102 is the most appropriate. The Ames test is a good tool for detecting gene

mutations, but it cannot detect clastogenic or aneugenic chromosome mutations. Nevertheless, this test allows the detection of gene mutations induced either by direct or indirect genotoxicants, the latter requiring a previous bioactivation leading to electrophilic metabolites, the true agents of genotoxicity. To distinguish these two mechanisms, the test can be carried out with or without an in vitro metabolizing mixture, S9 Mix, a rat microsomal liver fraction obtained after induction by Aroclor (an enzyme inducer of CYP450 mono-oxygenases), combined with NADP(H) generating cofactors.

The mouse lymphoma assay is carried out on the L5178Y mouse lymphoma cell line, heterozygous at the thymidine kinase locus (tk+/−). Deactivation of the tk+ allele leads to loss of heterozygosity, which induces resistance to toxic trifluorothymidine, allowing selection of the tk−/− mutants within the tk+/− cell population. This test can reveal gene mutations and clastogenic or aneugenic chromosome mutations.

The HPRT assay detects gene mutations at the HPRT locus in the V79 cell lines of Chinese hamster pulmonary fibroblasts or Chinese hamster ovary (CHO) cell lines. The test is based on the property of the HPRT enzyme to catalyze the phosphoribosylation of 6-thioguanine, leading to a cytotoxic monophosphate derivative. The mutations at the HPRT locus are assessed by counting clones resistant to 6-thioguanine.

The use of transgenic mouse strains has become a classical method for assessing genotoxicity in vivo. The BigBlue model contains the Lac I gene as a target for the genotoxicants and the Lac Z gene as a reporter.

Lac I represses β-galactosidase activity, which normally produces galactose by hydrolysis of the substrate X-Gal. The result is then the appearance of blue lysis plaques. Lac Z yields a functional β-galactosidase. Gene mutations of Lac I will result in a nonfunctional Lac*1 protein repressor, thereby allowing hydrolysis of X-Gal by β-galactosidase and the appearance of blue plaques. Another model of transgenic mouse has been developed, called Muta Mouse, which uses the gene Lac Z directly as a target. Based on the same principle as BigBlue, this method uses the toxicity of galactose on the bacteria Escherichia coli Cgal/E−, which cannot develop in the presence of galactose. Gene mutations of Lac Z prevent the production of galactose and thus allow the bacteria to develop normally.

Detection of Chromosome Mutations

Chromosome mutations involve tens of thousands of DNA bases or whole chromosomes according to two separate mechanisms: clastogenic mechanisms lead to direct structural (or qualitative) chromosome abnormalities like double-strand DNA breakages whereas aneugenic mechanisms result in numerical (or quantitative) abnormalities. The latter abnormalities consist of changes in the number of chromosomes by creating lesions in the proteins of the mitotic apparatus. To detect and evaluate chromosome structure and number damage, the *micronucleus test* is the most widely used. Micronuclei (MN) are nuclear entities independent of the main nucleus, 1–6 in number per cell, with diameters between 1/3 and 1/16 of that of the main nucleus. These MN are produced during cell division and may contain either acentric chromosome fragments, which cannot position on achromatic spindle (clastogenic effect), or whole chromosomes, which have been lost during the anaphase because of lesions of the spindle proteins (aneugenic effect). To specify whether the genotoxic substance induces a clastogenic and/or aneugenic effect, the micronucleus test can be combined with fluorescent in situ hybridization (FISH) using pancentromeric DNA probes, which provide a precise fluorescent visualization of the presence (aneugenic) or absence (clastogenic) of centromeres in the micronucleus. In several cell types, like human fibroblasts, the very slight level of centromeric signal obtained with the FISH technique leads to great difficulties in distinguishing the genotoxicity mechanisms of toxic substances. The micronucleus test, in combination with immunofluorescence staining of CENP-A (centromere protein A), can efficiently screen genotoxicants for their ability to induce clastogenic and/or aneugenic effects [7]. The micronucleus test can be carried out in vivo, *ex vivo*, and in vitro on CHO cell lines, mouse lymphoma L5175Y cell lines, V79 cell lines of Chinese hamster pulmonary fibroblasts, and primary cultured human cells (lymphocytes, fibroblasts, keratinocytes, melanocytes, enterocytes, etc.).

Genotoxic Mechanisms of NPs

The main mechanisms by which cells interact with NPs, generators of oxidative stress, are classified into three groups, according to the most widely accepted hypothesis: (1) direct involvement of the surface effect, (2) contribution of redox mechanisms due to transition metals released by NPs, (3) activation of membrane receptors like epidermal growth factor receptor (EGFR) by transition metals after their intracellular diffusion. In the first two cases, oxidative stress is combined with increased cytosolic calcium concentrations and activation of signaling pathways inducing activation of transcription factors, notably, nuclear factor-$\kappa\beta$ (NF-$\kappa\beta$) involved in the transcription of key genes. In the third case, the activation of membrane receptors leads to the mitochondrial distribution of NPs and concomitant production of oxidative stress [2].

Genotoxic Effects of NPs

This section is dedicated to recent examples of in vitro and in vivo studies that investigated the potential genotoxic effects of engineered NPs and reports the current data from the main scientific studies on two major categories of NPs: carbonaceous NPs and metal-containing NPs.

Carbon-based NPs

Carbon-based nanoscale materials exist in a variety of shapes and conformations. We focus on the genotoxicity of carbonaceous tube structures, including both single-wall carbon nanotubes (SWCNTs), multi-walled carbon nanotubes (MWCNTs), and carbon fullerenes, mainly C_{60} fullerenes.

SWCNTs

SWCNTs have been reported to induce many harmful cellular and animal toxicity reactions. The induction of genotoxic effects is mediated mainly by oxidative stress, but the exact mechanism of genotoxicity of these NPs remains to be identified.

For example, Kisin et al. [8] investigated the genotoxicity of SWCNTs having several dimensions (diameters from 0.4 to 1.2 nm, length of 1–3 mm and surface area of 1,040 m^2/g) and containing 99.7 wt% element carbon and iron levels of about 2.3 mg Fe/g sample (0.23 wt%). Exposure of Chinese hamster lung fibroblasts (V79 cell line) to SWCNTs for 3 h at 96 μg/cm^2 induced single and double-strand DNA breakages, as revealed by the comet assay, which also identified alkali-labile sites. However, for the same exposure, the micronucleus test did not reveal any significant increase in the number of

micronucleated cells, and the Ames test in Salmonella strain YG1024 or YG1029 was negative at concentrations of 60, 120, or 240 mg/plate, without S9 mix. The genotoxicity of SWCNTs is thus confirmed by the comet assay, which assesses primary DNA lesions, while the negative micronucleus test suggests that these primary lesions were effectively repaired.

In another example, Shvedova et al. [9] investigated the genotoxicity of SWCNTs (with diameters from 0.8 to 1.2 nm, and length of 100–1,000 nm). The in vivo experiment was carried out on C57BL/6 mice after exposure to SWCNTs, by using two forms of exposure: closed-chamber inhalation (5 mg/m^3 of SWCNTs, 5 h per day for 4 days) and aspiration through the lungs of a particle suspension deposited in the pharynx (5, 10, or 20 µg per mouse). Both exposure routes generated an immediate inflammatory reaction, oxidative stress, fibrosis, and hyperplasia of bronchial epithelium cells, but only inhalation exposure caused genotoxicity as measured by analyzing mutations of the k-ras gene, persisting after 28 days of exposure.

MWCNTs

Zhu et al. [10] reported in 2007 that suspensions of MWCNTs at concentrations of 5 and 100 µg/ml induced apoptosis in mouse embryo stem cells (ES cell line) by activation of p53 protein. This protein can arrest the cell cycle in presence of DNA lesions, thereby allowing the cell to implement DNA repair systems, but it can also induce apoptosis if these repair systems are insufficient or if the lesions cannot be repaired. The internet report by Michael Berger (Copyright 2007 Nanowerk LLC) mentioned in the review of Landsiedel et al. [11] suggests using p53 as a biomarker for preliminary screening of the genotoxicity of nanomaterials on the basis of the close relationship between p53 activation and DNA damage.

Moreover, the genotoxicity of well characterized MWCNTs has been investigated in rat lung epithelial cell cultures. DNA damage was evaluated by applying the micronucleus test at different concentrations of NPs: 10, 25, and 50 µg/ml. These NPs led to a significantly increased number of micronucleated cells. In addition, in MCF-7 cells centromere-positive and -negative micronuclei were produced, providing evidence that MWCNT are clastogenic and aneugenic, as identified using in situ hybridization of fluorescent pancentromeric probes. Moreover, intratracheal instillation of MWCNTs (0.5 or 2 mg) in rats over 3 days led to a dose-dependent increase in micronuclei in type II pneumocytes. This study provides the first proof of the in vivo genotoxicity of MWCNTs in a relevant target cell [12].

Fullerenes

One salient example of the difficulty in interpreting experimental results and the need to standardize methodologies is well illustrated by the following study: stable aqueous colloidal suspensions of C$_{60}$ fullerenes free of toxic organic solvents were prepared by two methods, (1) dispersing C$_{60}$ in ethanol and then redispersing it in water, (2) directly mixing C$_{60}$ in water. A primary culture of human lymphocytes was exposed to these two preparations, and primary DNA lesions were then identified by the comet assay. Both types of suspension induced a dose-dependent increase in DNA damage, but the suspension prepared by direct dispersion in water elicited higher genotoxic response than the one primarily prepared in ethanol [13]. It is rather likely than C$_{60}$ molecules directly dispersed in water spontaneously (slowly) got oxidized. Moreover, note that the dispersion of C$_{60}$ in tetrahydrofurane (THF) is debated, since it seems that the toxicity attributed to C$_{60}$ may have been caused by a decomposition product of THF [14]. These findings highlight the importance of a rigorous choice of NPs dispersion and of detection of artefacts caused by the material suspension procedures.

Metal and Oxide NPs

This section covers the current reports on genotoxicity of metal and oxide NPs: cerium dioxide, silver, titanium dioxide, iron oxide, silica, zinc oxide, cobalt and alloys, and copper oxide NPs.

Cerium Dioxide

Cerium, a lanthanide member of the rare earth metals, is a very reactive and strong oxidizing agent. Cerium can exist in either free metal or oxide form. Currently, engineered cerium dioxide (CeO$_2$) NPs are widely used in various applications and are considered one of the most interesting nanomaterials owing to their catalytic properties [15]. Contradictory results have been reported on the antioxidant or pro-oxidant effects of CeO$_2$ NPs. Some studies reported radio protective [16] activities of CeO$_2$ NPs. Conversely, others have shown these NPs have a strong genotoxic effect. Auffan et al. [17] reported that exposure of primary

human fibroblasts to CeO_2 NPs led to (1) significant and dose-dependent induction of micronuclei at environmentally relevant doses and (2) induction of SSBs (evaluated by the comet assay) at 0.6 µg/mL and above 60 µg/mL, and these SSBs were significantly reduced in presence of the antioxidant L-ergothioneine. On the other hand, Pierscionek et al. [18], found no DNA damage to eye lens epithelial cells at 5 or 10 µg/mL (measured by comet assay and sister chromatid exchange). In both studies, CeO_2 NPs were found in cytoplasmic vacuoles of human primary fibroblasts and human eye lens epithelial cells [17, 18]. Moreover, few data are available on CeO_2 genotoxicity for NPs ecological organisms. Lee et al. [19] reported that CeO_2 NPs were genotoxic toward aquatic sentinel species (freshwater crustacean *Daphnia magna* and the larva of the aquatic midge *Chironomus riparius*). DNA damage, particularly DNA strand breaks, was measured using comet assay. Olive tail moments increased in the two aquatic organisms exposed to CeO_2 NPs. Moreover, the smaller-sized CeO_2 NPs seemed to cause more DNA strand breaks, which might provoke higher-level consequences.

In addition, an in vivo toxicity study has demonstrated that exposure of male Sprague Dawley rats to CeO_2 NPs by a single intratracheal instillation (at doses of 0.15, 0.5, 1, 3.5, or 7 mg/kg) caused significant lung inflammation and toxicity to lung cells. Alteration of alveolar macrophages was significant at 1 day after exposure, but it was significantly increased, in a dose-dependent manner, at 10 and 28 days postexposure. The results show that CeO_2 NPs induced cytotoxicity, air/blood barrier damage, and phospholipidosis with enlarged alveolar macrophages. Despite these toxic effects, the comet assay indicated that exposure of rats to CeO_2 NPs did not cause any DNA damage in alveolar macrophages [20].

At this stage, it is difficult to distinguish between the protective and toxic effects of CeO_2 NPs. More studies should be undertaken to understand the genotoxicity of CeO_2 NPs for their safe use, particularly in biomedical applications.

Silver

Currently, silver nanoparticles (Ag NPs) have gained much interest owing to the broad spectrum of their antimicrobial effects. The induction of oxidative stress seems to be the main mechanism of Ag NPs -mediated genotoxicity.

AshaRani et al. [21] provided the first study on quantitative measurements of DNA damages and chromosomal aberrations in Ag NPs-treated human lung fibroblast cells (IMR-90) and human glioblastoma cells (U251). DNA damages, as measured by comet assay and cytokinesis-block micronucleus assay (CBMN), were dose-dependent and more prominent in cancer cells. In addition, the extensive damage of DNA was reflected in the cycle arrest of the IMR-90 and U251 cells in S and G2/M phases. After treatment with various doses of Ag NPs, the number of cells in the G2/M phase increased in dose-dependent manner possibly due to repair of damaged DNA. The authors also reported that Ag NPs reduced ATP content of the cell, leading to mitochondrial damage and increased production of ROS in a dose-dependent manner. Moreover, the TEM analysis indicated the presence of Ag NPs in the mitochondria and nucleus, implicating them directly in mitochondrial toxicity and DNA damage. The authors proposed a possible mechanism of Ag NPs toxicity, which involves disruption of the mitochondrial respiratory chain, leading to production of ROS and interruption of ATP synthesis, which in turn cause DNA damage.

Titanium Dioxide

Studies on the biological effects of titanium dioxide (TiO_2) NPs, coated or uncoated, have involved cultured animal and human cells. The exposure of Syrian hamster embryo (SHE) fibroblasts to TiO_2 NPs (20 nm) at different concentrations (from 0.5 to 10 mg/cm^2) induced clastogenic events. This genotoxic mechanism was demonstrated by using CREST antibodies in association with micronucleus assay. Interactions between the NPs and the fibroblast cell membranes might induce ROS production, which would lead to lipid peroxidation, intracellular Ca^{2+} homeostasis disturbance, and metabolic pathway alterations [22]. Exposure of HEpG-2 cells to TiO_2 NPs (anatase) at various concentrations (10, 20, 50, 100 µg/mL) induced DNA lesions (detected by comet assay) in dose-dependent manner. Moreover, micronucleus test showed significantly increased MN frequencies in HEp-2 cells exposed to a high concentration of NPs (50 µg/mL) [23].

Also, the genotoxic effects of 25 nm TiO_2 NPs (Degussa P25: 70–85% anatase/30–15% rutile) were evaluated in cultured peripheral human lymphocytes by the comet and micronucleus assays. The cells

incubated with NPs exhibited dose-dependent production of micronuclei and DNA strand breaks and increased ROS production, which could lead to breakage or loss of genetic material. Moreover, pre-treatment of lymphocytes with N-acetylcysteine significantly decreased ROS production and oxidative DNA damage. The results also demonstrated accumulation and activation of p53 proteins but without concomitant stimulation of its transactivation activity, required to arrest the cell cycle and trigger apoptosis [24]. Human lymphocytes and root or leaves of *Allium cepa* and *Nicotiana tabacum* incubated with TiO_2 NPs (CAS No.13463-67-7) and bulky TiO_2 exhibited dose-dependent production of DNA strand breaks. In addition, a significant production of MDA was observed in *Allium cepa*, after treatment with TiO_2 NPs. This indicated that lipid peroxidation could be involved as one of the mechanisms leading to DNA damage [25].

A study carried out in vivo (rats) by instillation of two types of 20 nm TiO_2 NPs (hydrophilic with untreated surface and hydrophobic with surface silanised by trimethoxyoctylsilane) at different doses (0.15, 0.3, 0.6, and 1.2 mg) for 90 days did not reveal any evidence of 8-OHdG production in the DNA of alveolar epithelial cells [26]. Conversely, the analysis of DNA damage and genetic instability using C57Bl/$6Jp^{un}/p^{un}$ (mice liver) after 5 days of oral exposure showed that TiO_2 NPs (75% anatase and 25% rutile) induced 8-OHdG, DNA double-strand breaks, and micronuclei in bone marrow and/or peripheral leukocytes. Moreover, these NPs increased DNA deletions during fetal development, which might be a result of oxidative genome damage. Also, an inflammatory reaction was observed in peripheral blood through changes in cytokine expression, suggesting an indirect genotoxicity pathway [27]. In addition, the genotoxic effects of TiO_2 NPs (anatase and rutile) at three doses (20, 50, and 100 µg/mL) on bottlenose dolphin leukocytes were tested by the comet assay. The results showed that both forms of TiO_2 statistically increased DNA fragmentation at 24 and 48 h exposures [28].

Micronucleus test on V79 cells (Chinese hamster lung fibroblasts) with TiO_2 NPs in anatase form, coated with vanadium pentoxide V_2O_5, showed a higher level of cytogenotoxicity than for the same NPs uncoated. The greater genotoxicity of the particles coated with V_2O_5 would thus appear to be mediated by oxidative stress resulting from elevation of ROS production [29].

Another study of TiO_2 NPs (<100 nm) with 10–100 µg/ml exposure in NIH 3 T3 and human fibroblast HFW cells showed, after short exposure (24–72 h), an increase in cell proliferation, ERK signaling activation, and ROS production. Long-term exposure (11 weeks) might affect chromosome segregation. Moreover, the results also indicated that the deregulation PLK1 function by TiO_2 NPs could cause formation of impaired centrosome size/number, which is closely correlated with aneuploidy and genomic instability in tumor development [30].

Finally, a study of various sizes of TiO_2 NPs (20, 50, 100, and 200 µg/ml) in NRK-52E cells showed a statistically lower cytotoxicity for the larger NPs (140 nm, 100% anatase) than the smaller ones (12 nm, 95% anatase; 25 nm, 89% anatase) by MTT assay. However, the size of these NPs did not influence the increase in cellular ROS production. In addition, the significant DNA damages induced by TiO_2 NPs (140 nm) indicated that the oxidative stress might not be the only cause of DNA damage [31].

All those studies revealed that the oxidative stress seems to be the main mechanism of TiO_2 NPs – mediated genotoxicity. However, the damage induced by TiO_2 NPs appears independent upon size and form of NPs. Unfortunately, the lack of details on the material and physicochemical characterization in some studies makes it difficult to compare the results and to obtain a clear understanding of TiO_2 NPs genotoxicity.

Iron Oxide

The cytogenotoxicity of maghemite (ΥFe_2O_3) NPs coated with dimercaptosuccinic acid (DMSA) in human fibroblasts was assessed at 10–100 µg/ml. Neither cytoxocitity nor genotoxicity were observed. The study also demonstrated the internalization of NPs via endocytosis vesicles but the DMSA coating, forming a stable organic layer, led to protecting cells against direct contact with the nanoparticles surface [32]. Moreover, Fe NPs (5.1 and 10.2 µg/cm^2) induced significant and dose-dependent increases in DNA damage in human monocyte cells (Mono Mac 6) and in rat alveolar macrophages exposed at air-tissue interface (using the comet test) [33].

The micronucleus test conducted on polychromatic erythrocytes of SWISS mice 12, 24, and 48 h after intravenous administration of magnetoliposomes (magnetite Fe_3O_4 coated with a lipid bilayer, 14 nm)

showed micronucleus induction at 24 h only [34]. Another similar study of long-term effect (1–30 days) of magnetite NPs coated with polyaspartic acid (8.5 nm) showed micronucleus induction from 1 to 7 days [35].

Silica

The genotoxic potential of amorphous and crystalline forms of silica NPs has been investigated by in vitro and in vivo studies. Exposure of human colon epithelial cell-line Caco-2 for 24 h to amorphous fumed SiO_2 NPs (14 nm, 20 and 80 mg/cm^2) induced cell mortality, significant oxidative DNA damage (Fpg- modified comet assay) and glutathione depletion [36].

Different sizes of amorphous SiO_2 NPs (16, 60, and 104 nm) have been used to investigate genotoxic effects in A549 lung carcinoma cells. The smallest particles showed only a slightly higher induction of micronuclei cells at non-cytotoxic doses. Considering the three NPs together, particle number and total surface greatly contribute to micronuclei induction. In addition, alkaline comet assay and FISH-centromeric probing of micronuclei showed a weak and not statistically significant induction of oxidative DNA damage, chromosome breakage, and chromosome loss [37].

Genotoxicity of ultrafine crystalline SiO_2 NPs (< 100 nm) was investigated by exposing human lymphoblastoid cells (WIL2-NS) during 6, 24, and 48 h at concentrations of 0, 30, 60, and 120 mg/ml. The micronucleus assay showed a dose-dependent increase in micronucleated cells and a decrease in cell proliferation. However, the comet assay gave no significant results, making it difficult to determine the genotoxic mechanism. Furthermore, these NPs (120 μg/ml) induced a significantly increased number of mutants, detected with the HPRT test [38].

In a recent in vivo study, rats inhaled freshly generated aerosolized amorphous SiO_2 NPs (37 and 83 nm, 3.7×10^7 and 1.8×10^8 particles/cm^3) during a short period (1 and 3 days). Pulmonary toxicity and genotoxicity endpoints (micronuclei induction) were estimated from 24 h up to 2 months after exposure. One- or 3-day aerosol exposures produced no significant pulmonary inflammation and no genotoxic or adverse lung histopathological effects in rats exposed to very high particle concentrations (1.8 or 86 mg/m^3) [39].

Finally, to explain the ability of NPs to carry and deliver pollutants, a Trojan horse effect has been suggested: silica NPs could facilitate cell penetration of the associated heavy metals and induce an up to eight times higher oxidative stress than the heavy metals alone in an aqueous solution [40].

Zinc Oxide

The genotoxic effects of ZnO NPs (40–70 nm) in presence and absence of UV light in human sperm and human lymphocytes have been investigated using comet assay. The results showed photogenotoxic events in these cells [41].

Two sizes (70 and 420 nm) of ZnO NPs in human bronchoalveolar carcinoma-derived cells (A549) induced ROS production, lipid peroxidation, and cell membrane leakage. DNA damage was observed by comet assay after 24 h exposure to 10, 12, 14 μg/ml of these NPs without size effect [42].

The influence of surface chemistry effects of ZnO NPs has been well studied: uncoated ZnO NPs and ZnO NPs coated with oleic acid (OA) or polymethacrylic acid (PMAA) were studied by comparing cytotoxicity, genotoxicity, and ROS generation in WIL-NS human lymphoblastoid cells. Results showed that coated NPs decreased ROS production and cytotoxicity. However, PMAA-coated and OA-coated ZnO NPs (10 mg/L) induced significant elevation in the frequency of micronuclei in cells exposed during 24 h [43]. The Ames test showed no mutations at any tested concentration for ZnO NPs coated with tetraethylammonium hydroxide in the TA98, TA100, TA1535, and TA1537 strains, with and without metabolic activation, and in the *Escherichia coli* strains WP2uvrA(−) [44].

Cobalt and Compounds

The genotoxicity of pure cobalt, cobalt-chromium, and cobalt-iron mixture NPs has been investigated in several studies. One study compared cobalt ions Co^{2+} and cobalt NPs (100 nm) in human leukocytes. Comet assay showed significantly higher DNA SSB for cobalt NPs than Co^{2+} ions, with a very significant cellular uptake of cobalt NPs. Conversely, the micronucleus test revealed a significantly increased frequency of micronucleated cells under exposure to Co^{2+} ions. The authors hypothesized that genotoxic effects leading to SSB are probably well repaired, but modulated by the salting out of Co^{2+} ions. In particular, the repair

gene hOGG1 (a gene coding for a protein that excises the oxidized base 8-OHdG) may also modulate the genotoxic response by more or less completely repairing all the ROS-induced SSB [45].

Moreover, the cytotoxic and genotoxic effects of cobalt-ferrite ($CoFe_2O_4$) particles, of nanometric (5.6 nm) and micrometric (10 and 120 μm) size, have been evaluated in cultures of human peripheral lymphocytes. The nanoparticles significantly reduced the cell proliferation index and significantly increased the frequency of micronucleated binucleated lymphocytes (MBNL), while the 10-mm microparticles increased only the frequency of MBNL. In addition, to see whether the genotoxicity might not have been caused by salted out ions, the NPs were coated in order to block this effect. The results showed that the level of micronucleated cells was four times lower than the one induced by bare NPs. This corroborates the hypothesis that Co^{2+} ions play a role in the genotoxicity of cobalt-ferrite NPs [46].

Furthermore, a novel approach of indirect genotoxicity mechanism brings to the participation of barrier cells to the DNA damage induced by CoCr alloy NPs, and the importance of their indirect genotoxic effects. Indeed, in the top layer of the barrier, CoCr nanoparticles could induce mitochondrial damage because Co^{2+} ions might mimic hypoxic conditions. These effects could induce a release of ATP, which could reach the second layer of cellular barrier via P2X receptors and pannexin and connexin hemichannels, and act on P2Y receptors or pass through connexin gap junctions. In the second cellular layer, ATP could induce an increase in intracellular calcium and subsequent ATP secretion. This ATP could cause DNA damage to fibroblasts. The multi-layered barrier used in this investigation was the BeWo (human trophoblast choriocarcinoma derived cell line) [47].

Copper Oxide

Few recent studies have investigated the cytotoxic and genotoxic potential of CuO NPs. After 4-h exposure of CuO NPs (40 μg/cm^2) in A549 cells, comet assay revealed that CuO NPs were more genotoxic than TiO_2 and ZnO NPs. Likewise, with the endonuclease FPG, the CuO NPs also displayed increasing levels of oxidative DNA damage.

Moreover, the effect of possible release Cu ions in the toxicity observed was evaluated, and the results showed that Cu ions were much less toxic than CuO NPs [48]. Mutagenicity of CuO NPs (<50 nm) was also evaluated by using Ames assay. Results showed a low mutagenic level in TA97a and TA100 strains and a significant growth inhibition of *Escherichia coli* WP2, trp, and uvra. The oxidative effect of CuO NPs seems to be the main mutagenic mechanism [49].

Conclusions

Analysis of the scientific literature brings out a likely genotoxic potency of nanoparticles. The observed genotoxic/mutagenic effects usually seem under mediation of oxidative stress, but direct genotoxicity remains possible notably as a result of adsorbed pollutants delivery. The inflammatory process also seems to play a key role in genotoxic mechanisms. Furthermore, conflicting results obtained in some studies highlight the necessity to reassess conventional toxicological methods in order to understand biological barrier-crossing mechanisms, accumulation in target tissues, and the precise nature of interactions with biological material. Additional studies must take into account (1) variations in the compositions and impurity levels of the studied nanocompounds, (2) the influence of the new length scale relevant to nanoparticles, and (3) the influence of the concentration of nanoparticles so as to avoid the two-phase results sometimes found as a function of the concentration. Finally, the genotoxicity of nanoparticles seems unlikely to be the only mechanism able to lead to mutagenesis and carcinogenesis. Some probably epigenetic mechanisms relating to gene regulators or micro-RNA constitute a new challenge and require further investigations.

Cross-References

▶ Cellular Mechanisms of Nanoparticle's Toxicity
▶ Ecotoxicology of Carbon Nanotubes Toward Amphibian Larvae
▶ Ecotoxicity of Inorganic Nanoparticles: From Unicellular Organisms to Invertebrates
▶ Effect of Surface Modification on Toxicity of Nanoparticles

References

1. Auffan, M.: Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective. Nat. Nanotechnol. **4**, 634–641 (2009)

2. Oberdorster, G.: Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ. Health Perspect. **113**, 823–839 (2005)

3. Gonzalez, L.: Genotoxicity of engineered nanomaterials: a critical review. Nanotoxicology **2**, 252–273 (2008)

4. Singh, N.: NanoGenotoxicology: the DNA damaging potential of engineered nanomaterials. Biomaterials **30**, 3891–3914 (2009)

5. Petersen, E.J.: Mechanisms and measurements of nanomaterial-induced oxidative damage to DNA. Anal. Bioanal. Chem. **398**, 613–650 (2010)

6. Güngör, N.: Activated neutrophils inhibit nucleotide excision repair in human pulmonary epithelial cells: role of myeloperoxidase. FASEB J. **21**, 2359–2367 (2007)

7. Benameur, L.: Detection of environmental clastogens and aneugens in human fibroblasts by cytokinesis-blocked micronucleus assay associated with immunofluorescent staining of CENP-A in micronuclei. Chemosphere **84**(5), 676–680 (2011)

8. Kisin, E.R.: Single-walled carbon nanotubes: geno- and cytotoxic effects in lung fibroblast V79 cells. J. Toxicol. Environ. Health A **70**, 2071–2079 (2007)

9. Shvedova, A.A.: Inhalation vs. aspiration of single-walled carbon nanotubes in C57BL/6 mice: inflammation, fibrosis, oxidative stress, and mutagenesis. Am. J. Physiol. Lung Cell. Mol. Physiol. **295**, L552–565 (2008)

10. Zhu, L.: DNA damage induced by multiwalled carbon nanotubes in mouse embryonic stem cells. Nano Lett. **7**, 3592–7 (2007)

11. Landsiedel, R.: Genotoxicity investigations on nanomaterials: methods, preparation and characterization of test material, potential artifacts and limitations – many questions, some answers. Mutat. Res. **681**, 241–258 (2009)

12. Muller, J.: Clastogenic and aneugenic effects of multi-wall carbon nanotubes in epithelial cells. Carcinogenesis **29**, 427–433 (2008)

13. Dhawan, A.: Stable colloidal dispersions of C_{60} fullerenes in water: evidence for genotoxicity. Environ. Sci. Technol. **40**, 7394–7401 (2006)

14. Nielsen, G.D.: In vivo biology and toxicology of fullerenes and their derivatives. Basic Clin. Pharmacol. Toxicol. **103**, 197–208 (2008)

15. Cassee, F.R.: Exposure, health and ecological effects review of engineered nanoscale cerium and cerium oxide associated with its use as a fuel additive. Crit. Rev. Toxicol. **41**, 213–229 (2011)

16. Tarnuzzer, R.W.: Vacancy engineered ceria nanostructures for protection from radiation-induced cellular damage. Nano Lett. **5**, 2573–2577 (2005)

17. Auffan, M.: CeO_2 nanoparticles induce DNA damage towards human dermal fibroblasts in vitro. Nanotoxicology **3**, 161–171 (2009)

18. Pierscionek, B.K.: Nanoceria have no genotoxic effect on human lens epithelial cells. Nanotechnology **21**, 035102 (2010)

19. Lee, S.W.: Genotoxicity and ecotoxicity assays using the freshwater crustacean Daphnia magna and the larva of the aquatic midge Chironomus riparius to screen the ecological risks of nanoparticle exposure. Environ. Toxicol. Pharmacol. **28**, 86–91 (2009)

20. Ma, J.Y.: Cerium oxide nanoparticle-induced pulmonary inflammation and alveolar macrophage functional change in rats. Nanotoxicology. (2010). http://www.ncbi.nlm.nih.gov/pubmed/20925443

21. AshaRani, P.V.: Cytotoxicity and genotoxicity of silver nanoparticles in human cells. ACS Nano **3**, 279–290 (2008)

22. Rahman, Q.: Evidence that ultra fine titanium dioxide induces micronuclei and apoptosis in Syrian hamster embryo fibroblasts. Environ. Health Perspect. **110**, 797–800 (2002)

23. Osman, I.F.: Genotoxicity and cytotoxicity of zinc oxide and titanium dioxide in HEp-2 cells. Nanomedicine **5**, 1193–1203 (2010)

24. Kang, S.J.: Titanium dioxide nanoparticles trigger p53-mediated damage response in peripheral blood lymphocytes. Environ. Mol. Mutagen. **49**, 399–405 (2008)

25. Ghosh, M.: Genotoxicity of titanium dioxide (TiO_2) nanoparticles at two trophic levels: plant and human lymphocytes. Chemosphere **81**, 1253–1262 (2010)

26. Rehn, B.: Investigations on the inflammatory and genotoxic lung effects of two types of titanium dioxide: untreated and surface treated. Toxicol. Appl. Pharmacol. **189**, 84–95 (2003)

27. Trouiller, B.: Titanium dioxide nanoparticles induce DNA damage and genetic instability in vivo in mice. Cancer Res. **15**, 8784–8789 (2009)

28. Bernardeschi, M.: Genotoxic potential of TiO_2 on bottlenose dolphin leukocytes. Anal. Bioanal. Chem. **396**, 619–623 (2010)

29. Bhattacharya, K.: Vanadium pentoxide-coated ultrafine titanium dioxide particles induce cellular damage and micronucleus formation in V79 cells. J. Toxicol. Environ. Health A **71**, 976–980 (2008)

30. Huang, S.: Disturbed mitotic progression and genome segregation are involved in cell transformation mediated by nano-TiO_2 long-term exposure. Toxicol. Appl. Pharmacol. **1**, 182–194 (2009)

31. Barillet, S.: Toxicological consequences of TiO_2, SiC nanoparticles and multi-walled carbon nanotubes exposure in several mammalian cell types: an in vitro study. J. Nanopart. Res. **12**, 61–73 (2010)

32. Auffan, M.: In vitro interactions between DMSA-coated maghemite nanoparticles and human fibroblasts: a

physicochemical and cyto-genotoxical study. Environ. Sci. Technol. **40**, 4367–4373 (2006)

33. Grigg, J.: DNA damage of macrophages at an air-tissue interface induced by metal nanoparticles. Nanotoxicology **3**, 348–354 (2009)

34. Garcia, V.A.P.: Magnetoliposome evaluation using cytometry and micronucleus test. Eur. Cells Mater. **3**, 154–155 (2002)

35. Sadeghiani, N.: Genotoxicity and inflammatory investigation in mice treated with magnetite nanoparticles surface coated with polyaspartic acid. J. Magn. Magn. Mater. **289**, 466–468 (2005)

36. Gerloff, K.: Cytotoxicity and oxidative DNA damage by NPs in human intestinal Caco 2 cells. Nanotoxicology **3**, 355–364 (2009)

37. Gonzalez, L.: Exploring the aneugenic and clastogenic potential in the nanosize range: A549 human lung carcinoma cells and amorphous monodisperse silica NPs as models. Nanotoxicology **4**, 382–395 (2010)

38. Wang, J.J.: Cytotoxicity and genotoxicity of ultrafine crystalline SiO_2 particulate in cultured human lymphoblastoid cells. Environ. Mol. Mutagen. **48**, 151–157 (2007)

39. Sayes, C.M.: Changing the dose metric for inhalation toxicity studies: short-term study in rats with engineered aerosolized amorphous silica NPs. Inhal. Toxicol. **22**, 348–354 (2010)

40. Limbach, L.K.: Exposure of engineered NPs to human lung epithelial cells: influence of chemical composition and catalytic activity on oxidative stress. Environ. Sci. Technol. **41**, 4158–4163 (2007)

41. Rajendran, C.: The effect of zinc oxide and titanium dioxide nanoparticles in the Comet assay with UVA photoactivation of human sperm and lymphocytes. Nanotoxicology **3**, 33–39 (2009)

42. Lin, W.: Toxicity of nano- and micro-sized ZnO particles in human lung epithelial cells. J. Nanopart. Res. **11**, 25–39 (2009)

43. Yin, H.: Effects of surface chemistry on cytotoxicity, genotoxicity, and the generation of reactive oxygen species induced by ZnO nanoparticles. Langmuir **26**, 15399–15408 (2010)

44. Yoshida, R.: Mutagenicity of water-soluble ZnO nanoparticles in Ames test. J. Toxicol. Sci. **34**, 119–122 (2009)

45. Colognato, R.: Comparative genotoxicity of cobalt nanoparticles and ions on human peripheral leukocytes in vitro. Mutagenesis **23**, 377–382 (2008)

46. Colognato, R.: Analysis of cobalt ferrite nanoparticles induced genotoxicity on human peripheral lymphocytes: comparison of size and organic grafting-dependent effects. Nanotoxicology **1**, 301–308 (2007)

47. Bhabra, G.: Nanoparticles can cause DNA damage across a cellular barrier. Nat. Nanotechnol. **4**, 876–883 (2009)

48. Karlsson, H.L.: Copper oxide nanoparticles are highly toxic: a comparison between metal oxide nanoparticles and carbon nanotubes. Chem. Res. Toxicol. **21**, 1726–1732 (2008)

49. Pan, X.: Mutagenicity evaluation of metal oxide NPs by the bacterial reverse mutation assay. Chemosphere **79**, 113–116 (2010)

Gold Nanorods

Yang Lu[1] and Jun Lou[2]
[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Department of Mechanical Engineering and Materials Science, Rice University 223 MEB, Houston, TX, USA

Synonyms

Cylindrical gold nanoparticles; Short (low aspect ratio) gold nanowires

Definition

A gold nanorod is defined as nanometer-sized gold particles that are rod shaped. These solid, cylindrical nanoparticles have uniform diameters ranging from a few nanometers to hundreds nanometers, with aspect ratios (length/diameter) larger than 1 but typically smaller than 10. Owing to their unique physical properties, gold nanowires/nanorods have many attractive applications in several areas, such as in life science, electronic and electromechanical devices, sensors, and solar cells. In the field of nanomedicine, the typical sizes for gold nanorods were around tens of nanometers in diameters and hundreds of nanometers in lengths. The crystalline structures of these gold nanomaterials can be either polycrystals or single crystals with different orientations, depending on the growth methods and conditions.

