Springer Series in Optical Sciences 191

Jingsong Wei

Nonlinear Super-Resolution Nano-Optics and Applications

Science Press Beijing

Springer

Springer Series in Optical Sciences

Volume 191

Springer Series in Optical Sciences

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief William T. Rhodes, Georgia Institute of Technology, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.
With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.
The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also www.springer.com/series/624

More information about this series at http://www.springer.com/series/624

Jingsong Wei

Nonlinear Super-Resolution Nano-Optics and Applications

Science Press
Beijing

Springer

Jingsong Wei
Shanghai Institute of Optics
and Fine Mechanics
Chinese Academy of Sciences
Shanghai
China

ISSN 0342-4111 ISSN 1556-1534 (electronic)
ISBN 978-3-662-44487-0 ISBN 978-3-662-44488-7 (eBook)
DOI 10.1007/978-3-662-44488-7

Library of Congress Control Number: 2014946407

Springer Heidelberg New York Dordrecht London

Jointly published with Science Press, Beijing
ISBN: 978-7-03-041948-4, Science Press, Beijing

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

With the development of electro-optical technology, nano-Sci and Tech., and life science and biology, the light imaging resolution, lithographic linewidth, and optical information mark size are required to reach down to subwavelength or even nanoscale. However, these are restricted by the Abbe limit due to the diffraction effect. Researchers have proposed a number of methods to fight against the Abbe limit. These methods can be classified into two kinds, one is to change the point spread function, such as scanning near-field optical probe microscopy, liquid (solid) immersion lens, and phase-only pupil filter, etc. The other is to improve the resolution through detecting the fluorescence signal, such as photo-activated localization microscopy (PALM) and stochastic-optical reconstruction microscopy (STORM), which are generally used in biomedical and life science. The stimulated emission depletion (STED) is an excellent combination of two kinds of techniques mentioned above. The STED is mainly used in biomedical imaging and life science, and is also explored to apply to nanolithography in recent years.

This book first introduces the principle and technical schematics of common methods for realizing nanoscale spot (Chap. 1), describes the third-order nonlinear effects and characterization methods (Chaps. 2 and 3), and then analyzes the strong nonlinear characteristics (including nonlinear absorption and refraction) of semiconductor and metal-doped semiconductor thin films (Chap. 4). Chapters 5–7 focus on nonlinearity-induced super-resolution effects, including nonlinear saturation absorption-induced aperture-type super-resolution, nonlinear refraction-induced self-focusing and interference-manipulation super-resolution, and the combination of nonlinear thin films and phase-only pupil filters to compress the side-lobe intensity and reduce the main spot size to nanoscale. Applications in high-density optical information storage, nanolithography, and high-resolution light imaging are presented in Chaps. 8 and 9, and some remarks are given at the end of the book.

I hope that this book can drive nano-optics and nanophotonics to continue to advance. The book is helpful for advanced undergraduates, graduate students, and researchers and engineers working in related fields of nonlinear optics, nano-optics and nanophotonics, information storage, laser fabrication, and lithography,

and light imaging etc. It is unavoidable that some errors and incorrectness may occur in this book, I hope that the readers can point them out. I also will further correct them and improve my work on future releases.

The work in this book is partially supported by the National Natural Science Foundation of China (Grant Nos. 51172253, 61137002, and 60977004). Here please allow me to express my appreciations to my family, Prof. Fuxi Gan, and Prof. Chenqing Gu due to their support in my work and life. It is a pleasure to thank my colleagues and students for their help. Last but not least, I am delighted to dedicate this book to my son, Yusen. Yusen is a smart boy, and he brings joy and happiness to our family.

Shanghai, China Jingsong Wei

Contents

Chapter 1
General Methods for Obtaining Nanoscale Light Spot

1.1 Introduction

In 1873, the German physicist Ernst Abbe realized that the resolution of optical imaging instruments is fundamentally limited by the diffraction of light. His finding indicated that ultimately the resolution of an imaging instrument is not constrained by the quality of the instrument, but by the wavelength of light used and the aperture of its optics [1]. That is, the resolution of a conventional optical-lens-based imaging system is restricted by the Abbe diffraction limit to $\sim\lambda/2\mathrm{NA}$, where λ and NA are the free-space wavelength of the imaging radiation and numerical aperture of the lens, respectively. This diffraction-limited phenomenon hindered the performance of optical microscopy for over a century and was considered a fundamental, unbreakable rule. Beating the diffraction limit and realizing nanoscale light spot have become one of the primary research focuses in modern optics [2]. Realizing nanoscale light spot is also always a hot subject due to its important applications in high-density data storage and nanolithography, etc. According to Abbe diffraction limit, to reduce the light spot size, a short-wavelength laser source and a high NA lens need to be used. However, in current far-field optical system, on one hand, it is difficult to shorten the laser wavelength further because of high product cost and low transmission for optical elements. On the other hand, the NA of a commercially available lens is close to the limit. Under such circumstances, numerous methods and techniques are proposed to overcome the Abbe limit. Let us give a brief introduction to the working schematics and principles of general methods for realizing nanoscale light spot.

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_1

1.2 Near-Field Scanning Probe Method

In far-field optical systems, the optical resolving power is generally constrained to half-wavelength, namely half-wavelength limit. This is because the high spatial frequencies of objects are such that the waves become evanescent in the certain direction. If the evanescent waves containing fine structure information can be used to illuminate the object, the resolving power of half-wavelength limit can be overcome. The generation of evanescent wave is key to break through the Abbe limit. In 1972, Ash and Nicholls [3] proposed a concept of the super-resolution aperture scanning microscope, where one can fabricate a small aperture of radius r, $(r/\lambda \ll 1)$ in a thin diaphragm, and the object information of a comparably small area can be obtained only if the object is located very close to the aperture. The super-resolution aperture scanning microscope is actually a primary configuration for generating and obtaining the evanescent wave because the distance between the aperture and object surface is smaller than λ.

In order to observe smaller fine structure of objects, Betzig and Trautman [4] proposed to establish the near-field optical microscopy where the irradiation wavelength is in the visible light. The basic configuration is a near-field pinhole probe, as shown in Fig. 1.1. Figure 1.1a is a schematic of the pinhole probe, where the fiber probe is coated with an Al thin film, and an aperture with a diameter of far smaller than incident light wavelength is at the apex of probe. The incident light passes through the fiber probe, and a very small light dot is formed at the aperture of the apex of the probe. The light dot size is determined by the aperture diameter (also see Fig. 1.1b). The light dot is evanescent wave and decays exponentially along the optical axis. In real applications, the distance between the light dot and object needs to be remained in the near-field range of less than the light wavelength, or else it is difficult to obtain a resolution below the diffraction limit.

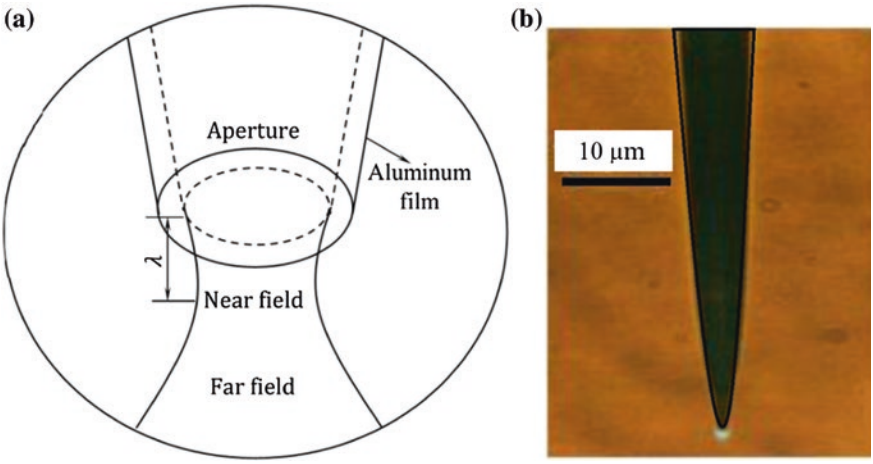

Fig. 1.1 Near-field fiber probe, **a** schematic configuration, and **b** real fiber probe

1.2.1 Aperture-Type Probe

In the configuration of near-field optical microscopy, the design and fabrication of the near-field probe are critical. The aperture-type probe is one of the probes that are most widely used. The aperture-type probe can be fabricated by heating and subsequent pulling method. An alternative method is by wet chemical etching, where bare optical fibers are dipped in an etchant solution to produce a taper with a sharp endpoint. Generally speaking, the probe surface needs to be coated with Al to create an aperture smaller than 100 nm. The advantages are follows: (1) the probe sharpness can be reduced down to 20 nm; (2) the light is easily delivered to the end of the probe; (3) the flat endpoint is beneficial in imaging; (4) the small lateral dimensions of the probe results in a good access to corrugated samples. However, some drawbacks are difficult to overcome, such as the grainy structures on the aluminum coatings resulting in poor optical imaging, highly asymmetric polarization behavior, pinholes at the taper region disturbing the polarization behavior of the tip, and low optical throughput and brightness etc.

In order to improve the probe performance, Veerman et al. [5] developed focused ion beam (FIB) fabrication method. The steps are as follows. (1) The fiber tips are pulled with a fiber puller, and the tip sharpness determines the minimum obtainable aperture size of the finished probe. (2) The Al is deposited on the probes using e-beam evaporation. It is noted that evaporation at an angle of approximately 75° with respect to the probe axis creates an aperture, while evaporation at an angle of 90° covers the entire fiber end. (3) The probes are then put into the FIB machine. In the FIB machine, the Gallium ions are used to remove a very thin slice of material from the tip end, which results in the formation of flattened aperture. The aperture size can be tuned by simply changing the thickness of the slice that is milled. Figure 1.2 shows images of two FIB-etched tips. Figure 1.2a is a flat-end face and well-defined circular aperture with a diameter of 120(±5) nm, a much smaller aperture of about 35(±5) nm is fabricated by tuning the slice thickness in Fig. 1.2b.

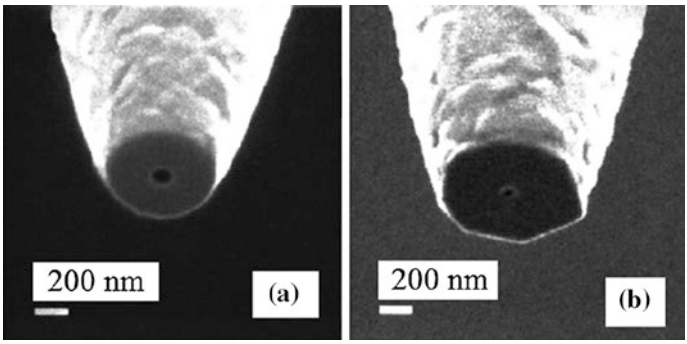

Fig. 1.2 The images of probes fabricated by FIB method, the aperture diameter of **a** 120 nm and **b** 35 nm. Reprinted with permission from [5]. Copyright 1998, American Institute of Physics

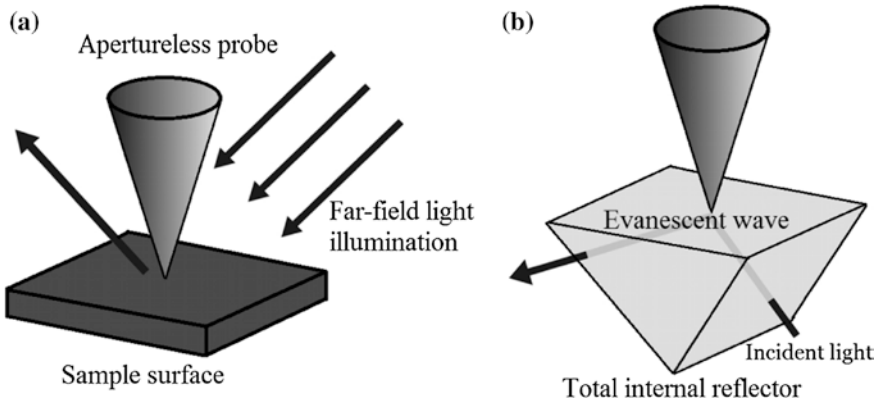

Fig. 1.3 Configurations of apertureless-type metal probe, the sample is illuminated **a** by far-field light, and **b** by an evanescent wave

1.2.2 Apertureless-Type Metal Probe [6]

An alternative to design the probe is a scattering-type tip operated in the apertureless configuration, where a sharp tip is illuminated by far-field light and scatters the light field locally at tip apex. The tip apex actually acts as a local source, as shown in Fig. 1.3a, where the relevant near-field light signal needs to be extracted from a large background of unwanted light from the sample surface.

One can decrease the stray light by using lock-in detection method, where the tip is vibrated at certain frequency, and the detected near-field light signal is demodulated with a lock-in scheme. Another method to reduce the unwanted scattered light is using evanescent wave illumination by total internal reflection technique, as shown in Fig. 1.3b, where the evanescent wave is generated over the sample surface and scattered by the probe, and the scattered light is collected by external optics. Apertureless-type metal probe can obtain much better resolution of below 10 nm by concentrating light field close to the tip apex, and the resolution is determined by the apex radius.

1.2.3 Tip-on-Aperture-Type Probe

For apertureless-type metal probes, the far-field illumination by a focused laser beam exposes a large area around the tip apex, which causes some problems, such as the generation of background scattering signal and large area bleaching for fluorescence imaging. In order to solve these problems, Frey et al. [7] designed and fabricated tip-on-aperture-type probe, where the light fields can be optimally concentrated at tip apex by fully exciting localized surface plasmons or antenna resonance effect. The tip-on-aperture-type probe combines the advantages of

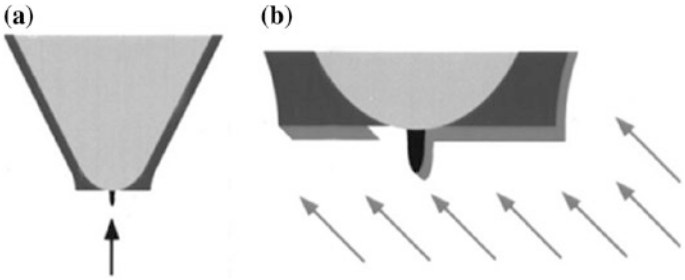

Fig. 1.4 Schematic of fabricating process of tip-on-aperture-type probe. **a** An electron-beam-deposited tip is grown on the aperture of a metallized glass fiber probe in a scanning electronic microscopy, **b** one-sided deposition of Al metalizes the tip and forms an elongated and reduced aperture in tip shadow. Reprinted with permission from [7]. Copyright 2002, American Institute of Physics

aperture and apertureless probes and achieves low background noise and high light throughput as well as topographical resolution. The fabrication process of tip-on-aperture-type probe is as follows.

(1) An optical single mode fiber is thinned in an etching solution.
(2) The fiber is etched in a fourfold diluted solution to reduce the diameter to 15 μm, which makes a 2 μm length sharp tip be obtained because the fiber core is etched more slowly than the cladding.
(3) The tip is covered with Cr adhesion layer with several nanometers and Au layer of about 200 nm due to good contrast in the scanning electron microscope.
(4) An aperture can be generated by pressing the tip on a glass surface and monitoring the far-field light throughput.
(5) The tip is obtained by focusing the electron beam to the center of the aperture for several seconds at 8 kV acceleration voltage, also see Fig. 1.4a.
(6) Cr thin film of 3.5 nm and A1 thin film of 33 nm are then sequencially deposited on the probe surface by evaporation technique at 45° incidence angle (also see Fig. 1.4b).

In this way, the electron-beam-deposited tip is metalized on one side, and the original aperture is reduced to a small and elongated aperture left by the tip's shadow, resulting in an asymmetric tip–aperture arrangement. The position of the tip coincides with one edge of the new aperture.

1.2.4 C-Aperture Encircled by Surface Corrugations on a Metal Film [8]

As reported, the light throughput of the 60-nm circular aperture follows an exponential decay with the distance from aperture due to evanescent wave effect.

However, the light throughput of C-aperture is decreased quadratically, which may be because the propagating wave contributes an improvement and delivers the energy into several hundred nanometers. The light throughput can be further increased by the C-aperture encircled by surface corrugations on a metal film where surface plasmon resonance is excited. It is well known that surface plasmon resonance depends on the momentum and energy matching conditions among the incident photon, grating vector, and surface plasmons. A surface corrugation (such as groove) corresponds to lots of grating vector components. Thus, the surface plasmon resonance can be excited as long as some of grating vector components fulfill the matching condition and the light throughput can be enhanced accordingly.

In principle, the energy of surface plasmon resonance is from two parts: one is propagating components parallel to the metal surface, and the other is an exponentially attenuated component perpendicular to the metal surface. Thus, the field distribution of the surface plasmon resonance is extended in the x–y plane but exponentially decayed in the z-direction. However, for C-aperture with and without the surface corrugations, the propagating component is toward the z-direction, which causes larger light throughput than circular aperture.

For C-aperture encircled by surface corrugations on a metal film, the light throughput is further enhanced compared with the C-aperture, which suggests that excited surface plasmon resonance coexists with propagating waves, i.e., the so-called hybrid effect. In the hybrid effect, the incident light can support propagating waves and generate the surface plasmon resonance, which results in constructive interference concomitantly. Chen et al. fabricated circular aperture, single C-aperture, and groove-encircled C-aperture by a FIB (as schematically shown in Fig. 1.5). They also calculated the hybrid effect and found that light throughput contributed by the hybrid effect can reach 6.79, a factor of 3.88 times larger than that obtained with the single C-aperture and 10^5 times larger than that obtained with a 60-nm circular aperture. These results were also verified by near-field scanning optical microscope.

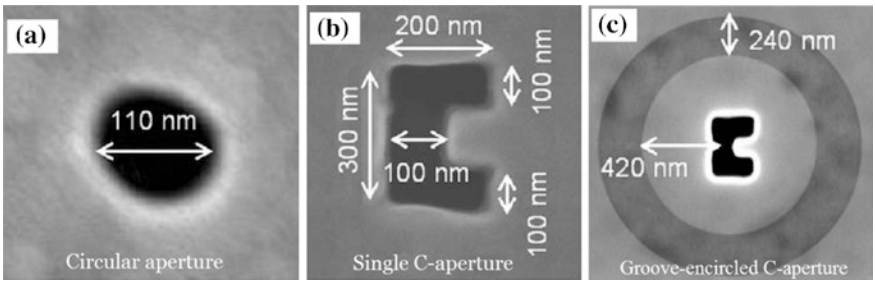

Fig. 1.5 Fabricated **a** circular aperture, **b** C-aperture, and **c** groove-encircled C-aperture. Reprinted with permission from [8]. Copyright 2006, Optical Society of America

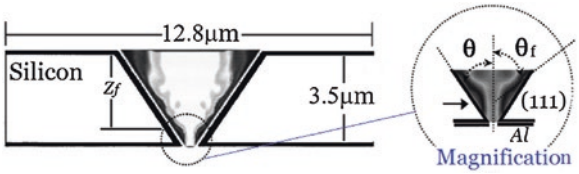

Fig. 1.6 Si-based nonlinear self-focusing probe by using As_2S_3 as nonliear materials. (The pyramidal waveguide is filled with As_2S_3, and the magnification image presents a parabolic beam shape caused by the self-focusing effect). Reprinted with permission from [9]. Copyright 2002, American Institute of Physics

1.2.5 Nonlinear Self-focusing Probe

A good method to increase near-field intensity at the probe aperture is using a type of material with a very high third-order nonlinear refraction coefficient and small linear and nonlinear optical losses. Figure 1.6 gives the basic concept of the Si-based nonlinear self-focusing aperture, where a parabolic shape beam with large cone angle near the focus is developed according to the self-focusing effect. Optical image is simulated by a theoretical calculation through the modified non-paraxial model at 1.3 mW laser power. The self-focusing angle θ_f, defined as the refractive angle of the self-focusing effect, is dependent on the incident beam power. If θ_f is larger than the θ, where θ is the angle of the Si (111) plane boundary, then the self-focusing distance z_f depends on the incident light power. The optical loss region is minimized until the beam size becomes in the order of half-wavelength, and the local effective near-field excitation can be achieved.

For As_2S_3, the self-focusing beam size could be in the order of half-wavelength due to the large nonlinear refraction coefficient at the wavelength in the Urbach tail region. Song et al. [9] took the As_2S_3 as the nonlinear material and fabricated nonlinear self-focusing probe on the basis of Si microfabrication technique. This is a very simple fabrication process without any geometrical change near the aperture. The experimental results indicated that the light throughput of the near-field aperture with an aperture diameter of 100 nm increases over 100 times compared with conventional tapered fiber tips.

1.3 Solid Immersion Lens Method [10]

Solid immersion lens (SIL) is generally formed by grinding and fabricating the transparent materials with high refractive index n. According to shape characteristics, there is two type of SIL. One is semi-sphere-type SIL, and the other is super semi-sphere SIL. Figure 1.7 presents the working principles. In Fig. 1.7a, the focusing illustration of ordinary lens is given, where a collimated light beam passes a lens and is focused into a diffraction-limited spot. The spot diameter is about

$$D \sim \lambda/(n_0 \sin\theta_1) \tag{1.1}$$

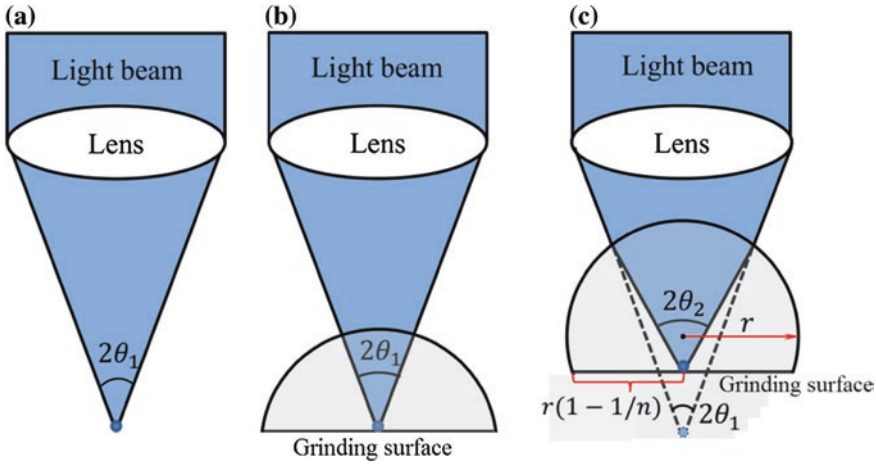

Fig. 1.7 Principle of solid immersion lens, focusing **a** with ordinary lens, **b** semi-sphere-type SIL, and **c** super semi-sphere-type SIL

where $n_0 \sin \theta_1 = \text{NA}$ is the numerical aperture of lens, n_0 is refractive index of medium, and θ_1 is maximum aperture half-angle of lens. λ is light wavelength.

Figure 1.7b is a semi-sphere-type SIL. The light beam goes through the ordinary lens and is focused onto the grinding surface of semi-sphere-type SIL. The light wavelength inside the SIL is changed into $\lambda_s = \lambda/n$; thus, the spot diameter onto the grinding surface of semi-sphere-type SIL is

$$D_1 \sim \frac{\lambda_s}{n_0 \sin \theta_1} = \frac{D}{n} \tag{1.2}$$

Formula (1.2) indicates that the spot size is reduced down to $1/n$ times of ordinary lens.

Figure 1.7c is a super semi-sphere SIL, where the radius of the grinding surface is $r(1 - 1/n)$, and r is the radius of sphere. The collimated light beam passes through the ordinary lens and is refocused onto the grinding surface by the super semi-sphere-type SIL. The focal point is also called as aplanatic foci. In super semi-sphere-type SIL, the light wavelength is $\lambda_s = \lambda/n$. The maximum aperture half-angle is θ_2. Based on geometrical optics, one has $\sin \theta_2 = n \sin \theta_1$. Thus,

$$D_2 \sim \frac{\lambda_s}{n_0 \sin \theta_2} = \frac{\lambda}{n \cdot n \cdot n_0 \sin \theta_1} = \frac{D}{n^2} \tag{1.3}$$

Formula (1.3) tells us that the spot size is reduced down to $1/n^2$ times of ordinary lens. The reduction of spot size is very useful for improving the capacity of optical memory. If the area density of optical memory is set to be Ω, then

$$\Omega \propto \frac{1}{D^2} \Rightarrow \Omega \propto n^4 \tag{1.4}$$

If one takes $n = 2.5$ of super semi-sphere SIL as an example, compared with ordinary optical head (lens), the areal density is increased up to about 40 times.

1.4 Surface Plasmonic Lens

Surface plasmonic lens can be designed into different types. Here, concentric ring grating-type plasmonic lens is introduced as an example. Figure 1.8a schematically shows two-dimensional array of plasmonic lenses (4×2) fabricated in a square lattice. The plasmonic lens consists of a nanoaperture surrounded by 15 through-rings on an aluminum film, where the aperture diameter, ring periodicity, ring width, and aluminum layer thickness are 100, 250, 50, and 80 nm, respectively. The plasmonic lens can obviously make the light become strong. For example, if the light with a wavelength of 365 nm is incident onto the plasmonic lens, the numerical simulation results show an intensity enhancement factor of 100 times and a focused spot of 80 nm. The spot size is obviously smaller than the incident light wavelength, which indicates that the plasmonic lens can improve the point spread function (PSF) of imaging or lithography system.

In order to generate high-throughput nanopatterning lithography using plasmonic lens, Srituravanich et al. [11] adopted high-speed flying plasmonic lens arrays, and the sample is placed on an air-bearing rotation stage, and the rotation speed can reach up to 4–12 m/s. The plasmonic lens array enables parallel writing with high throughput. Figure 1.8b presents the schematic of high-throughput maskless nanolithography using plasmonic lens array. The plasmonic lens array is mounted on the flying head. The ultraviolet laser ($\lambda = 365\,\text{nm}$) is incident on the plasmonic lens array and generates surface plasmons. The plasmonic lens then concentrate surface plasmons into sub-100-nm spot in the near-field of the lens. The sub-100-nm spot is directly coupled into the resist thin films to conduct nanolithography in the near-field range. The near-field range is generally maintained at 20 nm between the plasmonic lens and the resist thin films. The smallest lithographic line width obtained by the plasmonic lens can reach down to about 80 nm.

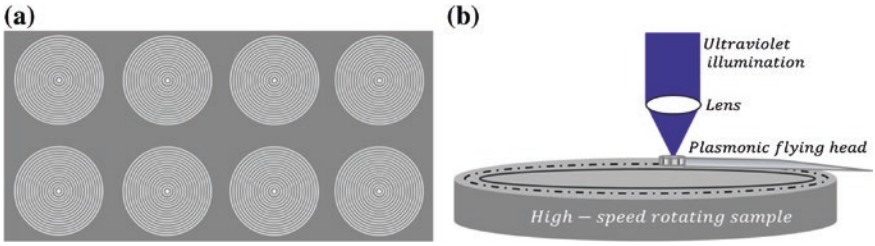

Fig. 1.8 Schematics of surface plasmonic lens and lithography, **a** an array of plasmonic lenses, and **b** schematics of high-throughput maskless nanolithography using plasmonic lens array head

1.5 Stimulated Emission Depletion Fluorescence Microscope Methods

Biologists need to have fine information images of cellular processes. Unfortunately, due to the Abbe limit, two objects become indistinguishable if the distance between them is close to the half-wavelength. One can improve the resolution using short-wavelength laser or electron beam. The problem with the approaches is that the higher energy associated with short-wavelength photon and electron tends to require a vacuum and kill cells, which cannot be applied for observing dynamic process or living cells. The microscope using laser-induced fluorescence as light source is a good tool for resolving objects about 200 nm apart. Usually, one calls the microscope as biological fluorescence microscope, because the fluorescence labeling is used here, which is different from the conventional light imaging techniques.

In 1994, Hell and Wichmann [12] proposed stimulated emission depletion (STED) microscope. Compared with other super-resolution fluorescence microscopy techniques, its unique feature is that it is a purely optical method that does not require any mathematical manipulation or image processing.

The basic idea of STED microscopy is to improve the PSF (Point Spread Function). In addition to the conventional laser beam that excites the fluorophores, a donut-shaped depletion laser beam is used to de-excite the peripheral fluorescence through stimulated emission, and thus, only fluorescence emission from the sub-diffraction-limited center is left out to be recorded. In other words, the resolution enhancement is essentially based on switching off the fluorescence molecules of stimulated emission using intense laser light in the outer region of the diffraction-limited excitation spot. This intense irradiation causes almost all of the excited molecules to return to the ground state, leaving only the region which is close to the center of the excitation spot. Fluorescence from the remaining excited dye molecules is then detected and used to form the high-resolution images.

Figure 1.9 presents the schematic of the basic principle of STED microscope. In Fig. 1.9a, a short pulse laser beam spot, called as stimulation light with a wavelength of 405 nm, irradiates onto the fluorescence labeling materials. For the fluorescence molecules in the spot, the electrons in S_0 band of ground state are excited to S_1^* band of the first excitation state. The electrons quickly jump to the lowest energy band S_1 of the first excitation state in the form of vibrational relaxation. Another hollow circular laser beam spot (called as STED light) with a wavelength of 635 nm, which has a long wavelength compared with the stimulation light, depletes excited state fluorescence molecules in the marginal area of stimulation light spot. Therefore, the electrons in the S_1 state return to the S_0^* band of ground state and then vibrationally relaxes to the S_0 band of the ground state, as shown in Fig. 1.9b. The resulting spots are schematically given in Fig. 1.9c, where the blue is stimulation spot, and the red is hollow circular STED spot. The combination of stimulation spot and STED spot causes the formation of a very small fluorescence spot marked in green.

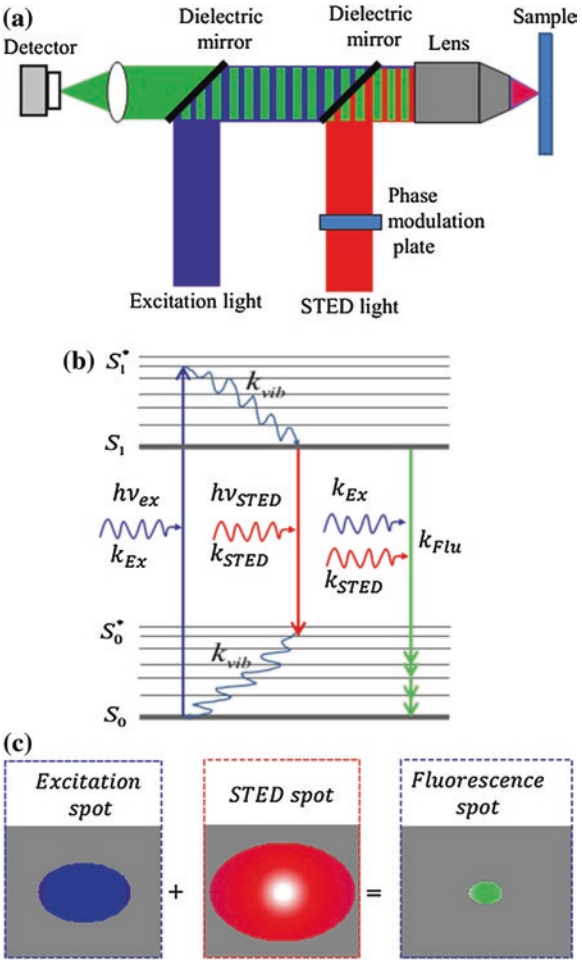

Fig. 1.9 Schematic of the STED principle, **a** optical setup, **b** energy band structure of fluorescence, and **c** PSF of the SETD microscope

For STED microscope, the small green fluorescence spot is actually PSF. One can use the green fluorescence spot to scan the sample and generate high resolving image. Compared with the stimulation and STED light spots, the green fluorescence spot is greatly reduced, and thus, the PSF is improved. The laternal resolution size of the green fluorescence spot can be calculated as

$$\Delta x = \frac{\lambda}{2\mathrm{NA}\sqrt{1 + I_{\max}/I_{\mathrm{sat}}}} \tag{1.5}$$

where NA is numerical aperture of imaging lens, $I_{\max}$ is the maximum focusing intensity of STED light, and I_{sat} is the saturation intensity. One can define the saturation factor $\xi = I_{\max}/I_{\mathrm{sat}}$. Formula (1.5) indicates that $\Delta x \to 0$ at $\xi \to \infty$; thus, STED resolution is theoretically unlimited by indefinitely increasing the depletion laser power. In order to improve the resolution, one can also reduce the I_{sat}.

The I_{sat} is inversely proportional to lifetime τ_{flu} and cross section of stimulation emission σ_{flu} of fluorescence molecules. Therefore, one can use the fluorescence labeling materials with long lifetime and large cross section of stimulation emission.

References

1. Beyond the diffraction limit, Nat. Photonics 3, 361 (2009). doi:10.1038/nphoton.2009.100
2. D.R. Mason, M.V. Jouravlev, K.S. Kim, Enhanced resolution beyond the Abbe diffraction limit with wavelength-scale solid immersion lenses. Opt. Lett. **35**(12), 2007–2009 (2010)
3. E.A. Ash, G. Nicholls, Super-resolution aperture scanning microscope. Nature **237**, 510–512 (1972)
4. E. Betzig, J.K. Trautman, Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit. Science **257**, 189–195 (1992)
5. J.A. Veerman, A.M. Otter, L. Kuipers, N.F. van Hulst, High definition aperture probes for near-field optical microscopy fabricated by focused ion beam milling. Appl. Phys. Lett. **72**, 3115–3117 (1998)
6. L. Aeschimann, *Apertureless Scanning Near-Field Optical Microscope Probe for Transmission Mode Operation*, a Dissertation at the Institute of Microtechnology, University of Neuchatel, Switzerland, 2004
7. H.G. Frey, F. Keilmann, A. Kriele, R. Guckenberger, Enhancing the resolution of scanning near-field optical microscopy by a metal tip grown on an aperture probe. Appl. Phys. Lett. **81**, 5030–5032 (2002)
8. Y. Chen, J. Fang, C. Tien, H.D. Shieh, High-transmission hybrid-effect-assisted nanoaperture. Opt. Lett. **31**, 655–657 (2006)
9. K. Song, J. Kim, K. Park, Technique to enhance the throughput on a near-field aperture by the use of self-focusing effect. Appl. Phys. Lett. **80**, 2827–2829 (2002)
10. J. Wei, Super-resolution optical effects of nanoscale nonlinear thin film structure and ultrahigh density information storage, in *Toward Functional Nanomaterials*, ed. by Z.M. Wang (Springer, 2009), pp. 257–283
11. W. Srituravanich, L. Pan, Y. Wang, C. Sun, D.B. Bogy, X. Zhang, Flying plasmonic lens in the near field for high-speed nanolithography. Nat. Nanotech. **3**, 733–737 (2008)
12. S.W. Hell, J. Wichmann, Breaking the diffraction resolution limit by stimulated emission. Opt. Lett. **19**(11), 780–782 (1994)

Chapter 2
Third-Order Nonlinear Effects

2.1 Introduction

The book focuses on utilizing the third-order nonlinear effects (including nonlinear absorption and refraction) to break the diffraction limit and form super-resolution nanoscale spot. In this chapter, let us give a brief introduction to the characteristics of the third-order effects.

When a light beam with a frequency of ω is incident on the isotropic nonlinear medium, the nonlinear effect occurs, and the second-order nonlinear susceptibility $\chi^{(2)}$ can be neglectable. The whole polarization is expressed as [1]

$$P[E(\omega)] = P^{(1)} + P^{(3)} = \epsilon_0 \left[\chi^{(1)} + 3\chi^{(3)} |E(\omega)|^2 \right] E(\omega) \tag{2.1}$$

where $P^{(1)}$ and $P^{(3)}$ the linear and third-order nonlinear polarization. $\chi^{(1)}$ and $\chi^{(3)}$ are the linear and third-order nonlinear susceptibility, respectively. The polarizations are marked into the real and imaginary parts as follows

$$\begin{cases} P^{(1)} = P_R^{(1)} + iP_I^{(1)} \\ P^{(3)} = P_R^{(3)} + iP_I^{(3)} \end{cases} \tag{2.2}$$

The $\chi^{(1)}$ and $\chi^{(3)}$ can be also marked into real and imaginary parts as follows

$$\begin{cases} \chi^{(1)} = \chi_R^{(1)} + i\chi_I^{(1)} \\ \chi^{(3)} = \chi_R^{(3)} + i\chi_I^{(3)} \end{cases} \tag{2.3}$$

where $\chi_R^{(1)}$ and $\chi_R^{(3)}$ are real part and directly related with the refraction. $\chi_I^{(1)}$ and $\chi_I^{(3)}$ are imaginary parts and related with absorption.

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_2

2.2 Nonlinear Refraction

For isotropic (homogeneous) materials, the nonlinear refraction can be considered to be from a four-wave interaction and the third-order nonlinearity is mainly field dependent. Accordingly, the nonlinear polarization is written as $P^{(3)}[E(\omega)] = 3\epsilon_0 \chi_R^{(3)} |E(\omega)|^2 E(\omega)$. Assuming that the refractive index is measured using a single laser beam method, such as single beam z-scan technique, the four-order and higher terms can be neglectable. Then, the total polarization of the material system is

$$P_R[E(\omega)] = P_R^{(1)} + P_R^{(3)} = \epsilon_0 \left[\chi_R^{(1)} + 3\chi_R^{(3)} |E(\omega)|^2 \right] E(\omega) \equiv \epsilon_0 \chi_R^{\text{eff}} E(\omega) \quad (2.4)$$

where $\chi_R^{(1)}$ is the real part of linear susceptibility. Here, one defines the effective susceptibility as $\chi^{\text{eff}} = \chi_R^{\text{eff}} + i\chi_I^{\text{eff}}$, where the χ_R^{eff} and χ_I^{eff} are the real and imaginary parts of effective susceptibility, respectively. According to formula (2.4), the χ_R^{eff} is written as

$$\chi_R^{\text{eff}} = \chi_R^{(1)} + \chi_{eR}^{(3)} |E(\omega)|^2 \quad (2.5)$$

with

$$\chi_{eR}^{(3)} |E(\omega)|^2 = 3\chi_R^{(3)} |E(\omega)|^2 \quad (2.6)$$

If all of the nonlinear susceptibility is thought to be effective susceptibility, one can obtain

$$\chi_{eR}^{(3)} |E(\omega)|^2 \sim \chi_R^{(3)} |E(\omega)|^2 \quad (2.7)$$

The electric displacement vector is

$$D = \epsilon_0 E + P = \epsilon_0 \left[1 + \chi_R^{(1)} + \chi_R^{(3)} |E(\omega)|^2 \right] E(\omega) \quad (2.8)$$

Based on $D = \epsilon E$, the total permittivity is

$$\epsilon = \epsilon_0 \left[1 + \chi_R^{(1)} + \chi_R^{(3)} |E(\omega)|^2 \right] \quad (2.9)$$

The relative permittivity is

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} = 1 + \chi_R^{(1)} + \chi_R^{(3)} |E(\omega)|^2 \quad (2.10)$$

The refractive index is defined as

$$n = \sqrt{\epsilon_r \mu_r} \quad (2.11)$$

where μ_r is relative permeability and generally equals to 1 for non-magnetic materials. Thus

$$n = \sqrt{\epsilon_r} \tag{2.12}$$

Substituting formula (2.10) into formula (2.12), one has

$$n = \sqrt{1 + \chi_R^{(1)} + \chi_R^{(3)}|E(\omega)|^2} \tag{2.13}$$

The linear refractive index is defined as

$$n_0^2 = 1 + \chi_R^{(1)} \tag{2.14}$$

Thus, formula (2.13) is rewritten as

$$n = n_0 \left[1 + \frac{\chi_R^{(3)}|E(\omega)|^2}{n_0^2}\right]^{\frac{1}{2}} \tag{2.15}$$

Generally, the $\chi_R^{(3)}|E(\omega)|^2/n_0^2$ term is much smaller than 1. Using the Taylor series expansion, the formula (2.15) can be expressed as

$$n \approx n_0 + \frac{\chi_R^{(3)}}{2n_0}|E(\omega)|^2 \tag{2.16}$$

Generally, the refractive index can also be written as

$$n = n_0 + \Delta n = n_0 + \gamma I \tag{2.17}$$

where Δn is the third-order nonlinear susceptibility-induced refractive index change and γ is nonlinear refraction coefficient. I is light intensity and can be defined as

$$I = \frac{v\epsilon_r\epsilon_0}{2}\langle\tilde{E}(\omega,t)^2\rangle \tag{2.18}$$

where $v = c/n_0$ is light speed in the medium. $\tilde{E}(\omega,t)$ is time-dependent optical field, and $\epsilon_r \sim n_0^2$, thus

$$I \approx \frac{1}{2}\epsilon_0 c n_0 \langle\tilde{E}(\omega,t)^2\rangle \tag{2.19}$$

If the optical field is of the form [2]

$$\tilde{E}(\omega,t) = E(\omega)\exp(-i\omega t) + c.c., \tag{2.20}$$

one has

$$\langle\tilde{E}(\omega,t)^2\rangle = 2E(\omega)E^*(\omega) = 2|E(\omega)|^2 \tag{2.21}$$

According to formulas (2.19) and (2.21), one has

Table 2.1 Typical values of γ and response time for linearly polarized light [2]

Mechanism	$\gamma(m^2/W)$	Response time (s)
Electronic polarization	$\sim 10^{-20}$	$\sim 10^{-15}$
Molecular orientation	$\sim 10^{-18}$	$\sim 10^{-12}$
Electrostriction	$\sim 10^{-18}$	$\sim 10^{-9}$
Saturated atomic absorption	$\sim 10^{-14}$	$\sim 10^{-8}$
Thermal effects	$\sim 10^{-10}$	$\sim 10^{-3}$
Photorefractive effects[a]	Large	Intensity dependent

[a] The photorefractive effect can cause large nonlinear effect, which cannot usually be described with nonlinear susceptibility $\chi^{(3)}$ (or γ). The nonlinear polarization process of the photorefractive effect is not the same as the other physical mechanisms listed

$$|E(\omega)|^2 = \frac{1}{\epsilon_0 c n_0} I \tag{2.22}$$

Substituting formula (2.22) into formula (2.16), one obtains

$$n \approx n_0 + \frac{\chi_R^{(3)}}{2\epsilon_0 c n_0^2} I \tag{2.23}$$

Based on formulas (2.17) and (2.23), the nonlinear refraction coefficient is written as

$$\gamma = \frac{\chi_R^{(3)}}{2\epsilon_0 c n_0^2} \tag{2.24}$$

According to formula (2.24), one can obtain γ value. The different physical mechanisms can cause different γ values. The typical γ value and response time are listed in Table 2.1. In formula (2.24), $\gamma < 0$ corresponds to self-defocusing effect and $\gamma > 0$ can lead to self-focusing effect. The self-focusing effect can generate the nanoscale spot, which is very useful for nanolithography and high-resolving light imaging, etc.

2.3 Nonlinear Absorption

Similar to the real part of polarization, the imaginary part of polarization is

$$P_I[E(\omega)] = P_I^{(1)} + P_I^{(3)} = i\epsilon_0 \left[\chi_I^{(1)} + \chi_I^{(3)} |E(\omega)|^2\right] E(\omega) \tag{2.25}$$

For isotropic medium, on the basis of the *slowly varying-envelope approximation*, one has

$$\frac{\mathrm{d}E}{\mathrm{d}z} = \frac{i\omega}{2\epsilon_0 c n_0} P_I \tag{2.26}$$

Substituting formulas (2.22) and (2.25) into formula (2.26), one has

$$\frac{\mathrm{d}E}{\mathrm{d}z} = -\left(\frac{\omega}{2cn_0}\chi_I^{(1)} + \frac{\omega}{2\epsilon_0 c^2 n_0^2}\chi_I^{(3)} I\right)E \tag{2.27}$$

One defines the following

$$\frac{\alpha}{2} = \frac{\alpha_0}{2} + \frac{\Delta\alpha}{2} = \frac{\omega}{2cn_0}\chi_I^{(1)} + \frac{\omega}{2\epsilon_0 c^2 n_0^2}\chi_I^{(3)} I \tag{2.28}$$

From the formula (2.28), the formula (2.27) is rewritten as

$$\frac{\mathrm{d}E}{\mathrm{d}z} = -\frac{\alpha}{2}E \tag{2.29}$$

So

$$E(L) = E_0 \exp\left(-\frac{\alpha}{2}L\right) \tag{2.30}$$

where L is medium thickness. The intensity decaying accordingly is

$$I(L) = I_0 \mathrm{e}^{-\alpha L} \tag{2.31}$$

Thus,

$$\begin{cases} \alpha_0 = \frac{\omega}{cn_0}\chi_I^{(1)} \\ \Delta\alpha = \frac{\omega}{\epsilon_0 c^2 n_0^2}\chi_I^{(3)} I \end{cases} \tag{2.32}$$

Thus, α_0 is defined as linear absorption coefficient, and $\Delta\alpha$ is nonlinearity-induced intensity-dependent absorption coefficient change and is

$$\Delta\alpha = \beta I \tag{2.33}$$

where β is nonlinear absorption coefficient and is

$$\beta = \frac{\omega}{\epsilon_0 c^2 n_0^2}\chi_I^{(3)} \tag{2.34}$$

For some materials, such as semiconductors, $\beta < 0$ is saturation absorption, and $\beta > 0$ means reverse saturation absorption or multi-photon absorption.

As described in formula (2.31), the light goes through the nonlinear materials, and the intensity changes with the traveling distance z inside the materials. The intensity decaying process can be rewritten as

$$\frac{\partial I}{\partial z} = -\alpha_0 I - \beta I^2 \tag{2.35}$$

The exiting light intensity from the materials is [3]

$$I_e = \frac{I_0 \exp(-\alpha_0 L)}{1 + \beta I_0 L_{\mathrm{eff}}} \tag{2.36}$$

with

$$L_{\text{eff}} = \frac{1 - \exp(-\alpha_0 L)}{\alpha_0}$$

For $\alpha_0 \to 0$, the linear absorption of materials is neglectable, and the third-order nonlinear absorption is dominant. In this case, $\exp(-\alpha_0 L) \sim (1 - \alpha_0 L)$, thus $L_{\text{eff}} \sim L$. Based on formula (2.36), the exiting light intensity becomes

$$I_e \sim \frac{I_0}{1 + \beta I_0 L} \tag{2.38}$$

An alternative way to obtain the exiting light intensity is through the formula (2.35) with $\alpha_0 = 0$, which yields

$$\frac{\partial I}{\partial z} = -\beta I^2 \tag{2.39}$$

Formula (2.38) can be solved by the separation of variables, which yields

$$I(z) = \frac{I_0}{1 + \beta I_0 z} \tag{2.40}$$

Formula (2.40) is consistent with formula (2.38) and can be used to calculate the light spot intensity profile after passing through the nonlinear absorption materials with weak linear absorption characteristics ($\alpha_0 \to 0$).

The nonlinear refraction can induce self-focusing super-resolution effect (for $\gamma > 0$), and the nonlinear absorption can lead to the generation of aperture-type super-resolution effect (for $\beta < 0$), or subwavelength energy absorption region (for $\beta > 0$), which can be applied to the nanolithography and nanoscale optical data storage, etc.

References

1. D. L. Mills, *Nonlinear Optics–Basic Concepts*, (Springer, Berlin, 1998)
2. R. W. Boyd, *Nonlinear Optics*, Academic (2003)
3. M. Sheik-Bahae, A. A. Said, T. H. Wei, D. J. Hagan, E.W. Van Stryland, Sensitive measurement of optical nonlinearities using a single beam. IEEE J. Quantum Electron. **26**, 760–769 (1990)

Chapter 3
Characterization Methods for Nonlinear Absorption and Refraction Coefficients

3.1 Introduction

There are lots of methods for measuring the nonlinear refraction and absorption coefficients of materials. The z-scan is one of the most simplest and sensitive methods and has been well developed since Sheik-Bahae et al. [1, 2] gave the principle and experimental setup, such as time-resolved z-scan [3], beam dimension measurement z-scan [4] two-beam multi-functional z-scan [5, 6]. In this chapter, the basic theory of z-scan method is introduced, and some issues for measurement of semi-transparent samples and thin films are pointed out. The improved z-scan schematics are presented, and the data processing methods are described accordingly.

3.2 Theory and Setup of Basic z-scan Method

3.2.1 Description of Basic Principle

In a basic z-scan setup, a collimated laser beam is focused through a lens, and the sample is scanned along the z-axis (optical axis) through the focus (as shown in Fig. 3.1). If there is nonlinearity in the sample, the light intensity at different z-scan position either increases or decreases depending on the positive or negative sign of the nonlinearity. The recorded z-scan curves are compared with the theoretically determined fitting curves, and the nonlinear index is subsequently extracted.

In z-scan measurement, the nonlinear refraction and absorption coefficients can be determined. For nonlinear refraction coefficient measurement, an aperture is placed in front of the detector to measure the transmitted light intensity. The sample displaying nonlinear refraction acts as a lens with variable focal length as it is moved along the optical axis (z-axis). Assuming that the sample thickness is less than the diffraction length (also named as Rayleigh length z_0) of the incident laser beam, the sample exhibits negligible nonlinear effect when it is far from the focus due to the low light

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_3

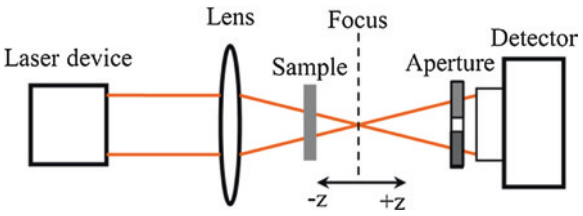

Fig. 3.1 The basic *z*-scan setup

intensity. For positive nonlinear refraction, the sample acts as a positive (or called as convex) lens diverging the light beam and shifting the beam waist as the sample is moved toward the focus, which causes a large spot at the aperture. The effect increases as the sample is moved to the focus because the light intensity increases; thus, a minimum transmittance occurs when the sample is just in front of the focus. The transmittance minimum (valley) moves to a maximum (peak) as the sample is moved further and the light beam becomes collimating as a result of the convex lensing by the sample (as shown in Fig. 3.2). The transmittance will return to the linear value as the sample is moved further far from the focus. For the sample possing a negative nonlinear refraction, the transmittance change is inverted.

However, in most cases, the nonlinear refraction occurs usually in conjunction with nonlinear absorption, which means the data of closed-aperture mode *z*-scan measurement contain both nonlinear refraction and nonlinear absorption. In order to exclude the nonlinear refraction, the aperture in front of the detector shown in Fig. 3.1 ought to be removed, which is called as open-aperture mode *z*-scan. The transmittance curve obtained in open-aperture mode *z*-scan measurement is independent on nonlinear refraction, and only dependent on nonlinear absorption. For positive nonlinear absorption, the curve forms a valley curve that is symmetric around the focus (shown in Fig. 3.3a), and a peak curve that is also symmetric around the focus can be generated for negative nonlinear absorption (shown in Fig. 3.3b). The pure nonlinear refraction curve can usually be obtained through dividing the data of closed-aperture mode (as shown in Fig. 3.3c) by those of open-aperture mode (shown in Fig. 3.3d).

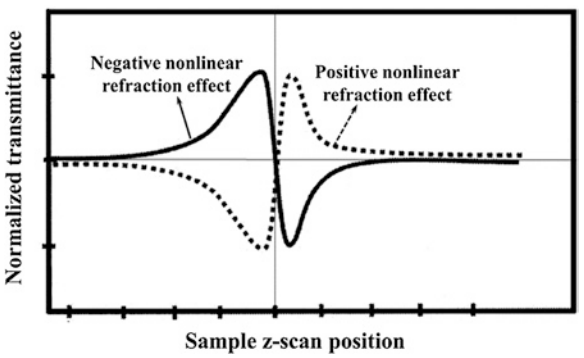

Fig. 3.2 Typical closed-aperture mode *z*-scan curves

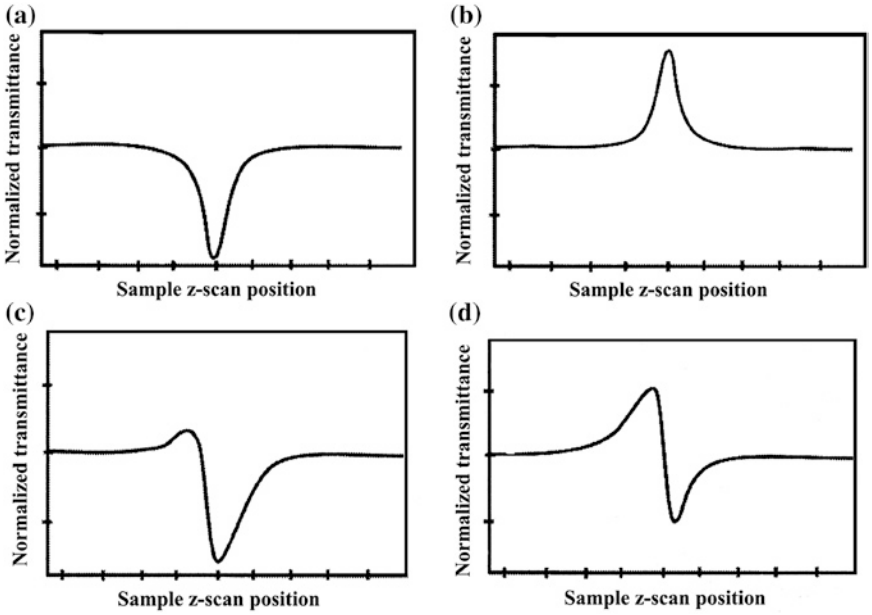

Fig. 3.3 The typical z-scan measurement curves. Open-aperture mode z-scan curve **a** for positive nonlinear absorption and **b** for negative nonlinear absorption; **c** closed-aperture mode z-scan curve, and **d** pure nonlinear refraction curve obtained through dividing of **c** by **a**

3.2.2 Data Analysis for z-scan Curves [1]

3.2.2.1 Nonlinear Refraction

In z-scan experiment, a TEM_{00} mode Gaussian laser beam with an electric field $E_i(z, r, t)$ is incident on the sample and induces a phase shift $\Delta\phi(z, r, t)$ inside the sample due to nonlinear refraction. The incident electric filed is written as

$$E_i(z, r, t) = E_0(t)\frac{w_0}{w(z)}\exp\left[-\frac{r^2}{w^2(z)} - \frac{ikr^2}{2R(z)}\right]\exp\left[-i\varphi(z, t)\right] \tag{3.1}$$

where $k = 2\pi/\lambda$ is the wave vector, and λ is the laser wavelength. $w(z)$ and $R(z)$ are sample position z-dependent beam radius and the radius of curvature of the wavefront, respectively.

$$w^2(z) = w_0^2\left(1 + \frac{z^2}{z_0^2}\right) \tag{3.2}$$

$$R(z) = z\left(1 + \frac{z^2}{z_0^2}\right) \tag{3.3}$$

with

$$z_0 = kw_0^2/2 \tag{3.4}$$

$E_0(t)$ contains the temporal envelope of laser pulse and denotes the electric field at the center of beam waist position. For z-scan measurement, only the radial phase variation $\Delta\varphi(r)$ is concerned; hence, $\exp[-i\varphi(z,t)]$ can be ignored in the data analysis.

Under the irradiation of laser beam, the refraction index of sample is expressed as

$$n = n_0 + \frac{n_2}{2}|E|^2 = n_0 + \gamma I = n_0 + \Delta n \tag{3.5}$$

where n_0 is the linear index of refraction, and I denotes the irradiance intensity (MKS) of the laser beam within the sample. n_2 and γ are related through the conversion formula

$$n_2(\text{esu}) = \frac{cn_0}{40\pi}\gamma\left(\frac{\text{m}^2}{\text{W}}\right) \tag{3.6}$$

where c (m/s) is the speed of light in vacuum, and W is laser power unit of Watt. The basic assumption conditions for data processing need to be considered as follows:

1. The sample is thin enough that the changes in beam diameter within the sample due to either diffraction or nonlinear refraction are negligible. In this case, the self-refraction process is referred to as "external self-action." For linear refraction, this implies that $L \ll z_0$, where L is sample thickness, while for nonlinear refraction, $L \ll z_0/\Delta\phi(0)$, which is automatically met since $\Delta\phi(0)$ is small in most experiments. Thus, $L \ll z_0$ is more restrictive than it needs to be for linear refraction, and it is sufficient to replace it with $L < z_0$.
2. Fresnel reflection losses from sample are usually ignored such that, for example, $I_0(t)$ is thought to be the irradiance within the sample.
3. The far-field condition for aperture plane $L_{\text{far-field}} \gg z_o$ needs to be satisfied, where $L_{\text{far-field}}$ is the traveling distance of exiting electric field from sample surface to the aperture plane.

The assumptions simplify the problems considerably; in *slowly varying envelope approximation* (SVEA), the changes of intensity and phase of the electric field inside the sample are expressed by the following differential formula

$$\frac{\text{d}\Delta\phi(z,r,t)}{\text{d}z'} = \Delta n(I)k \tag{3.7}$$

$$\frac{\text{d}I}{\text{d}z'} = -\alpha(I)I \tag{3.8}$$

where z' is the propagation depth within the sample. $\alpha(I)$, in general, includes linear and nonlinear absorption terms; here, it should be noted that z' should not be confused with the sample position z.

If the nonlinear absorption is negligible, a phase shift $\Delta\phi(z,r,t)$ formed by the nonlinear refraction at the exiting surface of sample is

$$\Delta\phi(z,r,t) = \Delta\phi_0(z,t)\exp\left[-\frac{2r^2}{w^2(z)}\right] \tag{3.9}$$

with

$$\Delta\phi_0(z,t) = \frac{\Delta\Phi_0(t)}{1+\frac{z^2}{z_0^2}} \tag{3.10}$$

where $\Delta\Phi_0(t)$ is the on-axis phase shift at the beam waist position and defined as

$$\Delta\Phi_0(t) = k\Delta n_0(t)L_{\text{eff}} \tag{3.11}$$

with

$$L_{\text{eff}} = \frac{1-\exp(-\alpha_0 L)}{\alpha_0} \tag{3.12}$$

$$\Delta n_0(t) = \gamma I_0(t) \tag{3.13}$$

where α_0 is linear absorption coefficient, and $I_0(t)$ is the on-axis intensity at the focus ($r = z = 0$). The complex electric field at the exiting surface of the sample containing the nonlinear phase distortion is

$$E_e(z,r,t) = E_i(z,r,t)\exp\left(-\alpha_0 L/2\right)\exp\left[i\Delta\phi(z,r,t)\right] \tag{3.14}$$

The far-field light beam pattern at the aperture plane can be obtained using the Huygens principle by conducting a zeroth-order Hankel transformation of $E_e(z,r,t)$. To simplify the calculation, the Gaussian decomposition (GD) approach is implemented by Sheik-Bahae et al. [1].