Key Research Findings

Surface Plasmon Resonance (SPR)

One of the most remarkable features of gold nanorods (NRs) is its unique and tunable localized surface plasmon resonance (LSPR) properties. When metals are reduced to nanoscale dimensions (a few nanometers to hundreds of nanometers), interesting physical properties such as LSPR emerge due to the restrictions imposed on their electron mobility, an effect

which also strongly depends on the overall shape of the metal. For nano-sized metallic particles, especially noble metal nanoparticles (such as gold, silver, platinum), they have the ability to confine resonant photons within their small particle size to induce the localized surface plasmon oscillations of conduction-band electrons. The surface plasmon is thus resonant at a specific frequency of the incident light and is called surface plasmon resonance (SPR). For spherical gold nanoparticles, the SPR properties can be described by Mie theory [1], which has one specific SPR frequency (wavelength ∼520 nm) for strong absorption. While for nanorods, the SPR oscillates along two directions of the rods: longitudinal and transverse directions. The excitation of the surface plasmon oscillation along the short (transverse) axis induces an absorption band in the visible light region at wavelength (520–530 nm) similar to that of spherical gold nanoparticles, referred to as the transverse band. The transverse band is insensitive to the aspect ratio of the nanorods. While the excitation of the surface plasmon oscillation along the longitudinal direction induces a much stronger absorption band in the longer wavelength region, referred to as the longitudinal band. The longitudinal SPR band of gold nanorods is very sensitive to the aspect ratio of the nanorods. Slight increases in the aspect ratio greatly red-shifts the absorption maximum (λ_{max}, nm) of the longitudinal band from visible to the near-infrared (NIR) region with a dramatic increase in the band intensity. According to the Gans' theory [2], which is an extension of Mie's theory to treat elongated ellipsoids, the absorption maximum of the longitudinal band is linearly dependent on the aspect ratio (R) of the nanorods by the following relationship $\lambda_{max} = 95R + 420$ [3]. This optical tunability of the nanorods in the NIR region via the varying rod aspect ratio provides great opportunities for in vivo medical applications.

Photoluminescence Properties

Usually bulk metals (such as gold, silver, and copper) have extremely weak fluorescence due to their rapid nonradiative electron–hole recombination. However, it has been observed that for short nanorods (aspect ratio < 3), the quantum efficiency of the photoluminescence (PL) is in the range of 10^{-4} to 10^{-3}, about six orders of magnitude greater than that found in the bulk [3]. In addition, the PL wavelength maximum is found to increase linearly with increasing aspect ratio (less than three). The quantum efficiency increases quadratically for aspect ratios below 3 and then begins to diminish thereafter. This can be understood by the fact that, as the SPR absorption shifts to longer wavelengths as the nanorod aspect ratio (AR) is increased, the extent of the overlap between the interband processes and the SPR absorption band will be constantly changing. When less than 3, increases of AR result in greater overlap responsible for an emission enhancement. At some point, such as AR ∼ 3.5 shown in the experimental works by Eustis and El-Sayed, further increases in AR diminish the degree of overlap resulting in an emission decline.

However, recently scientists have found that the quantum efficiency of long nanorods (length >200 nm, AR > 10) rapidly increases when the lengths are larger than 200 nm and their corresponding fluorescence intensity has been significantly enhanced [4]. In fact, the actual PL responses for metallic nanoparticles are much more complicated, and could be explained as multi-step processes [3, 4]. In brief, the strong dependency of photon emission on the nanorod length, combined with the fact that similar enhancements are not observed for sphere nanoparticles, provided evidences that the strong surface fields of the longitudinal surface plasmon oscillations are the origin of the fluorescence enhancement. This is also supported by the polarization of the emission being along the long rod axis [3]. The excitation of interband absorption simultaneously excites the longitudinal surface plasmon of gold nanorods, resulting in enhanced local electric fields amplifying both exciting and emission fields. The enhanced fluorescence of gold nanorods thus provides opportunities for imaging and diagnostics applications.

Photothermal Properties

Unlike aforementioned radiative properties, photothermal properties of gold nanorods are nonradiative processes, in which the absorbed light is converted into heat via a series of photo-physical processes [5]. Basically, these photo-physical processes begin with a fast loss of phase of the coherently excited electrons on the few femto-second timescale, then followed by energy loss to the phonon bath in the crystal lattice of gold nanorods. The relaxation process is size- and shape-independent and also independent for both the

transverse and longitudinal surface plasmon in the nanorods. Meanwhile, the lattice cools down by passing heat to the surrounding medium via phonon–phonon relaxation of $\sim$100 ps [3], which leads to the competitive process of cooling and heating of the whole nanorods. If the heating rate is much faster than the cooling rate, at appropriate timescale, heat can be accumulated within the lattice causing a local temperature rise of the nanorods in a short time, which could provide an effective means for selective photothermal therapy.

Synthesis and Surface Modifications

Although the synthesis of spherical gold nanoparticles can be dated back to 1857 by Faraday, fabrication of gold nanorods with well-controlled shape and crystalline structure emerged only during past decade [3]. Five typical gold nanorod/nanowire synthesis approaches will be described here: (1) seed-mediated growth [6, 7]; (2) electrochemical synthesis [8]; (3) photochemical reduction [9]; (4) lithographic methods [10, 11]; and (5) template-assisted growth [12].

Seed-Mediated Growth
In 2001, Murphy and coworkers [6] prepared colloidal gold nanorods by the addition of citrate-capped small gold nanospheres to a bulk $HAuCl_2$ growth solution obtained by the reduction of $HAuCl_4$ with ascorbic acid in the presence of cetyltrimethylammonium bromide (CTAB) surfactant and silver ions. This method was improved by the same group with a three-step procedure in the absence of silver nitrate for larger aspect-ratio (up to 25) nanorods. In a typical procedure, the gold nanospheres/nanorods in the previous step are sequentially used as "seeds" for the next step growth. Later on, El-Sayed and coworkers further improved the method by replacing sodium citrate with a stronger CTAB stabilizer in the seed formation process and utilizing silver ions to control the aspect ratio of gold nanorods [7]. These modifications resulted in reproducible formation of nanorods with aspect ratios ranging from 1.5 to 4.5 with desired lengths. To grow longer nanorods with aspect ratios ranging from 4.6 to 10, a binary surfactant mixture composed of benzyldimethylhexadecylammonium chloride (BDAC) and CTAB was used, with appropriate silver ion concentrations. The growth of gold

nanorods via seed-mediated method has been explained in two mechanisms [3]: One is the surfactant-preferential-binding-directed growth proposed by Murphy and coworkers based on high resolution TEM (HRTEM) imaging, which showed the atomic structure of the nanorods during various formation stages. It is believed that the intrinsic structural twinning in the face-centered cubic (fcc) metallic nanoparticles causes symmetry breaking to produce anisotropic nanoparticles. And the following anisotropic growth into nanorods is due to the preferential binding of the CTA^+ head group to faceted seed nanoparticles on the {100} faces [3]. The other theory is the electric-field-directed growth mechanism proposed by Mulvaney and coworkers, in which the $AuCl^{2-}$ ions bound to cationic micelles diffuse to seed nanoparticles at high curvature points due to higher electric double layer gradient at these points and thus leads to nanoparticle growth in rod shape. Finally, it might be noted that the effect of the silver ion content is not always to increase the aspect ratio of NRs. Nikoobakht et al. showed that by increasing the concentration of the $AgNO_3$ in the growth solution to more than a specific value, the length of the NRs started to decrease. The negative effect of the silver ions in higher concentration was probably due to their interaction with the bromide counter ion of the surfactant monomers [7].

Electrochemical Synthesis
Gold nanorods have also been prepared via electrochemical oxidation/reduction within a simple two-electrode type of cell [8]. A gold metal plate is used as the anode and a platinum plate is used as the cathode in the electrochemical cell. Both electrodes are immersed in an electrolytic solution consisting of a cationic surfactant, hexadecyltrimethylammonium bromide (CTAB), and a rod-inducing co-surfactant (tetraoctylammonium bromide, TOAB). The CTAB serves not only as the supporting electrolyte but also as the stabilizer for nanoparticles to prevent their further growth. During the synthesis, the bulk gold metal is converted from the anode to form gold nanoparticles mostly at the interfacial region of the cathode surface and within the electrolytic solution. A current-controlled electrolysis is used throughout the process and an appropriate amount of acetone added into the electrolytic solution is necessary [8]. The synthesis is conducted under an ultrasonication (for dissipating the nanorods away from cathode) and a controlled temperature (typically at 38°C). It is

suggested that the role of acetone is to facilitate the incorporation of cylindrical shape–inducing co-surfactant into the CTAB micellar framework and inducing the cylindrical growth to form the Au-CTAB-TOAB system [3, 8]. The typical rod-inducing co-surfactant (e.g., TOAB) employed is an overall much more hydrophobic cationic surfactant. Similarly, by inserting a silver plate into the solution and releasing silver ions, the aspect ratio of nanorods can be varied.

Photochemical Reduction

In this method, auric acid is bound to rod-like cationic micelle surfactants to form ion pairs, and then excited by UV light to reduce metal ions to form metals via electron transfer from metal ions to ligands [3]. It is proposed that the rod formation is due to a two-step aggregation process [9], that is, metal nuclei aggregation to form primary particles and then primary crystal aggregation to form rod-like structures due to specific crystal face stabilization by surfactant micelles. In this method, addition of sodium chloride electrolyte, the concentration of silver ions, and the UV light wavelength/intensity can adjust the growth speed, size, and structures of the nanorods.

Lithographic Methods

The most commonly utilized lithography technique for fabricating gold nanorods is electron-beam lithography (EBL) [10]. Nanoscale patterns written with an electron beam allow for the formation of nanometer-scale openings in the resist through which gold can be deposited. By lift-off of the remaining resist and the unwanted gold deposited on it, free-standing gold nanorods can be obtained with high-quality and well-controlled geometry. Another approach, so-called focused ion beam (FIB) lithography, utilizes a rastered ion beam to sputter away unwanted portions of a continuous film leaving behind nanostructures of the desired shape [11]. Both methods can fabricate samples with unparalleled control over the size, shape, and alignment. However, unlike previous mentioned chemical/photochemical/electrochemical synthesis methods, the crystalline structure cannot be well controlled. In addition, both methods are very time consuming and expensive.

Template-Assisted Growth

Template methods involve the use of nanoporous track-etched polycarbonate or anodized aluminum oxide (AAO) membrane [12] as the template where gold atoms are deposited by electrochemical reduction of gold ion precursor in an electrochemical cell. The subsequent chemical etching of the template membrane gives free-standing gold nanorods. The diameter of the nanorods is determined by the pore diameter of the membrane template. For AAO templates, uniform diameters ranging from tens of nanometers to hundreds of nanometers can be prepared by controlling the anodization process. The length of the rods is controlled by varying growth time and applied voltage/current (the amount of the gold deposited), and thus, the aspect ratio of the nanorods can be tuned accordingly. Similar to lithographic methods, crystalline structure is less controllable in this method. However, template method provided similar excellent control over the size (diameter, length, aspect ratio) at a much lower cost. In addition, it is possible to grow multi-segment nanorods (such as Au-Ni bimetal nanorods [13]) by switching templates between different electrolyte solutions, which brings opportunities to further introduce new functions to the original gold nanorods.

In summary, the first three approaches are bottom-up methods and the latter two are considered as top-down methods. Among these, the seed-mediated growth is the most popular method for the synthesis of colloidal gold nanorods due to the simplicity of the procedure, high quality and yield of nanorods, ease of particle size/aspect ratio controlling, and flexibility for structural modifications [3]. However, this method, as well as other bottom-up methods which involve surface surfactant, usually results in a double layer of cetyltrimethylammonium bromide (CTAB), which is problematic for bio-conjugation, nonspecific absorption of DNA, cytotoxicity, and stability [14]. These factors have severely limited the use of gold nanorods in biological and nanomedicine applications. Therefore, it is necessary to apply surface modification for gold nanorods fabricated by the bottom-up methods, to increase stability, facilitating surface chemistry, tuning properties, and broadening the practical applications.

Surface Modifications

Surface modifications of gold nanorods have attracted considerable interest in recent years, and generally two types of surface modifications have been developed.

One is to introduce or exchange surface surfactant (e.g., CTAB) by other chemical or biological

functional groups on nanorod surface. For example, Takahashi et al. [15] utilized phosphatidylcholine (PC) to replace the CTAB and successfully reduced the cytotoxicity. More recently, researchers applied ligand exchange and customization by using thiol group [14] and found that the functionalized nanorods had very strong binding with DNA which demonstrated potential applications in biomedical fields.

The other is to coat nanorods surface with hard inorganic materials. Firstly, gold nanorods themselves are ideal templates for further growth of other metallic nanomaterials. By further growth of silver, or platinum, or palladium thin layer(s), the surface properties of Au nanorods could be modified or improved by forming such core-shell structures. For example, silver coating can significantly enhance Raman scattering [3]. Secondly, one could also coat gold nanorod surface with iron oxide or silicon oxide hard layers. SiO_2 is of special interests since it is highly biocompatible, stable, and hydrophilic, and Au nanorods coated with a thin layer of silicon oxide can be easily adopted for further surface functionalizations [16].

Examples of Applications

Drug/Gene Delivery and Selective Release

The mechanism is to use the shape transformation and heating effect of gold nanorods under exposure to near-infrared (NIR) laser pulses to result in the drug or DNA release due to surface atom rearrangement or breaking of the chemical bonds. For example, Wijaya et al. used gold nanorods with two different lengths to carry two different DNAs [17]. When applied laser with appropriate wavelength, selective release of DNA could be made and the releasing rate was up to 50–80%. In the work by Wu et al. [18], the thiolated gene of enhanced green fluorescence protein (EGFP) was covalently conjugated to gold nanorods. Under laser irradiation, gold nanorods underwent significant shape transformation, resulting in the release of EGFP DNA.

Molecular Imaging

Gold nanorods can be ideal imaging probes or contrast agents for molecular imaging applications, due to their unusual optical properties at NIR region. Compared to conventional contrast agents – such as radionucleotide in

PET and SPECT, gadolinium compounds in MRI, and dye molecules in optical fluorescence imaging [3] – gold nanorods can act as NIR fluorescence probe and have two main advantages: firstly, near-IR laser introduces less damage to living cells; secondly, the penetrating depth for NIR laser in biological tissues is much larger than that for visible light, up to hundreds of mm. These make near-IR imaging with gold nanorods ideal for in vivo diagnostics. For example, Durr et al. modified the surface of gold nanorods with anti-epidermal growth factor receptor (anti-EGFR) antibody, and attached them onto cancel cell selectively [19]. By performing two-photon fluorescence (TPF) on the three-dimensional tissue with cancer cells, researchers found the fluorescence intensity due to gold nanorod labeling was about three orders of magnitude higher than the intensity of cancer cell fluorescence illumination [19].

Surface Plasmon Resonance Sensing

Due to the fact that the SPR frequency highly depends on the dielectric constant of the surrounding medium, increasing the dielectric constant of the surrounding medium leads to a red shift of the SPR wavelength. Therefore, the SPR wavelength either by absorption or scattering provides great opportunity to monitor the changes of the local environment of the nanoparticles, and thus can be used for sensing. Compared to gold nanospheres, the unique transverse and longitudinal SPR characteristics for gold nanorods provide higher accuracy: the longitudinal SPR absorption is very sensitive to the change of surrounding environment while the transverse SPR will have no obvious red shift and can be served as a reference. In addition, the characteristics SPR responses for gold nanorods with different aspect ratios can be utilized for higher efficient diagnostics. For example, Yu et al. attached three types of antibody molecules to three different gold nanorods (with different AR) and multiplexing detection of three targets was successfully realized via the observation of three LSPR red shifts [20].

Photothermal Therapy

The recent interest in the use of gold nanorods in photothermal therapy rooted from the advantages of their small size, excellent stability, high efficiency, and tunable absorption in the biological window. Gold nanorods become very popular also because they can absorb light in the NIR window (650–900 nm) where light penetration is optimal due to minimal absorption

by water and hemoglobin in the tissue as discussed earlier. In addition, these plasmonic nanoparticles could enable dual imaging and therapy functions at the same time. For example, Huang et al. provided an in vitro demonstration of gold nanorods as contrast agents for both molecular imaging and photothermal cancer therapy [3]. In their work, nanorods were conjugated to anti-epidermal growth factor receptor (anti-EGFR) monoclonal antibodies and incubated in cell cultures with a nonmalignant epithelial cell line (HaCat) and two malignant oral epithelial cell lines. The anti-EGFR antibody-conjugated nanorods binded specifically to the surface of the malignant-type cells with a much higher affinity due to the over expressed EGFR on the cytoplasmic membrane of the malignant cells [3]. As a result of the strongly scattered red light from gold nanorods in dark field, observed using a laboratory microscope, the malignant cells were clearly visualized and diagnosed from the nonmalignant cells. It was found that, after exposure to continuous red laser at 800 nm, malignant cells require about half of the laser energy to be photothermally destroyed than the nonmalignant cells. Thus, both efficient cancer cell diagnostics and selective photothermal therapy were realized at the same time.

Future Directions

Despite the fact that gold nanorods have already exhibited a wide range of promising applications and great potentials in disease diagnosis and treatment, there remain plenty of things to be done before they can actually advance into the realm of clinical trials. First of all, although some of the biomedical applications, such as photothermal cancer therapy, have been successfully demonstrated in animal experiments, more in vivo experimental data are definitely needed in order to establish a complete set of parameters for translational research and clinical practice. Secondly, considerable refinements need to be made, for example, by further introducing hyper-structures to the nanorods, such as porous gold nanostructures, or advanced surface coatings, so that drug/gene delivery efficiency and imaging/sensing mechanisms could be correspondingly tuned. Last but not the least, rigorous studies on the long-term effects, such as distribution and fate of gold nanorods in the human body, should be carried out carefully.

Cross-References

▶ Nanomedicine

References

1. Mie, G.: Considerations on the optics of turbid media, especially colloidal metal sols. Ann. Phys. **25**, 377 (1908)
2. Gans, R.: Form of ultramicroscopic particles of silver. Ann. Phys. **47**, 270 (1915)
3. Huang, X., Neretina, S., El-Sayed, M.A.: Gold nanorods: from synthesis and properties to biological and biomedical applications. Adv. Mater. **21**, 4880–4910 (2009)
4. Li, C.Z., Male, K.B., Hrapovic, S., Luong, J.H.T.: Fluorescence properties of gold nanorods and their application for DNA biosensing. Chem. Commun. **2005**, 3924–3926 (2005)
5. Link, S., El-Sayed, M.A.: Spectral properties and relaxation dynamics of surface plasmon electronic oscillations in gold and silver nanodots and nanorods. J. Phys. Chem. B **103**, 8410–8426 (1999)
6. Jana, N.R., Gearheart, L., Murphy, C.J.: Seed-mediated growth approach for shape controlled synthesis of spheroidal and rodlike gold nanoparticles using a surfactant template. Adv. Mater. **13**, 1389–1393 (2001)
7. Nikoobakht, B., El-Sayed, M.A.: Preparation and growth mechanism of gold nanorods (NRs) using seed-mediated growth method. Chem. Mater. **15**, 1957–1962 (2003)
8. Yu, Y.Y., Chang, S.S., Lee, C.L., Chris Wang, C.R.: Gold nanorods: electrochemical synthesis and optical properties. J. Phys. Chem. B **101**, 6661–6664 (1997)
9. Kim, F., Song, J.H., Yang, P.: Photochemical synthesis of gold nanorods. J. Am. Chem. Soc. **124**, 14316–14317 (2002)
10. Billot, L., Chapelle, L.D.L., Grimault, A.S., Barchiesi, A.V.D., Bijeon, J.L., Adam, P.M., Royer, P.: Surface enhanced Raman scattering on gold nanowire arrays: evidence of strong multipolar surface plasmon resonance enhancement. Chem. Phys. Lett. **422**, 303–307 (2006)
11. Babu Dayal, P., Koyama, F.: Polarization control of 0.85 μm vertical-cavity surface-emitting lasers integrated with gold nanorod arrays. Appl. Phys. Lett. **91**, 111107 (2007)
12. Sander, M.S., Tan, L.S.: Nanoparticle arrays on surfaces fabricated using anodic alumina films as templates. Adv. Funct. Mater. **13**, 393–397 (2003)
13. Salem, A.K., Searson, P.C., Leong, K.W.: Multifunctional nanorods for gene delivery. Nat. Mater. **2**, 668–671 (2003)
14. Wijaya, A., Hamad-Schifferli, K.: Ligand customization and DNA functionalization of gold nanorods via round-trip phase transfer ligand exchange. Langmuir **24**, 9966–9969 (2008)
15. Takahashi, H., Niidome, Y., Niidome, T., Kaneko, K., Kawasaki, H., Yamada, S.: Modification of gold nanorods using phosphatidylcholine to reduce cytotoxicity. Langmuir **22**(1), 2–5 (2006)
16. Pastoriza-Santos, I., Perez-Juste, J., Liz-Marzan, L.M.: Silica-coating and hydrophobation of CTAB-stabilized gold nanorods. Chem. Mater. **18**, 2465–2467 (2006)
17. Wijaya, A., Schaffer, S.B., Pallares, I.G., Hamad-Schifferli, K.: Selective release of multiple DNA oligonucleotides from gold nanorods. ACS Nano. **3**(1), 80–86 (2009)

18. Chen, C.C., Lin, Y.P., Wang, C.W., Tzeng, H.C., Wu, C.H., Chen, Y.C., Chen, C.P., Chen, L.C., Wu, Y.C.: DNA-gold nanorod conjugates for remote control of localized gene expression by near infrared irradiation. J. Am. Chem. Soc. **128**(11), 3709–3715 (2006)

19. Durr, N.J., Larson, T., Smith, D.K., Korgel, B.A., Sokolov, K., Ben-Yakar, A.: Two-photon luminescence imaging of cancer cells using molecularly targeted gold nanorods. Nano Lett. **7**(4), 941–945 (2007)

20. Yu, C., Irudayaraj, J.: Multiplex biosensor using gold nanorods. Anal. Chem. **79**(2), 572–579 (2007)

Graphene

Debdeep Jena
Department of Electrical Engineering,
University of Notre Dame, Notre Dame, IN, USA

Synonyms

Bilayer graphene; Graphite; Nanoribbons

Definition

Graphene is defined as an atomically thin planar allotrope of carbon. Each carbon atom in this two-dimensional (2D) crystal is bonded to three other carbon atoms, forming a hexagonal structure. Graphene is the building block of graphite, carbon nanotubes, and fullerenes. It is the thinnest crystal known at this time.

Overview

The various solid forms of carbon have been known and used since thousands of years. In three-dimensional (3D) form, charcoal is the prototype solid fuel used to sustain life through the ages. Lavoisier first showed in 1772 that diamond was composed entirely of carbon by "burning" diamond in oxygen and producing carbon dioxide. This 3D crystalline form of carbon is the hardest, electrically insulating, and most thermally conductive solid known. These properties of diamond form the basis for its many uses. Graphite is a layered solid form of carbon, and is more stable than diamond. It is remarkably different from diamond in its physical properties. It is an electrical conductor, exhibits high anisotropy in its electrical and thermal conductivities, and is very refractory. Because it is capable of carrying copious amounts of electrical current, graphite was instrumental in the discovery of the light bulb, which converts electrical energy into photons for illumination. It is still used in arc lamps and where high currents and high temperature operation is necessary. The uses of graphite are more diverse, ranging from neutron moderation in nuclear reactors to the pencil lead, which exploits its layered structure. The rich range of physical properties of the various forms of carbon emerges from seemingly minor differences in the microscopic arrangements of the carbon atoms.

With the advent of electron microscopy, it became possible to investigate the atomic structure of solids made of carbon. Lower dimensional forms of carbon were then discovered. In semiconductor heterostructures, such as those made of silicon or GaAs, man-made quantum-size effect materials were created starting from 3D bulk semiconductors to 2D quantum wells, 1D quantum wires, and 0D quantum dots. Thus, nanostructures based on semiconductor heterostructures developed from large to small. In sharp contrast, the first carbon nanostructure discovered was zero-dimensional (0D) fullerenes (C_{60}), followed by 1D carbon nanotubes. The 2D form of carbon, graphene, has been known since the 1940s to be the building block of graphite. It was also found subsequently to be the building block of fullerenes and nanotubes. But by itself, a single layer of graphene was believed to be thermodynamically unstable. Based on the Mermin–Wagner theorem in solid-state physics, the energy cost for spontaneously broken symmetries in 2D crystals is low. This means that a perfect "flat" 2D crystal can increase its entropy by forming ripples or crumples. Thus, to a certain extent, the isolation of 2D atomic crystals in 2004 was a surprise. Using scotch tape, single layer flakes of graphene were isolated from graphite, and their electronic properties were investigated [1]. More than the isolation of single layers of graphene, the findings of the nature of electron motion in it have far-reaching consequences on the fundamental understanding of the electronic properties of solids. These findings have opened a new door to aspects of symmetry in condensed matter physics. Rapid progress has occurred in exploiting these properties in applications. To understand the origins of the peculiar properties of graphene, its atomic structure is discussed next.

a

unit cell

basis

B

A

carbon atoms

1.24Å

Graphene Crystal

b

Hopping energy: $\gamma_0 \approx 3$ eV

y

x

sp^2 bonds

P_z orbitals

Chemical bonds in graphene
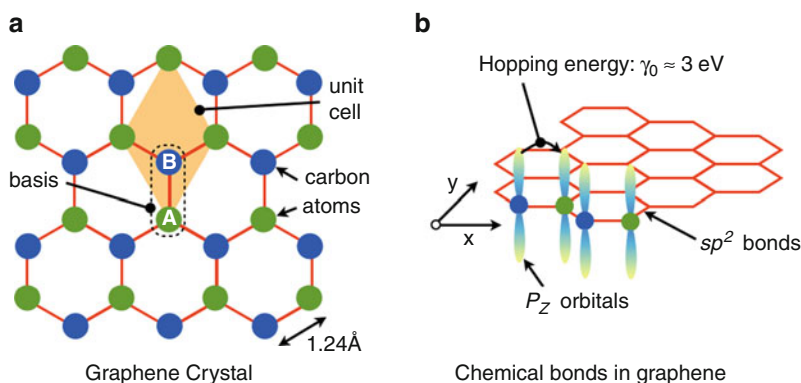
Crystal Structure of Graphene

Carbon is the sixth element in the periodic table, meaning it has six protons in its nucleus. Charge neutrality requires the presence of six electrons. The six electrons fill the lowest three orbitals in the arrangement $[1s]^2[2s]^2[2p]^2$, where the $|s\rangle$ orbitals are spherically symmetric, and $|p\rangle$ orbitals are elongated along a linear axis. When carbon atoms are brought close to each other, the two electrons in the innermost $[1s]^2$ orbitals remain close to the nucleus, and do not participate in the bonding process. The outermost orbitals start interacting, and are capable of forming chemical bonds. The four electrons in the outermost $[2s]^2[2p]^2$ orbitals can be shared between atoms in a variety of ways, each leading to a different form of carbon. For example, when four nearest carbon atoms share all four electrons of a carbon atom equally, the bond is sp^3. An sp^3 bond is tetrahedral, and is inherently 3D. Carbon atoms bonded in this fashion arrange periodically in 3D space, forming the crystal structure of diamond. Carbon is rather versatile in its bonding possibilities; when combined with hydrogen atoms, electrons in carbon atoms can form sp^2 or sp bonds as well. This means carbon atoms can form 2D planar molecules (e.g., the hexagonal molecule benzene or C_6H_6), or long 1-dimensional chains (e.g., in conductive polymers). This versatility of electrons in carbon is the root of the complexity of organic molecules that lead to the emergence of life itself.

When a carbon atom shares electrons with *three* nearest neighbors, the bonding is sp^2, written as $|sp^2\rangle = a|s\rangle + b|p_x\rangle + c|p_y\rangle$, where *a, b, c* are constants. This bonding leaves one electron behind for every carbon atom. The three chemical bonds are

separated by $120°$, and are planar. This is the crystal structure of graphene, depicted in Fig. 1. The energies of the electrons forming the sp^2 bonds bunch together into bands called the σ band. The σ band is responsible for the perfectly planar nature of graphene, and determines its structural and vibrational properties, such as its thermal conductivity and Young's modulus, among others. Thus, graphene is a 2D crystal with hexagonal arrangements of carbon atoms. The atomic spacing is $a_{cc} \sim 0.124$ nm. As indicated in Fig. 1, the smallest repeating unit that can be translated to cover the entire 2D space consists of not one, but *two nearest-neighbor carbon atoms*. Though the atoms are physically identical (they are both carbon), they are labeled as A and B in Fig. 1 since the "mechanical" symmetry can be broken by the electron wavefunction associated with them. The A-B atom combination is the two-atom "basis" of the graphene crystal, similar to that in 3D diamond or silicon. The primitive repeating unit is shown as a parallelogram. By virtue of being 1-atom thick, graphene is the "thinnest" crystal. It is impermeable to small gas molecules. The structural and chemical rigidity of graphene is protected by symmetry: to form a chemical bond to graphene, one needs to convert sp^2 bonds to sp^3. This requires buckling of the crystal since the bonds have to change from 2D to 3D, requiring the movement of atoms. If and when this is achieved, the material is no longer graphene. In addition, carbon is the lightest in group-IV of the periodic table, and the strong σ bonds lead to the highest in-plane thermal conductivity. However, the σ band does not directly influence the electronic properties of graphene. Electrons in sp^3 bonds and bands resulting from them determine the electronic *and* structural properties in conventional 3D semiconductor crystals (e.g., Si, GaAs, diamond).
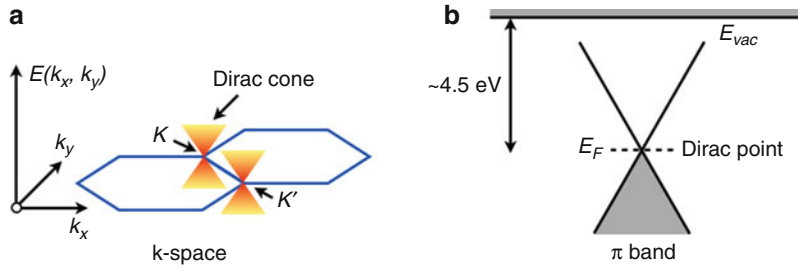
Graphene, Fig. 2 Allowed energies for electrons in graphene in the k-space, and bandstructure. (**a**) The hexagonal symmetry of electrons in the momentum space, and the Dirac cones originating from the mobile p electrons. (**b**) The bandstructure, showing electron states filled in the p band till the Dirac point energy, which is 4.5 eV below the vacuum energy level

In graphene, the electronic and structural properties arise from different bands. For understanding the electronic properties of graphene, attention is focused back to the *electron that was left out* in the formation of the sp^2 bond.

Electronic Properties of Graphene

The electrons left out in the chemical bonding process are from the $|p_z\rangle$ orbitals of the carbon atoms forming graphene. The linear axis of the $|p_z\rangle$ orbital protrudes outward from the plane, as shown in Fig. 1. These electrons can hop between the nearest neighbor carbon atoms. The energy characterizing this hopping is $\gamma_0 \sim 3.0$ eV. Thus, their wavefunctions become delocalized. Their allowed energies bunch together to form what are called π bands. Electrons in the π bands can be visualized as sliding over the hexagonal carbon atom structure of graphene. These electrons impart graphene its exceptional electrical conductivity.