If only small phase changes are considered, the $E_e(z,r,t)$ is decomposed into a summation of Gaussian beams through a Taylor series expansion of nonlinear phase term as follows

$$\exp\left[i\Delta\phi(z,r,t)\right] = \sum_{m=0}^{\infty}\left\{\frac{\left[i\Delta\phi_0(z,t)\right]^m}{m!}\exp\left[-\frac{2mr^2}{w^2(z)}\right]\right\} \tag{3.15}$$

Each Gaussian beam travels individually to the aperture plane where they are resumed to reconstruct the beam. Considering the initial beam curvature for the incident-focused beam, the resultant electric field pattern $E_a(z,r,t)$ at the aperture is as

$$E_a(z,r,t) = E_i(z,r=0,t)\exp\left(-\frac{\alpha L}{2}\right)\sum_{m=0}^{\infty}\left\{\frac{\left[i\Delta\phi_0(z,t)\right]^m}{m!}\frac{w_{m0}}{w_m}\exp\left(-\frac{r^2}{w_m^2}-\frac{ikr^2}{2R_m}+i\theta_m\right)\right\} \tag{3.16}$$

Setting $g = 1 + d/R(z)$, the remaining parameters in formula (3.16) are written as

$$w_{m0}^2 = \frac{w^2(z)}{2m+1} \tag{3.17}$$

$$w_m^2 = w_{m0}^2 \left(g^2 + \frac{d^2}{d_m^2} \right) \tag{3.18}$$

$$R_m = d \left(1 - \frac{g}{g^2 + d^2/d_m^2} \right)^{-1} \tag{3.19}$$

$$\theta_m = \tan^{-1} \left(\frac{d/d_m}{g} \right) \tag{3.20}$$

$$d_m = \frac{k w_{m0}^2}{2} \tag{3.21}$$

The transmitted laser power $P_T[\Delta\Phi_0(t)]$ through the aperture is obtained by spatially integrating $E_a(z, r, t)$ of formula (3.16) up to the aperture radius r_a and can be calculated as

$$P_T[\Delta\Phi_0(t)] = c\epsilon_0 n_0 \pi \int_0^{r_a} |E_a(z, r, t)|^2 r \mathrm{d}r \tag{3.22}$$

with ϵ_0 the permittivity of vacuum, the normalized z-scan transmittance is

$$T(z) = \frac{\int_{-\infty}^{+\infty} P_T[\Delta\Phi_0(t)]\mathrm{d}t}{S \int_{-\infty}^{+\infty} P_i(t)\mathrm{d}t} \tag{3.23}$$

with

$$P_i(t) = \frac{\pi w_0^2 I_0(t)}{2} \tag{3.24}$$

$$S = 1 - \exp\left(-\frac{2r_a^2}{w_a^2} \right) \tag{3.25}$$

where $P_i(t)$ is the instantaneous incident laser power (inside the sample), S is the linear transmittance through the aperture, and w_a is the beam radius at the aperture plane. Formula (3.23) tells us that for a given $\Delta\Phi_0$, the magnitude and shape of $T(z)$ are independent on the geometry and the light wavelength.

In the experimental setup, the aperture size S is important since large aperture reduces the variations of $T(z)$. For a very large or no aperture ($S = 1$), the variations of $T(z)$ do not occur and $T(z) = 1$ and thus cannot reflect the nonlinear refraction signal.

To simplify the analysis, an easily measurable quantity ΔT_{P-V} is defined as the difference between normalized peak and valley transmittances. ΔT_{P-V} is a function of $|\Delta \Phi_0|$ and linear dependence on a specific aperture size, and independent of the geometry and laser wavelength. In the accuracy range of ± 2 %, the dependence can be expressed as

$$\Delta T_{P-V} \approx 0.406(1-S)^{0.25}|\Delta \Phi_0| \quad \text{for} \quad |\Delta \Phi_0| < \pi \tag{3.26}$$

One can extend the analysis to the transient effects induced by the laser pulse using a time-averaged refraction index change $\langle \Delta n_0(t) \rangle$,

$$\langle \Delta n_0(t) \rangle = \frac{\int_{+\infty}^{-\infty} \Delta n_0(t) I_0(t) \mathrm{d}t}{\int_{+\infty}^{-\infty} I_0(t) \mathrm{d}t} \tag{3.27}$$

Considering that the nonlinearity has an instantaneous response, for a temporally Gaussian pulse, $\langle \Delta n_0(t) \rangle$ can be simplified as

$$\langle \Delta n_0(t) \rangle = \Delta n_0 / \sqrt{2} \tag{3.28}$$

where Δn_0 is the on-axis peak value change of refractive index at focus.

All these formulas are derived by analyzing the third-order nonlinearity. The higher-order nonlinearities can be also conducted in a similar way. Taking a fifth-order nonlinearity $\chi^{(5)}$ as an example, where $\chi^{(5)}$ can occur in semiconductors due to charge of carriers generated through two-photon absorption, the nonlinear refraction change is accordingly represented as

$$\Delta n = \eta I^2 \tag{3.29}$$

Following the same steps, a formula, similar to the third-order nonlinearity, can be achieved for the fifth-order nonlinearity

$$\Delta T_{P-V} \approx 0.21(1-S)^{0.25}|\Delta \Phi_0|, \Delta \Phi_0 = k\eta I^2 \left(\frac{1-e^{-2\alpha_0 L}}{2\alpha_0}\right) \tag{3.30}$$

3.2.2.2 Nonlinear Absorption

In general, the nonlinear refraction is companied with nonlinear absorption, which can be expressed by the third-order complex susceptibility

$$\chi^{(3)} = \chi_R^{(3)} + i\chi_I^{(3)} \tag{3.31}$$

with the relations to nonlinear refraction and nonlinear absorption coefficients

$$\chi_R^{(3)} = 2n_0^2 \epsilon_0 c \gamma \tag{3.32}$$

$$\chi_I^{(3)} = \frac{n_0^2 \epsilon_0 c^2}{\omega} \beta \tag{3.33}$$

where ω is frequency of incident laser beam, β the nonlinear absorption coefficient. For nonlinear sample, under the laser irradiation, the absorption coefficient is written as

$$\alpha(I) = \alpha_0 + \beta I \tag{3.34}$$

The z-scan measurement can be also used to determine β if the aperture of Fig. 3.1 is removed, which is called as open-aperture mode z-scan. The graph of transmittance versus sample position is symmetric around the focus.

Under the condition of considering the nonlinear absorption, the solutions of differential formulas (3.7) and (3.8) yield the irradiation intensity distribution and phase shift of the laser beam at the exiting surface of sample as follows

$$I_e(z, r, t) = \frac{I(z, r, t)\exp(-\alpha L)}{1 + q(z, r, t)} \tag{3.35}$$

$$\Delta\phi(z, r, t) = \frac{k\gamma}{\beta}\ln\left[1 + q(z, r, t)\right] \tag{3.36}$$

with

$$q(z, r, t) = \beta I L_{\text{eff}} \tag{3.37}$$

For nonlinear absorption, the aperture is removed in z-scan measurement, which causes the normalized z-scan curve to be insensitive to nonlinear refraction, and the normalized z-scan curve is only a function of the nonlinear absorption. The total transmitted power in that case ($S = 1$) can be obtained by spatially integrating formula 3.35) without the free-space propagation process. Integrating formula (3.35) at position z over r, the transmitted power is:

$$P(z, t) = P_i(t)\exp(-\alpha L)\frac{\ln\left[1 + q_0(z, t)\right]}{q_0(z, t)} \tag{3.38}$$

with

$$q_0(z, t) = \frac{\beta I_0(t) L_{\text{eff}}}{1 + z^2/z_0^2} \tag{3.39}$$

where $P_i(t)$ has been defined in formula (3.24). For temporally Gaussian pulse, the formula (3.38) can be time-averaged integrating to get normalized energy transmittance as

$$T(z, S = 1) = \frac{1}{\sqrt{\pi}q_0(z, 0)}\int_{-\infty}^{+\infty}\ln\left[1 + q_0(z, 0)e^{-\tau^2}\right]\mathrm{d}\tau \tag{3.40}$$

For $|q_0| < 1$, the formula (3.40) can be simplified as

$$T(z, S = 1) = \sum_{m=0}^{\infty}\frac{\left[-q_0(z, 0)\right]^m}{(m + 1)^{3/2}} \tag{3.41}$$

The nonlinear absorption coefficient β can be determined unambiguously by fitting the normalized transmittance curve from open-aperture mode z-scan measurement. The nonlinear refraction coefficient γ can be directly obtained through fitting the pure nonlinear refraction transmittance curve, which can usually be obtained through dividing the data of closed-aperture mode z-scan data by those of open-aperture mode z-scan data.

3.3 Generation and Elimination of Pseudo-nonlinearity in z-scan Measurement

Generally speaking, in the case of pure linear absorption, for ideal Gaussian beams, i.e., within the near-axis approximation, the light intensity should remain constant during the z-scan, which is a well-known basic principle for z-scan measurement. However, it is demonstrated the existence of pseudo-nonlinear absorption in z-scan measurements even for samples of pure linear absorption due to the incident angle change [7]. The shape of the discovered pseudo-nonlinear absorption curves appears to be similar to that of the nonlinear absorption curves of the open-aperture z-scan measurement. The false nonlinearities may be recorded especially for converging lenses with a relatively large numerical aperture.

3.3.1 Incident Angle as a Function of z-scan Position

In a practical z-scan measurement, a collimated laser beam is focused by a converging lens with an angular semi-aperture of θ_0. The light beam traveling along the z-axis is shown in Fig. 3.4. The angular semi-aperture θ_0 is only determined by the lens. The incident angle is $\theta_i = \theta_0$ when the sample is located far from the beam waist position (position 1 in Fig. 3.4a). The incident angle θ_i decreases with the sample movement from position 1 to position 2 along the optical z-scan axis (Fig. 3.4b). The incident angle reaches $\theta_i = 0$ when the sample moves to the beam waist position (focus) (position 3 in Fig. 3.4c).

The change of light beam radius $w(z)$ with sample position z obeys the hyperbolic formula,

$$\frac{w^2(z)}{w_0^2} - \frac{z^2}{z_0^2} = 1 \tag{3.42}$$

The dependence of the incident angle θ_i on the sample position z can be directly obtained from the derivative of formula (3.42),

$$\theta_i = \arctan\left[\frac{\partial w(z)}{\partial z}\right] = \arctan\left(\frac{z}{\sqrt{z_0^2 + z^2}}\frac{\lambda}{\pi w_0}\right) \tag{3.43}$$

Formula (3.43) shows that the θ_i changes with sample position z. At $z = 0$, $\theta_i = 0$, which corresponds to the beam waist position (position 3 in Fig. 3.4c), and at $z \to +\infty$, $\theta_i = \theta_0$, which corresponds to the angular semi-aperture (position 1 in Fig. 3.4a). Figure 3.5 shows an example that the incident angle θ_i changes with the sample position z. In the z-scan process from $z \to -\infty$ to $z = 0$, the incident angle decreases from $\theta_i = \theta_0 = 11.5°$ to $\theta_i = 0$ and then increases from $\theta_i = 0$ to $\theta_i = \theta_0 = 11.5°$ when the sample position changes from $z = 0$ to $z \to +\infty$.

The beam waist radius w_0 and angular semi-aperture θ_0 change with the numerical aperture (NA) of the lens, and θ_0 can be calculated as:

$$\theta_0 = \arcsin\left(\mathrm{NA}/n_{\mathrm{environ}}\right) \tag{3.44}$$

where n_{environ} is refractive index of environment, and $n_{\mathrm{environ}} = 1$ because the z-scan is usually carried out in an atmospheric environment. Formula (3.44) indicates that θ_0 increases with NA, which means that in z-scan measurement, θ_0 becomes large if the lens with a large NA is used.

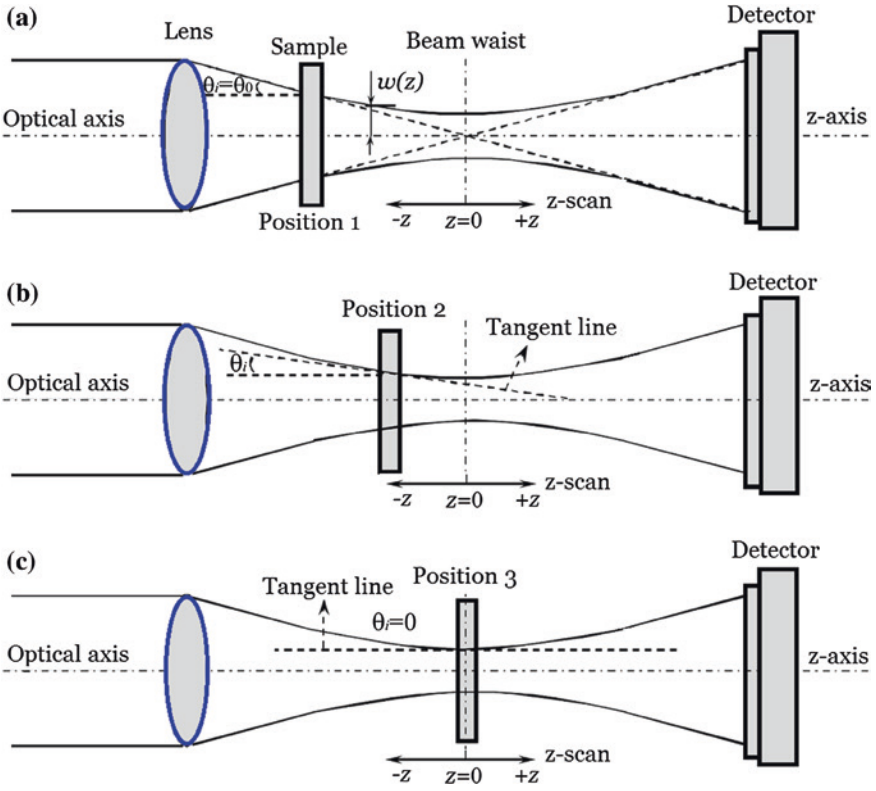

Fig. 3.4 Schematic of a z-scan measurement, sample position **a** far away from focal point, **b** close to focal point, and **c** at focal point

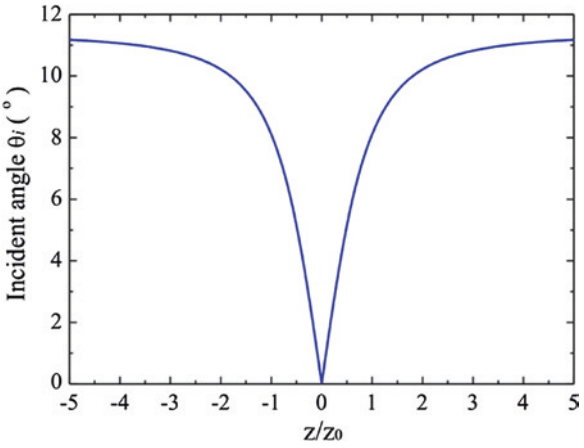

Fig. 3.5 Changes of calculated z-scan incident angle with position z ($\lambda = 633$ nm, $w_0 = 1$ μm)

3.3.2 *Dependence of Transmittance on Incident and Polarization Azimuth Angles*

It is known that from the textbook, the transmittance and reflectance of sample are dependent on the incident and polarization azimuth angles. For the polarization, two special cases of s- and p-polarizations are taken into account. Figure 3.6 shows the schematic of the transmission and reflection when a laser beam irradiates onto the sample surface. Note in passing that although the whole laser beam is generalized as a Gaussian beam with a propagation direction along the optical axis, i.e., in perpendicular to the sample surface, the field components of the beam remain vectors and the polarization of each component can always be decomposed into s- and p-polarizations according to the local surface and the inputting direction of the component, i.e., the polarization azimuth angles. Figure 3.6a shows the p-polarization incident light, where the electric vector of the incident light $E_{\parallel}^{(i)}$ is parallel to the incident plane, and Fig. 3.6b shows the s-polarization incident light, where the electric vector of the incident light $E_{\perp}^{(i)}$ is perpendicular to the incident plane. In the general case of an incident electric field with a polarization azimuth angle α, it is always possible to decompose the field $E^{(i)}$ into s-polarization and p-polarization components, as schematically shown in Fig. 3.6c, such that $E_{\parallel}^{(i)} = E^{(i)} \cos\alpha$, $E_{\perp}^{(i)} = E^{(i)} \sin\alpha$.

The light is incident from medium 1 with a refractive index of n_1 to medium 2 with a refractive index of n_2. The Poynting vectors of incident and refraction light are marked as $s^{(i)}$ and $s^{(t)}$, respectively. The incident and refraction angles are θ_i and θ_t, respectively. The refraction angle can be calculated as

$$\theta_t = \arcsin \frac{n_1 \sin\theta_i}{n_2} \tag{3.45}$$

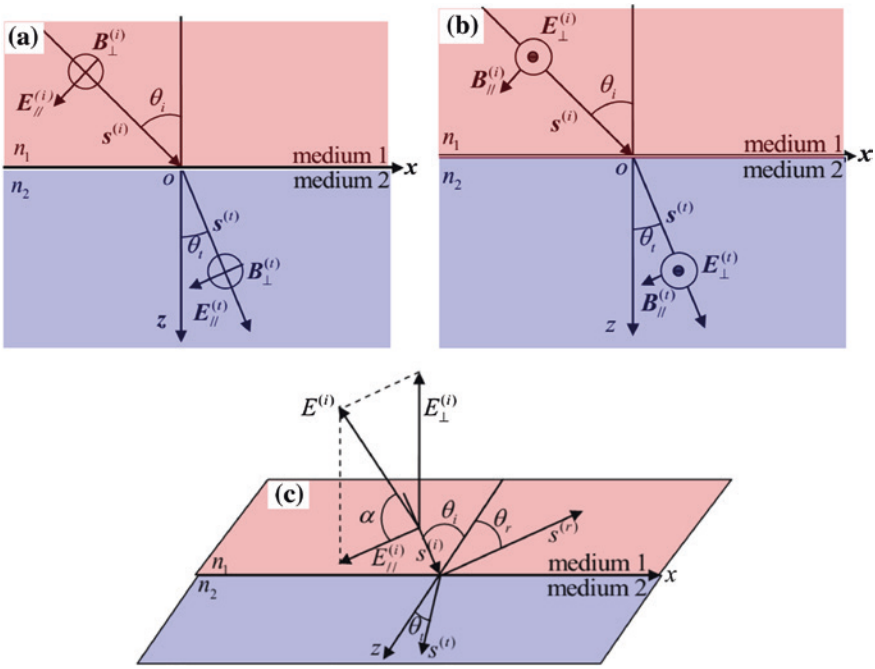

Fig. 3.6 Schematic of transmission. **a** *p*-polarization incident, **b** *s*-polarization incident, and **c** any polarization azimuth angle incident

and the electric field of the transmitted light can be expressed as:

$$\begin{cases} E_{\parallel}^{(t)} = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} E_{\parallel}^{(i)} \\ E_{\perp}^{(t)} = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} E_{\perp}^{(i)} \end{cases} \tag{3.46}$$

The horizontal component of the transmittance, defined as the ratio of the horizontal component of the transmitted light intensity to the horizontal component of the incident light intensity, then is,

$$T_{\parallel}(\theta_i) = \frac{n_2 \cos\theta_t}{n_1 \cos\theta_i} \frac{\left|E_{\parallel}^{(t)}\right|^2}{\left|E_{\parallel}^{(i)}\right|^2} = \frac{4n_1 n_2 \cos\theta_i \cos\theta_t}{(n_1 \cos\theta_t + n_2 \cos\theta_i)^2} \tag{3.47}$$

The vertical component of the transmittance, defined as the ratio of the vertical component of the transmitted light intensity to the vertical component of the incident light intensity, can be written as,

$$T_{\perp}(\theta_i) = \frac{n_2 \cos\theta_t}{n_1 \cos\theta_i} \frac{\left|E_{\perp}^{(t)}\right|^2}{\left|E_{\perp}^{(i)}\right|^2} = \frac{4n_1 n_2 \cos\theta_i \cos\theta_t}{(n_1 \cos\theta_i + n_2 \cos\theta_t)^2} \tag{3.48}$$

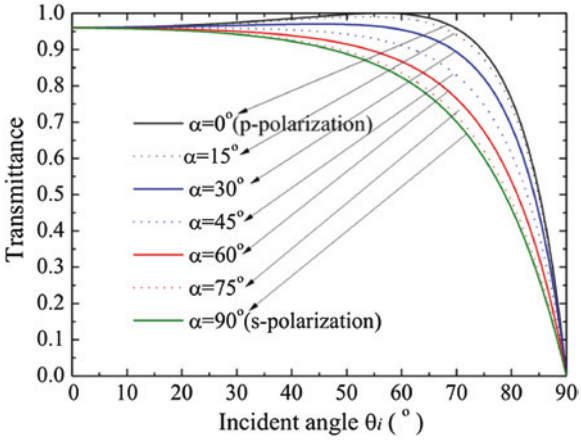

Fig. 3.7 Dependence of transmittance on the incident angle for different polarization azimuth angle

Finally, the total transmittance can be calculated by

$$T(\theta_i, \alpha) = T_{\parallel}(\theta_i)\cos^2\alpha + T_{\perp}(\theta_i)\sin^2\alpha \tag{3.49}$$

Figure 3.7 shows that the calculated transmittance is dependent on the incident angle for light with a given polarization azimuth angle. For $\alpha = 0$, the transmittance increases with the incident angle at $\theta_i < 56°40'$. The transmittance is 1 at $\theta_i = 56°40'$, and then decreases with the incident angle at $\theta_i > 56°40'$. For $\alpha < 45°$, the transmittance initially increases to a certain value and then decreases with the incident angle. For $\alpha > 45°$, the transmittance always decreases with the incident angle.

3.3.3 *Incident Angle Change-Induced Pseudo-nonlinear Absorption*

In z-scan measurement, the sample is moved between the $-z$ and $+z$ positions. For open-aperture mode z-scan, the Fourier transform of the amplitude of incident beam [from formula (3.1)] is called as the spatial spectrum $\tilde{E}_i(z,f)$, where f is the spatial frequency [8]:

$$E_i^a(z,r) = \int_{-\infty}^{+\infty} \tilde{E}_i(z,f)\exp(i2\pi fr)\mathrm{d}f \tag{3.50}$$

where $E_i^a(z,r)$ is the amplitude of electric field of incident beam. The $\exp(i2\pi fr)$ can be interpreted as a plane wave traveling in a direction that lies in the $r-z$ plane, making an angle θ_i with the z-axis such that

$$f = \sin\frac{\theta_i}{\lambda} \tag{3.51}$$

The $E_i^a(z,r)$ can be also obtained by taking the real part of formula (3.1) as follows

$$E_i^a(z,r) = E_0 \frac{w_0}{w(z)} \exp\left[-\frac{r^2}{w^2(z)}\right] \tag{3.52}$$

According to formulas (3.50) and (3.52), the corresponding spatial spectrum of incident light at the waist $w(z)$ is

$$\tilde{E}_i(z,f) = \sqrt{\pi} E_0 w_0 \exp\left\{-\left[\pi w(z) f\right]^2\right\} \tag{3.53}$$

The amplitude part of the transmission function depends sensitively on the incident angle θ_i of the various plane wave components and polarization azimuth angle α. The transmitted electric field $E^{(t)}(z,r,\alpha)$ can be expressed as

$$E_t(z,r,\alpha) = \int_{-\infty}^{+\infty} \tilde{E}_t(z,f,\alpha) \exp(i2\pi f r) \mathrm{d}f \tag{3.54}$$

The transmitted light intensity $I^{(t)}(z,r,\alpha)$ can be written as follows

$$I_t(z,r,\alpha) = E_t(z,r,\alpha) \times E_t^*(z,r,\alpha) \tag{3.55}$$

The actually detected light transmittance is an integration along the radial coordinate r. Finally, the normalized z-scan transmittance curve through the sample (such as glass substrate) can be approximately obtained by spatially integrating $I_{(t)}(z,r,\alpha)$, giving

$$T(z,\alpha) = \frac{P_t(z,\alpha)}{P_i(z)} = \frac{\int_0^{+\infty} I_t(z,r,\alpha) r \mathrm{d}r}{\int_0^{+\infty} I_i(z,r) r \mathrm{d}r} \tag{3.56}$$

where P_i is the incident power and P_t is the transmitted power. According to formula (3.1), the incident laser beam intensity $I_i = E \times E^*$ can be written as:

$$I_i(z,r) = I_0 \frac{w_0^2}{w^2(z)} \exp\left[-\frac{2r^2}{w^2(z)}\right] \tag{3.57}$$

with

$$w_0 = \frac{\lambda}{\pi \tan(\theta_0)} \tag{3.58}$$

3.3.4 Calculated Pseudo-nonlinear Absorption Curves

3.3.4.1 Pseudo-nonlinear Absorption Curves Induced by Changes of Incident Angle

Figure 3.8 shows the calculated pseudo-nonlinear absorption z-scan curves. Figure 3.8a is, for $\alpha = 90°$ (s-polarization), a peak configuration of this z-scan,

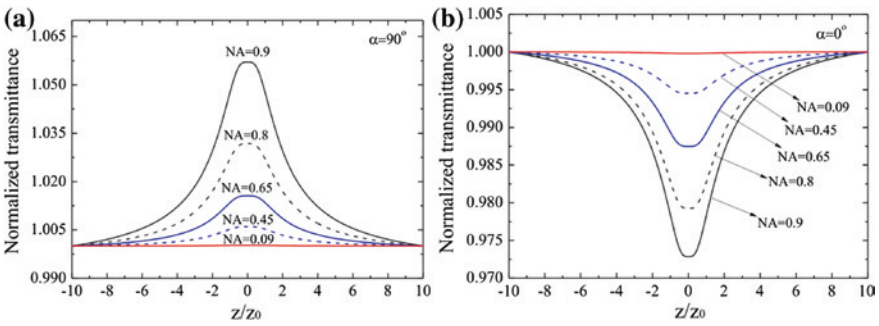

Fig. 3.8 Calculated result of $T(Z)$ for $\alpha = 90°$, $90°$ using $\lambda = 633$ nm, $n_0 = 1$, $n_1 = 1.5$, and NA $= 0.9, 0.8, 0.65, 0.45$, as well as 0.09: **a** *s*-polarization and **b** *p*-polarization

which is indicative of nonlinear saturation absorption characteristic, and the maximum transmittance variation increases with increased NA. Figure 3.8b is, for $\alpha = 0°$ (*p*-polarization), a valley configuration, which is indicative of nonlinear reverse saturation absorption characteristic, and the maximum transmittance variation also increases with an increased NA. Both Fig. 3.8a, b show that the pseudo-nonlinear absorption becomes substantial when NA increases.

3.3.4.2 Pseudo-nonlinear Absorption Curves Induced by Changes of Polarization Angle

According to formula (3.56), the polarization azimuth angle also has an important effect on the normalized transmittance, which causes pseudo-nonlinear absorption characteristics in the *z*-scan measurement. Figure 3.9 shows that the change of the polarization azimuth angle leads to pseudo-nonlinear absorption curves. Figure 3.9a, b show the results for NA of 0.90 and 0.45, respectively. The result for the NA of 0.09 is shown in Fig. 3.9c. Figure 3.9a shows that for NA $= 0.90$ when the polarization azimuth angles are $\alpha = 0°$, $35°$, $45°$, $60°$, and $90°$, the maximum transmittance variations are 2.72, 1.79, 0.075, 1.35, 3.49 and 5.7 %, respectively. Figure 3.9a shows that NA $= 0.45$, when the polarization azimuth angles are $\alpha = 0°$, $20°$, $35°$, $45°$, $60°$, and $90°$, the maximum transmittance variations are 0.6, 0.31, 0.02 0.18, 0.42 and 0.55 %, respectively. For NA$= 0.09$, the maximum variation in the transmittance is less than 0.02 %, as shown in Fig. 3.9c. Figure 3.9 also shows that the peak and valley configurations of the *z*-scan curves are indicative of transformation from pseudo-nonlinear saturation absorption to pseudo-nonlinear reverse saturation absorption due to different polarization azimuth angles.

Figure 3.9 indicates that for a fixed lens NA, the incident light beam with a critical polarization azimuth angle α_{critical} can minimize or eliminate the interference of the pseudo-nonlinear absorption on the experimental results. Figure 3.10 shows the dependence of α_{critical} on the NA. At NA $= 0.10$, $\alpha_{\text{critical}} = 45.2°$. α_{critical} decreases with increased NA and $\alpha_{\text{critical}} = 36.2°$ at NA $= 0.90$.

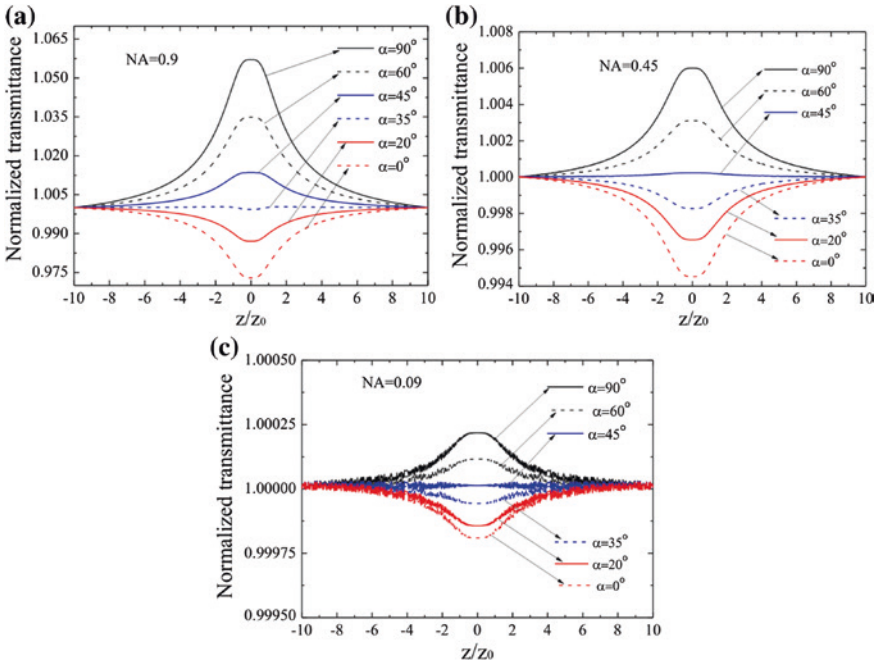

Fig. 3.9 Calculated results of $T(Z)$, using $\lambda = 633\,\text{nm}$, $n_0 = 1$, $n_1 = 1.5$. **a** NA = 0.09, **b** NA = 0.45, as well as **c** NA = 0.09 for $\alpha = 0°$ (s-polarization), 20°, 35°, 45°, 60°, and 90° (p-polarization)

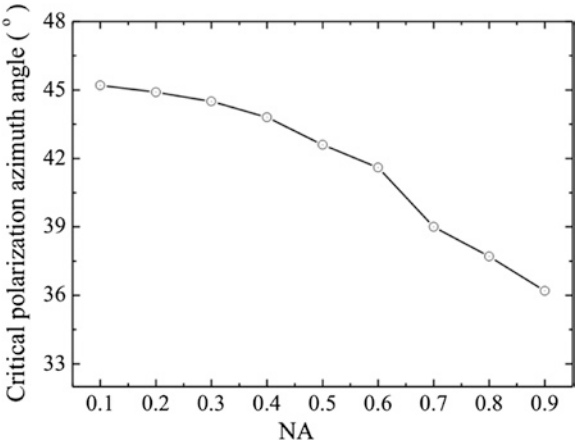

Fig. 3.10 Calculated dependence of critical polarization azimuth angle on NA

3.3.5 Reduction or Elimination of Pseudo-nonlinear Absorption

Figures 3.8, 3.9 and 3.10 also indicate that the pseudo-nonlinear absorption effect can be reduced or avoided by choosing the incident light beam with a critical polarization azimuth angle or using a lens with a low NA in z-scan measurement. In order to verify these, an open-aperture mode z-scan setup is established. In this setup, the cw Ar^+ laser beam with a wavelength of 488 nm is used as the light source, a half-wave plate is adopted to change the laser beam polarization azimuth angle, and the angular semi-aperture is tuned by the lens with different NA.

The experimental results are shown in Fig. 3.11 where a 0.6 mm-thickness BK7 glass slice sample is examined. Figure 3.11a depicts the results of polarization azimuth angle $\alpha = 90°$ where the three cases of NA = 0.09, 0.45 and 0.65 are presented, respectively. The experimental curves are indicative of nonlinear saturation absorption characteristics. The pseudo-nonlinear absorption is so small that changes of the transmittance are not significant when NA is not more than 0.09. The maximum variation of the transmittance increases with increased NA and can reach 0.36 and 0.57 % for NA = 0.45 and 0.65, respectively.

Figure 3.11b shows the experimental data for NA = 0.45, and the maximum transmittance variations are 0.36 and 0.47 % for s-polarization ($\alpha = 90°$) and p-polarization ($\alpha = 0°$), respectively. In the z-scan measurement, the glass slice sample moves along the optical axis, the normalized transmittance curves slowly increase especially in the range of Rayleigh length, the normalized transmittance can basically remain unchanged, and the calculated curve shows a platform. In addition, the polarization angle can be adjusted to the corresponding critical angle to eliminate the influence of pseudo-nonlinear absorption. A half-wave plate is used to tune the critical polarization azimuth angle $\alpha_{\text{critical}} = 43°$; the results are presented in curve marked with black square symbol in Fig. 3.11b, and the pseudo-nonlinear absorption becomes very small and can be ignored.

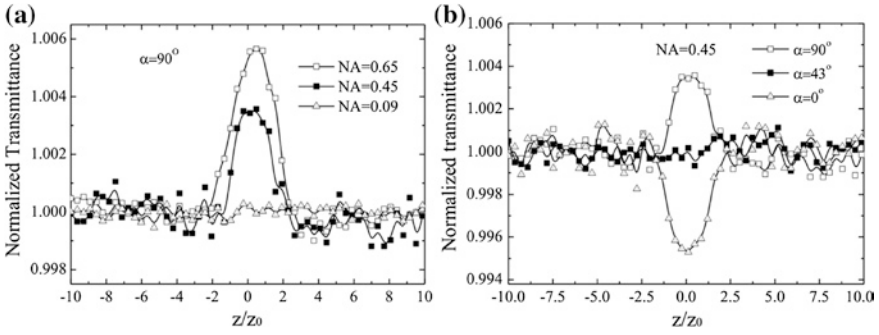

Fig. 3.11 Normalized z-scan transmittance of BK7 glass sample measured using *cw* laser with $\lambda = 488$ nm. **a** $\alpha = 90°$, NA = 0.09, 0.45, 0.65, **b** $\alpha = 90°, 43°, 0°$, NA = 0.45

The change of the incident angle in the z-scan measurement can cause the pseudo-nonlinear absorption, and the significance of the effects depends on the NA of the lens. When the laser is focused by a lens with a large NA, the interference of the pseudo-nonlinear absorption due to the incident angle change cannot be ignored. If a lens with a low NA is used, the error induced by the pseudo-nonlinear absorption is negligible. For an incident laser beam focused by a lens with a large NA, the polarization azimuth angle of the incident light can be adjusted to the corresponding critical angle to eliminate the influence of pseudo-nonlinear absorption.

3.4 Eliminating the Influence from Reflection Loss on z-scan Measurement

3.4.1 Fresnel Reflection Loss in the z-scan Measurement

In z-scan measurement, when optical nonlinear refraction and absorption coefficients are small enough the Fresnel reflection losses can be neglected. However, the Fresnel reflection loss is significant and cannot be neglected for some samples, such as the nanofilms or strong nonlinear materials. If there are significant nonlinear absorption and refraction, the effective incident light intensity I_{eff} within the sample at different z-scan position is no longer a constant, but is a function of sample position z, and the incident light intensity within the sample should be expressed as $I_{\text{eff}}(z)$. For a typical z-scan configuration as shown in Fig. 3.12, a collimated Gaussian laser beam with electric field intensity E_i is used as the light

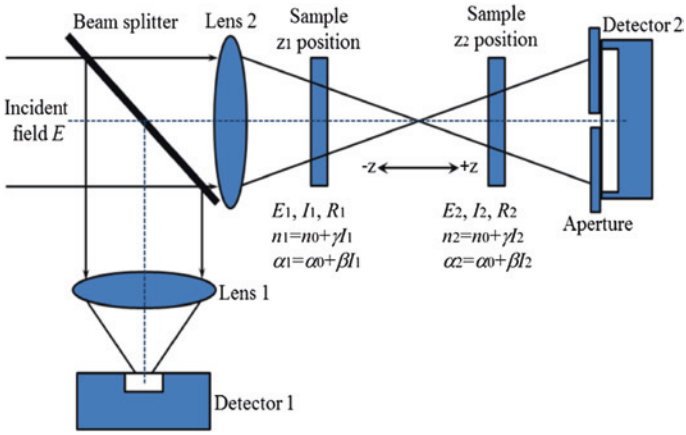

Fig. 3.12 The conventional configuration of z-scan setup (E_1 electric field, I_1 intensity, R_1 reflection, n_1 refractive index, α_2 absorption coefficient at position z_1; E_2 electric field, I_2 intensity, R_2 reflection, n_2 refractive index, α_2 absorption coefficient at position Z_2). Reprinted from Ref. [9], Copyright 2013, with permission from Elsevier

source. The light beam is split into two parts by a beam splitter, one is focused onto detector 1 to monitor the beam power stability, and the other is focused by lens 2 to function as nonlinear z-scan measurement light beam. The sample is z-scanned along the optical axis near the focus point. The light beam through the sample is fully collected by detector 2 for nonlinear absorption measurement, and nonlinear refraction is measured by placing an aperture at the front of detector 2.

However, there is large surface reflection for certain kind of samples, and the Fresnel surface reflection loss is different at different position z at the z-scan measurement, resulting from the laser beam-induced nonlinear absorption and nonlinear refraction within the sample. For example, when the sample z-scans to position z_1, the incident laser beam intensity is I_1. The laser beam radius and the electric field intensity are marked as r_1 and E_1, respectively. The refractive index $n_1 = n_0 + \gamma I_1$, and absorption coefficient $\alpha_1 = \alpha_0 + \beta I_1$ accordingly. The Fresnel reflection loss is marked as R_1. However, when the sample z-scans to position z_2, the incident laser intensity becomes I_2 because the laser beam radius is r_2, and $r_1 \neq r_2$ leads to $I_1 \neq I_2$. At position z_2, the laser beam-induced refractive index and absorption coefficient are $n_2 = n_0 + \gamma I_2$, and $\alpha_2 = \alpha_0 + \beta I_2$, respectively. The Fresnel reflection loss R_2 at position z_2 is different with the R_1 at position z_1 that is, $R_2 \neq R_1$ because $n_2 \neq n_1$ and $\alpha_2 \neq \alpha_1$. Therefore, the Fresnel reflection loss at the z-scan measurement cannot be ignored because the reflection loss is not a constant, but is a function of position z, that is $R(z)$. Here, two kinds of samples are analyzed, one is thin samples, and the other is thin films with a thickness of nanoscale, which is named as nanofilms.

3.4.2 The Case of Thin Samples

3.4.2.1 Theoretical Model for Characterization of Nonlinear Refraction and Absorption

For thin samples, the thickness is small enough that change of the beam diameter within the sample due to either diffraction or nonlinear refraction can be neglected. The sample is considered to be "thin" sample, in which case the self-refraction process is referred to as "external self-action."

When the sample moves to position z, the effective laser intensity inside the sample should be a function of z and is set to be $I_{\text{eff}}(z)$. Both the induced refractive index n and absorption coefficient α are also the functions of position z and can be expressed as

$$n(z) = n_0 + \gamma I_{\text{eff}}(z) \tag{3.59}$$

$$\alpha(z) = \alpha_0 + \beta I_{\text{eff}}(z) \tag{3.60}$$

The Fresnel reflection loss can be considered to be from two respects. One is induced by the internal multi-interference effect for the thin sample with a

thickness of $L <\sim 4.65/\alpha$, the "very thin" samples are called as the nanofilms. This case will be analyzed in the next section. The other is from the surface (or interface) reflection, the samples are called as "thin samples." The internal multi-interference effect is negligible if the sample thickness is $L >\sim 4.65/\alpha$, based on the textbook of Born and Wolf [10]. For example, for the film sample with an absorption coefficient $\alpha = 4 \times 10^7 \mathrm{m}^{-1}$, the film thickness should meet the condition of L $\geq$ 120 nm if the internal multi-interference effect needs to be ignored. The reflection loss from sample surface at position z can be calculated as

$$R(z) = \frac{[n(z)-1]^2 + \left[\frac{\alpha(z)}{2k}\right]^2}{[n(z)+1]^2 + \left[\frac{\alpha(z)}{2k}\right]^2} \tag{3.61}$$

At position z, the effective laser intensity $I_{\mathrm{eff}}(z)$ is key to characterize the induced $n(z)$ and $\alpha(z)$. The effective laser beam intensity can be calculated as

$$I_{\mathrm{eff}}(z,r) = [1 - R(z)] \times I_i(z,r) \tag{3.62}$$

Substituting formula (3.57) into formula (3.62), $I_{\mathrm{eff}}(z,r)$ can be calculated as

$$I_{\mathrm{eff}}(z,r) = [1 - R(z,r)] \times \left\{ I_0 \frac{w_0^2}{w^2(z)} \exp\left[-\frac{2r^2}{w^2(z)}\right] \right\} \tag{3.63}$$

Based on the formulas (3.61)–(3.63), one can see that it is difficult to extract the $I_{\mathrm{eff}}(z)$ as an explicit function. The $R(z)$ can be approximately calculated using $I_{\mathrm{eff}-}(z)$ instead of the $I_{\mathrm{eff}}(z)$, where $I_{\mathrm{eff}-}(z)$ is the effective intensity inside the sample placed at the position z_-, and z_- is defined as the previous position close to position z. Thus,

$$I_{\mathrm{eff}}(z,r) \simeq \left[1 - R_-(z_-,r)\right] \times \left\{ I_0 \frac{w_0^2}{w^2(z)} \exp\left[-\frac{2r^2}{w^2(z)}\right] \right\} \tag{3.64}$$

with

$$R_-(z_-,r) = \frac{\left\{\left[n_0 + \gamma I_{\mathrm{eff}\_}(z,r)\right] - 1\right\}^2 + \left[\frac{\alpha_0 + \beta I_{\mathrm{eff}\_}(z,r)}{2k}\right]^2}{\left\{\left[n_0 + \gamma I_{\mathrm{eff}\_}(z,r)\right] + 1\right\}^2 + \left[\frac{\alpha_0 + \beta I_{\mathrm{eff}\_}(z,r)}{2k}\right]^2} \tag{3.65}$$

For thin samples with a thickness $z_0 >> L > 4.65/\alpha$, the *slowly varying envelope approximation* (SVEA) can be adopted. All phase changes are neglected because the transmitted light intensity is concerned. The effective laser intensity inside the sample $I_{\mathrm{eff}}(z)$ decays exponentially along the optical axis z-direction as

$$\frac{\mathrm{d}I_{\mathrm{eff}}(z,r)}{\mathrm{d}z'} = -\alpha I_{\mathrm{eff}}(z,r) \tag{3.66}$$

The intensity at the exiting surface of the sample can be calculated as

$$I_e(z,r) = \frac{I_{\text{eff}}(z,r)\exp(-\alpha_0 L)}{1+\beta L_{\text{eff}} I_{\text{eff}}(z,r)} \tag{3.67}$$

where L_{eff} is the effective sample thickness, which is calculated as

$$L_{\text{eff}} = \frac{1-\exp(-\alpha_0 L)}{\alpha_0} \tag{3.68}$$

The total transmitted light intensity (i.e., open-aperture mode z-scan $S = 1$) can be obtained by spatially integrating formula (3.67). The normalized transmitted z-scan curve is

$$\begin{aligned} T(z,S=1) &= \frac{\pi\int_0^{+\infty} I_e(z,r)r\mathrm{d}r}{\pi\int_0^{+\infty} I_i(z,r)r\mathrm{d}r} = \frac{\int_0^{+\infty}\frac{I_{\text{eff}}(z,r)\exp(-\alpha_0 L)}{1+\beta L_{\text{eff}} I_{\text{eff}}(z,r)}r\mathrm{d}r}{\int_0^{+\infty} I_0\exp\left(-\frac{2r^2}{w_0^2}\right)r\mathrm{d}r} \\ &= \frac{\int_0^{+\infty}\frac{I_{\text{eff}}(z,r)\exp(-\alpha_0 L)}{1+\beta L_{\text{eff}} I_{\text{eff}}(z,r)}r\mathrm{d}r}{\frac{1}{4}I_0 w_0^2} \end{aligned} \tag{3.69}$$

The normalized open-aperture mode reflected z-scan curve is

$$R(z,S=1) = \frac{\pi\int_0^{+\infty}[1-R(z)]I_i(z,r)r\mathrm{d}r}{\pi\int_0^{+\infty} I_0\exp\left(-\frac{2r^2}{w_0^2}\right)r\mathrm{d}r} = \frac{\int_0^{\infty}[1-R(z)]I_i(z,r)rdr}{\frac{1}{4}I_0 w_0^2} \tag{3.70}$$

In order to obtain the nonlinear absorption coefficient β and nonlinear refraction coefficient γ, the z-scan measurement data are processed by the following steps.

1. The open-aperture mode reflected z-scan measurement is performed to obtain the $R(z)$ curve experimentally. The measured $R(z)$ curve is theoretically determined by both the nonlinear refraction coefficient γ and nonlinear absorption coefficient β.
2. According to the measured $R(z)$ curve, the effective laser intensity $I_{\text{eff}}(=[1-R(z)]I_i)$ inside the sample is calculated by formula (3.63).
3. The open-aperture mode transmitted z-scan measurement is then carried out to obtain $T(z)$ curve experimentally.
4. The measured $T(z)$ is fitted by formula (3.69) and I_{eff}, the nonlinear absorption coefficient β can be firstly obtained.
5. The measured open-aperture mode reflected z-scan curve is fitted by the formula (3.70) and given β value, and the nonlinear refraction coefficient γ is finally determined.

3.4.2.2 Experimental Data Fitting

In order to verify the theoretical analysis, an experimental measurement system with open-aperture mode reflected and transmitted z-scan measurements is established. Figure 3.13 describes the schematics.

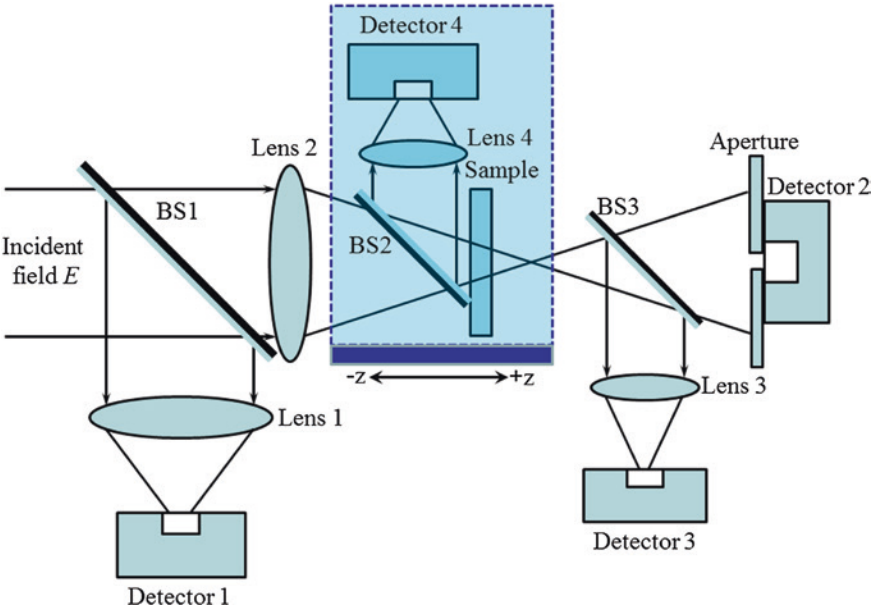

Fig. 3.13 Experimental measurement setup. The reflected light is collected by detector 4. (To avoid the reflected spot size changes drastically with the *z*-scan movement, the reflected light collection unit, which includes BS2, lens 4, and detector 4, synchronously conducts *z*-scan movement with the sample, where BS is beam splitter.) Reprinted from Ref. [9], Copyright 2013, with permission from Elsevier

In this measurement system, a laser beam with a wavelength of 405 nm is used as the incident light source, and the incident light is divided into two beams. One small beam is focused into detector 1 to monitor the power stability. The other beam is converged to the sample by the lens 2. The sample is placed at the stepper motor to scan along the optical axis *z*-direction. The light passes through the sample to induce the nonlinear refraction and absorption, which is called as transmitted *z*-scan measurement. The transmitted light is divided into two parts by a beam splitter, one is totally collected as the open-aperture mode transmitted *z*-scan measurement, and the other is centrally collected as the closed-aperture mode transmitted *z*-scan measurement. The reflected light from the sample surface is also collected to be used as open-aperture mode reflected *z*-scan measurement. Finally, all collected signals are put into computer and conducted the data fitting process.

An In-doped Sb thin film is taken as an example, where the In-doped Sb thin film is generally used as active layer of the nonlinear optical information. The experimental results of both open-aperture mode transmitted and reflected *z*-scan measurements are shown in Fig. 3.14, where a 30 nm-thickness In-doped Sb deposited on glass substrate is examined.

Figure 3.14a depicts the open-aperture mode transmitted *z*-scan data at the total irradiation power of 6.2 mW, and Fig. 3.14b gives the open-aperture mode

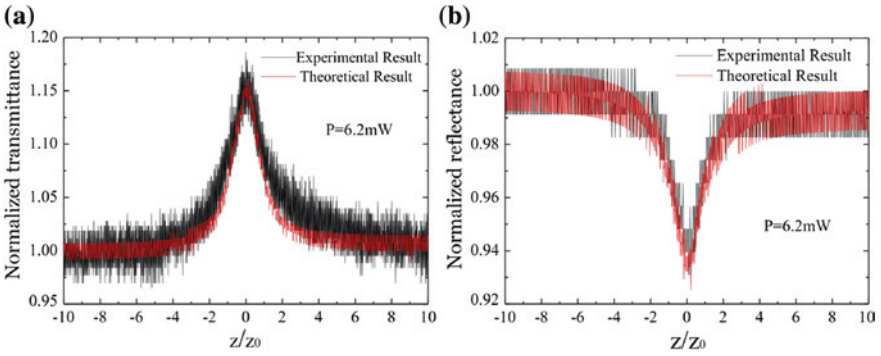

Fig. 3.14 The experimental results of nonlinear z-scan measurement. **a** Normalized open-aperture mode transmitted z-scan *curves*. $\beta = -5.3 \times 10^{-2}$ m/W is obtained by theoretical fitting. **b** Normalized open-aperture mode reflected z-scan *curves*. $\gamma = -1.0 \times 10^{-9}$ m^2/W is obtained by theoretical fitting. Reprinted from Ref. [9], Copyright 2013, with permission from Elsevier

reflected z-scan date. Firstly, $R(z)$ is experimentally obtained from the Fig. 3.14b. Secondly, the effective laser intensity inside the sample is calculated according to $R(z)$ and formula (3.63). Thirdly, the Fig. 3.14a is fitted with the red curve based on $I_{\text{eff}}(z, r)$ and formula (3.67), and the nonlinear absorption coefficient $\beta = -5.3 \times 10^{-2}$ m/W is obtained. Finally, the Fig. 3.14b is fitted with the red curve based on the given β value and formula (3.70), and the nonlinear refraction coefficient is $\gamma = -1.0 \times 10^{-9}$ m^2/W. Figure 3.14 shows that the fitting curves are very consistent with the measurement curves, which indicates that the theoretical model and experimental measurement system are reasonable and feasible in the real application.

3.4.3 The Case of Nanofilm Samples

3.4.3.1 Theoretical Model

For a very thin sample with a thickness $L <$ ~4.65/α, the Fresnel reflection loss is mainly determined by the internal multi-interference effect. The exiting electric field should be written as

$$E_e(z, r) = E_i(z, r)\tilde{t}(z, r) \exp\{-i[\Delta\phi(z, r)]\} \tag{3.71}$$

where $\tilde{t}(z, r)$ is the complex transmission coefficient of the incident electric field due to internal multi-interference, and $\Delta\phi(z, r)$ is the phase variation of the lens-like phase aperture induced by the nonlinear refraction. $\Delta\phi(z, r)$ is written as, again

$$\Delta\phi(z, r) = \frac{4\gamma P_0}{\lambda\alpha w^2(z)}\left[1 - \exp(-\alpha L)\right] \exp\left[-\frac{2r^2}{w^2(z)}\right] \tag{3.72}$$

with P_0 denoting the laser irradiation power. In general, for z-scan samples, the thin films are directly deposited on substrates. Therefore, a z-scan sample can be modeled as a two-interface layered system. The transmission for a two-interface layered system in normal incidence is

$$\tilde{t}(z,r) = \frac{t_{12}(z,r)t_{23}(z,r)\exp\left[i\frac{2\pi}{\lambda}\tilde{n}(z,r)L\right]}{1 + r_{12}(z,r)r_{23}(z,r)\exp\left[i\frac{4\pi}{\lambda}\tilde{n}(z,r)L\right]} \tag{3.73}$$

with

$$\begin{cases} t_{12}(z,r) = \frac{2n_1}{n_1 + \tilde{n}(z,r)} \\ t_{23}(z,r) = \frac{2\tilde{n}(z,r)}{\tilde{n}(z,r) + n_3} \\ r_{12}(z,r) = \frac{n_1 - \tilde{n}(z,r)}{n_1 + \tilde{n}(z,r)} \\ r_{23}(z,r) = \frac{\tilde{n}(z,r) - n_3}{\tilde{n}(z,r) + n_3} \end{cases} \tag{3.74}$$

and

$$\tilde{n}(z,r) = n(z,r) + iK(z,r) \tag{3.75}$$

$$K(z,r) = \frac{\lambda}{4\pi}\alpha(z,r) \tag{3.76}$$

where n_1 and n_2 are marked as the refractive index of air medium and substrate, respectively. The inclusion of the complex refractive index, the local nonlinear absorption, and the multi-interference in formula (3.71) enable us to calculate the exiting field distribution rigorously for nanofilm samples. The schematic of the exiting electric field is depicted in Fig. 3.15 [11]. Figure 3.15a, b illustrate that the internal multi-interference causes both intensity and phase variations in the exiting electric field and consequently brings about a distortion in the exiting electric field. Figure 3.15c shows the phase shift of the lens-like phase aperture that is induced by the Gaussian-like intensity profile of the incident light beam, which is also considered in general z-scan theory. In this case, the local refractive index as well as the local absorption coefficient becomes a function of the light intensity. Since the intensity varies with the sample position, the nonlinear refraction and nonlinear absorption cause variant transmittance.

Finally, the transmitted light is collected by a detection system at a distance L_1 away from the sample. At each point of the entrance pupil, the field is a superposition of the irradiated field from each point of the exiting field distribution on the sample surface.

Assuming there is a circular detector, the field at a point s from the center on the detection plane with the Fresnel–Kirchhoff diffraction integral (the geometrical relation is shown in Fig. 3.16) is [12]

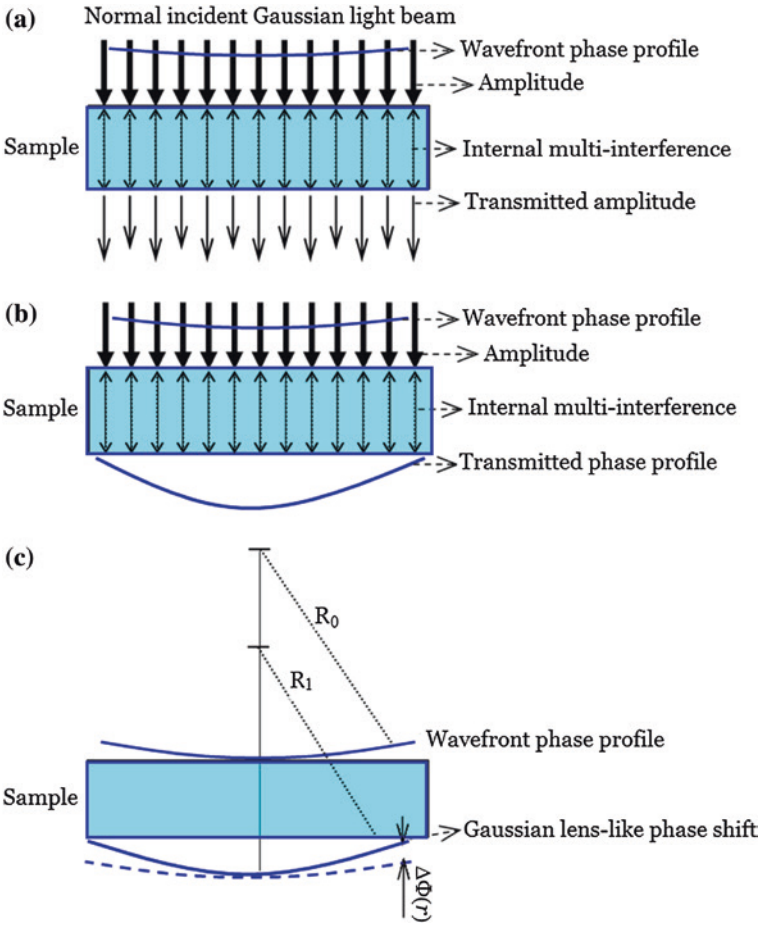

Fig. 3.15 Schematic showing that the electric field is modified by the nonlinear nanofilm samples in **a** electric field amplitude, **b** phase of complex transmission due to internal multi-interference, and **c** phase shift of the lens-like phase aperture

$$E_{\text{Detection}}(z,s) = -\frac{i}{\lambda}\int_0^{\infty}\int_0^{2\pi}\left(\frac{1}{2}+\frac{L_1}{2L_2}\right)E_e(z,r)\frac{\exp\left(i\frac{2\pi}{\lambda}L_2\right)}{L_2}r\mathrm{d}r\mathrm{d}\varphi \quad (3.77)$$

with

$$L_2 = \sqrt{L_1^2 + (r\cos\varphi - s)^2 + (r\sin\varphi)^2} \quad (3.78)$$

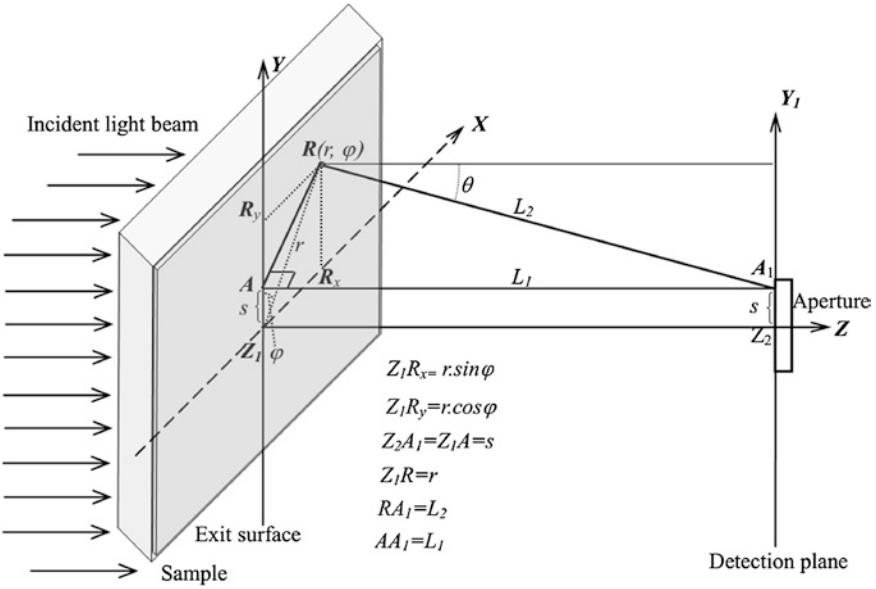

Fig. 3.16 The geometrical relation in the Fresnel–Kirchhoff diffraction integral

For circular detector of radius R_a, the total collected light power is

$$P(z) = 2\pi \int_0^{R_a} |E_{\text{Detection}}(z, s)|^2 s \mathrm{d}s \tag{3.79}$$

3.4.3.2 Numerical Simulation

The numerical tests are conducted by taking into account the multi-interference. As an example, the simulation parameters are chosen as follows: $\gamma = 5 \times 10^{-10}$ m^2/W, P = 500 mW, $w_0 = 5$ μm, $\lambda = 532$ nm, $\alpha_0 = 4 \times 10^7$ m^{-1}, $\beta = 1 \times 10^{-7}$ m/W, $n_{20} = 2$, $L_1 = 0.5$ m, $n_1 = 1$, $n_3 = 1.5$, and $R_a = 0.2$ mm. The numerical results of the simulation are presented in Figs. 3.17 and 3.18 [11]. Figure 3.17 presents three group of curves with respect to the focal point at the smallest beam size w_0. In Fig. 3.17a, the amplitude of the transmission $\tilde{t}(z = 0, r)$ is computed for a set of film thickness. The transmission is not a constant and changes with the radial coordinate r. The relative variation becomes increasingly significant when the thickness decreases. In Fig. 3.17b, c, the phase shifts are presented. Figure 3.17c contains the phase shift of the lens-like phase aperture, which is usually considered in conventional z-scan theory, whereas Fig. 3.17b contains the extra phase change due to the internal multi-interference. In light of the numerical test in Fig. 3.17, one expects that the effects from multi-interference and nonlinear absorption change the exiting electric field significantly.

The complete z-scan fitting curve is then obtained, and the results are presented in Fig. 3.18a. In Fig. 3.18a, the black solid line represents the fitting curve obtained with the present model calculation. The two principal contributions to

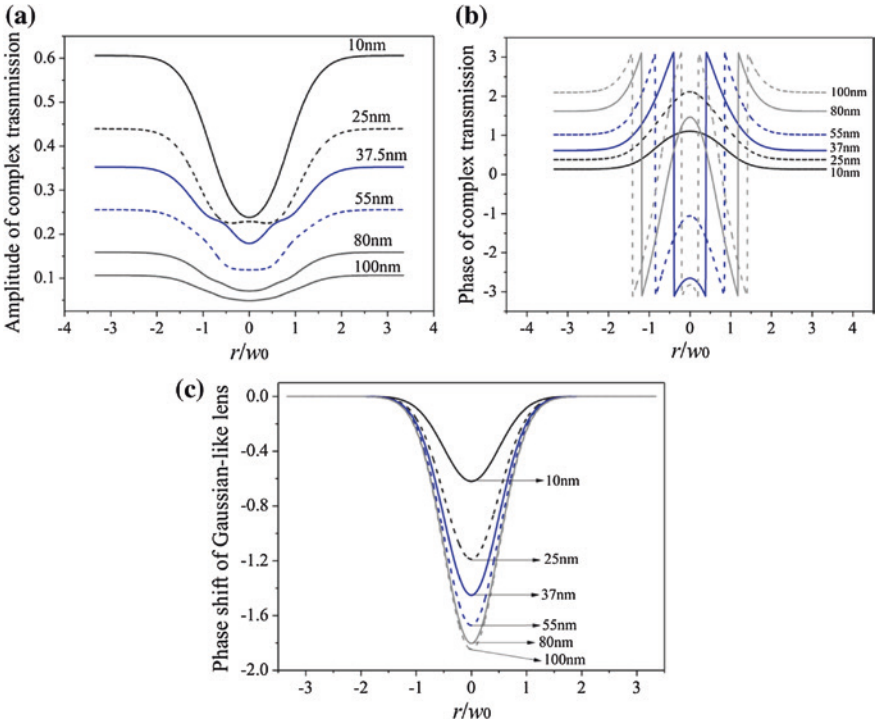

Fig. 3.17 Complex transmission and lens-like phase shift in radial direction for film thickness from 10 to 100 nm: **a** electric field amplitude, **b** phase of complex transmission due to internal multi-interference, and **c** phase shift of the lens-like phase aperture

the fitting curve are from the phase shift and the complex transmission. In order to have a better understanding and comparison, the phase shift contribution curve is plotted with the intensity scale $\times 10^{-4}$, i.e., the right-hand side y-axis in Fig. 3.18a, while the complex transmission contribution is plotted with the intensity scale $\times 10^{-6}$, i.e., the left-hand side y-axis in Fig. 3.18a.

Note that the phase shift contribution curve in Fig. 3.18a actually represents the fitting curve in previous z-scan calculation in which the transmission is usually assumed to be constant. Therefore, the curves in Fig. 3.18a provide a direct comparison between the present model and the conventional z-scan calculation. One realizes that the fitting curve from the present model calculation deviates significantly from the curve of conventional z-scan theory, both in magnitude and in shape, and the reason for the deviation stems from the inclusion of the internal multi-interference and nonlinear absorption into the complex transmission. In order to have a close look at the fitting curves and make an analysis on the stability of the numerical process, the fitting curves of Fig. 3.18a are amplified by a factor of 10, i.e., the discrete step in the numerical process is changed from $dz = 0.05$ to $dz = 0.005$. The amplified curves are presented in Fig. 3.18b. One can see that the numerical process appears to be stable near the focal area. Therefore, provided enough discrete points are taken in the area, the numerical integrations in formulas (3.77) and (3.79) are well convergent and stable.

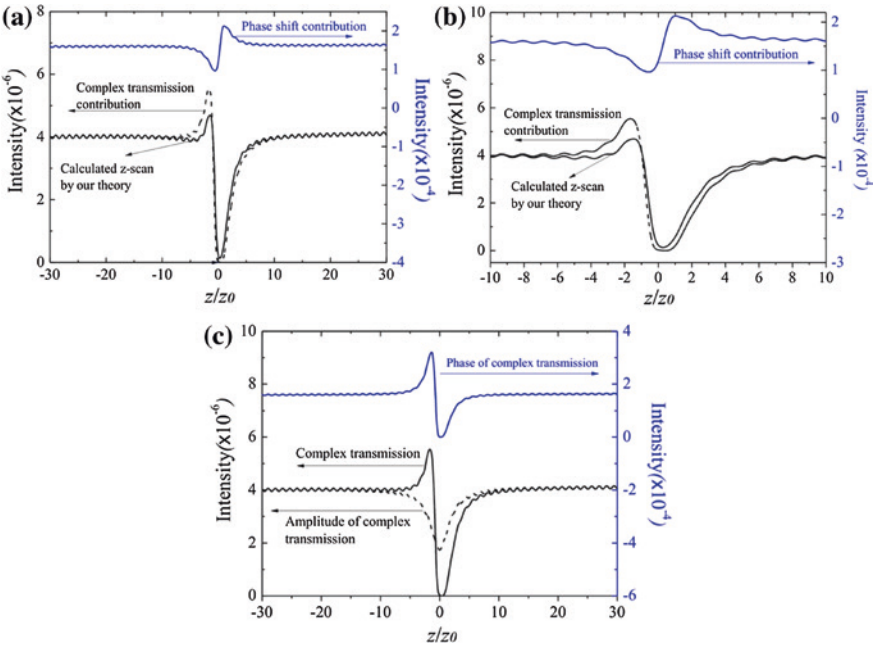

Fig. 3.18 Numerical results for $L = 80$ nm. **a** The fitting curve as well as the contribution from the complex transmission and the lens-like phase shift (conventional z-scan) contribution, **b** 10 times amplified view of the fitting curve in **a**, **c** the complex transmission curve as well as the contributions from the amplitude and phase of the complex transmission

Furthermore, within the complex transmission curve, there are contributions from the amplitude variation and the phase change. In order to further identify these two contributions, additional numerical calculations for the same fitting curve are given in Fig. 3.18c, it is clear that both terms contribute significantly to the fitting curve. In light of the numerical results in Fig. 3.18, one concludes that the effects of multi-interference and nonlinear absorption do indeed change the exiting electric field significantly for nanofilms. One notices that some small oscillations occur in the curves of Fig. 3.18. The oscillations stem from three sources in the formulas: (1) the $\exp\left[\frac{ikr^2}{2R(z)}\right]$ term in the incident Gaussian wave in formula (3.1); (2) the phase aperture $\Delta\phi(z, r)$ in formula (3.71) and (3.72); and (3) the axial dependence of the beam size $w(z) = w_0\sqrt{1 + z^2/z_0^2}$.

3.5 Influence of Feedback Light on z-scan Measurement

The laser device is one of the critical components, and the stability of laser power is very important to accurately get the nonlinear index. Generally speaking, the gas laser devices are stable compared with the solid-state laser devices and semiconductor laser

diode (LD) devices; thus, the gas laser devices are generally adopted in z-scan setup. However, in recent years, with the development of semiconductor technology, the semiconductor laser devices and semiconductor laser pump solid-state laser devices have been widely applied in the z-scan experiments because of small size, compactness, long service life, cheapness, and being modulated and integrated.

In z-scan measurement, the optical beam is normally incident on the sample surface, which causes the reflected light from sample surface to go back to the resonant cavity of LD. The output stability of LD is very sensitive to the feedback light; thus, the feedback light leads to the instability of the laser output [13, 14] and has an ill-effect on the accurate testing of the nonlinear index. That is, the laser is also reflected back to LD inevitably, which causes the instability of laser power. The instability causes a bad influence on the z-scan measurement. It is well known that the feedback light from sample surface, which is called as external cavity feedback, can increase the reflectivity of LD inner cavity surface [15, 16], reduce the output coupling, and thus result in lower output power [17]. In z-scan measurement, the feedback light has large influence on nonlinear index measurement results because the intensity changes as the sample moves along the z-axis.

3.5.1 *Influence of Feedback Light on Semiconductor Laser Devices [18]*

3.5.1.1 Inner Cavity Characteristics of Laser Devices

Inside LD, the medium junction interface forms a Fabry–Perot (F–P) resonator with two cleavage planes whose reflectivities are R_1 and R_2, respectively, as shown in Fig. 3.19. P_1 is laser power from cleavage plane R_1, and P_1' is the laser power reflected by cleavage plane R_2. The F–P resonator can be considered as an inner cavity of LD. The light reflected back and forth inside the inner cavity is amplified, which causes the laser output with a power of P_2.

In inner cavity, the light energy loss includes free-carrier absorption loss α_{fc}, diffraction loss α_{diff}, and the output end loss $\alpha_T = (1/2l)\ln[1/(R_1R_2)]$, where l is the length of the F–P resonator. The total loss α is:

$$\alpha = \alpha_{\mathrm{fc}} + \alpha_{\mathrm{diff}} + \alpha_T = \alpha_0 + \frac{1}{2l}\ln\frac{1}{R_1R_2} \tag{3.80}$$

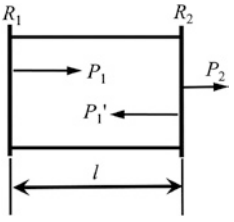

Fig. 3.19 Inner cavity scheme of LD

where $\alpha_0 = \alpha_{fc} + \alpha_{diff}$. When the balance between gain and loss is reached, the threshold gain g_{th} equals to the threshold loss, which can be written as:

$$g_{th} = \alpha_0 + \frac{1}{2l} \ln \frac{1}{R_1 R_2} \tag{3.81}$$

The threshold current is proportional to the threshold gain, $I_{th} = C' g_{th}$, and the output power P_2 of LD can be written as follows:

$$P_2 = (1 - R_2)\frac{C}{g_{th}}(I - I_{th}) = (1 - R_2)\left(\frac{CI}{\alpha_0 + \frac{1}{2l} \ln \frac{1}{R_1 R_2}} - C'\right) \tag{3.82}$$

where C and C' are constants only related with semiconductor LD itself, and I is the injected current. Formula (3.8) shows that the output power of LD is positive proportional to the injected current I, and there is laser output only when $I > I_{th}$.

3.5.1.2 Influence of External Feedback Light on Output Characteristics of Laser Devices

In real applications of LD, there is usually external feedback light into LD, and the feedback light inevitably has an influence on the output stability of LD. Figure 3.20 gives the schematics of influence of external feedback light on LD output power, where the external feedback light path can be considered as external cavity. In Fig. 3.20a, R_3 is reflectivity of external cavity output mirror. P_2 and P_2' are the laser powers irradiating on and reflected from external cavity mirror R_3, respectively. Generally speaking, the length of the inner cavity l is in submillimeter magnitude while the external cavity length L_{ex} is in decimeter magnitude, thus the inner cavity length l can be ignored compared with external cavity length L_{ex}. According to the Fresnel reflection, one can get [17]:

$$P_1' = R_2 P_1 + (1 - R_2) P_2' \tag{3.83}$$

$$P_2 = (1 - R_2) P_1 + R_2 P_2' \tag{3.84}$$

$$P_2' = R_3 P_2 \tag{3.85}$$

In order to simplify the analysis, the inner cavity and external cavity can be combined into an effective internal LD structure, which is presented in Fig. 3.20b, where the effective reflectivity R_{eff} of the output plane of internal LD structure can be written as:

$$R_{eff} = \frac{P_1'}{P_1} = \frac{R_2 + R_3 - 2R_2 R_3}{1 - R_2 R_3} \tag{3.86}$$

Figure 3.21a gives the dependences of R_{eff} on R_2 and R_3. When R_2 is close to 1, R_{eff} basically remains unchanged with R_3. However, when R_2 is much smaller than 1, R_3 has much influence on R_{eff}, and the larger the R_3 is, the faster the growth rate of R_{eff} is

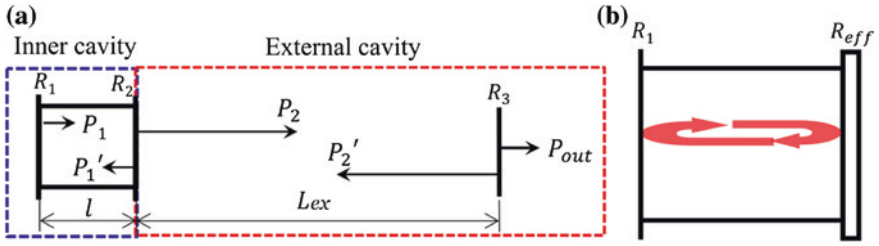

Fig. 3.20 Schematic of far-field external cavity feedback influence on LD. **a** Detailed external feedback structure, and **b** effective internal LD structure [18]

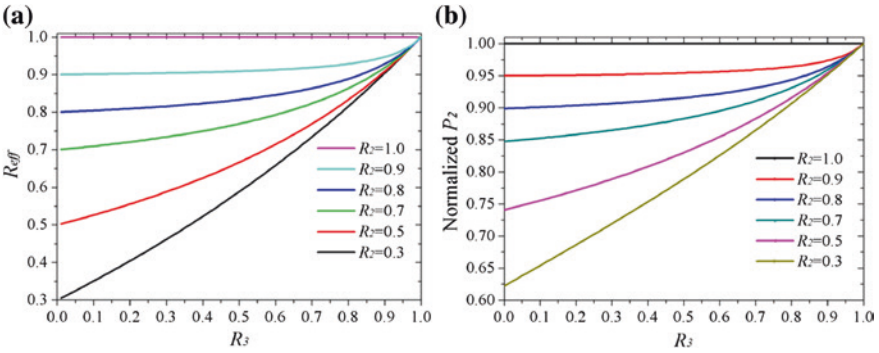

Fig. 3.21 Relationship among P_2, R_{eff}, and R_3 for different R_2, **a** dependence of R_{eff} on R_3, and **b** dependence of P_2 on R_3 [18]

Replacing R_2 of formula (3.81) with R_{eff}, the threshold gain g_{th} of effective internal LD structure can be obtained as follows:

$$g_{\text{th}} = \alpha_0 + \ln \frac{1}{R_1} \frac{1 - R_2 R_3}{R_2 + R_3 - 2R_2 R_3} \tag{3.87}$$

One can see from formula (3.87) that the threshold gain g_{th} is actually determined by R_2 and R_3. R_2 is fixed for a given LD, and R_3 is the reflectivity of the external cavity. Thus, the external cavity characteristic has an important effect on the threshold gain g_{th}. Substituting formula (3.87) into formula (3.82), one can get the influence of external cavity feedback light on the laser output of LD as follows,

$$P_2 = (1 - R_2)\left(\frac{CI}{\alpha_0 + \ln \frac{1}{R_1} \frac{1 - R_2 R_3}{R_2 + R_3 - 2R_2 R_3}} - C' \right) \tag{3.88}$$

Figure 3.21b gives the dependence of output power P_2 on the reflectivity R_3 of external cavity output mirror according to formula (3.88). It can be seen that, in order to get a stable output of LD, one needs to decrease R_3 as much as possible in experiment.

3.5.1.3 Dynamic Characteristics of External Feedback in *z*-scan Measurement

In *z*-scan measurement, the sample always moves along the optical axis. Figure 3.1 gives the scheme of traditional *z*-scan measurement, where the incident light is from LD and irradiates on the sample after passing through the lens. The sample moves near the focal plane of the lens from $-z$ to $+z$ direction along the optical axis, where $-z$ and $+z$ means that the sample is on the left and right side of the focal plane, respectively. When the sample moves along the optical axis, the nonlinearity of the sample causes the change of light propagation. Thus, the power detected by the detector varies with sample position z and the nonlinear index can be calculated accordingly. However, in the measurement, some part of the incident light is reflected back to LD by the sample, which causes the instability of LD. The intensity of reflected light into LD changes with sample position z.

Figure 3.22 gives an illustration of the reflected light change with the sample position z, where the propagation locus of the edge line of the laser beam is schematically

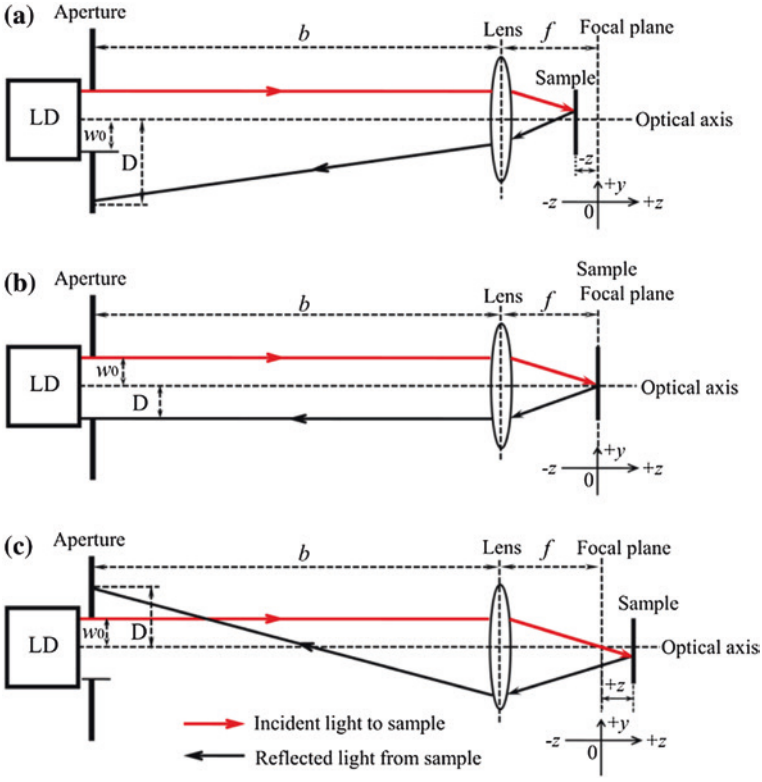

Fig. 3.22 Geometric simplification of laser propagation at different sample position, sample **a** before the focal plane, **b** on the focal plane, and **c** after the focal plane [18]

presented while sample moves along the z-direction. w_0 is the aperture radius of LD, b is the distance from LD to the lens, f is the focal length of the lens, and z is the distance between sample and the focal plane of lens. In z scan measurement, if the sample is on the focal plane, as shown in Fig. 3.22b, all reflected light from sample should be collected by the lens and go through the LD aperture and back to LD. However, if the sample is on the left of focal plane (that is, $-z$ direction), only partial reflected light goes back to LD because of the limited aperture size of LD, as shown in Fig. 3.22a. If the sample situates on the right side of the focal plane (that is, $+z$-direction), the reflected light is focused before entering LD, similar to Fig. 3.22a, only part of the reflected light goes back to LD, as shown in Fig. 3.22c.

In z-scan measurement, a collimated Gaussian laser beam is emitted from the LD, and the electric field intensity is written as:

$$E(r) = E_0 \exp\left(-\frac{r^2}{w_0^2}\right) \tag{3.89}$$

where E_0 is the electric field amplitude at $r = 0$. w_0 is the beam radius, which is the same as the aperture radius of LD. By complex mathematical operation, the reflected beam radius in LD plane can be written as:

$$D = \frac{2w_0}{f^2}|z|\left(b - \frac{f(f+2|z|)}{2|z|}\right) \tag{3.90}$$

Integrating the laser intensity from the center to aperture radius w_0, the laser power actually reflected back into LD P_{reflect} is:

$$P_{\text{reflect}} = 2\pi E_0^2 \frac{w_0^2}{D^2} \int_0^{w_0} \exp\left(-\frac{2r^2}{D^2}\right) r\mathrm{d}r = \frac{1}{2}\pi w_0^2 E_0^2 \left[1 - \exp\left(-\frac{2w_0^2}{D^2}\right)\right] \tag{3.91}$$

Substituting formula (3.90) into formula (3.91), one can get,

$$P_{reflect} = \frac{1}{2}\pi w_0^2 E_0^2 \left\{1 - \exp\left(-\frac{2f^4}{[2bz - f(f+2z)]^2}\right)\right\} \tag{3.92}$$

Formula (3.92) shows that the power of reflected light into LD changes with the sample position z. The dependence of P_{reflect} on sample position z can be seen by taking $b = 500$ nm and $f = 12$ mm as an example. Figure 3.23 presents the normalized power P_{reflect} of reflected light into LD. The P_{reflect} dramatically changes when the sample moves backward and forward near the focal plane.

According to Fig. 3.20, in the z-scan, the sample can be considered as an external feedback mirror, and the reflectivity is R_3. Thus, $P_{\text{reflect}} = R_3 P_{\text{in}}$, where $P_{\text{in}} = (1/2)\pi w_0^2 E_0^2$ is incident light power regardless of propagation energy loss. From formula (3.92), one can get

$$R_3 = 1 - \exp\left(-\frac{2f^4}{\left[2bz - f(f+2z)\right]^2}\right) \tag{3.93}$$

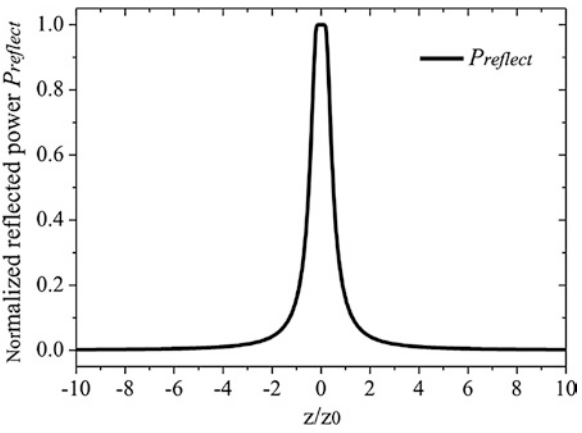

Fig. 3.23 Normalized reflected light power P_{reflect} in z-scan measurement

Formula (3.93) indicates that the R_3 changes with the sample position z. According to formula (3.86), one can use the effective reflectivity R_{eff} to analyze feedback light influence on the z-scan measurement.

3.5.1.4 Feedback Influence on Laser Device Output in z-scan Measurement

In z-scan measurement, the incident light is reflected back to LD by the sample, and the reflected light intensity is a function of sample position z, which causes the instability of LD and induces pseudo-nonlinearity effect. In order to monitor the instability of LD and detect the pseudo-nonlinearity effect, three detectors are put into the optical path of traditional z-scan setup (shown in Fig. 3.24). Detector 1 is used to collect the feedback light from sample, detector 2 is used to detect the noise, and detector 3 collects the transmitted light through the sample.

In the experiment, LD laser beam with a wavelength of 658 nm is used as light source in z-scan setup, a BK7 glass of about 1 mm thickness is used as the sample. It is well known that BK7 glass has no nonlinearity at low power irradiation. However, in open-aperture mode z-scan experiment, the reflected light R_3 from BK7 glass presents false nonlinear saturation absorption effect, as shown in Fig. 3.25, where z_0 is the Rayleigh length of the lens. P_{noise} from LD output power is almost consistent with R_3; thus, one can see that the LD output is unstable and is a function of sample position z. The change tendency of R_3 and P_{noise} with sample position z is similar to that of P_{reflect} in Fig. 3.23. It indicates that the instability of LD is closely related with external cavity reflectivity R_3, and the feedback light from sample causes LD to become unstable. The larger the external cavity reflectivity R_3 is, the larger the reflectivity R_{eff} is, and the larger the LD instability is.

Besides, Fig. 3.26a gives the change P_{noise} with sample position z at different LD output power P. It is found that the lower the LD output power, the larger the feedback light intensity is. The P_{noise} even goes up to 260 % when LD power is

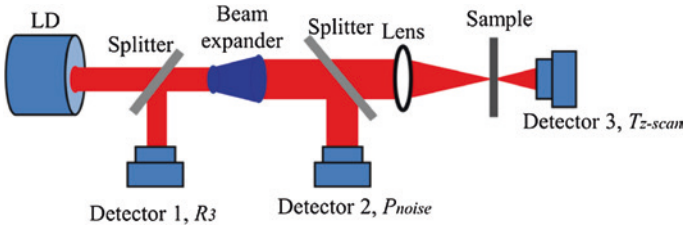

Fig. 3.24 Improved scheme of z-scan setup [18]

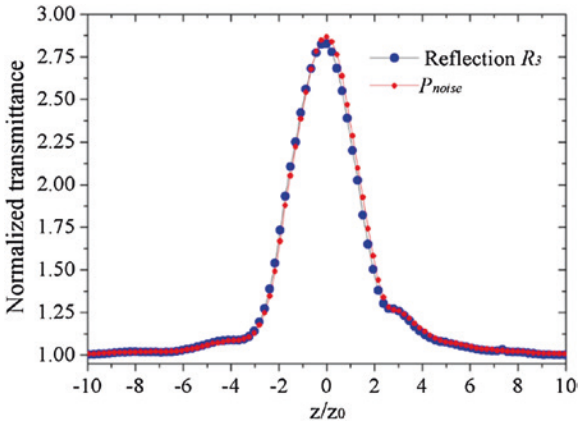

Fig. 3.25 Dependences of R_3 and P_{noise} on sample position in z-scan measurement for BK7 glass

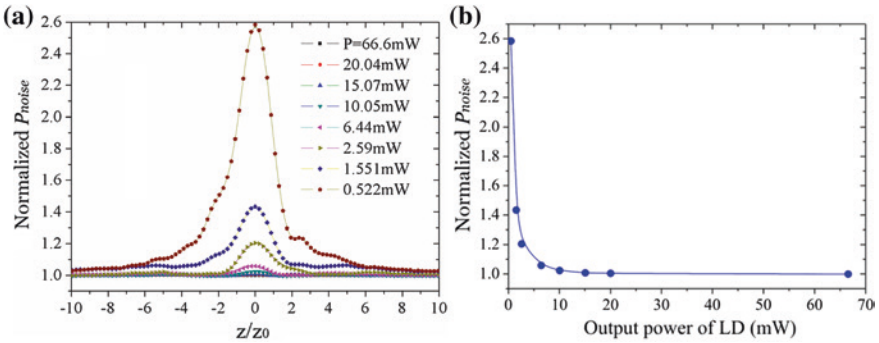

Fig. 3.26 Relationship among LD output power, P_{noise} and sample position z. **a** Normalized P_{noise} in z-scan, and **b** dependence of P_{noise} on LD power [18]

0.522 mW. One further plots the dependence of P_{noise} on the laser power of LD, as shown in Fig. 3.26b, the P_{noise} decreases exponentially with LD power increasing, that is, the influence of feedback light on LD output instability can be ignored when LD power exceeds to certain value, which is useful for eliminating the reflected light influence on LD in z-scan measurement.

3.5.2 Elimination of Feedback Light Influence on z-scan Measurement

Both theoretical analysis and experimental data show that, in z-scan measurement, the feedback light from sample inevitably has a large influence on nonlinearity measurement, which can induce pseudo-nonlinearity effect. Thus, some experimental methods should be taken to decrease and eliminate the feedback light influence on z-scan measurement.

3.5.2.1 Decreasing Feedback Light Influence by Adding an Attenuator in z-scan Setup

Figure 3.26b shows that the influence of feedback light on LD output instability can be reduced and ignored when the LD power exceeds to certain value; thus in the z-scan setup, one can place an attenuator after the LD, as shown in Fig. 3.27.

Figure 3.28 gives a detailed illustration, where R_1 and R_2 are the reflectivity of cleavage plane 1 and cleavage plane 2, respectively. R_{att} and R_s are reflectivity of attenuator and sample, respectively. The attenuator is a critical element for decreasing the feedback light because of its distinctive advantages. Assuming that the reflectivity of attenuator is $R_{att} = 0.9$, the reflectivity of sample $R_s = 0.3$, and the nonlinearity excitation power of sample is required to be P. Regardless of energy loss, the LD output should be set to be P. When there is without attenuator, the feedback light entering LD is $0.3P$, that is, about 30 % light energy goes back to LD. While when there is an attenuator, the LD output power should be set to be $10P$ due to $R_{att} = 0.9$; thus, transmittance $T_{att} = 0.1$, and the feedback light going into LD should be $0.03P$, that is, only about 0.3 % light energy is back to

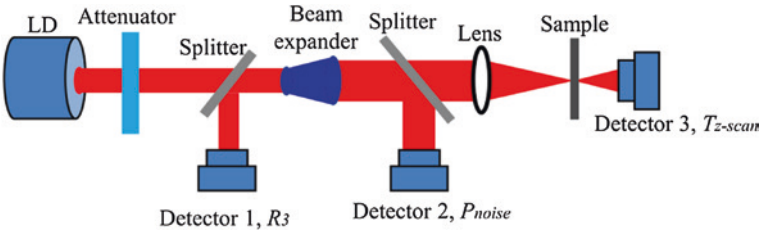

Fig. 3.27 z-scan setup with attenuator

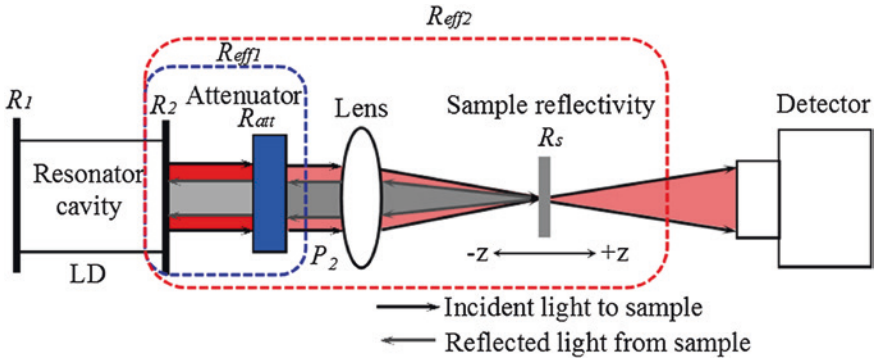

Fig. 3.28 Detailed illustration of effective LD output with an attenuator [18]

LD. By Comparison, one can see that the attenuator can reduce the feedback light to only 1 % that of without attenuator. Besides, as shown in Fig. 3.26b, the larger the LD output power is, the less the influence from feedback light on LD is, thus placing an attenuator can decrease the ill-influence from feedback light in z-scan measurement.

In Fig. 3.28, LD and the attenuator can be considered as a new F–P resonator with an effective cavity end reflectivity R_{eff1}, as shown in the blue dashed box. According to formula (3.86), the effective cavity end reflectivity can be written as:

$$R_{\text{eff1}} = \frac{R_2 + R_{\text{att}} - 2R_2R_{\text{att}}}{1 - R_2R_{\text{att}}} \tag{3.94}$$

According to Fig. 3.20, taking R_s into consideration, the new F–P resonator and the sample together make up a new effective internal LD structure with a reflectivity R_{eff2}, as shown in the red dashed box of the Fig. 3.28. Replacing R_2, R_3 by R_{eff1} and R_s, respectively, in formula (3.86) the reflectivity of new effective internal LD structure is,

$$R_{\text{eff2}} = \frac{R_2 + R_{\text{att}} + R_s - 2R_2R_{\text{att}} - 2R_2R_s - 2R_{\text{att}}R_s + 3R_2R_{\text{att}}R_s}{1 - R_2R_{\text{att}} - R_2R_s - R_{\text{att}}R_s + 2R_2R_{\text{att}}R_s} \tag{3.95}$$

Figure 3.29a presents the dependence of R_{eff2} on the sample reflectivity R_s at $R_2 = 0.3$. One can see that for a fixed R_2, the larger the R_{att} is, the more stable the R_{eff2} becomes. That is, the stability of LD can be improved by placing an attenuator after LD. If $R_{\text{att}} = 0.9$, the R_{eff2} is almost unchanged at $R_s < 0.5$, which is easily met in z-scan because the sample is generally transparent or semi-transparent. It means that the external feedback influence caused by sample can be greatly reduced or even ignored in z-scan measurement.

According to formula (3.88), the output power P_2 of the new effective internal LD structure can be written as:

$$P_2 = (1 - R_{\text{eff1}})\left(\frac{CI}{\alpha_0 + \frac{1}{2l}\ln\frac{1}{R_1R_{\text{eff2}}}} - C'\right) \tag{3.96}$$

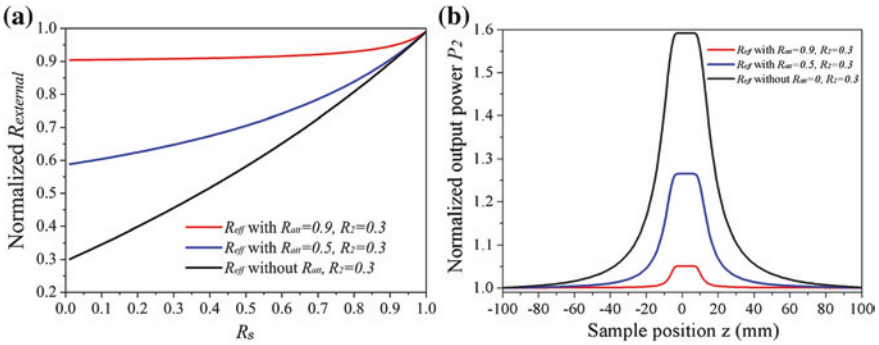

Fig. 3.29 Relations of R_{eff2} and P_2 with different attenuation in z-scan when $R_2 = 0.3$. **a** Dependence of R_{eff} on R_3 and **b** dependence of P_2 on z position [18]

Substituting formulas (3.94) and (3.95) into formula (3.96), one can get the relationship of P_2 and the sample position z, as shown in Fig. 3.29b, the larger the R_{att} is, the more stable the LD output power P_2 becomes. Compared with without attenuator ($R_{\text{att}} = 0$), the instability of LD output P_2 is reduced to 0.83 %, that is, the stability of LD output is greatly improved in the z-scan measurement.

Figure 3.30 gives the open-aperture mode z-scan measurement results with the BK7 glass sample for different R_{att} by using the setup in Fig. 3.27, where the LD output power is set at $P = 1$ mW. One can see from the black dotted line that the interface reflectance from glass sample causes a strong false nonlinear saturation absorption effect without attenuator in the optical path. Comparison among the curves of Fig. 3.30 shows that the attenuator can obviously decrease the false nonlinear saturation absorption effect, and the false nonlinear saturation absorption effect decreases as the reflectivity of attenuator R_{att} increases. The false nonlinear saturation absorption effect is almost eliminated at $R_{\text{att}} = 99.5$ %.

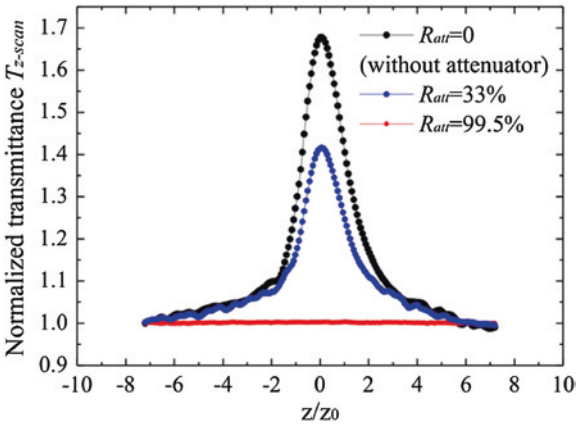

Fig. 3.30 Comparison of feedback light influence with different attenuation [18]

3.5.2.2 Decreasing the Influence from Feedback Light by Adding an Opto-isolator Unit

Here, another method to decrease and eliminate the influence from feedback light is given. An opto-isolator unit, which consists of a 1/2 wave plate, a polarized beam splitter (PBS), and a 1/4 wave plate, is put into z-scan setup, as shown in Fig. 3.31a. The laser beam passes through the 1/2 wave plate and becomes p-polarization light, and the p-polarization light travels through the PBS and 1/4 wave plate and becomes circular polarization light. The circular polarization light is then split into two beams, one is detected by the noise detector, which monitors the instability of LD, and the other is focused by the lens and irradiates on the sample to induce nonlinear effect. The sample moves through the focal plane of lens from $-z$ to $+z$ direction, and then, the transmitted light through sample is collected by detector 3. Besides, some part of the circularly polarized light reflected by the sample passes through the 1/4 wave plate again and becomes s-polarization. The s-polarization light is perpendicular to the incident p-polarization light and thus cannot pass through the PBS and go back to LD.

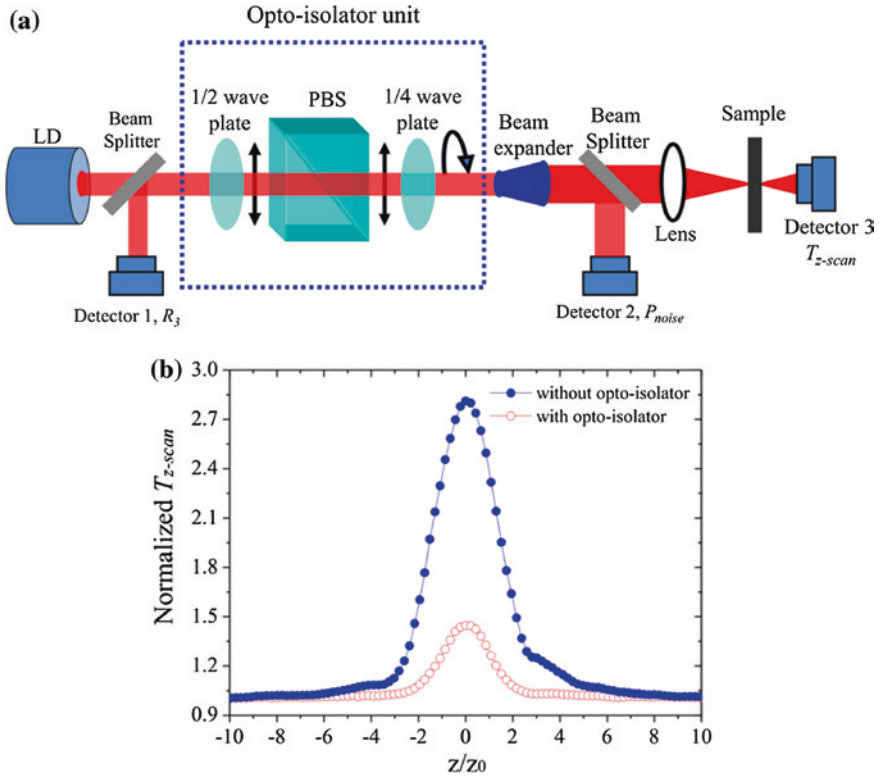

Fig. 3.31 Reduction of feedback light influence on z-scan measurement by opto-isolator unit, **a** z-scan setup, **b** z-scan measurement with and without opto-isolator unit at $P = 1.52$ mW [18]

The opto-isolator unit realizes the isolation of reflected light from sample, and the feedback light influence on z-scan measurement is reduced and even eliminated theoretically.

Figure 3.31b gives typical experimental results, where BK7 glass is used as sample. It can be found that, when without opto-isolator unit is put into optical path, the z-scan measurement of BK7 glass presents strong nonlinear saturation absorption effect, and the normalized peak of T_{zscan} is about 2.8. While the opto-isolator unit is inserted into the z-scan setup, the normalized peak of T_{zscan} is reduced to about 1.4. Thus, the opto-isolator unit is very useful for decreasing the influence of feedback light on z-scan measurement.

Figure 3.31 also shows that it is hard to thoroughly eliminate the influence of feedback light on z-scan measurement because in experiment a little leakage light inevitably feeds back to LD and has an ill-effect on measurement accuracy. Thus, in experimental setup of Fig. 3.31a, a noise detector 2 is inserted before lens. The detector 2 is used to monitor and record the instability of LD output P_{noise}. The nonlinear measurement accuracy can be further improved by numerical operation of $T_{\text{zscan}}/P_{\text{noise}}$. Figure 3.32 shows typical P_{noise} and T_{zscan} curves for BK7 glass sample. The T_{zscan} curve has a similar tendency to P_{noise} curve. $T_{\text{zscan}}/P_{\text{noise}}$ curve (marked in blue color) is also plotted accordingly. The T_{zscan} peak of ~1.45 is reduced to $T_{\text{zscan}}/P_{\text{noise}}$ of ~1.00133, which is very close to 1. That is, the z-scan measurement of BK7 glass sample is without nonlinear effect, and the false nonlinear absorption effect induced by the feedback light from sample is almost eliminated. Therefore, it is obvious that the numerical calculation of $T_{\text{zscan}}/P_{\text{noise}}$ can further decrease the influence of feedback light on z-scan measurement.

It is very important to characterize the nonlinear absorption and nonlinear refraction coefficients of thin film samples, especially the semi-transparent semiconductor thin films. The z-scan is one of the most simplest and sensitive methods and has been applied in real practice. The basic principle of z-scan is described, and the occurrence of pseudo-nonlinearity effect due to the incident angle change with sample movement at the z-scan measurement is theoretically

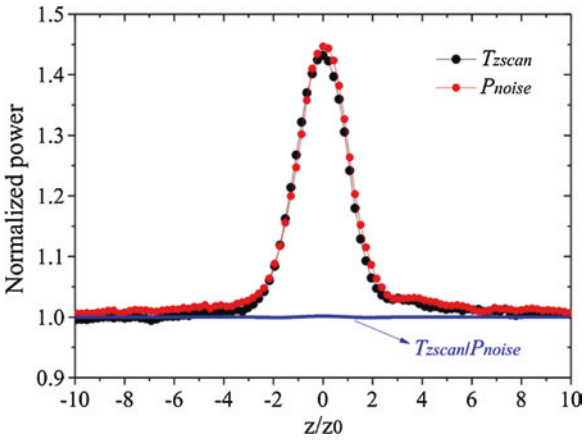

Fig. 3.32 Numerical operation for further eliminating the feedback light influence

and experimentally analyzed. The influences from surface reflection on z-scan measurement (such as instability of laser devices and variation of reflection loss with different z-scan position) are analyzed, and some methods for eliminating the ill-effects are given accordingly.

References

1. M. Sheik-Bahae, A.A. Said, T.H. Wei, D.J. Hagan, E.W. Van Stryland, Sensitive measurement of optical nonlinearities using a single beam. IEEE J. Quantum Electron. **26**, 760–769 (1990)
2. M. Sheik-Bahae, A.A. Said, E.W. Van Stryland, High sensitivity single beam n_2 measurement. Opt. Lett. **14**, 955–958 (1989)
3. J. Wang, M. Sheik-Bahae, A.A. Said, D.J. Hagan, E.W. Van Stryland, Time-resolved z-scan measurements of optical nonlinearities. J. Opt. Soc. Am. B **11**, 1009–1017 (1994)
4. G. Tsigaridas, M. Fakis, I. Polyzos, P. Persephonis, V. Giannetas, Z-scan technique through beam radius measurements. Appl. Phys. B **76**, 83–86 (2003)
5. J. Liu, J. Wei, Single beam two-mode z-scan equipment with real-time observation and tunable power and pulse width. China Invention Patent: ZL 20091 0054823.1 (2009)
6. X. Ma, J. Wei, Two-beam multi-functional z-scan optical nonlinearity measurement equipment. China Invention Patent: ZL 20101 0139025.1 (2010)
7. R. Wang, J. Wei, M. Xiao, Pseudo-nonlinear absorption in z-scan measurement. J. Opt. **15**, 025204 (2013)
8. R.P. Riesz, R. Simon, Reflection of a Gaussian beam from a dielectric slab. J. Opt. Soc. Am. A **2**, 1809–1817 (1985)
9. R. Wang, J. Wei, Accurate determination of optical nonlinear absorption coefficients by the combination of reflection and transmission open-aperture z-scan measurements. Opt. Commun. **297**, 121–124 (2013)
10. M. Born, E. Wolf, *Principle of Optics (the seventh version)* (Cambridge University, Cambridge 1999)
11. J. Wei, M. Xiao, A z-scan model for optical nonlinear nanometric films. J. Opt. A: Pure Appl. Opt. **10**, 115102 (2008)
12. J. Wei, in *Nonlinear Performance and Characterization Methods in Optics*, ed. by J. Wei. Thermo-optical Coefficient Measurement with Laser Pulse Width Tunable z-scan Method (Nova, New York, 2013)
13. R. Lang, K. Kobayashi, External optical feedback effects on semi- conductor injection laser properties. IEEE J. Quantum Electron. **QE-16**, 347–355 (1980)
14. B. Tromborg, J.H. Osmundsen, H. Olesen, Stability analysis for a semiconductor laser in an external cavity. IEEE J. Quantum Electron. **QE-20**, 1023–1032 (1984)
15. J. Ge, J. Chen, A. Hermerschmidt, H.J. Eichler, Achieving single-lobed far-field patterns of broad area laser diode with external cavity feedback. Chin. Opt. Lett. **1**, 334–336 (2003)
16. Y. Guo, Y. Wu, Y. Wang, Method to identify time delay of chaotic semiconductor laser with optical feedback. Chin. Opt. Lett. **10**, 061901 (2012)
17. C. Liu, J. Ge, J. Chen, Influence of external cavity feedback on the oscillating characteristics of a semiconductor laser. Acta. Phys. Sin. **55**, 5211–5215 (2006). (in Chinese)
18. H. Yan, J. Wei, False nonlinear effect in z-scan measurement based on semiconductor laser devices: theory and experiments. Photon. Res. **2**(2), 51–58 (2014)

Chapter 4
Optical Nonlinear Absorption and Refraction of Semiconductor Thin Films

4.1 Introduction

In nonlinear super-resolution nano-optics, in order to obtain below diffraction-limited spot, the nonlinear absorption and refraction coefficients of samples are required to be large. The semiconductor thin films generally possess large nonlinear coefficients due to the manipulations from free-carriers, gap shrinkage, and thermal effect. In this chapter, the nonlinear physical processes and characteristics of some typical semiconductor thin films are described.

4.2 Theoretical Basis

4.2.1 Two-Band Model for Free-Carriers-Induced Nonlinear Effects

4.2.1.1 Single-Photon-Absorption-Induced Free-Carrier Nonlinear Effects

(1) **Nonlinear absorption and refraction based on the rate formulas** [1]

In most semiconductors, the energy band consists of conduction band and valence band, which is usually called as two-band model. Figure 4.1 gives an ideal schematic of two-band model.

At the excitation of laser pulse with a light intensity of I and frequency of ω, the electrons at the ground state S_0 will be excited to the excited state S_1 with an absorption cross section of σ_{01}, and the holes are left in the valence band. The electrons and holes are called as free-carriers. The electrons can relax to the ground state through spontaneous radiation, non-radiative transition, or the non-radiative recombination, and the relaxation time is τ_{21}.

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_4

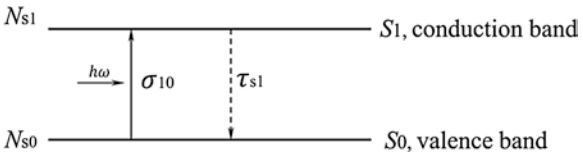

Fig. 4.1 Two-band model of semiconductors

Assuming that the numbers of electrons at ground state S_0 and excited state S_1 are N_{s0} and N_{s1}, respectively, and neglecting the diffusion of free-carriers, the rate formula of free-carrier numbers can be expressed as

$$\frac{\mathrm{d}N_{s0}}{\mathrm{d}t} = -\frac{\sigma_{10}N_{s0}I}{\hbar\omega} + \frac{N_{s1}}{\tau_{S1}} \tag{4.1}$$

Assuming that the number of free-carriers is $N = N_{s0} + N_{s1}$, the N can be approximately calculated as

$$N = \frac{I\tau\alpha(\hbar\omega)}{\hbar\omega} \tag{4.2}$$

The α is absorption coefficient. α and τ are dependent on N, and formula (4.2) needs to be solved self-consistently. When the free-carriers are mostly photo-generated free-carriers, the number of electrons equals to that of holes, and the relaxation time can be approximately written by a polynomial

$$\frac{1}{\tau_{21}(N)} = A + BN + CN^2 \tag{4.3}$$

In the right side of formula (4.3), the first term is from non-radiative recombination through defect states in the band gap, which is called as Shockley-read-hall recombination. Shockley-read-hall recombination has particularly simple form when empty defect levels capture electrons much more rapidly than holes or vice versa. The second term comes from radiative transition, and the third term is from Auger recombination.

When the laser spot has a non-uniform profile, an effective intensity should be used in formula (4.2). For large laser spot irradiation, the photo-generated free-carrier diffusion can be neglected, and the effective laser intensity can be written as

$$I_{\mathrm{eff}} = \frac{P}{\pi(r_{1/e^2})^2} \tag{4.4}$$

where r_{1/e^2} is the radius at $1/e^2$ intensity of the spot.

For small laser spot irradiation, the photo-generated free-carrier diffusion out of spot should be considered. If the photo-generated free-carriers diffuse only laterally, the laser spot has a Gaussian profile, and the recombination time is approximately constant, then the effective laser intensity is written as

$$I_{\mathrm{eff}} = \frac{P}{\pi(r_{1/e^2})^2 + 4\pi\,\mathrm{d}\tau} \tag{4.5}$$

At steady-state condition, that is, $t_p > \tau_{s1}$, where t_p is laser pulse width, the $\frac{\partial N_{s0}}{\partial t} = 0$, and neglecting the diffusion term; thus, formula (4.1) becomes

$$\frac{\sigma_{10} N_{\text{so}} I}{\hbar\omega} = \frac{Ns_1}{\tau_{S1}} \tag{4.6}$$

Setting

$$I_{\text{sat}} = \frac{\hbar\omega}{\sigma_{10}\tau_{S1}} \tag{4.7}$$

Thus

$$\frac{I}{I_{\text{sat}}} = \frac{Ns_1}{N_{\text{so}}} \Rightarrow \frac{I}{I_{\text{sat}}} + 1 = \frac{Ns_1 + N_{\text{so}}}{N_{s0}} \Rightarrow \frac{I}{I_{\text{sat}}} + 1 = \frac{N}{N_{s0}} \Rightarrow N_{s0} = \frac{N}{1 + \frac{I}{I_{\text{sat}}}} \tag{4.8}$$

Defining the absorption coefficient as $\alpha = \sigma_{10} N_{s0}$ and the linear absorption coefficient $\alpha_0 = \sigma_{10} N$, and according to formula (4.8), the α can be written as

$$\alpha = \frac{\alpha_0}{1 + \frac{I}{I_{\text{sat}}}} \tag{4.9}$$

It is well known that $\alpha(\hbar\omega, N)$ is directly related to photo-generated free-carrier density and can be also approximately modeled in the form of free-carrier density:

$$\alpha(\hbar\omega, N) = \frac{\alpha_0(\hbar\omega)}{1 + \frac{N}{N_{\text{sat}}(\hbar\omega)}} \tag{4.10}$$

where $N_{\text{sat}}(\hbar\omega)$ is saturation density. Better fits are often obtained by adding an unsaturation term to formula (4.10),

$$\alpha(\hbar\omega, N) = \frac{\alpha_0(\hbar\omega)}{1 + \frac{N}{N_{\text{sat}}(\hbar\omega)}} + \alpha_{\text{unsatur}}(\hbar\omega) \tag{4.11}$$

When the excitation is with short optical pulses, free-carrier density depends on light fluence F. Then, instead of formula (4.10), one may use

$$\alpha(\hbar\omega, F) = \frac{\alpha_0(\hbar\omega)}{1 + \frac{F}{F_{\text{sat}}(\hbar\omega)}} \tag{4.12}$$

$F_{\text{sat}}(\hbar\omega)$ is the saturation light fluence.

Near exciton absorption peaks, the absorption should be modeled with two saturation terms. The first term, which saturates rapidly, represents the excitonic contribution to the absorption. The second term represents slow saturation background absorption from energy bands.

$$\alpha(\hbar\omega, I) = \frac{\alpha_0^{\text{ex}}(\hbar\omega)}{1 + \frac{I}{I_{\text{sat}}^{\text{ex}}(\hbar\omega)}} + \frac{\alpha_0^{\text{bg}}(\hbar\omega)}{1 + \frac{I}{I_{\text{sat}}^{\text{bg}}(\hbar\omega)}} \tag{4.13}$$

Similarly, the nonlinear refractive index is modeled as

$$\Delta n(\hbar\omega, N) = n_s(\hbar\omega)\frac{N/N_{\text{sat}}(\hbar\omega)}{1 + N/N_{\text{sat}}(\hbar\omega)} \tag{4.14}$$

or

$$\Delta n(\hbar\omega, I) = n_s(\hbar\omega)\frac{I/I_{\text{sat}}(\hbar\omega)}{1 + I/I_{\text{sat}}(\hbar\omega)} \tag{4.15}$$

(2) **Band-filling model for direct band semiconductors**

In two-band model, the electrons at valence band (ground state) are excited to conduction band (excitation state) and the holes are left in the valence band, which is the origin of nonlinear saturation absorption. For direct band semiconductor with small band gap, the energy band can be dealt with parabolic two-band model, shown in Fig. 4.2. The photo-excited electrons fill into the conduction band and cause the nonlinear effect; thus, the optical nonlinearity of semiconductor from photo-excited carriers is called as band-filling nonlinearity.

In band-filling nonlinearity for III–VI semiconductors, the absorption coefficient at photon energy $\hbar\omega$ near the absorption edge is [2, 3]

$$\alpha(\hbar\omega) = \frac{8\sqrt{2}m^{1/2}e^2}{3\hbar^2 c}\left(\frac{\mu}{m}\right)^{3/2}\frac{1}{n_0}\frac{mP_M^2}{\hbar^2}\frac{\left(\hbar\omega - E_g\right)^{1/2}}{\hbar\omega}B(\hbar\omega), \tag{4.16}$$

where m is static electron mass, E_g band gap, and n_0 linear refractive index. P_M is the momentum matrix elements and defined by Kane as follows

$$P_M = -\left(\frac{i\hbar}{m}\right)\langle S|p_z|Z\rangle \tag{4.17}$$

μ is called as the reduced effective mass, with

$$\mu = \frac{m_c m_v}{m_c + m_v} \tag{4.18}$$

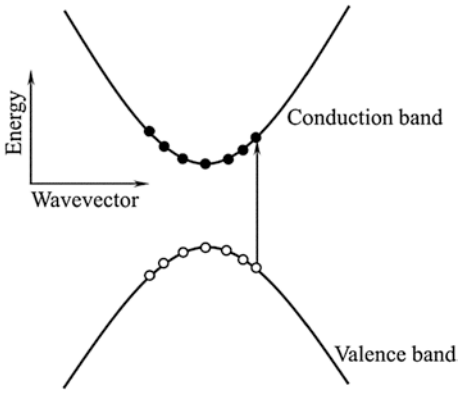

Fig. 4.2 Parabolic type two-band model for direct band gap semiconductor

where m_c and m_v are the effective masses of conduction band and valence band, respectively.

The $B(\hbar\omega)$ is a blocking factor that accounts for band (state) filling:

$$B(\hbar\omega) = \{1 - f_h[E_v(\hbar\omega)] - f_e[E_c(\hbar\omega)]\} \tag{4.19}$$

where $E_v(\hbar\omega)$ and $E_c(\hbar\omega)$ are the valence band and conduction band energy levels, respectively, with

$$E_c(\hbar\omega) = \frac{m_v}{m_c + m_v}\left(\hbar\omega - E_g\right) \tag{4.20}$$

$$E_v(\hbar\omega) = \frac{m_c}{m_c + m_v}\left(\hbar\omega - E_g\right) \tag{4.21}$$

The occupation probabilities $f_e(E_c)$ and $f_h(E_v)$ are Fermi–Dirac statistics distribution functions for holes in the valence band and electrons in the conduction band, respectively.

$$f_e(E_c) = \frac{1}{e^{(E_c - E_F)/k_B T} + 1} \tag{4.22}$$

$$f_h(E_v) = \frac{1}{e^{(E_F - E_v)/k_B T} + 1} \tag{4.23}$$

where E_F is Fermi energy level and k_B is Boltzmann constant.

Electrons in the conduction band dominate the blocking effect. If the excitation is not too strong, the change of the occupation probability for states in a parabolic band $\Delta f[E(k)]$ can be approximately obtained by Boltzmann statistics in the presence of the radiation field as follows

$$\Delta f_e[E_c(\hbar\omega)] = 4\pi^{3/2}\left(\frac{\hbar^2}{2m_c k_B T}\right)^{3/2} N e^{-\frac{E_c(\hbar\omega)}{k_B T}} \tag{4.24}$$

$$\Delta f_h[E_v(\hbar\omega)] = 4\pi^{3/2}\left(\frac{\hbar^2}{2m_v k_B T}\right)^{3/2} N e^{-\frac{E_v(\hbar\omega)}{k_B T}} \tag{4.25}$$

where N is the density of photon-generated free-carriers and defined in formula (4.2).

Because the effective mass of holes is always 10 times larger than that of electrons in direct band gap III–VI compound semiconductors, the Δf_h can be neglected in comparison with Δf_e for all the energy range of interest, and the change of the absorption coefficient at energy $\hbar\omega$ is

$$\Delta\alpha(\hbar\omega) \approx -\frac{16\pi^{3/2}}{3}\frac{e^2\hbar}{m_c}\left(\frac{\mu}{m_c}\right)^{3/2}\frac{mP_M^2}{\hbar^2}\frac{1}{n_0}\frac{N}{(k_B T)^{3/2}} \frac{\left(\hbar\omega - E_g\right)^{1/2}}{\hbar\omega}\exp\left[-\frac{\mu\left(\hbar\omega - E_g\right)}{m_c k_B T}\right] \tag{4.26}$$

According to $\Delta\alpha = \beta I$ (β nonlinear absorption coefficient), and substituting formula (4.2) into formula (4.26), β is expressed as

$$\beta \approx -\frac{16\pi^{3/2}}{3}\frac{e^2\hbar}{m_c}\left(\frac{\mu}{m_c}\right)^{3/2}\frac{mP_M^2}{\hbar^2}\frac{1}{n_0}\frac{1}{(k_BT)^{3/2}}\frac{\tau\alpha}{\hbar\omega}\frac{(\hbar\omega - E_g)^{1/2}}{\hbar\omega} \exp\left[-\frac{\mu(\hbar\omega - E_g)}{m_c k_B T}\right] \tag{4.27}$$

From the Kramers–Krönig relation, the change of refractive index at photon energy $\hbar\omega$ can be written as

$$\Delta n(\hbar\omega) = \frac{\hbar c}{\pi}\int_0^\infty \frac{\Delta\alpha(\hbar\omega')}{(\hbar\omega')^2 - (\hbar\omega)^2}d(\hbar\omega') \tag{4.28}$$

If $|\hbar\omega - E_g| \ll E_g$ and $k_BT \ll E_g$, according to formulas (4.2), (4.26), and (4.28), the nonlinear refraction coefficient $n_2(= \Delta n/I)$ can be calculated by

$$n_2(\hbar\omega) \approx -\frac{8\sqrt{\pi}}{3}\frac{e^2\hbar^2}{m}\frac{\mu}{m_c}\frac{mP_M^2}{\hbar^2}\frac{1}{n_0}\frac{1}{k_BT}\frac{\alpha(\hbar\omega)\tau_R}{(\hbar\omega)^3}J\left(\mu\frac{\hbar\omega - E_g}{m_c k_B T}\right) \tag{4.29}$$

where

$$J(a) = \int_0^\infty \frac{x^{1/2}e^{-x}}{x - a}\mathrm{d}x \tag{4.30}$$

In the present Boltzmann approximation, $\Delta n \propto I$. However, at high intensity where the free-carrier density exceeds the limit of the Boltzmann approximation, the nonlinearity itself will saturate and can only be described with the full Fermi–Dirac statistics.