The allowed energies of the delocalized π electrons depend on their wavelength, in accordance with the rules of quantum mechanics. According to the de-Broglie relation, the wavelength is inversely proportional to momentum. The momentum is given by $\hbar k$, where $\hbar$ is Planck's constant and $k = 2\pi/\lambda$, where λ is the wavelength. The allowed energies of electrons $E(k_x, k_y)$ can be then plotted in the k-space, as shown in Fig. 2. By taking into account the basis of two atoms, the wavefunctions of electrons in graphene are represented as "spinors," which are two-component matrices of the form $\frac{exp[i\mathbf{k}\cdot\mathbf{r}]}{\sqrt{2}}\begin{pmatrix} 1 \\ exp[i\theta] \end{pmatrix}$, where $\mathbf{k} = \mathbf{x}k_x + \mathbf{y}k_y$, and $tan\theta = k_y/k_x$. Using a tight-binding procedure, the energy spectrum $E(k_x, k_y)$ is given by the equation $\hbar v_F(\vec{\sigma}\cdot\mathbf{k})|\psi\rangle = E(k_x, k_y)|\psi\rangle$, where $\vec{\sigma} = \mathbf{x}\sigma_x + \mathbf{y}\sigma_y + \mathbf{z}\sigma_z$ is a vector composed of the three Pauli spin matrices. The term $v_F = 3a_{cc}\gamma_0/2\hbar \sim 10^8$ cm/s is a parameter characterizing the bandstructure, with dimensions of velocity. This Hamiltonian has a closer analogy to the Dirac equation rather than the Schrodinger equation used for conventional semiconductors. The Dirac Hamiltonian consists of Pauli spin matrices in the kinetic energy term, which implies that electrons have spin-like properties based on their momentum. This property is termed "pseudo-spin," and may be pictured as a vector whose direction depends on the momentum. In traversing a path around a constant-energy circle in the Dirac cone, the wavefunction accumulates a phase of π due to the pseudospin. This phase is called Berry's phase, and has consequences on scattering and on quantum Hall effects.

The energy bandstructure obtained from this "Dirac equation" for the π band is given by $E(k_x, k_x) = E_F \pm \gamma_0\sqrt{1 + 4cos(\sqrt{3}k_xa/2)cos(k_ya/2) + 4cos^2(k_ya/2)}$, where $a = \sqrt{3}a_{cc}$ is the lattice constant of graphene. Here, E_F is the Fermi energy, which for pristine undoped graphene is $\sim$4.5 eV below the vacuum energy level. This is the electron affinity or the work function of pure graphene (Fig. 2). The positive sign is for the conduction band that lies above E_F, and the negative sign is for the valence band. One can note that $E(0, 4\pi/3a) = E_F \pm 0$ eV, which implies that there is *no energy bandgap* in the π band of graphene. The points in the k-space such as $K = (0, 4\pi/3a)$ where the conduction and valence bands meet are called the Dirac points. These points form a hexagonal network in the k-space as indicated in Fig. 2, with the

nearest-neighbor points labeled as K and K'. In analogy to traditional semiconductors, graphene is a direct bandgap semiconductor. Since each K point is shared between three hexagons, there are two distinct points in the smallest repeating unit cell in k-space, called the Brillouin zone. Thus, in analogy to traditional semiconductors, graphene has a valley degeneracy of 2.

The Fermi energy is at the Dirac point for pure graphene, meaning at zero temperature and in the absence of doping, all electronic states below the Dirac point are occupied, and all above it are empty. Electrical conductivity thus occurs in graphene due to transitions of electrons between allowed states close to the Dirac point. In analogy to the effective mass approximation for semiconductors, expanding the energy dispersion around the Dirac point, say around $K = (0, 4\pi/3a)$, yields the rather interesting relation $E(k_x, k_x) - E_F \approx \hbar v_F \sqrt{k_x^2 + k_y^2} = \hbar v_F |\mathbf{k}|$. This is the equation of a cone, and owing to its origin in the Dirac-like Hamiltonian, referred to as the Dirac cone. The conical form of the energy dispersion remains valid for energies exceeding a few eVs from the Dirac point energy, which is the range of interest for all low-energy electronic and optical properties. A few intriguing aspects of this energy dispersion can be identified by comparison with traditional semiconductors.
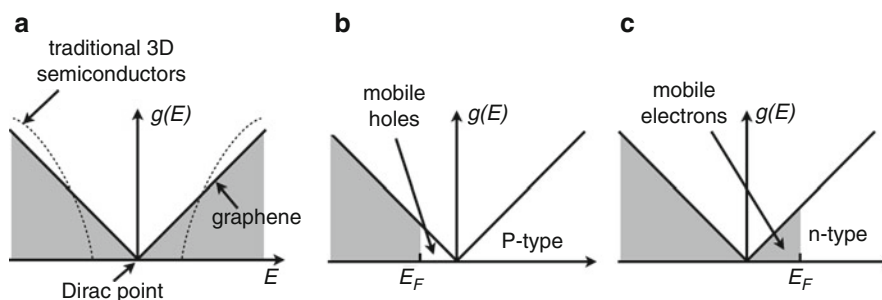
In traditional semiconductors, the kinetic energy of electrons near a band edge E_C is proportional to the square of the electron quasi-momentum $E(k) - E_C = (\hbar k)^2/2m^*$, with the magnitude of the kinetic energy parameterized by an effective mass distinct from the mass of free electrons. The effective mass m^* is a parameter that absorbs the details of the crystal potential. It replaces the complex motion of free electrons in a crystal by that of an electron with a different mass in free space. This is a considerable simplification, one that is heavily leveraged in the design of electronic and optical devices such as transistors, light-emitting diodes, and lasers based on traditional semiconductors. In stark contrast, the kinetic energy of electrons in graphene depends on its momentum *linearly* $E(k_x, k_x) = v_F \cdot \hbar |\mathbf{k}|$, where Fermi energy at the Dirac point is used as reference. Wallace originally derived this linear form of the bandstructure of graphene in 1947 [2].

A linear dependence of kinetic energy on momentum is characteristic of light or photons. The energy dispersion of electrons in graphene is identical to that of photons, with the speed of light $c = 3 \times 10^8$m/s replaced by the "Fermi velocity" $v_F \approx 10^6$m/s. The maximum speed of electrons in graphene is $v_F/c = 1/300$ that of light in free space. In traditional semiconductors, particularly those of small bandgaps, the velocity of carriers can also approach these speeds. A consequence of the linear energy dispersion is the difficulty in associating an effective mass to electrons in graphene. The effective mass is *defined* as the curvature of the energy dispersion. By definition, the curvature of a linear curve is zero. Thus, electrons in graphene are referred to as "massless Dirac fermions." The "massless" feature should be understood in the light of the discussion above; electrons in graphene are identical to electrons in any other material. Unlike *truly massless photons*, they need external driving forces such as electric or magnetic fields to carry currents. *What is unique about graphene is the linear energy dispersion of electrons*. In the dispersion of electrons are hidden a number of symmetries, each leading to many of its unique features.

Carrier Statistics in Graphene

The absence of an energy bandgap in graphene is related to the light-like linear dispersion. The linear dispersion also implies that the transport properties of filled states in the conduction band (see Fig. 2) and empty states in the valence bands (holes) are similar. The conduction band is a mirror-reflection of the valence band. This symmetry is special. Two-dimensional electron or hole gases in semiconductor quantum wells have multiple subbands to occupy, and can be truly 2D at very low energies or at low temperatures as long as only one subband is occupied. In stark contrast, electrons and holes in 2D graphene are truly 2D at all temperatures, since there are no higher subbands near the Dirac point. Based on the energy dispersion, the density of states (DOS) of graphene is given by $g(E) = 2|E|/\pi(\hbar v_F)^2$, which is linear in energy. Three-dimensional semiconductors have density of states that depend on energy as $g(E) \sim \sqrt{E - E_0}$, whereas the density of states of 2D subbands in quantum wells are energy-independent. The DOS of 2D graphene is shown in comparison to traditional 3D semiconductors in Fig. 3. Based on the DOS, the electron and hole densities in pristine graphene are given by $n_0 = p_0 = (\pi/6) \cdot (kT/\hbar v_F)^2$, where T is the temperature and k is the Boltzmann constant. At room temperature, there are $\sim 8 \times 10^{10}$/cm^2 electrons in the

a

traditional 3D
semiconductors

$g(E)$

graphene

Dirac point E

b

mobile
holes

$g(E)$

P-type

E_F

c

mobile
electrons

$g(E)$

n-type

E_F

Graphene, Fig. 3 Density of states of pure graphene and p-type and n-type graphene. (**a**) Density of states of graphene is linear and symmetric around the Dirac point energy. It has no energy gap, unlike conventional semiconductors. (**b**) and (**c**) P-type and n-type graphene created by moving the Fermi level into the valence and conduction band, respectively

conduction band due to interband thermal excitation, leaving behind an equal number of holes in the valence band. These carriers are mobile, and lead to a finite conductivity of graphene at room temperature even in the absence of intentional doping. They set the *lower limit* of mobile carrier concentration in 2D graphene: with any changes such as doping or field-effect, the net mobile carrier density can only exceed this number. The $\sim T^2$ dependence of intrinsic interband thermal generation of carriers is in stark contrast to traditional semiconductors, where the density depends on the bandgap as $exp(-E_G/2kT)$ where E_G is the bandgap.

When graphene is electrostatically doped, the Fermi energy moves away from the Dirac point into the valence or the conduction band, making it p-type or n-type. Under these conditions, the carrier concentrations are given by $n = E_F^2/\pi(\hbar v_F)^2$. Thus, the Fermi energy changes as the square root of the carrier concentration, $E_F = \hbar v_F \sqrt{\pi n}$. At a 2D carrier concentration of 10^{12}/cm^2, $E_F \sim 0.12$ eV from the Dirac point, and increases to $E_F \sim 0.37$ eV at 10^{13}/cm^2. The Fermi energy is tunable over a wide range, which implies a tunable work function for 2D graphene. The electrons or holes thus formed in graphene are responsible for its electrical conductivity. The current per unit width that flows in a graphene sheet in response to a voltage applied across contacts made to graphene is given by $I/W = qn\langle v \rangle$, where $\langle v \rangle$ is the ensemble-averaged velocity. In the absence of any scattering, the velocity of *all* carriers in graphene is the Fermi velocity v_F, by virtue of its linear bandstructure. However, since a net current flows only along the electric field, the effective velocity is only a fraction of the Fermi velocity obtained appropriately by angular averaging. The current drive in clean defect-free graphene is typically high, owing to the high Fermi velocity. In the presence of defects, however, the conductivity suffers due to scattering.

Carrier Transport in Graphene

The mobility of electrons in graphene is typically high, if the material is pure. Electron and hole mobilities well in excess of 50,000 cm^2/Vs have been measured at room temperature in the cleanest samples [3]. The high mobility is typically degraded due to scattering from impurities that are not inherent to graphene itself. For example, charged impurities located outside the plane of graphene result in Coulombic scattering of mobile carriers. Similarly, integrating a dielectric in close proximity to graphene allows electrons in graphene to excite phonon modes in them, resulting in "remote" phonon scattering [4]. When these scattering processes are suppressed, the mobility of carriers in graphene is found to increase. The high mobilities lead to long mean-free paths, and ballistic transport is possible over micrometer length scales. At high voltages, when electrons gain kinetic energies high enough to excite the intrinsic optical phonon modes of graphene, they collide with the carbon atoms and create optical phonons. By virtue of the lightness of carbon atoms, the optical phonon energy is high ($\sim$160 meV), implying that carriers cannot emit optical phonons until they acquire an equivalent kinetic energy, resulting in high velocities.

The high mobility and velocity of carriers leads to high currents in field-effect transistor-like device applications. Due to the lack of a bandgap, the demonstrated current modulation in 2D graphene typically does not
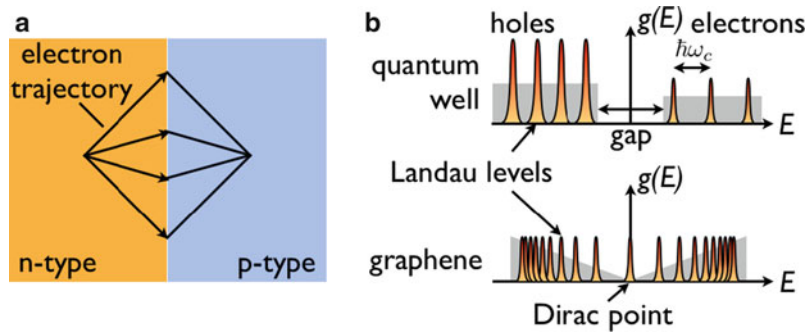
Graphene, Fig. 4 Electron "lensing" in graphene, and Landau levels in traditional semiconductors compared to those in 2D graphene. (**a**) Electron transport in graphene mimics light; a p-n junction behaves as a lens for electrons. (**b**) At high magnetic fields, the density of states of graphene collapses into Landau levels that are quite different from what is observed in 2D electron–hole systems in conventional semiconductors

exceed a factor of 10X, currently restricting the feasibility of 2D graphene in electronic digital switching. However, applications that do not require a high current modulation stand to benefit from the high current drive and the high speed. Unlike traditional semiconductors, the symmetry of the bandstructure and the absence of a bandgap implies that graphene is as conductive when p-type as when it is n-type. Thus, when the Fermi energy is tuned from the valence band to the conduction band, the current characteristics change from hole conduction to electron conduction symmetrically; this property is termed ambipolar. The ambipolar symmetry is useful for passive applications in electronic circuits such as mixers to generate higher harmonics of oscillatory signals. In most traditional semiconductors, the valence band effective mass is higher than the conduction band effective mass, and the ambipolar nature is not protected by symmetry, as in graphene.

Adjacent regions of graphene can be electrostatically gated with opposite polarity, forming p-n junctions as shown in Fig. 4. In such a case, when electron waves are launched from a point in the n-region at various angles to the interface, they have to conserve energy and momentum when they enter the p-region. Since they have to tunnel from the conduction band in the n-side to the valence band in the p-side, the component of momentum parallel to the junction is flipped, resulting in the collimation of the electron waves. The tunneling process in graphene is unique owing to its Dirac dispersion, and is called Klein tunneling. This sort of electron "lensing" is a consequence of the light-like energy dispersion of graphene [5]. It requires the phase of the electron waves to be coherent throughout the transport process, and thus requires long mean free paths and low scattering. Similarly, electrons can be reflected from similar interfaces, and by combining junctions, can be "waveguided" from one region to another through all-electrical controls. The similarity of the electron energy dispersion to that of light makes such processes possible in graphene.

The quantum Hall effect in graphene offers a signature of its energy dispersion. As shown in Fig. 4, in traditional semiconductor quantum wells, 2D electron and hole gases have different effective masses, and hence different density of states. Upon the application of a high magnetic field B, the DOS collapse into uniformly spaced Landau levels, with energies peaking at $E_n = \hbar\omega_C(n + \frac{1}{2})$ for each band, where $\omega_C = qB/m^*$ is the cyclotron frequency and $n = 0, 1, 2, \ldots$ is an integer. As the height of each Landau level and their separation is changed by varying the magnetic field, the Fermi level passes through peaks and minima of the Landau level, causing the longitudinal magnetoresistance to oscillate. The transverse resistance is the Hall resistance. The quantized Hall *conductance* in traditional semiconductor 2DEGs is given by $\sigma_{xy} = \frac{2q^2}{h} n$. The Landau levels of 2D graphene, on the other hand, are given by $E_n = \pm v_F \sqrt{2q\hbar B} \sqrt{n}$, bunching closer as n increases, as shown in Fig. 4. The resulting Hall conductance is given by $\sigma_{xy} = \frac{4q^2}{h}(n + \frac{1}{2})$. For $n = 0$, there is a Landau level at the Dirac point for graphene, and the half-integer quantum Hall conductance is a signature of both the Dirac cone dispersion and the Berry phase of

Graphene,
Fig. 5 Bandstructure of bilayer graphene in the absence (**a**), and the presence (**b**) of an external field. Bilayer graphene is one of the first materials that has an electrically tunable energy bandgap
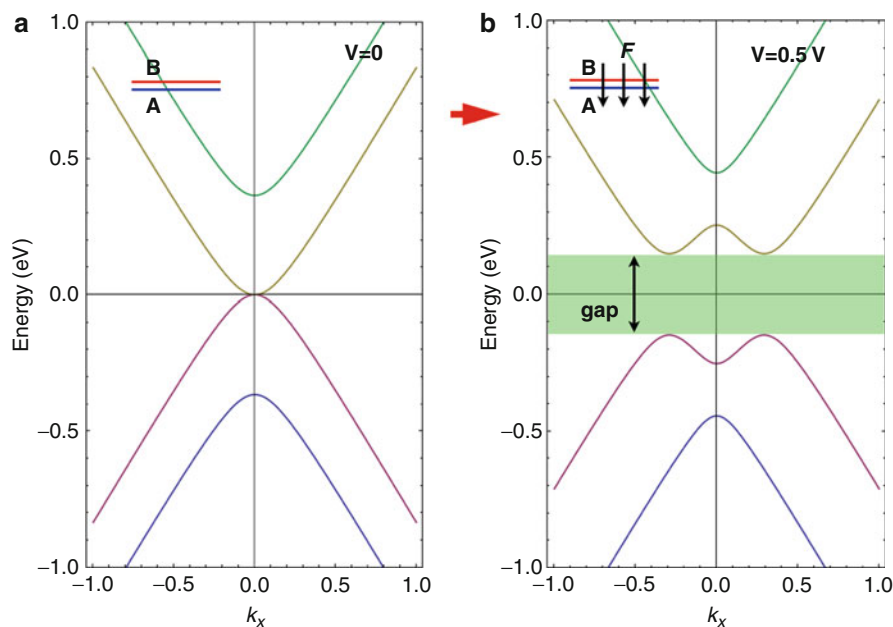
π for electron wavefunctions [6, 7]. The half-integer quantum Hall effect has been observed in single layer graphene grown by various methods, and the precision of the quantization is high.

Bilayer Graphene

When two layers of graphene are stacked, the mutual interaction changes the electron energy dispersion, and hence its electronic properties. Bilayer graphene should not be considered as two individual layers of graphene, but one composite electronic system. Bernal stacking is a particular type of stacking, where the A atoms in one layer (Fig. 1) are vertically above the B atoms of the other. For this form of bilayer graphene, the energy dispersion changes from the Dirac cone feature of a single layer to a parabolic bandstructure characterized by an effective mass $m^* \approx 0.033m_0$, but still with a zero bandgap. This bandstructure is shown in Fig. 5. A remarkable phenomenon occurs upon the application of an electric field vertical to the graphene bilayer. The bandstructure deforms to that shown in Fig. 5b. An energy bandgap appears, which is tunable with the applied electric field [8]. The electric field breaks the symmetry of electron wavefunctions in the direction perpendicular to the bilayer, and this broken symmetry results in a bandgap. Such field-tunable energy bandgaps are rare in solid state systems, and

can be expected to lead to novel applications. Energy bandgaps opened in this fashion up to ~0.2 eV have been experimentally measured [9].

The symmetry of the bandstructure of graphene has also led to the proposal of interlayer tunneling devices such as the BISFET [10]. This device mimics Bose-condensation of electron–hole pairs (excitons) in two graphene layers close enough to allow strong interactions, yet far enough to prevent the formation of a bilayer. It exploits collective many-body effects in an all-electronic system at room temperature. As the number of graphene layers stacked increases, the energy bandstructure evolves toward that of bulk graphite. Graphite is a semimetal, where the conduction band minimum and the valence band maximum occur at different points in the k-space, and they overlap in energy. The semimetallic nature of graphite is responsible for its high current carrying capacity.

Graphene Nanostructures

When graphene sheets are "cut" into thin nanoscale ribbons, the size confinement results in the selection of a finite number of modes that can propagate freely along the axis of the ribbon. These structures are called graphene nanoribbons (GNRs) [11]. The edge of
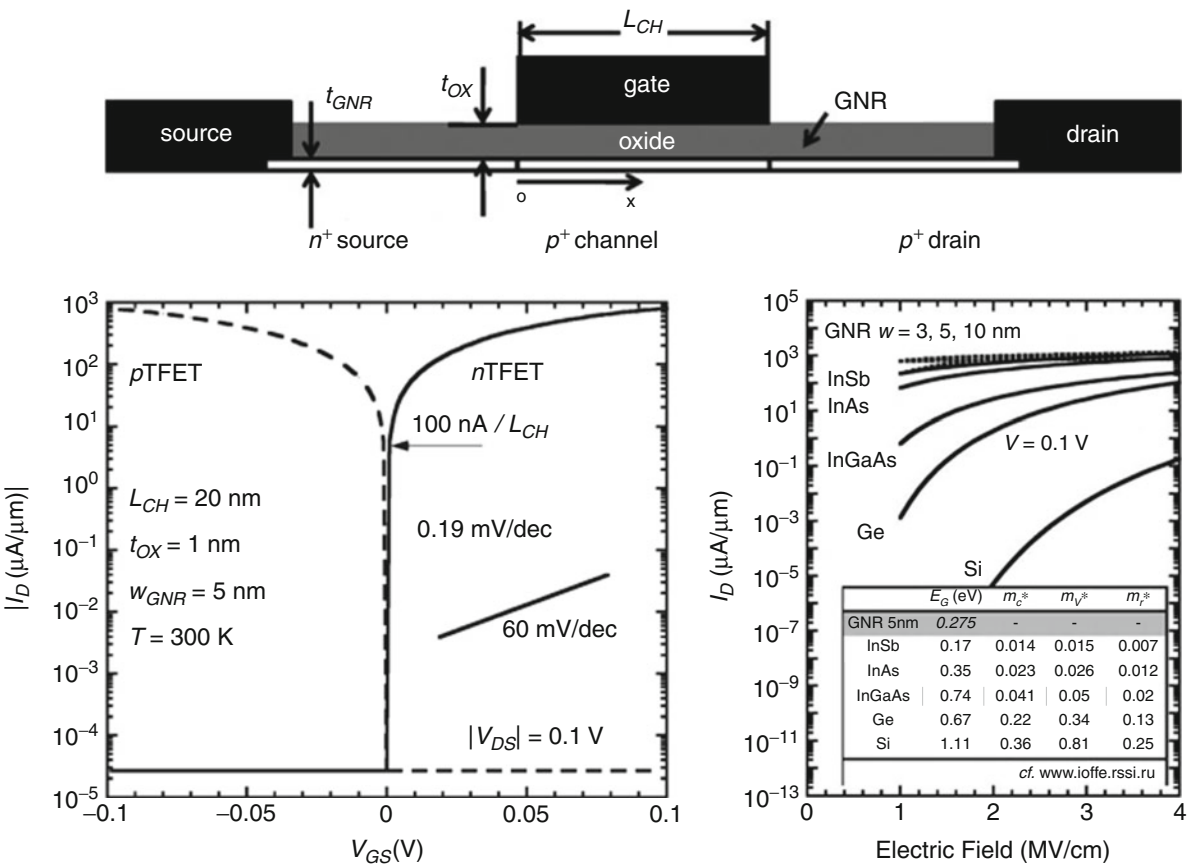
Graphene, Fig. 6 Graphene nanoribbon–based tunneling field-effect transistors. Proposed device structure and predicted performance showing vast advantages over state-of-the-art silicon CMOS devices

a GNR can be armchair or zigzag, and depending on the edge and the width, the allowed propagating modes can be metallic or semiconducting. In this aspect, GNRs are similar to carbon nanotubes (CNTs), except that the nanotube has no edges. A CNT is a rolled up GNR. Semiconducting GNRs have a bandgap $E_g \approx 1.4/W$ eV where W is the width of the GNR in nm. Since GNRs lie flat on substrates, they can dissipate heat efficiently, and are capable of high current drives. In addition, the existence of a bandgap implies they can be used for electronic switching. GNRs retain the symmetry of the bandstructure of the graphene sheet they are carved from, in the sense that the valence band of a GNR is a mirror image of the conduction band. Thus, semiconducting GNRs can be ambipolar, with a high current modulation. In the process of fabrication, if the edges of GNRs are rough, scattering from this disorder can reduce the electron mobility and

carrier velocity. Nevertheless, GNRs can prove attractive for making electronic devices in the future.

If a p-n junction can be fabricated from GNRs, then the structure is very attractive for an unconventional field-effect transistor. Recently, a lot of research has focused on lowering the energy cost of digital switching by transistors. An attractive candidate for replacing conventional silicon field-effect transistor relies on the interband tunneling of electrons. The device is called a tunneling field-effect transistor or TFET; it promises digital switching of current at smaller voltages [12]. High current drives are required in transistors to maintain high performance. The tunneling currents in traditional semiconductors are typically low due to the wide bandgaps. However, devices based on GNRs can deliver among the highest current drives among all homojunction TFETs [13]. Figure 6 shows the concept of the device,

along with its predicted energy band diagram and switching performance. Owing to the symmetry of the bandstructure, a GNR TFET makes complementary operation feasible, a feature difficult to attain in similar devices made of traditional semiconductor heterostructures. In addition, the absence of out-of-plane bonds and the planar nature implies that GNRs and graphene are simpler to integrate with existing semiconductor technologies.

Optical Properties of Graphene

Interaction of light with graphene is understood from its energy spectrum. Single layers of graphene are found to absorb a well-defined fraction of visible light passing through them by interband excitation of electrons. This fraction is given by $\pi q^2 / \hbar c = \pi \cdot \alpha \sim 2.3\%$ where α is the fine structure constant [14]. The absorption coefficient is *independent* of the photon energy over a large spectral window. The photogenerated electrons and holes can lead to photocurrent, and aided by the high mobility, 2D graphene can be used for fast photodetectors that operate over a large spectral window. The ease of integration of graphene with traditional semiconductors makes it attractive for applications in optical communication devices.

Electrons in graphene can also absorb photons of very long wavelengths and small energy (in the THz spectral window). This happens by the process of free-carrier absorption, and requires momentum scattering, as in a traditional semiconductor. Since the number of photons absorbed is related to the circumference of the Fermi circle in the k-space, by tuning the Fermi energy of graphene, the number of free carriers available for such absorption can be tuned over a large range. Thus, graphene is attractive for making THz modulators at a given frequency [15]. Further, graphene can be patterned into periodic structures similar to photonic crystals to enhance the absorption of certain modes and orientations of incident THz light. Such structures called graphene metamaterials are similar to photonic crystals made of dielectrics, and hold much promise for devices exploiting light–matter interaction.

Since graphene has zero-bandgap with a high mobility, an electromagnetic wave can create waves of electrons and holes that support its propagation. Such modes where free carriers on the surface of a conductor carry part of the electromagnetic energy of an electromagnetic wave are called surface plasmon modes. The high mobility leads to a low loss, and plasmons are thus useful for confining and controlling the flow of light in dimensions shorter than its wavelength. Such sub-wavelength waveguiding is attractive for on-chip optical communications in the future. This can enable transistors far from each other on a microprocessor chip, or between chips to exchange information by using waveguided light instead of electrons in conductors that can be lossy and have *RC* bandwidths.

Methods of Fabrication

Single-layer graphene flakes were first isolated by mechanical exfoliation from graphite. They were transferred on to substrates with $\sim$300 nm SiO_2 on conductive silicon. The dielectric contrast of the substrate at optical wavelengths made it possible to observe the atomically thin crystal under a microscope. Since this initial breakthrough, several methods of preparing graphene have been developed [16]. Large-area single-layer graphene is formed on the surface of single crystal SiC wafers by sublimation. Under specific conditions, the few layers of Si and C on the surface dissociate, Si atoms leave the crystal, and the remaining carbon atoms rearrange to form graphene. A larger-scale method of fabricating single-layer graphene is by chemical-vapor deposition of carbon on heated copper (or other metallic) foils [17]. Owing to the low solubility of carbon in copper, it "precipitates" out, forming a single layer of graphene on the surface. The process is self-limiting, and leads to single-layer graphene of large dimensions. The graphene is subsequently transferred on to insulating substrates by removing the metallic foil. Growth and fabrication of graphene are still at their infancy. The methods used to grow graphene, and the resulting crystalline quality is expected to improve with time.

Other Applications

At present, the high conductivity of graphene and its ultrathin nature makes it attractive for transparent conductive electrodes for flexible electronics. The high in-plane thermal conductivity of graphene can be harnessed for effective heat spreading in various nano-scale structures; graphene can scale this usage of graphite

down to nanoscale dimensions. The light mass of the atoms in graphene, the high bond strength, and atomically thin structure implies that microelectro-mechanical (MEMs) or nanoelectro-mechanical (NEMs) structures made from graphene can operate at very high frequencies, and extend the range of applicability of existing devices. The ease of field-effect of graphene also makes it especially conducive to the fabrication of gas and charge sensors in the field-effect geometry.

Other 2D Crystals and Future Directions

The symmetry of the bandstructure of graphene led to the proposal, and eventual discovery of a new class of materials called topological insulators [18]. These materials have a graphene-like Dirac cone energy dispersion for surface states, whereas the bulk states have energy bandgaps as in traditional semiconductors. The valley degeneracy of such materials is one, so in a sense they are "half" of graphene. Electron motion in the surface bands of topological insulators are protected against scattering by geometric constraints, and are currently under much scrutiny.

The discovery of graphene also opened the door to the search for other 2D crystals, and a number of them have been identified. Among them, boron nitride (BN) is an excellent example. Boron and nitrogen straddle carbon in the periodic table, and is the compound semiconductor analogue of graphene. However, the broken symmetry in the crystal basis itself creates a very large bandgap, approaching ∼5.4 eV. Thus, BN is an excellent insulator, and quite distinct from other insulating dielectrics since it is composed of sp^2 bonds. The absence of out-of plane or broken bonds implies that it is chemically and electronically more inert than its 3D insulating counterparts. Indeed, graphene when placed on BN in planar fashion exhibits high electron mobilities and this is attributed to the reduced electronic coupling between the two crystals. The transition-metal dichalcogenides (such as MoS_2) are an analogous family of 2D crystals that exhibit semiconducting as well as metallic properties. However, these 2D crystals are not perfectly planar, as their chemical bonds are different from sp^2; they involve d-orbitals. The presence of d-orbitals leads to a richer range of electronic phenomena, ranging from metallic, semiconducting [19], magnetic, to superconducting. Combining this rich range of electronic properties of 2D crystals in heterostructures is expected to lead to new generations of more efficient and fundamentally new devices in the future. Combining 2D crystals by stacking is expected to be simpler than 3D materials since there are no out-of-plane bonds to join, or break [20].

Cross-References

▶ Active Carbon Nanotube-Polymer Composites
▶ Active Plasmonic Devices
▶ Carbon Nanotube-Metal Contact
▶ Carbon Nanotubes for Chip Interconnections
▶ Carbon-Nanotubes
▶ Flexible Electronics
▶ Fullerenes for Drug Delivery
▶ Light Localization for Nano-Optical Devices
▶ Modeling Thermal Properties of Carbon Nanostructure Composites
▶ Nanostructure Field Effect Transistor Biosensors
▶ Nanostructured Materials for Sensing
▶ Nanostructures for Photonics
▶ Nanotechnology
▶ Physical Vapor Deposition
▶ Terahertz Technology for Nano Applications
▶ Thermal Conductivity and Phonon Transport

References

1. Novoselov, K.S., Jiang, D., Schedin, F., Booth, T.J., Khotkevich, V.V., Morozov, S.V., Geim, A.K.: Two-dimensional atomic crystals. Proc. Natl. Acad. Sci. **102**, 10451–10453 (2005)
2. Wallace, P.R.: The band theory of graphite. Phys. Rev. **71**, 622–634 (1947)
3. Dean, C.R., Young, A.F., Meric, I., Lee, C., Wang, L., Sorgenfrei, S., Watanabe, K., Taniguchi, T., Kim, P., Shepard, K.L., Hone, J.: Boron nitride substrates for high-quality graphene electronics. Nat. Nanotechnol. **5**, 722–726 (2010)
4. Konar, A., Fang, T., Jena, D.: Effect of high-κ gate dielectrics on charge transport in graphene-based field effect transistors. Phys. Rev. B **82**, 115452–115458 (2010)
5. Cheianov, V.V., Fal'ko, V., Altshuler, B.L.: The focusing of electron flow and a veselago lens in graphene p-n junctions. Science **315**, 1252–1255 (2007)
6. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I., Grigorieva, I.V., Dubonos, S.V., Firsov, A.A.: Two-dimensional gas of massless Dirac fermions in graphene. Nature **438**, 197–200 (2005)
7. Zhang, Y., Tan, Y.-W., Stormer, H.L., Kim, P.: Experimental observation of the quantum hall effect and Berry's phase in graphene. Nature **438**, 201–204 (2005)
8. McCann, E.: Asymmetry gap in the electronic band structure of bilayer graphene. Phys. Rev. B **74**, 161403–161406 (2006)

9. Zhang, Y., Tang, T., Girit, C., Hao, Z., Martin, M.C., Zettl, A., Crommie, M.F., Shen, Y.R., Wang, F.: Direct observation of a widely tunable bandgap in bilayer graphene. Nature **459**, 820–823 (2009)

10. Banerjee, S.K., Register, L.F., Tutuc, E., Reddy, D., MacDonald, A.H.: Bilayer pseudospin field-effect transistor (BiSFET): a proposed new logic device. IEEE Electron Device Lett. **30**, 158–160 (2009)

11. Han, M.Y., Ozyilmaz, B., Zhang, Y., Kim, P.: Energy bandgap engineering of graphene nanoribbons. Phys. Rev. Lett. **98**, 206805 (2007)

12. Zhang, Q., Zhao, W., Seabaugh, A.: Low-subthreshold-swing tunnel transistors. IEEE Electron Device Lett. **27**, 297–299 (2006)

13. Zhang, Q., Fang, T., Seabaugh, A., Xing, H., Jena, D.: Graphene nanoribbon tunnel transistors. IEEE Electron Device Lett. **29**, 1344–1346 (2008)

14. Nair, R.R., Blake, P., Grigorenko, A.N., Novoselov, K.S., Booth, T.J., Stauber, T., Peres, N.M.R., Geim, A.K.: Fine structure constant defines visual transparency of graphene. Science **320**, 1308 (2008)

15. Sensale-Rodriguez, B., Fang, T., Yan, R., Kelly, M., Jena, D., Liu, L., Xing, H.: Unique prospects of graphene-based THz modulators. Appl. Phys. Lett. **99**, 113104–113106 (2011)

16. Berger, C., Song, Z., Li, X., Wu, X., Brown, N., Naud, C., Mayou, D., Li, T., Hass, J., Marchenkov, A.N., Conrad, E.H., First, P.N., de Heer, W.A.: Electronic confinement and coherence in patterned epitaxial graphene. Science **312**, 1191–1196 (2006)

17. Li, X., Cai, W., An, J., Kim, S., Nah, J., Yang, D., Piner, R., Velamakanni, A., Jung, I., Tutuc, E., Banerjee, S.K., Colombo, L., Ruoff, R.S.: Large-area synthesis of high-quality and uniform graphene films on copper foils. Science **324**, 1312–1314 (2009)

18. Hasan, M.Z., Kane, C.L.: Topological insulators. Rev. Mod. Phys. **82**, 3045–3067 (2010)

19. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., Kis, A.: Single-layer MoS_2 transistors. Nat. Nanotechnol. **6**, 147–150 (2011)

20. Novoselov, K.S.: Nobel lecture: graphene: materials in the flatland. Rev. Mod. Phys. **83**, 837–849 (2011)

Graphene Synthesis

Graphite

Greases

Green Tribology and Nanoscience

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

Synonyms

Ecotribology

Definition

Tribology (from the Greek word τρίβω "tribo" meaning "to rub") is the branch of science and technology concerned with interacting surfaces in relative motion and with associated matters (as friction, wear, lubrication, and the design of bearings). Green tribology is the science and technology of the tribological aspects of ecological balance and of environmental and biological impacts.