(3) **Saturation level**

(a) **For the energy band absorption**

In Sect. 4.2.1.1(1), the saturation free-carrier density $N_{sat}(\hbar\omega)$ and saturation excitation light intensity I_{sat} are used, which is because at moderate or high excitation, most of the required fluence or intensity goes to blocking the absorption of the energy bands. The conduction band states fill first. The density of states for electrons at energy E, measured from the bottom of a parabolic conduction band in a bulk semiconductor, is [1]

$$D(E) = \frac{1}{2\pi^2}\left(\frac{2m_{eff}}{\hbar^2}\right)^{3/2}E^{1/2} \tag{4.31}$$

where m_{eff} is effective mass of photo-generated free-carriers. Most photo-generated electrons will occupy states that are within a few k_BT of the bottom of the

band. Half the number of these states gives an estimate for the number of electrons required to reach the saturation level, so [1]

$$N_{\text{sat}}^{\text{bg}} \approx \frac{1}{2}\int_0^{3k_{\text{B}}T} D(E)\mathrm{d}E = \frac{1}{6\pi^2}\left(\frac{6m_{\text{eff}}k_{\text{B}}T}{\hbar^2}\right)^{3/2} \tag{4.32}$$

and

$$I_{\text{sat}}^{\text{bg}} \approx \frac{1}{6\pi^2}\left(\frac{6m_{\text{eff}}k_{\text{B}}T}{\hbar^2}\right)^{3/2}\frac{\hbar\omega}{\alpha_0\left(N_{\text{sat}}^{\text{bg}}\right)\tau\left(N_{\text{sat}}^{\text{bg}}\right)} \tag{4.33}$$

For example, using the typical values for GaAs, that is, $\hbar\omega = 1.43$ eV, $m_{\text{eff}} = 0.067\, m_0$, $\alpha_0 \approx 2{,}500/\text{cm}$, and $\tau = 10$ ns, one can get $N_{\text{sat}}^{\text{bg}} = 8.5 \times 10^{17}/\text{cm}^3$, and $I_{\text{sat}}^{\text{bg}} = 7.5 \times 10^3$ W/cm^2.

(b) **For exciton absorption**

The exciton saturation density can be estimated with simple physical arguments. Saturation occurs when the density of photo-generated free-carriers (electrons and holes) is about one charge free-carrier per exciton volume ($4\pi r_{\text{ex}}^3/3$), where r_{ex} is exciton radius. The picture gives [1]

$$N_{\text{sat}}^{\text{ex}}(\hbar\omega) = \frac{3}{8\pi r_{\text{ex}}^3} \tag{4.34}$$

and

$$I_{\text{sat}}^{\text{ex}}(\hbar\omega) = \frac{3}{8\pi r_{\text{ex}}^3}\frac{\hbar\omega}{\alpha\left(N_{\text{sat}}^{\text{ex}}\right)\tau\left(N_{\text{sat}}^{\text{ex}}\right)} \tag{4.35}$$

For example, using typical values for GaAs, $r_{\text{ex}} = 120$ Å, $\hbar\omega = 1.43$ eV, $\alpha_0 \approx 7{,}500/\text{cm}$, and $\tau = 10$ ns, one can get $N_{\text{sat}}^{\text{ex}} = 7 \times 10^{16}/\text{cm}^3$, and $I_{\text{sat}}^{\text{ex}} = 200$ W/cm^2.

4.2.1.2 Mixed Multi-Photon-Absorption-Induced Free-Carrier Nonlinearity

(1) **Mixed multi-photon absorption**

The single-photon absorption can generate the nonlinear absorption and refraction due to photo-generated free-carriers effect. The two-photon absorption (2PA) or multi-photon absorption (MPA) can also cause the generation of nonlinear absorption and refraction of the free-carriers. The nonlinearities are different from the bound electronic nonlinear effects because they present cumulative characteristics

(with a decaying given by the free-carriers lifetime). In a single-beam z-scan experiment, if a number of free-carriers are generated, the nonlinear absorption can be written as

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\alpha_0 I - \beta_{\mathrm{two}} I^2 - \beta_{\mathrm{three}} I^3 - (\sigma_e N_e + \sigma_h N_h) I \tag{4.36}$$

where $\sigma_{e,h}$ is the free-carriers absorption cross section for electrons and holes, respectively. α_0 is the linear absorption coefficient. β_{two} and β_{three} are two-photon and three-photon absorption coefficients, respectively.

In two-band approximation, one can take $N_e = N_h = N$ and define $\sigma = \sigma_e + \sigma_h$. The formula (4.36) is rewritten as

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\alpha_0 I - \beta_{\mathrm{two}} I^2 - \beta_{\mathrm{three}} I^3 - N\sigma\, I \tag{4.37}$$

In most cases, there are single-photon absorption, two-photon absorption, and even multi-photon absorption, and all the absorptions generate free-carriers. The rate formula of free-carrier density is approximately written as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\alpha_0 I}{\hbar\omega} + \frac{\beta_{\mathrm{two}} I^2}{2\hbar\omega} + \frac{\beta_{\mathrm{three}} I^3}{3\hbar\omega} - \frac{N}{\tau_R} \tag{4.38}$$

Under steady state $\frac{\mathrm{d}N}{\mathrm{d}t} = 0$, one can obtain

$$N = \left(\frac{\alpha_0 I}{\hbar\omega} + \frac{\beta_{\mathrm{two}} I^2}{2\hbar\omega} + \frac{\beta_{\mathrm{three}} I^3}{3\hbar\omega} \cdots \right) \tau_R \tag{4.39}$$

Substituting formula (4.39) into formula (4.37), one can get

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\alpha_0 I - \left(\beta_{\mathrm{two}} + \frac{\alpha_0 \sigma \tau_R}{\hbar\omega} \right) I^2 - \left(\beta_{\mathrm{three}} + \frac{\beta_{\mathrm{two}} \sigma \tau_R}{2\hbar\omega} \right) I^3 - \cdots \tag{4.40}$$

Formula (4.40) indicates that the free-carrier absorption is mixed and accumulative effect.

(2) **Mixed multi-photon absorption-induced free-carriers nonlinear refraction**

In the light beam traveling into sample, the effect of free-carriers on the phase Φ of light beam is proportional to the density of free-carriers,

$$\frac{\mathrm{d}\Phi}{\mathrm{d}z} = k\gamma_{\mathrm{be}} I + k\sigma N \tag{4.41}$$

In formula (4.41), k is wavenumber, and γ_{be} is nonlinear refraction coefficient of bound electrons. The first term of right side is from bound electronic contribution, and the second term is from the free-carriers refraction contribution. The free-carriers refraction can be described through two different band-filling models: one is Aronov–Auston model, and the other is Moss–Burstein model.

(a) **Aronov–Auston model**

In Aronov–Auston model, the nonlinear refraction originating from free-carriers is directly calculated from the real part of the complex dielectric function. The generation of free-electrons with a density N in the conduction band, which is often referred to as the Drude contribution, is accompanied by an elimination of the bound electrons with a density N in the valence band, which is often referred to as a Lorentz contribution to the change of the dielectric constant. The overall change in the refractive index is given by [4]

$$\Delta n(\omega; N) = -\frac{\mathrm{N}e^2}{2\varepsilon_0 n_0 \omega^2} \frac{1}{\mu} \frac{E_g^2}{E_g^2 - (\hbar\omega)^2} \tag{4.42}$$

Substituting formula (4.39) into formula (4.42), one has

$$\Delta n(\omega; N) = -\left(\frac{\alpha_0 I}{\hbar\omega} + \frac{\beta_{\mathrm{two}} I^2}{2\hbar\omega} + \cdots\right) \frac{\tau_R e^2}{2\varepsilon_0 n_0 \omega^2} \frac{1}{\mu} \frac{E_g^2}{E_g^2 - (\hbar\omega)^2} \tag{4.43}$$

The first term at the right side of formula (4.43) is the contribution from linear absorption, and the second term is the two-photon absorption contribution. The two-photon absorption contribution can cause the fifth-order nonlinear refraction, and the one-photon contribution leads to the third-order nonlinear refraction. Neglecting the fifth-order nonlinear refraction, the formula (4.43) becomes

$$\Delta n(\omega; N) = -\frac{\alpha_0 \tau_R e^2}{2\hbar\varepsilon_0 n_0 \omega^3} \frac{1}{\mu} \frac{E_g^2}{E_g^2 - (\hbar\omega)^2} I \tag{4.44}$$

According to the relation of $\Delta n = \gamma I$, the nonlinear refraction coefficient γ is

$$\gamma = -\frac{\alpha_0 \tau_R e^2}{2\varepsilon_0 n_0 \hbar\omega^3} \frac{1}{\mu} \frac{E_g^2}{E_g^2 - (\hbar\omega)^2} \tag{4.45}$$

Formula (4.45) shows that if $\hbar\omega < E_g$, then $\gamma < 0$; this is self-defocusing effect. If $\hbar\omega > E_g$, then $\gamma > 0$, which causes the self-focusing effect.

(b) **Moss–Burstein model**

The Moss–Burstein model is a dynamic model with Boltzmann statistics [5, 6]. In this model, the free-carriers induced the change of refractive index is dependent on the means of free-carrier generation. The free-carriers block the absorption at frequencies higher than the energy gap by filling the available states in the conduction and valence bands. The model can obtain the absorption coefficient change by Kramers–Kronig integral. The refractive index change, which can be analyzed using the band model shown in Fig. 4.3, including the contributions from electrons in conduction band, heavy holes, and light holes, is given as [6]

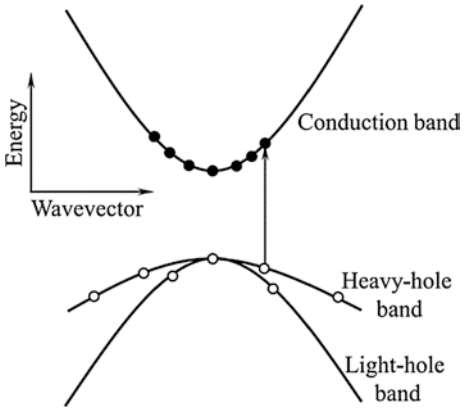

Fig. 4.3 The band structure schematic for Moss–Burstein model

$$\Delta n = -\frac{e^2}{2\varepsilon_0 n_0 \omega^2}\left\{\frac{N_c}{m_c}\left[1+Z\left(\frac{m_{ch}}{m_0}J_{hc}+\frac{m_{cl}}{m_0}J_{lc}\right)\right]\right.$$
$$\left.+\frac{Q_h}{m_h}\left(1+Z\frac{m_{ch}}{m_0}J_{hh}\right)+\frac{Q_l}{m_l}\left(1+Z\frac{m_{cl}}{m_0}J_{ll}\right)\right\} \tag{4.46}$$

with

$$Z = \frac{2}{3\sqrt{\pi}}\frac{E_p}{k_B T}\left(\frac{\hbar\omega}{E_g}\right)^2 \tag{4.47}$$

$$J_{ij} = \int_0^x \frac{x^2 \exp\left(-x^2\right)}{x^2 + a_{ij}}\mathrm{d}x \tag{4.48}$$

$$a_{ij} = \frac{E_g - \hbar\omega}{k_B T}\frac{m_{ci}}{m_j} \tag{4.49}$$

where m_0 is the free electron mass. $E_p = 2|p_{cv}(k=0)|^2/m_0$ is the Kane energy and approximately 21 eV for most semiconductors. N and Q are the photo-generated electron and hole densities, respectively. The subscripts c, h, and l represent the conduction, heavy-hole, and light-hole band, respectively. m_j is the effective mass of the band j. m_{ij} denotes the reduced effective mass of the ij band pair, and the dummy subscripts of i and j represent c, h, or l.

The first term in formula (4.46) is from the electronic contribution, which includes blocking caused by electron transitions from heavy-hole band and the light-hole band, and the change of the electron population in the conduction band. The other two terms containing Q_h and Q_l in formula (4.46) give the contributions of the holes from respective transitions.

The Q_h and Q_l are given by

$$\frac{N_c}{Q_h} \cong 1 + \left(\frac{m_l}{m_h}\right)^{3/2} \tag{4.50}$$

$$\frac{N_c}{Q_l} \cong 1 + \left(\frac{m_h}{m_l}\right)^{3/2} \tag{4.51}$$

Formula (4.48) is an approximation adequate for near-resonance radiation. Off resonance, such as two-photon absorption process, the J_{ij} should be replaced by F_{ij}, and $\mathrm{F_{ij}}$ is defined as

$$F_{ij} = -2J\left(\frac{m_{ci}}{m_j}\frac{E_g}{k_\mathrm{B}T}\right) + J\left(\frac{m_{ci}}{m_j}\frac{E_g - \hbar\omega}{k_\mathrm{B}T}\right) + J\left(\frac{m_{ci}}{m_j}\frac{E_g + \hbar\omega}{k_\mathrm{B}T}\right) \tag{4.52}$$

The J defines the integral as in formula (4.48). Formula (4.52) shows that for $\hbar\omega \cong E_g$ and $E_g \gg k_\mathrm{B}T$, the first and the third terms are very small compared with the second term; thus,

$$F_{ij} \approx J\left(\frac{m_{ci}}{m_j}\frac{E_g - \hbar\omega}{k_\mathrm{B}T}\right) \tag{4.53}$$

4.2.2 *Three-Band Model for Nonlinear Absorption and Refraction*

4.2.2.1 Excited-State Absorption

A realistic depiction of single-photon absorption-induced excited-state absorption (ESA) is shown in Fig. 4.4 with a three-band system of vibrational–rotational manifolds in the excited states. Although many bands are given, only two dominant transitions take place. For example, for molecules with rotational and vibrational degree of freedom coupled to the electronic transitions, the excitation to higher-lying states in each manifold is rapidly followed by intra-system relaxation to the bottom

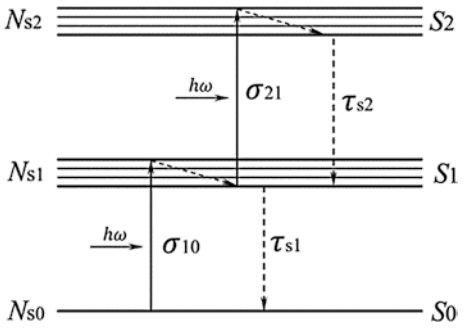

Fig. 4.4 Three-band model for excited state nonlinear absorption

of the band. The ESA then proceeds from this relaxed excited state into a higher rotational–vibrational manifold. Rapid inter-system relaxation returns the electrons in the second excited-state manifold to the bottom of the first excited band, where they can again absorb photons. Thus, a single excited-state absorber can efficiently absorb multiple times even for pulsed laser. If the ESA cross section is larger than that of the ground state, $\sigma_{21} > \sigma_{10}$, where the absorption cross sections are understood to be averaged over the vibrational–rotation manifolds, and the absorption process is also referred to be reverse saturation absorption (RSA). The model of three-band system is useful for describing the ESA in lots of organic molecules and semiconductors with free-carrier absorption phenomena.

In transient case, the absorption coefficient is dependent on time t under the radiation of laser pulse. In the three-band system, S_0, S_1, and S_2 are ground state, the first excited state, and the second excited state, respectively. The numbers of electrons at S_0, S_1, and S_2 states are N_{s0}, Ns_1, and Ns_2, respectively. The absorption cross sections of $S_0 \rightarrow S_1$ state and $S_1 \rightarrow S_2$ state are σ_{10} and σ_{21}, respectively. The total number of the electrons is $N(= N_{s0} + Ns_1 + Ns_2)$. Generally speaking, the electron at S_0 state absorbs one photon and transitions from S_0 to S_1 state. One part of the electrons at S_1 state will continue to absorb photons and transition from S_1 to S_2 state, and other part will relax to S_0 state through non-radiative transition, and the lifetime of the S_1 state is τ_{S1} accordingly. The electrons at S_2 state will relax to S_1 state through non-radiative transition, and the lifetime of the S_2 state is τ_{S2}. The number of electrons at each energy band meets the rate formulas as follows [7, 8]:

$$\frac{\mathrm{d}N_{S0}}{\mathrm{d}t} = -\frac{\sigma_{10}N_{\mathrm{so}}I}{\hbar\omega} + \frac{Ns_1}{\tau_{S1}} \tag{4.54}$$

$$\frac{\mathrm{d}N_{S1}}{\mathrm{d}t} = \frac{\sigma_{10}N_{\mathrm{so}}I}{\hbar\omega} - \frac{\sigma_{21}N_{s1}I}{\hbar\omega} - \frac{Ns_1}{\tau_{S1}} + \frac{Ns_2}{\tau_{S2}} \tag{4.55}$$

$$\frac{\mathrm{d}N_{S2}}{\mathrm{d}t} = \frac{\sigma_{21}N_{S1}I}{\hbar\omega} - \frac{Ns_2}{\tau_{S2}} \tag{4.56}$$

$$N = N_{s0} + Ns_1 + Ns_2 \tag{4.57}$$

Generally speaking, the lifetime τ_{S2} at excited state S_2 is very short, and the electrons at S_2 state will rapidly relax to the S_1 state. Thus, the electrons cannot be accumulated at S_2 state; the number of electrons at S_2 state is almost 0; that is,

$$\frac{\mathrm{d}N_{S2}}{\mathrm{d}t} = 0 \Rightarrow \frac{\sigma_{21}N_{S1}I}{\hbar\omega} = \frac{Ns_2}{\tau_{S2}} \tag{4.58}$$

According to formula (4.58), formulas (4.54)–(4.57) can be rewritten as

$$\frac{\mathrm{d}N_{S0}}{\mathrm{d}t} = -\frac{\sigma_{10}N_{\mathrm{so}}I}{\hbar\omega} + \frac{Ns_1}{\tau_{S1}} \tag{4.59}$$

$$\frac{\mathrm{d}N_{S1}}{\mathrm{d}t} = \frac{\sigma_{10}N_{\mathrm{so}}I}{\hbar\omega} - \frac{Ns_1}{\tau_{S1}} \tag{4.60}$$

$$N = N_{S0} + Ns_1 \tag{4.61}$$

The absorption coefficient from ground state S_0 can be written as

$$\alpha_{10} = \sigma_{10} N_{S0} \tag{4.62}$$

Before decaying back to the ground state S_0, the electrons at the first excited state S_1 can transition to a second higher state S_2 by absorbing another photon. The process produces what is called excited-state absorption (ESA) with an absorption cross section of σ_{21}. The absorption coefficient from S_1 excited state to S_2 excited state (marked as α_{21}) is simply written as

$$\alpha_{21} = \sigma_{21} N_{S1} \tag{4.63}$$

where N_{S1} is the electron density of the first excited state.

Case 1 ($\tau_{s1} \gg t_p$) If the $\tau_{s1} \gg t_p$ (t_p is laser pulse width), the electrons at S_1 state can retain for a long time, then in formulas (4.59) and (4.60)

$$\frac{\sigma_{10} N_{\mathrm{so}} I}{\hbar\omega} \gg \frac{Ns_1}{\tau_{S1}} \tag{4.64}$$

One can get

$$\frac{\mathrm{d}N_{S1}}{\mathrm{d}t} \approx \frac{\alpha_{10} I}{\hbar\omega} \tag{4.65}$$

Integrating formula (4.65) yields [9]

$$N_{S1}(t) = \frac{\alpha_{10}}{\hbar\omega} \int_{-\infty}^{t} I\left(t'\right) dt' = \frac{\alpha_{10}}{\hbar\omega} F(t) \tag{4.66}$$

where $F(t)$ is defined as the time-dependent fluence (energy per unit area). The light intensity change along sample thickness is generally expressed as

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\sigma_{10} N_{S0} I - \sigma_{21} N_{S1} I \tag{4.67}$$

with $F(t)$ instead of I; combining formulas (4.62), (4.66), and (4.67), one can get

$$\frac{\mathrm{d}F(t)}{\mathrm{d}z'} = -\sigma_{1g} N_g F(t) - \frac{N_g}{\hbar\omega} \sigma_{1g} \sigma_{21} F^2(t) = -\sigma_{1g} N_g F(t) \left[1 + \frac{\sigma_{21} F(t)}{\hbar\omega}\right] \tag{4.68}$$

Case 2 (*At steady state*) According to formulas (4.59) and (4.60), $\mathrm{d}N_{S0}/\mathrm{d}t = \mathrm{d}N_{S1}/\mathrm{d}t = 0$, one can get

$$N_{S1} = \frac{\sigma_{10} N_{\mathrm{so}} I \tau_{S1}}{\hbar\omega} \tag{4.69}$$

Substituting formula (4.62) into formula (4.69), one has

$$N_{S1} = \frac{\alpha_{10} I \tau_{S1}}{\hbar\omega} \tag{4.70}$$

The absorption coefficient of excited state can be obtained by the combination of formulas (4.63) and (4.70)

$$\alpha_{21} = \sigma_{21}\alpha_{10}\frac{I\tau_{S1}}{\hbar\omega} \tag{4.71}$$

The light intensity change along sample thickness direction is

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\alpha\, I = -(\sigma_{10}N_{s0} + \sigma_{21}N_{S1})I \tag{4.72}$$

According to formulas (4.61) and (4.72), the absorption coefficient α can be rewritten as

$$\alpha = \sigma_{21}N + (\sigma_{10} - \sigma_{21})N_{S0} \tag{4.73}$$

In three-band system, from formula (4.73), one can find that $\sigma_{21}N$ is a constant for a fixed sample. The N_{s0} decreases with the light intensity increasing; thus, for $\sigma_{10} > \sigma_{21}$, the sample shows saturation absorption, and nonlinear absorption coefficient $\beta < 0$. For $\sigma_{10} < \sigma_{21}$, $\beta > 0$, and the sample shows reverse saturation absorption.

4.2.2.2 Excited-State Absorption-Induced Free-Carriers Refraction

In three-band system, the rate formula of electron population at every energy state and the light intensity change are described by formulas (4.54)–(4.57), and the phase change of light beam is accordingly expressed as

$$\frac{\mathrm{d}\phi}{\mathrm{d}z'} = \frac{2\pi}{\lambda}\Delta n = k_{S0}N_{S0} + k_{S1}N_{S1} + k_{S2}N_{S2} \tag{4.74}$$

where Δn and λ are the refractive index change and light wavelength, respectively. k_{S0}, k_{S1} and k_{S2} are refraction degrees of S_0 state, S_1 state, and S_2 state, respectively. In general, the lifetime τ_{S2} of S_2 state is very short, and the electrons at S_2 state will rapidly relax to the S_1 state; thus, $N_{S2} \sim 0$. Similar to the excited-state absorption, the phase change of light beam in the traveling inside the sample is simplified as

$$\frac{\mathrm{d}\phi}{\mathrm{d}z'} = \frac{2\pi}{\lambda}\Delta n = k_{S0}N_{S0} + k_{S1}N_{S1} \tag{4.75}$$

Combining formulas (4.61) and (4.75), one can get

$$\frac{\mathrm{d}\phi}{\mathrm{d}z'} = k_{S1}N + (k_{S0} - k_{S1})N_{S0} \tag{4.76}$$

Formula (4.76) shows that the first term $k_{S1}N$ is a constant for a fixed sample. N_{s0} decreases with the increase in light intensity; the sample presents self-defocusing for $k_{S0} > k_{S1}$, and nonlinear refractive coefficient $\gamma < 0$. The sample presents self-focusing effect for $k_{S0} < k_{S1}$, and the $\gamma > 0$.

4.2.3 Thermally Induced Nonlinear Absorption and Refraction

4.2.3.1 Laser Excitation-Induced Temperature Change

(1) **Temperature change in two-band system**

In two-band system (also see Fig. 4.1), at the approximation of thin sample, the light intensity change through the sample meets the following formula

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\sigma_{10} N_{S0} I \tag{4.77}$$

where z' is traveling length of light inside the sample. In the process of electrons at S_1 state relaxation to S_0 state, the released energy is transformed into heat energy Q, which meets

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \sigma_{10} N s_0 I \tag{4.78}$$

The heat will diffuse inside the sample, and the heat conduction formula can be expressed as

$$\rho_0 C_p \frac{\partial \Delta T}{\partial t} - \kappa \nabla \Delta T = \frac{\mathrm{d}Q}{\mathrm{d}t} \tag{4.79}$$

where ρ_0 and C_p are the density and heat capacity of sample, respectively. κ and ΔT are thermal conductivity and temperature rise of sample, respectively. The laser pulse is short, and the thermal diffusion is negligible during the laser pulse, that is, $\kappa \nabla \Delta T \approx 0$. Thus, combining formulas (4.78) and (4.79), one can obtain [8]

$$\Delta T = \frac{\sigma_{10} N s_0 I}{\rho_0 C_p} \mathrm{d}t \tag{4.80}$$

(2) **Temperature change in three-band system**

In the three-band system (as shown in Fig. 4.4), there are non-radiative transitions under the excitation of laser pulse, and the non-radiative transitions can cause the heat generation.

At the approximation of thin sample, the change of light intensity through the sample meets the following formula

$$\frac{\mathrm{d}I}{\mathrm{d}z'} = -\sigma_{10} N_{S0} I - \sigma_{21} N_{S1} I \tag{4.81}$$

The relaxed quantity of heat per unit volume is accordingly

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \sigma_{10} N s_0 I + \sigma_{21} N_{S1} I \tag{4.82}$$

Similar to the two-level system, the temperature change for three-level system is [8]

$$\Delta T = \frac{1}{\rho_0 C_p}(\sigma_{10} N s_0 I + \sigma_{21} N_{S1} I)dt \tag{4.83}$$

Formulas (4.80) and (4.83) show that the excited-state absorption can cause the temperature rise. The temperature rise will lead to thermal expansion and band gap shrinkage for most semiconductors; thus, the thermally induced nonlinear refraction occurs.

4.2.3.2 Thermally Induced Change of Absorption Coefficient in Band-Filling Model

By considering the changes of E_g and $B(\hbar\omega)$ with temperature, the thermally induced change of absorption coefficient can be obtained through the derivation operation on temperature T at both sides of formula (4.16) as follows [10]

$$\begin{aligned}\frac{\mathrm{d}\alpha(\hbar\omega)}{\mathrm{d}T} = &-\frac{\mathrm{d}E_g}{\mathrm{d}T}\frac{\alpha(\hbar\omega)}{2\left(\hbar\omega - E_g\right)} \\ &- \frac{\alpha(\hbar\omega)}{\{1 - f_h[E_v(k)] - f_e[E_c(k)]\}}\left\{\frac{\mathrm{d}f_e[E_c(k)]}{\mathrm{d}T} + \frac{\mathrm{d}f_h[E_v(k)]}{\mathrm{d}T}\right\}\end{aligned} \tag{4.84}$$

In intrinsic semiconductors, the Femi energy level E_F is located within the band gap, generally speaking,

$$E_c - E_F \gg k_\mathrm{B}T, E_F - E_v \gg k_\mathrm{B}T \tag{4.85}$$

Based on formula (4.85), formulas (4.22) and (4.23) are approximately written as

$$f_e(E_c) \approx e^{-(E_c - E_F)/k_\mathrm{B}T} \tag{4.86}$$

$$f_h(E_v) \approx e^{-(E_F - E_v)/k_\mathrm{B}T} \tag{4.87}$$

That is, if the electrons in conduction band are located at the large energy gap or the temperature is not very high, the $f_e(E_c)$ and $f_h(E_v)$ are approximated to Bolztmann distribution, and the occupation probability of conduction band by thermally excited electrons is very small; substituting formulas (4.86) and (4.87) to formula (4.84), one can obtain

$$\begin{aligned}\frac{\mathrm{d}\alpha(\hbar\omega)}{\mathrm{d}T} = &-\frac{\mathrm{d}E_g}{\mathrm{d}T}\frac{\alpha(\hbar\omega)}{2\left(\hbar\omega - E_g\right)} \\ &- \frac{\alpha(\hbar\omega)}{\{1 - f_e[E_c(k)] - f_h[E_v(k)]\}}\left\{\frac{E_c - E_F}{k_\mathrm{B}T^2} f_e[E_c(k)] - \frac{E_v - E_F}{k_\mathrm{B}T^2} f_h[E_v(k)]\right\}\end{aligned} \tag{4.88}$$

For intrinsic semiconductors, assume that the E_F is at the middle of the energy gap, that is,

$$E_g = E_c - E_v, \quad \text{and} \quad E_F = (E_c + E_v)/2 \tag{4.89}$$

Thus,

$$E_c - E_F = E_F - E_v = E_g/2 \tag{4.90}$$

Substituting formula (4.90) into formulas (4.86) and (4.87), respectively, one can obtain

$$f_e(E_c) = f_h(E_v) = e^{-E_g/2k_BT} \tag{4.91}$$

According to formulas (4.89), (4.90), and (4.91), the formula (4.88) can be simplified as

$$\frac{d\alpha(\hbar\omega)}{dT} = -\frac{dE_g}{dT}\frac{\alpha(\hbar\omega)}{2(\hbar\omega - E_g)} - \frac{\alpha(\hbar\omega)}{1 - 2e^{-E_g/2k_BT}}\frac{E_g}{k_BT^2}e^{-E_g/2k_BT} \tag{4.92}$$

The first term on right side of formula (4.92) represents the contribution from energy gap change with temperature, and the second term is the contribution from occupation probability change with temperature.

For most semiconductors, the temperature dependence of band gap is well known as [11]

$$E_g(T) = E_g(0) - \frac{\xi T^2}{T + \eta} \tag{4.93}$$

where $E_g(0)$ is band gap at 0 K, ξ and η are factors with adjustable values, for example, $\xi \sim 5 \times 10^{-4}$, and $\eta \sim 250$ for InSb thin films above 300 K [12]. Thus, the band gap decreases as the temperature increases, i.e., $dE_g/dT < 0$ for most semiconductors.

4.2.3.3 Thermally Induced Change of Refractive Index

In the process of interaction of laser pulse with sample, the light absorption leads to a temperature change ΔT, which in turn also causes a change of local density $\Delta\rho$. The change of local density causes the generation of sound waves, which effectively relieve the stresses induced by the density change. On the other hand, the density change causes the structural change, such as volume expansion, which in turn also leads to shrinkage or broadening of energy band gap. Therefore, generally speaking, for solid semiconductors, the temperature-dependent refractive index change can be expressed as

$$\Delta n = \left(\frac{\partial n}{\partial T}\right)_{d\rho/dT}\Delta T + \left(\frac{\partial n}{\partial T}\right)_{dE_g/dT}\Delta T \tag{4.94}$$

where $\mathrm{d}\rho/\mathrm{d}T$ represents the density change with temperature, and $\mathrm{d}E_g/\mathrm{d}T$ the change of energy band gap with temperature. Here, $(\partial n/\partial T)_{\mathrm{d}\rho/\mathrm{d}T}$ describes the refractive index changes due to thermal expansion, and $(\partial n/\partial T)_{\mathrm{d}E_g/\mathrm{d}T}$ is temperature-dependent refractive index change from energy band gap, that is, the temperature-induced shifting of absorption edge. The $(\partial n/\partial T)_{\mathrm{d}E_g/\mathrm{d}T}\Delta T$ term can be calculated from thermally induced shift of the absorption edge, which is comparable with band filling ($n_2 \sim 10^{-6}$ cm^2/W), but is slower. The optic–thermal nonlinearity occurs when sufficient heat is deposited. The sample recovers when heat diffuses from the excited region. The time of turn-on and turn-off depends on optical power, optical spot size, and the way sample is heated, but typical values are from 0.1 to 10.0 μs. The $\mathrm{d}n/\mathrm{d}T$ is in the order of 10^{-4}/K for most semiconductors.

(1) $(\frac{\partial n}{\partial T})_{\mathrm{d}\rho/\mathrm{d}T}$ **term contribution to thermally induced nonlinear refraction**

The relation of refractive index n with density ρ of solid semiconductor can be expressed by the relation [13, 14]

$$\left(\frac{n^2-1}{n^2+2}\right)\frac{1}{\rho} = R \tag{4.95}$$

where R is a constant called specific refraction. Formula (4.95) is rewritten as

$$n^2 = \frac{2R\rho+1}{1-R\rho} \tag{4.96}$$

Formula (4.96) gives the relation of the refractive index with the density ρ and shows that the change of density of sample can lead to the change of refractive index. Conducting the derivation operation on ρ at both sides of formula (4.96), one can get the change of refractive index with density $\mathrm{d}n/\mathrm{d}\rho$ as

$$\frac{\mathrm{d}n}{\mathrm{d}\rho} = \frac{3R}{(1-R\rho)^2} \tag{4.97}$$

Substituting formula (4.95) into formula (4.97), one can find

$$\frac{\mathrm{d}n}{\mathrm{d}\rho} = \frac{\left(n^2-1\right)\left(n^2+2\right)}{6n\rho} \tag{4.98}$$

The density change can be caused by the electrostriction, magnetostriction, or thermal expansion. Here, one is only concerned about the thermo-optic coefficient induced by the temperature-dependent density change, i.e., thermal expansion, which can be calculated as

$$\left(\frac{\mathrm{d}n}{\mathrm{d}T}\right)_{\Delta\rho} = \frac{\mathrm{d}n}{\mathrm{d}\rho}\frac{\mathrm{d}\rho}{\mathrm{d}T} = \frac{\left(n^2-1\right)\left(n^2+2\right)}{6n\rho}\frac{\mathrm{d}\rho}{\mathrm{d}T} \tag{4.99}$$

The density can be written as

$$\rho = \frac{M}{V} \tag{4.100}$$

where M and V are mass and volume of sample, respectively.

$$\frac{\mathrm{d}\rho}{\mathrm{d}T} = -\frac{M}{V}\frac{\mathrm{d}V/V}{\mathrm{d}T} = -\rho\beta_T \tag{4.101}$$

where $\beta_T (= \frac{\mathrm{d}V/V}{\mathrm{d}T})$ is defined as the volume thermal expansion coefficient. The relation of linear thermal expansion coefficient α_T with the volume thermal expansion coefficient β_T is

$$\beta_T = 3\alpha_T \tag{4.102}$$

Substituting formulas (4.101) and (4.102) into formula (4.99), one can obtain

$$\left(\frac{\mathrm{d}n}{\mathrm{d}T}\right)_{\Delta\rho} = -\frac{\left(n^2-1\right)\left(n^2+2\right)}{2n}\alpha_T \tag{4.103}$$

Formula (4.103) is the density change-induced thermo-optic coefficient, which is negative value for most materials due to $\alpha_T > 0$. However, for aqueous sample with a negative thermal expansion coefficient $\alpha_T < 0$, $(\mathrm{d}n/\mathrm{d}T)_{\Delta\rho} > 0$, which can cause a positive thermo-optic coefficient.

(2) $(\mathrm{d}n/\mathrm{d}T)_{\Delta\rho} > 0$, **Band gap shrinkage contribution to thermally induced nonlinear refraction**

For the semiconductor, the thermo-optic coefficient is from two kind of contributions [15]: one is density-dependent thermal expansion, and the other is from energy band gap change with temperature and can also be theoretically expressed as

$$\frac{\mathrm{d}n}{\mathrm{d}T} = -\frac{m_e a_l c_0}{\sqrt{1+2c_0 y_B}}\alpha_T - \frac{c_0 y_B}{2\sqrt{1+2c_0 y_B}}\frac{1}{E_g}\frac{\mathrm{d}E_g}{\mathrm{d}T} \tag{4.104}$$

where a_l is lattice constant, c_0 and y_B are constants for a given semiconductor, and m_e is electron mass. The first term of formula (4.104) can be actually considered as density change-induced thermo-optic coefficient and is equivalent to formula (4.103) under certain condition. That is,

$$\left(\frac{\mathrm{d}n}{\mathrm{d}T}\right)_{\Delta\rho} = -\frac{m_e a_l c_0}{\sqrt{1+2c_0 y_B}}\alpha_T = -\frac{\left(n^2-1\right)\left(n^2+2\right)}{2n}\alpha_T \tag{4.105}$$

Thus

$$\frac{m_e a_l c_0}{\sqrt{1+2c_0 y_B}} = \frac{\left(n^2-1\right)\left(n^2+2\right)}{2n} \tag{4.106}$$

The second term of formula (4.104) mainly originates from the thermally induced energy band gap change and can be written as

$$\left(\frac{\mathrm{d}n}{\mathrm{d}T}\right)_{\mathrm{d}E_g/\mathrm{d}T} = -\frac{c_0 y_B}{2\sqrt{1+2c_0 y_B}}\frac{1}{E_g}\frac{\mathrm{d}E_g}{\mathrm{d}T} \tag{4.107}$$

For covalently bonding semiconductor compounds, the contribution of α_T (thermal expansion) to $\mathrm{d}n/\mathrm{d}T$ is almost negligible compared with the contribution from thermally induced energy band gap change. For most covalently bonding semiconductors, based on formula (4.93), the band gap decreases as the temperature increases (i.e., $\mathrm{d}E_g/\mathrm{d}T < 0$); thus, the thermo-optic coefficient is generally $(\mathrm{d}n/\mathrm{d}T)_{\mathrm{d}E_g/\mathrm{d}T} > 0$, which leads to self-focusing effect.

4.3 Nonlinear Absorption and Refraction of Semiconductor Thin Films

4.3.1 Nonlinear Saturation Absorption of c-Sb-Based Phase-Change Thin Films

4.3.1.1 Experimental Measurements

The nonlinear absorption of 20-nm *c*-Sb, *c*-Sb_7Te_3, and *c*-Sb_2Te_3 films is measured at 405 nm light wavelength using a single-beam *z*-scan technique, where c is logogram of crystalline. Figure 4.5a shows typical open-aperture (OA) *z*-scan results for the *c*-Sb, *c*-Sb_7Te_3, and *c*-Sb_2Te_3 films with a 3 mW incident laser power and a 50 ns pulse width. The curves show that all thin films exhibit saturation absorption characteristic, indicating that the transmitted light intensity increases as the sample moves closer to the focal plane. Moreover, the observed

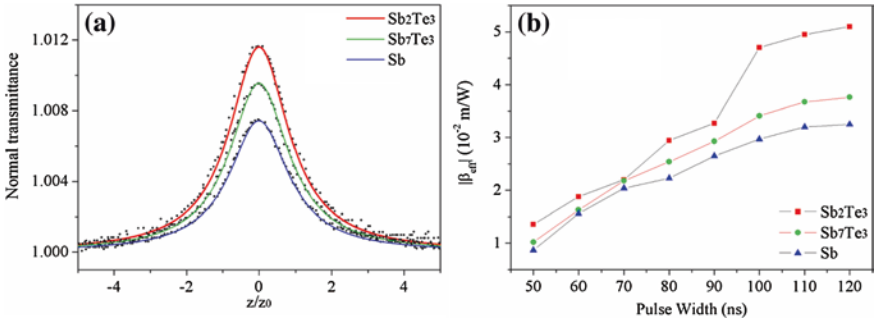

Fig. 4.5 Nonlinear absorption curves, **a** open aperture *z*-scan results for the *c*-Sb_2Te_3 (*red line*), *c*-Sb_7Te_3 (*green line*), and *c*-Sb (*blue line*) at 50 ns, **b** the effective nonlinear absorption coefficient β_{eff} for *c*-Sb_2Te_3 (*squares*), *c*-Sb_7Te_3 (*circles*), and *c*-Sb (*triangles*) as a function of pulse width. Reprinted with permission from Ref. [16]. Copyright 2012, American Institute of Physics

saturation absorption is highly repeatable, and no change in surface morphology of the sample is observed after the irradiation. Therefore, the nonlinearity of *c*-Sb, *c*-Sb_7Te_3, and *c*-Sb_2Te_3 can be safely attributed to their inherent mechanism, and the effect of morphological change is excluded. Data fitting shows that the effective nonlinear absorption coefficients β_{eff} of *c*-Sb, *c*-Sb_7Te_3, and *c*-Sb_2Te_3 are determined as -0.867×10^{-2}, -1.02×10^{-2}, and -1.36×10^{-2} m/W, respectively. The experimental results show that the β_{eff} is in the order of 10^{-2} m/W. The *z*-scan measurement for these samples is also performed through changing the laser pulse width from 50 to 120 ns to get the effect of pulse width on β_{eff}, and the fitting value for β_{eff} is plotted in Fig. 4.5b. β_{eff} shows a quick increasing initially, followed by a gradual increasing trend with pulse width. Generally speaking, the thermal effect becomes obvious as pulse width increases. Thus, according to the dependence of β_{eff} on pulse width, the thermal effect plays an important role in determining nonlinear absorption of *c*-Sb-based thin films.

Variable-temperature spectroscopic ellipsometry measurements are performed for *c*-Sb, *c*-Sb_7Te_3, and *c*-Sb_2Te_3 in the temperature range from 300 to 480 K in vacuum to further confirm the thermal origin of β_{eff}. Figure 4.6a–c show the plots of the temperature dependence of the absorption coefficient at 405 nm light wavelength. The absorption coefficients decrease with increasing temperature. Using

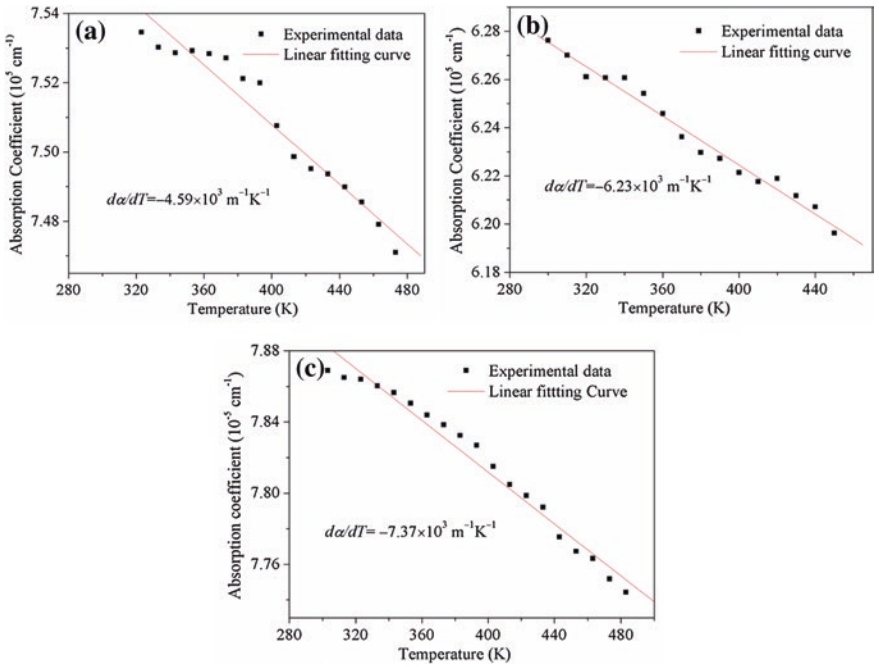

Fig. 4.6 Temperature dependence of absorption coefficients in the range from 300 to 480 K, **a** *c*-Sb, **b** *c*-Sb_7Te_3, and **c** *c*-Sb_2Te_3. Reprinted with permission from Ref. [16]. Copyright 2012, American Institute of Physics

Table 4.1 Thermo-physical parameters in the numerical calculation

	$\alpha(10^5\ \mathrm{cm}^{-1})$	$\rho(\mathrm{Kg/m}^3)$	C_p(J/Kg K)
Sb	7.50	6.68	210.00
$Sb_{70}Te_{30}$	6.32	6.65	209.25
Sb_2Te_3	7.78	6.50	205.49

the linear fitting, the $\mathrm{d}\alpha/\mathrm{d}T$ of c-Sb, c-Sb_7Te_3, and c-Sb_2Te_3 can be determined as -4.59×10^3, -6.23×10^3, and $-7.37 \times 10^3\ \mathrm{m}^{-1}\ \mathrm{K}^{-1}$, respectively. The thermally induced nonlinear absorption coefficient β_t under short laser pulse irradiation (tens of nanosecond) can be approximately expressed as follows [17]:

$$\beta_t = \frac{\alpha_0 t_p}{2\rho C_p}\frac{\mathrm{d}\alpha}{\mathrm{d}T} \tag{4.108}$$

where C_p the heat capacity and $\mathrm{d}\alpha/\mathrm{d}T$ is the change of α with respect to temperature. Inserting these values and the parameters listed in Table 4.1 into formula (4.108), the β_t values for c-Sb, c-Sb_7Te_3, and c-Sb_2Te_3 are calculated to be -0.608×10^{-2}, -0.702×10^{-2} and -1.08×10^{-2} m/W, respectively. The calculated nonlinear absorption coefficients agree well with z-scan experimental results, providing unambiguous proof that the saturation absorption observed in the z-scan measurement is from the thermal effect.

Figures 4.5 and 4.6 indicate that for c-Sb-based phase-change thin films, the nonlinear saturation absorption results from thermal effects. As is well known, the c-Sb-based phase-change thin films are constructed by resonant bonding. In z-scan measurement, the laser pulse heats the sample and causes the weakening of resonant bonding, and the weakening of resonant bonding is the essence of nonlinear saturation absorption.

4.3.1.2 Thermally Induced Weakening Mechanism of Resonant Bonding

Here, the Sb_2Te_3 phase-change material is taken as an example. Figure 4.7a shows that Sb_2Te_3 is crystallized into an A7-type structure with a D_{3d} point group, where the atoms are periodically stacked in a nearly octahedral coordination. The Sb and Te atoms possess three and four p-orbital electrons, respectively. The number of electrons is insufficient to form six saturation electron-pair bonds, and the resonant bonding predominates. As is schematically shown in Fig. 4.7b, Φ_1 and Φ_2 denote the two possible limiting bonding states for a simplified one-dimensional atom chain. In this bonding state, the p-orbital electrons are mainly localized in bonding area between two adjacent atoms. The real bonding state is a superposition of Φ_1 and Φ_2 due to the lack of enough p-orbital electrons, where p-orbital electrons are more delocalized and result in a large value in the electronic polarizability. Furthermore, atoms along certain direction align very well, which is favorable for polarization of p-orbital electrons. The alignment strengthens the resonant bonding and enhances the electronic polarizability. However, the crystalline lattice size and inter-atomic spacing expand with increasing temperature, resulting in

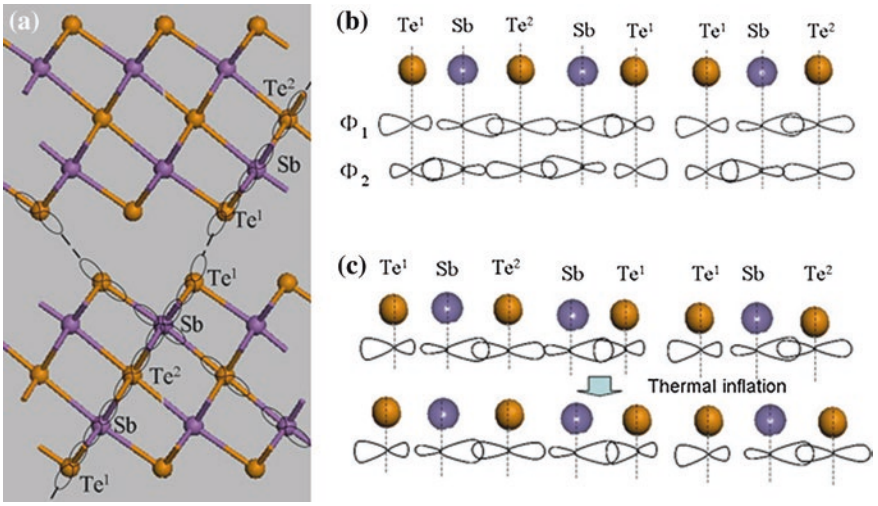

Fig. 4.7 **a** Local order in crystalline Sb_2Te_3 and how to align along two primary directions for the p orbits. **b** Schematic demonstration of limiting bonding state of the resonant bonding. **c** Electronic mixing weakens with lattice inflation. Reprinted with permission from Ref. [16]. Copyright 2012, American Institute of Physics

the weakening of the mixing among *p*-orbital electrons, as shown in Fig. 4.7c. The properties of Sb_2Te_3 from 300 to 500 K temperature range,[13] which is the temperature region involved in low laser irradiation intensity, are investigated with first-principle theory to determine the effect of temperature on the bonding state.

The resonant bonding is characterized by a high optical dielectric constant ε_∞ and a large Born effective charge Z^*. Monitoring the variation in the resonant-related physical quantities is an effective method to measure the resonant bonding strength. The Z^* and ε_∞ values of Sb_2Te_3 are calculated using the first-principle code ABINIT within the generalized gradient approximation (GGA), as well as the plane wave method and Perdew–Burke–Ernzerhof (PBE)-type pseudo-potential potentials, whereas the lattice parameters at the higher temperature are obtained by assuming a constant thermal expansion coefficient of 12×10^{-6} K^{-1}. Despite of the lattice expansion, the relative position of the atoms in the unit cell remains constant at all temperatures. The dominant effect from temperature in the low temperature range is the lattice expansion, and the change of relative position only plays an important role in higher temperature.

Figure 4.8a, b give the plots of the calculated Z^* and ε_∞ versus temperature, respectively. Sb_2Te_3 is a hexagonal type lattice with two different components (parallel and perpendicular), which can be averaged. In addition, the stacking sequence in Sb_2Te_3 is $-Te^1 - Sb - Te^2 - Sb - Te^1$; three set of Born effective charges corresponding to Te^1, Te^2, and Sb exist. Figure 4.8a shows that the Z^* of Te^2 and Sb is much larger than that of Te^1, which is because there are six atoms in the first neighboring shell for Te^2 and Sb, whereas for Te^1 only three Sb atoms are

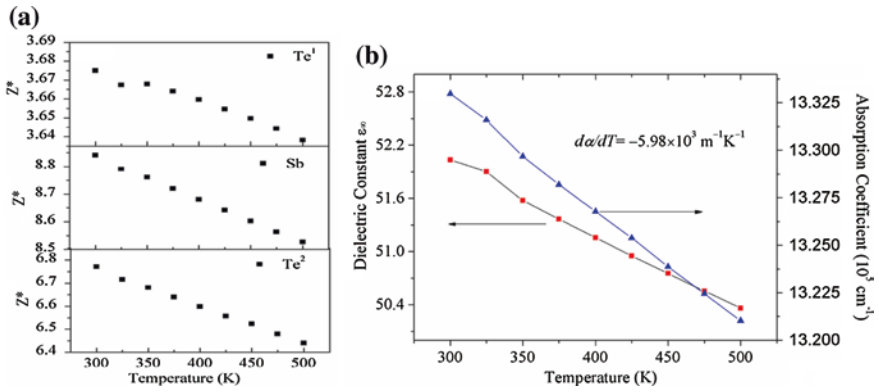

Fig. 4.8 **a** Calculated Born effective charge, and **b** optical dielectric constant ε_∞ (*squares*), and derivative of absorption coefficient (*triangles*) of *c*-Sb_2Te_3 in the temperature range from 300 to 500 K, respectively. Reprinted with permission from Ref. [16]. Copyright 2012, American Institute of Physics

closely bonded and the other three Te^1 atoms are weakly bonded by weak Van der Waals force. As a result, the p-orbital electrons of Te^2 and Sb atoms are more non-locally distributed, resulting in a larger value of Born effective charge. Besides, the most obvious feature is that Z^* decreases with the increasing temperature for all atoms. The decrease degree in the Z^* for Te^2 at 500 K is as much as 3 % compared with the value at the room temperature, corresponding to a bond length expansion of 2.4 %. This is a significant value, indicating that Z^* is sensitive to the lattice expansion. It can be understood that the spacing between atoms expands with the increasing temperature; thus, the p-orbital mixing between adjacent Sb and Te atoms becomes smaller and the Coulomb interaction becomes weaker, leading to a weakening of resonant bonding and a smaller value of the Born effective charge. Similar to Z^*, the dielectric constant ε_∞ also shows an unusual large value resulting from the large optical matrix elements of resonant bonding, as well as a similar decreasing trend with increasing temperature. ε_∞ decreases by as much as 3.2 % when the temperature increases from 300 to 500 K. Given that the large optical dielectric constant characteristics is due to the high electronic polarizability, the decline of ε_∞ also reflects the decrease in the resonant bonding strength.

The gradual weakening of the resonant bonding with increasing temperature is thought to induce a giant nonlinearity. The process can be understood as follows: The temperature at the center of the irradiation spot is higher due to larger central intensity when a Gaussian profile laser irradiates. The temperature decreases along the radial direction, and the resonant bonding at the center of the irradiation spot is disturbed the most, whereas those at the edge of the spot are the least affected. Note that the linear absorption decreases as resonant bonding is disturbed. It needs to be pointed out that the mechanism responsible for the weakening of resonant bonding is attributed to lattice expansion. The thin films with this kind of absorption characteristics exhibit a higher transmittance when irradiated by a stronger

laser. This model can thus account for the giant saturation absorption observed in the *c*-Sb-based phase-change materials.

The absorption coefficient α of Sb_2Te_3 at 405 nm light wavelength is calculated and plotted in Fig. 4.8b to determine the dependence on the temperature accurately. The α monotonously decreases with increasing temperature. This result is consistent with fact that the resonant bonding becomes weaker as the temperature increases. Using a simple linear fitting, $\mathrm{d}\alpha/\mathrm{d}T$ of Sb_2Te_3 is determined as $-5.89 \times 10^3\ \mathrm{m}^{-1}\ \mathrm{K}^{-1}$. A β_t value of the order of 10^{-2} m/W is obtained by inserting the calculated $\mathrm{d}\alpha/\mathrm{d}T$ and the typical parameters from Table 4.1 into formula (4.108). This value is quite large compared with the nonlinearity induced either by the Kerr effect or by the redistribution of the electron population, implying different physical mechanisms. Thermally induced nonlinearity is conventionally not preferred because of long response time. However, the response time of the thermally induced nonlinearity in *c*-Sb-based films is found to be in the order of 10 ns. The short response time is possible because the disturbing of resonant bonding does not require thermal diffusion and long distance atom movement.

4.3.2 Nonlinear Reverse Saturation Absorption and Refraction of c-InSb Thin Films

4.3.2.1 Experimental Data

c-InSb is an n-type semiconductor that belongs to the III–V family materials. Bulk *c*-InSb is a direct-transition-type semiconductor with narrow band gap of approximately 0.18 eV at 300 K. To obtain the band energy gap of *c*-InSb thin films, Fourier transform infrared spectroscopy (FT-IR) measurement is conducted at room temperature, and the dependence of absorption on photon energy $h\nu$ is plotted in Fig. 4.9 [18]. The plot shows the band gap of $E_g = \sim 0.26$ eV for the 30-nm *c*-InSb thin film. This band gap is a little larger than that of bulk crystal, which may be ascribed to two aspects. One is the quantum confinement effect of thin films. The blueshift of band gap energy due to quantum confinement effect can be estimated through an effective mass approximation

$$E_{qc} = \frac{\pi^2\hbar^2}{2L^2}\left[\frac{1}{m_e} + \frac{1}{m_h}\right] \tag{4.109}$$

Taking into account the parameters listed in Table 4.2, the quantized energy of the electrons in $L = 30$ nm thin film is calculated as ~0.031 eV. The other is the formation of nanoscale *c*-InSb particles in the preparation of the thin films. The nanoscale crystalline particles can be considered to be quantum dots, and the quantum dot effect also leads to blueshift of the band gap. Therefore, the two aspects cause the band gap of InSb thin films to be a little larger than the InSb bulk.

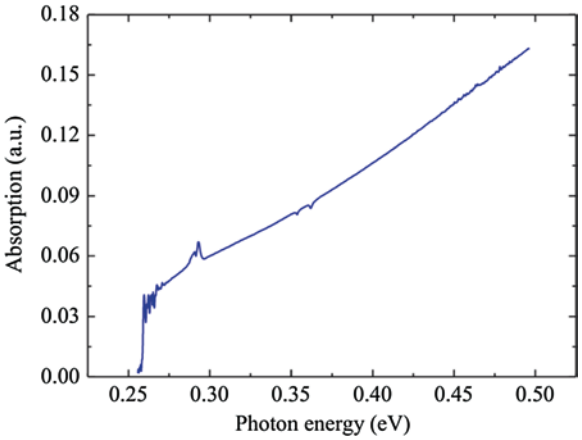

Fig. 4.9 FT-IR absorption versus photon energy $h\nu$

Table 4.2 Thermo-physical parameters of InSb

$m_c(m_e)$	$m_v(m_h)$	ρ(g/cm^3)	C_p(J/g K)	n_0	τ(ns)	α_0(m^{-1})
$0.014\,m_0$	$0.43\,m_0$	~5.50	~0.23	~2.34	50	$\sim 6.10 \times 10^7$
a_l(Å)	η	y_B	c_0	dE_g/dT(eV/K)	E_g(eV)	α_T(K^{-1})
6.478	250	3.063	3.422	$\sim 4.0 \times 10^{-4}$	~0.26	4.64×10^{-6}

The "glass substrate/c-InSb(30 nm)/ZnS–SiO_2(25 nm)" sample is employed in the z-scan measurement. ZnS–SiO_2 is used to ensure that the measured nonlinearity is not influenced by oxidation and thermal expansion of the c-InSb thin films. The optical nonlinearity is measured at a laser wavelength of 405 nm through the z-scan setup, where the duration of laser pulse is 100 ns, and the pulse width is $t_p = 50$ ns, accordingly. Figure 4.10a presents the open-aperture-mode z-scan measurement for nonlinear absorption, and the inset in Fig. 4.10b presents the closed-aperture mode z-scan measurement for mixture of nonlinear absorption and refraction [18]. Pure nonlinear refraction is obtained through dividing the closed-aperture mode data by the open-aperture data, which is displayed in Fig. 4.10b. The data indicate that the sample shows obvious reverse saturation absorption (RSA) and self-focusing refraction characteristics at 405-nm laser pulse irradiation. All of the measurements are repeatable; that is, no structural change in the irradiation area is observed. Therefore, the intensity-dependent transmittance (z-scan curve) originates from laser-induced internal nonlinearity rather than the structural change-induced nonlinearity.

Theoretical fitting demonstrates that the effective nonlinear absorption coefficient β_{eff} is $+4.061 \times 10^{-2}$ m/W, while the effective nonlinear refractive index n_2 is $+7.63 \times 10^{-9}$ m^2/W. The giant optical nonlinearity could play an important role in the nonlinearity-induced super-resolution effect.

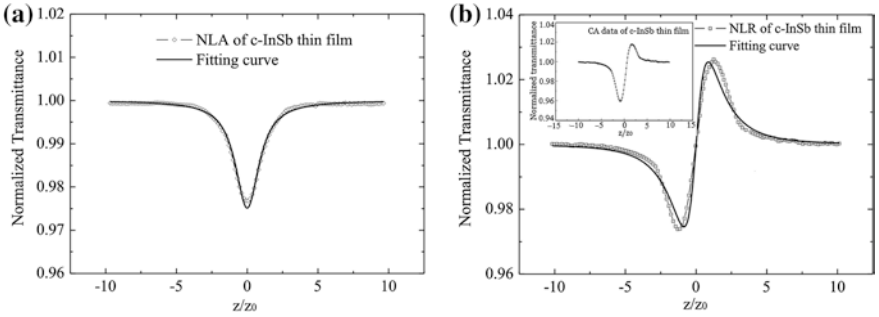

Fig. 4.10 Nonlinear z-scan measurement results at an excitation intensity of 8.15×10^7 W/m^2, **a** optical nonlinear absorption, **b** nonlinear refraction (division of closed-aperture data by open-aperture data). The *inset* of **b** is normalized transmittance of closed-mode measurement

4.3.2.2 Physical Mechanism of Nonlinear Absorption

The intensity-dependent absorption coefficient $\alpha(I)$ for reverse saturation absorption can be written as

$$\alpha(I) = \frac{\alpha_0}{1 - \frac{I}{I_t}} \tag{4.110}$$

where I_t is the maximal intensity of reverse saturation absorption. If the excitation laser intensity is much lower than the maximal intensity, i.e., $I \ll I_t$, then formula (4.110) can be rewritten as

$$\alpha(I) \approx \alpha_0(1 + I/I_t) = \alpha_0 + \beta I \tag{4.111}$$

where $\beta = \alpha_0/I_t$ represents the optical nonlinearity absorption process. Generally speaking, for semiconductor materials under nanosecond laser pulse irradiation, the nonlinear reverse saturation absorption effect is mainly from two aspects: one is thermal effect and the other is electronic processes.

(1) **Thermal contribution to nonlinear absorption**

The nonlinear absorption measurements are carried out at the laser wavelength of 405 nm, and the photo-generated free-carriers energy at 405 nm light wavelength is 3.06 eV, which is much larger than the InSb band gap of ~0.28 eV. The photo-generated free-carriers need to lose thermally ~2.8 eV. The temperature rise occurs in the z-scan measurements due to the much thermal loss of photo-generated free-carriers, and the influence of thermal effect on the nonlinear absorption contribution should be considered, accordingly.