Research Areas

Green tribology deals with the ecological aspects of technology involving friction, lubrication, wear, solid surfaces, coatings, as well as applications of this technology. This makes green tribology closely related to the novel areas of tribology: nanotribology, biotribology, and biomimetics. The term "green tribology" was suggested in about 2009 by several engineers and economists. In 2010, the first scientific volume was published by Philosophical Transactions of the Royal Society [1]. A number of symposia and conference sections devoted to ecological and green tribology took place starting from approximately the same time, as well as books prepared for publication [2].

Green or environment-friendly tribology emphasizes the aspects of interacting surfaces in relative motion, which are of importance for energy or environmental sustainability or which have impact upon today's environment. This includes tribological technology that mimics living nature (biomimetic surfaces) and thus is expected to be environment-friendly,

the control of friction and wear that is of importance for energy conservation and conversion, environmental aspects of lubrication (including natural oil lubrication and self-lubrication), surface modification techniques, and tribological aspects of green applications, such as wind-power turbines, tidal turbines, or solar panels. A number of tribological problems could be put under the umbrella of "green tribology," and they are of mutual benefit to one another.

Green tribology can be viewed in the broader context of two other "green" areas: green engineering and green chemistry. The U.S. Environmental Protection Agency (EPA) defines green engineering as "the design, commercialization, and use of processes and products that are technically and economically feasible while minimizing (1) generation of pollution at the source (2) risk to human health and the environment." Green chemistry, also known as sustainable chemistry, is defined as the design of chemical products and processes that reduce or eliminate the use or generation of hazardous substances.

Biomimetic and Self-Lubricating Materials/Surfaces

Biomimetic materials are also usually environmentally friendly in a natural way since they are a natural part of the ecosystem. For this reason, the biomimetic approach in tribology is particularly promising. In the area of biomimetic surfaces, a number of ideas have been suggested, including the Lotus-effect, gecko-effect, micro- and nanostructured surfaces, texturing, self-lubrication, hydrophobicity, oleophobicity, etc.

Environmental engineers have only just started paying attention to biomimetic surfaces. One research indicates that while the use phase benefits are apparent, production burdens can outweigh them when compared with other cleaning methods, so a more thoughtful and deliberate use of bio-inspiration in sustainable engineering is needed. Clearly, more studies are likely to emerge in the near future.

Biodegradable Lubrication

In the area of environment-friendly and biodegradable lubrication, several ideas have been suggested: This includes the use of natural (e.g., vegetable-oil-based or animal-fat-based) biodegradable lubricants for engines, hydraulic applications, and metal cutting applications. In particular, corn, soybean, coconut oils have been used so far (the latter is of particular interest in tropical countries such as India). These lubricants

are potentially biodegradable, although in some cases chemical modification or additives for best performance are required. Vegetable oils can have excellent lubricity, far superior than that of mineral oil. In addition, they have a very high viscosity index and high flash/fire points. However, natural oils often lack sufficient oxidative stability, which means that the oil will oxidize rather quickly during use, becoming thick and polymerizing to a plastic-like consistency. Chemical modification of vegetable oils and/or the use of antioxidants can address this problem.

Powder lubricants and, in particular, boric acid lubricants, in general, tend to be much more ecologically friendly than the traditional liquid lubricants. Boric acid and MoS_2 powder can also be used as an additive to the natural oil. Friction and wear experiments show that the nanoscale (20 nm) particle boric acid additive lubricants significantly outperformed all of the other lubricants with respect to frictional and wear performance. In fact, the nanoscale boric acid powder-based lubricants exhibited a wear rate more than an order of magnitude lower than the MoS_2 and larger-sized boric acid additive based lubricants.

Self-replenishing lubrication that uses oil-free environmentally benign powders for lubrication of critical components such as bearings used in fuel cell compressors and expanders.

Tribology of Renewable Energy Sources

The tribology of renewable sources of energy is a relatively new field of tribology. Today, there are meetings and sections devoted to the tribology of wind turbines at almost every tribology conference, and they cover certain issues specific for these applications. Unlike in the case of the biomimetic approach and environment-friendly lubrication, it is not the manufacturing or operation, but the very application of the tribological system which involves "green" issues, namely, environmentally friendly energy production. The following issues can be mentioned [1].

Wind power turbines have a number of specific problems related to their tribology and constitute a well-established area of tribological research. These issues include water contamination, electric arcing on generator bearings, issues related to the wear of the mainshaft and gearbox bearings and gears, the erosion of blades (solid particles, cavitation, rain, hail stones), etc.

Tidal power turbines are another important way of producing renewable energy, which involves certain

tribiological problems. Tidal power turbines are especially popular in Europe (particularly, in the U.K.), which remains the leader in this area, although several potential sites in North America have been suggested. There are several specific tribological issues related to the tidal power turbines, such as their lubrication (seawater, oils, and greases), erosion, corrosion, and biofouling, as well as the interaction between these modes of damage.

Besides tidal, the ocean water flow energy and river flow energy (without dams) can be used with the application of special turbines, such as the Gorlov helical turbine, which provides the same direction of rotation independent of the direction of the current flow. These applications also involve specific tribological issues.

Geothermal energy plants are used in the USA (in particular, at the Pacific coast and Alaska); however, their use is limited to the geographical areas at the edges of tectonic plates. In 2007, they produced 2.7 GW of energy in the USA, with Philippines (2.0 GW) and Indonesia (1.0 GW) at the second and third place. There are several issues related to the tribology of geothermal energy sources which are discussed in the literature [1].

Principles

Nosonovsky and Bhushan [1] formulated the principles of green tribology, which belong to the three areas, suggested in the preceding section. Some principles are related to the design and manufacturing of tribological applications (3–10), while others belong to their operation (1–2 and 11–12). Tradition is followed and limited the number of principles is limited to 12.

1. *Minimization of heat and energy dissipation*. Friction is the primary source of energy dissipation. According to some estimates, about one third of the energy consumption in the USA is spent to overcome friction. Most energy dissipated by friction is converted into heat and leads to the heat pollution of atmosphere and the environment. The control of friction and friction minimization, which leads to both the energy conservation and prevention of the damage to the environment due to the heat pollution, is a primary task of ribology. It is recognized that for certain tribological applications (e.g., car brakes and clutches) high friction is required; however, ways of effective use of energy for these applications should be sought as well.

2. *Minimization of wear* is the second most important task of tribology which has relevance to green tribology. In most industrial applications, wear is undesirable. It limits the lifetime of components and therefore creates the problem of their recycling. Wear can lead also to catastrophic failure. In addition, wear creates debris and particles which contaminate the environment and can be hazardous for humans in certain situations. For example, wear debris generated after human joint replacement surgery is the primary source of long-term complications in patients.

3. *Reduction or complete elimination of lubrication and self-lubrication*. Lubrication is a focus of tribology since it leads to the reduction of friction and wear. However, lubrication can also lead to environmental hazards. It is desirable to reduce lubrication or achieve the self-lubricating regime when no external supply of lubrication is required. Tribological systems in living nature often operate in the self-lubricating regime. For example, joints form essentially a closed self-sustainable system.

4. *Natural lubrication* (e.g., vegetable-oil based) should be used in cases when possible since it is usually environmentally friendly.

5. *Biodegradable lubrication* should also be used when possible to avoid environmental contamination.

6. *Sustainable chemistry and green engineering principles* should be used for the manufacturing of new components for tribological applications, coatings, and lubricants.

7. *Biomimetic approach* should be used whenever possible. This includes biomimetic surfaces, materials, and other biomimetic and bio-inspired approaches since they tend to be more ecologically friendly.

8. *Surface texturing* should be applied to control surface properties. Conventional engineered surfaces have random roughness, and the randomness is the factor which makes it extremely difficult to overcome friction and wear. On the other hand, many biological functional surfaces have complex structures with hierarchical roughness, which defines their properties. Surface texturing provides a way to control many surface properties relevant to making tribo-systems more ecologically friendly.

9. *Environmental implications of coatings* and other methods of surface modification (texturing, depositions, etc.) should be investigated and taken into consideration.

10. *Design for degradation* of surfaces, coatings, and tribological components. Similar to green chemistry applications, the ultimate degradation/utilization should be taken into consideration during design.
11. *Real-time monitoring*, analysis, and control of tribological systems during their operation should be implemented to prevent the formation of hazardous substances.
12. *Sustainable energy applications* should become the priority of the tribological design as well as engineering design in general.

Cross-References

▶ Biomimetics
▶ Nanotribology

References

1. Nosonovsky, M., Bhushan, B.: Green tribology: principles, research areas and challenges. Philos. Trans R. Soc. A. **368**, 4677–4694 (2010)
2. Nosonovsky, M., Bhushan, B. (eds.): Green Tribology Biomimetics, Energy Conservation, and Sustainability. Springer, Heidelberg (2012)

Green's Function Integral Equation Method

Thomas Søndergaard
Department of Physics and Nanotechnology, Aalborg University, Aalborg Øst, Denmark

Synonyms

Lippmann–Schwinger integral equation method

Definition

The Green's function integral equation method (GFIEM) is a method for solving linear differential equations by expressing the solution in terms of an integral equation, where the integral involves an overlap integral between the solution itself and a Green's function. In particular, within nanotechnology the method is frequently applied to calculate scattering of light.

Overview

The Green's function integral technique can be used to solve a linear inhomogeneous differential equation in real space such as

$$\theta_0 \varphi(\mathbf{r}) = \pm f(\mathbf{r}), \tag{1}$$

where θ_0 is an operator, $\mathbf{r}$ is the position, $f(\mathbf{r})$ is a given source term, and $\varphi(\mathbf{r})$ is the function that should be calculated such that Eq. 1 is satisfied. A solution can in principle be obtained straightforwardly from a Green's function $g(\mathbf{r}, \mathbf{r}')$ of the operator θ_0. The Green's function satisfies the equation

$$\theta_0 g(\mathbf{r}, \mathbf{r}') = \pm \delta(\mathbf{r} - \mathbf{r}'), \tag{2}$$

where $\delta(\mathbf{r} - \mathbf{r}')$ is the Dirac delta function. A solution to the inhomogeneous equation (1) can be given straightforwardly by an overlap integral between the Green's function and the source term, i.e.,

$$\varphi(\mathbf{r}) = \int g(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') d^3 r. \tag{3}$$

It should be recognized that there can be many solutions to Eq. 2 and thereby many possible Green's functions. Usually, it is also required that the solution $\varphi(\mathbf{r})$ must satisfy certain boundary conditions, and in many cases it is possible to choose the specific Green's function that will automatically make the expression (3) satisfy these boundary conditions. An example of the above procedure, which is usually not discussed in the context of Green's functions, is the electrostatic potential $\varphi(\mathbf{r})$ due to a distribution of charges in free space, in which case the potential must satisfy the equation [1]

$$\nabla^2 \varphi(\mathbf{r}) = -\rho(\mathbf{r})/\varepsilon_0, \tag{4}$$

where $\rho(\mathbf{r})$ is the charge distribution, and ε_0 is the vacuum permittivity. The appropriate Green's function for finding the solution being zero at infinite distance from the source region is given by

$$g(\mathbf{r}, \mathbf{r}') = 1/4\pi|\mathbf{r} - \mathbf{r}'|, \qquad (5)$$

which is a solution to

$$\nabla^2 g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'), \qquad (6)$$

and it satisfies the boundary condition that it becomes zero at infinity. Using Eqs. 3–5 it is now straightforward to write up the well-known solution for the potential [1]

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r'. \qquad (7)$$

This may all seem very simple. However, this is only true if the Green's function is known and is in a form that is easy to use. Analytic expressions for Green's functions are usually only known for a few simple cases. Furthermore, in many cases the Green's function has a singularity that requires special attention in a numerical solution scheme. Consider an equation of the form

$$(\theta_0 + \Delta\theta)(\varphi_0(\mathbf{r}) + \Delta\varphi(\mathbf{r})) = 0, \qquad (8)$$

where the operator $\theta_0 + \Delta\theta$ is the sum of an unmodified operator with a known Green's function having appropriate boundary conditions, and a modification $\Delta\theta$. The solution $\varphi(\mathbf{r}) = (\varphi_0(\mathbf{r}) + \Delta\varphi(\mathbf{r}))$ should be in the form of the sum of a given solution $\varphi_0(\mathbf{r})$ for the case that $\Delta\theta = 0$ and a modification $\Delta\varphi(\mathbf{r})$ that must disappear when $\Delta\theta = 0$. This can be rewritten in the form

$$\theta_0(\varphi(\mathbf{r}) - \varphi_0(\mathbf{r})) = -\Delta\theta\varphi(\mathbf{r}), \qquad (9)$$

since $\theta_0\varphi_0(\mathbf{r}) = 0$. Equation 9 now has the form of an "inhomogeneous" equation, and a solution can be obtained using Eq. 3, i.e.,

$$\varphi(\mathbf{r}) = \varphi_0(\mathbf{r}) + \int g(\mathbf{r}, \mathbf{r}')\Delta\theta(\mathbf{r}')\varphi(\mathbf{r}')d^3 r', \qquad (10)$$

which is an integral equation since $\varphi(\mathbf{r})$ appears on both sides. One possible solution scheme is to divide the domain in which $\Delta\theta \neq 0$ in small elements and to assume that φ is constant within each element, in which case the discretized equation (10) becomes a linear system of equations in the discrete values of φ that can be put on matrix form and solved numerically on a computer.

In the following, integral equations of the form (10) shall be constructed for a range of simple cases of interest for scattering problems in optics and electromagnetics. The first case that will be considered is the scalar wave equation in one and two dimensions. Then, the numerically more challenging case of the vector wave equation in three dimensions shall be considered. The examples are for the case of monochromatic light with angular frequency ω. The real solution is obtained from the complex solutions obtained in the following by multiplying with $\exp(i\omega t)$ and then taking the real part of the resulting expression. Throughout this chapter the factor $\exp(i\omega t)$ will be suppressed.

Scalar Green's Function Integral Equation Method for One-Dimensional Scattering Problems

While the type of problems considered in this section can readily be solved by other methods such as, e.g., a transfer-matrix method [2], the case of a scalar one-dimensional scattering problem nevertheless provides a very instructive and transparent exemplification of the Green's function method, and serves as a good introduction to the more complex cases. As an example, consider light propagating from left to right being normally incident on a dielectric barrier (Fig. 1), a case that was previously considered in Ref. [3].

The incident field $E_0(x)$ is a solution to the wave equation for a structure where the dielectric barrier is not present, i.e.,
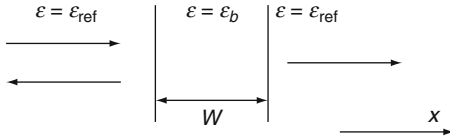
$$\left(\frac{\partial^2}{\partial x^2} + k_0^2\varepsilon_{ref}\right)E_0(x) = 0. \qquad (11)$$

A natural choice of incident field could be a plane wave propagating along the positive x-axis, i.e.,

$$E_0(x) = E_0 e^{-ik_0 n_{ref} x}, \qquad (12)$$

where $n_{ref} = \sqrt{\varepsilon_{ref}}$. The equation to be solved is the wave equation for the full structure including the barrier, namely,

$$\left(\frac{\partial^2}{\partial x^2} + k_0^2\varepsilon(x)\right)E(x) = 0, \qquad (13)$$

Green's Function Integral Equation Method, Fig. 1 Illustration of scattering of light by a dielectric barrier of dielectric constant ε_b and width w. The dielectric constant of the reference medium (without the barrier) is $\varepsilon_{\mathrm{ref}}$

Green's Function Integral Equation Method, Fig. 2 Electric field magnitude normalized relative to the incident field, i.e., $|E(x)/E_0(x)|$ for an incident wave with wavelength 633 nm (red light) on a dielectric barrier of width 500 nm and dielectric constant with and without absorption losses, namely, $\varepsilon = 4$ or a case with losses $\varepsilon = 4-i$

with the boundary condition that the solution must be the sum of the given incident field $E_0(x)$ and a scattered field component that propagates away from the barrier. Equations 11 and 13 can be rewritten in the form of the following "inhomogeneous" equation

$$\left(\frac{\partial^2}{\partial x^2} + k_0^2 \varepsilon_{ref}\right)(E(x) - E_0(x))$$
$$= -k_0^2\big(\varepsilon(x) - \varepsilon_{ref}\big)E(x). \tag{14}$$

The relevant Green's function for the reference medium must satisfy the equation

$$\left(\frac{\partial^2}{\partial x^2} + k_0^2 \varepsilon_{\mathrm{ref}}\right)g(x,x') = -\delta(x-x') \tag{15}$$

along with the radiating boundary condition, namely, that the solution must propagate away from the position x'. Equation 15 and this boundary condition uniquely determines the Green's function which is given by

$$g(x,x') = g(x-x') = \frac{1}{2ik_0 n_{\mathrm{ref}}} e^{-ik_0 n_{\mathrm{ref}}|x-x'|}. \tag{16}$$

A solution of Eq. 14 can then be obtained by solving the integral equation

$$E(x) = E_0(x)$$
$$+ \int g(x,x')k_0^2(\varepsilon(x') - \varepsilon_{\mathrm{ref}})E(x')dx'. \tag{17}$$

Due to the properties of the chosen Green's function, Eq. 17 will give the solution of interest since it satisfies the appropriate boundary condition, namely, that the total field must be the sum of the given incident field and a scattered field component propagating away from the scattering geometry.

An advantage in a numerical solution scheme of Eq. 17 compared with other methods, such as finite-

difference methods, is that initially it is sufficient to consider only the x, x' where $\varepsilon - \varepsilon_{\mathrm{ref}} \neq 0$. A simple approach is to, e.g., discretize the barrier in Fig. 1 in N elements sampled at positions x_1, x_2,...x_N with distance Δx between sampling points, and to assume that within each element the field is constant and given by $E_1, E_2,...E_N$, respectively. In that case, Eq. 17 can be approximated with the discrete linear system of equations

$$E_i = E_{0i} + \sum_j g_{ij}k_0^2\big(\varepsilon_j - \varepsilon_{\mathrm{ref}}\big)E_j\Delta x, \tag{18}$$

where $g_{ij} = g(x_i, x_j)$. This equation can be put in matrix form and readily solved on a computer. The calculated field inside the barrier can then be used in Eq. 17 to obtain the field at all other desired positions outside the barrier by direct integration over the field inside the barrier. Since a homogeneous reference medium is considered, the transmission through the barrier can, e.g., be calculated by considering $E(x)$ for a position x being larger than the barrier positions from the ratio $T = |E(x)/E_0(x)|^2$. Similarly, the reflection can be calculated by considering a position x being smaller than the barrier positions from the ratio $R = |(E(x) - E_0(x))/E_0(x)|^2$.

Finally, a few examples of calculations obtained with this method are presented that may also serve as
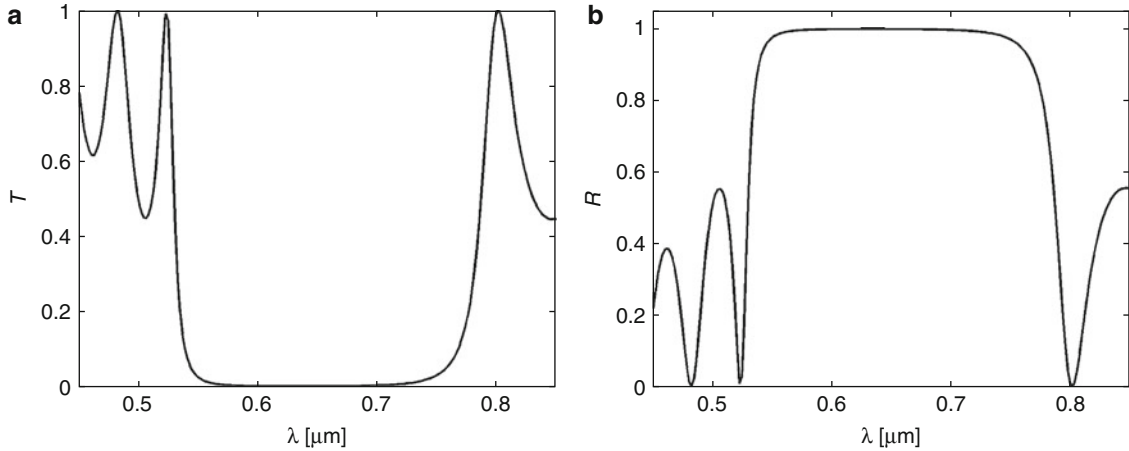
Green's Function Integral Equation Method, Fig. 3 Reflection and transmission spectra for light of wavelength λ being incident on a layered structure with alternating seven layers of

high refractive index $n_H = 2.32$ and thickness 633 nm/$4n_H$ and six layers of a smaller refractive index $n_L = 1.38$ and thickness 633 nm/$4n_L$. The surrounding material is air

an exercise for the reader, namely, the electric field in the case of scattering (reflection) from a dielectric barrier (Fig. 2), and reflection and transmission spectra for light incident on a layered structure or grating designed to strongly reflect red light (Fig. 3). These examples are very similar to instructive examples in Ref. [3].

Scalar Green's Function Integral Equation Method for Two-Dimensional Scattering Problems

In this section, the method from the previous section is extended to a two-dimensional scattering problem, namely, scattering of light by a very long (nano) wire (Fig. 4). Consider a situation with light propagating in the xy-plane and a wire oriented along the z-axis, and that the electric field is invariant along the z-axis and polarized along the z-axis, i.e., $\mathbf{E}(\mathbf{r}) = \hat{z}E(x, y)$. In that case the wave equation for the electric field reduces to the following scalar wave equation:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2\varepsilon(x, y)\right)E(x, y) = 0, \quad (19)$$

where $\varepsilon(x, y)$ is the dielectric constant of the total structure including the wire. Outside the wire the dielectric constant is given by ε_{ref}. Thus, also in this case is $(\varepsilon(x, y) - \varepsilon_{\text{ref}})$ the modification made to
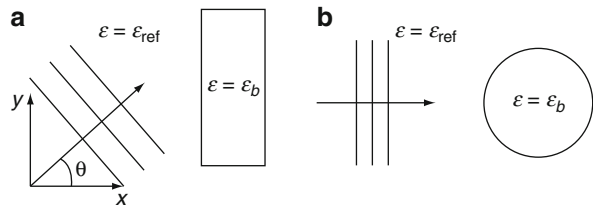
Green's Function Integral Equation Method, Fig. 4 Schematic of a plane wave being incident on a wire with (**a**) rectangular and (**b**) circular cross section

a reference structure, and it is only nonzero for positions inside the wire. The boundary condition when solving the differential equation (19) is also in this case that the total field $E(x, y)$ should be the sum of a given incident field $E_0(x, y)$ and a scattered field, where the scattered field is propagating away from the scattering object. The given incident field must be a solution to the wave equation in the case when there is no nanowire, i.e.,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2\varepsilon_{\text{ref}}\right)E_0(x, y) = 0. \quad (20)$$

For simplicity, the incident field can, e.g., be a plane wave propagating in the direction given by $\hat{s} = \hat{x}\cos\theta + \hat{y}\sin\theta$, i.e.,

$$E_0(x, y) = E_0 e^{-ik_0 n_{\text{ref}}(x\cos\theta + y\sin\theta)}. \quad (21)$$

From Eqs. 19 and 20, the following equation is obtained:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2 \varepsilon_{\text{ref}}\right)(E(x,y) - E_0(x,y))$$
$$= -k_0^2(\varepsilon(x,y) - \varepsilon_{\text{ref}})E(x,y). \tag{22}$$

While the method can be used for wires with any geometry, examples will be given for a wire with (a) rectangular and (b) circular cross section (see Fig. 4).

The appropriate Green's function is a solution to the equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2 \varepsilon_{\text{ref}}\right)g(x,y;x',y')$$
$$= -\delta(x-x')\delta(y-y'), \tag{23}$$

with the boundary condition that the solution must correspond to a field propagating away from the position $\mathbf{r}' = (x', y')$. With these requirements, the Green's function is uniquely determined to be the cylindrical wave given by

$$g(\mathbf{r}, \mathbf{r}') = \frac{1}{4i} H_0^{(2)}(k_0 n_{\text{ref}} |\mathbf{r} - \mathbf{r}'|), \tag{24}$$

where $H_0^{(2)}(z)$ is the Hankel function of order 0 and second kind. From Eqs. 22 and 23, the following integral equation is straightforwardly obtained:

$$E(\mathbf{r}) = E_0(\mathbf{r}) + \int g(\mathbf{r}, \mathbf{r}') k_0^2(\varepsilon(\mathbf{r}) - \varepsilon_{\text{ref}})E(\mathbf{r}')d^2r', \tag{25}$$

where the field has the desired form of the sum of the given incident field and a scattered part propagating away from the scatterer due to the properties of the chosen Green's function.

The simplest approach for solving Eq. 25 is also in this case to discretize the structure in, e.g., N square shaped area elements all having the same area ΔA and center at $\mathbf{r_1}, \mathbf{r_2}, \ldots \mathbf{r}_N$, respectively. As an approximation, it can be assumed that the field values are constant within each element and denoted $E_1, E_2, \ldots, E_N$. Thereby, the following linear system of equations is obtained:

$$E_i = E_{0,i} + \sum_j g_{ij} k_0^2 (\varepsilon_j - \varepsilon_{ref})E_j \Delta A, \tag{26}$$

where it is, e.g., possible to use $g_{ij} = g(\mathbf{r}_i, \mathbf{r}_j)$ if $i \neq j$ and $\varepsilon_j = \varepsilon(\mathbf{r}_j)$. It may be slightly more accurate to use

$g_{ij} = \frac{1}{\Delta A} \int\limits_{\text{element } j} g(\mathbf{r}_i, \mathbf{r}')d^2r'$. However, this does not make a significant difference unless $i = j$. In that case, however, due to the singularity of the Green's function it is not possible to use $g_{ij} = g(\mathbf{r}_i, \mathbf{r}_j)$. Instead it can be acceptable to approximate the square element with a circular element having the same area and center, in which case the integral can be calculated analytically. The radius of the circle then is $a = \sqrt{\Delta A/\pi}$. This leads to the approximation

$$g_{ii} \approx \frac{2\pi}{\Delta A} \int\limits_{r=0}^{a} \frac{1}{4i} H_0^2(k_0 r) r dr$$
$$= \frac{1}{2i(k_0 a)^2}\left[k_0 a H_1^{(2)}(k_0 a) - i\frac{2}{\pi}\right]. \tag{27}$$

The linear system of Eq. 26 can be formulated in matrix form and solved on a computer. As a small example that may also serve as an exercise for the reader, a few results are presented for scattering from the rectangular nanowire (Fig. 5) and a circular nanowire (Fig. 6).

Dyadic Green's Function Integral Equation Method for Three-Dimensional Problems

For a general three-dimensional scattering problem where light is scattered by one or more objects such as, e.g., a nanoparticle, the total electric field vector $\mathbf{E}$ must satisfy the equation

$$\left(-\nabla\nabla \cdot + \nabla^2 + k_0^2\varepsilon(\mathbf{r})\right)\mathbf{E}(\mathbf{r}) = \mathbf{0}, \tag{28}$$

where $\varepsilon(\mathbf{r})$ is the dielectric constant of the total structure consisting of a reference structure and one or more modifications representing the scattering objects. The incident field $\mathbf{E}_0$, which is the field in the case when no modification has been made to the reference structure, is a solution to the equation

$$\left(-\nabla\nabla \cdot + \nabla^2 + k_0^2\varepsilon_{\text{ref}}\right)\mathbf{E}_0(\mathbf{r}) = \mathbf{0}. \tag{29}$$

A dyadic Green's tensor $\mathbf{G}$ for the reference medium is defined as a solution to the equation

$$\left(-\nabla\nabla \cdot + \nabla^2 + k_0^2\varepsilon_{\text{ref}}\right)\mathbf{G}(\mathbf{r}, \mathbf{r}') = -\mathbf{I}\delta(\mathbf{r} - \mathbf{r}'), \tag{30}$$
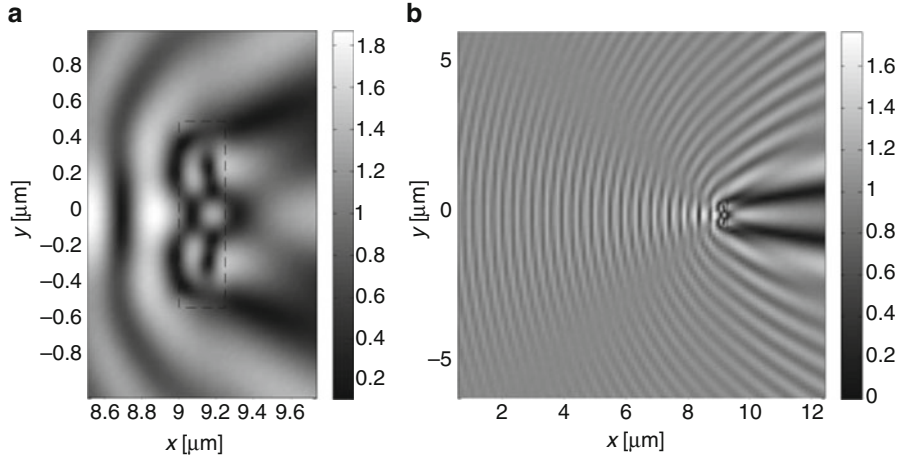
Green's Function Integral Equation Method, Fig. 5 Magnitude of electric field normalized with the magnitude of the incident plane wave for a plane wave being scattered by a rectangular nanowire. The light wavelength is $\lambda = 633$ nm and the nanowire widths are 0.25 and 1.0 μm, respectively. The nanowire has dielectric constant $\varepsilon = 4$ and is located in free space with dielectric constant $\varepsilon_{\text{ref}} = 1$. The calculation was carried out with 10×40 area elements inside the scatterer
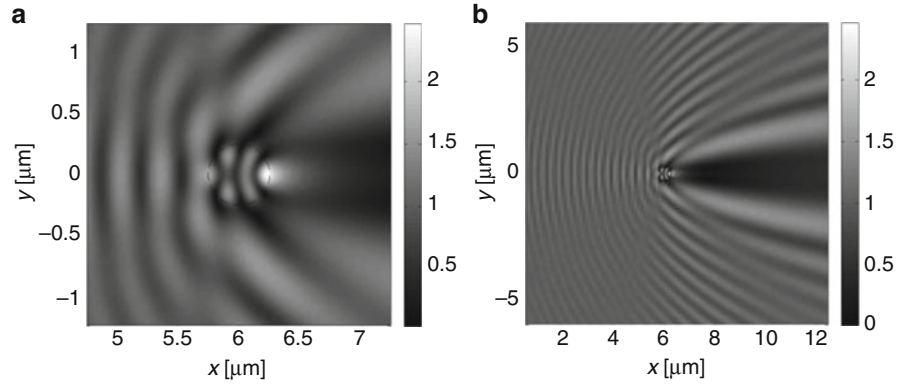
Green's Function Integral Equation Method, Fig. 6 The same as Fig. 5 except that light is incident on a circular nanowire with diameter 0.5 μm

where $\mathbf{I}$ is the unit dyadic tensor ($\mathbf{I} \cdot \mathbf{A} = \mathbf{A}$ for any vector function $\mathbf{A}$). In the case of a homogeneous reference medium (ε_{ref} being independent of the position) the dyadic Green's tensor can be expressed in terms of the scalar Green's function

$$g(\mathbf{r}, \mathbf{r}') = \frac{e^{-ik_0 n_{\text{ref}}|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|} \qquad (31)$$

satisfying

$$\left(\nabla^2 + k_0^2 \varepsilon_{\text{ref}}\right) g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'). \qquad (32)$$

The dyadic Green's function is for this special case given by

$$\mathbf{G}(\mathbf{r}, \mathbf{r}') = \left(\frac{1}{k_0^2 \varepsilon_{\text{ref}}} \nabla\nabla + \mathbf{I}\right) g(\mathbf{r}, \mathbf{r}') \qquad (33)$$

which satisfies Eq. 30 and the radiating boundary condition that it corresponds to propagation away from the position $\mathbf{r}'$.

The resulting integral equation for the scattering problem is then given by

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + \int \mathbf{G}(\mathbf{r}, \mathbf{r}') k_0^2 (\varepsilon(\mathbf{r}') - \varepsilon_{\text{ref}})$$
$$\cdot \mathbf{E}(\mathbf{r}') d^3 r'. \qquad (34)$$

An approach for solving Eq. 34 which is similar to the discrete-dipole approximation method [4] is again

to discretize Eq. 34 into a number of volume elements where the electric field vector and the dielectric constant are assumed constant. This leads to the discretized version of Eq. 34 given by

$$\mathbf{E}_i = \mathbf{E}_{0,i} + \sum_j \mathbf{G}_{i,j} \cdot \left(\varepsilon_j - \varepsilon_{\mathrm{ref},j}\right)\mathbf{E}_j, \qquad (35)$$

where

$$\mathbf{G}_{i,j} = \int_{V_j} \mathbf{G}(\mathbf{r}, \mathbf{r}') k_0^2 d^3 r'. \qquad (36)$$

The integral in Eq. 36 is rather difficult in the case when $i = j$ due to the singularity of the dyadic Green's tensor. An useful approach for calculating the integral (36) in that case is to exploit the expressions (32) and (33) and transform the volume integral into a surface integral with the surfaces placed at a distance from the singularity, i.e.,

$$\mathbf{G}_{i,i} = -\frac{\mathbf{I}}{\varepsilon_{ref}} + \frac{1}{\varepsilon_{ref}} \oint_{\text{surface of } V_i} (\hat{n}'\nabla' - \mathbf{I}\hat{n}'\cdot\nabla')g(\mathbf{r}, \mathbf{r}')d^2 r', \quad (37)$$

where $\hat{n}'$ is the surface outward unit normal vector. In the case when $i \neq j$, it might be acceptable to use

$$\mathbf{G}_{i,j} \approx \mathbf{G}(\mathbf{r}_i, \mathbf{r}_j) k_0^2 V_j. \qquad (38)$$

In the event that the difference in the dielectric constant in Eq. 34 is large, such as is, e.g., the case with metal particles in a homogeneous dielectric medium, the numerical solution of Eq. 34 becomes challenging. If the structure is divided into cubic volume elements, the representation of a curved surface becomes poor since it will be represented with a stair-cased approximation. In this case, convergence becomes slow at best but may not be achievable with any reasonable manageable number of discretization elements [7]. However, convergence can be achieved for this case as well if discretization elements are chosen that accurately represents the curved surfaces [7].

Further Reading

For a deeper study of the Green's function integral equation method for electromagnetic scattering problems, the reader is referred to the Refs. [5–7]. For a deeper study of the theory of Green's functions, the reader is referred to Refs. [8–10]. In the case of

scattering from a two-dimensional periodic array of scatterers being, e.g., placed on a planar surface the appropriate Green's function is presented in, e.g., Ref. [11]. Another Green's function integral equation method where integrals are over the surface of scatterers instead of over the interior of the scatterers is available for two-dimensional scattering problems in, e.g., Refs. [6, 7] and more recently for three-dimensional scattering problems in Ref. [12]. The Green's function for a reference medium involving two semi-infinite homogeneous media separated by a planar interface is given in, e.g., [5, 7].

Cross-References

▶ Ab initio DFT Simulations of Nanostructures
▶ Finite Element Methods for Computational Nano-Optics
▶ Finite-Difference Frequency-Domain Technique
▶ Finite-Difference Time-Domain Technique

References

1. Reitz, J.R., Milford, F.J., Christy, R.W.: Foundations of Electromagnetic Theory, 4th edn. Addison-Wesley, New York (1992)
2. Klein, M.V., Furtak, T.E.: Optics, 2nd edn. Wiley, New York (1986)
3. Martin, O.J.F., Dereux, A., Girard, C.: Iterative scheme for computing exactly the total field propagating in dielectric structures of arbitrary shape. J. Opt. Soc. Am. A **11**, 1073–1080 (1994)
4. Draine, B.T.: The discrete-dipole approximation and its application to interstellar graphite grains. Astrophys. J. **333**, 848–872 (1988)
5. Novotny, L., Hecht, B.: Principles of Nano-optics. Cambridge University Press, Cambridge (2007)
6. Jin, J.: The Finite Element Method in Electromagnetics, 2nd edn. Wiley, New York (2002)
7. Søndergaard, T.: Modeling of plasmonic nanostructures: Green's function integral equation methods. Phys. Stat. Sol. (b) **244**, 3448–3462 (2007)
8. Tai, C.-T.: Dyadic Green's Functions in Electromagnetic Theory. Intext Educational Publishers, London (1971)
9. Morse, P.M., Feshbach, H.: Methods of Theoretical Physics. Mc Graw-Hill, New York (1953)
10. Economou, E.N.: Green's Functions in Quantum Physics. Springer, Berlin (1979)
11. Kobidze, G., Shanker, B., Nyquist, D.P.: Efficient integral-equation-based method for accurate analysis of scattering from periodically arranged nanostructures. Phys. Rev. E **72**, 056702 (2005)
12. Kern, A.M., Martin, O.J.F.: Surface integral formulation for 3D simulations of plasmonic and high permittivity nanostructures. J. Opt. Soc. Am. **26**, 732–740 (2009)

Growth of Carbon Nanotubes

Growth of CNTs

Growth of Oxide Nanowires

Growth of Silica Nanowires

Praveen Kumar Sekhar[1] and Shekhar Bhansali[2]
[1]Electrical Engineering, School of Engineering and Computer Science, Washington State University Vancouver, Vancouver, WA, USA
[2]Department of Electrical and Computer Engineering, Florida International University, Miami, FL, USA

Synonyms

Growth of oxide nanowires; Growth of silicon dioxide nanowires; Manufacturing silica nanowires; Nucleation of silica nanowires

Definition

The evolution of nano (one-billionth of a meter)-sized material made of silicon and oxygen in a wire-like morphology is referred to as growth of silica nanowires. The evolution may be from a substrate or freestanding.

Introduction

Nanotechnology continues to induce a paradigm shift in electronic, chemical, biomedical, sensor, and related industries [1]. The use of nanoscience enables (a) superior performance of the materials and devices, (b) increased sensitivity levels in sensors, (c) faster response in devices, and (d) high-density synthesis of materials and technology platforms leading to cost-effective manufacturing. These properties stem from the fact that nanostructures have enhanced surface area of volume ratios, allowing surface phenomenon to dominate inertial phenomenon. Consequently a small variation in their dimensions alters the surface area/volume ratio and leads to significant changes in physical, optical, and electrical properties, even resulting in new class of materials and devices. Additionally, the large surface-to-volume ratios and their low dimensions make them ideal candidates to probe events at submolecular and atomic scale. One of the prominent areas of research in nanotechnology is the synthesis of nanomaterials [2].

Of the numerous one-dimensional (1D) nanostructures, silica (SiO_2) nanowires are a subject of intense research [3, 4] for numerous applications such as localization of light, low-dimensional waveguides for functional microphotonics, scanning near field optical microscopy (SNFOM), optical interconnects in integrated optical microchip, sacrificial templates, biosensors, and optical transmission antennae. Silica nanowires are made up of silicon and oxygen, the two most abundant elements in the earth's crust. Their combination as bulk silica has been used in a wide variety of applications ranging from colloid chemistries to MEMS and CMOS applications [5, 6].

The surface of silica can react with various compounds such as amine, carboxyl, or thiol groups to make it functional for biosensing applications [7]. Although silica nanowires have been observed to grow at the ends of CVD reactors since the 1960s [4], they were not actively exploited till 2004. The properties of silica nanowires are unique compared to bulk or colloidal silica [8]. As nanowires would be a critical building block of any nanosystem, the ability to synthesize them at a desired location in a controlled manner is of significant importance.

This chapter reviews silica nanowire growth and the growth mechanisms. First, a general overview of the "top-down" and "bottom-up" synthesis of nanowires is presented. After brief discussion of the various growth mechanisms in "top-down" and "bottom-up" synthesis, the vapor liquid solid (VLS) technique, a preferred method to grow silica nanowires, is described in detail.

Nanowire Growth Techniques

Top-Down   Bottom-Up

Templated Synthesis   Self-Assembly
                          Vapor Phase
                          Carbothermal
                          Solution Grown
                          Oxide Assisted Growth
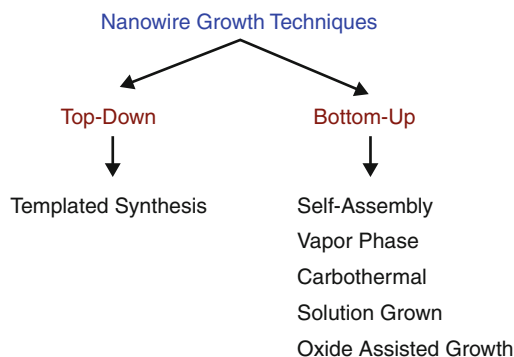
Growth of Silica Nanowires, Fig. 1 Classification of nanowire growth methodologies

Next, two new approaches in selective growth of silica nanowires using a combination of VLS and oxide-assisted growth (OAG) mechanism are presented along with a discussion on the mass manufacturing aspects of the nanoscale wires.

Nanowire Nucleation Techniques

Understanding the growth mechanism of nanowires is critical for enabling high-density and controlled synthesis of silica nanowires. The structure and morphology of the silica nanowires is dictated by the growth mechanism used. Figure 1 schematically illustrates the different techniques of growing nanowires in general. Top-down nanowires are routinely grown using templates. The preferred templates for the growth are anodized aluminum oxide (AAO), track etched polymers, and in some instances porous silicon. Bottom-up approach is an additive process, which builds on material from an atomic scale (molecule by molecule or atom by atom) to evolve one-dimensional nanostructures. Bottom-up approaches include vapor phase growth, solution-grown mechanism, oxide-assisted growth, carbothermal reduction technique, and self-assembly process. All of these growth techniques require nucleation of the wire.

Nucleation Techniques: Top Down and Bottom Up
There exist two generic methods for the growth of nanostructures, namely, the "top-down (TD)" and the "bottom-up (BU)" approach. TD approach involves machining and etching techniques to remove material from a bulk form to a nanoscale size. Templated method falls in the category of the TD approach in the formation

of nanowires. On the other hand, BU approach is an additive process, which builds on material from an atomic scale (molecule by molecule or atom by atom) to evolve one-dimensional nanostructures.

Top Down: Templated Method
The "template" is a fundamental structure within which a network forms. Deposition of the material in the template and the removal of the template leaves a nanowire made from deposited material. The external morphology of the resulting nanowire can, when deposited carefully, replicate the structural details of the template. Various predefined structures such as polymer membranes, anodic alumina, and porous silicon of fixed geometry have been used as templates. Once the template is fabricated, nanowires of desired geometry (diameter dictated by the pore size and length determined by the pore length) are derived. Electrochemical methods, electroless techniques, chemical vapor deposition (CVD), sol-gel and chemical polymerization–based deposition mechanism are used to fill the pores. The porous part or the scaffold is selectively removed using appropriate etching chemistry leaving behind the nanowires. The template method is convenient and simple. Nanowires of different compounds can be easily produced harnessing the properties pertaining to specific applications. Christian [9] reported a novel iron-based surfactant to form rod-like micelle structures, which may be, used to template the formation of silica nanowires with diameters less than 5 nm. The main drawbacks of this method are the inability to produce single crystalline wires and limited geometries (i.e., fixed diameter and length).

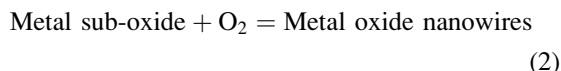### Bottom Up: Self-assembly Process
Self-assembly process is a viable bottom-up route for manufacturing nanostructures. It is defined as the creation of material from its constituent components in a spontaneous natural manner, i.e., by an interaction between the components or by a specific rearrangement of them that proceeds naturally without any special external impetus. Kievsky and Sokolov [10] reported a novel modification of acidic synthesis of nanoporous silica fibers that assembled fibers of uniform in size and shape. Each fiber was shaped like a hexagonal cylinder of approximately 2.0 μm in diameter and approximately 4.8 μm in length. The yield of the assembled fibers was virtually 100%.

BU: Solution-Grown Method

In this method, molecular capping agents (which control the kinetics of crystal growth by allowing selective nucleation on certain facets) along with suitable seed materials are dispersed in a solvent (a suitable chemical compound comprising material of interest) medium. The solvent is then heated and pressurized until it becomes a supercritical fluid. When the concentration of the solvent exceeds solubility limit, the seed ejects the elemental compound in the solvent as nanowires. This forms the basis of solution-grown nanowires. Kita et al. [11] successfully synthesized silicon oxide nanowires using a low-temperature hydrothermal treatment (200°C) without the use of metallic catalysts. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) revealed nanowires having uniform distribution and straight morphology with diameters ranging from 50 to 100 nm and lengths up to ca. 60 μm. The nanowires were produced by hydrothermal treatment based on the original concept of slow structure formation by the dissolution of bulk glass source materials (borosilicate glasses), which can be used to supply a continuous source solution.

BU: Carbothermal Reduction Technique

In this method, elevated temperatures are used which causes activated carbon to behave as a catalyst thus reducing metal oxides/nitrides/carbides to their corresponding nanowires with release of carbon monoxide. Equations 1 and 2 represent the carbothermal reduction technique.

$$\text{Metal oxide} + \text{C} = \text{Metal sub-oxide} + \text{CO} \quad (1)$$

$$\text{Metal sub-oxide} + \text{O}_2 = \text{Metal oxide nanowires} \quad (2)$$

Lin and Lin [12] achieved the growth of SiO_2 nanowires via carbothermal reduction of CuO/C powders at a temperature of 1,000–1,100°C in flowing Ar with the flow rate of 100 sccm.

BU: Oxide-Assisted Growth (OAG) Mechanism

Metal catalysts are not required for the synthesis of nanowires by this bottom-up fabrication method. Stoichiometric or non-stoichiometric silicon oxide (Si_xO_y) in its vapor form induces the nanowire growth.

For example, the growth of SiO_2 nanowires [13] was promoted by SiO vapor derived from Si powder targets or SiO_2 thin films. Equations explaining the oxide-assisted growth mechanisms are as follows:

$$Si_xO \rightarrow Si_{x-1} + SiO(x > 1) \quad (3)$$

$$2SiO \rightarrow Si + SiO_2 \quad (4)$$
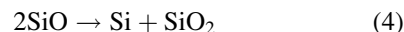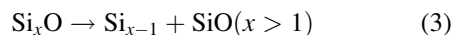
These decompositions result in the precipitation of Si nanoclusters, which act as the nuclei of the silicon nanowires covered by shells of silicon oxide.

BU: Vapor Phase Techniques

Typically, nanowires or one-dimensional nanostructures can be fabricated by promoting the crystallization of solid-state structures along one direction. This can be achieved either by the growth of intrinsically anisotropic crystal structure or the use of templates with one-dimensional morphologies directing the formation of one-dimensional nanostructures or the introduction of a liquid-solid interface thereby reducing the symmetry of a seed or by self-assembly of zero-dimensional nanostructures. The latter phenomenon forms the basis for vapor phase nucleation techniques. The growth of nanowires from the vapor phase techniques requires that the starting reactants be a gas phase species. Though there are many subcategories of vapor phase techniques, the preferred mechanism for nanowire growth referred to as the vapor liquid solid (VLS) technique will be discussed in the next section.

Vapor Liquid Solid (VLS) Mechanism   Wagner and Ellis first discovered the VLS method in 1964 [14]. In the VLS method, the nanowire is grown in a desired orientation when a metal catalyst reacts with the "nanowire material" in its gaseous phase at high temperatures. This method is based on the eutectic solidification of a metal and semiconductor. The catalyst also known as a second-phase material is used to confine the crystal growth onto a specific orientation. The catalyst alloys itself with the growth material. This acts as a trap to the growth species, which results in the precipitation at the growth surface in a single direction. Figure 2 depicts a typical VLS growth.

The VLS growth is facilitated by (a) the mass transport of whisker growth species in the vapor phase to the vapor liquid phase, (b) the chemical reaction on the vapor-liquid interface, (c) the dissolution and diffusion
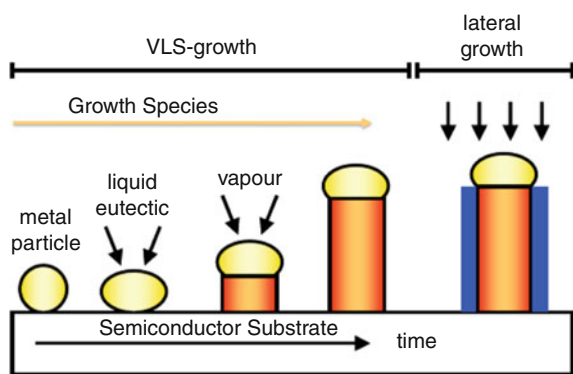
Growth of Silica Nanowires, Fig. 2 Schematic of the VLS growth mechanism

through the liquid alloy phase, and (d) the precipitation on the solid-liquid interface.

Common conditions for a VLS mediated growth are as follows:

(a) The catalyst must form a liquid solution with the crystalline material to be grown at the deposition temperature.

(b) The distribution coefficient of the catalyst must be less than unity at the deposition temperature.

(c) The equilibrium vapor pressure of the catalyst over the liquid droplet must be very small.

(d) The catalyst or impurity must be inert chemically.

(e) The interfacial energy, i.e., the wetting characteristics, plays a very important role. For a given volume of liquid droplet, a small wetting angle results in a large growth area, leading to a large diameter of nanowires.

(f) For a compound nanowire growth, one of the constituents can serve as the catalyst.

(g) For controlled unidirectional growth, the solid-liquid interface must be well defined crystallographically.

In VLS growth, the growth species in its vapor form passes through the substrate in the presence of a catalyst (metal particle), which then diffuses into a liquid droplet. The liquid's surface has a large accommodation coefficient, so it is a preferred site for deposition. Saturated growth species in the droplet diffuse and precipitate at the interface between the substrate and liquid thus growing into a nanowire. The VLS growth method is not restricted by the type of substrate materials and catalysts, thus enabling the nanowires to be either single crystalline,

polycrystalline, or amorphous. In general, the catalyst plays a key role in the VLS process.

It forms a liquid droplet as a preferred site to adsorb the vapor pieces of the source materials. The desired one-dimensional (1D) nanostructure precipitates, nucleates, and grows from the saturated liquid droplet. The selected catalysts in the VLS process should meet two demands. First, the catalyst should form the liquid eutectic alloy with a 1D nanostructure material under the growth temperature. The liquid alloy phase reduces the activation energy at the liquid-solid interface and enhances the growth rate appreciably. Second, the catalyst should be more stable than the components of the desired 1D nanostructure. This sustains the continuous growth as long as the vapor species is available.

VLS Kinetics Typically a catalytic agent is coated onto a growth substrate and is placed in a vacuum furnace of high temperature. Vapor atoms of the nanowire material are passed through it. The lowest temperature and pressure are chosen such that they cause the catalyst to form alloyed eutectic phase seeds with the lowest concentration of the growth species (GS). This causes the tip of the seed to be semi-molten. The semi-molten structure of the tip has a soft lattice and loose interatomic configurations to allow adsorption of the GS vapor.

If there is a lateral temperature gradient, or a sudden change of temperature, it leads to random breakup of the catalyst droplets. Also if there is an oversupply of the GS, it leads to oscillation of surface energies, which enables multiple nucleations. Lastly, the gas of the GS can cause hydrodynamic stress on the droplet thus causing droplet breakup. The temperature and pressure are very crucial for the VLS mechanism to succeed. The growth of the nanowires will proceed only when the concentration of the GS is above the equilibrium solubility. Although the evaporation of the catalyst does not change the composition of the saturated liquid composition, it reduces the total volume of the liquid droplet. Unless more catalyst is supplied, the volume of the liquid droplet reduces and growth will stop when all the catalyst is evaporated.

The orientation of the nanowires depends on the substrate texture and the diameter of wire. Moreover, it also depends on the nanowire composition, as materials prefer to orient in a direction minimizing their enthalpy of formation.

Silica Nanowires: A Review of Growth Investigations

Over years, silica nanowires have been synthesized using growth methodologies discussed in the previous sections. Among the various techniques, vapor phase approaches have been routinely used in applications that require mass manufacturing and high-density growth. A comprehensive review [15] on silica nanowire growth (for the past 12 years) is presented in Fig. 3.

Selective Growth of Silica Nanowires: New Approaches

Ion Implantation–Based Seeding Approach

Of the several methods used to produce nanowires, the VLS mechanism proposed by Wagner and Ellis for silicon whisker growth has proved to be very reliable for growing silica nanowires. A simulation study [16] indicated the feasibility of forming catalyst nanoclusters in silicon using ion implantation. Ion implantation is a low-temperature technique for the introduction of impurities (dopants) into semiconductors, and is preferred over diffusion. It is a process of introducing energetic and charged particles into a substrate. Implantation of ions into a target may alter the physical or chemical properties of the host via damage induced by the energy loss of the incident ions. In this surface modification technique, ions are extracted and accelerated toward the target. Before penetrating the host, the charged ions are separated according their mass-to-charge ratio. Then the desired species is directed at the target. The foreign atoms enter the crystal lattice, collide with the host atoms, lose energy, and finally come to rest at some depth within the substrate. The average penetration depth is determined by the species, substrate material, and the initial energy of the ions. This depth has some distribution, as the collisions with the target atoms are random. Figure 4 shows the ion implantation setup.

In this section, ion implantation as a new method to form catalyzing seed nanoclusters at or below the surface of Si and subsequent growth of silica nanowires is presented [17]. The ion implantation–based seeding approach possesses several advantages over conventional methods of catalyst placement such as (a) the size and distribution of the catalyst can be accurately controlled by varying the ion dose, energy, and dose rate, (b) the placement of catalyst being an integral part of the substrate lattice lends for an impurity-free nanowire growth, and (c) majority of the element across the periodic table can be implanted giving rise to hybrid nanostructures (metal/dielectric). Such an approach was demonstrated by implanting Pd ions into Si through a mask.

Methodology In this investigation, prime grade Si (100) wafers of $2''$ diameter were used as substrates for Pd ion implantation performed with a metal vapor vacuum arc (MEVVA) ion implanter. The MEVVA ion source is a high current device, which can readily produce ion beam currents of several hundred milliamperes on target. Extraction potential up to 50 kV is used to implant metallic elements. Ion beams of most metallic elements either individually or in combination can be generated with a MEVVA source.

In this study, the lowest possible extraction voltage to produce a stable ion beam in the implanter (of 10 kV) was chosen to ensure maximum surface concentration of implanted Pd. Since the MEVVA ion source produces ions with multiple charge states, the energy of the ions emanating from the source depends on the charge state. For the Pd ion implantation, the energy was estimated to be 19 keV using a mean charge state of 1.9. Circular shadow mask (Al) was used to define identical and well-separated zones of ion-implanted regions. Dose was varied from 5×10^{12} to 3×10^{16} Pd ions/cm^2. The sample was introduced into the furnace around 1,100°C and heated for the desired time (7–60 mins) in Ar followed by cooling to room temperature. Further, the growth temperature was varied.

Discussion After heating at 1,100°C for 60 min, the ion-implanted region showed a dense growth of rather long and interwoven nanowires as shown in Fig. 5. The inset shows the unimplanted area at the same magnification confirming the absence of nanowires due to unavailability of Pd seeds. Also, Fig. 5a demonstrates the selective nucleation of nanowires on the implanted region. The interface region is very narrow and sharp indicating the ability to directly locate nanowires on the implanted areas. The nanowires grow with diameters ranging from about 15 to 90 nm, as for example seen in Fig. 5b. The length of the nanowires was observed to vary up to about 50 μm corresponding to an aspect ratio of up to 3,000. The morphology of the nanowires depends on the growth time as evident from

| Substrate	Type of Catalyst and Preparation	Growth Conditions			Geometry		Mechanism	Reference et al.
		Temperature (C)	Time (min)	Ambient	Dia. (nm)	Height and Orientation		
Glass, Si,Al$_2$O$_3$	Spu-Au, 15–90nm	CVD 350–1000	30	Ar	72	Spring Morphology	VLS	Wang L., Nanotech. 2006
Si (111)	Ga, Ni Power Ni(NO$_3$)$_2$.H$_2$O	950	30	Ar/NH$_3$	30	20 nm–20 µm, Jelly fish, core-shell architecture	VLS	Cai, X.M., JAP 2005
Si	Ga Molten Droplets	700–950	10–60	NH$_3$	50	> 30 µm, Self assembled cone	VLS	Luo, S., Small 2006
Si	NiO Powder	1000–1100	50–100	NH$_3$	50–100	Ultra long mm	SLS	Park, B., Surf.Rev. Lett 2003
Fused Silica Capillary	2 Chloro-1,1,2 TriFluoroEthylMe thyl Ether	400	720	-	10–100	5–200 nm	VLS/VS	Woldegio rgis A., Electroph oresis 2004
Silica	Si, C and Co	1500	60	Ar	20–120	10–250 µm, Flower Like Nanofibers	Solid Phase/ Gas solid Reaction	Zhu, Y.Q., J.Mat.Chem 1998
Silica	Si/Si Powder	1400,Pyrolysis	12	N$_2$	80	[110], Core-Shell	VS	Z.L. Wang, Adv Mat 2000
AAO	SiCl$_4$, Wash CCl$_4$	Hydrolysis			P.Dia	P.Length	Surface Chem, Solgel	Kovtyukh ova, Adv Mat 2003
Fused Silica/Si (111)	-	1800, HFCVD	20	SiCl$_4$+H$_2$	80	3.5 µm, Catalyst Free, Template Free	-	Jih-Jen Wu, Av Mat 2002
Si	Molten Ga	520–900	5–300	Ar/O$_2$	15	10 µm	Extended VLS	L.Dai. Phy.Lett A 2005
Si (111)	Sn	980	2	Ar	100–200	50–100	VLS	S.H.Sun, Solid. State. Commun. 2003
Si	MgO+Graphite Powder and Au	1100	45	Ar	60–90	Long	VLS	S.Kar, Solid. State. Commun. 2005
Si	Ga/In$_2$O$_3$	400	120	Ar/NH$_3$	60	15	VLS	J.C. Wang., Solid. State. Commun. 2003
Si	Si/Al$_2$O$_3$	1150	8–60	Ar	20–100	mm	VLS	X.S.Fang, Appl.Phys A 2005
SiO$_2$	Au	1100	60	Ar	25–100	mm	SVLS/OAG	Houlier Jnl of Phase Equ and Diffusion 2007
Si (100)	Co NPs	1100	120	Ar/CH$_4$	5–60	-	Coils	F.Wang, JPC, 2007
Crucible	SiO (P) + Fe	850	240	Ar	15	mm	VLS	D.P.Yu, APL, 1998
Si	GaN	1150	300	Ar	15–300	10–40 µm	Extended VLS	Z.L.Wang, JACS 2001
Pellets	SiO$_2$, FeNO$_3$,9H$_2$O, Active C Ball Mill	1350	180	Ar	60–110	>100 µm	Carbothermal Redn +VLS	X.C. Wu, Chem.Phy. Lett 2001

Growth of Silica Nanowires, Fig. 3 (continued)

Substrate	Type of Catalyst and Preparation	Growth Conditions			Geometry		Mechanism	Reference et al.
		Temperature (C)	Time (min)	Ambient	Dia. (nm)	Height and Orientation		
Si		800–1300	30	H_2/H_2O	50	Long NWs Aligned	Thermal Growth	J.Q.Hu, Chem.Phy. Lett 2003
AAO	TEOS + Ethanol	1 Day Aging, 200 C-1 Day	-	-	22–30	6 µm		Zhang, J.Mat Sci.Lett 1999
Si	Pd/Au, 1:1 Spu	1100	40	N_2	100		VLS	Elechiguerra, Applyphys A 2001
Si (111)	Ni–40 nm	950		Ar/H_2	20	10 mm	SLS	Yan, Chem Phys Lett 2000
TiN/Ni/SiO$_2$/ Si		1050	20	H_2/Ar		-	Solid state reaction	K-H Lee, Chem Phy Lett 2004
Si + PMMA Mask	Ni	1100	20	Ar/H_2	30	-	SLS	Ju Hung Hsu, Nanotech nology 2006
Si	NiO, WO$_3$	1100	60	Ar	20–100	-	SLS + Carbothermal	B.T.Park, Nanotech, 2006
Si	Sn + SiO	1320	420	Ar/H_2	230	mm	VLS+OAG	H.Wang Angew. Chem.Int 2005
Si	Ga	1050	60	Ar/H_2	10–20	0.5–1.5 µm	VLS	Yiqin Chen, Nanotech 2006
Si	CuO/C	1000	60	Ar	5–500	-	VLS + Carbothermal	Yu-Chiao Lin, IEEE Nano 2005
Si	Thermal Evap, Laser Ablation, SiO Powder	800–1300	30	Ar	60–120	10 µm	[111]	J-Q.Hu, Adv.Mat 2003
Si	Ni + In$_2$O$_3$	650	45	N_2/NH_3	10–20	5–20 µm	VLS	Renzlin Ma, Chem.Phy. Lett 2003
Si (111)	Ni Sputter	1000	60	Ar/SiH_4	9	10 µm	VLS	Niu, Physica E 2004
Crucible	SiO + SnO	1180	240	Ar	2500	100	VLS	J.Hu, Small 2006
Si	Zinc Carbonate, Hydroxide and Graphite	1100	60		100	Several µm	VLS	Q.Wei, Solid.State Communc 2006
Graphite	Si Powder	1200	30	Ar/O_2	300	500	VLS	K.S Wenger, J.Mat.Chem 2003
Colloidal Particles	TEOS, Ethanol, HCl, H$_2$O	80	30		1000		Solution Grown	S-S Choi, J.Mat.Sci Lett 2003

Growth of Silica Nanowires, Fig. 3 (continued)

Substrate	Type of Catalyst and Preparation	Growth Conditions			Geometry		Mechanism	Reference et al.
		Temperature (C)	Time (min)	Ambient	Dia. (nm)	Height and Orientation		
Ceramic Boat	SiO Powder + H_2O	1200	60	Ar	100	mm	VLS	Y.Hao, APL, 2005
Alumina	Si Powder, SiO_2, Pbo, C	1200	90	Ar/O_2	710	Various Morphology	VLS	J.Zhang, Jnl.Cryst. Growth, 2001
Si	Au	1000	180	N_2/H_2	30–150	10000	SLS	Paulose M., Jnl.Nanosci and Nanotech.
Si	Fe	1200			8–20	> 1 μm, [111]	Core-shell	Lieber, Science 1998

Growth of Silica Nanowires, Fig. 3 Literature review of silica nanowire growth
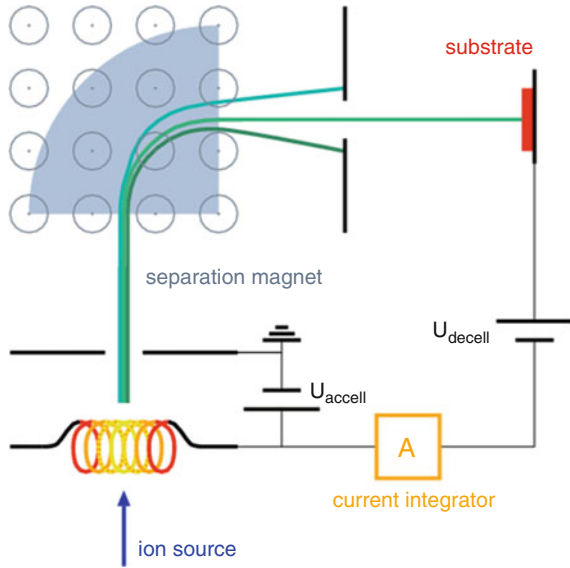
Growth of Silica Nanowires, Fig. 4 Ion implantation setup (Courtesy: Wikipedia)

Fig. 5c (after 7 min) captured at the early stage of the growth with nanowires emanating from their parent seeds. Figure 5d is a composite constructed from an overlay of two SEM images recorded in backscattered mode (pink color) and secondary mode (green color). The pink areas therefore show atomic number contrast. The pink oval seed is confirmed to be rich in Pd by spot energy dispersive spectroscopy (EDS) measurements on and away from it.

Figure 5d therefore captures the "birth" of a silica nanowire (green color) growing from a Pd-rich seed.

A much thinner nanowire can also be seen to emerge from the same seed. The composition of the nanowires was measured with EDS (at low electron energy of 5 kV to ensure that majority of X-rays come from nanowires and not from Si substrate) and compared with that of a control sample of quartz analyzed under identical EDS conditions (Fig. 6a). Thus the composition was found to be close to SiO_2 with a slight deficiency of oxygen.

The phase identification was done by X-ray diffraction (XRD) on the same sample (Fig. 6b), which did not show the peak observed in quartz sample indicating that the nanowires are amorphous. The temperature was also found to be a critical factor for the growth of nanowires. At a lower temperature of 1,010°C for the same growth time of 60 mins, only a very early stage of nucleation and growth was observed. At still lower temperature of 910°C (Fig. 7a), no growth of nanowires is observed. Pd ion dose was also found to be another factor critical for nucleation of Pd-rich seeds and nanowires.