The variable-temperature ellipsometric spectroscopy measurements are conducted in the temperature range of 400–600 K at Ar inert atmosphere. The temperature dependence of the measured extinction coefficient at the light wavelength of 405 nm is plotted in Fig. 4.11a. The corresponding temperature dependence

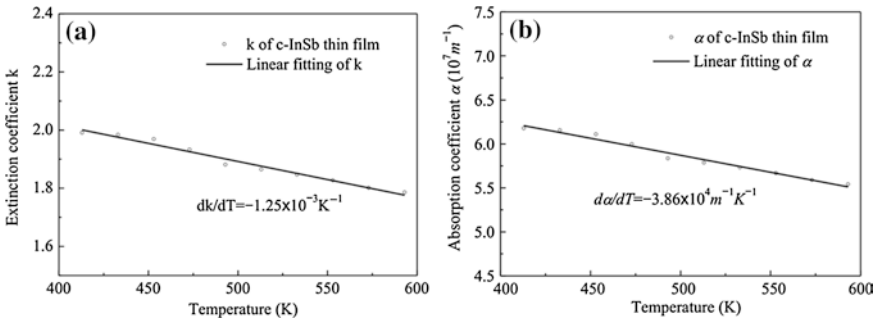

Fig. 4.11 Dependence of **a** the extinction coefficient and **b** absorption coefficient on temperature at 405 nm wavelength in the temperature range from 400 to 600 K

of the absorption coefficient is presented in Fig. 4.11b. The fitted value of $\mathrm{d}\alpha/\mathrm{d}T$ is -3.86×10^4 m^{-1} K^{-1}. The dependence of absorption on temperature may originate from thermally induced lattice constant change, which can be estimated through first-principle calculation. Using formula (4.108) and inserting the parameters in Table 4.2, one can obtain $\beta_t = \sim -10^{-2}$ m/W. The thermally induced nonlinear absorption coefficient β_t has a sign opposite to that of the z-scan measurement of β_{eff}, which indicates that the reverse saturation absorption characteristic mainly originates from electronic contribution induced by free-carriers absorption.

(2) **Electronic contribution to nonlinear absorption**

The reverse saturation absorption can be interpreted with mixed multi-photon absorption effect. The formula (4.40) can be rewritten as

$$\mathrm{d}I/\mathrm{d}z = -\beta_{\mathrm{eff}1} I - \beta_{\mathrm{eff}2} I^2 - \beta_{\mathrm{eff}3} I^3 - \cdots = -\sum_{g=1}^{N} \beta_{\mathrm{eff}} I^g \tag{4.112}$$

Solving formula (4.112), one can then calculate the z-scan transmission by

$$T(z) = \frac{1}{\left\{1 + (g-1)\beta_{\mathrm{eff}} L_{\mathrm{eff}} \left[\frac{I_0}{1+(z/z_0)^2}\right]^{g-1}\right\}^{\frac{1}{g-1}}} \tag{4.113}$$

where g is effective multi-photon absorption level, $L_{\mathrm{eff}} (= \frac{1-\exp(-\alpha_0 L)}{\alpha_0})$ is the effective thickness of sample, and L is thin film thickness. Fitting the curve in Fig. 4.10a with formula (4.113) provides the value of effective nonlinear absorption coefficient $\beta_{\mathrm{eff}} = +4.061 \times 10^{-2}$ m/W. The fitted effective photon absorption number $g = 1.976$, which is very close to the equivalent two-photon absorption number ($g = 2$). The fitting curve is in accordance with experimental data. Thus, the nonlinear absorption of the c-InSb thin films is caused by the laser-induced free-carriers absorption effect rather than thermal effect.

4.3.2.3 Physical Mechanism of Nonlinear Refraction

The intensity-dependent nonlinear refraction is always accompanied with intensity-dependent nonlinear absorption in semiconductors. In consideration of the laser intensity and pulse duration employed in the z-scan measurements, the nonlinear refraction is mainly from electronic nonlinearity γ_e and thermal effect γ_{thermal} for the c-InSb semiconductor. The nonlinear refractive coefficient can be written as

$$\gamma = \gamma_e + \gamma_{\text{thermal}} \tag{4.114}$$

(1) **Thermal contribution to nonlinear refraction**

The time-averaged steady-state temperature rise in the z-scan measurements can be written as

$$\langle \Delta T \rangle = \frac{\alpha_0 t_p}{2\rho C_p} I_0 \tag{4.115}$$

By substituting the thermo-physical parameters in Table 4.2, the temperature rise $\langle \Delta T \rangle$ is calculated to be 218 K in z-scan measurement by ignoring the heat loss through heat conduction, radiation, and natural convection. The temperature of laser irradiation area in z-scan measurement changes from 300 K (room temperature) to approximately 518 K. The thermal contribution to nonlinear refraction in the temperature range can be experimentally analyzed by variable-temperature ellipsometric measurement at 405 nm light wavelength. The results are presented in Fig. 4.12 [18]. The refractive index increases with increased temperature from 400 to 600 K. The corresponding thermo-optic coefficient $\mathrm{d}n/\mathrm{d}T$ derived from Fig. 4.12 is 8.03×10^{-4} K^{-1}.

The nonlinear refractive coefficient from thermal contribution can be calculated through [17]

$$n_2(\Delta T) = \frac{\mathrm{d}n}{\mathrm{d}T} \frac{\Delta T}{I_0} \tag{4.116}$$

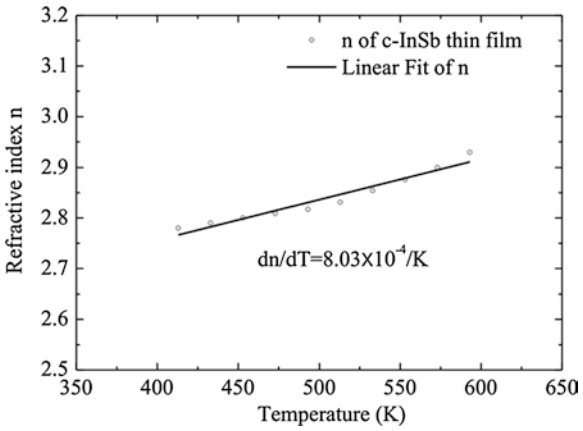

Fig. 4.12 Temperature dependence of refractive index at 405 nm light wavelength in the temperature range from 400 to 600 K

With the values of dn/dT and ΔT, one can calculate the nonlinear refractive coefficient approximately $+2.32 \times 10^{-9}$ m^2/W, which is a positive contribution to the nonlinear refraction, and basically consistent with the z-scan experimental data in the same order of magnitude. One can conclude that the self-focusing feature of c-InSb films is partially due to thermal effect.

The thermal contribution is from two aspects: one is due to density fluctuation (volume change), and the other is temperature rise-induced band gap shrinkage. Using the physical parameters listed in Table 4.2, the temperature dependence of the refractive index caused by thermally induced volume change is $(dn/dT)\Delta\rho = -3.32 \times 10^{-5}$ K^{-1}, which is an order of magnitude smaller than the temperature-variable ellipsometric measurement results shown in Fig. 4.12. The contribution from band gap shrinkage is $(dn/dT)_{dE_g/dT} = 2.22 \times 10^{-3}$ K^{-1}, which is in reasonable agreement with the experimental results of the variable-temperature ellipsometric measurement shown in Fig. 4.12. These results indicate that the thermally induced nonlinear refraction is due to band-shrinking effect.

(2) **Electronic contribution to nonlinear refraction**

The nonlinear reverse saturation absorption of c-InSb thin films is partially from mixed multi-photon absorption-induced free-carriers effect. The nonlinear refraction is also companied in the process of mixed multi-photon absorption. According to Aronov–Auston model ($\gamma_e = -\frac{\alpha_0 \tau e^2}{2\varepsilon_0 n_0 \hbar\omega^3} \frac{1}{\mu} \frac{E_g^2}{E_g^2 - (\hbar\omega)^2}$) and the parameters listed in Table 4.2, $\gamma_e \sim +1.69 \times 10^{-9}$ m^2/W which is also a positive contribution to the nonlinear refraction, and close to the z-scan experimental result, and the self-focusing feature of c-InSb films is partially due to electronic contribution. Therefore, the nonlinear refraction of the c-InSb thin films is from two parts: one is from electronic transition contribution, and the other is from thermal effect.

4.3.3 Nonlinear Reverse Saturation Absorption of AgInSbTe Thin Films

Amorphous AgInSbTe (AIST) thin films are deposited on the glass substrate by the radio frequency magnetron-control sputtering method at the room temperature. The crystalline state is obtained by laser irradiating the amorphous AIST thin films. Nonlinearity is measured using the z-scan setup where He–Ne laser with a wavelength of 633 nm is used. Continuous He–Ne laser is modulated into pulse light by an acoustic–optic modulator; in this measurement $t_p = 50$ ns.

4.3.3.1 Optical Nonlinearity of Amorphous AIST Thin Films

Figure 4.13 shows the typical open-aperture-mode z-scan measurement of the amorphous AIST thin films with the light wavelength of 633 nm and pulse width

of $t_p = 50$ ns. It is obviously observed that the sample is of the reverse saturation absorption. The laser pulse width of about 50 ns is in the timescale that free-carriers dominate in the absorption process. The band gap of amorphous AIST is 1.42 eV, which is smaller than the photon energy of 633 nm laser; the nonlinearity may be from free-carriers absorption process.

Utilizing the formula (4.113), the fitted β_{eff} value in Fig. 4.13 is 7.53×10^{-3} m/W. The fitting curve also gives the $g = 1.722$, indicating that the reverse saturation absorption of amorphous AIST thin film comes from the mixed multi-photon absorption (including single-photon absorption and two-photon absorption process).

Figure 4.14 is dependence of nonlinear absorption coefficient on the light intensity. β_{eff} gradually increases with the light intensity, which may result from the temperature-induced band gap shrinkage. The g decreases gradually with increasing laser intensity, which indicates that single-photon absorption plays more important role in the equivalent mixed multi-photon absorption process.

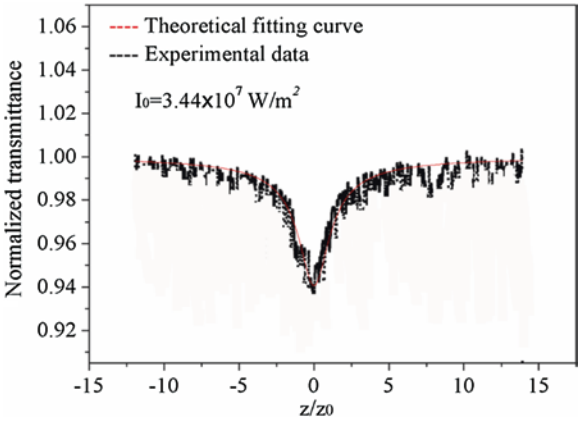

Fig. 4.13 Experimental (*black*) and theoretical fitting (*red*) *z*-scan curves of the amorphous AIST thin film. Reprinted with permission from Ref. [19]. Copyright 2009, American Institute of Physics

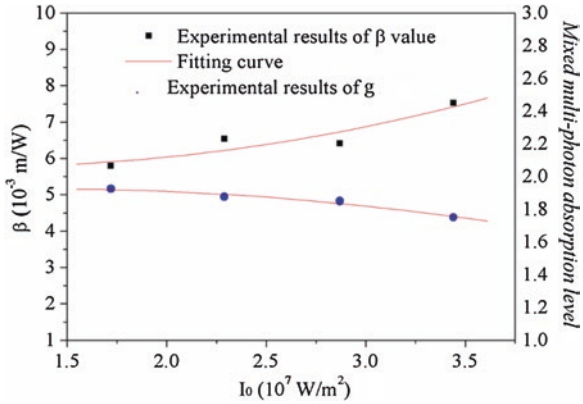

Fig. 4.14 Dependences of nonlinear absorption coefficient β_{eff} and g on the light intensity. Reprinted with permission from Ref. [19]. Copyright 2009, American Institute of Physics

4.3.3.2 Optical Nonlinearity of Crystalline AIST Thin Films

Figure 4.15 shows a typical open-aperture-mode z-scan measurement result of the crystalline AIST thin films with the light wavelength of 633 nm and pulse width of $t_p = 50$ ns. It is obviously observed that the sample is of the reverse saturation absorption. Using formula (4.113), the effective nonlinear absorption coefficient $\beta_{\text{eff}} = 3.5 \times 10^{-2}$ m/W at the laser intensity $I_0 = 1.15 \times 10^8$ W/m^2. The band gap of crystalline state of AIST is about 1.39 eV, smaller than the photon energy, and the experimental effective multi-photon level $g = 1.7011$.

Figure 4.16 shows the relationship between the laser intensity I_0 and the effective nonlinear absorption coefficient β_{eff} as well as the effective mixed multi-photon absorption level g of crystalline AIST. β_{eff} is gradually ascending tendency for the increasing laser intensity, and the g is slight descending tendency with increasing laser intensity.

The large nonlinear absorption coefficient of c-AIST thin films comes from the laser-induced free-carriers effect. The AIST thin films are metallic-type semiconductor; the free-carriers can be easily created by laser irradiation because of

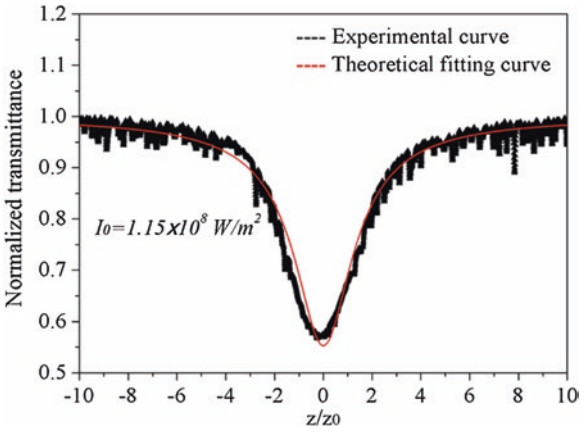

Fig. 4.15 Open aperture z-scan curves for c-AIST thin films ($I_0 = 1.15 \times 10^8$ W/m^2). Reprinted with permission from Ref. [19]. Copyright 2009, American Institute of Physics

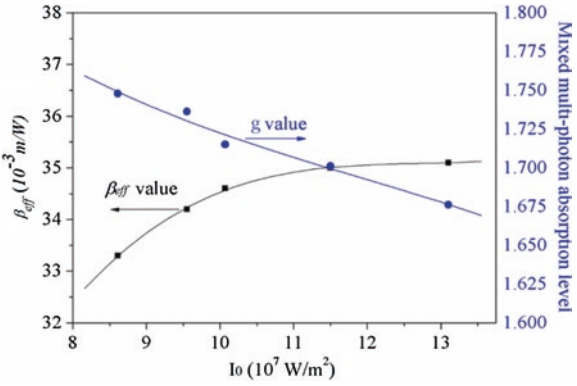

Fig. 4.16 Relationship between the light intensity I_0 and the effective absorption coefficient β_{eff} as well as the effective multi-photon absorption level g for c-AIST. Reprinted with permission from Ref. [19]. Copyright 2009, American Institute of Physics

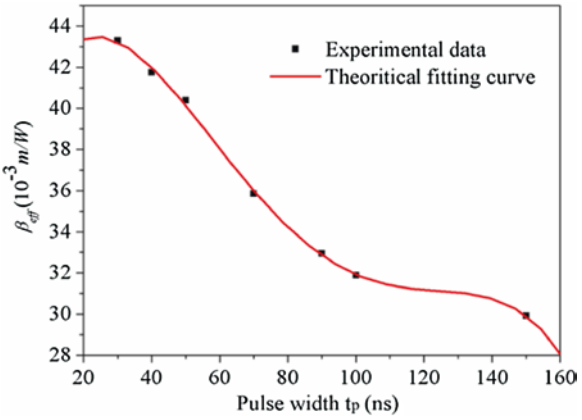

Fig. 4.17 Dependence of β_{eff} on laser pulse width t_p for c-AIST thin films. Reprinted with permission from Ref. [19]. Copyright 2009, American Institute of Physics

their semiconductor-like energy band structure. Despite the population decays in the pulse time, the free-carriers density is locally balanced and approaches to the saturation state when laser intensity $I(z)$ increases to the threshold. When the laser intensity increases above the threshold, the free-carriers absorb photons to higher states, namely intra-band absorption. Then, these higher-state free-carriers release excess energy by the phonons of crystal lattice in a longer time range than the laser pulse, releasing the energy to the lattice. Figure 4.17 is the dependence of β_{eff} on laser pulse width. It is obvious that the free-carrier diffusion depends on the irradiation laser pulse width; thus, the short laser pulse can cause small free-carriers diffusion, and obtain larger nonlinear effect, accordingly. The β_{eff} decreases with increasing laser pulse width.

4.3.4 Nonlinear Absorption Reversal of c-$Ge_2Sb_2Te_5$ Thin Films

4.3.4.1 Nonlinear Absorption Performance of c-$Ge_2Sb_2Te_5$ Thin Films

The samples with about 40 nm thickness are deposited from a nominal $Ge_2Sb_2Te_5$ target using direct-current magnetron sputtering method and then annealed in vacuum for 30 min at 180 °C to obtain the crystalline $Ge_2Sb_2Te_5$ (c-GST). Optical nonlinearity is measured using an improved single-beam z-scan technique, where a semiconductor laser with a wavelength of 660 nm is used as excitation light source. The laser beam is modulated into 40-ns pulse light using signal generator.

Figure 4.18a shows typical open-aperture z-scan results for c-GST films with the laser intensities of $0.02, 0.07, 0.17$, and $0.31\ \text{GW/m}^2$. At the excitation intensity $0.02\ \text{GW/m}^2$, the open-aperture z-scan measurement exhibits an saturation absorption. As the intensity increases from 0.07 to $0.17\ \text{GW/m}^2$, the nonlinear absorption response changes from saturation absorption to reverse saturation absorption. As the intensity further increases to $0.3\ \text{GW/m}^2$, the curve shows

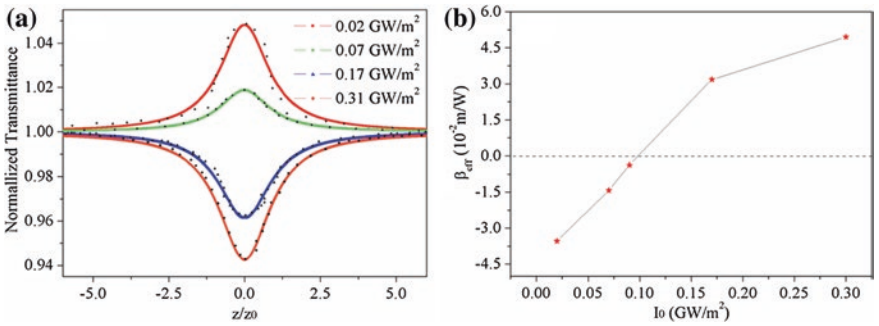

Fig. 4.18 **a** Open-aperture z-scan results for the c-GST at excitation intensities of 0.02 GW/m^2 (*squares*), 0.07 GW/m^2 (*circles*), 0.17 GW/m^2 (*triangles*), and 0.31 GW/m^2 (*diamonds*). The solid lines are the best-fitting curves calculated by using the z-scan theory. **b** Effective nonlinear absorption coefficient β_{eff} as a function of intensity. Reprinted with permission from Ref. [20]. Copyright 2011, American Institute of Physics

an enhanced reverse saturation absorption. No obvious nonlinear refraction is observed. When the excitation intensity is lower than the threshold of approximately 0.32 GW/m^2, the nonlinear absorption reversal becomes repeatable. The reversal results from electron-associated process and not from laser-induced structural change.

Generally, there exist saturation absorption and reverse saturation absorption for semiconductors upon irradiation, the saturation absorption is usually attributed to band-filling effect, and reverse saturation absorption can be induced by mixed multi-photon absorption effects. The z-scan transmittance curve is fitted by

$$T_t(z) = \frac{1}{\sqrt{\pi} q_0(z, t=0)} \int_{-\infty}^{+\infty} ln\left[1 + q_0(z, t=0)\right] e^{-\tau^2} \mathrm{d}\tau \qquad (4.117)$$

where $q_0(z, t=0) = \beta_{\text{eff}} I_0 L_{\text{eff}} (1 + z^2/z_0^2)$, I_0 and L_{eff} are the peak intensity at the focal plane and the effective thickness of sample, respectively. z_0 is Rayleigh length of z-scan setup. Utilizing formula (4.117), the open aperture mode z-scan measurements for intensities of 0.02, 0.07, 0.17, and 0.31 GW/m^2 are fitted, the fitted nonlinear absorption coefficients are -3.54×10^{-2}, -1.43×10^{-2}, $+3.18 \times 10^{-2}$, and $+4.95 \times 10^{-2}$ m/W, respectively, as shown in Fig. 4.18b.

The reversal of nonlinear absorption suggests that the dominant nonlinear process changes at different intensities. The optical absorption spectrum of c-GST shown in Fig. 4.19 reflects the electronic transition characteristics. The absorption coefficient at the excitation wavelength of 660 nm is about 5.55×10^7/m, indicating that c-GST has a strong absorption at 660 nm light wavelength in the form of single-photon absorption. More importantly, the Fig. 4.19 also shows that the absorption coefficient at the wavelength of 330 nm, equivalent to two photons at 660 nm in energy, is larger than that of excitation wavelength 660 nm, indicating that c-GST has the potential to absorb two photons at the excitation laser

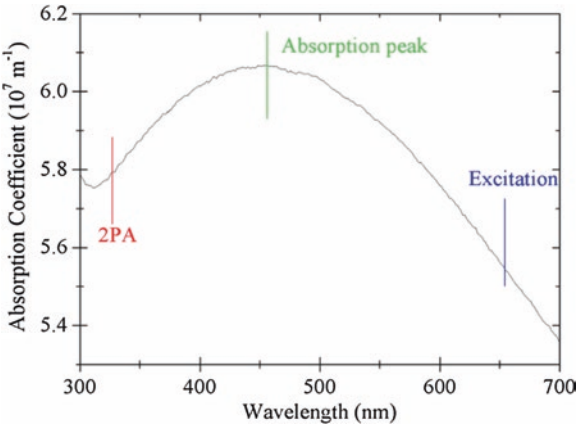

Fig. 4.19 Dependence of absorption coefficient on wavelength for *c*-GST thin films. Reprinted with permission from Ref. [20]. Copyright 2011, American Institute of Physics

wavelength at one time. In addition, there is an absorption maximum between 660 and 330 nm wavelength, and this absorption peak band can provide available states for the relaxation process of the electrons during the two-photon absorption, and the peak band can prevent the excited electrons from relaxing rapidly to the ground state. Consequently, the ground-state electron bleaching may occur, and the dominant electronic transition responsible for nonlinear absorption can be changed from ground-state absorption to free-carriers absorption. This physical picture provides a possible explanation for nonlinear absorption reversal.

4.3.4.2 Five-Band Model for the Reversal of Nonlinear Absorption Performance

The total density of states (TDOS) is presented in the inset of Fig. 4.20a, where one is only interested in the levels that should be involved in the electronic transition regarding the excitation photon energy of 1.88 eV. In Fig. 4.20a, three DOS peaks locating around 1.7, 3.75, and 2.82 eV are marked as S_1, S_2 and S_3, respectively. Each DOS peak indicates a great number of available electronic states. In the electronic transition from the ground state to the valence band, the electrons at the ground state are excited to the S_1 energy band only if the energy is more than 1.7 eV; this can be easily realized by absorbing one photon with energy of 1.88 eV. The excited electron relaxes to ground state in a very short time.

However, if the ground electron absorbs two photons with energy of 3.76 eV, it will be directly excited to the S_2 energy band with energy gap of 3.75 eV. The excited electron relaxes firstly toward the S_3 band, which can delay the excited electrons relaxing back to the ground state. As the excitation intensity increases, the delay effect will be more pronounced and induce the bleaching of the ground states. Thus, three DOS peaks of S_1, S_2 and S_3 can be used to construct a five-level

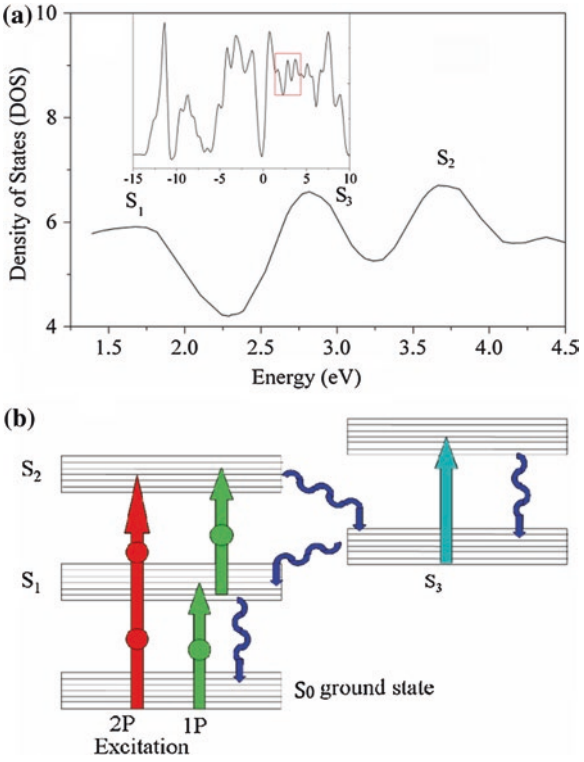

Fig. 4.20 **a** Total density of states (TDOS) for *c*-GST at the excitation range, three DOS peaks labeled as S_1, S_2, and S_3, which can be used to construct five-level model structure. **b** Five-level model showing transition of electrons for reverse saturation absorption. Reprinted with permission from Ref. [20]. Copyright 2011, American Institute of Physics

model, which can be used to explain the reverse saturation absorption process and demonstrate the reversal of nonlinear absorption. The corresponding simplified model is shown in Fig. 4.20b. At low excitation intensity, the $S_0 \rightarrow S_1$ transition in the form of single-photon absorption prevails and the further single-photon absorption will be blocked when most of states in S_1 band are occupied; this process induces a saturation absorption effect, which agrees with the *z*-scan measurement for intensities of 0.02 and 0.07 GW/m^2. For higher excitation light intensity, the probability of $S_0 \rightarrow S_2$ transition in the form of two-photon absorption becomes important, and the transition is conducted either by direct ground-state electron absorption or by two-step process. The first transition is $S_0 \rightarrow S_1$, and the second transition is from $S_1 \rightarrow S_2$ by excited free-carriers absorption. The large number of vacant states at S_3 band can accept excited electrons relaxed from S_2 band and delay the excited electrons from relaxing toward ground state, leading to the decrease of the number of ground state electrons. As the excitation light intensity increases, the two-photon absorption gradually dominates over single-photon absorption, which causes the nonlinear absorption to change from saturation absorption to reverse saturation absorption process. As the excitation intensity increases further, the electrons at the ground state tend to bleach and cause an enhanced reverse saturation absorption; this process is manifested by the *z*-scan measurement for intensities of 0.17 and 0.31 GW/m^2.

Table 4.3 Thermo-physical parameters for *c*-GST

$m_c(m_e)$	$m_v(m_e)$	$\alpha_0(\mathrm{m}^{-1})$	c_p(J/Kg)	$\rho(\mathrm{Kg/m}^3)$	E_g(eV)	$mP_M^2/\hbar^2$(eV)	n_0
0.014	0.4	4.9×10^7	210	6.4×10^3	0.5	10.5	2.62

In the above analysis, saturation absorption effect is attributed to the blocking of single-photon absorption by free-carriers in the conduction band. To check whether the band-filling model can quantitatively account for the saturation absorption for *c*-GST at low excitation light intensity, by referring to formulas (4.16)–(4.27) and inserting the parameters in Table 4.3 [20], the nonlinear absorption coefficient $\beta_{\mathrm{eff}} = -3.2 \times 10^{-2}$ m/W, which agrees well with *z*-scan result of -3.54×10^{-2} m/W at the laser intensity 0.02 GW/m^2.

4.3.4.3 Thermal Contribution to Nonlinear Absorption

The thermal effect also plays an important role in determining the nonlinearity of *c*-GST, especially for nanosecond laser pulse excitation. According to formula (4.115) and inserting parameters in Table 4.3, the calculated temperature rises are 14.6, 51, 124, and 220 K for intensities of 0.02, 0.07, 0.17, and 0.31 GW/m^2 in the *z*-scan measurement, respectively. The *z*-scan measurement is performed at 300 K; thus, the real sample temperature is in the range from 314.6 to 520 K accordingly, as shown in Fig. 4.21a.

According to formula (4.108), the thermally induced nonlinear absorption coefficient β_t is proportional to $\mathrm{d}\alpha/\mathrm{d}T$. If $\mathrm{d}\alpha/\mathrm{d}T < 0$, then $\beta_t < 0$, which corresponds to a nonlinear saturation absorption process. If $\mathrm{d}\alpha/\mathrm{d}T > 0$, then $\beta_t > 0$, which corresponds to a nonlinear reverse saturation absorption process. To obtain $\mathrm{d}\alpha/\mathrm{d}T$ for *c*-GST, a variable-temperature ellipsometric spectroscopy measurement is performed in the temperature range from 300 to 530 K in vacuum. The temperature dependence of the absorption coefficient at the light wavelength of 660 nm is plotted in Fig. 4.21b. Figure 4.21b shows that the absorption coefficient curve can be divided into three regions: The region A is in the temperature range from

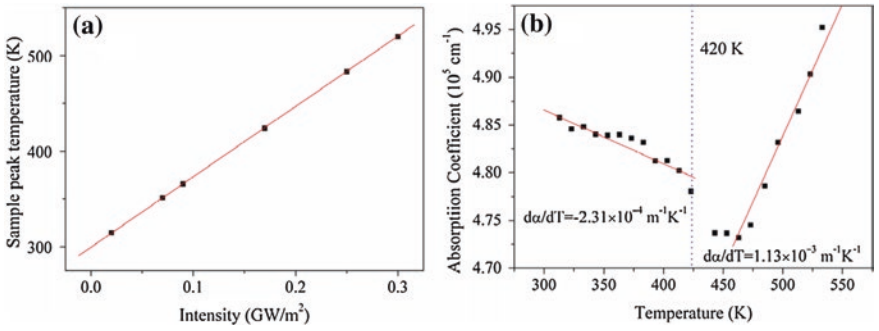

Fig. 4.21 **a** Relation between sample peak temperature and irradiation intensity. **b** Temperature dependence of absorption coefficient within the temperature range from 300 to 530 K. Reprinted with permission from Ref. [20]. Copyright 2011, American Institute of Physics

300 to 420 K, α decreases with temperature, and the linear fitting indicates that $d\alpha/dT = -2.31 \times 10^{-4}$ m^{-1} K^{-1}. The region B is in the temperature from 420 to 470 K; α remains almost constant. The region C is in the temperature range from 420 to 530 K; α increases with temperature, the fitting $d\alpha/dT = 1.13 \times 10^{-3}$ m^{-1} K^{-1}. Combining Fig. 4.21a and b, one can find that for laser intensities of 0.02 and 0.07 GW/m^2, the sample temperature is in the range from 310 to 420 K, and the z-scan measurement is found to be saturation absorption. For the laser intensity of 0.17 and 0.31 GW/m^2, the sample temperature is in the range 420–530 K, and the z-scan measurement is found to be reverse saturation absorption. Thus, the thermally induced absorption coefficient change agrees with the z-scan measurement.

Based on Fig. 4.21b and formula (108), in the temperature range from 310 to 420 K, the $\beta_t = -6.9 \times 10^{-3}$ m/W and in the temperature range from 420 to 530 K, the $\beta_t = +3.38 \times 10^{-2}$ m/W. It is worth to point out that at $T < 420$ K, the sign of β_t agrees with z-scan measurement at the low laser intensity of 0.02 and 0.07 GW/m^2, while the value is one order smaller than the experimental result. At $T > 470$ K, both the sign and value of β_t agree well with the z-scan measurement result at high laser intensity of 0.31 GW/m^2.

In order to exclude the thermally induced nonlinearity, which is usually indicative of thermal lens effect, another z-scan measurement with 21-ps laser pulse is performed for c-GST at the laser intensities of 1,256, 2,109, 3,1813, and 4,710 GW/m^2, respectively. The rise time of thermal lens in a solid is determined by the acoustic transmit time $\tau_{\text{transient}} = w_0/v_s$, where w_0 is the beam waist radius and v_s is the propagation speed of sound in the solid. For c-GST with $v_s = 3.2 \times 10^3$ m/s and $w_0 = 4.3$ um for the z-scan setup, one obtain $\tau_{\text{transient}} = 1.3$ ns, which is almost two order of magnitude larger than the 21-ps laser pulse. Thus, the thermally induced nonlinearity can be neglected in the picosecond z-scan. The measured results are plotted in Fig. 4.22; the β

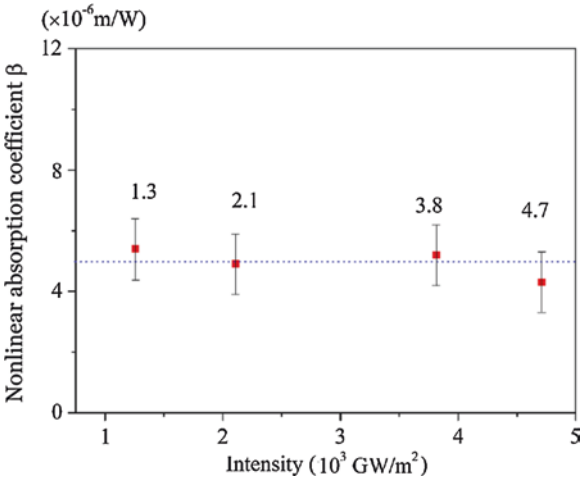

Fig. 4.22 Effective nonlinear absorption coefficient β measured by picosecond z-scan at different excitation intensities, the dashed line showing the average magnitude of β measured at the picosecond timescale. Reprinted with permission from Ref. [20]. Copyright 2011, American Institute of Physics

is at the order of 10^{-6} m/W, which is much smaller than the nanosecond z-scan result. Therefore, the picosecond z-scan experimental results give a solid support for the influence of thermal effect on nonlinearity in the nanosecond z-scan experiment.

4.3.5 Nonlinear Saturation Absorption and Refraction of Ag-doped Si Thin Films

4.3.5.1 Microscopic Structures and Optical Spectrum Characteristics

The Ag-doped Si thin films are deposited on BK7 glass substrate by cosputtering of Ag and Si targets. The Ag concentration is adjusted from 10 to 90 % by changing sputtering power of the Ag target.

The film morphology is characterized with an optical microscope and a scanning electron microscope (SEM). Both optical and SEM topographic images are presented in Fig. 4.23. The bright particles on the SEM pictures are the Ag particles, and segregated Ag nanoparticles are formed and randomly distributed on the surface with a large volume fraction. The SEM images in Fig. 4.23b, c are taken from different positions (indicated in Fig. 4.23a) of the same sample. The sample is irradiated by a 633-nm-wavelength laser pulse with a duration of 50 ns and power intensity of about 9×10^7 W/m^2. In Fig. 4.23a, the blank point is the area upon which the laser beam is not pointed, while the dark point is the area upon which the laser beam is pointed. The obvious difference between the pre-laser irradiation and post-laser irradiation images is that the concentration and tightness of the Ag clusters increase after their laser pulse irradiation. That is, there is a laser-induced modification of the cluster concentration and configuration, and the metal particles are aggregated with smaller distances after the laser irradiation.

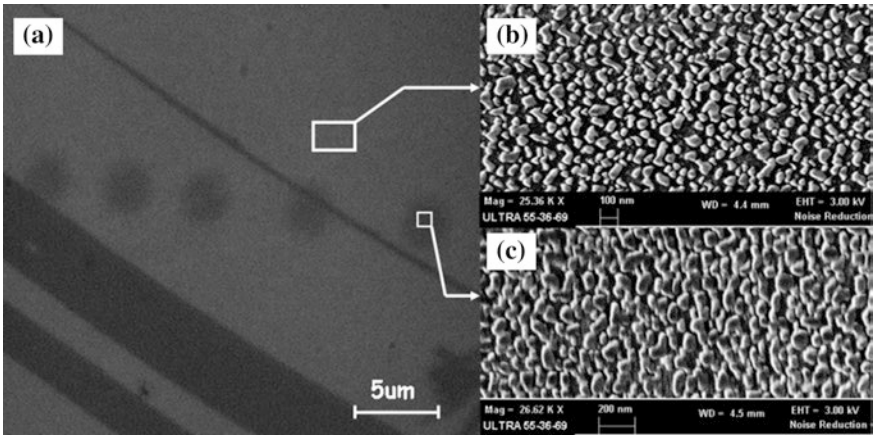

Fig. 4.23 Optical and SEM images of Ag-doped Si thin films, **a** optical image of the thin film with laser irradiation points; **b** SEM image of the blank area on (**a**); **c** SEM image of the irradiation area on (**a**). Reprinted from Ref. [21], with kind permission from Springer Science+Business Media

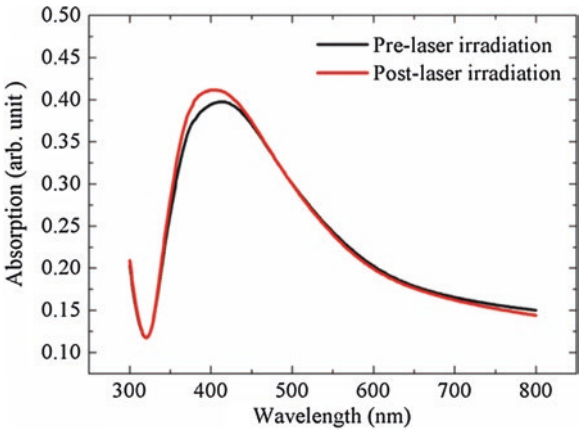

Fig. 4.24 Optical absorption spectrum of pre-laser irradiation and post-laser irradiation. Reprinted from Ref. [21], with kind permission from Springer Science+Business Media

Optical absorption spectrum is then measured in a wavelength range from 300 nm to 800 nm. The resulting curves for pre-laser irradiation and post-laser irradiation samples are presented in Fig. 4.24. Obviously, there is an absorption peak at about 415 nm wavelength, and the excitation at 415nm can be identified as the surface plasmon resonance absorption (SPR) peak, stemming from the surface plasmon between metal particles and matrix. In addition, the red curve in Fig. 4.24 has a higher peak value than that of the black one, indicating that the SPR absorption is enhanced after the laser irradiation. As is shown in Fig. 4.23, the local clusters become more concentrated and the inter-particle distances are reduced after the sample is exposed to the laser beam.

4.3.5.2 Optical Nonlinearity of Ag-doped Si Thin Films

In z-scan measurement, when the Ag concentration is below or above about 50 %, there are small nonlinear responses whether in the pre-laser irradiation area or in the post-laser irradiation area. Large nonlinear responses are recorded when samples are near 50 % Ag concentration. The measured results are shown in Fig. 4.25.

The measured z-scan curves corresponding to the pre-laser irradiation area are drawn in Fig. 4.25a, and the curves corresponding to the post-laser irradiation area are illustrated in Fig. 4.25b. The black curves represent the normalized closed-aperture-mode z-scan measurements, and the red curves represent the open-aperture-mode z-scan measurements. On the other hand, the green curves are the results of normalized closed-aperture mode divided by normalized open aperture mode, which indicates the optical nonlinear refraction. It is clear from Fig. 4.25a that, for the pre-laser irradiation z-scan measurement, there is no evidence of significant nonlinear refraction. In addition, there is an inflexion in the open-aperture-mode measurement, indicating that there exists a threshold, at which the laser triggers Ag particle aggregation.

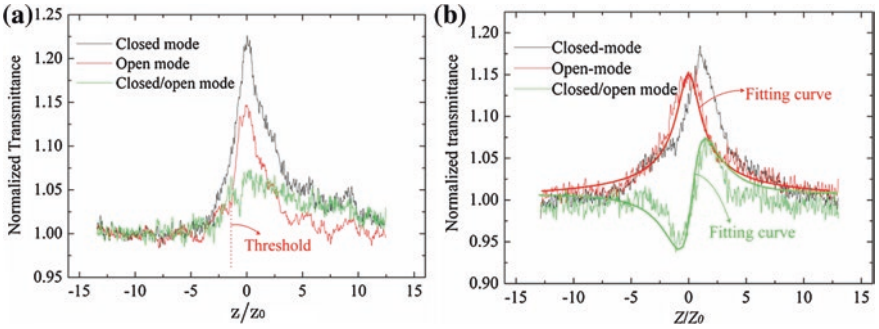

Fig. 4.25 Initial **a** and repeated **b** normalized transmittance of closed-aperture-mode (*black*) and open-aperture-mode (*red*) z-scan measurement, and the *green curves* are the result of normalized closed-aperture mode divided by normalized open aperture mode. Reprinted from Ref. [21], with kind permission from Springer Science+Business Media

For post-laser irradiation, there is no inflexion in the open-aperture-mode z-scan measurement, and the curve is symmetric. This behavior indicates an apparent nonlinear refraction and suggests that the thin film functions as a convex lens under laser irradiation, i.e., a self-focusing effect. The theoretical fittings of the nonlinear absorption and refraction in Fig. 4.25b indicate nonlinear saturation absorption. The effective nonlinear absorption coefficient is $\beta_{\text{eff}} = -8.086 \times 10^{-2}$ m/W. The nonlinear refraction coefficient $\gamma = 1.47 \times 10^{-9}$ m^2/W, which is several orders of magnitude larger than most of the reported metal-doped composites. The z-scan measurements for the post-laser irradiation area are taken repeatedly, and the results of each measurement remain almost unchanged, suggesting that this nonlinearity is reversible.

4.3.5.3 Concentration Resonance-Induced Large Nonlinearity

The large nonlinear signals could only be recorded with a Ag concentration in a narrow range near 50 %, and out of the concentration range, the signals are weak. The nonlinear absorption can reach saturation, while the nonlinear refraction creates self-focusing effect. The process of generating nonlinear signals is accompanied with a laser-induced reconstruction of the Ag nanoparticles. Without the modification of particle distribution, no signal could be generated.

Due to the fact that shows correlation between the enhancement and the redistribution of the Ag particles, it is reasonable to conjecture that the interaction among the particles would play an important role and be principal cause of the giant enhancement. The enhancement responses come mainly from the scattering field of Ag particles. If the local field driving the particles would strongly be enhanced due to the strong interaction among the particles, the optical response of the whole system would increase strongly.

The numerical calculation of the local field is carried out and in turn the scattering field goes toward the observation area. The sample is modeled as a cluster of 50 Ag particles. The distance among the particles and the size of the particles is randomly generated. To simplify the model, the rest of the sample is considered to be a surface beneath the cluster of Ag particles and the particles are two-dimensionally distributed on the surface as depicted in Fig. 4.26a, b. A suitable parameter to take into account the degree of the inter-particle couplings is the tightness of the cluster, i.e., an overall account of the distances among the particles. For example, in Fig. 4.26a, the particles are tightly distributed, whereas in Fig. 4.26b the particles are loosely redistributed in a larger area. One understands that the interaction would be enhanced once the tightness increases. Actually, one finds out that the situation would be more complicated.

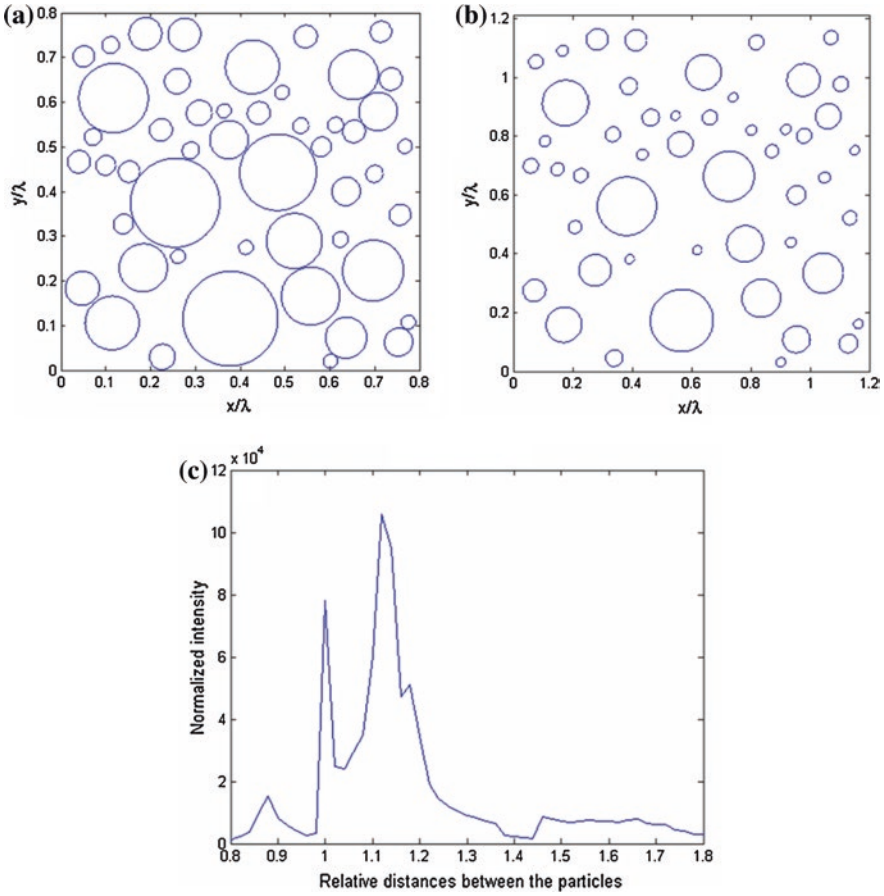

Fig. 4.26 Cluster of 50 Ag particles with **a** tight distribution and **b** loose distribution, and **c** the optical response as a function of the tightness, from tight distribution to loose distribution. Reprinted from Ref. [21], with kind permission from Springer Science+Business Media

A microscopic theory to calculate the optical response as a function of the tightness of the cluster is employed. The calculation takes two steps, namely solving the self-consistent set of formulas that account for all microscopic interactions among the particles, and relating the local field to the scattering field in the observation area. The two typical formulas, respectively, for each step can be written as follows:

$$\vec{E}_i(\omega) = \vec{E}_i^0(\omega) - \mu\omega^2 \sum_{j=1}^{N} \overleftrightarrow{G}(\vec{r}_i, \vec{r}_j, \omega) \cdot \left[\overleftrightarrow{\alpha}_j(\omega) \cdot \vec{E}_j(\omega)\right], \quad j = 1, 2, 3, \ldots, N \tag{4.118}$$

$$\vec{E}\left(\vec{R}, \omega\right) = -\mu\omega^2 \sum_{i=1}^{N} \overleftrightarrow{G}\left(\vec{R}_i, \vec{r}_i, \omega\right) \cdot \left[\overleftrightarrow{\alpha}_i(\omega) \cdot \vec{E}_i(\omega)\right] \tag{4.119}$$

In the first formula, the local field at each particle site is a sum of the contributions from the rest of the particles with polarizability strength $\overleftrightarrow{\alpha}_j(\omega)$ and the interaction among particles is included in the field propagator $\overleftrightarrow{G}(\vec{r}_i, \vec{r}_j, \omega)$. Solving the set of formulas self-consistently as a standard eigenvalue problem, one expects that the local field can be accurately calculated with the possibility to have resonance. Once the local field is calculated, the observed field at a point far away from the cluster at $\vec{R}$ is calculated using the second formula representing the optical response of the cluster. The details of the field propagator as well as other parameters involved in the numerical simulation can be found in [22]. In the calculation, the length is normalized to the wavelength. Therefore, the normalized length marked in Fig. 4.26a, b for a wavelength of 633 nm corresponds to a range about 500–800 nm, respectively.

In Fig. 4.26c, the calculated optical response is presented as a function of the tightness. A notable feature in the curve is the resonance. When the particles are far away from each other, the responses are small. When the tightness increases, there are possibilities to excite the resonances and the scattering field is enhanced strongly of several orders of magnitude. If the particles are further pushed together, the field strength decreases again. The above behavior of the optical response appears to be consistent with experimental results; that is, the optical response increases with the density of the Ag clusters, and the enhancement reaches a peak before becoming saturated. That is, the large nonlinearity occurs at suitable Ag concentration, which is attributed to the laser-triggered concentration resonances.

4.4 Summary

In nonlinear super-resolution nano-optics, in order to obtain the below diffraction-limited spot, the nonlinear absorption and refraction coefficients of samples are required to be large. According to the manipulations from free-carriers, gap shrinkage, and thermal effect, the semiconductor thin films (specially for chalcogenide

phase-change thin films) generally possess large nonlinear effects, and the nonlinear absorption coefficient can reach up to the magnitude of $10^{-2} \sim 10^{-3}$ m/W, and nonlienar refraction coefficient is about $10^{-10} \sim 10^{-9}$ m^2/W. Thus, these semiconductor materials can be used to obtain optical super-resolution effect.

References

1. E. Garmire, A. Kost, *Nonlinear Optics in Semiconductors: Semiconductors and Semimetals* (Academic, USA, 1999)
2. D.A.B. Miller, C.T. Seaton, M.E. Prise, S.D. Smith, Band-gap-resonant nonlinear refraction in III-V semiconductors. Phys. Rev. Lett. **47**(3), 197–200 (1981)
3. J. Liu, S. Liu, J. Wei, Origin of the giant optical nonlinearity of Sb_2Te_3 phase change materials. Appl. Phys. Lett. **97**, 261903 (2010)
4. D.H. Auston, S. McAffee, C.V. Shank, E.P. Ippen, O. Teschke, Picosecond spectroscopy of semiconductors. Solid State Electron. **21**, 147–150 (1978)
5. T.S. Moss, Theory of intensity dependence of refractive index. Phys. Stat. Solidi B **101**, 555–561 (1980)
6. B.S. Wherrett, A.C. Walker, F.A.P. Tooley, Nonlinear refraction for CW optical bistability, in *Optical Nonlinearities and Instabilities in Semiconductors*, ed. by H. Haug (Academic, New York, 1999), p. 239
7. M. Shui, *Excited state nonlinear refractive dynamics in semiconductor materials*. Dissertation for the doctoral degree in science of Harbin Institute of Technology, 2011
8. Y. Fang, *Research on thermal optical nonlinearity induced by nonlinear absorption*. Dissertation for the master degree in science of Harbin Institute of Technology, 2007
9. D.N. Christodoulides, I.C. Khoo, G.J. Salamo, G.I. Stegeman, E.W. Van Stryland, Nonlinear refraction and absorption: mechanisms and magnitudes. Adv. Opt. Photon. **2**, 60–200 (2010)
10. X. Cai, *Optical and thermal characteristics of InSb semiconductor thin films*, Dissertation for the master degree of University of Chinese Academy of Sciences, 2013
11. Y.P. Varshni, Temperature dependence of the energy gap in semiconductors. Physica **34**, 149–154 (1967)
12. M. Levinshtein, S. Rumyantsev, M. Shur, *Handbook Series on Semiconductor Parameters*, vol. 1 (World Scientific, London, 1995)
13. L. Prodhomme, A new approach to the thermal change in the refractive index of glasses. Phys. Chem. Glasses **1**(4), 119–122 (1960)
14. J. Wei, F. Zhou, Y. Wang, F. Gan, Y. Wu, D. Gu, Optical near-field simulation of Sb thin film thermal lens and its application in optical recording. J. Appl. Phys. **97**, 073102 (2003)
15. B. Jensen, A. Torabi, Temperature and intensity dependence of the refractive index of a compound semiconductor. J. Opt. Soc. Am. B **2**(9), 1395–1401 (1985)
16. S. Liu, J. Wei, F. Gan, Nonlinear absorption of Sb-based phase change materials due to the weakening of the resonant bond. Appl. Phys. Lett. **100**, 111903 (2012)
17. X. Cai, J. Wei, H. Yan, Optical nonlinear characteristics of amorphous InSb thin film and its super-resolution effect. Acta Opt. Sin. **33**, 0931002 (2013). (in Chinese)
18. X. Cai, J. Wei, Optical nonlinearity characteristics of crystalline InSb semiconductor thin films. J. Phys. D Appl. Phys. **46**, 435101 (2013)
19. J. Liu, J. Wei, Optical nonlinear absorption characteristics of AgInSbTe phase change thin films. J. Appl. Phys. **106**, 083112 (2009)

20. S. Liu, J. Wei, F. Gan, Optical nonlinear absorption characteristics of crystalline $Ge_2Sb_2Te_5$ thin films. J. Appl. Phys. **110**, 033503 (2011)
21. J. Wei, J. Liu, M. Xiao, Giant optical nonlinearity of silver-doped silicon thin film at low power input: laser triggered cluster resonance. Appl. Phys. A **104**, 1031–1037 (2011)
22. M. Xiao, A. Zayats, J. Siqueiros, Scattering of surface-plasmon polaritons by dipoles near a surface: optical near-field localization. Phys. Rev. B **55**, 1824–1837 (1997)

Chapter 5
Nanoscale Spot Formation Through Nonlinear Refraction Effect

5.1 Introduction

The diffraction limit of an optical lens system refers to the smallest focused spot that the lens can produce, which is related to the wavelength and the numerical aperture (NA) of the lens system as $D \sim 1.22\lambda/\mathrm{NA}$. For example, for a laser beam of $\lambda = 650$ nm and a lens of NA $= 0.6$, one estimates the size of the diffraction limit spot $D \sim 1.321$ μm. Overcoming the diffraction limit and obtaining a super-resolution spot have been hot topic because of the important demands in the fields of nanolithography, ultrahigh-density data storage, and nanoscale-resolved optical imaging and detection. Numerous methods and techniques have been proposed to overcome the limit, including the use of a metamaterial lens [1], microsphere-based microscopic lens [2], and fluorescence labeling methods [3, 4], among others. One of the promising approaches to break through the diffraction limit and form a super-resolution spot is taking advantage of nonlinear effects.

For the nonlinear effects, nonlinear refraction is well-known phenomena in most optical materials. Light beam with Gaussian intensity profile is incident on nonlinear materials and causes the refractive index to present Gaussian distribution. The Gaussian distribution of refractive index can restrict light into nanoscale spot by self-focusing behavior or interference manipulation [5, 6]. The nanoscale spot can be directly coupled into the samples in near-field range to conduct high-density data storage, high-resolving light imaging, and nanolithography. In this chapter, the models of nanoscale spot formation, such as interference manipulation model and multilayer thin lens self-focusing model, are presented based on different conditions, and some application schematics are given, accordingly.

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_5

5.2 Interference Manipulation-Induced Nanoscale Spot

In nonlinear super-resolution effect of nanofilm sample with nonlinear refraction and without absorption, some approximations can be used, the first is that the incident light is thought to be a collimated beam and does not diverge or broaden in propagation inside the sample. The second is that the refractive index n remains unchanged along sample thickness direction. According to the approximations, the irradiation of Gaussian beam spot with initial intensity distribution is

$$I_{inc}(r) = I_0 \exp\left(-2r^2/w_0^2\right), \quad \text{with} \quad I_0 = 2P/\pi w_0^2 \tag{5.1}$$

where P is incident laser power, and w_0 is spot radius at $1/e^2$ central intensity. The refractive index is marked as $n(r)$,

$$n(r) = n_0 + \gamma I_{\text{inc}}(r) \tag{5.2}$$

where n_0 and γ are the linear refractive index and nonlinear refraction coefficient, respectively. The self-action effects (such as self-focusing effect or self-defocusing effect) can be neglected; the contribution to super resolution is mainly from interference manipulation when the thin film thickness is smaller than incident light wavelength. The interference manipulation can be approximately dealt with nonlinear Fabry–Perot cavity structure.

5.2.1 Nonlinear Fabry–Perot Cavity Structure Model

The approach to deal with the interference effect of nonlinear nanofilms is that the nonlinear nanofilms can be considered as nonlinear Fabry–Perot cavity structure. It is well known that nonlinear Fabry–Perot cavity containing a nonlinear material can become bistable and multistable at a sufficiently high incident intensity. Under suitable conditions, the light transmitted from nonlinear absorptive or dispersive medium in an optical cavity can exhibit a hysteresis cycle. For this reason, the properties have attracted a great deal of attention. For example, the bistable mode can be used as a switching element in an optical circuit. Here, the nonlinear Fabry–Perot cavity structure can considered into an optical super-resolution device because the interference manipulation can induce nano-optical hot spot formation.

5.2.1.1 The Theoretical Derivation

The designed nonlinear Fabry–Perot cavity structure is shown in Fig. 5.1, where the region 2 is filled with a nonlinear thin film material, and the nonlinear thin film material is sandwiched by dielectric mirrors situated in region 1 and region 3. Figure 5.1 gives the schematic of light beam traveling through the nonlinear

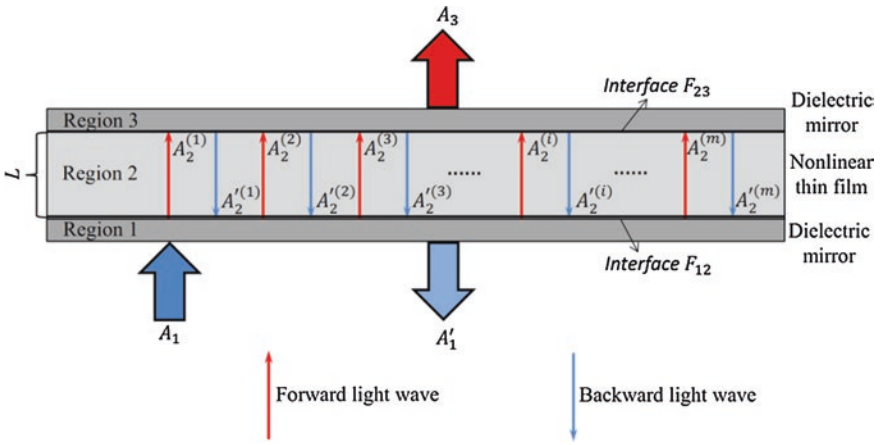

Fig. 5.1 The schematic of light beam traveling through the nonlinear Fabry–Perot cavity [7]

Fabry–Perot cavity, where the self-action and diffraction effects of light beam are ignorable due to the thickness of nonlinear Fabry–Perot cavity structure is only less than 1 μm. The nonlinear Fabry–Perot cavity is made up of three regions, where region 1 and region 3 are filled with transparent dielectric materials and form two dielectric mirrors. The nonlinear thin film material is filled into region 2, which is sandwiched by region 1 and region 3. The interface between region 1 and region 2 is marked as F_{12}, and the interface between region 2 and region 3 is marked as F_{23}.

Assuming that a light beam with an electric field intensity of A_1 is normally incident into the nonlinear Fabry–Perot cavity structure from region 1, the A_1 is written as

$$\mathrm{A}_1 = E_0 \exp\left(-r^2/w_0^2\right) \tag{5.3}$$

where w_0 is spot radius, and E_0 is electric field intensity at the center of spot. The light intensity can be accordingly expressed by formula (5.1).

The amplitude transmission and reflection at the interface F_{12} are assumed to be τ_{12} and ρ_{12}, and at the interface F_{23} are assumed to be τ_{23} and ρ_{23}, respectively. For normally incident TE light beam [8],

$$\begin{cases} \rho_{12} = \frac{n_1 - \tilde{n}}{n_1 + \tilde{n}} \\ \rho_{23} = \frac{\tilde{n} - n_3}{\tilde{n} + n_3} \\ \tau_{12} = \frac{2n_1}{n_1 + \tilde{n}} \\ \tau_{23} = \frac{2\tilde{n}}{\tilde{n} + n_3} \end{cases} \tag{5.4}$$

whereas for TM light beam,

$$\begin{cases} \rho_{12} = \frac{(1/n_1)-(1/\tilde{n})}{(1/n_1)+(1/\tilde{n})} \\ \rho_{23} = \frac{(1/\tilde{n})-(1/n_3)}{(1/\tilde{n})+(1/n_3)} \\ \tau_{12} = \frac{(2/n_1)}{(1/n_1)+(1/\tilde{n})} \\ \tau_{23} = \frac{(2/\tilde{n})}{(1/\tilde{n})+(1/n_3)} \end{cases} \tag{5.5}$$

The n_1 and n_3 are the refractive index of region 1 and region 3, respectively. In general, the dielectric mirror of region 1 is identical with that of region 3, and $n_1 = n_3$. $\tilde{n}$ is complex refractive index of nonlinear thin film in region 2.

$$\tilde{n} = n + i\frac{\lambda}{4\pi}\alpha \tag{5.6}$$

where n and α are refractive index and absorption coefficient of region 2, respectively.

The reflectance R and transmittance T can be simply written as

$$\begin{cases} R = R_{12} = |\rho_{12}|^2 = R_{23} = |\rho_{23}|^2 = |\rho|^2 \\ T = T_{12} = \Re\left[\frac{\tilde{n}}{n_1}\right]|\tau_{12}|^2 = T_{23} = \Re\left[\frac{n_3}{\tilde{n}}\right]|\tau_{23}|^2 \end{cases} \tag{5.7}$$

with

$$R + T = 1$$

The $\Re[\cdot]$ denotes the operation of taking the real part.

The A_1 passes through the interface F_{12} and travels to the interface F_{23}. The light field becomes $A_2^{(1)}$, which is calculated as

$$A_2^{(1)} = A_1 \tau e^{ikL - \frac{1}{2}\alpha L} \tag{5.8}$$

where $k = 2\pi n/\lambda$ is propagation constant, which is also called as wavenumber inside the sample. The $e^{[ikL-(\alpha L/2)]}$ means the variations of phase and intensity of light field in the traveling from F_{12} to F_{23}. It is noted that for $e^{[ikL-(\alpha L/2)]}$, the mean-field approximation is implicitly made, in other words, n and α are assumed to be unchanged along the thin film thickness direction position z. If such is not the case, the $\left[\left(i\frac{2\pi}{\lambda}n - \frac{1}{2}\alpha\right)L\right]$ should be replaced by $\int_0^L \left[i\frac{\pi}{\lambda}n(z) - \frac{1}{2}\alpha(z)\right]dz$ [9].

The $A_2^{(1)}$ is partially reflected to interface F_{12} after traveling a distance of L, and light field becomes $A_2^{\prime(1)}$

$$A_2^{\prime(1)} = A_1 \tau e^{ikL - \frac{1}{2}\alpha L} \rho e^{ikL - \frac{1}{2}\alpha L} \tag{5.9}$$

The $A_2^{\prime(1)}$ is also partially reflected to the interface F_{23} after traveling a distance of L, and the light field becomes $A_2^{(2)}$

$$A_2^{(2)} = A_1 \tau e^{ikL - \frac{1}{2}\alpha L} R e^{2ikL - \alpha L} \tag{5.10}$$

The $A_2^{(2)}$ is partially reflected to interface F_{12} after traveling a distance of L, and the light field becomes $A_2^{\prime(2)}$

$$A_2^{\prime(2)} = A_1 \tau e^{ikL-\frac{1}{2}\alpha L} \rho \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right) \tag{5.11}$$

Similar to $A_2^{(2)}$, the 3rd forward light wave $A_2^{(3)}$ should be written as

$$A_2^{(3)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^2 \tag{5.12}$$

The 3rd backward light wave $A_2^{\prime(3)}$ should be written as

$$A_2^{\prime(3)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \rho \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^2 \tag{5.13}$$

The ith forward light wave $A_2^{(i)}$ should be written as

$$A_2^{(i)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^{(i-1)} \tag{5.14}$$

The ith backward light wave $A_2^{\prime(i)}$ should be written as

$$A_2^{\prime(i)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \rho \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^{(i-1)} \tag{5.15}$$

After coming and going m times (m is integer), the mth forward light wave $A_2^{(m)}$ should be written as

$$A_2^{(m)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^{(m-1)} \tag{5.16}$$

The mth backward light wave $A_2^{\prime(m)}$ should be written as

$$A_2^{\prime(m)} = A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \rho \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \left(R \mathrm{e}^{2ikL-\alpha L} \right)^{(m-1)} \tag{5.17}$$

After the light beam is reflected for m times between interface F_{12} and interface F_{23}, a stable effective field can be obtained. The effective field consists of forward light wave and backward light wave. The forward light wave is marked as A_2 and the backward light wave is mark as $A_2^{\prime}$. Figure 5.2 presents the schematic, where A_3 and $A_1^{\prime}$ are exiting and reflected light fields from the nonlinear Fabry–Perot cavity structure, respectively. Here, one only considers the exiting light field A_3.

The forward light field A_2 is a superposition and can be calculated as

$$A_2 = A_2^{(1)} + A_2^{(2)} + A_2^{(3)} + \cdots + A_2^{(i)} + \cdots + A_2^{(m)} = \frac{A_1 \tau \mathrm{e}^{ikL-\frac{1}{2}\alpha L}}{1 - R \mathrm{e}^{2ikL-\alpha L}} \tag{5.18}$$

The backward light field $A_2^{\prime}$ is also a superposition and can be calculated as

$$A_2^{\prime} = A_2^{\prime(1)} + A_2^{\prime(2)} + A_2^{\prime(3)} + \cdots + A_2^{\prime(i)} + \cdots + A_2^{\prime(m)} = A_2 \rho \mathrm{e}^{ikL-\frac{1}{2}\alpha L} \tag{5.19}$$

Formula (5.18) describes the properties of the nonlinear Fabry–Perot cavity structure. The intensity of forward light wave A_2 is

$$I_2 = |A_2|^2 = \frac{TI_1 e^{-\alpha L}}{1 - 2Re^{-\alpha L}\cos(4\pi nL/\lambda) + R^2 e^{-2\alpha L}} \tag{5.20}$$

As is shown in Fig. 5.2, the exiting light field from the nonlinear Fabry–Perot cavity structure is A_3, which can be calculated as $A_3 = \tau A_2$. The intensity is accordingly

$$I_3 = |A_3|^2 = TI_2 = \frac{T^2 I_1 e^{-\alpha L}}{1 - 2Re^{-\alpha L}\cos(4\pi nL/\lambda) + R^2 e^{-2\alpha L}} \tag{5.21}$$

In formula (5.21), the n and α are induced by the internal cavity field inside region 2, which is named as A_{cavity}. The A_{cavity} consists of two parts, one is the forward light wave A_2, and the other is the backward light wave A_2', thus $A_{\text{cavity}} = A_2 + A_2'$. For simplicity, one can ignore the standing wave within the cavity structure; the light intensity I_{cavity} in region 2 is

$$I_{\text{cavity}} = |A_2|^2 + \left|A_2'\right|^2 = I_2 + I_2' \tag{5.22}$$

According to formula (5.19), the I_2' can be calculated as

$$I_2' = \left|A_2'\right|^2 = \left|A_2 \rho e^{ikL - \frac{1}{2}\alpha L}\right|^2 = I_2 R e^{-\alpha L} \tag{5.23}$$

Hence

$$I_{\text{cavity}} = I_2(1 + Re^{-\alpha L}) \tag{5.24}$$

If the n and α remain unchanged along the film thickness direction z and are only function of radial coordinate r, which is reasonable because the self-action and diffraction effects are ignored. The n and α are expressed as $n(r)$ and $\alpha(r)$,

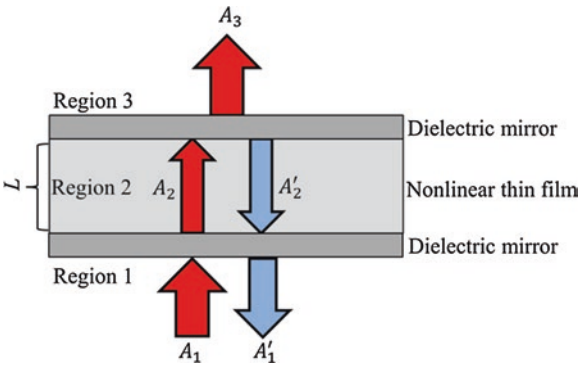

Fig. 5.2 Simplified schematic of light beam propagation using the effective light field [7]

respectively. The internal cavity field-induced refractive index and absorption coefficient are, respectively

$$\begin{cases} n(r) = n_0 + \gamma I_{\text{cavity}}(r) \\ \alpha(r) = \alpha_0 + \beta I_{\text{cavity}}(r) \end{cases} \tag{5.25}$$

where α_0 is linear absorption coefficient. β is nonlinear absorption coefficient. The internal cavity field-induced complex refractive index is written as

$$\tilde{n}(r) = \left[n_0 + \gamma I_{\text{cavity}}(r)\right] + i\frac{\lambda}{4\pi}\left[\alpha_0 + \beta I_{\text{cavity}}(r)\right] \tag{5.26}$$

One can obtain the intensity distribution of the exiting spot from the nonlinear Fabry–Perot cavity structure with formula (5.21). Here, it should be pointed out that, for the thin film samples with both nonlinear absorption and nonlinear refraction, the super-resolution effect from the contribution of nonlinear absorption is stronger than that from the contribution of nonlinear refraction because the nonlinear absorption easily generates an optical pinhole channel [10, 11]. In addition, the absorption also causes interference effect to become weak inside the nonlinear Fabry–Perot cavity structure. Hence, in the following, let us consider a special case for nonlinear thin film without absorption, i.e., $\alpha = 0$ in order to analyze the interference manipulation effect inside the nonlinear Fabry–Perot cavity structure. The thickness of dielectric mirror is assumed to be less than 100 nm. The light intensity distribution in region 2 can be directly mapped onto the exiting surface of the nonlinear Fabry–Perot cavity structure because the thickness of dielectric mirror is in the optical near-field range. Based on formula (5.1), the formula (5.21) can be rewritten as

$$I_3(r) = |A_3|^2 = \frac{T^2 I_0 \exp\left(-\frac{2r^2}{w_0^2}\right)}{1 - 2R\cos\left[\frac{4\pi n(r)L}{\lambda}\right] + R^2} \tag{5.27}$$

The formula (5.27) is actually implicit equation and cannot be directly obtained. However, one can have the exiting spot intensity distribution through numerical calculations with initial values of $n = n_0$ for nonlinear thin film. When the calculated I_{cavity} is stable, the calculation is finished.

5.2.1.2 Nanoscale Optical Hot Spot Formation by Constructive Interference Manipulation

The As_2S_3 thin film filled into the region 2 is taken as an example due to giant positive nonlinear refraction coefficient of $\gamma = 8.65 \times 10^{-10}$ m^2/W and ignorable linear and nonlinear absorption coefficient $\alpha_0 \approx \beta \approx 0$ at the light wavelength of $\lambda = 633$ nm [12]. $n_0 \approx 2.5$ is chosen because of large linear refraction for

chalcogenide glass materials. The incident laser power is fixed at $P = 0.2$ mW, and $w_0 = 300$ nm, the incident light intensity can be obtained by $I_0 = 2P/(\pi w_0^2)$.

Based on formula (5.27), in order to obtain nanoscale spot, $I_3(r)$ should be maximum at $r = 0$, which requires $4\pi nL/\lambda = 2q\pi$ with $q = 0, 1, 2, \ldots$. After the incident laser power is fixed, one can change the nonlinear thin film thickness to obtain strongest interference field at $r = 0$ [13]. Thus

$$L = q\frac{\lambda}{2n},\ q = 0, 1, 2, \ldots \tag{5.28}$$

The rough internal cavity field intensity is estimated by assuming the $L = 100$ nm and shown in Fig. 5.3a. One can find that $I_{\text{cavity}} = 8.55 \times 10^8$ W/m^2 at $r = 0$. The internal cavity field-induced refractive index is plotted in Fig. 5.3b, where the $n \approx 3.24$ at $r = 0$.

According to formula (5.24) the I_{cavity} has a little fluctuation at $4\pi nL/\lambda = 2q\pi$ with the nonlinear thin film thickness. In order to get the interference enhancement at $r = 0$, by substituting $n \approx 3.24$ into formula (5.28) and optimizing the calculation, one can obtain a series of L values as follows.

$$L = 286\text{ nm}, 386\text{ nm}, 484\text{ nm}, 586\text{ nm}, 687\text{ nm}, \ldots$$

Figure 5.4 presents normalized two-dimensional spot intensity distribution. Figure 5.4a is incident spot itself, which is a typical Gaussian profile. Figure 5.4b is the exiting spot at $L = 286$ nm; one can find that the spot is obviously smaller than the incident spot. When the thickness is increased to $L = 386$ nm, the spot is further smaller than the incident spot. At $L = 484$ nm, a central spot with a size of about 100 nm occurs, as shown in Fig. 5.4d. Increasing $L = 586$ nm, a spot of about 70 nm appears in the central region (also see Fig. 5.4e). A very small spot with a size of 40 nm occurs in the central region when $L = 687$ nm; as is shown in Fig. 5.4f, the spot size is only about $\lambda/16$, which is very useful in nanolithography and high-resolving light imaging.

In order to further analyze the spot characteristic, Fig. 5.5 gives the normalized three-dimensional spot intensity distribution. Figure 5.5a is the incident spot.

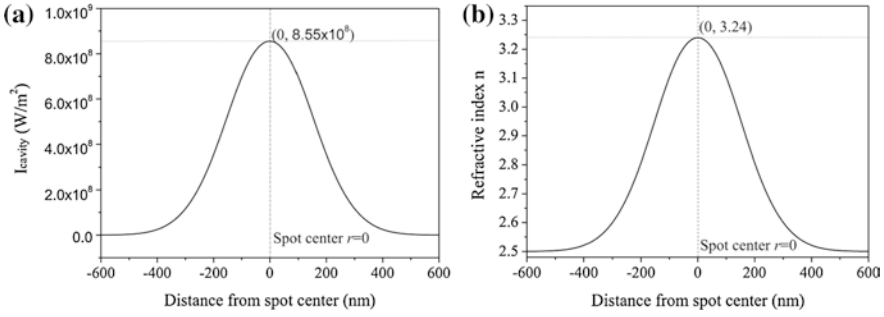

Fig. 5.3 The distribution of **a** cavity field intensity and **b** refractive index at incident laser power of 0.2 mW and $L = 100$ nm [7]

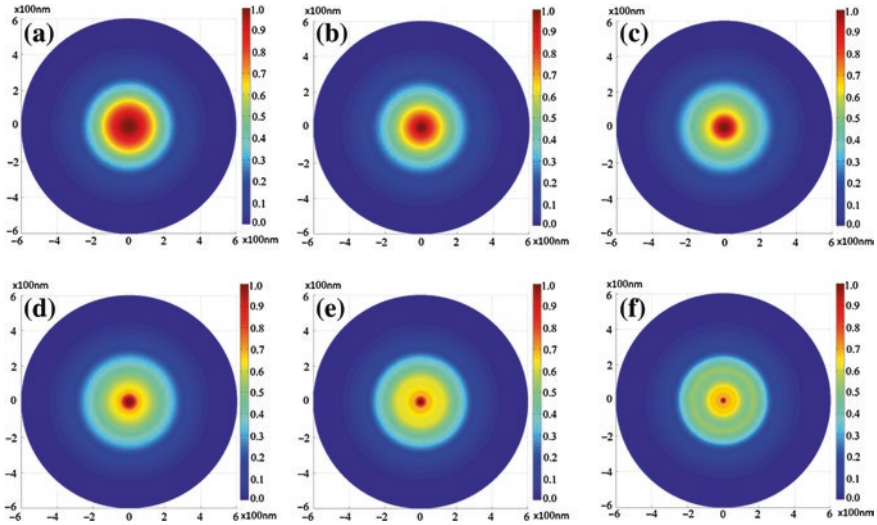

Fig. 5.4 The two-dimensional spot intensity distribution **a** incident spot, **b** $L = 286$ nm, **c** $L = 386$ nm, **d** $L = 484$ nm, **e** $L = 586$ nm, and **f** $L = 687$ nm [7]

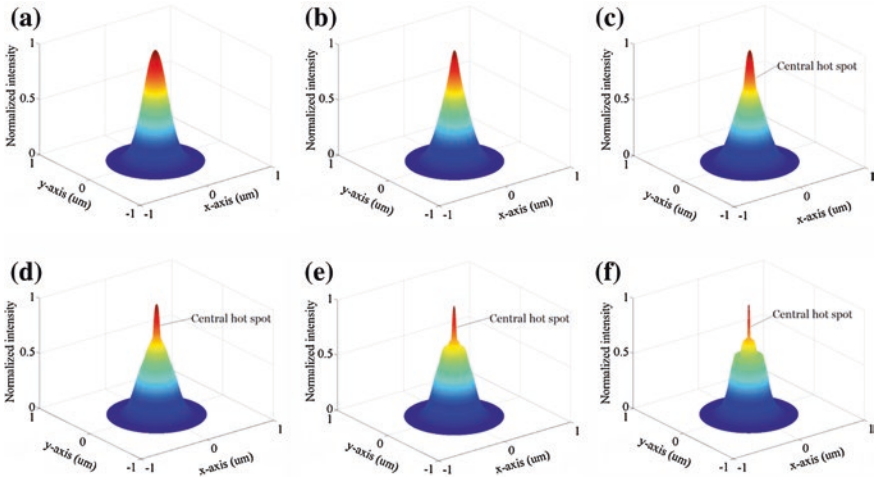

Fig. 5.5 The three-dimensional spot intensity distribution **a** incident spot, **b** $L = 286$ nm, **c** $L = 386$ nm, **d** $L = 484$ nm, **e** $L = 586$ nm, and **f** $L = 687$ nm [7]

Figure 5.5b is exiting spot at $L = 286$ nm; compared with Fig. 5.5a, one can find that the spot becomes sharp. Figure 5.5c is the exiting spot at $L = 386$ nm; the spot becomes shaper and a central hot spot occurs at the position of normalized intensity of about 2/3. Figure 5.5d shows that at $L = 484$ nm, the central hot spot

becomes sharper and the size is also smaller than those of Fig. 5.5c. Figure 5.5e shows that with the nonlinear thin film thickness increasing to $L = 586$ nm, the central hot spot size obviously decreases compared with Fig. 5.5d. The central hot spot size is further reduced at $L = 687$ nm, as is given in Fig. 5.5f, and there are two inflection points at the exiting spot intensity distribution. One is at about 1/2 of normalized intensity, and the other is at about 2/3 of normalized intensity. At the position of about 2/3 normalized intensity, a very sharp central hot spot occurs, and the central hot spot looks like a tip of optical probe. The central hot spot size is far smaller than the incident spot itself.

Figure 5.6a shows the cross-section profile of the normalized spot intensity along the radial direction for different nonlinear thin film thickness. The increase of nonlinear thin film thickness can sharpen the exiting spot, and the central optical hot spot occurs at $L \geq 386$ nm. The central hot spot size is reduced to about 40 nm at $L = 687$ nm, which is a nanoscale optical hot spot applicable in nano-optical data storage, high-resolving imaging, and direct laser writing lithography.

In spot simulation, one can find that the central hot spot size is very sensitive to nonlinear thin film thickness. In order to understand the sensitivity, the $L = 687$ nm is taken as an example, and the dependence of central hot spot size on nonlinear thin film thickness at $L = 687$–690 nm is analyzed.

Figure 5.6b presents the calculated results, the central hot spot size linearly increases with nonlinear thin film thickness for $L = 687$–690 nm, which indicates that the hot spot size is sensitive to nonlinear thin film thickness, and the accuracy is required to be up to nanometer or even subnanometer scale, which is very large challenging for thin film deposition method and process. Fortunately, some advanced thin film deposition technique can prepare the thin film materials with nanoscale (or even subnanometer scale) accuracy. For example, the atomic layer deposition technique, also called as atomic layer epitaxy, is a good method to produce high-quality large-area thin film materials with perfect structure and process controllability, and the accuracy can reach up to single atom layer accordingly [14].

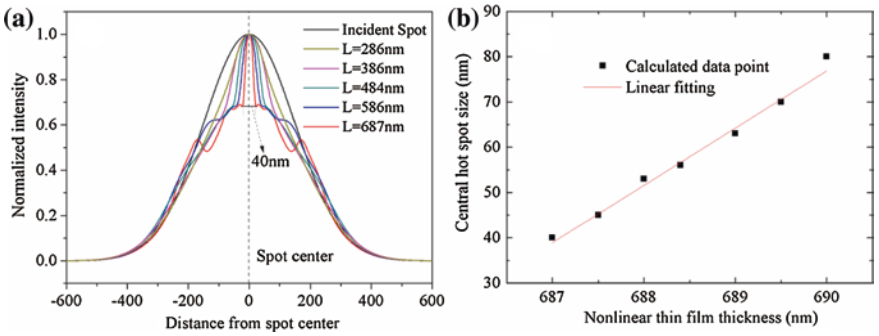

Fig. 5.6 The dependence of spot characteristics on nonlinear thin film thickness, **a** normalized intensity distribution of exiting light spot from nonlinear Fabry–Perot cavity structure for different nonlinear thin film thickness, and **b** sensitivity of central spot size to nonlinear thin film thickness for $L = 687 \sim 690$ nm and $P = 0.2$ mW [7]

The advanced thin film deposition technique, such as atomic layer epitaxy, is indeed good method for obtaining high-accuracy nonlinear thin films. Actually, there is also an alternative method for realizing nanoscale central optical hot spot, that is, slightly changing the incident laser power can obtain the same performance as the high-accuracy manipulation of nonlinear thin film thickness, which is because the super-resolution spot is also sensitive to the incident laser intensity. The Fig. 5.4f with $L = 687$ nm and $P = 0.2$ mW is taken as an example. The central hot spot with a size of about 40 nm can be obtained. However, the central hot spot immediately becomes large, and the spot top becomes flat when the nonlinear thin film thickness becomes $L = 690$ nm, as is seen in the two-dimensional intensity distribution of Fig. 5.7a. Compared with Fig. 5.4f, the central hot spot obviously becomes large although the thickness difference is only 3 nm. The cross-section profile curve is shown with red curve in Fig. 5.7b, and the central hot spot becomes about 80 *nm*, accordingly. In order to obtain the same performance as Fig. 5.4f, one can slightly tune the incident laser power from

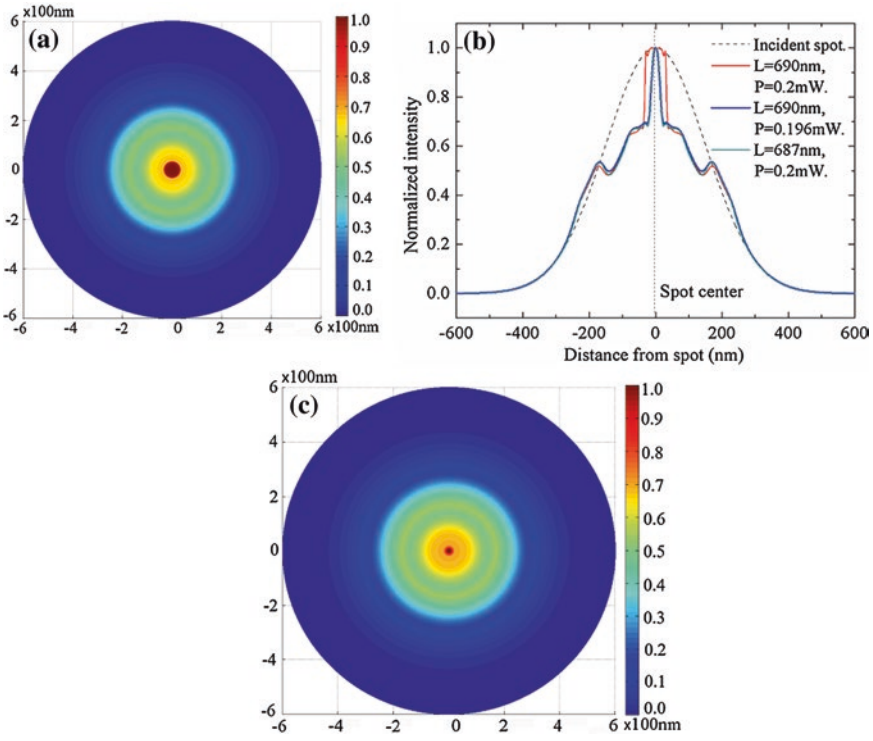

Fig. 5.7 Laser power manipulation for obtaining nanoscale central optical hot spot, **a** the two-dimensional intensity distribution for $L = 690$ nm and $P = 0.2$ mW, **b** comparison of spot intensity among $L = 890$ nm and $P = 0.2$ mW, $L = 690$ nm and $P = 0.196$ mW, and $L = 687$ nm and $P = 0.2$ mW, **c** the two-dimensional intensity distribution for $L = 690$ nm and $P = 0.196$ mW [7]

$P = 0.2$ mW to $P = 0.196$ mW. Figure 5.7c gives the two-dimensional intensity distribution accordingly. Compared with Fig. 5.7a, the central hot spot obviously becomes small by tuning incident laser power of $P = 0.196$ mW, and the cross-section profile curve is shown with blue curve in Fig. 5.7b, and the central hot spot becomes about 40 nm, accordingly. Figure 5.7b also gives the intensity profile of Fig. 5.4f with dark cyan curve (with $L = 687$ nm and $P = 0.2$ mW), where the dark cyan curve is almost coincident with the blue curve (with $L = 690$ nm and $P = 0.196$ mW). Therefore, Fig. 5.7 indicates that the slightly changing the incident laser power is also an alternative method for compensating for drawback of low thickness accuracy of nonlinear thin films.

The nanoscale optical hot spot can be reversibly generated because the nonlinear characteristics of thin film in region 2 are reversible. For practical applications, such as nanolithography and optical surface imaging, generation and movement of the nanoscale central hot spot are possible by scanning the incident light beam through a galvanometer mirror or by moving the nonlinear Fabry–Perot cavity structure. The resist thin films or the samples to be imaged are directly deposited on the surface of the nonlinear Fabry–Perot cavity structure. The nanoscale central hot spot is directly coupled into the resists or samples to be imaged because the thickness of the dielectric mirror is less than 100 nm, which is in the near-field range.

However, the central hot spot size cannot decrease infinitely with increasing L. That is, the q value of formula (5.28) cannot increase infinitely. For example, further increasing the q value such that L = 781 nm at P = 0.2 mW. Figure 5.8 illustrates the optimized exiting spot characteristics. Figure 5.8a shows the radial distribution of the normalized spot intensity. The central hot spot becomes coarse and the size becomes large compared with the case at L = 687 nm and $P = 0.2$ mW. Figure 5.8b shows the normalized two-dimensional intensity image. Compared with Fig. 5.4f where L = 687 nm and $P = 0.2$ mW, the spot morphology is no longer a single central hot spot, but is an annular spot, and the annular

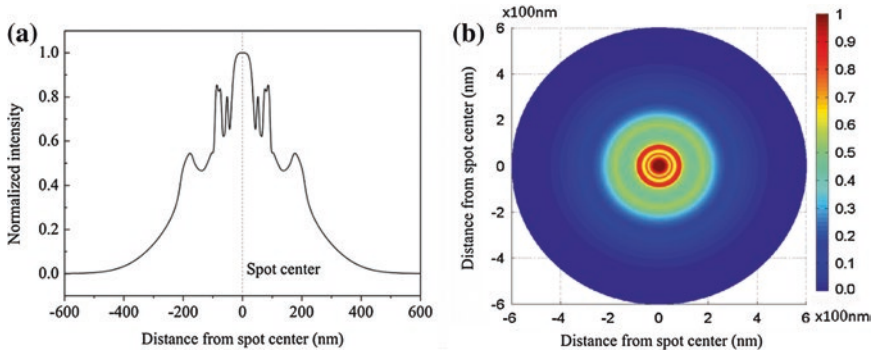

Fig. 5.8 The normalized spot intensity distribution for $L = 781$ nm and $P = 0.2$ mW, **a** cross-section profile along radial direction, and **b** two-dimensional spot intensity profile [7]

central spot size is about 80 nm. Such size is obviously larger than the central hot spot at $L = 687$ nm. Figure 5.8 indicates, to obtain a nanoscale central hot spot by interference manipulation, an optimum nonlinear thin film thickness for the nonlinear Fabry–Perot cavity structure needs to be adjusted. In the example, the optimum thickness is $L = 687$ nm for $P = 0.2$ mW, and the nano-optical hot spot size is 40 nm.

Here, one needs to notice that, based on the theoretical model, different laser power can induce different refractive index profile; thus, the super-resolution performance is also different for different laser power irradiation. That is, the super-resolution effect can be regulated by changing the laser power.

It needs to be pointed out that in the theoretical model, some reasonable approximations are used, such as mean-field approximation or ignoring the standing wave effect. In addition, Ref. [15] also gave some calculation and experimental results for reducing the spot through the nonlinear Fabry–Perot resonators.

5.2.1.3 Test and Application Schematic Designs of Nonlinear Fabry–Perot Cavity Structures

According to the simulation analysis above, the super-resolution spot can be obtained through the constructive interference manipulation in the nonlinear Fabry–Perot cavity structures when the As_2S_3 thin film is chosen as the nonlinear thin film. Figure 5.9 presents a test schematic design for the super-resolution spot. The nonlinear thin film is sandwiched in between the dielectric mirrors, where the thickness of dielectric mirrors is less than 100 nm. A collimated laser beam passes through the neutral density optical filter and is focused by a converging lens. The laser power can be finely regulated by the neutral density optical filter. The focused spot is normally incident into the nonlinear Fabry–Perot cavity structure and induces the formation of super-resolution spot. The super-resolution spot only exists at the surface of nonlinear Fabry–Perot cavity structure and is difficult to propagate into the air. Therefore, in order to test the super-resolution spot, a near-field spot scanning method needs to be designed. The super-resolution spot can be scanned and tested by near-field probe scanning head, where the distance regulation between the fiber tip of scanning head and the sample surface is controlled by utilizing a tuning fork, and the regulated distance should be maintained at ~50 nm, which is a typical near-field range [16].

The application schematics can be roughly designed as follows. The resists and optical recording materials with obvious threshold effect (such as structural change and molten ablation) are directly deposited onto the nonlinear Fabry–Perot cavity structure, as shown in Fig. 5.10a. The thickness of dielectric mirrors in the nonlinear Fabry–Perot cavity structure is less than 100 nm. A collimated laser beam is focused and normally incident onto the nonlinear Fabry–Perot cavity structure, and the super-resolution spot can be generated and directly coupled into the resists or optical recording media in the near-field range. The near-field distance can be regulated by the thickness of dielectric mirror. The laser power can be

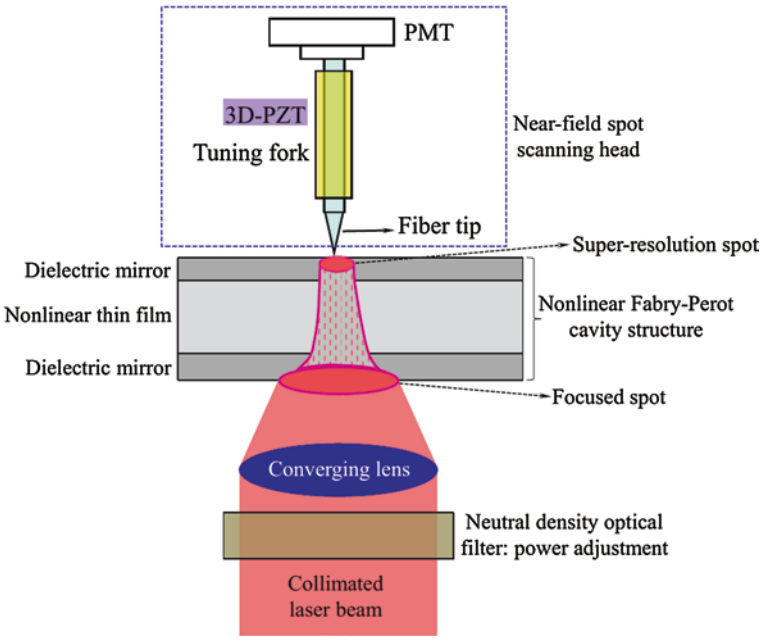

Fig. 5.9 Super-resolution spot test schematic of nonlinear Fabry–Perot cavity structures [7]

finely adjusted by the neutral density optical filter. When the spot intensity of 70 % central maximum induces the threshold effect of resists or recording media, the nanoscale lithography or optical recording can take place, as shown in Fig. 5.10b.

Here, it should be explained that in the example, the photo-induced or thermally induced expansion effect is very weak due to the low laser power of milliwatt magnitude, which is enough for laser direct writing lithography, optical

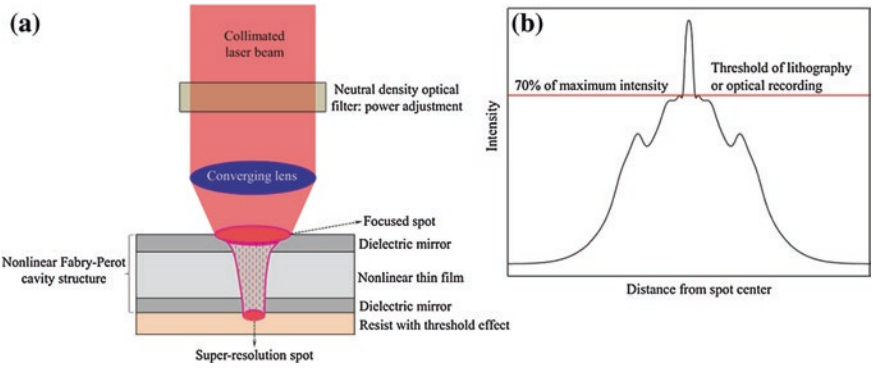

Fig. 5.10 Application schematics of the nonlinear Fabry–Perot cavity structure in nanolithography, **a** formation schematic of super-resolution spot, **b** the threshold effect for lithography or optical recording [7]

recording, and imaging. If there is thermal expansion, which induces the thickness change of nonlinear thin films, one can regulate the laser power a little by the neutral density optical filter to eliminate the expansion influence on super-resolution effect.

When the resists or recording media are deposited onto the nonlinear Fabry–Perot cavity structures, the interface loss may occur and has a little influence on the super-resolution spot performance. The influence can be reduced or eliminated by slightly adjusting the neutral density optical filter to change the laser power a little.

Therefore, one can manipulate the nonlinear thin film thickness of the nonlinear Fabry–Perot cavity structure to make the position at $\mathrm{r} = 0$ become constructive interference. The nanoscale optical hot spot can be produced in the central region. The hot spot size is sensitive to nonlinear thin film thickness, and the accuracy is required to be up to nanometer or even subnanometer scale, which is very large challenging for thin film deposition technique; however, slightly changing the incident laser power can compensate for drawbacks of low thickness accuracy of nonlinear thin films. Taking As_2S_3 as the nonlinear thin film, the central hot spot with a size of 40 nm is obtained at suitable nonlinear thin film thickness and incident laser power. The central hot spot size is only about $\lambda/16$, which is very useful in superhigh-density optical recording, nanolithography, and high-resolving optical surface imaging.

5.3 Self-focusing Effect-Induced Nanoscale Spot Through "Thick" Samples

For nonlinear "thick" samples, the internal multiple interference effect among interfaces can be neglected. The refractive index change along sample thickness direction needs to be considered, that is, the n is not only function of radial coordinate r, but also vertical coordinate z. For nonlinear refraction samples with $\gamma > 0$, the irradiation of Gaussian intensity beam spot induces self-focusing, and the converging effect of light beam occurs inside nonlinear sample. In the process of converging, the light intensity increases along sample thickness direction z, and the refractive index should be marked as $n(r, z)$.

The positive nonlinear refraction can induce self-focusing effect and form small spot; however, the diffraction effect also needs to be considered when the light beam becomes very small. The diffraction effect causes the beam to broaden, while the wavefront curvature induced by nonlinear refraction tends to pull the beam toward central region with higher intensity. When the nonlinear effect is weak, the diffraction is dominant over self-focusing effect. With the increase of laser power, a point might be reached when the self-focusing becomes balance with the diffraction effect, and the beam can then travel over long distance without any noticeable expansion or contraction, that is, a self-focusing needle produces inside the nonlinear samples.

The self-focusing needle can be physically considered to be an effective waveguide for itself and propagates without spreading. One can also take advantage of the self-focusing needle to obtain nanoscale spot at the bottom of nonlinear medium by manipulating experimental conditions and some parameters. The nanoscale self-focusing spot can be directly coupled into the samples, such as optical recording thin films and resists, in near-field range, and realize the ultra-high-density data storage and nanolithography, etc.

The formation process of nanoscale spot inside nonlinear "thick" film samples can be obtained by solving the nonlinear Schrodinger equation; however, it is difficult to directly obtain analytical solution. Some numerical methods are usually used, such as finite difference method and pseudo-spectrum analyzing method. In real applications, the numerical methods are required to conduct complex mathematical operations.

The self-focusing spot can be also roughly estimated by considering the nonlinear medium as converging lens, as shown in Fig. 5.11. In Fig. 5.11a, the light beam irradiates into the nonlinear sample and forms a self-focusing spot. In practice, one can generally consider that the refractive index only changes along the radial direction r and remains unchanged in the thickness z position. The nonlinear sample is usually considered as an equivalent lens, as shown in Fig. 5.11b, and the focal spot size and focal distance are calculated with simple geometrical optical formulas, accordingly.

However, when the sample is thick, the light beam converging effect occurs inside nonlinear sample. In the process of converging, the light intensity increases along sample thickness direction, and the refractive index should be marked as $n(r, z)$. In addition, the beam broadening resulting from diffraction effect needs to be considered, that is, in the propagation, the light beam might diverge with increasing propagation length. Here, a multilayer thin lens self-focusing model is presented to understand the formation and propagation characteristics of self-focusing beam spot inside nonlinear "thick" samples. The As_2S_3 is still taken as an example; the nanoscale self-focusing spot is analyzed as follows.

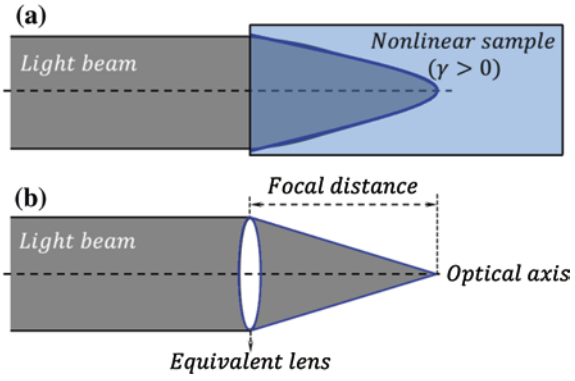

Fig. 5.11 Schematics of equivalent lens, **a** light beam is incident into nonlinear sample and converging into a point, **b** the nonlinear sample is equivalent to converging lens

5.3.1 Multilayer Thin Lens Self-focusing Model [17]

A collimated light beam with Gaussian intensity profile is focused onto the "thick" sample with positive nonlinear refraction coefficient ($\gamma > 0$) and causes the refractive index n to be maximum at spot center and decrease with the radial coordinate r increasing. The traveling speed of light at central region of the spot ($r = 0$) is the slowest and then increases along radial coordinate r, and the traveling speed at the rim of the spot is the fastest. Thus, a self-focusing lens is formed inside the nonlinear "thick" samples. The nonuniform profile of traveling speed causes the equiphase surface of wavefront to be curved surface, and the curved surface changes with the sample thickness position z due to different refractive index profile, as shown in Fig. 5.12. The propagation direction of light ray is perpendicular to equiphase surface, which causes the light ray to bend in the process of traveling inside nonlinear sample, and the spot becomes smaller and smaller; thus, a nanoscale self-focusing spot is formed at the bottom side of nonlinear sample.

To understand the beam propagation inside the nonlinear "thick" film, the multilayer (M-L) "thin" lens self-focusing model is proposed. The nonlinear "thick" sample is divided into m layers, and the thickness of each layer is ΔL. $m \times \Delta L = L$, where L is thickness of nonlinear sample. Every layer can be considered as "thin" lens when m is large enough. Inside every layer, n is uniform along thickness direction and is only a function of radial coordinate r. The focal length of self-focusing thin lens is marked as f_i, with $i = 1, 2, 3, \ldots, m$. Figure 5.13 gives the schematic of multilayer thin lens self-focusing model for positive nonlinear refraction sample, where Fig. 5.13a is Gaussian intensity profile of incident light beam spot, and Fig. 5.13b is the light beam self-focusing process.

A collimated light beam is focused into a diffraction-limited spot by a lens and incident to the nonlinear sample surface. The phase change of wavefront inside every thin layer is neglected because the thickness of thin layer ΔL is much less

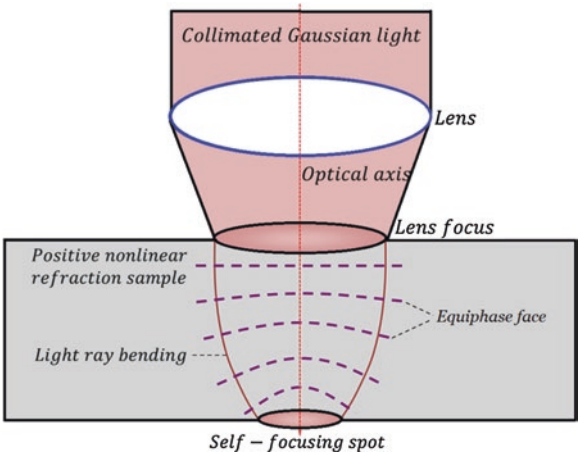

Fig. 5.12 The schematic of beam propagation inside positive nonlinear refraction sample

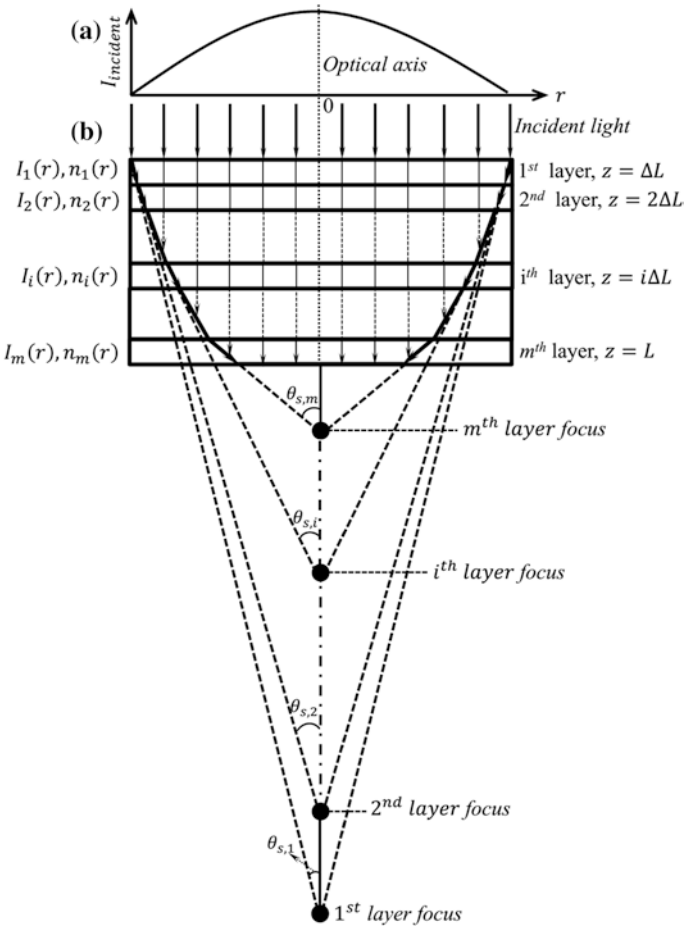

Fig. 5.13 Schematic of multilayer thin film model for positive nonlinear refraction sample, **a** intensity profile of incident Gaussian light beam spot, and **b** light beam self-focusing

than the Rayleigh length z_0. The light ray at the rim converges to center of self-focusing thin lens according to Fermat principle, that is, the light rays at the rim and the center have the same optical path. Thus, the light ray at the rim deflects toward optical axis at an angle of $\theta_{s,i}$ $(i = 1, 2 \ldots m)$, where the $\theta_{s,i}$ is called as self-focusing angle of the *i*th layer.

When the diffraction-limited spot goes into the 1st layer and induces redistribution of refractive index, the central refractive index of self-focusing thin lens is

$$n_{c,1} = n_0 + \gamma I_0 \tag{5.29}$$

At the rim of self-focusing thin lens, the refractive index is $n_{r,1} = n_0 + \gamma I_0/\mathrm{e}^2$. Based on Fermat principle, one has

$$n_{c,1} \times \Delta L = n_{r,1} \times \Delta L / \cos\theta_{s,1} \Rightarrow \theta_{s,1} = \cos^{-1}\left(n_{r,1}/n_{c,1}\right) \tag{5.30}$$

The spot radius after passing through the 1st layer is changed into

$$w_1 = w_0 - \Delta L \times \tan\theta_{s,1} \tag{5.31}$$

Considering the linear absorption effect α_0, which has no influence on beam radius variation, the central light intensity at the exiting surface of the first layer is

$$I_1 = I_0 \mathrm{e}^{-\alpha_0 \Delta L} w_0^2 / w_1^2 \tag{5.32}$$

At the 2nd layer, $n_{r,2} = n_0 + \gamma I_1 / e^2$. The central refractive index $n_{c,2}$, self-focusing angle $\theta_{s,2}$, spot radius w_2, and light intensity after passing through the 2nd layer are, respectively

$$\begin{aligned} n_{c,2} = n_0 + \gamma I_1,\ \theta_{s,2} = \cos^{-1}\left(n_{r,2}/n_{c,2}\right), w_2 = w_1 - \Delta L \tan\theta_{s,2}, \\ I_2 = I_1 \mathrm{e}^{-\alpha_0 \Delta L} w_1^2 / w_2^2 \end{aligned} \tag{5.33}$$

Similarly, at the ith layer, $n_{r,i} = n_0 + \gamma I_{i-1} / e^2$. The central refractive index $n_{c,i}$, self-focusing angle $\theta_{s,i}$, spot radius w_i, and light intensity I_i after passing through the ith layer are

$$\begin{aligned} n_{c,i} = & n_0 + \gamma I_{i-1},\ \theta_{s,i} = \cos^{-1}\left(n_{r,i}/n_{c,i}\right),\ w_i = w_{i-1} - \Delta L \tan\theta_{s,i}, \\ & I_i = I_{i-1} \mathrm{e}^{-\alpha_0 \Delta L} w_{i-1}^2 / w_i^2 \end{aligned} \tag{5.34}$$

At the last layer (the mth layer), $n_{r,m} = n_0 + \gamma I_{m-1} / \mathrm{e}^2$,

$$\begin{aligned} n_{c,m} = n_0 + \gamma I_{m-1},\ \theta_{s,m} = \cos^{-1}\left(n_{r,m}/n_{c,m}\right),\ w_m = w_{m-1} - \Delta L \tan\theta_{s,m}, I_m \\ = I_{m-1} \mathrm{e}^{-\alpha_0 \Delta L} w_{m-1}^2 / w_m^2 \end{aligned} \tag{5.35}$$

In the process of light traveling inside every layer of nonlinear sample, the self-focusing is always accompanied with diffraction effect. The diffraction angle $\theta_{d,i}$ is defined as

$$\theta_{d,i} = 0.61\lambda / (n_{r,i} w_{i-1}) \tag{5.36}$$

When the self-focusing angle $\theta_{s,i}$ is larger than diffraction angle $\theta_{d,i}$, $\theta_{s,i} > \theta_{d,i}$, the self-focusing effect is dominant over the diffraction effect; the light beam size and intensity are calculated by formula (5.34). In reverse, $\theta_{s,i} < \theta_{d,i}$, the diffraction effect is dominant over the self-focusing effect, and the beam radius and central intensity should be recalculated by

$$\begin{aligned} \theta_{d,i} = 0.61\lambda / (n_{r,i} w_{i-1}),\ w_i = w_{d,i} = w_{i-1} + \Delta L \ \tan\theta_{d,i}, \\ I_i = I_{i-1} \mathrm{e}^{-\alpha_0 \Delta L} w_{i-1}^2 / w_i^2 \end{aligned} \tag{5.37}$$

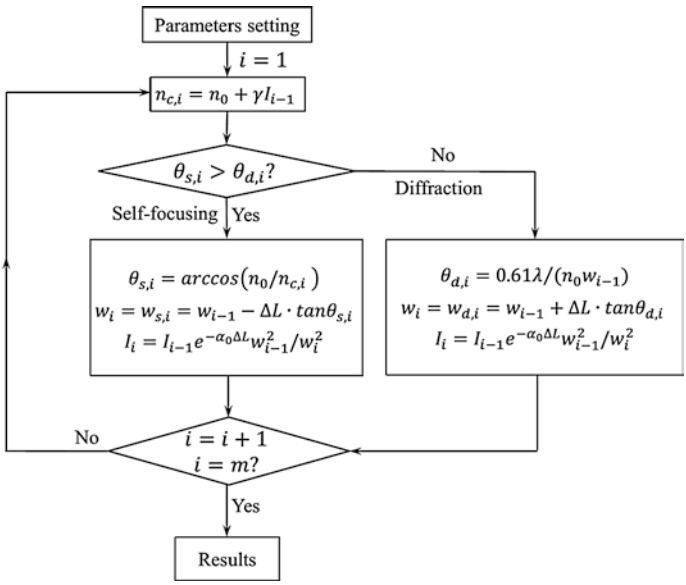

Fig. 5.14 Calculation flow of the light beam traveling inside the nonlinear refractive samples

5.3.2 Light Traveling Inside Positive Nonlinear Refraction Samples

5.3.2.1 Calculation Flow Process

Arsenic trisulfide (As_2S_3) film is still taken as an example due to strong nonlinear refractive coefficient, but little absorption at 633 nm wavelength. The calculation parameters are as follows: $\lambda = 633$ nm,$w_0 = 2.15$ μm,$\alpha_0 = 1000\ m^{-1}$, $\gamma = 8.64 \times 10^{-10}$ m^2/W [12], $L = 1.7$ μm, $\Delta L = 1$ nm, and $n_0 = 2.32$. The initial incident laser power density is $I_0 = 2P/\pi w_0^2$, where P is laser power. The numerical calculation flow is presented in Fig. 5.14. First, one needs to set the initial parameters, compute the central refractive index of the first layer ($i = 1$), and then compare the self-focusing and diffraction effects for the ith layer. If the self-focusing is dominant, then the light beam size and intensity are calculated using formula (5.34) or else should be calculated by formula (5.37). Then, one calculates the $(i + 1)$th layer by setting $i = i + 1$. The exiting spot from the sample bottom can be obtained when $i = m$.

5.3.2.2 Self-focusing Spot

The typical exiting spot intensity profiles at the bottom of sample are shown in Fig. 5.15. Figure 5.15a is the incident spot intensity profile, the spot diameter is

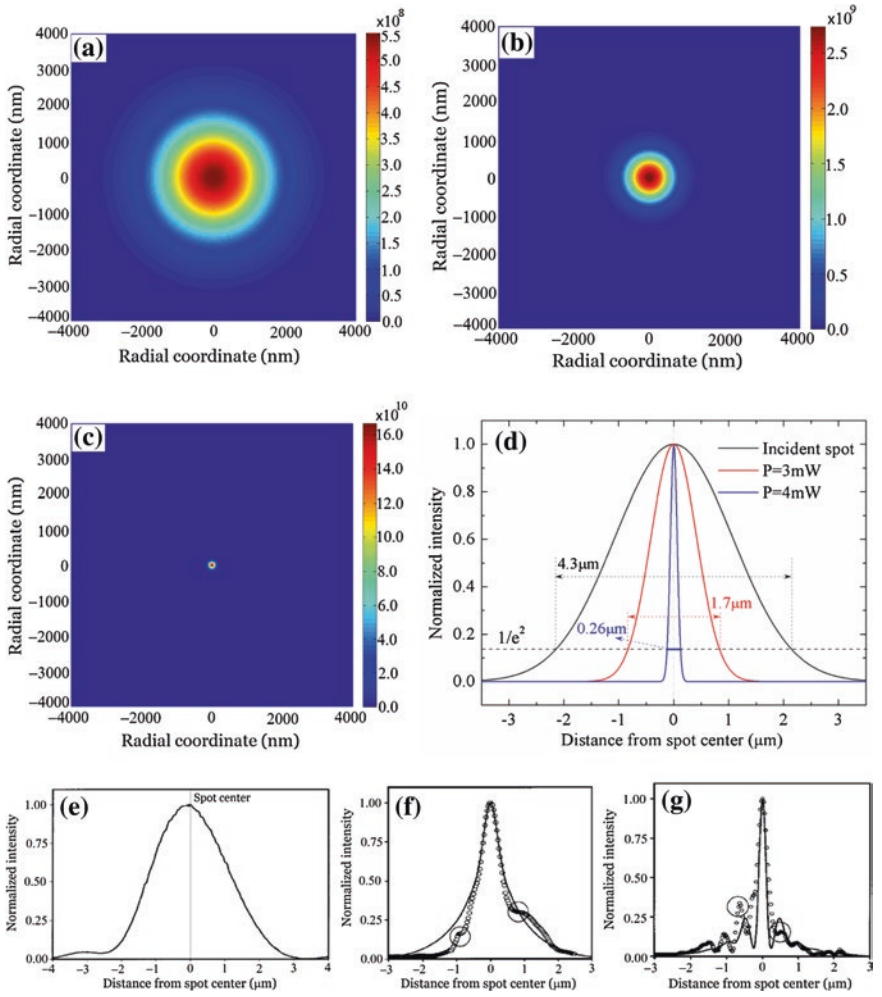

Fig. 5.15 The intensity profile of exiting spot at the bottom of sample, **a** incident spot, **b** $P = 3.0$ mW, and **c** $P = 5.0$ mW, and **d** the cross-section curves of **a**–**c**. The experimental observation of self-focusing spot intensity profiles from Ref. [18], **e** initial incident spot, self-focusing spot **f** at $P = 3.0$ mW, and **g** at $P = 5.0$ mW. Reprinted with permission from Ref. [18]. Copyright 2003, American Institute of Physics

about 5.0 μm, and central intensity is about 5.5×10^8 W/m^2. Figure 5.15b is the exiting spot at $P = 3$ mW; compared with the incident spot, the spot size is reduced, and the central intensity is about 2.7×10^9 W/m^2, which indicates that the self-focusing effect takes place. Figure 5.15c is the exiting spot at $P = 5$ mW, a very small spot occurs; the spot size is greatly reduced compared with the Fig. 5.15a, b, and the central intensity is about 1.67×10^{11} W/m^2, which demonstrates that the self-focusing effect at $P = 5$ mW is obviously stronger than that at $P = 3$ mW.

Figure 5.15d is the cross-section intensity profile of the Fig. 5.15a–c. The incident spot diameter is about 5.0 μm. The self-focusing spot diameter for $P = 3$ mW is reduced to about 1.7 μm; the reduction ratio of spot size is about 66 %. At $P = 5.0$ mW, the self-focusing spot diameter at $1/e^2$ intensity is about 0.26 μm, the size is about the order of the wavelength inside the sample, and the ratio of spot size reduction is about 95 %. The self-focusing spot size of 0.26 μm is close to the diffraction-limited spot size inside the sample and has been reduced to the minimum for the As_2S_3 sample at the 633 nm laser wavelength irradiation. The experimental results from Ref. [18] are also presented in Fig. 5.15e–g, where the experimental conditions are the same as the simulation parameters. The circles are the experimental points, and the solid lines are the fitting curves based on the numerical calculation of the non-paraxial approximation method from Ref. [19]. In the spot scanning experiment, the initial incident spot size is about 5.0 μm (also see Fig. 5.15e) and then is reduced to about 2.0 μm at $1/e^2$ intensity for $P = 3.0$ mW; the reduction ratio is about 60 % (also see Fig. 5.15f). At $P = 5.0$ mW, the self-focusing spot size at $1/e^2$ intensity is reduced to about 0.28 μm; the reduction ratio is close to 95 % (also see Fig. 5.15g).

5.3.2.3 Light Beam Self-focusing Behavior Inside the Nonlinear Samples

The self-focusing behavior of light beam inside the nonlinear sample can be obtained using the multilayer thin lens self-focusing model. Figure 5.16a gives the dependence of the full width at half maximum (FWHM) on incident laser power. The FWHM basically decreases linearly with the laser power increasing at $0 \leq P \leq 4$ mW. At $P > 4$ mW, the FWHM is about 0.15 μm and remains basically unchanged with laser power increasing. That is, there is no more reduction of the beam diameter at 4.0 mW $\leq P \leq$ 5.0 mW. In the range of the incident laser power, the reduction of the beam diameter is saturated and has reached the limit at which the self-focusing effect cannot overcome diffraction. This means that the self-focusing beam is arrested. The minimum beam FWHM is about 0.15 μm,

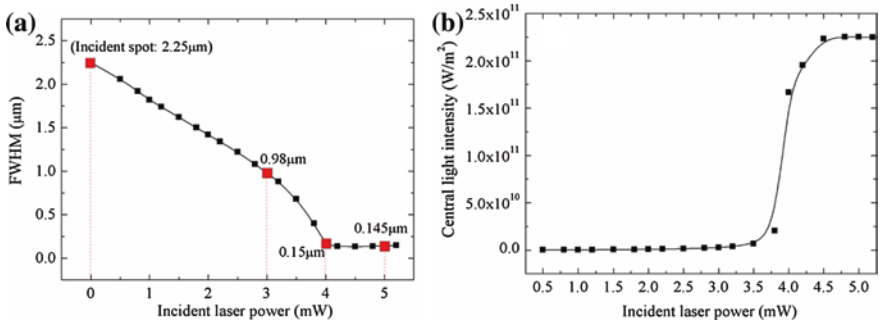

Fig. 5.16 The dependence of **a** FWHM and **b** central light intensity on incident laser power

and a series of focusing–defocusing cycles do not occur in the simulation results. The simulations are also consistent with the experimental results reported in Ref. [18]. The central light intensity is first gradually increased with laser power increasing at $P < 3.5$ mW, as shown in Fig. 5.16b, and then abruptly increases 3.5 mW $\leq P \leq$ 4.5 mW, and remains almost stable at P > 4.5 mW.

The light beam intensity profile inside the nonlinear samples is presented in Fig. 5.17, where the light beam incident is along sample thickness direction. Figure 5.17a is for $P = 3$ mW, the light beam slowly converges to the sample bottom, and the light intensity also becomes stronger. Figure 5.17b is for $P = 4$ mW, a very strong spot is formed at the sample bottom. When the incident laser power is increased to 5 mW, a self-focusing needle is generated, and the needle is extended to the sample bottom.

In order to intuitively observe the light beam self-focusing process inside the nonlinear sample, Fig. 5.18 presents the profile of light intensity normalized by the optical axis center intensity for different sample thickness position. One can see the light beam traveling and self-focusing behavior inside the sample.