As the dose was reduced from 3E16/cm^2 (Fig. 8a) to 1E16/cm^2 (Fig. 8b), coiled nanowires of reduced density were seen. The density of the wires was estimated by counting the wires in five 4 × 4 μm (high dose), 7 × 7 μm regions (low dose) of the SEM and averaging them. The observation show an estimated 2E8 wires/cm^2 in sample implanted with a dose of 3E16/cm^2 (Fig. 8a) and 1E8 wires/cm^2 in sample implanted with at a dose of 1E16/cm^2 (Fig. 8b). These observations confirm that implantation dose is a critical factor in determining the density of the

Growth of Silica Nanowires, Fig. 5 Selective growth of nanowires at a constant Pd ion dose of 3×10^{16} ions/cm^2 after heating at 1,100°C for 60 and 7 min. (**a**) Dense wire growth after 60 min; *inset*: Selective growth on implanted regions, inset: *I* Implanted region showing nanowire growth, *IF* Interface, *U* Unimplanted region devoid of nanowire growth, (**b**) variation in morphology of wires in (**a**), (**c**) early stage of nanowire growth (after 7 min), and (**d**) composite of backscattered and scanning electron image highlights the elemental contrast between seed and wire
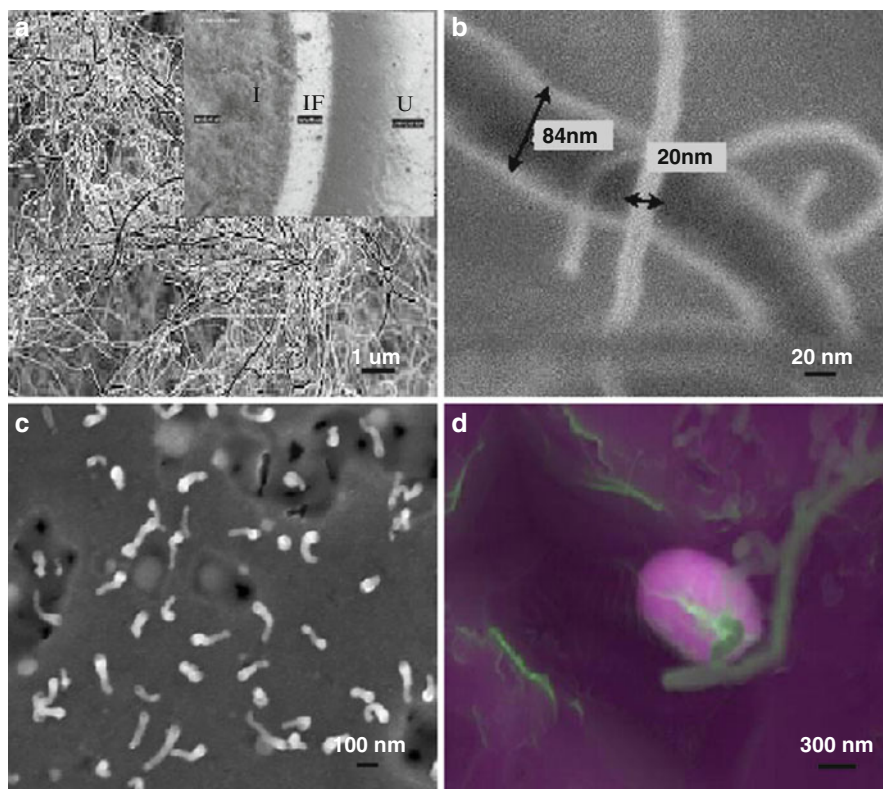
Growth of Silica Nanowires, Fig. 6 Elemental composition and texture analysis of the synthesized nanowires. (**a**) Comparison of EDS spectra on quartz with EDS on nanowire sample revealing SiO_x ($x < 2$) stoichiometry and (**b**) XRD showing their amorphous nature
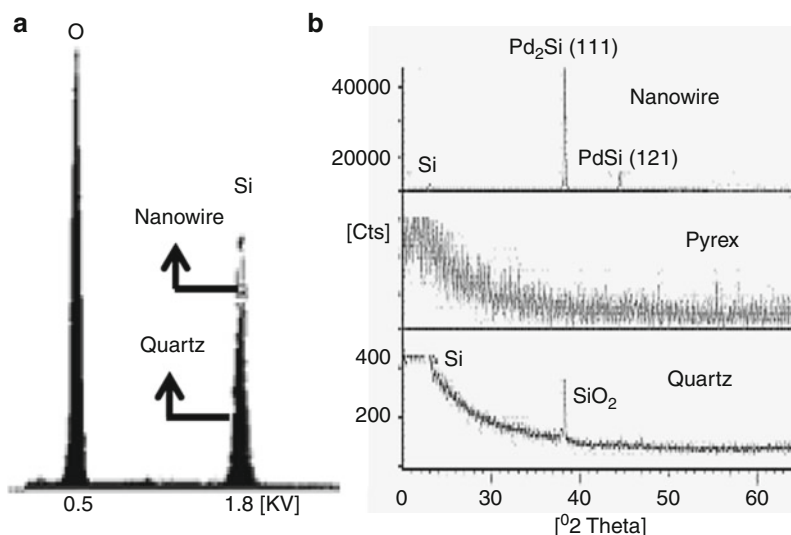
nanowires. However, at a dose of 5E12 ions/cm^2, no growth of nanowires (inset in Fig. 8b) or Pd-rich seeds was observed.

Moreover, the flow of Ar gas was found to be critical in producing growth of silica nanowires. In another experiment at 1,100°C, the Ar gas was shut off after the furnace temperature reached 1,100°C, and then the sample was introduced into the furnace and kept there for 60 min, after which the furnace was allowed to cool down to room temperature and
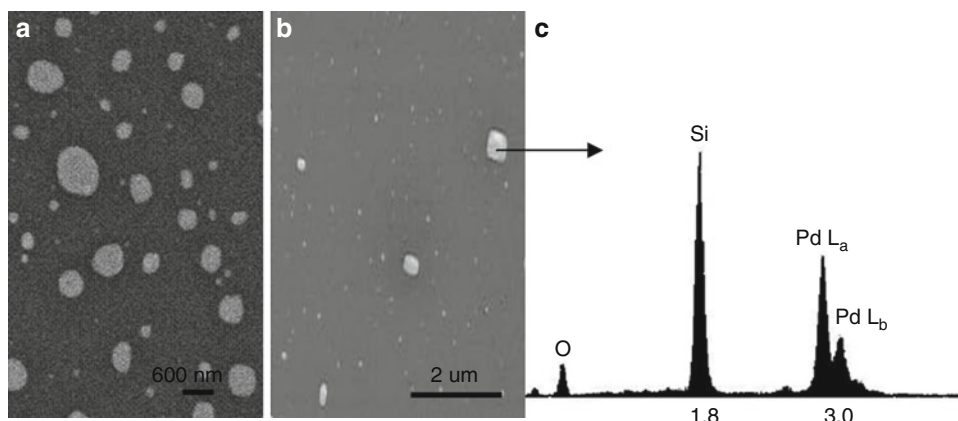
Growth of Silica Nanowires, Fig. 7 Conditions of no nanowire growth at a constant Pd ion dose of 3×10^{16}/cm^2. (**a**) 910°C, Ar flow – absence of nucleation, (**b**) 1,100°C, no Ar flow – formation of large whitish islands, and (**c**) spot EDS confirming the presence of palladium silicides on the large islands

Growth of Silica Nanowires, Fig. 8 Effect of implanted dose on the growth of nanowires heated at 1,100°C for 60 min. (**a**) 3×10^{16} ions/cm^2 – high-density thinner nanowires and (**b**) 1×10^{16} ions/cm^2 – thicker coiled nanowires of reduced density, *inset*: 5×10^{12} ions/cm^2 – absence of nanowires
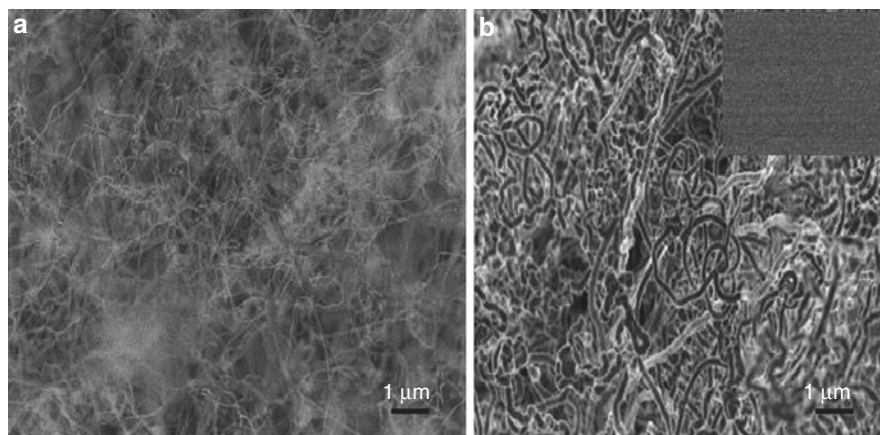
the sample was taken out for further study by SEM and XRD. The SEM results showed islands of bright appearance.

Larger islands with rectangular morphology (Fig. 7b) were confirmed to be consisting of Pd$_2$Si composition by performing a spot EDS (Fig. 7c) on them. The composition of smaller islands could not be determined accurately because their size was much smaller compared to the excitation volume for the EDS. They appear to be the Si-rich PdSi phase. Away from islands, no Pd was found. There were no nanowires on this sample. XRD obtained from this sample shows the presence of Pd$_2$Si and PdSi phases. The absence of silica nanowires proves the role of Ar as a carrier gas in transporting Si vapor to the reaction site resulting in the whisker growth by a VLS mechanism.

The physical model to explain the nanowire growth is (Fig. 9) as follows: the ion dose needs to be high enough to produce nanoclusters of Pd within the silicon matrix amorphised by ion damage. As the implanted sample is heated in Ar atmosphere, the interdiffusion at the Pd-Si interface produces two phases of PdSi and Pd$_2$Si, which eventually become liquid as the melting point of the silicides is approached at 972°C or 1,331°C, respectively. Since the maximum temperature used in the study is 1,100°C, only liquid drops of PdSi phase are formed. These liquid drops may now coalesce to form larger sized liquid drops and migrate toward the surface of the solid Si substrate. The amorphised surface layer of silicon would also crystallize concurrently. When these liquid drops are exposed to the Si vapor being swept across by the carrier gas (Ar), the VLS model of whisker growth comes to play and Si nanowires begin to grow.
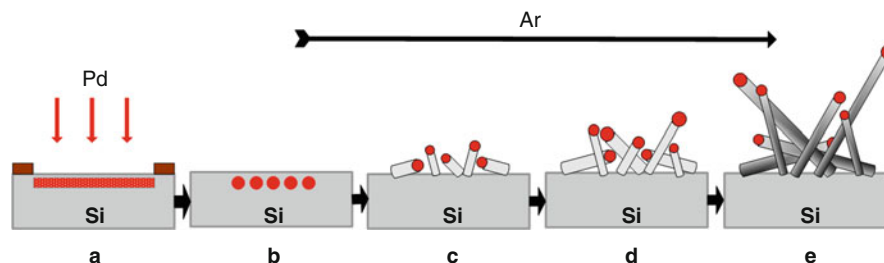
Growth of Silica Nanowires, Fig. 9 Schematic of the proposed physical model explaining ion implantation–based seeding approach. (**a**) Implanted Pd on Si (100) through a mask, (**b**) formation of PdSi cluster, and (**c–e**) VLS growth of silica nanowires

Some nanoclusters, especially those on which nanowire growth has not yet started, agglomerate by diffusion/Oswald ripening processes, become bigger in size and nucleate wires. The growth process proceeds continuously by diffusion and precipitation of silicon atoms across the liquid-nanowire interface. Since small amount of oxygen is inevitably present in the ambient as impurity, these thin silicon nanowires may oxidize rapidly from outside and eventually get fully oxidized.

The observed oxygen deficit from the stoichiometric SiO_2 may be consistent with some unoxidized silicon around the center of the nanowires. The occurrence of selective VLS growth of nanowires only on the Pd ion–implanted regions has been validated by a careful selection of ion dose, growth temperatures, and Ar gas flow. The VLS growth can stop when either of the following conditions are met – no more vapors of Si arrive or the temperature drops below the melting point of the material of the liquid drop, or the implanted ion dose is below that required for Pd cluster formation. However, the origin of the observed morphology of silica nanowires is unclear. The nanowires are produced with a variety of diameters and have much larger lengths forming a network of entangled structures. Since the physical model is based on VLS growth of Si nanowires followed by their oxidation, one can expect the dominant growth direction to be (110) on a (100) substrate surface so that the nanowires would point out at $\pm 45°$ with respect to surface plane. However, the SEM views of the early stage of growth do not show such large angles. They appear to be coming out almost along the surface at much smaller angles. This observation appears to be in agreement with recent reports of diameter-dependent growth directions of epitaxial Si nanowires grown on Si (100) substrates using Au as liquid catalyst facilitating VLS growth [18]. It was found that Si nanowires grown epitaxially on Si (100) via the VLS growth mechanism change their growth direction from <111> to <110> at a crossover diameter of approximately 20 nm.

A model has been proposed [19] for the explanation of this phenomenon. Based on this model, it can be hypothesized that most of the nanowires with diameters greater than 20 nm grow along <211> or <111> or a mixture of these growth directions and therefore lie close to the substrate plane (within about $\pm 18°$). As the diameter of the nanowires changes, so does the growth direction leading to the observed entanglement of the nanowires.

A similar change with decreasing diameter of the nanowires from <111> to <110> direction has also been observed for non-epitaxial Si nanowires. In summary, a new method for selective growth of silica nanowires on silicon wafers using ion implantation–based seeding approach is presented. Since ion implantation is compatible with the standard top-down CMOS process, this method holds a great potential for facilitating directed bottom-up growth of nanowires on silicon wafers. It may also provide an opportunity to combine the top-down and bottom-up technologies for creation of robust nanoscale devices. It may therefore open doorway to a large number of applications in wide areas of new technologies such as on-chip optoelectronics, biosensors, micro-antennae, and metallic nanotubes. The only limitation of this technique is scarce availability of ion implanters.

Metal Thin Film–Based Catalysis

A metal thin film catalysis–based silica nanowire growth circumvents the need for ion implantation as a tool for the placement of the catalyst. Also, the ability to produce stable nanowires at desired locations in a controlled

Growth of Silica Nanowires, Fig. 10 Platinum thin film deposited on a silicon substrate

manner on planar substrates is of critical importance for fabricating nanodevices such as nanoFETs, nanotransistors, and microphotonic nanosystems.

In this section, selective growth of silica nanowires in silicon using Pt thin film is presented [20]. This synthesis approach for localization offers several advantages such as the tunable size and distribution of the parent nanoclusters and a cost-effective solution for growing large-scale arrayable nanowires.

Methodology Prime grade 2″ n-type silicon wafers were used as substrates for subsequent Pt deposition.

Target Pt thickness of 2, 3, 5, 10, and 100 nm was sputtered onto Si substrate (Fig. 10).

The temperature of heating is chosen based on the Pt-Si phase diagram. High-purity Ar was chosen as carrier gas and was set to 30 sccm throughout the course of the experiment. Initially the furnace was flushed with Ar to minimize interference from gaseous impurities. Then the sample is introduced into the furnace and ramped up to 1,200°C. Once the furnace reached 1,200°C, the heating process was timed.

Discussions After heating at 1,200°C for an hour, regions deposited with Pt (20 nm) showed dense growth of interwoven nanowires. Figure 11a demonstrates selective growth of nanowires identifying the nucleation area and regions devoid of nanowire due to absence of Pt film. A higher magnification SEM micrograph taken in the growth region (Fig. 11b) illustrates nanowires of varying morphology. The nanowires grew at an average diameter of 200–500 nm (Fig. 11c) with lengths varying from 20 to 75 μm giving rise to high aspect ratio nanostructures for biological applications. Figure 11d is a schematic capture of birth of several nanowires from their parent seed. The composition of these nucleated nanostructures from low-energy EDS was found to be SiO_2 closely matching with the EDS spectra of a control sample of quartz analyzed under similar test conditions. Phase identification through XRD revealed

the amorphous (as compared to a quartz sample) nature of the wires along with the formation of PtSi, which acts as a seed catalyzing the nanowire growth. The thickness of platinum layer was found to be a critical factor affecting nanowire growth. As the thickness was increased to 3 nm (keeping other experimental conditions constant), the number density of the nanowires reduced. Further reduction in number density was observed when scaled up to 5 nm. At a thickness of 10 nm, silicide nanoclusters were formed indicating the absence of nanowire nucleation. When the deposited platinum thickness was around 100 nm, micron-sized clusters were seen.

Figure 12 illustrates the statistical nature of the growth process through a frequency histogram. All the nanowire samples with different Pt thicknesses were heated to 1,200°C for 1 h. For a Pt thickness of 2 nm, the ensemble diameter average was found to be 500 nm with a standard deviation (σ) of 6.7 over the 79% range. A uniform distribution with an average wire diameter of 319 nm ($\sigma = 9.1$) was observed for the 3 nm Pt sample. As the platinum thickness increased to 5 nm, the nucleated nanowires exhibited a wide range of (60–220 nm) diameter variation explained by the Oswald ripening process. The average diameter was found to be 146 nm ($\sigma = 7.4$) over the 62% range. Exploiting this selectivity of nanowire growth to platinum film thickness, a Si substrate with differential Pt coating (differential thickness achieved through repeated application of shadow mask) resulted in the nucleation of nanowires while simultaneously forming optically conducting silicide in a single furnace treatment. This would facilitate a low-loss hybrid nanointerconnect scheme in silicon for future integrated optical systems. The heating temperature was also found to be a critical factor influencing nanowire growth.

At a lower growth temperature of 950°C for 1 h (Fig. 13a), spherical islands appeared surrounded by nanoclusters. These islands were identified as Pt_3Si particles through XRD. Testing the role of ambient in nanowire growth, the flow of Ar was intentionally shut off while ramping up the furnace to 1,200°C. When the sample was analyzed in a SEM, rectangular islands of bright appearance (Fig. 13b) were observed. Larger islands were confirmed to be PdSi by composition through an EDS on them. Smaller islands could not be resolved elementally due to limitation on the excitation volume for the EDS. The absence of silica nanowires proves the role of Ar as a carrier gas in

Growth of Silica Nanowires, Fig. 11 Selective growth of nanowires after heating at 1,200°C for 60 min for a Pt thickness of 2 nm. (**a**) Localized growth of nanowires, (**b**) higher magnification image in area of nanowire growth indicating the varied morphology of nanowires, (**c**) size of nanowires – diameter ranging from 50 to 500 nm, and (**d**) early stage of nanowire growth captured after heating at 1,200°C for 15 min
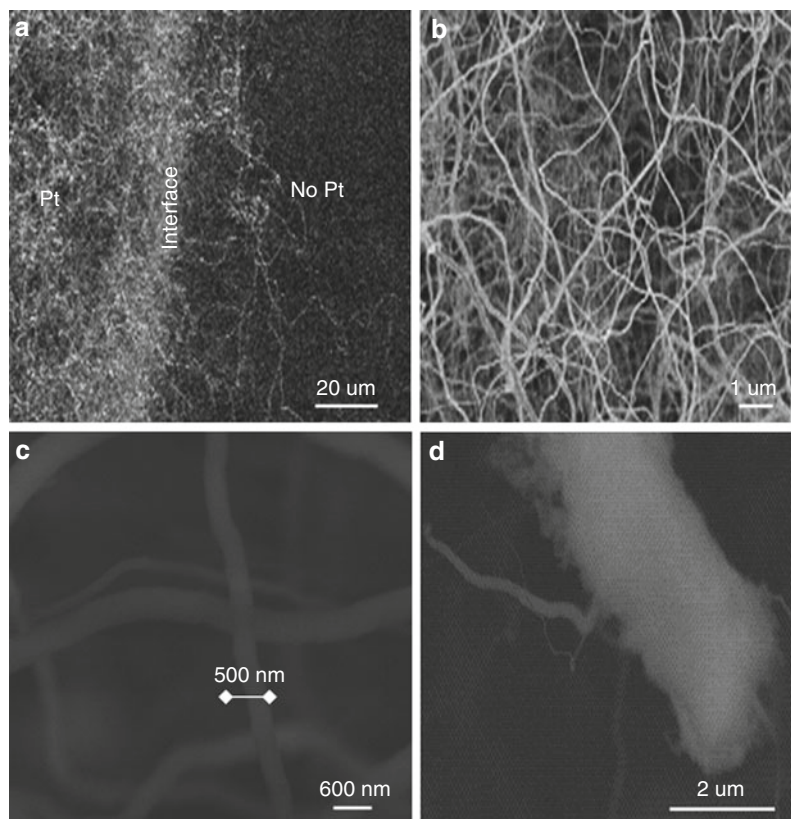
transporting Si vapor to the reaction site resulting in the whisker growth by a VLS mechanism.

In summary, selective growth of silica nanowires on silicon catalyzed by Pt thin film was demonstrated. Platinum silicide acts as the seed facilitating the bottom-up nucleation of the oxide nanowires. The nanoscale optical wires nucleate following a VLS model validated through a careful set of growth conditions. This approach represents a simple, flexible, and cost-effective solution for synthesizing large-scale silica nanowires for wide variety of future applications that include functional microphotonics, high-density optical nanointerconnects, high-resolution optical heads, and nanoantennae. In the next section, the manufacturing aspects of silica nanowires will be discussed.

Manufacturing Aspects of Silica Nanowire Growth

Overview

Exploring the manufacturing aspects of nanostructures can enable the transition from a laboratory-based research into a commercial product. Scalable manufacturing methods are required to harness the benefits of nanotechnology for a wide variety of applications. Understanding the manufacturing aspects of nanomaterials requires growth model and growth rate estimation. This section details on (a) a growth model derived from studying silica nanowire nucleation as a function of heating time, (b) nanowire growth rate estimation via weight differential of the Si substrate before and after growth, and (c) reusability of the Si substrate for nanowire growth. The growth mechanism of silica nanowires on Si via Pd metal thin film catalysis was investigated. A growth model was derived by sequentially varying the nanowire growth time at a constant temperature. The derived growth model for mass manufacturing nanowires was verified by a combination of experiments such as (a) annealing the metal thin film–coated silicon wafer in the absence of gas flow, (b) removing the nanowires in HF to inspect the substrate after growth, and (c) vary the Ar flow rate. In addition, the growth rate was estimated by weight differential technique while identifying

**Growth of Silica Nanowires,
Fig. 12** Percentage
histogram indicating the
statistical nature of the growth
process with varying
diameters for different
Pt thicknesses. (**a**) Pt – 2 nm:
ensemble average
diameter = 500 nm, σ = 6.7;
(**b**) Pt – 3 nm: ensemble
average diameter = 319 nm,
σ = 9.1; and (**c**) Pt – 5 nm:
ensemble average
diameter = 146 nm, σ = 7.4
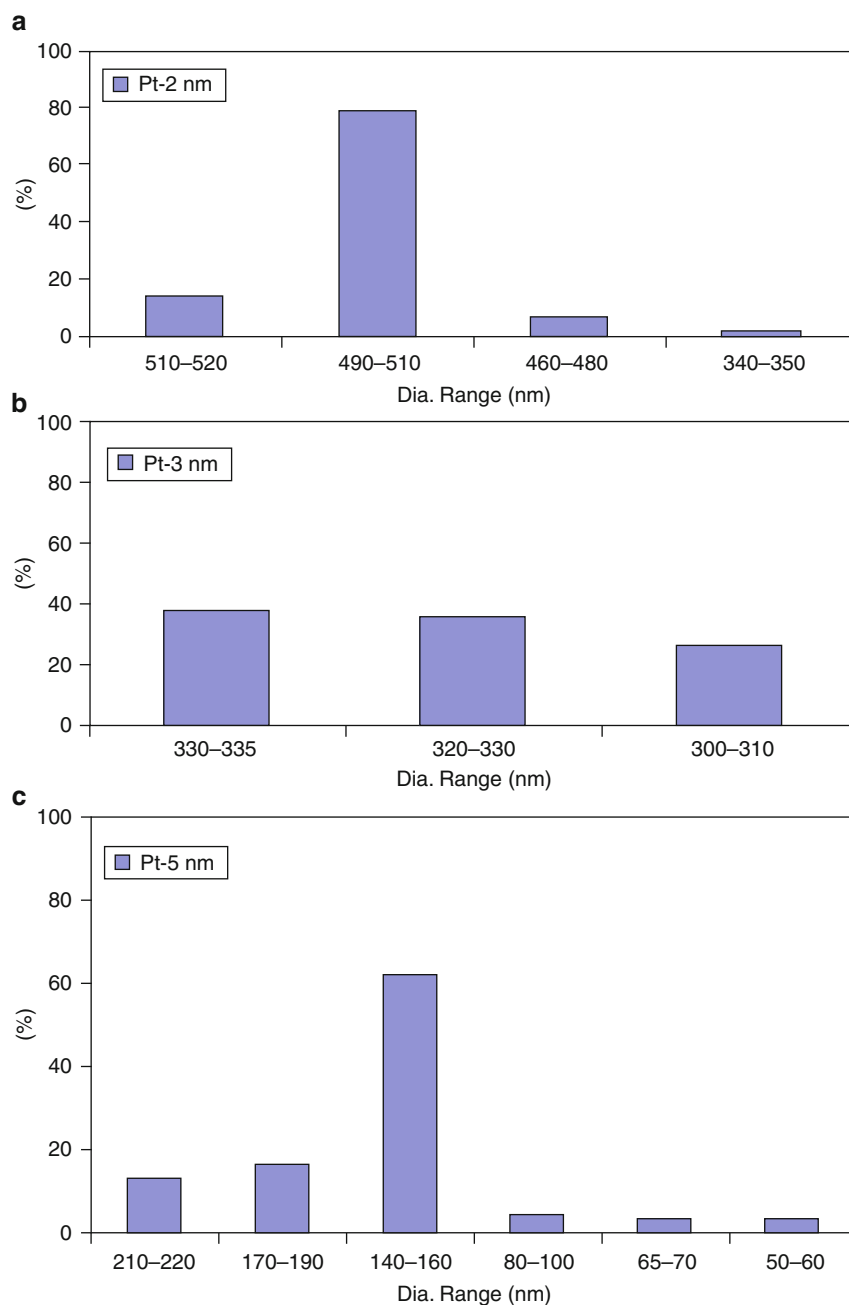
distinct nanowire growth regimes. Finally, the reusability of the silicon substrate for further nanowire growth was evaluated.

Methodology

Prime grade Si wafer pieces (2 × 2 cm) were used as substrates for metal catalyst deposition. Pd of thickness of 5 nm was sputtered onto Si substrate, which served as the catalyst. The growth temperature of nanowires was chosen to be 1,100°C based on Pd–Si phase diagram. High-purity Ar was chosen as carrier gas in an open tube furnace and was set to 30 sccm throughout the course of the experiment. Initially the furnace was flushed with Ar to minimize interference from gaseous impurities. The furnace set up with the sample for nanowire growth has been illustrated in Fig. 14.

Growth of Silica Nanowires, Fig. 13 Conditions of no nanowire growth for samples deposited with 2 nm Pt thickness. (**a**) 950°C, 1 h – XRD indicating the formation of Pt_3Si (**c**) and (**b**) 1,200°C, 1 h, No Ambient – EDX spectra revealing PtSi atomic stoichiometry of seeds (**d**)
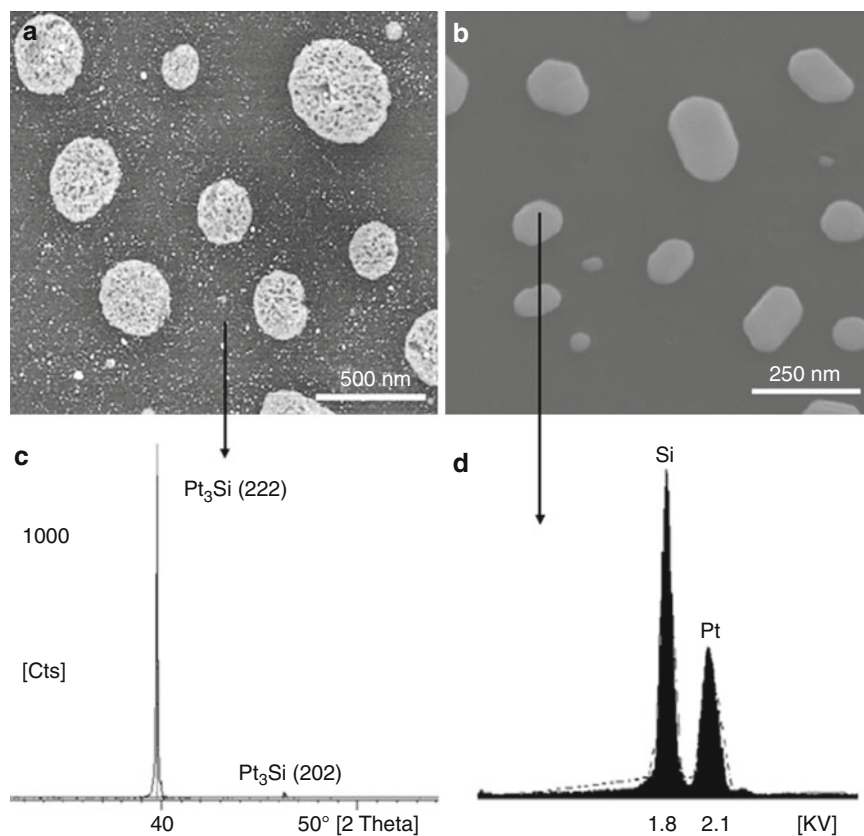
Growth of Silica Nanowires, Fig. 14 Schematic of the furnace and experimental setup
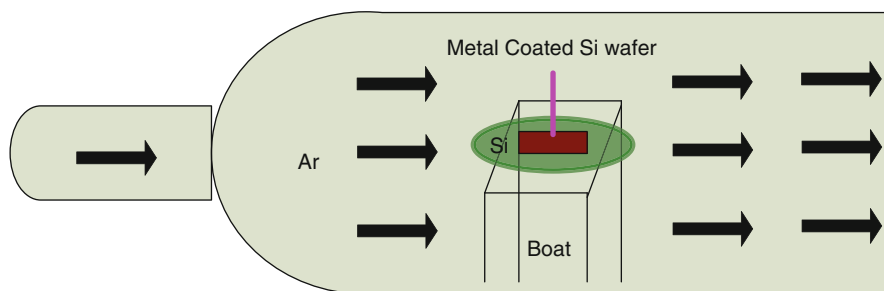
The base support Si wafer acts as Si vapor source. Then, the heating process was timed after the furnace reached the desired temperature. The growth time was varied between 10 and 90 min in order to derive the growth model. For the estimation of growth rate, the heating time was extended up to 210 min. Further, the growth model was validated by (a) heating the metal coated Si substrate in the absence of Ar flow, (b) removing the nanowires via HF treatment and then inspecting the substrate, and (c) varying the Ar flow rate to observe the change in nanowire growth, if any. Finally, the reusability of the Si substrate was evaluated by removing the nanowires grown on Si substrate by HF solution. Then, the Si substrate was again introduced into the furnace to check for nanowire growth. After the heating experiments, the samples were observed under Hitachi S480 SEM. EDS was also performed to identify elemental compositions in addition to XRD.

Discussions

Figure 15 is a SEM composite micrograph showing the morphology of Pd coated Si substrate heated at 1,100°C for various time intervals.

Growth of Silica Nanowires,
Fig. 15 SEM micrograph
illustrating the morphology of
the Pd-coated Si wafer
heated at 1,100°C. (**a**) 10 min,
(**b**) 15 min, (**c**) 20 min, and
(**d**) 30 min
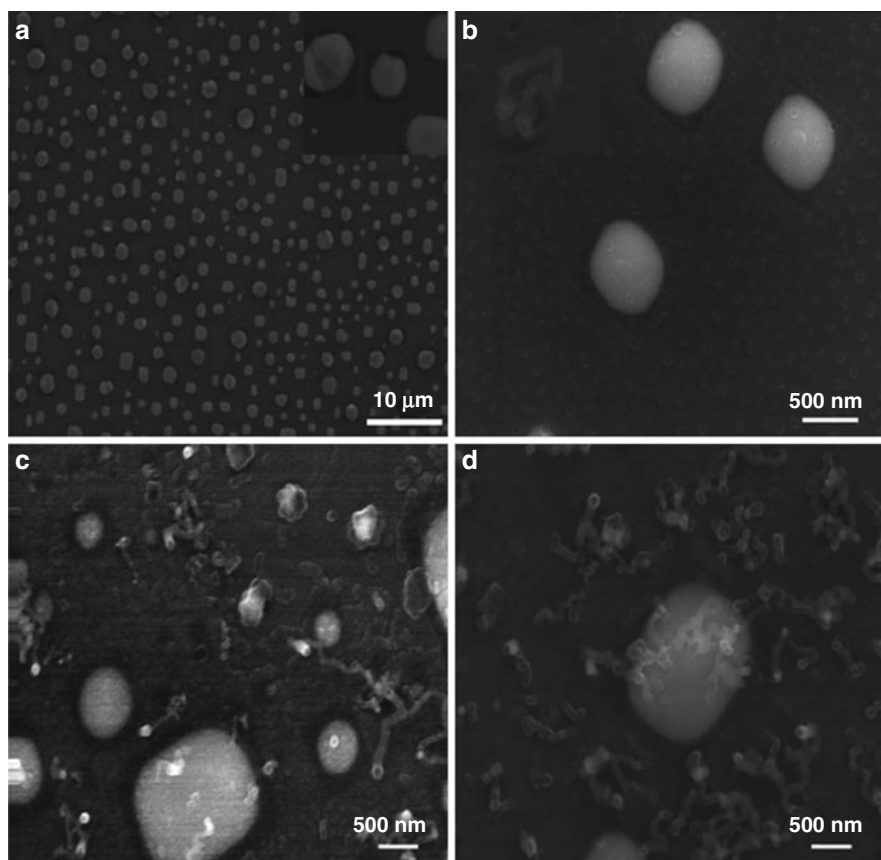
Inspecting the sample after heating 10 min, nanoclusters of various shapes and sizes were observed (shown in Fig. 15a). An EDS inspection on one of the nanoclusters indicates the presence of Pd, Si, and O. Further, the XRD data indicates the presence of crystalline Pd as well as the formation of palladium silicide. Also, a small layer (Fig. 15a inset) covering the wafer surface surrounding the island was noticed. An EDS analysis on the thin layer with quantification indicated the layer to be a SiO_2 film. The extremely low partial pressure and concentration of oxygen in high-purity Ar tank was responsible for the formation of the oxide layer at 1,100°C. A deliberate introduction of oxygen at 0.01 sccm in addition to Ar completely oxidized the substrate without forming the islands or the nanowires. Hence, it is postulated that the thermodynamic equilibrium between the atmosphere (in an open tube furnace) and the low oxygen partial pressure in the tube is critical for nanowire growth. The sample inspected after a heating cycle of 15 min revealed

surface roughening (Fig. 15b) in the form of etched pits both on the surface of Si and palladium silicide island. This has been attributed to the disassociation of the thin oxide layer to volatile SiO with continuous supply of oxygen.