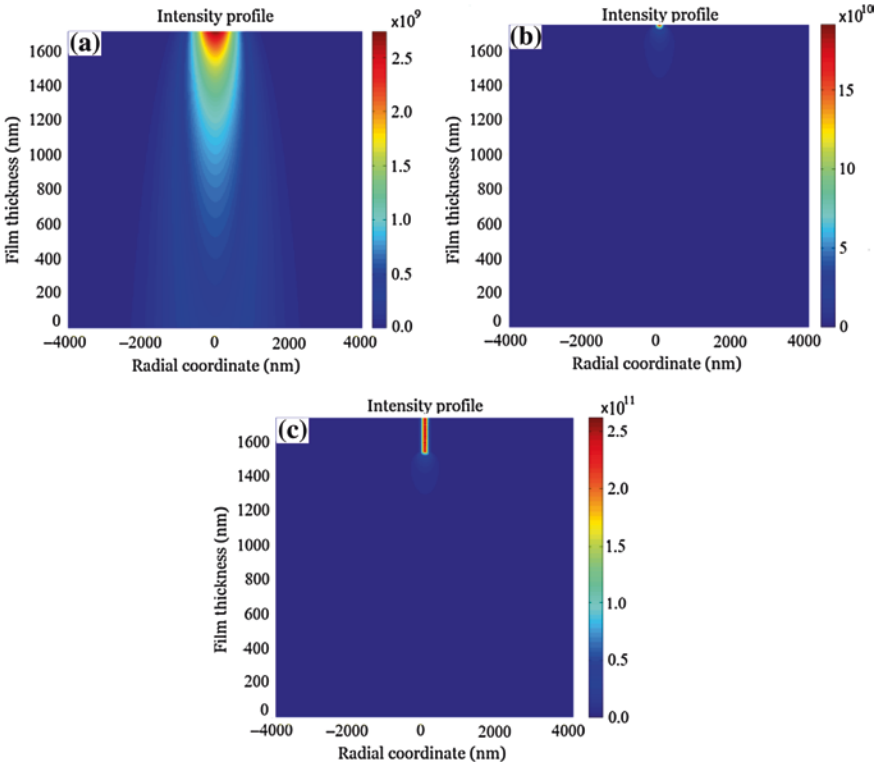

Fig. 5.17 Light beam intensity profile inside the nonlinear sample at different laser powers, **a** $P = 3$ mW, **b** $P = 4$ mW, and **c** $P = 5$ mW

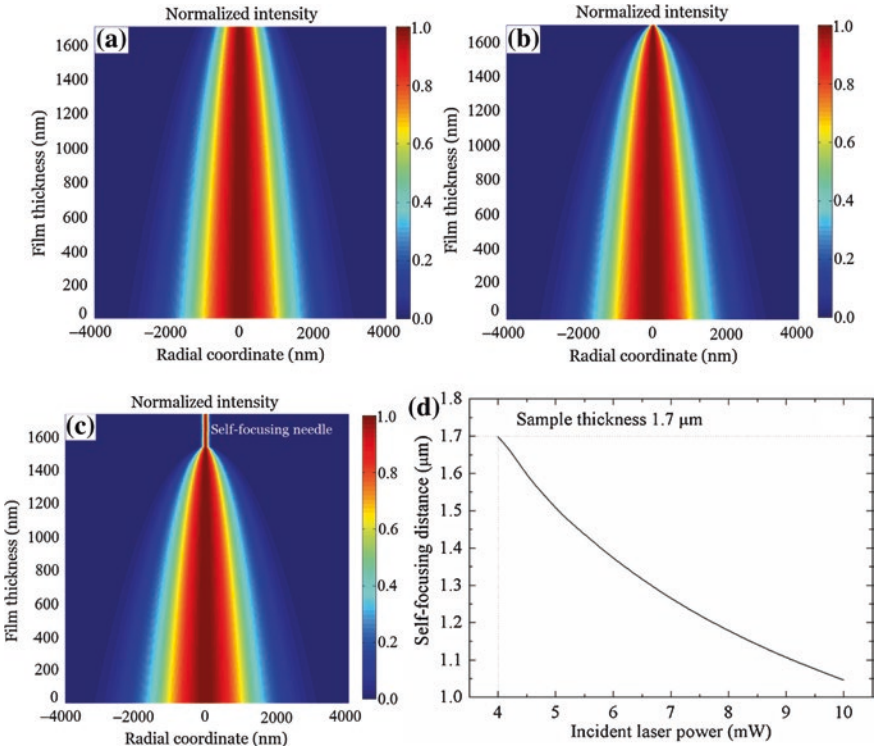

Fig. 5.18 Normalized intensity of self-focusing behavior inside the sample, **a** $P = 3.0$ mW, **b** $P = 4.0$ mW, **c** $P = 5.0$ mW, and **d** dependence of self-focusing spot position on incident power

Figure 5.18a is at $P = 3.0$ mW; the light beam size slowly becomes small with traveling distance. Figure 5.18b is at $P = 4.0$ mW, the light beam is self-focused into a diffraction-limited spot at the bottom of the sample, which indicates that the self-focusing effect becomes saturation, and a diffraction-limited self-focusing spot is formed inside the sample. A self-focusing needle is generated after the diffraction-limited spot when one continues to increase the laser power to $P = 5.0$ mW (as shown in Fig. 5.18c), which shows that the self-focusing reaches balance with the diffraction effect, and the self-focusing needle size remains unchanged. By comparison, the size of exiting spot in Fig. 5.18c is the same as Fig. 5.18b. The difference is that in Fig. 5.18c, the diffraction-limited self-focusing spot is generated inside the sample, and the light beam travels in the form of self-focusing needle. In Fig. 5.18b, the diffraction-limited self-focusing spot is generated at the bottom of sample. That is, the position of the diffraction-limited self-focusing spot can be manipulated by changing the laser power. Figure 5.18d presents the dependence of self-focusing distance on incident laser power, the diffraction-limited self-focusing spot occurs at the bottom of the sample when P $\approx$ 4.0 mW, and then, the self-focusing distance linearly decreases with the laser power increasing, and the higher the laser power, the shorter the self-focusing distance.

5.3.3 Comparison with Equivalent Converging Lens Model

Generally, if ignoring the diffraction effect and nonlinear refractive index variation inside the sample, the self-focusing spot can be roughly estimated using an equivalent converging lens method. That is, the Gaussian laser beam irradiation-induced nonlinear refraction effect can be considered as converging lens, as is shown in Fig. 5.19. The light beam with Gaussian intensity profile is normally incident onto the nonlinear sample, and an equivalent converging lens with self-focusing angle of θ_{sf} is produced, accordingly. The light beam radius is w_0. The equivalent converging lens focuses the light into a spot with a radius of w_{sf}.

For the converging lens, the central refractive index can be rewritten as

$$n_{sf} = n_0 + \gamma I_0 \tag{5.38}$$

The self-focusing angle is also rewritten as,

$$\theta_{sf} = cos^{-1}(n_0/n_{sf}) \tag{5.39}$$

According to Fig. 5.19, one has

$$\cos\theta_{sf} = \frac{n_0}{n_{sf}} = \frac{z_{sf}}{\sqrt{z_{sf}^2 + (\Delta w)^2}} \tag{5.40}$$

with

$$\Delta w = w_0 - w_{sf} \tag{5.41}$$

By mathematical operation, the self-focusing distance is,

$$z_{sf} = \frac{n_0(w_0 - w_{sf})}{\sqrt{n_{sf}^2 - n_0^2}} \tag{5.42}$$

According to Refs. [20, 21], using the lens formula, the diameter of self-focusing spot can be calculated as

$$D_{sf} = 2w_{sf} = \frac{0.61\frac{\lambda}{n_0}}{w_0} z_{sf} = \frac{0.61\lambda}{n_0 w_0} z_{sf} \tag{5.43}$$

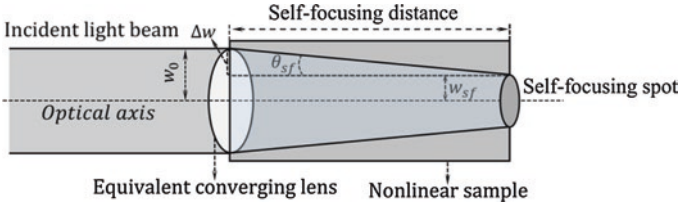

Fig. 5.19 Schematics of equivalent self-focusing converging lens

Substituting formula (5.42) into formula (5.43), and by a series of mathematical operation, one has

$$w_{sf} \approx \frac{0.3\lambda w_0}{w_0\sqrt{n_{sf}^2 - n_0^2} + 0.3\lambda} \tag{5.44}$$

The self-focusing spot intensity can be calculated as

$$I_{sf} = I_0 e^{-\alpha_0 L} w_0^2 / w_{sf}^2 \tag{5.45}$$

The $P = 5.0$ mW is taken as an example. The incident light spot radius $w_0 \sim 2.15$ μm , $\gamma = 8.64 \times 10^{-10}\text{m}^2/W$, $\alpha_0 = 1{,}000\ \text{m}^{-1}$, $L = 3.0$ μm, and $n_0 = 2.32$. By mathematical computing, one can obtain $I_0 = 2P/\pi w_0^2 \sim 5.5 \times 10^8\ \text{W/m}^2$, $n_{sf} \sim 2.91$, $\theta_{sf} = 37.13°$, and $w_{sf} = 0.103$ μm. The spot diameter is $2 \times w_{sf} = 0.206$ μm , accordingly. The self-focusing distance and intensity are $z_{sf} \sim 2.70$ μm and $I_{sf} \sim 2.9931 \times 10^{11}\ \text{W/m}^2$, respectively. Compared with the multilayer thin lens self-focusing model, where the spot radius and intensity are about 0.28 μm and $1.67 \times 10^{11}\ \text{W/m}^2$, respectively, one can find that the difference is very obvious. The spot radius obtained by equivalent converging lens model is also obviously different with the experimental reports of Fig. 5.15g. These indicate that for nonlinear thick sample, the equivalent converging lens model is not suitable, and the multilayer thin lens self-focusing model is a good alternative for analyzing and understanding the self-focusing behavior and spot performances.

5.3.4 Application Schematic Design

The self-focusing spot can be reduced to nanoscale by manipulating the incident laser power, and the nanoscale self-focusing spot is useful for high-density data

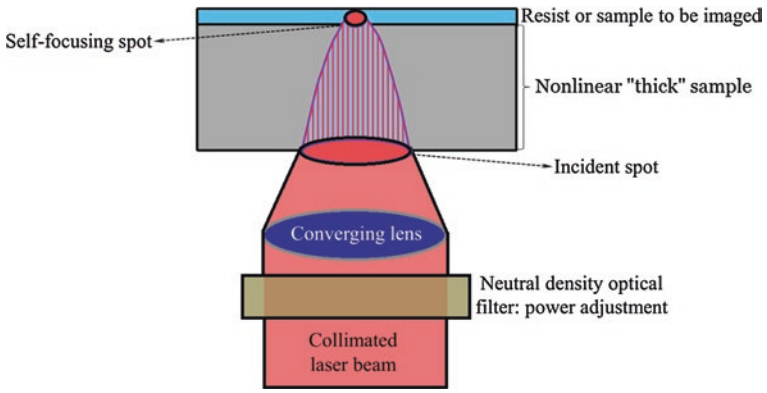

Fig. 5.20 Designed application schematics

storage, nanolithography, and high-resolution light imaging. The application schematics, similar to Fig. 5.10, can be roughly designed as follows. The resists or the samples to be imaged are directly deposited (or placed) onto the nonlinear samples, as shown in Fig. 5.20. A collimated laser beam is focused and normally incident onto the nonlinear thick film. A self-focusing spot is formed at the bottom of sample and directly coupled into the resists or the samples to be imaged. The self-focusing spot size and focal distance can be finely adjusted by the neutral density optical filter. The high-density data storage, nanolithography, and high-resolution light imaging can be carried out using the self-focusing spot to scan the samples.

5.4 Summary

Nonlinear refraction effect can generate nanoscale spot, which can be realized by self-focusing behavior or interference manipulation. For "thin" film samples, the refractive index changes only along radial direction and almost remains unchanged along sample thickness direction. Taking advantage of the internal multiple constructive interference effect, one can obtain the nanoscale spot through designing a nonlinear Fabry–Perot structure. For "thick" sample, considering the refractive index change along sample thickness direction and diffraction effect, the self-focusing behavior is dominant over interference effect and one can analyze the light field characteristics using multilayer thin lens self-focusing model. Here, it should be noticed that for "thick" sample, the self-focusing spot is still determined by the diffraction limit; thus, in order to obtain small self-focusing spot, the linear refraction is also required to be as strong as possible.

References

1. W. Yan, N. Mortensen, M. Wubs, Hyperbolic metamaterial lens with hydrodynamic nonlocal response. Opt. Exp. **21**(12), 15026–15036 (2013)
2. Z. Wang, W. Guo, L. Li, B. Luk'yanchuk, A. Khan, Z. Liu, Z. Chen, M. Hong, Optical virtual imaging at 50 nm lateral resolution with a white-light nanoscope. Nat. Commun. **2**, 218 (2011)
3. E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, H.F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution. Science **313**(5793), 1642–1645 (2006)
4. M.J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods **3**, 793–795 (2006)
5. K. Konno, H. Suzuki, Self-focussing of laser beam in nonlinear media. Phys. Scr. **20**, 382–386 (1979)
6. S. Kima, M. Kanga, and S. Choia, Nonlinear characteristics of super resolution chalcogenide thin film, Materiaux Resource Society Symposium Proceedings 803, HH6.6.1-HH6.6.5 (2003)
7. J. Wei, R. Wang, H. Yan, Y. Fan, Formation of super-resolution spot through nonlinear Fabry-Perot cavity structures: theory and simulation. Opt. Exp. **22**(7), 7883–7897 (2014)

8. M. Born and E. Wolf, *Principle of Optics*, 7th edn. (Cambridge University, Cambridge, 1999)
9. Robert W. Boyd, *Nonlinear optics*, 2nd edn. Academic (2003)
10. X. Ma, J. Wei, Nanoscale lithography with visible light: optical nonlinear saturation absorption effect induced nanobump pattern structures. Nanoscale **3**, 1489–1492 (2011)
11. J. Wei, S. Liu, Y. Geng, Y. Wang, X. Li, Y. Wu, A. Dun, Nano-optical information storage induced by the nonlinear saturable absorption effect. Nanoscale **3**, 3233–3237 (2011)
12. K.B. Song, J. Lee, J.H. Kim, K. Cho, S.K. Kim, Direct observation of self-focusing with sub-diffraction limited resolution using near-field scanning optical microscope. Phys. Rev. Lett. **85**(18), 3842 (2000)
13. J. Wei, M. Xiao, F. Zhang, Super-resolution with a nonlinear thin film: beam reshaping via internal multi-interference. Appl. Phys. Lett. **89**, 223126 (2006)
14. N. Pinna and M. Knez, *Atomic Layer Deposition of Nanostructured Materials*, WILEY-VCH (2011)
15. M. Kreuzer, H. Gottschling, R. Neubecker, T. Tschudi, Analysis of dynamic pattern formation in nonlinear Fabry-Perot resonators. Appl. Phys. B **59**, 581–589 (1994)
16. J. Wei, J. Liu, Direct observation of below-diffraction-limited optical spot induced by nonlinear saturation absorption of Ag-doped Si nanofilms. Opt. Lett. **35**, 3126–3128 (2010)
17. J. Wei and H. Yan, Gaussian beam spot induced nanoscale spot through nonlinear thick sample: multi-layer thin lens self-focusing model. J. Appl. Phys. 116, 063107 (2014)
18. J. Lee, S. Park, K. Cho, K. Song, Analysis of self-focusing beams of concentric symmetry in As_2S_3 glass using a near-field scanning optical microscope. Appl. Phys. Lett. **82**, 2416 (2003)
19. M.D. Feit, J.A. Fleck Jr, Beam nonparaxiality, filament formation, and beam breakup in the self-focusing of optical beams. J. Opt. Soc. Am. B **5**, 633–640 (1988)
20. A.A. Mityureva, V.V. Smirnov, O.A. Vorobiev, A nanohole as an atomic lens for atoms and molecules. J. Phys. D Appl. Phys. **34**(13), L65–L69 (2001)
21. J. Wei, Y. Wang, W. Xu, F. Zhou, F. Zhang, F. Gan, Working mechanism of Sb thin films in super-resolution near-field structure. Opt. Commun. **224**(4–6), 269–273 (2003)

Chapter 6
Optical Super-Resolution Effect Through Nonlinear Saturation Absorption

6.1 Basic Description of Nonlinear Saturation Absorption-Induced Super-Resolution Effect

Nonlinear saturation absorption is a well-known phenomenon in most optical materials. Taking advantage of nonlinear saturation absorption thin films to reduce the spot to nanoscale can avoid the difficulty of controlling the near-field distance between optical probe and sample surface, which has been experimentally applied to nano-optical data storage and nanoscale lithography [1, 2]. The basic physical picture of this technique is presented in Fig. 6.1. A collimated Gaussian laser beam is focused on the nonlinear saturation absorption thin films and induces nonlinear saturation absorption effect. For nonlinear saturation absorption, a higher laser intensity results in a larger transmittance, which causes the transmittance to be the largest in the central region of the spot and then gradually decrease along the radial direction. A below-diffraction-limit optical pinhole channel is produced in the nonlinear saturation absorption thin films. The laser beam goes through the optical pinhole channel, and a super-resolution spot is obtained at the back side of the nonlinear saturation absorption thin film. Nano-optical storage and nanolithography are realized by interaction of the super-resolution spot with the information-recording medium or photoresist by near-field coupling. The optical pinhole channel can automatically close when the laser beam is removed, and the nonlinear saturation absorption thin film returns to its original state. Thus, the formation of below-diffraction-limited optical pinhole channel is dynamic and reversible.

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_6

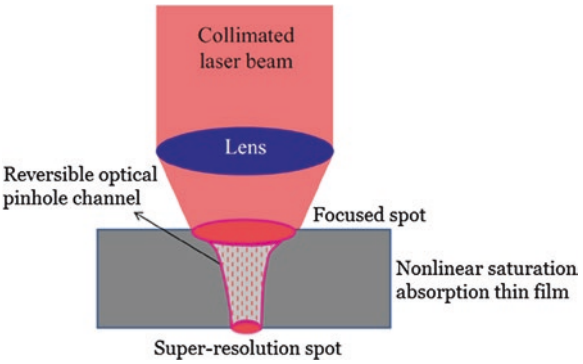

Fig. 6.1 Schematics for optical pinhole channel generation and super-resolution spot formation

6.2 Beer–Lambert Model for Thin (or Weak) Nonlinear Saturation Absorption Sample

6.2.1 Beer–Lambert Analytical Model

In nonlinear saturation absorption-induced super-resolution effect, for thin (or weak) nonlinear samples, some approximations can be used. The first is that the incident light is thought to be a collimated beam and does not diverge or broaden in propagation inside the sample. The second is that the absorption coefficient remains unchanged along sample thickness direction. According to the approximations, the irradiation of Gaussian incident laser beam spot with initial intensity distribution is

$$I_{\text{inc}}(r) = I_0 \exp\left(-2r^2/w_0^2\right), \quad \text{with} \quad I_0 = 2P/\pi w_0^2 \tag{6.1}$$

where P is incident laser power and w_0 is spot radius at $1/e^2$ central intensity, the absorption coefficient is marked as

$$\alpha(r) = \alpha_0 + \beta I_{\text{inc}}(r) \tag{6.2}$$

where α_0 and β are linear and nonlinear absorption coefficients, respectively. Formula (6.2) shows that the absorption coefficient presents a reverse Gaussian spatial profile for nonlinear saturation absorption samples with $\beta < 0$.

For the samples with nonlinear absorption without nonlinear refraction, there is no self-action effect, and diffraction effect is also ignorable. The light is incident on the thin (or weak) nonlinear saturation absorption sample, and the intensity decays exponentially along thin film thickness direction. The exiting spot from the sample can be approximately analyzed using the Beer–Lambert formula (abbreviated as B–L model).

For thin film samples irradiated by a laser beam with an initial intensity profile of $I_{\text{inc}}(r)$, the intensity decaying can be expressed as follows:

$$\mathrm{d}I/\mathrm{d}z = -\alpha I \tag{6.3}$$

where $\mathrm{d}I$ is the variation of the laser intensity along the sample thickness direction, $\mathrm{d}z$. The transmitted beam spot intensity can be obtained by integrating formula (6.3) as follows:

$$I_t(r) = I_{\text{inc}}(r)e^{-\alpha L} \tag{6.4}$$

where L is the thin film thickness. Formula (6.4) is the B–L model. For most samples, $\alpha = \alpha_0$ is constant. The normalized transmitted beam spot intensity profile, $I_t(r)$, is spatially the same as the initial incident laser intensity, $I_{\text{inc}}(r)$.

The beam spot passing through the nonlinear thin film sample can be calculated by substituting formula (6.2) into (6.4) to obtain

$$I_t(r) = I_{\text{inc}}(r)\exp\{-[\alpha_0 + \beta I_{\text{inc}}(r)]L\} \tag{6.5}$$

Generally, the initial incident beam intensity presents Gaussian profile and can be expressed by formula (6.1). Substituting formula (6.1) into (6.5) yields the transmitted beam spot intensity through the nonlinear thin film sample:

$$I_t(r) = I_0 \exp\left(-\frac{2r^2}{w_0^2}\right) \exp\left\{-\left[\alpha_0 + \beta I_0 \exp\left(-\frac{2r^2}{w_0^2}\right)\right]L\right\} \tag{6.6}$$

For nonlinear saturation absorption thin film samples, the intensity at $r = 0$ is the strongest, and the normalized transmitted beam spot is derived as

$$I_{\text{trans.normalized}}(r) = \frac{I_t(r)}{I_t(0)} = \frac{\exp\left(-\frac{2r^2}{w_0^2}\right)\exp\left[-\beta I_0 \exp\left(-\frac{2r^2}{w_0^2}\right)L\right]}{\exp(-\beta I_0 L)} \tag{6.7}$$

where $I_{\text{trans.normalized}}(r)$ is normalized spatial intensity profile of the transmitted spot, and the transmitted spot is exactly the super-resolution spot.

Formula (6.7) is also called as (nonlinear) B–L model. The B–L model needs to meet two requirements. One is that the absorption coefficient is thought to be uniform along the sample thickness direction and is only a function of radial coordinate r, marked as $\alpha(r)$. The other is that the radius of focused spot w_0 should remain unchanged along the sample thickness direction coordinate z. In general, the two requirements are easily met for thin film with a thickness of less than 100 nm.

Formula (6.7) shows that the size of the super-resolution spot is determined by nonlinear absorption coefficient β, thin film thickness L, initial incident spot radius w_0, and laser power density I_0. For a given nonlinear thin film device, β and L are fixed, and w_0 is determined by the testing system. Thus, the super-resolution spot size can only be tuned by changing the laser power density, I_0, namely the laser power P because $I_0 = 2P/(\pi w_0^2)$.

6.2.2 Experimental Observation of Super-Resolution Spot

6.2.2.1 Ag-Doped Si Thin Films Used as Nonlinear Saturation Absorption Samples

In order to verify the ability to reduce the spot to below diffraction limit by nonlinear saturation absorption effect, the Ag-doped Si thin films are prepared by magnetron-controlled co-sputtering of Si target and Ag target using Ar^+ ions. The Ag

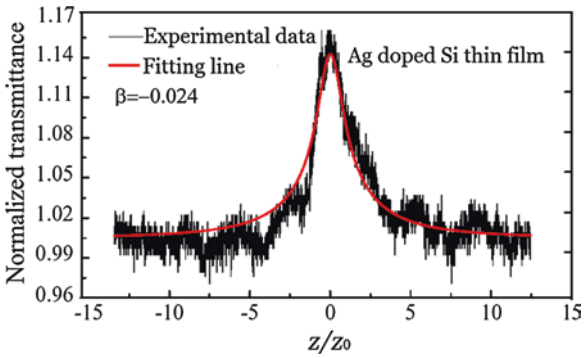

Fig. 6.2 Nonlinear measurement (z_0 is the Rayleigh length of z-scan setup). Reprinted with permission from Ref. [4]. Copyright 2010, Optical Society of America

concentration is tuned by changing the sputtering power. The nonlinear properties are measured by a z-scan method where a He–Ne laser with a wavelength (λ) of 633 nm is used. The He–Ne laser is modulated into pulse light by an acoustic–optic modulator, where the laser pulse width can be randomly tunable. The laser power can be changed by a variable attenuator, and detailed experimental setup schematic can be found in Ref. [3] and Chap. 3. The laser pulse width is fixed at 50 ns, and a converging lens with a numerical aperture (*NA*) of 0.09 is used to conduct the z-scan measurement. The laser power P is fixed at 2.5 mW. The laser intensity can be calculated by $I_0 = 2P/(\pi w_0^2) \approx 8.625 \times 10^7\ \mathrm{W/m^2}$ accordingly, where $w_0 (= 1.22\lambda/2NA)$ is the beam waist radius. The Ag-doped Si thin film thickness L is about 100 nm.

The measured results are shown in Fig. 6.2. The Ag-doped Si thin film presents an obvious nonlinear saturation absorption characteristic. The theoretical fitting for the measured data indicates that the nonlinear saturation absorption coefficient β is about -0.024, which is very large value compared with most nonlinear absorption materials [4]. Here, it needs to be noted that the nonlinear saturation absorption coefficient of $\beta = -0.024$ is a little smaller than reported in Chap. 4, and no obvious nonlinear refraction effect occurs. These may result from different Ag particle concentration. The Ag particle concentration is generally determined by the sputtering parameters. In this experiment, the percentage composition of Ag particles may be a little deviation from the resonance concentration. The detailed physical essence of the nonlinear effects can be found in Chap. 4 and Ref. [5].

In order to understand the nonlinear saturation absorption-induced super-resolution effect, the sample structure as "Glass substrate with a thickness of 1.2 mm/Ag-doped Si thin film with a thickness of 20 nm" is firstly prepared. The diffraction-limited focused spot with a Gaussian intensity profile passes through the glass substrate and irradiates on the nonlinear saturation absorption layer surface, as shown in Fig. 6.3. For the nonlinear saturation absorption materials, the higher the laser beam intensity, the larger the transmittance is. Thus, there is the largest transmittance at the central part of spot, which causes a small optical channel to

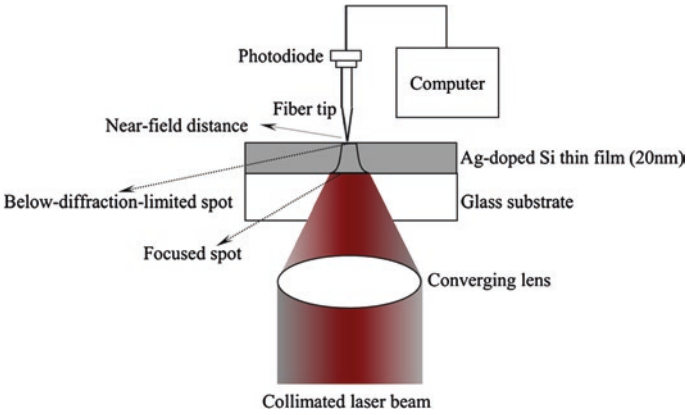

Fig. 6.3 Schematic of near-field spot scanning with fiber tip. Reprinted with permission from Ref. [4]. Copyright 2010, Optical Society of America

be formed at the center of spot. The focused laser beam passes through the channel, and a below-diffraction-limited spot is generated at the surface of Ag-doped Si thin film. The reduction degree of the spot can be tuned by changing the laser intensity.

In order to further carry out a direct observation of the nonlinear saturation absorption-induced below-diffraction-limited spot, a near-field optical spot scanning setup is established. As is shown in Fig. 6.3, a laser diode at wavelength 635 nm is applied as the light source, and the converging lens with an NA of about 0.55 is used to focus the spot. Here, it needs to be explained that a high-NA converging lens being not adopted is because the working distance of the higher-NA lens is shorter than the thickness of glass substrate, which cannot realize the focusing on the Ag-doped Si thin film. The theoretical full width at half-maximum (FWHM) of the diffraction-limited spot can be calculated by $D = 0.61\lambda/NA$, for setup in Fig. 6.3; the theoretical FWHM is about 0.704 μm. The laser beam passes through the glass substrate and is focused on the Ag-doped Si thin film. After the thin film, a metal-coated fiber tip affiliated onto the tuning fork is used to scan the laser beam spot for the intensity distribution. The tuning fork is attached on a three-dimensional piezoelectric transducer (PZT), the lateral maximum moving distance of the PZT is 60 μm, and the lateral scanning resolution is about 100 nm limited by the aperture of the fiber probe. The distance between the sample surface and the fiber tip is modulated by an atomic force microscopy controller, and when the fiber tip gets close to the sample surface, the resonance frequency of the tuning fork would shift due to the generated shear force. The controller receives this shift signal as error and regulates the z-axis voltage onto the PZT. The below-diffraction-limited spot intensity profile information is collected by the scanning fiber tip and then transferred to the photodiode and produces the signal in the format of voltage to the computer.

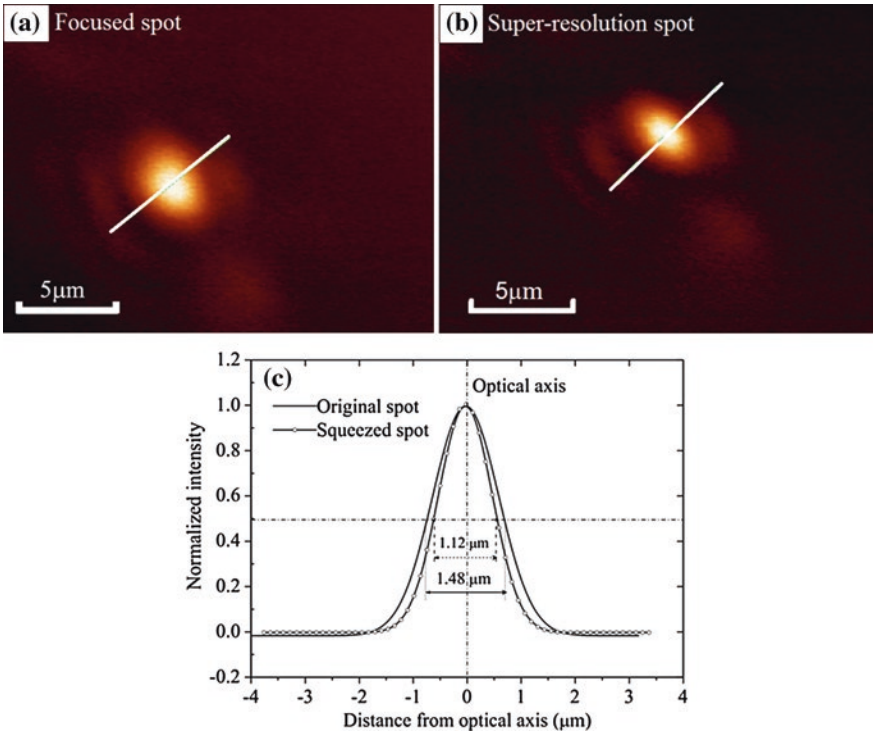

Fig. 6.4 Normalized real spot images obtained by near-field fiber probe scanning method, **a** the diffraction-limited focused spot, **b** the squeezed spot at laser power of 11.5 mW, and **c** cross-sectional comparisons of the squeezed spot with the diffraction-limited focused spot. Reprinted with permission from Ref. [4]. Copyright 2010, Optical Society of America

Before scanning the nonlinear saturation absorption thin films, the diffraction-limited focused spot through the glass substrate with a thickness of 1.2 mm is firstly scanned, and the result is shown in Fig. 6.4a. It can be seen that the focused spot is an elliptical shape, and the FWHM of the short axis is about 1.48 μm; this size is larger than theoretical FWHM, which may be from three aspects. The first is that the glass substrate (1.2 mm thickness) affects the light path due to the refraction; the second is that the semiconductor diode laser beam is not collimated and has a large divergence angle; the third is that the setup may be not very perfect; for example, the real NA of the converging lens is a little smaller than theoretical $NA(= 0.55)$. However, the measurement mainly focuses on the squeezed spot size ratio; thus, the theoretical spot size of the setup has little effect on the experimental results. The formation of the elliptical spot could be experienced due to the laser diode's junction. It is possible to avoid this elliptical spot by overfilling the back aperture of the converging lens. However, the effective laser power decreases greatly, which has an ill-effect on the direct observation of spot size reduction.

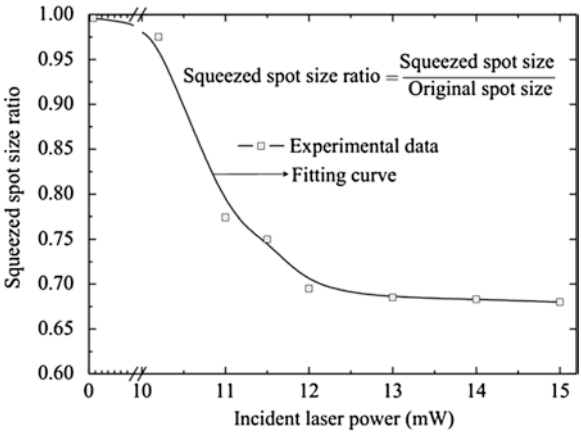

Fig. 6.5 Dependence of the squeezed spot size ratio on laser power. Reprinted with permission from Ref. [4]. Copyright 2010, Optical Society of America

Then, the glass substrate is replaced by "Glass substrate with a thickness of 1.2 mm/Ag-doped Si thin film with a thickness of 20 nm" sample, and the spot is scanned along the Ag-doped Si thin film surface. The laser power is fixed at about 11.5 mW. The spot intensity profile is presented in Fig. 6.4b; the spot shape is still elliptical, which is the same as Fig. 6.4a. Whether at the short-axis direction or at the long-axis direction, the spot size is obviously squeezed; that is, the below-diffraction-limited spot is generated by the Ag-doped Si thin film due to the nonlinear saturation absorption effect. In order to compare more effectively and conveniently, the cross-sectional distribution curves of Fig. 6.4a, b at the short-axis direction (marked in white color line) are plotted in Fig. 6.4c. The FWHM of the short-axis size of the squeezed spot is reduced to about 1.12 μm, which is about 75 % FWHM of the diffraction-limited spot size. In other words, the squeezed spot size ratio is about 0.75, where the squeezed spot size ratio is defined as the ratio of squeezed spot size to the diffraction-limited focused spot size.

By changing the laser power, the squeezed spot size can be tuned as well; the results are given in Fig. 6.5. The results indicate that the squeezed spot ratio decays exponentially with laser power increasing. The exponential decaying seems to accord with the B–L model. That is, the reduction degree of the squeezed spot size increases with laser power, and the higher the laser power, the smaller the spot size is. However, from Fig. 6.5, the spot does not be squeezed until the laser power increases to about 10 mW. In other words, there is an abrupt reduction of spot size at laser power of ~10 mW, which may be due to the nonlinearity excitation with threshold effect. The spot cannot be squeezed infinitely because it is impossible to infinitely increase the laser power due to strong laser intensity leading to the damage of the materials. The smallest spot can be squeezed to 68 % the diffraction-limited focused spot size, which is very useful for high-density optical recording and super-resolution optical lithography.

6.2.2.2 Sb_2Te_3 Thin Films Used as Nonlinear Saturation Absorption Samples

The Ag-doped Si thin films are replaced with the Sb_2Te_2 thin films. The nonlinear absorption coefficient and thickness of Sb_2Te_3 thin film are approximately $\beta = -2.5 \times 10^{-2}$ m/W [6] and $L = 20$ nm, respectively. Similar to Sect. 6.2.2.1, the super-resolution spot is also measured. Figure 6.6a shows the morphology and intensity profile of the spot passing through glass substrate. The focused spot of the testing setup is elliptical, where the NA of lens in the testing setup is smaller than that in Sect. 6.2.2.1.

Subsequently, the glass substrate is replaced with the "glass/Sb_2Te_3 (20 nm)" sample and scanned along the Sb_2Te_3 surface. The laser power is fixed at 8 mW. The spot intensity profile is presented in Fig. 6.6b, where the spot shape remains elliptical. At both short- and long-axis directions, the spot size is observed to be squeezed; that is, the super-resolution spot is generated by the Sb_2Te_3 thin film due to the nonlinear saturation absorption effect. The cross-sectional profile curves of Fig. 6.6a, b at the short-axis direction (marked as green color line) are

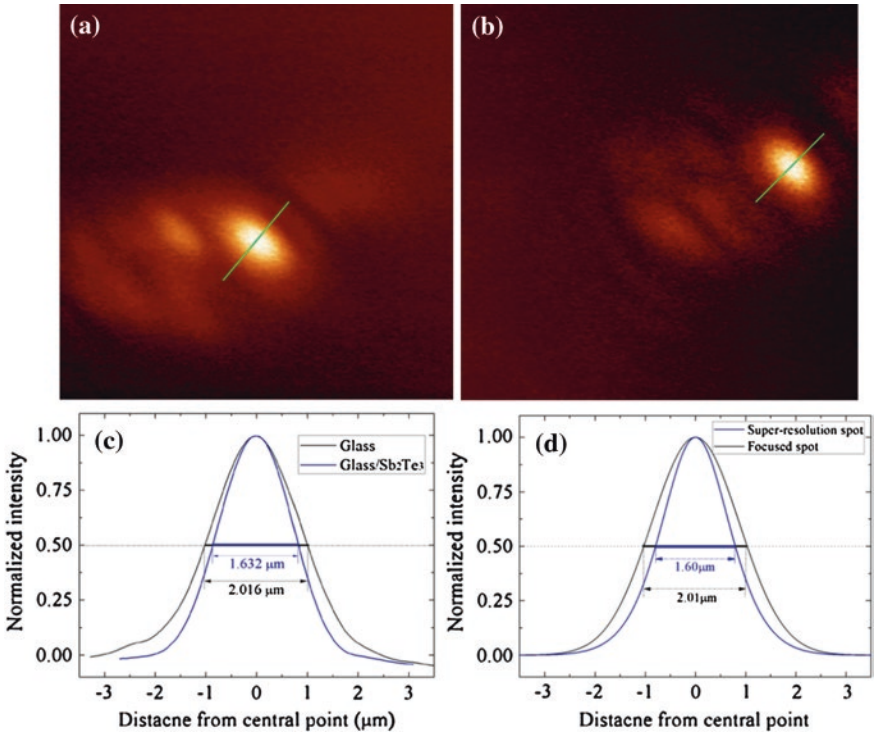

Fig. 6.6 Spot analysis by the near-field optical probe scanning method: **a** through glass substrate and **b** through "glass/Sb_2Te_3" sample. Comparison of the normalized intensity profile along the cross-sectional direction: **c** experimental measurement and **d** theoretical calculation by formula (6.7). Reprinted from Ref. [7], Copyright 2013, with permission from Elsevier

plotted in Fig. 6.6c to facilitate comparison. The FWHM of the short-axis size of the focused spot is approximately 2.016 μm, whereas the spot through the "glass/Sb_2Te_3 (20 nm)" sample is reduced to approximately 1.632 μm. The results indicate that the super-resolution spot can be produced by the Sb_2Te_3 thin film. A theoretical calculation of the super-resolution spot is shown in Fig. 6.6d. The calculation, which aims to verify the analytical model of formula (6.7), is performed by setting the experimental parameters of $\lambda = 650\,\text{nm}$, $L = 20\,\text{nm}$, and $P = 8\,\text{mW}$. $\beta = -2.5 \times 10^{-2}\,\text{m/W}$, following a related study, w_0 is set to be 1.75 μm based on the data from Fig. 6.6c. The radius size w_0 is at the position of $1/e^2$ intensity of the spot through the glass substrate. The focused spot size is approximately 2.01 μm, and the super-resolution spot has the theoretical value of 1.60 μm. Comparison with Fig. 6.6c shows that the theoretical model from formula (6.7) is reasonable for thin (or weak) nonlinear saturation absorption sample.

6.3 Multi-layer Model for Thick (or Strong) Nonlinear Saturation Absorption Samples

In nonlinear super-resolution effect, for thick (or strong) nonlinear saturation absorption samples, the absorption coefficient is not only function of radial coordinate r, but also vertical coordinate z (along sample thickness direction). The light intensity decays exponentially along sample thickness direction; the absorption coefficient should be marked as $\alpha(r, z)$. The formation of the super-resolution spot can no longer be roughly analyzed using the B–L model; more rigorous method needs to be developed. Here, the multi-layer analytical model is introduced by considering the absorption coefficient change along sample thickness direction.

6.3.1 Multi-layer Analytical Model for Formation of Pinhole Channel

In the process of super-resolution spot formation, the optical pinhole channel is the core. To determine the morphology of the optical pinhole channel, the sample with a thickness L is dealt into a multi-layer system (e.g., m layers), and the thickness of every layer is ΔL, $m * \Delta L = L$. The light intensity and absorption coefficient of ith layer are marked as I_i and α_i, where $i = 1, 2, 3, \ldots, m$, respectively. The light is focused on the sample surface. On the nonlinear thin film surface, $z = 0$, $w(z = 0) = w_0$, and $I_{\text{inc}} = I_0 \exp\left[-2r^2/w_0^2\right]$. The sample is considered to be a multi-layer system, as shown in Fig. 6.7. The light goes through the 1st layer and decays into the second layer due to the absorption effect. As the m is chosen to be sufficiently large, the B–L model is suitable for each layer.

The 1st layer can be considered to be a sample surface, $z = z_1 = 1\Delta L$

$$\alpha_1 = \alpha_0 + \beta I_{\text{inc}}, \quad I_1 = I_{\text{inc}} e^{-\alpha_1 \Delta L} \tag{6.8}$$

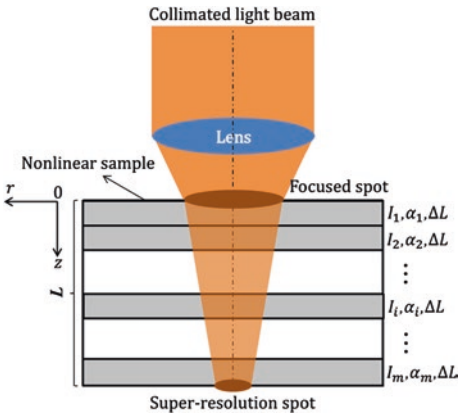

Fig. 6.7 Schematic of multi-layer analytical model for super-resolution spot formation

At the 2nd layer, $z = z_2 = 2\Delta L$

$$\alpha_2 = \alpha_0 + \beta I_1, \quad I_2 = I_1 e^{-\alpha_2 \Delta L}, \tag{6.9}$$

Similarly, at the ith layer, $z = z_i = i\Delta L$,

$$\alpha_i = \alpha_0 + \beta I_{i-1}, \quad I_i = I_{i-1} e^{-\alpha_i \Delta L}, \tag{6.10}$$

Finally, at the mth layer, $z = z_m = m\Delta L$,

$$\alpha_m = \alpha_0 + \beta I_{m-1}, \quad I_m = I_{m-1} e^{-\alpha_m \Delta L} \tag{6.11}$$

The light intensity profile of the super-resolution spot can actually be approximated to the light field of the last layer (mth layer) and can be simplified as

$$I_{\text{super-resolution spot}} = I_m = I_{\text{inc}} \exp\left(-\sum_{i=1}^{m} \alpha_i \Delta L \right) \tag{6.12}$$

When m goes to infinity, theoretically accurate integral can be obtained, and the formula (6.12) can be written as

$$I_{\text{super-resolution spot}} = I_{\text{inc}} \exp\left[-\int_0^L \alpha(r, z) \mathrm{d}z \right] \tag{6.13}$$

Formula (6.13) is the expected expression for analyzing the pinhole channel and light field characteristics of the super-resolution spot.

6.3.2 Super-Resolution Effect Analysis Using Multi-layer Model

Based on the above theoretical analysis, numerical simulations are conducted to verify the multi-layer model (M-L model). In order to intuitively compare with B–L model, the crystalline Sb_2Te_3 sample is used as an example in simulation because of strong

nonlinear saturation absorption; the sample thickness is chosen as $L = 150\,\text{nm}$. The other parameters are as follows: $\alpha_0 = 5.32 \times 10^7\,\text{m}^{-1}$, $\beta = -6.63 \times 10^{-2}\,\text{m/W}$ at $\lambda = 405\,\text{nm}$ [8], $w_0 = 1{,}000\,\text{nm}$, and $I_0 = 8 \times 10^8\,\text{W/m}^2$.

6.3.2.1 Formation of Optical Pinhole Channel Inside Nonlinear Samples

The optical pinhole channel can be essentially reflected by the absorption coefficient profile. That is, the essence of the optical pinhole channel is due to the small absorption coefficient compared with other regions. Figure 6.8 gives the absorption coefficient distribution [9, 10]. Figure 6.8a is obtained by B–L model, where α is only a function of the radial coordinate r. The optical pinhole channel with ~1 μm diameter is produced and does not change with sample thickness position, which is because the absorption coefficient is considered to remain unchanged along sample direction for B–L model.

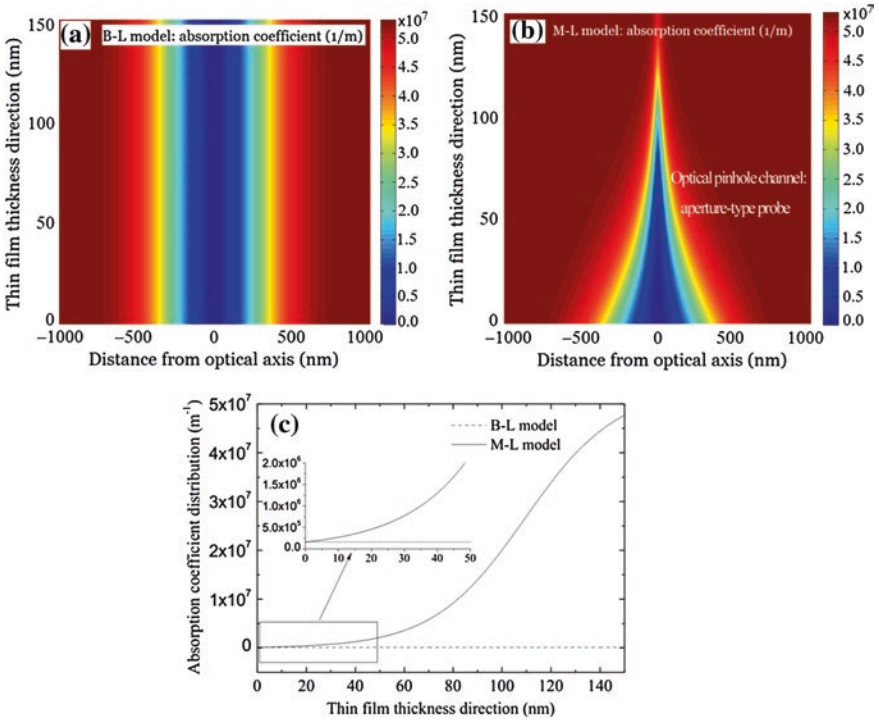

Fig. 6.8 Absorption coefficient distribution inside nonlinear sample, the three-dimensional profile along r–z plane for **a** B–L model and **b** M-L model, and **c** variation of $\alpha(0, z)$ along thickness direction

Figure 6.8b is the three-dimensional distribution of the absorption coefficient at r–z plane for M-L model. The $\alpha(r, z)$ varies with radial coordinates r and sample thickness position z and has the smallest value at the center of the incident plane ($z = 0$ position). It is obvious that an optical pinhole channel, looking like an aperture-type optical probe, is generated at the central region along the sample thickness direction. The pinhole channel size decreases with the sample thickness increasing, and a very tiny pinhole is formed at the apex of optical channel.

Figure 6.8c shows the variation of $\alpha(0, z)$; in order to give a direct comparison, the $\alpha(0, z)$ calculated by the B–L model is also presented. For B–L model, the $\alpha(0, z)$ remains unchanged with the variation of thin film thickness. For the M-L model, $\alpha(0, z)$ is almost unchanged for small sample thickness at $L < 20\,\text{nm}$. At $L > 20\,\text{nm}$, $\alpha(0, z)$ exponentially increases with the thickness position. Comparison of B–L model and M-L model shows that at $L < 20\,\text{nm}$, the $\alpha(0, z)$ of M-L model is basically consistent with the B–L model. In other words, the B–L model can be used for thin sample with $L < 20\,\text{nm}$ under this simulation.

This difference between B–L and M-L models is due to the fact that light intensity attenuation inside the nonlinear thin film is considered in the M-L model; thus, the absorption coefficient should increase with sample thickness increasing, accordingly. That is, the absorption coefficient is no longer a fixed value for different thickness position, but changes with thickness position.

6.3.2.2 Light Intensity Distribution Inside the Nonlinear Saturation Absorption Samples

While light is traveling through the optical pinhole channel, the light intensity changes inside the sample, which finally forms the super-resolution spot. Figure 6.9 shows the light intensity distribution $I(r, z)$ [9, 10]. In Fig. 6.9a, the intensity obtained by B–L model slowly decreases along the thin film thickness direction and slightly becomes narrow in radial direction. This is because in B–L model, the absorption coefficient $\alpha(r, z)$ is calculated using a fixed light intensity I_{inc} for different sample thickness position and the light intensity attenuation along the sample thickness direction is ignored; it varies only radially; more intensity is absorbed on the edge of the Gaussian beam, while little attenuation in the central region; thus, the laser beam becomes narrow slightly. However, in Fig. 6.9b, mutual influence between intensity and absorption coefficient is considered by the M-L model; the light intensity attenuation results in the absorption coefficient increasing along sample thickness direction.

Figure 6.9c shows the variation of central intensity ($r = 0$) along the sample thickness direction. At $L < 20\,\text{nm}$, the light intensity change in the B–L model is almost the same as that in the M-L model. However, at $L > 20\,\text{nm}$, the difference between the B–L model and M-L model increases dramatically with increased sample thickness. At thickness position $z = 150\,\text{nm}$, the central intensity decreases to only 10 % of the incident light intensity in M-L model, while it almost remains unchanged in B–L model.

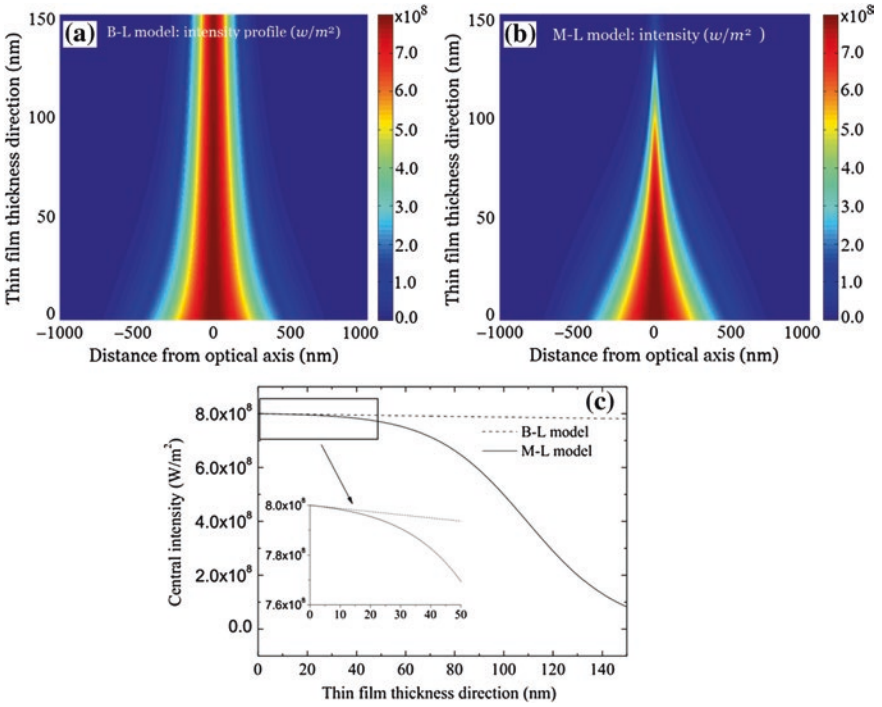

Fig. 6.9 The light internal intensity distribution for **a** B–L model and **b** M-L model, and **c** central intensity variation along the sample thickness direction at $r = 0$

From the analysis above, one can see that B–L model is consistent with M-L model only at the $L < 20\,\text{nm}$. The divergence increases with the thickness position for the strong nonlinear absorption sample. Thus, the B–L model approximation is no longer suitable, and the M-L model can give a more accurate theoretical simulation for the thick (or strong) nonlinear saturation absorption samples.

6.3.2.3 Light Beam Size Variation Inside the Nonlinear Saturation Samples

The light travels through optical pinhole channel and can be restricted into a very small beam. Figure 6.10 shows the variation of light beam size along the thickness direction inside nonlinear sample [9, 10]. In the B–L model, the light beam radius change is marked in the dotted curve, whereas the solid curve depicts the beam radius variation in the M-L model. The light beam radius exponentially decreases along the sample thickness direction. At $L < 20\,\text{nm}$, the calculated results with B–L model is almost the same as the M-L model. At $L > 20\,\text{nm}$, the variation in the M-L model is larger than that in the B–L model, and large difference occurs with increasing sample thickness. For example, at $z = 40\,\text{nm}$, the laser spot

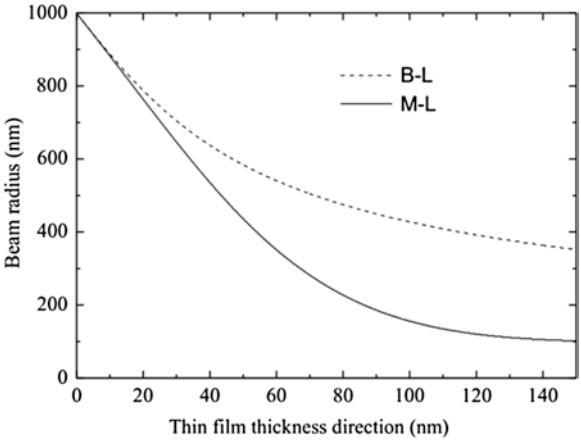

Fig. 6.10 Light beam radius variation along the sample thickness direction

radius is 640 nm in B–L model, but only 530 nm in M-L model; the difference is 110 nm. At $z = 150\,\text{nm}$, the beam radius is about 100 nm in M-L model, while in B–L model, the beam radius is about 350 nm, and the difference is 250 nm. Considering the mutual influence between light intensity and the absorption coefficient, the theoretical light beam size in M-L model is more reliable.

6.3.2.4 Super-Resolution Spot Formation at the Exiting Surface of Nonlinear Sample

The optical pinhole channel restricts the light into very small beam, and the super-resolution spot can be generated at the back surface of nonlinear sample. Figure 6.11 presents the characteristics of super-resolution spot at different sample thickness position [9, 10]. At $L < 20\,\text{nm}$, the super-resolution spot intensity and size calculated by the B–L model are consistent with those obtained by the M-L model, as shown in Fig. 6.11a. However, with increased thin film thickness, the differences (including intensity and size) between the B–L model and M-L model become obvious, as shown in Fig. 6.11b–d. At $L = 150\,\text{nm}$, the central intensity of the super-resolution spot in M-L model is only 10 % that in the B–L model (also see Fig. 6.11d), which also indicates that the B–L model is no longer suitable for analyzing the characteristics of the super-resolution spot for $L > 20\,\text{nm}$.

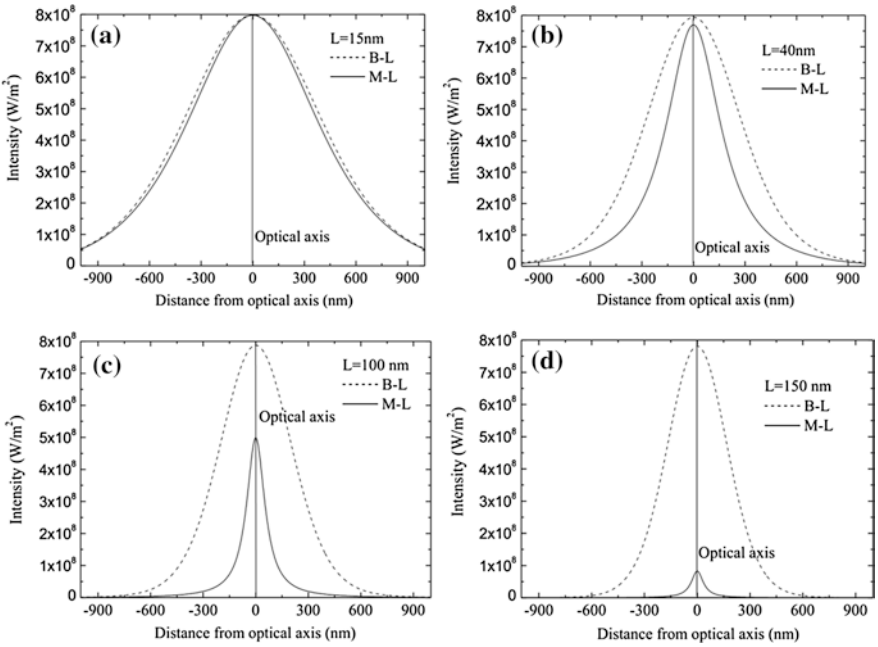

Fig. 6.11 Characteristics of super-resolution spot, **a** $L = 15$ nm, **b** $L = 40$ nm, **c** $L = 100$ nm, and **d** $L = 150$ nm

The super-resolution spots from the exiting surface of sample are shown in Fig. 6.12. Figure 6.12a is the incident spot intensity profile, and the diameter is about 2 μm. If the super-resolution spot is calculated using B–L model, the spot diameter is about 1.2 μm as shown in Fig. 6.12b. However, the super-resolution spot is actually much smaller than the size obtained by B–L model because in B–L model the absorption coefficient change along sample thickness direction is not considered. Figure 6.12c gives the super-resolution spot intensity profile calculated through M-L model where the absorption coefficient attenuation along thickness direction is considered. One can see that the super-resolution spot is about 300 nm, and obviously smaller than the B–L model. The M-L model is more accurate for analyzing the super-resolution of nonlinear saturation absorption samples.

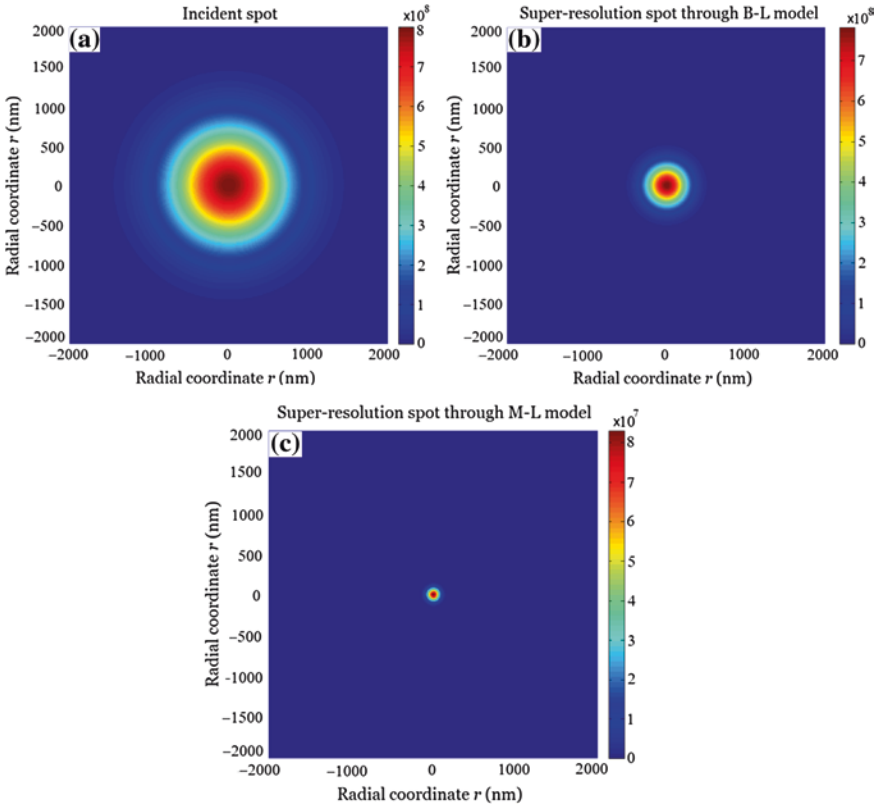

Fig. 6.12 Spot intensity profiles, **a** incident spot, obtained by **b** B–L model, and **c** M-L model

6.4 Summary

In nonlinear saturation absorption-induced super-resolution effect, for thin (or weak) nonlinear saturation absorption samples, the absorption coefficient can be considered to be almost unchanged along sample thickness direction. If the interference effect between front and back surfaces of sample can be neglected, the super-resolution effect can be analyzed using B–L model. However, for thick (or strong) nonlinear saturation absorption samples, the light intensity decays exponentially along sample thickness direction, and the absorption coefficient should be marked as $\alpha(r, z)$. The formation of the super-resolution spot needs to be analyzed with the M-L analytical model. For real application, the super-resolution spot can be directly coupled into the materials in the near-field range to conduct high-density data storage, high-resolution non-fluorescence labeling imaging, and nanolithography, etc.