A similar observation of surface roughening of a silicon wafer during annealing in argon ambient at an extremely low oxygen partial pressure was reported by Suzuki [21]. The author concluded that the surface roughening in the form of square pits occur when the oxygen partial pressure is lower than 3×10^{-5} atm. Figure 15c shows the SEM micrograph of a sample heated for 20 min. The nucleation of nanowires could be observed along with ripening of the seeds to induce its birth. Localized Si etching was also observed along the catalyst periphery and subsequent sinking action of the nanoclusters was seen. As the eutectic alloy formation proceeds at high temperature, a tiny amount of Si diffuses from the substrate. The silicide formation consumes silicon from the surrounding which gives

**Growth of Silica Nanowires,
Fig. 16** SEM micrograph
illustrating the morphology of
the Pd-coated Si wafer
heated at 1,100°C. (**a**) 45 min,
(**b**) 60 min, (**c**) 75 min, and
(**d**) 90 min
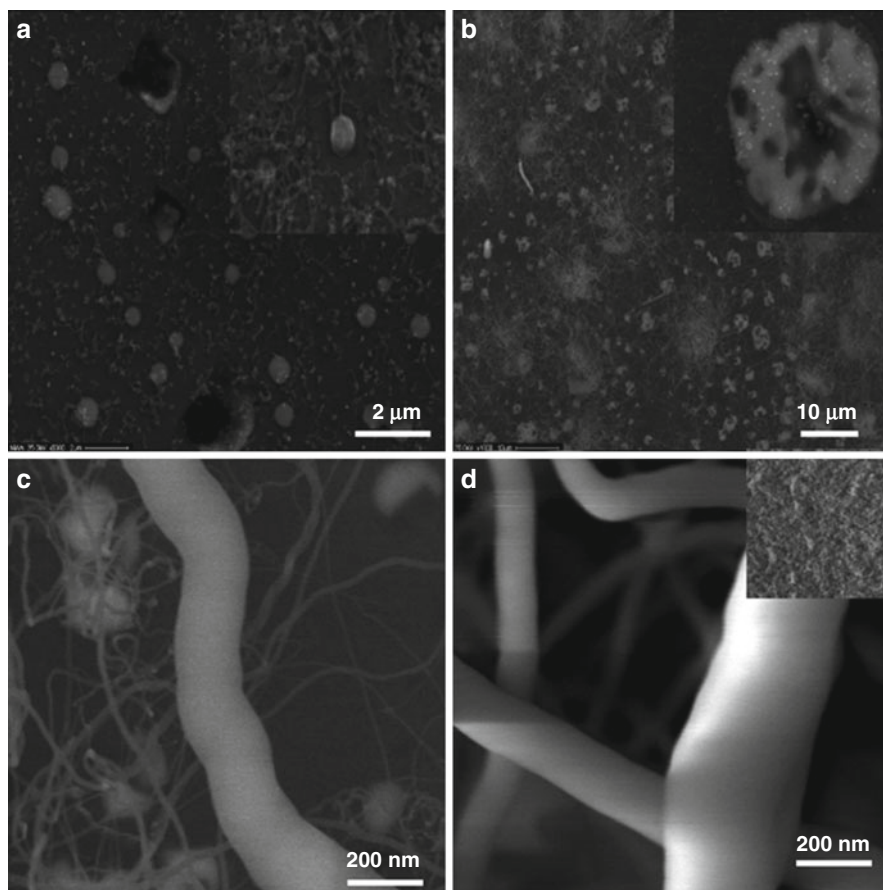
**Growth of Silica Nanowires,
Fig. 17** TEM micrograph
showing migration of Pd
catalyst particles captured at an
early growth stage (3E16 ions/
cm^2, 1,100°C, 7 min).
(**a**) "Birth" of nanowires,
nucleated wires of various sizes
and shapes concentrated with
Pd particles, *Inset*: scanning
electron microscopy (SEM)
micrograph highlighting the
early growth stage, Ostwald
ripening of the seeds observed.
(**b**) Magnified image of a single
nanowire for estimating the
number density of the wires:
1×10^{13} particles/cm^2
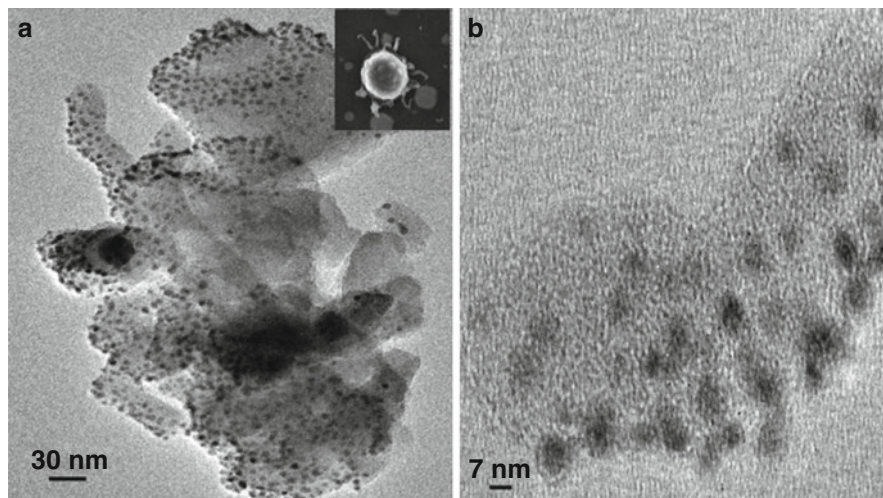
the appearance of Si etching. Peng et al. [22] utilized
the metal-induced Si etching behavior along with
nanolithography to synthesize ordered Si nanowires.
As the nanoclusters vary in size distribution (several

μm to nm), some of the nanowires seem to originate
from silicon. The number density of nanowires
increased twofold upon inspecting a sample heated at
1,100°C for 30 min. EDS spectra on a bunch of

Growth of Silica Nanowires, Table 1 Comparison of number density of silica nanowires as a function of heating time

Heating time (mins)	Number density of nanowires ($\times 10^{12}$ wires/m^2)
20	1
30	2
45	3
60	5
75	6
90	8

nanowires away from the palladium nanocluster indicate Si and O composition. Extracting the stoichiometry by background subtraction and in comparison with a standard quartz substrate reveals SiO_x ($1.8 < x < 2$) stoichiometry indicating a deficiency in silicon.

The number density of nanowires increased for samples heated for 45 min (Fig. 16a). Also, the inset in Fig. 16a shows the reduction in the size of nanoclusters at the expense of nanowire growth. Upon inspected by transmission electron microscope, the catalyst seems to travel up the nanowire as it grows (Fig. 17). Table 1 indicates the progressive increase in the number density of nanowires with increase in heating time. Figure 16b shows bunches of nanowires along with secondary nucleation (inset) from the nanocluster. High aspect ratio nanostructures were seen on samples heated to 75 min (Fig. 16c) along with size reduced parent nanoclusters. Samples heated for 90 min at 1,100°C show intertwined nanowires Fig. 16d devoid of nanoclusters. The bushy growth of nanowires has been shown in the inset.

Next the growth rate (Fig. 18) was estimated by measuring the weight differential of Si substrate before and after nanowire growth as function of growth time. Five distinct growth regimes, namely, (1) thin oxide growth, (2) Si etch pit formation, (3) nanowire nucleation, (4) nanowire growth, and (5) nanowire growth saturation, have been demarcated.

Based on the weight differential data, it can be concluded that SiO_x nanowires are conducive to mass manufacturing when heated at 1,100°C for a heating time above 100 min. Further, the reusability of the substrate was investigated by etching the nanowires off the Si substrate in hydrofluoric acid (HF) and using the processed substrate again for nanowire growth (1,100°C, 90 min). The nanowires grow at the expense of the catalyst in

the sense the catalyst gets consumed inside the nanowires and the Si surface is depleted of the catalyst nanoparticles. Based on these experimental evidences, a growth model was derived and is illustrated in Fig. 19.

Some of the key features of silica nanowire growth via metal thin film-based catalysis and implantation-based seeding approach as inferred from the growth model are as follows:

(a) The low concentration level of oxygen in Ar is the key to the growth of silica nanowires in an open tube furnace.

Absence of nanowire formation was noted in the case of a minute increase in the oxygen level as mentioned earlier. Though the concentration magnitude could not be estimated, the partial pressure of oxygen must be below 3×10^{-5} atm. obtained from earlier report [19] for nanowire growth.

(b) The source of silicon originates both from the substrate (when the metal consumes Si for the alloy formation leaving exposed facets of Si) and the support wafer. Also, the initial oxide layer is transformed into silicon monoxide upon heating the sample at 1,100°C. Presence of SiO vapor aids nanowire growth by supplying both vapor phase Si and O. Such hypothesis is concurrent with the prior investigation on thermodynamic equilibrium between Si and SiO at 1,100°C [23].

(c) A combination of growth models governs nanowire growth. The presence of metal catalyst at the tip of nanowires, formation of silicide in tandem with the need for Ar suggests Vapor Liquid Solid (VLS) model. To rule out the Solid Liquid Solid (SLS) model, the nanowires were mechanically removed and then the substrate was inspected for any silicide nanocluster. The absence of catalyst at the sample surface negates the possibility of SLS model. Also, the formation of a thin oxide layer and role of SiO inducing nanowire growth signifies oxide-assisted growth (OAG) mechanism.

Conclusions

In summary, the manufacturing aspect of silica nanowire growth was investigated. Nanowires were found to nucleate following a combination of VLS and OAG mechanism. For a given thickness of Pd thin film and heating temperature of 1,100°C, the nanowire growth seems to saturate after a heating time

Growth of Silica Nanowires, Fig. 18 Schematic illustrating the growth rate (weight differential) of silica nanowires as a function of heating time
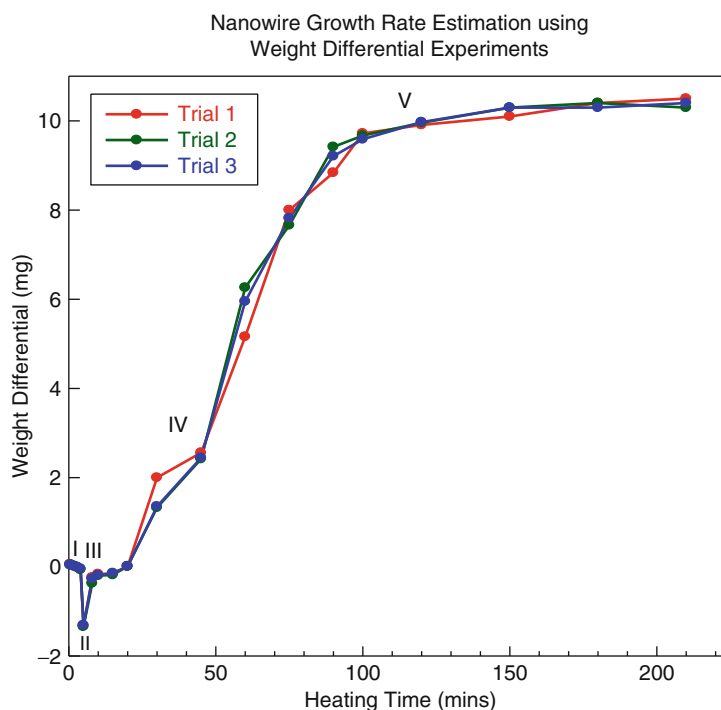

Growth of Silica Nanowires, Fig. 19 Growth model for silica nanowires on Si (**a**) metal catalyst placement in Si using ion implantation or sputtered thin film, (**b**) after sample introduction into furnace – silicide formation along with a thin oxide layer, (**c**) disassociation of oxide layer to form SiO and formation of Si etch pits with increasing heating time, (**d**) birth of silica nanowires with metal catalyst on top, and (**e**) growth of silica nanowires of high aspect ratio
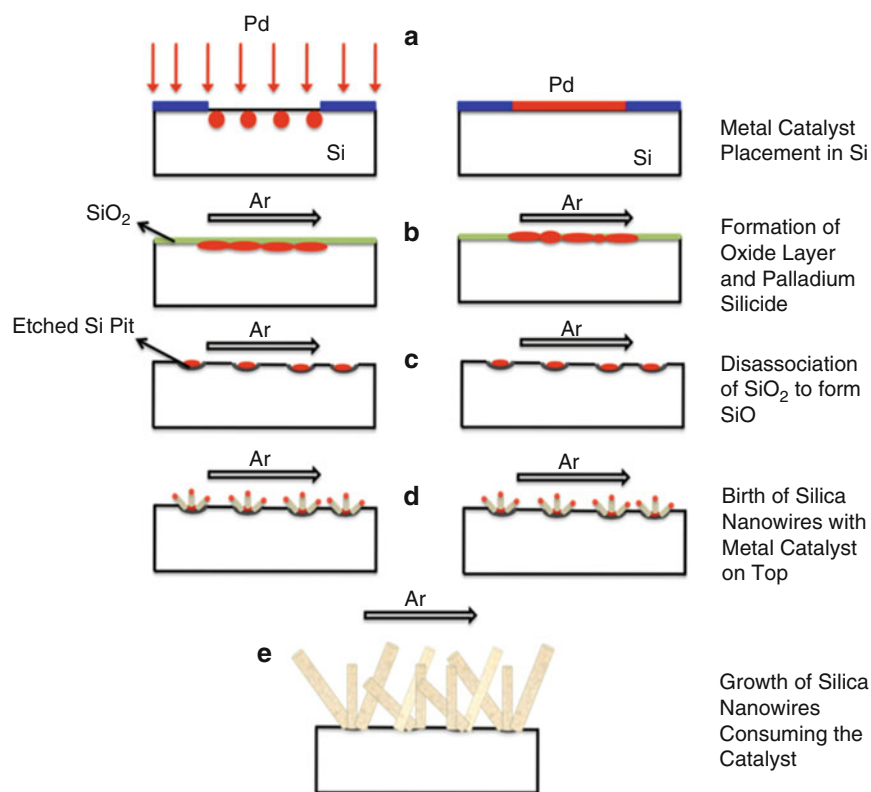
of 100 min. It has been observed that the nanowires grow at the expense of the catalyst negating the reusability for the substrate.

Summary

In this entry, silica nanowire growth was discussed in view of their potential for several unique applications such as localization of light, low-dimensional waveguides for functional microphotonics, SNFOM, optical interconnects on a futuristic integrated optical microchip, sacrificial templates, biosensors, and optical transmission antennae. The "bottom-up" and "top-down" approaches in growing nanowires were discussed with emphasis on the VLS-mediated growth. Further, the research on selective growth of silica nanowires on silicon using metal ion implantation–based seeding approach and metal thin film catalysis was presented. While the former technique has the advantage of precise placement of catalyst and near to uniform growth, the latter offers an inexpensive way to mass manufacture them. A unique feature tying both the synthesis methodologies is the use of simple open tube furnace in a high-purity Ar ambient unlike the use of harmful gases like silane or complex instrumentation such as a CVD system.

Time sequence study of the silica nanowire growth reveals a combination of VLS and OAG mechanism responsible for nanowire nucleation. Silica nanowires grown using such a combinatorial mechanism exhibit high aspect ratio ($\sim$3,000:1) and are amorphous in nature. Growth rate characteristics indicate an initiation period, prolific growth regime, and a steady state phase.

Structural characterization at a high resolution indicates catalyst metal nanoclusters along its body suggesting the formation of a hybrid (metal/dielectric) nanostructure, critical for sensing applications. Finally, a brief highlight on manufacturing aspects of silica nanowires was presented.

Future Directions

Producing an ordered array of vertically aligned nanowires is critical for many applications. In particular, the biosensing applications require silica nanowires to be grown vertically. Vertically aligned nanowires are easier to manipulate and offer greater surface coverage. Silica nanowires can be grown vertically or in an aligned fashion using electric field–assisted assembly, porous alumina templates, and self-assembly using bioengineered interfaces. Also, growing nanowires at low temperature on various substrates is critical for device applications. Apart from the growth aspects, investigating silica nanowires using Local Electrode Atom Probe (LEAP) will yield interesting results on the elemental composition and texture along the length of the wire. The atom probe is a combination of a field ion microscope and a mass spectrometer. Looking at the nanowires at their early stage of growth using LEAP will provide fundamental details on nucleation and the role of catalyst inducing growth. Moreover, these results can give an insight on the growth model.

Cross-References

▶ Biomimetic Synthesis of Nanomaterials
▶ Electric-Field-Assisted Deterministic Nanowire Assembly
▶ Gas phase Nanoparticle Formation
▶ Nanostructured Functionalized Surfaces
▶ Nanostructured Materials for Sensing
▶ Nanostructures for Surface Functionalization and Surface Properties
▶ Optical Techniques for Nanostructure Characterization
▶ Scanning Electron Microscopy

References

1. Foster, L.E.: Nanotechnology: Science, Innovation, and Opportunity. Prentice Hall, Upper Saddle River (2009)
2. Cao, G.: Nanostructures and Nanomaterials: Synthesis, Properties, and Applications. Imperial College Press, London (2004)
3. Sui, C.-H., Wu, P.-H., Zhao, P.: Waveguiding properties and applications of silica nanowires in various environments. Optoelectron. Lett. **5**, 241–243 (2009)
4. Choi, D., McIlroy, D., Nagler, J., Aston, E., Hrdlicka, P., Gustin, K., Hil, R., Stenkamp, D., Branen, J.: One-dimensional silica structures and their applications to the biological sciences. In: Kumar, C. (ed.) Nanostructured Oxides, pp. 83–103. Wiley-VCH, Weinheim (2009)
5. Stathis, J.: Reliability limits for the gate insulator in CMOS technology. IBM J. Res. Dev. **46**, 265–285 (2002)
6. Ollier, E.: Optical MEMS devices based on moving waveguides. IEEE J. Sel. Topics. Quantum Electron. **8**, 155–162 (2002)
7. Sekhar, P.K., Ramgir, N.S., Bhansali, S.: Metal-decorated silica nanowires: an active surface-enhanced Raman

substrate for cancer biomarker detection. J. Phys. Chem. C **112**, 1729–1734 (2008)

8. Tsuneyuki, S., Tsukada, M., Aoki, H.: First-principles interatomic potential of silica applied to molecular dynamics. Phys. Rev. Lett. **61**, 869–872 (1988)

9. Christian, P.: Preparation of ultrafine silica nanowires using iron based surfactant. Langmuir **26**, 1405–1407 (2010)

10. Kievsky, Y., Sokolov, I.: Self-assembly of uniform nanoporous silica fibers. IEEE Trans. Nanotechnol. **4**, 490–494 (2005)

11. Kita, T., Nishimoto, S., Kameshima, Y., Miyake, M.: Characterization of silicon oxide nanowires prepared by hydrothermal treatment using slow dissolution of bulk glass source materials. J. Am. Ceram. Soc. **93**, 2427–2429 (2010)

12. Lin, Y.-C., Lin, W.-T.: Growth of SiO_2 nanowires without a catalyst via carbothermal reduction of CuO powders. Nanotechnology **16**, 1648–1654 (2005)

13. Rao, C.N.R., Deepak, F.L., Gundaiah, G., Govindaraj, A.: Inorganic nanowires. Prog. Solid State Chem. **31**, 5–147 (2003)

14. Wagner, R.S., Ellis, W.C.: Vapor-liquid-solid mechanism of single crystal growth. Appl. Phys. Lett. **4**, 89–90 (1964)

15. Sekhar, P.K., Abd Bhansali, S.: Manufacturing aspects of oxide nanowires. Mat. Lett. **64**, 729–732 (2010)

16. Muralidhar, G.K., Bhansali, S., Pogany, A., Sood, D.K.: Electron microscopy studies of ion implanted silicon for seeding electroless copper films. J. Appl. Phys. **83**, 5709–5713 (1998)

17. Sood, D.K., Sekhar, P.K., Bhansali, S.: Ion implantation based selective synthesis of silica nanowires on silicon wafers. Appl. Phys. Lett. **88**, 143110–143113 (2006)

18. Schmidt, V., Senz, S., Gosele, U.: Diameter-dependent growth direction of epitaxial silicon nanowires. Nano. Lett. **5**, 931–935 (2005)

19. Schmidt, V., Senz, S., Gosele, U.: The shape of epitaxially grown silicon nanowires and the influence of line tension. Appl. Phys. A **80**, 445–450 (2005)

20. Sekhar, P.K., Sanbandam, S.N., Sood, D.K., Bhansali, S.: Selective growth of silica nanowires in silicon catalysed by Pt thin film. Nanotechnology **17**, 4606–4613 (2006)

21. Suzuki, T.: Effect of annealing a silicon wafer in argon with a very low oxygen partial pressure. J. Appl. Phys. **88**, 6881–6886 (2000)

22. Peng, K., Zhang, H., Lu, A., Wong, N.B., Zhang, R., Lee, S.-T.: Ordered silicon nanowires via nanosphere lithography and metal-induced etching. App. Phys. Lett **90**, 163123–163125 (2007)

23. Bahloul-Hourlier, D., Perrot, P.: Thermodynamics of the Au–Si–O system: application to the synthesis and growth of silicon–silicon dioxide nanowires. J. Phase Equilib. Diff. **28**, 150–157 (2007)

Growth of Silicon Dioxide Nanowires

▶ Growth of Silica Nanowires

Gyroscopes

Alexander A. Trusov
Department of Mechanical and Aerospace Engineering, The University of California, Irvine, CA, USA

Synonyms

Inertial sensors; Motion sensors; Orientation sensors; Rate sensors; Rotation sensors

Definition

Gyroscopes are physical sensors that detect and measure the angular motion of an object relative to an inertial frame of reference. The term "gyroscope" is attributed to the mid-nineteenth century French physicist Leon Foucault who named his experimental apparatus for Earth's rotation observation by joining two Greek roots: *gyros*, rotation and *skopeein*, to see. Unlike rotary encoders or other sensors of relative angular motion, the unique feature of gyroscopes is the ability to measure the absolute motion of an object without any external infrastructure or reference signals. Gyroscopes allow untethered tracking of an object's angular motion and orientation and enable standalone Attitude Heading Reference Systems (AHRS). Combining three gyroscopes with three accelerometers in a complete 6-axis Inertial Measurement Unit (IMU) enables self-contained Inertial Navigation Systems (INS) for navigation, guidance, and dead reckoning.

All *gyroscopes* can be divided into two main categories, depending on whether the angular velocity or orientation is being measured [1]. *Rate gyroscopes* measure the angular velocity, or the rate of rotation of an object. *Angle gyroscopes*, also called *Whole Angle* or *Rate Integrating* gyroscopes, measure the angular position, or orientation of an object directly. While devices sensitive to the angular acceleration are used in some applications, these sensors are typically not referred to as gyroscopes, but rather as angular accelerometers. Essentially all currently existing Micro-Electro-Mechanical-Systems (MEMS) gyroscopes are of the rate measuring type and are typically employed for motion detection (e.g., in consumer electronics and automotive safety devices) and motion
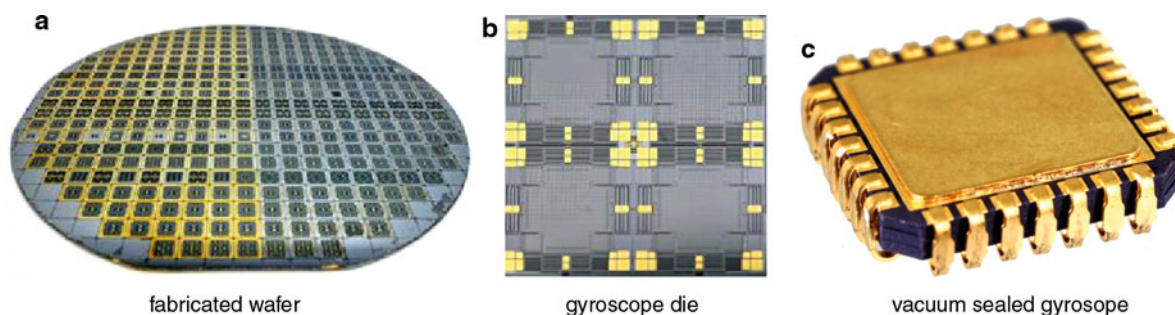
Gyroscopes, Fig. 1 Photographs of a wafer-level fabricated silicon-on-isolator gyroscope with capacitive transduction designed, fabricated, and packaged at the University of California, Irvine MicroSystems Laboratory, http://mems.eng.uci.edu/

stabilization and control (e.g., in smart automotive steering and antenna/camera stabilization systems).

True INS and AHRS applications rely on the continuous tracking of the object's orientation. Measurement of the angular position can be accomplished either by numerical integration of a rate gyroscope's output, or by using an angle gyroscope which effectively integrates the rotation rate by virtue of its internal dynamics and outputs the angle information directly. When a rate gyroscope is used to track the orientation, its output signal is integrated over time together with the associated errors and noise, leading to fast buildup of the orientation angle drifts (e.g., white noise in the angular rate signal results in $1/f^2$ drift, or random walk, of angle). Successful realization of standalone gyroscope-based INS and AHRS thus requires either angle gyroscopes or rate gyroscopes with extremely stable and low noise output.

Historical Overview of Vibratory Gyroscope Technologies

Early work by Leon Foucault during the mid-nineteenth century explored two different design paradigms for angle measuring mechanical gyroscope based on either a spinning or vibrating mass. While the spinning mass approach was the dominant method of mechanical gyroscope construction from its inception well into the second half of the twentieth century, it is not well suited for MEMS implementation due to the technological limitations in the manufacturing of precision, low friction bearings. Few designs of spinning mass MEMS gyroscopes using electrostatic levitation have been reported in the literature [2, 3] without

yet achieving commercial success due to the inherent instability of the mechanical system and necessity for a sophisticated control system.

The vibrating mass approach, illustrated by the popular Foucault pendulum experiment, exploits the exchange of energy between different axes of vibration due to the Coriolis effect. This architecture, at present referred to as the Coriolis Vibratory Gyroscope (CVG) [4] remained largely a scientific curiosity for almost a century until the introduction of a functional vibratory gyroscope by Sperry in the mid-twentieth century [5] followed by successful commercialization of quartz tuning fork gyroscopes by BEI Technologies in the late-twentieth century [6], and very high volume deployment of silicon MEMS CVGs in the early twenty-first century. Today, silicon vibratory rate gyroscopes with capacitive transduction comprise the majority of MEMS gyroscopes in development and production, with some research groups and manufacturers pursuing quartz devices with piezoelectric transduction or silicon devices with alternative transduction mechanisms such as inductive or electromagnetic [7]. Figure 1 shows photographs of a wafer-level batch fabricated silicon-on-insulator (SOI) gyroscope with capacitive actuation and detection designed, fabricated, and packaged at the University of California, Irvine [8].

## Coriolis Vibratory Gyroscope Dynamics

In this section, the principles of operation of Coriolis vibratory gyroscopes are derived from the basic concepts of classical mechanics. While a planar, z-axis vibratory gyroscope is the focus of the explanation, the discussions are generic in nature and equally

apply to devices with other architectures, including torsional, out-of-plane, and sensors with multiple modes of vibration for simultaneous detection of rotation in several axes. Let $S\xi\eta\zeta$ denote an inertial reference frame, in which a non-inertial frame $Oxyz$ is moving with a linear acceleration $A_0 = (A_{0x}, A_{0y}, A_{0z})$ and an angular velocity $\Omega = (\Omega_x, \Omega_y, \Omega_z)$, where the components of the vectors are given with respect to the moving frame. Coordinates of a point mass in the inertial and non-inertial reference frames are given by vectors $\rho = (\xi, \eta, \zeta)$ and $r = (x, y, z)$, respectively. Newton's equation of motion of the point mass m under the action of force vector $F = (F_x, F_y, F_z)$ with respect to the inertial reference frame $S\xi\eta\zeta$ are given by $\rho'' = A_{abs} = F/m$, where the prime denotes derivative with respect to time. According to the rules of differentiation in a moving frame:

$$
\begin{aligned}
r'' = A_{rel} \\
= F/m - 2[\Omega \times r'] - (A_0 + [\Omega \times [\Omega \times r]] \\
+ [\Omega' \times r]).
\end{aligned} \tag{1}
$$

This differential vector equation provides a mathematical foundation for the Coriolis vibratory gyroscopes. The Coriolis cross-product of the angular velocity Ω and the coordinate vector $r = (x, y, z)$ governs the coupling and exchange of energy between x, y, and z non-inertial axes. This effect allows measuring the input angular velocity Ω by observing the vibration pattern of the proof mass m relative to the non-inertial device reference frame $Oxyz$.

A theoretical model of a single axis Coriolis vibratory gyroscope is shown in Fig. 2. A point mass m forms the Coriolis sensitive proof mass and is constrained to motion in the Oxy plane under the influence of elastic forces of the suspension and inertial forces, caused by the motion of the $Oxyz$ non-inertial reference frame. To simplify the conceptual discussion, it is assumed that the proof mass suspension produces a linear field of elastic forces $F_s = (k_x x, k_y y)$. Also neglected are the projection of the gravity field on the Oxy plane, out-of-plane Oz axis dynamics, and acceleration of the origin O. These assumptions are justified by high out-of-plane stiffness of typical bulk micromachined structures and large frequency separation between the external and Coriolis accelerations. After linearization with respect to the input angular velocity vector Ω components Ω_x, Ω_y, Ω_z, Eq. 1 can be simplified to the following:
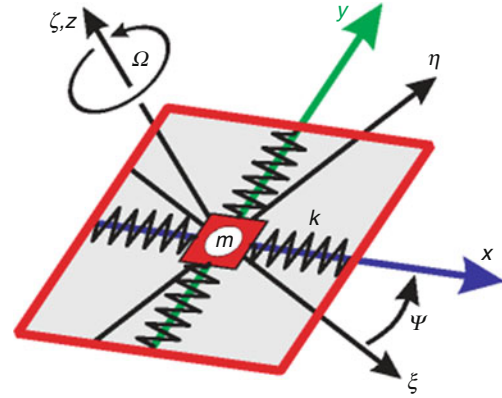
Gyroscopes, Fig. 2 Theoretical model of a single z-axis vibratory gyroscope consisting of a proof mass m suspended in the x-y plane. The xyz non-inertial frame of reference associated with the sensor is moving relative to the inertial frame $\xi\eta\zeta$ with an angular velocity $\Omega = (0, 0, \Omega_z)$. Coriolis force coupling between the x and y coordinates causes energy exchange which is used to detect the input rate Ω_z

$$
\begin{aligned}
x'' + x\omega_x^2 - y\Omega'_z - 2\Omega_z y' = F_x, \\
y'' + y\omega_y^2 + x\Omega'_z + 2\Omega_z x' = F_y,
\end{aligned} \tag{2}
$$

where ω_x and ω_x are the mechanical natural frequencies of the x- and y-mode, respectively. The following discussion will consider two different modes of operation of Coriolis vibratory gyroscopes – rate measuring and angle measuring.

Rate Gyroscope Operation

Rate gyroscopes are operated with inherent nonsymmetry between the Ox axis, designated as the drive-mode, and the Oy axis, designated as the sense-mode. The drive-mode is operated in the forced vibrations mode, where the excitation force F_x is a sinusoidal waveform with amplitude f and angular frequency ω_d. The sense-mode of a rate gyroscope can be operated either open-loop or in a force-to-rebalance closed loop, where a feedback force is generated to suppress the sense-mode vibrations and is simultaneously used as the measure of the input rate. Dynamics of a rate gyroscope with an open-loop sense-mode can be derived from Eq. 2 as:

$$
\begin{aligned}
x'' + x\omega_x^2 - y\Omega'_z - 2\Omega_z y' = f/m \, sin(\omega_d t), \\
y'' + y\omega_y^2 + x\Omega'_z + 2\Omega_z x' = 0.
\end{aligned} \tag{3}
$$
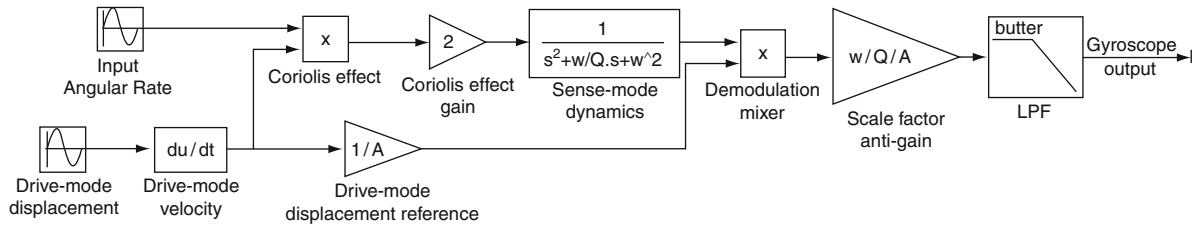
Gyroscopes, Fig. 3 Block diagram of a Coriolis vibratory rate gyroscope with closed-loop drive- and open-loop sense-modes. Input rotation rate is amplitude modulated by the drive-mode velocity due to the Coriolis effect

In practical implementations of vibratory rate gyroscopes, stabilization of the drive-mode velocity is desired to reduce the effect of manufacturing tolerances and operational conditions on the scale factor of the sensor. This is accomplished by an electronic feedback system which regulates the amplitude f of the driving force by means of an *Automatic Gain Control (AGC)* to maintain constant amplitude of the drive-mode motion. The frequency ω_d of the drive-mode feedback force is generated by a *Phase-Lock Loop (PLL)* which locks to the drive-mode mechanical resonance frequency ω_x. An alternative approach for drive-mode operation is through positive feedback of the velocity signal x', which effectively destabilizes the drive-mode and maintains resonance of the drive-mode. In the velocity feedback configuration, the need for a separate frequency source is eliminated, however often at the cost of increased jitter and noise in the Coriolis signal demodulation block. Simplified equations for a vibratory rate gyroscope with a closed-loop drive-mode and an open-loop sense-mode (see Fig. 3) are derived from Eq. 3 as:

$$
\begin{aligned}
x &= sin(\omega_x t), \\
y'' + y'(\omega_y/Q_y) + y\omega_y^2 &= -2\Omega_z x',
\end{aligned}
\tag{4}
$$

Here, an energy dissipation term, defined by the sense-mode quality factor Q_y, is now included in the sense-mode dynamics, while the term $x\Omega_z'$ is omitted assuming a slow-varying input Ω_z.

Figure 4 illustrates the motion of the proof mass relative to the device reference frame in the presence of input rotation. The drive-mode is a resonator driven to a constant amplitude of vibrations at a fixed frequency by means of a feedback system. During rotation of the device, the Coriolis effect causes coupling of energy from the drive-mode to the sense-mode. The Coriolis
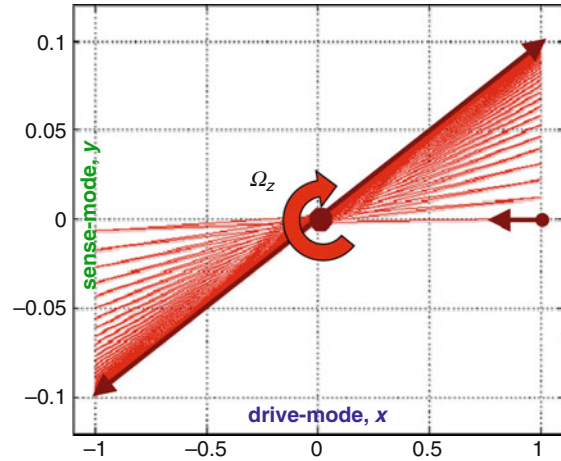
Gyroscopes, Fig. 4 Operation of a vibratory rate gyroscope is based on the excitation of the sense-mode by the Coriolis force, stemming from the cross-product of the drive-mode velocity and the input rotation rate. Sense-mode vibrations present an Amplitude Modulated (AM) signal of the input rotation rate with the drive-mode frequency serving as the carrier

force applied to the sense-mode is given by the cross-product of and is proportional to the drive-mode velocity x' and the input angular rate Ω_z. Since the drive-mode velocity is a sinusoidal signal with a fixed frequency, the Coriolis effect results in the *Amplitude Modulation (AM)* of the input angular rate by the drive-mode velocity. To measure the input angular rate, displacement (or, equivalently, velocity) of the Coriolis force induced sense-mode vibrations is measured by means of an amplitude demodulation.

Coriolis vibratory rate gyroscopes are commonly divided into two classes depending on the relative matching or mismatching between the drive-mode and sense-mode resonant frequencies. When the drive frequency is within the bandwidth of the sense-mode resonance peak, the mechanical sensitivity of the gyroscope is maximized proportionally to the sense-mode
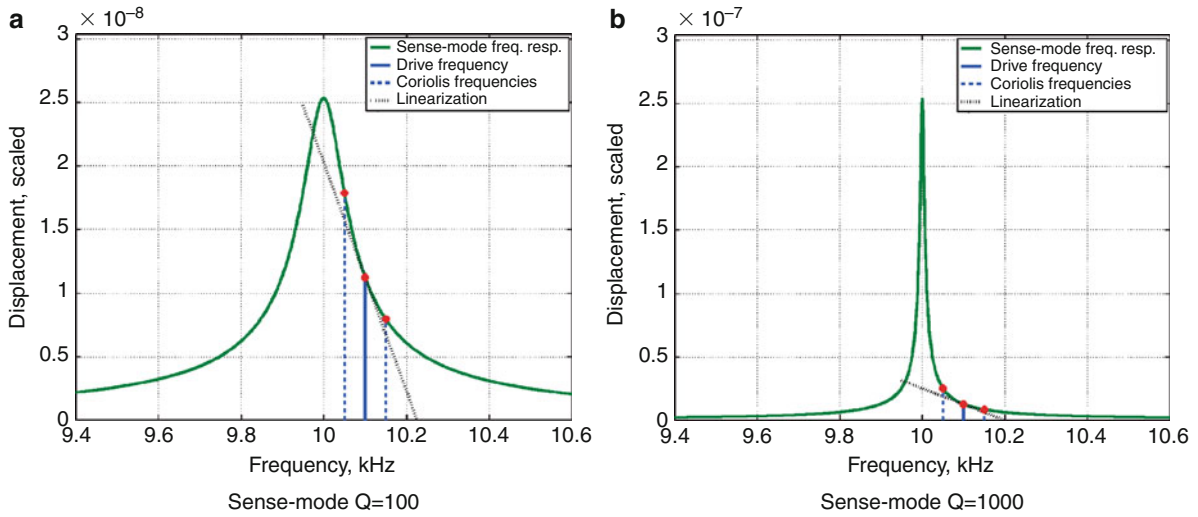
Gyroscopes, Fig. 5 Frequency domain representation of a mode-mismatched vibratory rate gyroscope response to a sinusoidal input angular rate with a 50 Hz frequency. The simulation assumes sense-mode resonant frequency of 10 kHz and drive-mode excitation at 10.1 kHz. (**a**) Sense-mode Q = 100. (**b**) Sense-mode Q = 1,000

quality factor Q_y. However, the sensitivity and *Signal-to-Noise (SNR)* advantages of this configuration come at the trade-off of the open loop rate sensor bandwidth, which in this case is inversely proportional to the sense-mode Q_y. An alternative design approach is the mismatching of the drive frequency from the sense-mode resonance by a prescribed amount.

In order to illustrate the dynamic response of a mode-mismatched rate gyroscope to a time-varying input angular velocity is considered in the form of $\Omega_z = cos(\omega_{in}t)$ and the output scale factor is studied as a function of the input frequency ω_{in}. According to Eq. 4, the sense-mode is excited by $-2\Omega_z x' = \omega_d[cos((\omega_d - \omega_{in})t) + cos((\omega_d - \omega_{in})t)]$. This means that for a given $\Omega_z = cos(\omega_{in}t)$ input, the sense-mode has two forcing components of equal amplitudes at frequencies $(\omega_d \pm \omega_{in})$. This concept is illustrated with numerical modeling in Fig. 5, where the sense-mode frequency is set to 10 kHz with a quality factor of 100 and 1,000, the drive-mode frequency is 10.1 kHz (i.e., 1% mismatching), and the input angular rate frequency is 50 Hz.

As shown in Fig. 3, the signals at two Coriolis frequencies $(\omega_d \pm \omega_{in})$ both pass through the sense-mode dynamics and are then mapped back into a single frequency ω_{in} by the demodulation block. Figure 6 shows the input-output frequency response of a mode-mismatched vibratory gyroscope highlighting

two distinct cases defined by the relationship between ω_y, Q_y, and ω_d. For a relatively low sense-mode Q_y, the gyroscope's 3-dB rate bandwidth includes the resonant peak at the $(\omega_d - \omega_y)$ frequency and the bandwidth exceeds the frequency mismatch between the drive- and sense-modes. For relatively high sense-mode Q_y, the resonant peak at the $(\omega_d - \omega_y)$ frequency escapes the gyroscope bandwidth, which in this case is given by approximately half of the drive- and sense-mode frequency mismatch. A detailed discussion on the mechanical design of vibratory rate gyroscopes is given in [9], where a multi-degree of freedom mechanical design of drive- and sense-modes is proposed as a way to increase the sensor robustness and bandwidth. A different approach to resolving the quality factor versus bandwidth trade-off is the instrumentation of a vibratory gyroscope for direct angle measurements, which is based on the *x-y* symmetric free vibrations regime.

Angle Gyroscope Operation

The resolution and sensitivity of MEMS vibratory gyroscopes are often improved by maximizing Q-factors and reducing frequency separation between the two modes. Mode-matched silicon gyroscopes with Q of 100,000 have been demonstrated to provide

Sense-mode at 10 kHz, Q=100, 10.1 kHz drive

Sense-mode at 10 kHz, Q=1000, 10.1 kHz drive
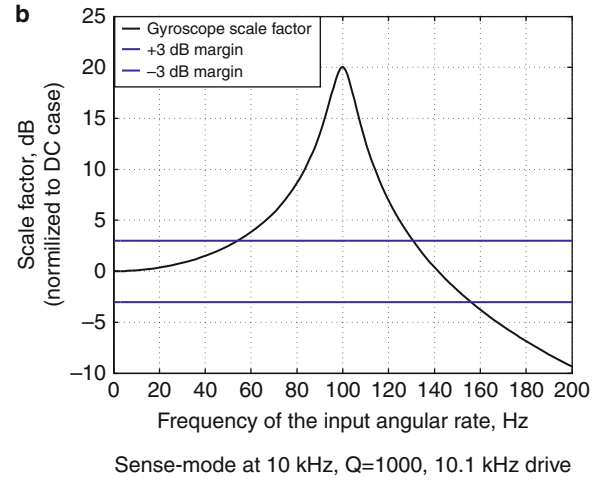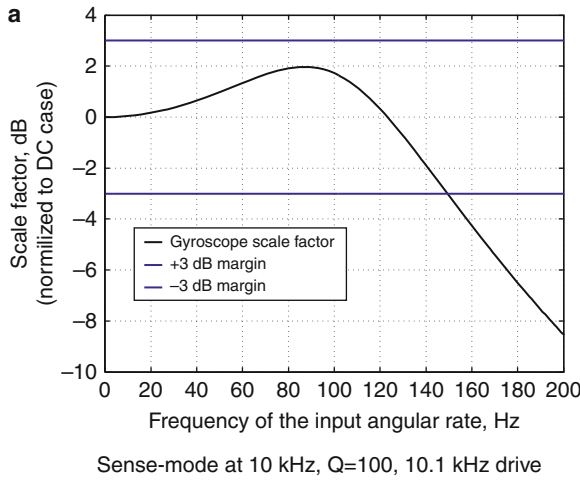
Gyroscopes, Fig. 6 Input rate bandwidth of a mode-mismatched vibratory rate gyroscope. The simulation assumes sense-mode resonant frequency of 10 kHz and drive-mode excitation at 10.1 kHz. (**a**) Sense-mode at 10 kHz, Q = 100, 10.1 kHz drive (**b**) Sense-mode at 10 kHz, Q =1000, 10.1 kHz drive

sub-degree-per-hour bias stability [10]. However, mode matching of a high-Q rate gyroscope typically limits the open-loop sense-mode range to below 50 deg/s and rate bandwidth to 1 Hz. The traditional approach is forced-to-rebalance closed-loop operation of the sense-mode. While this approach extends the input-output bandwidth of a rate gyroscope, the signal-to-noise advantage of the high-quality factor is only preserved at low frequencies of the input angular rate. Two solutions have been proposed. The first approach is to design a relatively low-Q gyroscope with sub-micron capacitive gaps required for efficient transduction at a very high resonant frequency of 3 MHz [11]. The second approach is to operate an ultrahigh Q-factor gyroscope in the free vibrations mode, enabling direct angle measurements with mechanically unlimited sensing range and bandwidth.

Whole Angle operation dictates stringent requirements for the structural frequency and Q-factor symmetry, calling for new design architectures and fabrication methods. For example, the macroscale hemispherical resonator gyroscope (HRG) with sub-arcsecond angle resolution operates based on a mechanical element with isotropic Q-factors as high as 26 million [12]. Achieving the required level of damping and stiffness symmetry across process variations of conventional silicon MEMS technologies is very challenging. The latest advances in design and packaging of silicon MEMS devices enabled vibratory gyroscopes with Q-factors of one million [8], invigorating the development of angle measuring MEMS gyroscopes.

An ideal angle gyroscope, such as the classical Foucault pendulum, is a 2-D isotropic mass-spring system vibrating with the natural frequency ω, where $\omega = \omega_x = \omega_y$. In the presence of the inertial rotation with the rate Ω_z, the equations of motion can be derived from Eq. 2 by assuming very low dissipation and absence of any external forces as:

$$
\begin{aligned}
x'' + x\omega^2 - y\Omega'_z - 2\Omega_z y' &= 0, \\
y'' + y\omega^2 + x\Omega'_z + 2\Omega_z x' &= 0.
\end{aligned}
\tag{5}
$$

The whole-angle mode is based on the free precession of the axis of vibrations in response to the inertial rotation, Fig. 7. Orientation of the free vibrations axis remains fixed in the inertial space, providing an instantaneous orientation reference. Unlike the conventional rate measuring mode, where the axis of vibration is effectively locked to the intended drive direction, the whole-angle mode is not mechanically limited in rotation rate range or bandwidth. Changes in the precession angle Θ can be detected by monitoring x and y displacement components according to $\Theta = atan\,(y/x)$.

The above conceptual analysis of the whole-angle operating mode assumes isotropic stiffness and negligible damping. Detailed error model analysis shows that
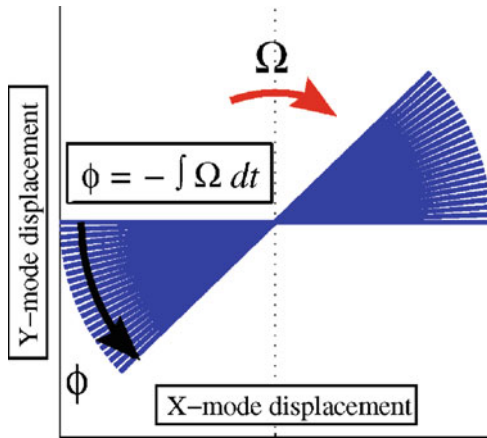
Gyroscopes, Fig. 7 Illustration of the whole-angle mode: the orientation of the vibrations axis remains fixed in the inertial space, enabling direct angle detection

Gyroscopes, Table 1 Typical performance parameters of state-of-the-art commercial MEMS rate gyroscopes [16]

Parameter	Value
Rate resolution	5–10 deg/h/√Hz
Temperature sensitivity of bias	10 deg/h/deg C
Bias repeatability (temperature, turn on)	30 deg/h
Temperature sensitivity of scale factor	250 ppm/°C
Scale factor repeatability (temperature, turn on)	400 ppm

the key sources of the angle error are the frequency mismatch $\Delta\omega = \omega_x - \omega_y$ and energy dissipation time constant mismatch $\Delta(1/\tau) = 1/\tau_x - 1/\tau_y$. Minimization of the damping asymmetry drift requires very low dissipation structures with isotropic Q-factors. For instance, the damping asymmetry $\Delta(1/\tau)$ of 1e-5 rad/s with τ of 160 s (equivalently, $\Delta Q/Q$ of 0.1%) is equivalent to the angle drift of 1°/h. These considerations motivate the advancement of fabrication of flat silicon structures with post-fabrication trimming and electrostatic tuning. Novel 3-D fabrication approaches for axisymmetric shell resonators with inherent smoothness and isotropy provide another promising possibility [13].

Historical Milestones of MEMS Gyroscopes

Since their initial introduction two decades ago in 1991 [14], the resolution of silicon micromachined gyroscopes has been improving by approximately a factor of 10 every 2 years [15]. Most recent literature on state-of-the-art vibratory MEMS gyroscopes with electrostatic/capacitive transduction reports precision on the order of 10 deg/h [16] with a trend of achieving better than 1 deg/h in controlled conditions [10] (Table 1). Despite the remarkable continuing improvement in the angular rate resolution achieved by micromachined vibratory gyroscopes, measurement repeatability, accuracy, and long-term drift still limit deployment in a variety of field applications where autonomous and repeatable operation is required over extended time

and varying environments. Currently, there are several companies producing tactical grade inertial sensors and systems based on the silicon MEMS technology, Table 2; inertial grade systems are yet to be demonstrated. A comprehensive review of MEMS vibratory rate gyroscope commercialization trends are presented in [9]; here, some of the most significant milestones published in literature are reviewed.

In 2002, Analog Devices reported a surface micromachined polysilicon gyroscope with a 50 deg/h random bias instability achieved at the averaging time below 1 min [17]. The single-chip integrated device was not vacuum packaged, which in combination with a 4 μm thin structural layer translates into a relatively low sense-mode quality factor on the order of 20. This family of gyroscopes has since gained a widespread acceptance and popularity in consumer and automotive markets, with new gyroscopes introduced every few years. According to Table 2, this device falls into the middle of the Rate grade category, as defined in [15]. Publication [17] also projected that the random bias can be improved by approximately 80% by introducing new electronics design, suggesting a mechanical-thermal limit on the level of 10 deg/h. The newer generation design of the gyroscope has four electrically coupled, low quality factor surface micromachined polysilicon Coriolis vibratory gyroscopes integrated on the same die together with the electronics in order to suppress the effect of external vibrations; improvement of the bias instability was also reported.

A higher performance polysilicon micromachined gyroscope from Bosch was reported in 2007 in [18]. The device has a 1 deg/h random bias instability, showing its potential for the Tactical grade performance category, Table 2. However, this high sensitivity occurs only at a long, averaging times on the order 160 min. The minimal detectable signal at practical bandwidths was lower (30 deg/h for a 10 Hz

Gyroscopes, Table 2 Performance requirements for three grades of gyroscopes [15]

Parameter, unit	Rate grade	Tactical grade	Inertial grade
Angle random walk (ARW), deg/√h	>0.5	0.5–0.05	<0.001
ARW-equivalent rate resolution, deg/h/√Hz	>30	30–3	<0.06
Bias drift, deg/h	10–1,000	0.1–10	<0.01
Scale factor accuracy, ppm	100–1,000	10–100	<1
Full scale linear range, deg/s	50–1,000	>500	>400
Bandwidth, Hz	>70	100	100

bandwidth, and 100 deg/h for a 100 Hz bandwidth). This device was not integrated with electronics on the die level (hybrid integration of the MEMS and the electronics dies was done on a package level), and its low noise level is attributed to vacuum packaging and high sense-mode quality factor.

A 2006 publication [16] from Charles Stark Draper Laboratory provides an insightful review of performance characteristics of a few additional commercial MEMS gyroscopes. According to this paper, the uncompensated thermal sensitivity of higher performance quartz tuning fork gyroscopes from Custom Sensors and Technologies (formerly Systron Donner/BEI) was reported at 300 ppm/°C, achieved primarily due to extremely low sensitivity of the material's properties to temperature. Analog Devices' automotive grade angular rate sensor ADXRS150 scale factor uncompensated thermal sensitivity was quoted as 1,700 ppm/°C. Honeywell's dissolved wafer, single crystal silicon on glass, device based on the Draper laboratory design achieves a 250 ppm/°C scale factor temperature coefficient.

Surveying the literature confirms that in recent years performance of the commercial MEMS gyroscopes continued to improve, reaching the criteria of Tactical grade category; however, performance of these gyroscopes is still two to three orders of magnitude away from the Inertial grade. The reported progress often comes with costly refinements of the fabrication and vacuum packaging processes. Most reported commercial gyroscopes have intricate electronic compensation for such parameters as temperature drift and aging. From the mechanical structure point of view, the majority of commercial gyroscopes consist of several coupled, anti-phase-driven vibratory gyroscope elements (2- or 4-tine tuning fork architecture) to suppress external vibrations.

Striking breakthroughs in very large scale commercialization of very low cost (<$1 per axis) MEMS rate gyroscopes and accelerometers have been achieved during 2005–2010 by InvenSense, ST Microelectornics, and other suppliers targeting consumer electronics devices. In contrast to the automotive or defense applications with stringent requirements for reliability and high accuracy, size and price are the main drivers for the consumer electronics applications. Current emphasis in low-cost inertial sensor commercialization is on a System-in-a-Package (SiP) and ultimately single-chip integration of 6 axis of inertial sensing augmented with a 3 axis magnetometer and a pressure sensor. Kalman filter-based sensor fusion is employed to produce useful 6-axis position and orientation information by taking advantage of the inherent sensory redundancy in the 10 degrees of freedom produced by these sensors.

Systematic Performance Parameters

In this section an explanation is provided of several most important terms used to describe and quantify the characteristics of Coriolis vibratory gyroscopes [4] with emphasis on deterministic and stochastic aspects of performance or rate sensors.

Scale factor (SF) is the ratio between the output and input signals in the linear regime of measurements. For rate gyroscopes with analog voltage output, *SF* is often expressed in V/(deg/s) or V/(deg/h). Several nonideal properties of *SF* are typically characterized, including sensitivity to temperature (in ppm/°C) and linear accelerations (in ppm/g), and *SF* turn-on repeatability (in ppm for 1-σ or 3-σ statistics). Common sources of *SF* error in MEMS gyroscopes are fabrication imperfections, drifts of drive- and sense-mode resonant frequencies and quality factors, and drifts of electronics gains, especially due to turn-on self-heating.

Bandwidth (BW) is the range of the input signal's frequencies which can be measured well by the gyroscope. BW is often specified assuming a ±3 dB margin

around the gain (Scale Factor) of the gyroscope for a nonzero constant input rate. In general, the gyroscope's input-output BW should be distinguished from the individual mechanical bandwidths of the drive- or the sense-modes of vibration. A more complete picture of the gyroscopes dynamic response is given by a frequency response plot.

Bias is the apparent output of the rotation sensor for zero input rate assuming specific fixed operational conditions (e.g., temperature). For rate gyroscopes *Bias* is expressed in deg/s or deg/h units. Typical associated nonidealities are the sensitivity of bias to temperature (in deg/s/°C or deg/h/°C) and acceleration (deg/s/g or deg/h/g), *Bias* turn-on repeatability (in deg/s or deg/h for 1-σ or 3-σ statistics). A major sources of *Bias* error in MEMS gyroscopes is mechanical quadrature – the undesired sense-mode vibration caused by the fabrication imperfections. While phase-sensitive demodulation suppresses the quadrature signal by several orders of magnitude, minute phase drifts in the system cause so-called quadrature spillover error – the erroneous change of bias due to the partially unsuppressed quadrature. Even though random noise and drift effects are often referred to as bias, it is important to draw a clear distinction between systematic and random changes of output.

Random Noise Performance Parameters

Bias Instability (BI) refers to the additive error in the gyroscope's output with stochastic or random characteristics. Different random noise modes dominate the overall gyroscope noise level depending on the averaging time, which in turn defines the useful sensor bandwidth. The Allan deviation method is often used to identify and quantify random noise modes with different autocorrelation properties and expose their effect when the output signal is averaged or integrated over time.

Allan deviation analysis of a time domain signal $w(t)$ consists of computing its Allan deviation (also know as root Allan variance) $\sigma(\tau)$ for different integration time constants τ and then analyzing the characteristic regions and log–log scale slopes of the $\sigma(\tau)$ curve to identify different noise modes, i.e., random additive components of the signal with different autocorrelation power laws. The first step of the analysis is to acquire a time history $w(t)$ of the gyroscope's output using an

experimental setup or a numerical computer simulation. The second step is to fix the integration time constant τ, divide the time history of the signal into K number of clusters of width τ, then average the signal over each cluster to obtain $w_{i=1...K}$, and finally compute the Allan variance defined as one half of the average of the squares of the differences between the successive averaged values: $\sigma^2(\tau) = 0.5 < (w_{i+1} - w_i)^2 >$. The described sequence of steps yields an estimated value of the Allan deviation $\sigma(\tau)$ for the chosen integration time constant τ, Fig. 8. To obtain the whole $\sigma(\tau)$ curve, the computation is repeated multiple times for a sequence of τ values. Typically, the integration time constants τ are iterated through the multiples or powers of the time discretization step (inverse of the signal sampling rate). While an Allan deviation curve for a signal contains essentially the same information as the signal's power spectral density (*PSD*), it presents the information in an alternative and often more convenient way. The power laws of *PSD* and Allan deviation are related to each other in the following way:

$$PSD(f) \propto f^\alpha \text{ is equivalent to, } \sigma(\tau) \propto \tau^\beta, \\ \text{where } \beta = -(\alpha+1)/2 \text{ and } f = 1/\tau. \quad (6)$$

Table 3 presents a classification of the main noise components in MEMS gyroscopes based on their power spectral density and Allan deviation power laws.

White noise is a random signal with a constant power spectral density. While this definition presents a mathematical abstraction not physically possible due to its infinite total power, it gives a useful model of random processes with very short characteristic autocorrelation times. For example, the random motion of particles due to mechanical-thermal noise at room temperature has a characteristic cutoff frequency of approximately 6 THz. Important characteristic properties of a white noise type random process in time, frequency, and Allan deviation domains are shown in Fig. 9 using a simulated signal. Power spectral density of the white noise is independent of frequency, i.e., $PSD(f) \propto f^\alpha$ with $\alpha = 0$. As expected from Eq. 6, the numerically estimated Allan deviation of the white noise is governed by $\sigma(\tau) \propto \tau^\beta$ with $\beta = -1/2$. In the Allan deviation plot, the rate white noise is identified by fitting the $-1/2$ sloped part of the log-log scale curve with $y = ax^{-1/2}$ and estimating the value of a. In rate gyroscopes, the amount of white noise is

Fig. 8 Illustration of the
Allan deviation analysis
procedure using a simulated
random process with a white
noise ($-1/2$ slope in the Allan
deviation) and a random walk
component (integral of white
noise, $+1/2$ slope in the Allan
deviation)
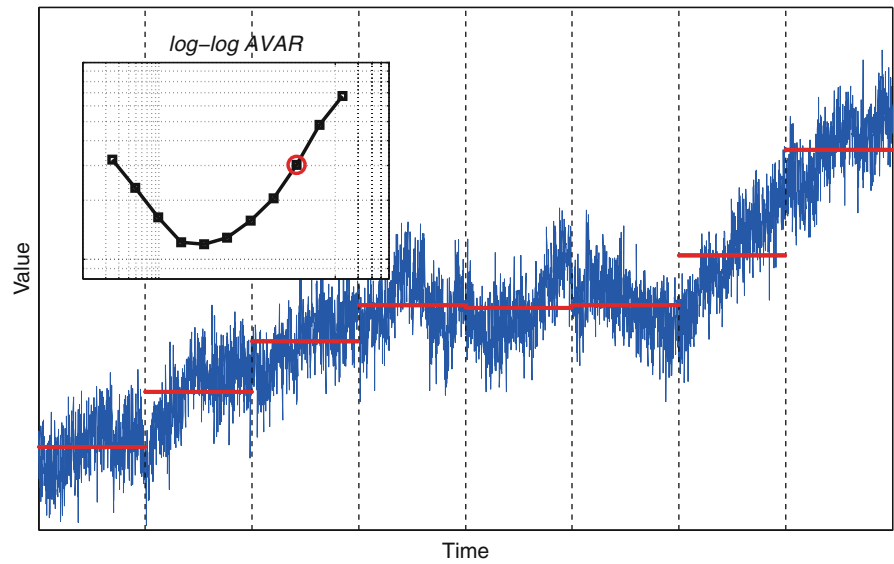
Gyroscopes, Table 3 Classification of random noise modes as applied to rate gyroscopes

Spectral type	Example sources	$PSD(f)$ power law f^{α}	$\sigma(\tau)$ power law τ^{β}	Averaging over time	Associated gyroscope parameters
White	Johnson–Nyquist thermal noise	f^0	$\tau^{-1/2}$	Good	Rate resolution in deg/s/√Hz or deg/h/√Hz
					Angle random walk (ARW) in deg/√h
Pink	Electronics flicker	f^{-1}	τ^0	Neutral	Flicker, bias instability in deg/s or deg/h
Red (Brown)	White noise accumulation	f^{-2}	$\tau^{+1/2}$	Bad	Rate Random Walk (RRW) in deg/s*√Hz or deg/h*√Hz

quantified in terms of its power spectral density with deg/s/√Hz or deg/h/√Hz units. The value is often referred to as the *rate resolution* (same units) or *Angle Random Walk* (converted to deg/√h).

Random processes with $1/f$ *PSD* are naturally the next class of noise to consider. *Flicker noise*, commonly encountered in electronic circuits, falls into this category. Figure 10 illustrates time, frequency, and Allan deviation domain properties of $1/f$, or pink noise using a computer simulation. A pink noise can be obtained by low-pass filtering a white noise with a -3 dB per octave filter. Pink noise can be distinguished from a white noise by more visible low-frequency fluctuations. As expected from Eq. 6, Allan deviation has essentially a flat profile in the case of pink $1/f$ noise. In rate gyroscopes, $1/f$ noise often defines the smallest value of the Allan deviation, which is referred to as the *Bias Instability* and has units of deg/s or deg/h.

The third important type of random noise is red noise characterized by $1/f^2$ *PSD*. This type of random process is called a *Random Walk* and occurs whenever a white noise is accumulated over time. While not correct from the color spectral parallel, red $1/f^2$ noise is sometimes referred to as Brown noise, in honor of Robert Brown who first studied Brownian motion. Figure 11 illustrates time, frequency, and Allan deviation domain properties of $1/f^2$ *Random Walk* obtained by numerical integration of a white noise signal. Another way to obtain a random walk process is to filter white noise with a -6 dB per octave filter. Random walk signal shows strong low-frequency fluctuations. The power spectral density of a random walk is given by $PSD(f) \propto f^{\alpha}$ with $\alpha = -2$ and its root Allan deviation profile is governed by $\sigma(\tau) \propto \tau^{\beta}$ with $\beta = +1/2$. In the Allan deviation plot, the rate random walk is identified by fitting the $+1/2$ sloped part of the curve with $y = bx^{+1/2}$ and estimating the value of b.

In rate gyroscopes, $1/f^2$ noise is typically referred to as the *Rate Random Walk* (RRW) and quantified in deg/s*√Hz or deg/h*√Hz.

When a rate gyroscope is used to track or maintain the angular orientation of an object, the rate signal is integrated over time together with the associated white noise of rate, flicker noise of rate, and random walk of rate. Upon integration, the three rate noise sources with $1/f^0$, $1/f^1$, and $1/f^2$ densities produce, respectively, drifts of angle with $1/f^2$, etc., power spectral densities. Each of these errors of angle grows fast with time, leading to degradation of angle accuracy. Angle measuring (rate integrating) micromachined gyroscope can potentially provide much better accuracy and stability since the time integration of the noise is avoided.

Gyroscopes,
Fig. 11 Illustration of the $1/f^2$ ("red" or "Brown") noise properties in time, frequency, and Allan deviation domains using numerical example
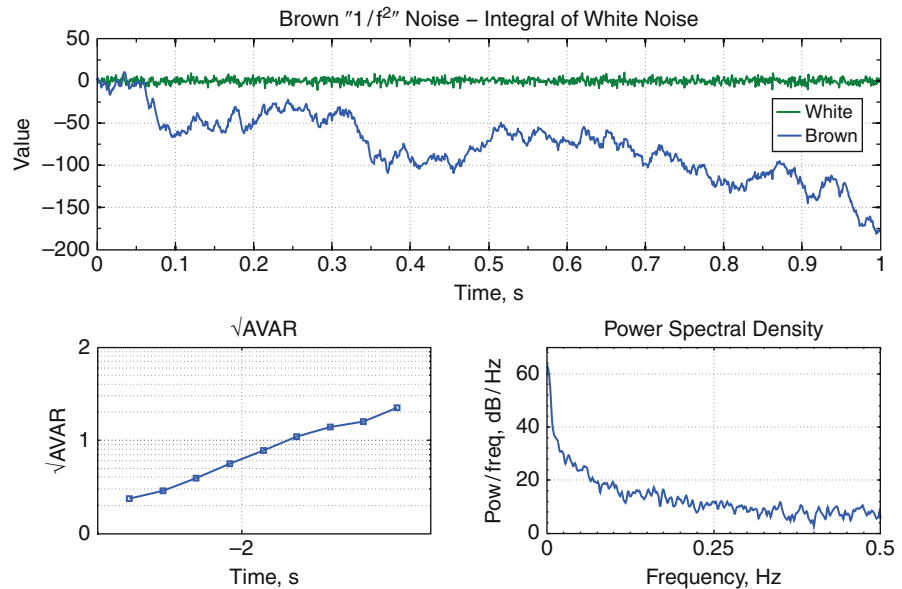
Cross-References

▶ Basic MEMS Actuators
▶ Capacitive MEMS Switches
▶ Laterally Vibrating Piezoelectric Resonators
▶ MEMS Packaging
▶ Piezoelectric MEMS Switch

References

1. Shkel, A.M.: Type I and Type II micromachined vibratory gyroscopes. In: Position, Location, and Navigation Symposium, San Diego, 25–27 Apr 2006, pp. 586–593. IEEE/ION, doi: 10.1109/PLANS.2006.1650648. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1650648/isnumber=34615

2. Nakamura, S.: MEMS inertial sensor toward higher accuracy and multi-axis sensing. In: Sensors, 30 Oct – 3 Nov 2005, pp. 4. IEEE, 2005. doi: 10.1109/ICSENS.2005.1597855. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1597855&isnumber=33608

3. Damrongsak, B., Kraft, M.: A micromachined electrostatically suspended gyroscope with digital force feedback. In: Sensors, 30 Oct–3 Nov 2005, pp. 4. IEEE, CA, 2005. doi: 10.1109/ICSENS.2005.1597720. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1597720&isnumber=33608

4. IEEE standard specification format guide and test procedure for coriolis vibratory gyros. IEEE Std 1431–2004, pp. 0_1-69 (2004). doi: 10.1109/IEEESTD.2004.95744. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1394849&isnumber=30352

5. Barnaby, R.E., Chatterton, J.B., Gerring, F.H.: General theory and operational characteristics of Gyrotron angular rate tachometer. Aeronaut. Eng. Rev. **12**(11), 31–36 (1953)

6. Madni, A.M., Costlow, L.E., Knowles, S.J.: Common design techniques for BEI GyroChip quartz rate sensors for both automotive and aerospace/defense markets. IEEE Sensor. J. **3**(5), 569–578 (2003). doi: 10.1109/JSEN.2003.817728. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1234893&isnumber=27682

7. Acar, C., Schofield, A.R., Trusov, A.A., Costlow, L.E., Shkel, A.M.: Environmentally robust MEMS vibratory gyroscopes for automotive applications. IEEE Sensor. J. **9**(12), 1895–1906 (2009). doi: 10.1109/JSEN.2009.2026466. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5297817&isnumber=5290388

8. Trusov, A.A., Prikhodko, I.P., Zotov, S.A., Schofield A.R., Shkel, A.M.: Ultra-high Q silicon gyroscopes with interchangeable rate and whole angle modes of operation. IEEE Sensors 2010, pp. 864–867, 1–4 Nov 2010. doi: 10.1109/ICSENS.2010.5690867. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5690867&isnumber=5689839

9. Acar, C., Shkel, A.M.: MEMS vibratory gyroscopes: structural approaches to improve robustness. Springer, New York/London (2009). ISBN 0387095357, 9780387095356

10. Zaman, M.F., Sharma, A., Zhili Hao., Ayazi, F.: A mode-matched silicon-yaw tuning-fork gyroscope with subdegree-per-hour Allan deviation bias instability. J. Microelectromech. Syst. **17**(6), 1526–1536, Dec 2008. doi: 10.1109/JMEMS.2008.2004794. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4636724&isnumber= 4681904

11. Sung, W.-K., Dalal, M., Ayazi, F.: A 3MHz spoke gyroscope with wide bandwidth and large dynamic range. In: 2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS), Wanchai,

Hong Kong, 24–28 Jan 2010, pp. 104–107. doi: 10.1109/MEMSYS.2010.5442554. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5442554&isnumber= 5442276

12. Rozelle, D.: The hemispherical resonator gyro: from wineglass to the planets. In: Proceeding 19th AAS/AIAA Space Flight Mechanics Meeting, pp. 1157–1178 (2009)

13. Zotov, S.A., Prikhodko, I.P., Trusov, A.A., Shkel, A.M.: 3-D micromachined spherical shell resonators with integrated electromagnetic and electrostatic transducers. In: Solid-state sensors, actuators, and microsystems workshop 2010, Hilton Head Island, South Carolina, 6–10 Jun 2010

14. Greiff, P., Boxenhorn, B., King, T., Niles, L.: Silicon monolithic micromechanical gyroscope. In: 1991 international conference on solid-state sensors and actuators. Digest of technical papers, TRANSDUCERS '91, San Francisco, 24–27 Jun 1991, pp. 966–968. doi: 10.1109/SENSOR.1991.149051. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=149051&isnumber=3940

15. Yazdi, N., Ayazi, F., Najafi, K.: Micromachined inertial sensors. Proc. IEEE. **86**(8), 1640–1659 (1998). doi:

10.1109/5.704269. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=704269&isnumber=15216

16. Weinberg, M.S., Kourepenis, A.: Error sources in in-plane silicon tuning-fork MEMS gyroscopes. J. Microelectromech. Syst. **15**(3), 479–491 (2006). doi: 10.1109/JMEMS.2006.876779. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1638473&isnumber=34353

17. Geen, J.A., Sherman, S.J., Chang, J.F., Lewis, S.R.: Single-chip surface micromachined integrated gyroscope with 50°/h Allan deviation. IEEE J. Solid-State Circ. **37**(12), 1860–1866 (2002). doi: 10.1109/JSSC.2002.804345. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1088117&isnumber=23646

18. Neul, R., Gomez, U.-M., Kehr, K., Bauer, W., Classen, J., Doring, C., Esch, E., Gotz, S., Hauer, J., Kuhlmann, B., Lang, C., Veith, M., Willig, R.: Micromachined angular rate sensors for automotive applications. IEEE Sensor. J. **7**, 302–309, (2007). doi: 10.1109/JSEN.2006.888610. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4066977&isnumber=4049782