References

1. M. Kuwahara, T. Nakano, J. Tominaga, M.B. Lee, N. Atoda, A new lithography technique using super-resolution near-field structure. Microelectron. Eng. **53**(1–4), 535–538 (2000)
2. J. Tominaga, T. Nakano, N. Atoda, An approach for recording and readout beyond the diffraction limit with an Sb thin film. Appl. Phys. Lett. **73**(15), 2078–2080 (1998)
3. J. Liu, J. Wei, Optical nonlinear absorption characteristics of AgInSbTe phase change thin films. J. Appl. Phys. **106**, 083112 (2009)
4. J. Wei, J. Liu, Direct observation of below-diffraction-limited optical spot induced by nonlinear saturation absorption of Ag-doped Si nanofilms. Opt. Lett. **35**, 3126–3128 (2010)
5. J. Wei, J. Liu, M. Xiao, Giant optical nonlinearity of silver-doped silicon thin film at low power input: laser triggered cluster resonance. Appl. Phys. A **104**, 1031–1037 (2011)
6. J. Liu, S. Liu, J. Wei, Origin of the giant optical nonlinearity of Sb_2Te_3 phase change materials. Appl. Phys. Lett. **97**, 261903 (2010)
7. J. Wei, On the dynamic readout characteristic of nonlinear super-resolution optical storage. Opt. Commun. **291**, 143–149 (2013)
8. S. Liu, J. Wei, F. Gan, Nonlinear absorption of Sb-based phase change materials due to the weakening of the resonant bond. Appl. Phys. Lett. **100**, 111903 (2012)
9. H. Yan, J. Wei, Super-resolution spot formation for strong nonlinear saturation absorption sample: a multi-layer analytical model (unpublished)
10. H. Yan, Optical super-resolution and imaging based on strong nonlinear film, A Dissertation of University of Chinese Academy of Sciences, 2014

Chapter 7
Resolving Improvement by Combination of Pupil Filters and Nonlinear Thin Films

7.1 Introduction

A number of studies have focused on optical far-field super-resolution spot using pupil filters. Pupil filters with diffractive optical elements have attracted much attention due to simple and relatively easy fabrication processes under current available techniques. Since Toraldo di Francia proposed the concept of super-resolving pupil filters [1], a number of pupil filters including amplitude-only, phase-only, and hybrid ones have been proposed [2–4]. Phase-only pupil filters offer a significant advantage over amplitude-only pupil filters in that they deliver most portion of incident light intensity to the sample. So far, some issues remain for these super-resolving pupil filters. These issues include the fact that the side lobe becomes non-negligible when the main lobe is squeezed to certain extent while being used in optical storage and confocal scanning microscopy. The second issue is that the reduction of the super-resolving spot to nanoscale is difficult. In this chapter, the design methods and examples on the pupil filters for different polarized light are given, and then some resolving improvement techniques are presented by combination with nonlinear effect of thin films.

7.2 Super-Resolution with Pupil Filters

7.2.1 Binary Optical Elements as Pupil Filters: Linearly Polarized Light Illumination

In real applications, the optical elements are required to be low cost, easy fabrication, and simple operation. The binary pupil filter is one of the simple optical elements. In addition, the linearly polarized laser beam can be directly obtained from laser

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_7

devices [5, 6]. Here, a design of linearly polarized light modulated by annular binary pupil filter is used to obtain super-resolution spot.

7.2.1.1 Three-Zone Annular Binary Phase-Type Pupil Filters

A three-zone annular binary phase-type pupil filter is used as an example to obtain a super-resolving optical spot due to a large light transmittance at the center of the focal plane. In this design schematic, the GaN semiconductor laser device with a wavelength of 405 nm is used as the incident light source. This is because on one hand, the GaN semiconductor laser device is cheap and easy to obtain. On the other hand, the wavelength of 405 nm is close to the transmission cutoff frequency of common optical elements. The lens with a numerical aperture (NA) of 0.95 is used to focus the light into the diffraction-limited spot. The NA of focusing lens is also close to the theoretical limit of the far-field optical lens system, which is easy to operate in practical application. The designed three-zone annular binary phase pupil filter is placed in front of the focusing lens (shown in Fig. 7.1). Figure 7.1a describes that the incident light emitted from the GaN semiconductor laser device passes through three-zone annular binary phase filter and is focused by the lens with a NA of 0.95 into a super-resolving spot. The super-resolving spot consists of the main and side lobes. Generally, the intensity of main lobe is several times that of the side lobe. Another important aspect is that the depth of focus (DOF) of the super-resolving spot is elongated by optimizing the design of three-zone annular binary phase pupil filter. Figure 7.1b displays the front view of the three-zone annular binary phase pupil filter, where the radius is defined as r_1, r_2, and r_3, and the phase is set to be $\varphi_1 = 0$, $\varphi_2 = \pi$, and $\varphi_3 = 0$. The transmission function of pupil filter is as follows.

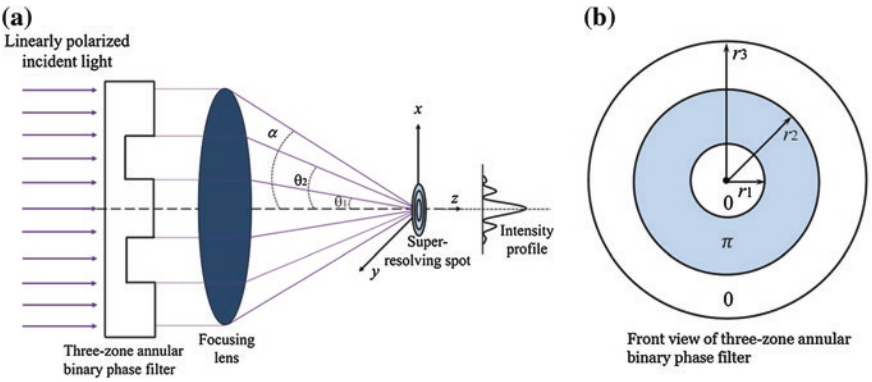

Fig. 7.1 Design schematics of the three-zone annular binary phase pupil filter. **a** Incident light passes through three-zone annular binary phase filter and is focused by high-NA lens. **b** The front view of the three-zone annular binary phase filter. Reprinted from [7], Copyright 2013, with permission from Elsevier

$$t_1(\theta) = \begin{cases} \cos\varphi_1 = +1 & 0 \le \theta < \theta_1 \\ \cos\varphi_2 = -1 & \theta_1 \le \theta < \theta_2 \\ \cos\varphi_3 = +1 & \theta_2 \le \theta < \alpha \end{cases} \tag{7.1}$$

For linearly polarized light illumination, according to the formulas (7.69) and (7.70) in section, the electric field components of any point P on image space can be calculated as

$$\begin{cases} \mathrm{e}_x(P) = -ic[I_0 + I_2\cos(2\phi_P)] \\ \mathrm{e}_y(P) = -icI_2\sin(2\phi_P) \\ \mathrm{e}_z(P) = -2cI_1\cos\phi_P \end{cases} \tag{7.2}$$

with

$$\begin{cases} I_0 = I_0(kr_P,\ \theta_P,\ \psi) \\ \quad = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(1+\cos\theta)\left[J_0(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \\ I_1 = I_1(kr_P,\ \theta_P,\ \psi) \\ \quad = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(\sin\theta)\left[J_1(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \\ I_2 = I_2(kr_P,\ \theta_P,\ \psi) \\ \quad = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(1-\cos\theta)\left[J_2(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \end{cases} \tag{7.3}$$

For linearly polarized light illumination, formulas (7.2) and (7.3) give the electric field of any point $P(r_P, \theta_P, \phi_P)$ on image space. For three-zone annular binary phase filter system, the electric field components of any point P on image space can also be calculated by formula (7.2), only if the I_0, I_1, and I_2 are rewritten as follows accordingly

$$\begin{aligned} I_0 = I_0(kr_P,\theta_P,\psi) &= \int_0^{\psi} t(\theta)\left(\sqrt{\cos\theta}\sin\theta\right)(1+\cos\theta) \\ &\left[J_0(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \\ I_1 = I_1(kr_P,\theta_P,\psi) &= \int_0^{\psi} t(\theta)\left(\sqrt{\cos\theta}\sin\theta\right)(\sin\theta) \\ &\left[J_1(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \\ I_2 = I_2(kr_P,\theta_P,\psi) &= \int_0^{\psi} t(\theta)\left(\sqrt{\cos\theta}\sin\theta\right)(1-\cos\theta) \\ &\left[J_2(kr_P\sin\theta\sin\theta_P)\mathrm{e}^{ikr_P\cos\theta\cos\theta_P}\right]\mathrm{d}\theta \end{aligned} \tag{7.4}$$

where $t(\theta)$ is the transmission of the pupil filter and can be replaced by $t_1(\theta)$ of formula (7.1). The intensity at point $P(x, y, z)$ can be calculated as (" $*$ " denotes complex conjugate)

$$I = \mathrm{e}_x \mathrm{e}_x^* + \mathrm{e}_y \mathrm{e}_y^* + \mathrm{e}_z \mathrm{e}_z^* \tag{7.5}$$

7.2.1.2 Optical Field Distribution of Three-Zone Annular Binary Phase Filters

Figure 7.2 presents the simulated results. Figure 7.2a shows a diffraction-limited spot in the focal plane. The spot is an ellipse that is longer in the x-axis direction than in the y-axis direction. Figure 7.2b presents the super-resolving spot squeezed in the y-axis direction. An evident side lobe occurs both at x-axis and y-axis directions. The intensity of the side lobe along the y-axis direction is stronger than that along the x-axis direction.

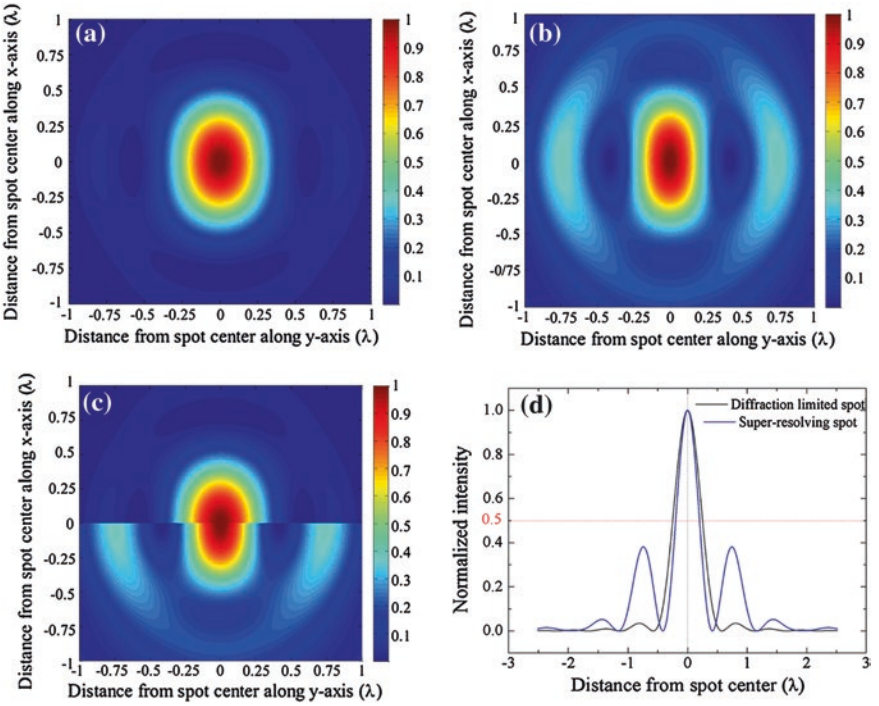

Fig. 7.2 Cross-sectional intensity distribution of optical spot at focal plane (x–y plane). **a** Diffraction-limited. **b** Super-resolving. **c** comparison of super-resolving spot (*lower half*) and diffraction-limited spot (*upper half*). **d** Comparison of y-axis normalized intensity distribution (*black* diffraction-limited spot, *blue* super-resolving spot). Reprinted from [7], Copyright 2013, with permission from Elsevier

To obtain an intuitive judgment of the super-resolving effect, the super-resolving spot is compared with the diffraction-limited spot, as shown in Fig. 7.2c, where the upper half is the diffraction-limited spot and the lower half is the super-resolving spot. The size of the main lobe of the super-resolving spot is evidently smaller than that of the diffraction-limited spot at the *y*-axis direction. A strong side lobe occurs, especially at the *y*-axis direction.

Figure 7.2d shows the comparison of the normalized intensity of the super-resolving spot and that of the diffraction-limited spot along the *y*-axis. The FWHM of the diffraction-limited spot and the main lobe of the super-resolving spot are approximately 204 and 163 nm, respectively. The size of the main lobe is reduced to approximately 80 % of the diffraction-limited spot. The three-zone annular pupil filter causes the size of the FWHM to decrease, but the reduction proportion is not sufficient in real applications. Meanwhile, the intensity of the side lobe is more than 38 % of the main lobe intensity. Such strong side lobe is not effective for practical applications and should be further reduced or eliminated.

The DOF can be observed along the *z*-axis direction in the *y*–*z* meridian plane. Figure 7.3 gives the simulated results. In Fig. 7.3a, the diffraction-limited spot in

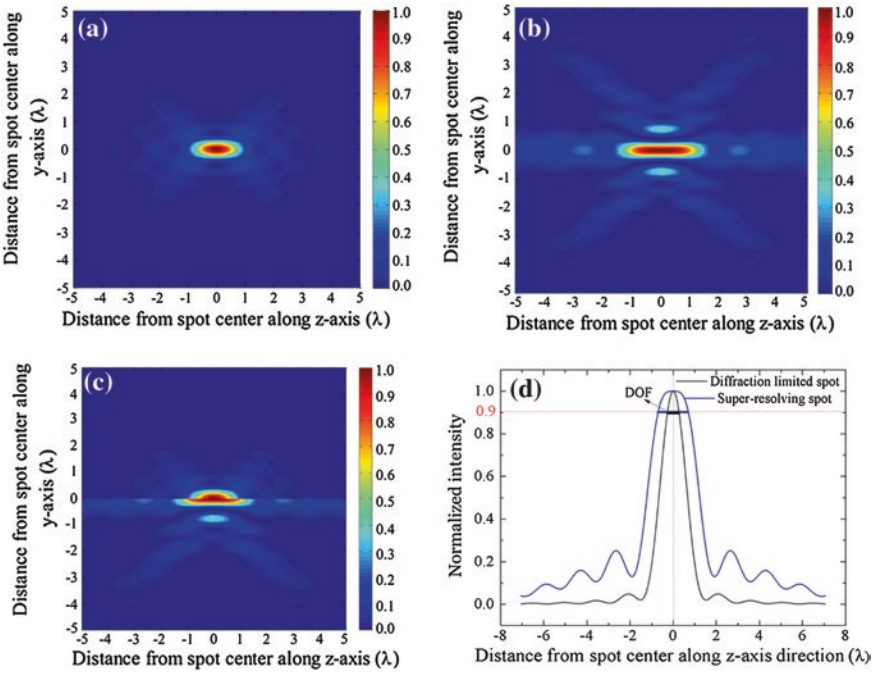

Fig. 7.3 Depth of focus along *y*–*z* meridian plane. **a** Diffraction-limited spot. **b** Super-resolving spot. **c** DOF comparison of super-resolving spot (*lower half*) and diffraction-limited spot (*upper half*). **d** Comparison of DOF between diffraction-limited and super-resolving spots along *z*-axis direction (*black* diffraction-limited spot, *blue* super-resolving spot). Reprinted from [7], Copyright 2013, with permission from Elsevier

the y–z meridian plane is also an ellipse, and the DOF can be observed along the z-axis direction. In Fig. 7.3b, the super-resolving spot is an elongated ellipse in the y–z meridian plane, that is, it is elongated along the z-axis direction. The DOF increases accordingly. In addition, the side lobe of the super-resolving spot can also be observed clearly.

7.2.2 *Ternary Optical Elements as Pupil Filters: Radially or Circularly Polarized Light Illumination*

Far-field super-resolution sometimes requires that both the spot size is below-diffraction-limited and the spot shape is circular. For lots of applications, such as lithography, optical data storage, the DOF needs to be extended. The radially polarized Bessel–Gauss (BG) beam and circularly polarized light are good candidate illumination modes due to excellent spot intensity uniformity. The ternary optical element with central block shows good performance in extending the DOF. Thus, the combination of radially polarized BG beam or circularly polarized light illumination mode and ternary optical element with central block is an alternative for realizing the far-field super-resolution.

Figure 7.4 presents a radially polarized BG beam or circularly polarized light being incident on a ternary optical element, as shown in Fig. 7.4a. The ternary optical element is presented in Fig. 7.4b and consists of one zero transmission belt at the central area (marked in grayscale) and four belts with transmissions of -1 and $+1$ alternately in the outer region. The yellow color represents transmission of -1, and the white color represents the transmission of $+1$.

In the focal region of the lens, a super-resolution focal spot with a uniform size and an ultralong DOF is obtained. The angle $\theta_i (i = 1, 2, 3, 4)$ corresponds to the aperture half-angle of the ith belt and is computed as $\theta_i = \sin^{-1}(r_i \mathrm{NA})$, where r_i is the normalized radius of the ith belt. The maximum aperture half-angle is $\psi = \sin^{-1}(\mathrm{NA}/n)$, where NA is the numerical aperture of focusing lens and n is the refractive index of medium, for air $n = 1$. The transmission function of the ternary optical element is

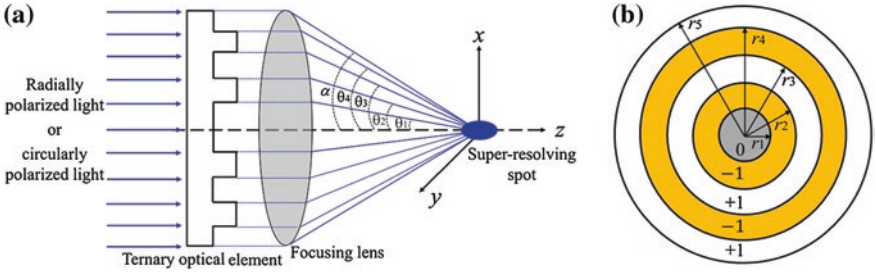

Fig. 7.4 Radially or circularly polarized light illumination for the ternary optical element. **a** Diagram of the light beam propagation. **b** Configuration of the ternary optical element

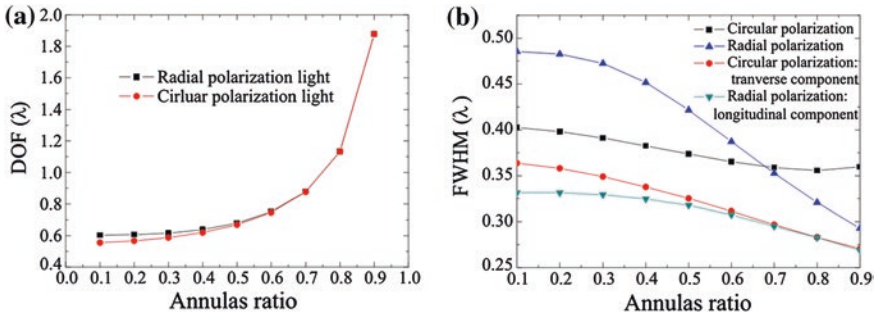

Fig. 7.5 Dependence of the (**a**) DOF and (**b**) FWHM on the annulus ratio for the circular and radial polarization [8]

$$t_2(\theta) = \begin{cases} 0 & \text{for } 0 \le \theta < \theta_1 \\ -1 & \text{for } \theta_1 \le \theta < \theta_2, \theta_3 \le \theta < \theta_4 \\ +1 & \text{for } \theta_2 \le \theta < \theta_3, \theta_4 \le \theta < \alpha \end{cases} \tag{7.6}$$

The central block area makes the incident beam become annular beam, which means that the portion of high-frequency band increases, the focused spot becomes small, and the DOF increases. The four belts in the outer region are a compromise between the longitudinal field strength and the optical efficiency. The radius of the ith zone is normalized to the outermost pupil radius of the five-zone annular ternary optical element, and $r_i = \sin\theta_i / \sin\psi$.

In order to obtain small spot, one can use an oil lens of NA = 1.4 and refractive index of oil $n = 1.5$, which is easy to obtain in industry. The dependences of the FWHM and DOF on the annulus ratio need to be first determined, where the annulus ratio is defined as the inner radius divided by the outer radius, Fig. 7.5 gives the results. Figure 7.5a shows that the DOF for the circularly polarized light is almost the same as that for the radially polarized light when the annulus ratio is over 0.6. Figure 7.5b shows that the FWHM of the radial polarization is larger than the circular polarization when the annulus ratio is less than 0.7, and it is smaller when the annulus ratio is more than 0.7. The FWHM of the longitudinal component of the radial polarization is smaller than the transverse component of the circular polarization when the annulus ratio is below 0.7, and continuously increasing the annulus ratio results in almost the same value of the two components. Thus, comparison with the radially polarized illumination indicates that, in practical applications, the circularly polarized illumination is also a good candidate for obtaining a small spot at an annulus ratio of over 0.7 if the materials are only sensitive to the transverse component of the circular polarization light.

7.2.2.1 Circularly Polarized Light Illumination

The circularly polarized light can form uniform and circular focused spot under converging lens system and has been used in the optical data storage. Further,

reducing the spot size and extending the DOF are useful for nanotechnology, and modulating the circularly polarized light with central block ternary optical element is a feasible way.

1. *Design of ternary optical elements for circularly polarized light illumination*

Based on Fig. 7.5, the proposed system configuration is shown in Fig. 7.4. A circularly polarized beam passes through the ternary optical element and is then focused by an oil lens with an NA of 1.4 and refractive index of oil $n_{\text{imsp}} = 1.5$. For circularly polarized light illumination, similar to the linearly polarized light, the electric field of any point P on the image space can be written as [8]

$$\begin{cases} \mathrm{e}_x = -ic\{I_0 + I_2[\cos(2\phi_P) + i\sin(2\phi_P)]\} \\ \mathrm{e}_y = -ic\{iI_0 - iI_2[\cos(2\phi_P) + i\sin(2\phi_P)]\} \\ \mathrm{e}_z = -2cI_1(\cos\phi_P + i\sin\phi_P) \end{cases} \tag{7.7}$$

where ϕ_P is the azimuthal angle of point P. The maximum aperture half-angle is $\psi = \arcsin(\text{NA}/n_{\text{imsp}})$. I_0, I_1, and I_2 are integrals evaluated over the aperture half-angle ψ and can be written as

$$\begin{cases} I_0 = \int_0^{\psi} t(\theta)\sqrt{\cos\theta}\sin\theta(1+\cos\theta)J_0(kr\sin\theta)\exp(ikz\cos\theta)\mathrm{d}\theta \\ I_1 = \int_0^{\psi} t(\theta)\sqrt{\cos\theta}\sin^2\theta J_1(kr\sin\theta)\exp(ikz\cos\theta)\mathrm{d}\theta \\ I_2 = \int_0^{\psi} t(\theta)\sqrt{\cos\theta}\sin\theta(1-\cos\theta)J_2(kr\sin\theta)\exp(ikz\cos\theta)\mathrm{d}\theta \end{cases} \tag{7.8}$$

where $t(\theta)$ is replaced by $t_2(\theta)$ of formula (7.6). J_0, J_1, and J_2 are the zero-order, first-order, and second-order Bessel functions of the first kind, respectively.

The radius of ternary optical elements can be designed and calculated through formulas (7.7) and (7.8). In order to save the computing time, one can take two-step method computing strategy, which is also suitable for designing other pupil filters. Firstly, by using exhaustive searching algorithm, one can obtain a certain number of potential solutions under the confinement of uniformity of the beam along the z-axis in the focal region and the FWHM of the focal spot, where one restricts them to certain range. The combination of the two confinements allows a fast searching speed. The uniformity along the z-axis is defined as the ratio of the intensity difference to the summation between the maximum and minimum intensities in the certain range along z-axis. Secondly, the FWHM is sorted from small to large and some favorable results are selected. In order to obtain an ultralong and uniform subwavelength light beam, the two decimal digits of the tolerance are extended to four decimal digits manually according to experience, while it does not take too much time. The most important feature of the computing strategy is time-saving.

2. *Super-resolution non-diffraction beam*

One of the set of optimized solutions is $r_1 = 0.8721$, $r_2 = 0.8869$, $r_3 = 0.9117$, $r_4 = 0.9864$, and $r_5 = 1$. Correspondingly, the electric energy density distributions on the y–z plane are shown in Fig. 7.6.

For comparison, the electric energy density distribution of the clear aperture is shown in Fig. 7.6a. Figure 7.6b shows that the DOF is elongated by the ternary optical element. The DOF modulated by the ternary optical element is approximately 8.283 λ, which is approximately 15-fold larger than that generated by the clear aperture focusing. Compared with the clear aperture case shown in Fig. 7.6a where the original beam has an FWHM of 0.4043 λ and DOF of 0.552 λ, the light beam modulated by the ternary optical element in the optimized design has an FWHM of 0.3562 λ and DOF of 8.283 λ, which means that the beam propagates without divergence over 8.283 λ. Thus, the beam is a highly localized super-resolution non-diffraction beam.

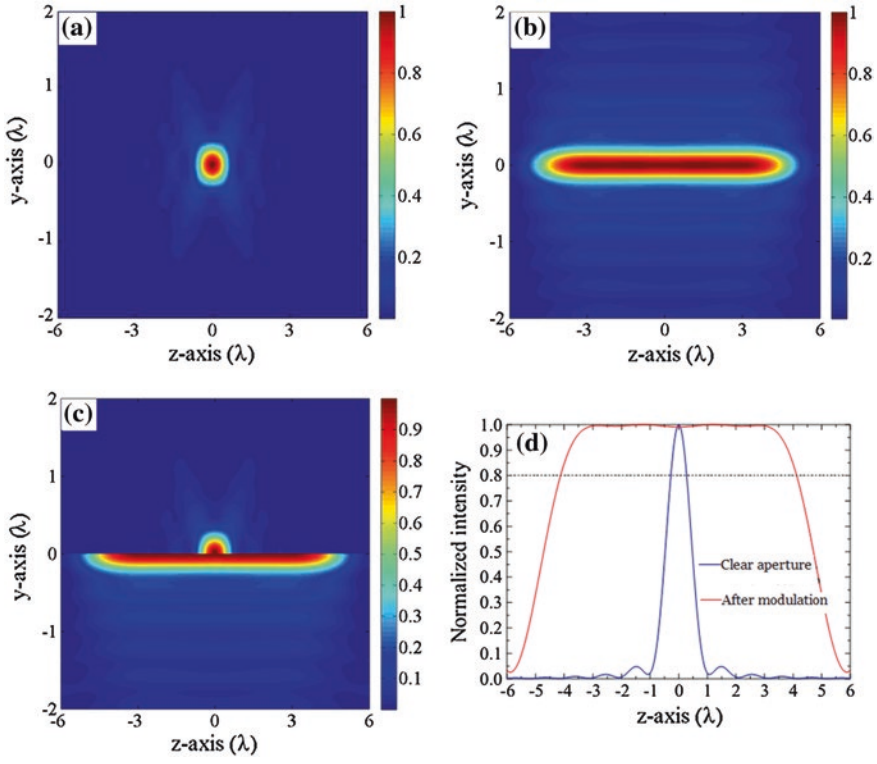

Fig. 7.6 Electric energy density distributions on the y–z plane. **a** Clear aperture. **b** Modulated by the ternary optical element. **c** Comparison between the upper half of **a** and lower half of **b**. **d** Electric energy density distribution along the z-axis of the clear aperture and that modulated by the ternary optical element [8]

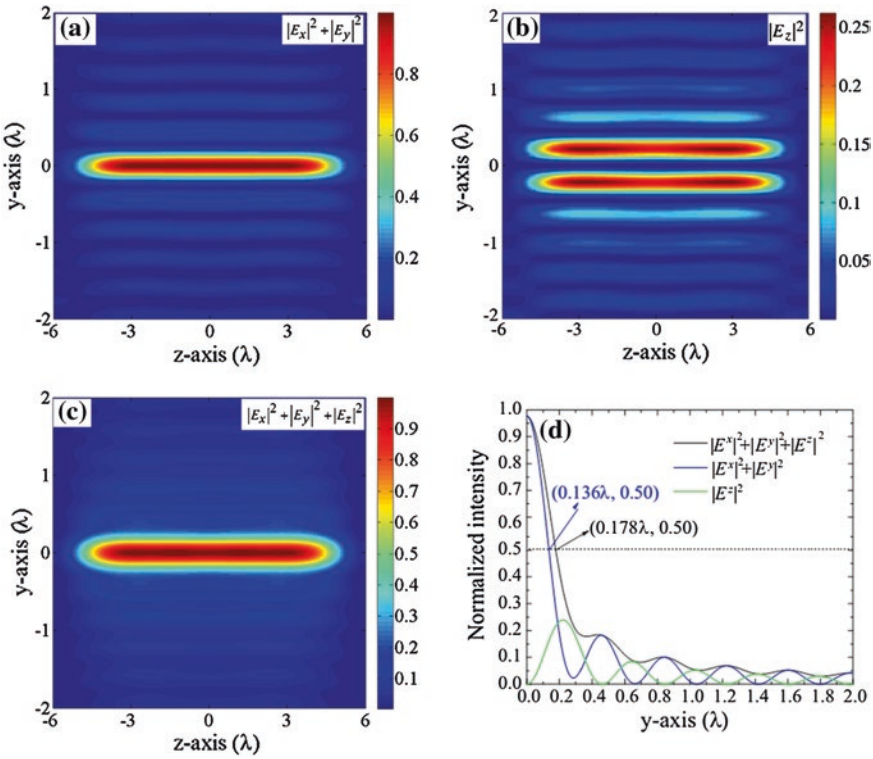

Fig. 7.7 Electric energy density distributions on the *y*–*z* plane for the ternary optical element. **a** Transverse. **b** Longitudinal components. **c** Total electric energy density. **d** Electric energy density component profile along the *y*-axis [8]

To further understand the optical characteristics of the super-resolution non-diffraction beam modulated by the ternary optical element, Fig. 7.7 shows the radial, longitudinal, and total electric energy density distributions on the *y*–*z* plane. The transverse component (shown in Fig. 7.7a) occupies a dominant portion of the total electric energy (shown in Fig. 7.7c) and is localized. The longitudinal component (shown in Fig. 7.7b) is shaped into a donut, which makes the spot size in the focal plane become larger. Figure 7.7d shows that the super-resolution non-diffraction beam has an FWHM of 0.3562 λ, whereas that of the transverse component is 0.273 λ, which means that the longitudinal component increases the focal spot size, and the transverse component plays a dominant role in the spot size reduction inside the media. The spot size of the super-resolution non-diffraction beam inside the media can theoretically become 0.273 λ.

7.2.2.2 Radially Polarized Light Illumination [9, 10]

1. *Optical field calculation*

According to formulas (7.81) and (7.82), the effect of the ternary optical element on the focusing system can be calculated through replacing $l(\theta)$ with $l(\theta)t_2(\theta)$ in formula (7.81). Here, the formula (7.81) is rewritten as

$$\begin{cases} e_\rho^{(s)}(\rho_s, z_s) = c\int_0^\psi \sqrt{\cos\theta}\sin(2\theta)l(\theta)t_2(\theta)J_1(k\rho_s\sin\theta)\mathrm{e}^{ikz_s\cos\theta}\mathrm{d}\theta \\ e_z^{(s)}(\rho_s, z_s) = 2ic\int_0^\psi \sqrt{\cos\theta}\sin^2\theta l(\theta)t_2(\theta)J_0(k\rho_s\sin\theta)\mathrm{e}^{ikz_s\cos\theta}\mathrm{d}\theta \end{cases} \tag{7.9}$$

with

$$l(\theta)t_2(\theta) = \begin{cases} 0 & \text{for } 0 \le \theta < \theta_1 \\ -\exp\left[-\kappa^2\left(\frac{\sin\theta}{\sin\alpha}\right)^2\right]J_1\left(2\kappa\frac{\sin\theta}{\sin\alpha}\right) & \text{for } \theta_1 \le \theta < \theta_2, \theta_3 \le \theta < \theta_4 \\ \exp\left[-\kappa^2\left(\frac{\sin\theta}{\sin\alpha}\right)^2\right]J_1\left(2\kappa\frac{\sin\theta}{\sin\alpha}\right) & \text{for } \theta_2 \le \theta < \theta_3, \theta_4 \le \theta < \alpha \end{cases} \tag{7.10}$$

Formulas (7.9) and (7.10) can be used to analyze the optical field distribution of radially polarized light modulated by the central block ternary optical element.

2. *The $\boldsymbol{\kappa}$ value determination*

For radially polarized illumination, the electric field distribution in the vicinity of the focal region can be analyzed by formula (7.81). The $\boldsymbol{\kappa}$ value first needs to be determined. The dependences of DOF and FWHM on the κ in clear aperture focusing conditions are shown in Fig. 7.8. The FWHM increases, and the DOF

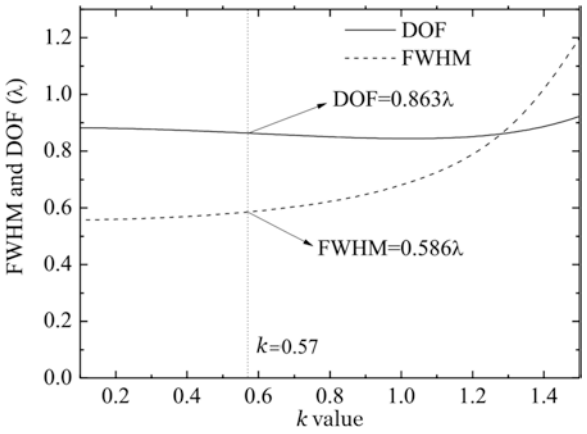

Fig. 7.8 Dependence of FWHM and DOF on κ in clear aperture focusing conditions

first decreases slightly and then increases as κ increases. Here, the DOF is defined as the two-point distance, where the intensity maximum $I_{\max}$ along the z-axis decreases to $0.8I_{\max}$. $\kappa = 0.57$ corresponds to the peak intensity of the radially polarization Bessel–Gauss beam at the rim of the focusing lens.

Figure 7.8 shows that the selection of $\kappa = 0.57$ is a reasonable compromise between optical efficiency and the resolution of the focused beam; further, decreasing the κ would increase the resolution a little. However, more light would be outside the aperture of the focusing lens.

3. *Super-resolution radially polarized beam*

The computing strategy is similar to circularly polarized light illumination. As an example, an optimized normalized radii are obtained as follows: $r_1 = 0.8875$, $r_2 = 0.8965$, $r_3 = 0.9191$, $r_4 = 0.9481$, and $r_5 = 1$. Considering the rotational symmetry, the electric energy density distributions of the radial, longitudinal, and total fields on the y–z plane are presented in Fig. 7.9. Figure 7.9a shows that the axial electric field energy density is very uniform. The beam has a size of $0.3995\,\lambda$ and it propagates without divergence over $12.83\,\lambda$ which is a highly localized

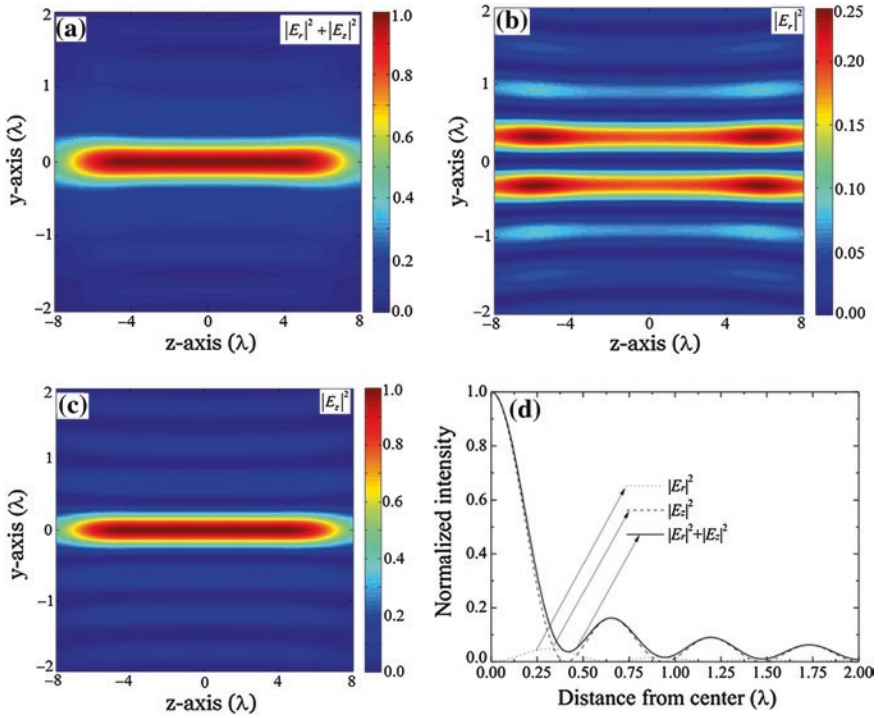

Fig. 7.9 Electric energy density distributions on the y–z-plane after phase modulation of ternary optical elements. **a** Total energy density distribution. **b** Radial component. **c** Longitudinal component. **d** The radial $|E_r|^2$, longitudinal $|E_z|^2$, and the total $|E_r|^2 + |E_z|^2$ electric energy density distribution on the focal plane (x–y plane)

super-resolution non-diffraction beam. The extended DOF is almost 15 times larger than that generated with clear aperture. Compared with the customarily used axial extent of the ordinary focus, which is defined as λ/NA^2 (for $\mathrm{NA} = 0.95$ it gives value $1.11\,\lambda$), the obtained DOF of $12.83\,\lambda$ is extended to about 11.56 times. The electric energy density profile of the transversal behavior of this beam on the focal plane is shown in Fig. 7.9d. The FWHM of the total electric density profile is $0.3995\,\lambda$, which is smaller than that of $0.58\,\lambda$ obtained through direct focusing of Bessel–Gauss beam with $\kappa = 0.57$. The area of the super-resolution spot is $0.126\,\lambda^2$. As a longitudinally polarized beam, the maximum electric density of the transversal polarization is only 5 % that of the longitudinal electric density. The way to evaluate the quality of this beam is to look at the energy contained in longitudinal polarization in the whole focused beam, which is defined as the figure of merit as:

$$\eta = \Phi_z/(\Phi_z + \Phi_r) \tag{7.11}$$

with

$$\begin{cases} \Phi_z = 2\pi \int_0^{r_0} |E_z(r,0)|^2 r\mathrm{d}r \\ \Phi_r = 2\pi \int_0^{r_0} |E_r(r,0)|^2 r\mathrm{d}r \end{cases} \tag{7.12}$$

where $r_0 (= 0.65\,\lambda)$ is the first zero point in the distribution of radial electric intensity, as shown in the $|E_r|^2$ curve of Fig. 7.9d. The figure of merit of the beam is $\eta = 0.876$.

Comparing circularly polarized light with radially polarized light, one can find that in Fig. 7.9 for the radially polarized light, the longitudinal component is localized and occupies most part of the total electric energy, and the parasitic radial component is donut-shaped [11], and the radially polarized light illumination is only suitable for subsurface imaging due to the deterioration of resolution for a longitudinally polarized beam inside materials, where the longitudinal component plays a dominant role, this may be solved by a type of material that is sensitive only to the longitudinal component [12, 13]. However, for circularly polarized light illumination (as shown in Fig. 7.7), the transverse component is dominant in the spot and does not decay rapidly inside the sample, which is very beneficial for optical data storage, optical coherent tomography, and three-dimensional confocal imaging and nanolithography, etc., since in these applications, the super-resolution non-diffraction beam needs to go into the media.

7.3 Combination of Pupil Filters with Nonlinear Absorption Thin Films

Optical far-field super-resolution with pupil filters can realize below-diffraction-limited spot and long DOF; however, there are some concerns that the side lobe becomes non-negligible when the main lobe is squeezed to a certain extent, and it

is difficult to reduce the super-resolving spot to nanoscale. Thus, the combination of super-resolving pupil filters with nonlinear absorption thin films is a good way to reduce the main lobe size to nanoscale and lower the intensity ratio of side lobe to main lobe, where the pupil filters are applied to extend the DOF, and the nonlinear absorption thin films are used to further reduce the spot size to the nanoscale.

7.3.1 Combination of Nonlinear Saturation Absorption Thin Films with Three-Zone Annular Binary Phase Filters: Linearly Polarized Light Illumination

In real applications, the optical elements are required to be low cost, easy fabrication, and simple operation. The binary pupil filter is one of the simple optical elements, and the linearly polarized laser beam can be directly obtained from laser devices. Here, a design of linearly polarized light modulated by annular binary pupil filter is used to obtain super-resolving spot. In order to simplify the simulation process, the three-zone annular binary pupil filter is used to combine with the nonlinear saturation absorption thin film (NSATF). The super-resolution spot irradiates onto the NSATF and results in the generation of the nanoscale spot.

The combination of three-zone annular binary phase filter with NSATF is shown in Fig. 7.10. The super-resolving spot is incident to the NSATF. For NSATF, the higher laser intensity results in a larger transmittance. The intensity presents an approximate Gaussian profile for both the main lobe and side lode of the super-resolving spot.

The super-resolving spot is incident onto the NSATF, the side lobe is squeezed, and the main lobe is further increased comparatively because the main lobe intensity is several times larger than that of the side lobe. A nanometric optical aperture is produced in the central region of the main lobe because for the NSATF

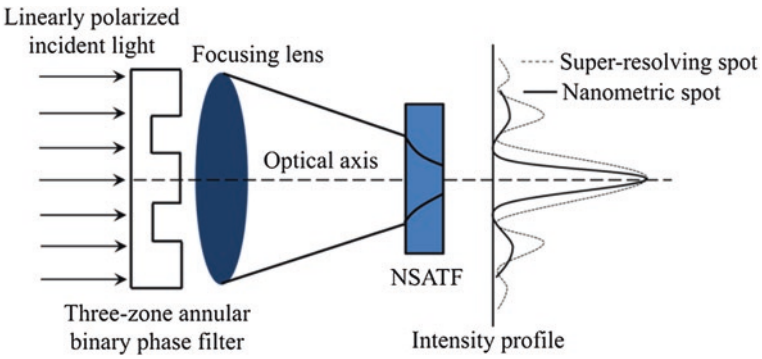

Fig. 7.10 Combination of the NSATF with the three-zone annular binary phase filter. Reprinted from [7], Copyright 2013, with permission from Elsevier

the higher laser intensity results in larger transmittance. Subsequently, the super-resolving spot goes through the nanometric optical aperture and a nanoscale spot is formed in the center of the main lobe. The side lobe is eliminated or squeezed. Finally, the nanoscale spot occurs only at the back side of the NSATF. The production of the nanometric optical aperture is dynamic and reversible, wherein the aperture can be open or closed when the laser beam is turned on or off.

7.3.1.1 Super-Resolution Performances

1. *Theoretical calculation analysis*

Here, the GaN laser with a wavelength of 405 nm is used as illumination light. In order to produce the nanometric optical aperture in the NSATF, the nonlinear saturation absorption coefficient β is required to reach up to the order of magnitude of -10^{-2} to -10^{-3} m/W. The Sb_2Te_3 thin films with a thickness of 50 nm are chosen as NSATF. The linear and nonlinear absorption coefficients at 405 nm wavelength are $\alpha_0 = 7.78 \times 10^7/\text{m}$ and $\beta = -2 \times 10^{-2}$ m/W, respectively [14]. The NSATF is combined with the three-zone annular binary phase filter to improve the super-resolving effect further. The combination schematic is described in Fig. 7.10, where the incident laser power is set at 2.0 mW.

It is well known that when a laser beam spot with an initial incident intensity profile of $I(x, y)$ irradiates onto a thin-film sample, the intensity decaying can be roughly expressed as:

$$\frac{\mathrm{d}I(x,y)}{\mathrm{d}z} = -\alpha I(x,y) \tag{7.13}$$

where α is the absorption coefficient of the sample and $\mathrm{d}I$ is the variation of the laser intensity along the sample thickness direction, $\mathrm{d}z$. For three-zone annular binary phase filter, the $I(x, y)$ can be obtained by formula (7.5). The transmitted beam spot intensity can be obtained by integrating formula (7.13) as follows:

$$I_t(x,y,L) = I(x,y)\exp(-\alpha L) \tag{7.14}$$

where L is the thin-film thickness. Formula (7.14) is the Beer–Lamber formula. For general samples, $\alpha = \alpha_0$ is constant and the normalized transmitted beam spot intensity profile, $I_t(x, y, L)$, is spatially the same as the initial incident laser intensity, $I(x, y)$. However, the α changes with laser intensity for the nonlinear absorption sample. In order to simplify the calculation, assuming that the α remains unchanged along sample thickness; thus,

$$\alpha(x,y) = \alpha_0 + \beta I(x,y) \tag{7.15}$$

Formula (7.15) shows that the absorption coefficient presents similar spatial profile to the incident laser intensity for nonlinear absorption samples. The nanometric spot after passing through the NSATF can be calculated by substituting formulas (7.5) and (7.15) into formula (7.14) to obtain

$$I_t(x, y, L) = I(x, y) \exp\left\{-\left[\alpha_0 + \beta I(x, y)\right] L\right\} \tag{7.16}$$

Formula (7.16) shows that the size of the nanometric spot is determined by the nonlinear absorption coefficient β, thin-film thickness L, and incident laser density I. For a given NSATF, β and L are fixed. Thus, the nanometric spot size formed at the back of the NSATF can only be tuned by changing the incident laser intensity I, namely the laser power P_L, because $I \propto P_L$.

2. *Super-resolution spot*

The simulated results are presented in Fig. 7.11, where Fig. 7.11a gives the spot in the x–y plane. The spot through the NSATF is still an ellipse, and a weak side lobe occurs near the main lobe. Figure 7.11b shows a comparison of the spot through the NSATF and the super-resolving spot, where the upper half is the super-resolving spot and the lower half is the spot through the NSATF. The size and side lobe intensity of the spot through the NSATF are, respectively, smaller and weaker than

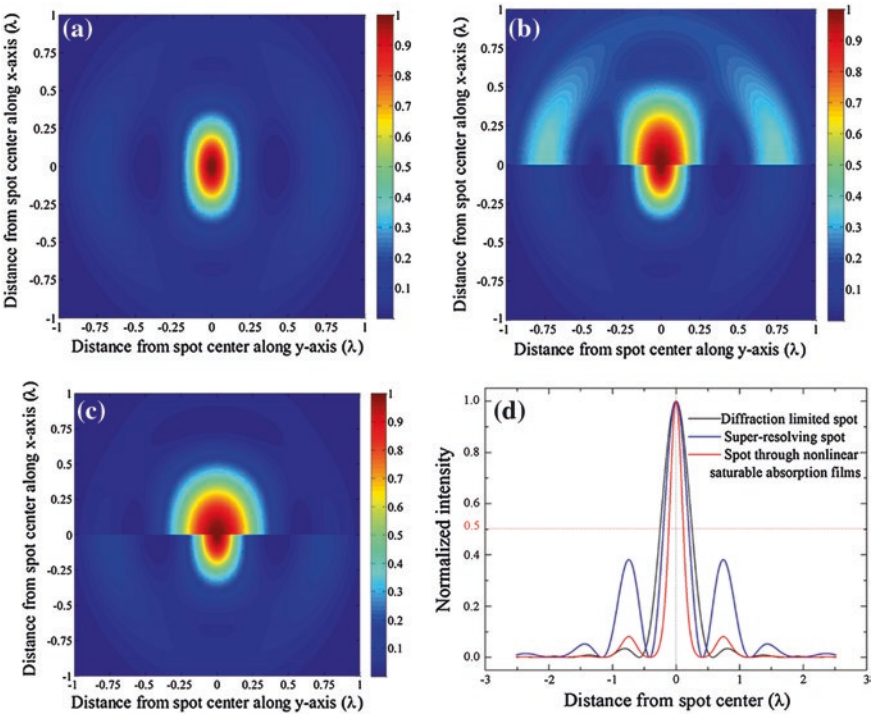

Fig. 7.11 Comparison of transverse intensity distribution at the focal plane (x–y plane). **a** Through NSATF. **b** Comparison of spot through NSATF (*lower half*) and super-resolving spot (*upper half*). **c** comparison of spot through NSATF (*lower half*) and diffraction-limited spot (*upper half*). **d** Comparison of y-axis normalized intensity distribution (*black* diffraction-limited, *blue* super-resolving, and *red* through NSATF). Reprinted from [7], Copyright 2013, with permission from Elsevier

those of the super-resolving spot. The diffraction-limited spot is compared further in Fig. 7.11c, where the upper half is the diffraction-limited spot and the lower half is the spot through the NSATF. The spot through the NSATF is approximately 50 % of the diffraction-limited spot, and the side lobe intensity is slightly larger than that of the diffraction-limited spot.

Figure 7.11d presents the comparison among the spot through the NSATF, the super-resolving spot, and the diffraction-limited spot along the *y*-axis distribution. The FWHM of the spot at the *y*-axis direction is reduced to a nanoscale size of approximately 100 nm, which is smaller than the FWHM of the diffraction-limited spot of ~204 nm , as shown in Fig. 7.11d. The size of the spot through the NSATF is the smallest among the three spots, and the side lobe is effectively squeezed to approximately 0.12 times of the main lobe intensity. Thus, the size of the main lobe is reduced, and the side lobe is also squeezed by the combination of three-zone annular binary phase filter and NSATF.

7.3.1.2 Application Schematic Design

The FWHM of the main lobe size in combined NSATF and three-zone annular binary phase filter can reduce to nanometric size, and the side lobe intensity can also be squeezed to approximately 12 % of the main lobe intensity. The nanometric spot is formed at the back of the NSATF. However, the nanometric spot exponentially decays along the *z*-axis direction and becomes difficult to propagate to far-field distance. Figure 7.12 shows a feasible application schematics to solve this problem. The samples (such as chalcogenide resists and phase-change materials) are directly deposited on the substrate. Then, a transparent dielectric thin film with a thickness of 20–50 nm, which is used as isolation layer and near-field distance controlling layer, is deposited on the samples. Last, the NSATF is deposited on the transparent dielectric film. In the application schematic, the nanometric spot is directly coupled into the samples in near-field range of approximately 20–50 nm. Thus, the nanolithography, nano-optical data storage, and nano-optical imaging can be realized by the combined NSATF and three-zone annular binary phase filter.

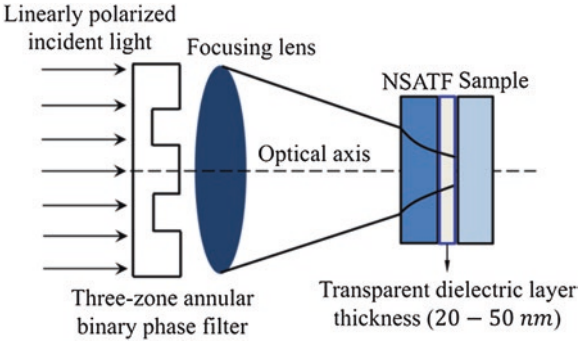

Fig. 7.12 Feasible schematic of combined NSATF and three-zone annular binary phase filter. Reprinted from [7], Copyright 2013, with permission from Elsevier

The applicable schematic is easy to operate because the distance between the focusing lens and the sample remains in the far-field range. Moreover, the tolerance distance between the super-resolving spot and the sample can be larger than the DOF of the diffraction-limited spot, where the tolerance distance is defined as the tunable distance of the sample along the z-axis direction, where the nanometric spot intensity remains almost unchanged within the tolerance distance. Figure 7.13 shows the tolerance distance. Figure 7.13a gives the tolerance distance at the y–z meridian plane. Figure 7.13b presents a comparison of the tolerance distance of nanometric spot and the DOF of the super-resolving spot. The tolerance distance is a little smaller than the DOF of the super-resolving spot. Figure 7.13c shows the comparison of the tolerance distance and DOF of the diffraction-limited spot. The tolerance distance is larger than the DOF of the diffraction-limited spot. Figure 7.13d is the comparison along the z-axis direction in Fig. 7.13a, b, and c. The tolerance distance is a slightly smaller than the DOF of the super-resolving spot, but twice more

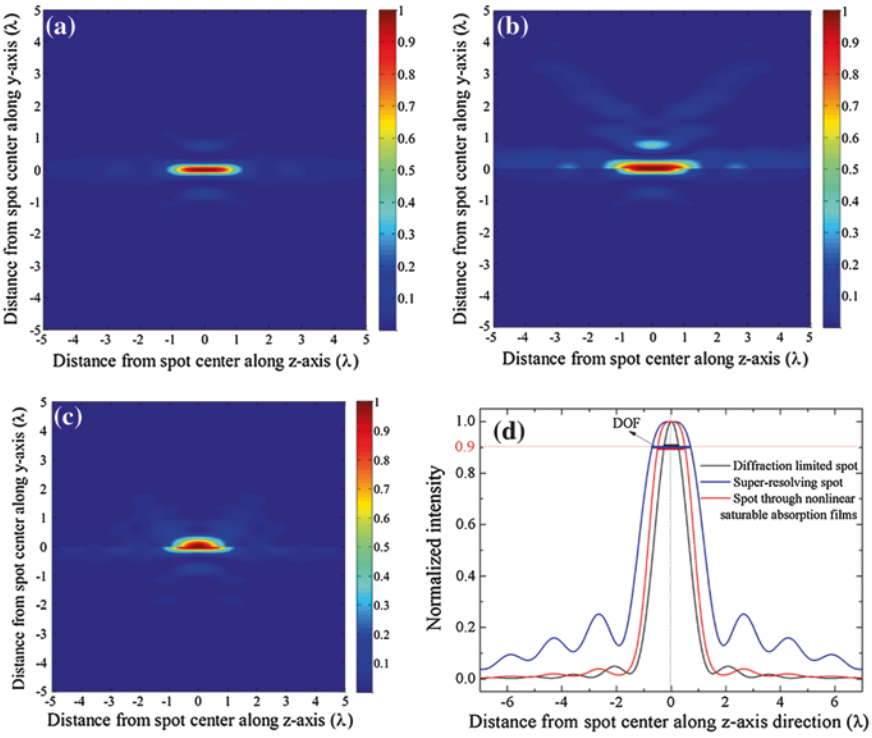

Fig. 7.13 Tolerance distance and DOF at y–z meridian plane. **a** Tolerance distance of nanometric spot. **b** Comparison of tolerance distance (*lower half*) and DOF of super-resolving spot (*upper half*). **c** Comparison of tolerance distance (*lower half*) and DOF of diffraction-limited spot (*upper half*). **d** Comparison of tolerance distance and DOF of diffraction-limited spot, and super-resolving spot along z-axis direction (*black* diffraction-limited spot, *blue* super-resolving spot, and *red* nanometric spot). Reprinted from [7], Copyright 2013, with permission from Elsevier

the DOF of the diffraction-limited spot. The intensity fluctuation at both sides of the center of the nanometric spot is much lower than the super-resolving spot and close to the intensity fluctuation of the diffraction-limited spot.

7.3.2 Combination of Nonlinear Reverse Saturation Absorption Thin Films with Five-Zone Binary Pupil Filter: Circularly Polarized Light Illumination

As is analyzed in lots of literatures, the circularly polarized light can produce uniform and circular spot, which is very useful for maskless direct laser writing lithography. Here, under the illumination of circularly polarized light, the combination of super-resolution pupil filter and nonlinear reverse saturation absorption inorganic resist thin films is designed to obtain nanolithography.

7.3.2.1 Design Method and Principle

The design schematic is given in Fig. 7.14. A collimated circularly polarized laser beam passes through the super-resolution pupil filter and enters converging lens. The converging lens focuses the laser beam into a super-resolution spot (marked in green area), whose size is reduced compared with the focused spot (marked in red

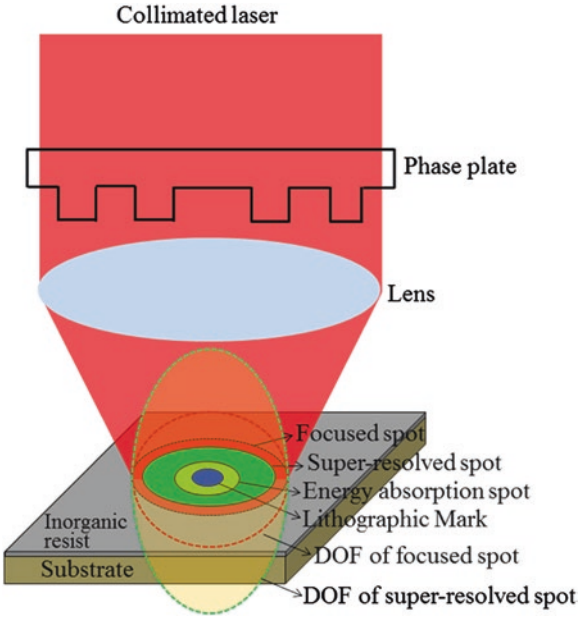

Fig. 7.14 Combination of pupil filter and nonlinear reverse saturation absorption films. Reprinted from [15], Copyright 2013, with permission from Elsevier

area). Meanwhile, the DOF (marked in green dotted line) of the super-resolving spot is extended compared with the DOF (marked in red dotted line) of focused spot. The super-resolution spot is incident on the inorganic resist thin films, which are deposited on the substrate. The inorganic resist thin films have an obvious nonlinear reverse saturation absorption characteristic. The higher the laser intensity, the larger the absorptivity is. The largest absorptivity is in the central part of the super-resolution spot due to the Gaussian profile of the light intensity, which causes a smaller energy absorption spot at the center of the super-resolving spot (marked in yellow area). The inorganic resist thin film absorbs the spot energy and is heated. When the energy exceeds certain threshold, the inorganic resist at the central region of the energy absorption spot experiences a structural transformation and forms the lithographic marks. The feature size of the lithographic mark is far smaller than the focused spot.

The combination of super-resolution pupil filter and nonlinear reverse absorption inorganic resist thin films does not only extend the DOF, which is very helpful for the auto-focusing and tracking servo controlling during the laser direct writing lithography, but also reduces the lithographic feature size to being far smaller than the focused spot itself.

Here, a five-zone annular phase-only binary optical element is chosen as pupil filter because higher strehl ratio than the central block ternary optical element, where strehl ratio is defined as the ratio of peak intensity of the main lobe of super-resolving spot to that of the diffraction-limited focused spot. Figure 7.15 gives the schematic. Figure 7.15a presents the incident light traveling through five-zone annular binary phase filter and being focused by high-NA lens. Figure 7.15b is the front view of the five-zone annular binary phase filter. A monochromatic circularly polarized plane wave with a wavelength of 635 nm passes through the five-zone annular phase-only binary pupil filter and is focused by the oil lens with a numerical aperture of 1.4 into a super-resolving spot. The radii of the five-zone

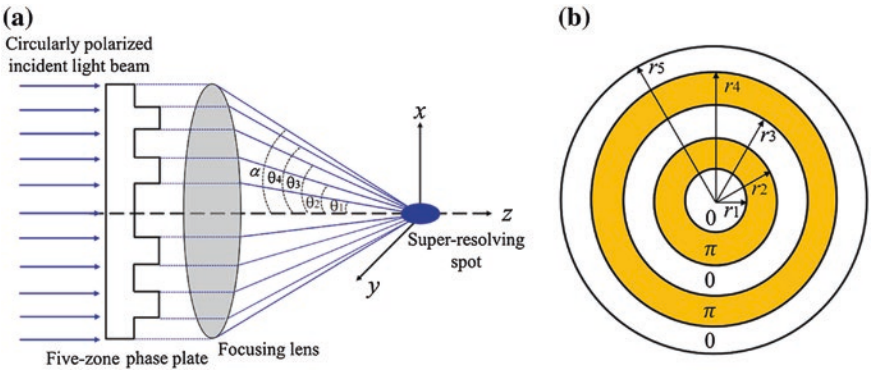

Fig. 7.15 The schematic of the annular five-zone phase-only binary pupil filter. **a** Incident light passes through five-zone annular binary phase filter and focused by high NA lens. **b** The front view of the five-zone annular binary phase filter

annular phase-only binary pupil filter are defined as r_1, r_2, r_3, r_4, and r_5. The phase of the five-zone is set to be $0 - \pi - 0 - \pi - 0$ structure. Correspondingly, the transmission function is

$$t_3(\theta) = \begin{cases} +1 & \text{for} \quad 0 \le \theta < \theta_1 \\ -1 & \text{for} \quad \theta_1 \le \theta < \theta_2, \theta_3 \le \theta < \theta_4 \\ +1 & \text{for} \quad \theta_2 \le \theta < \theta_3, \theta_4 \le \theta < \alpha \end{cases} \tag{7.17}$$

The optical field distribution of super-resolving spot can be calculated by formulas (7.7) and (7.8) in Sect. 7.2.2.1, where $t(\theta)$ is replaced by $t_3(\theta)$ of formula (7.17).

Besides the annular five-zone phase-only binary pupil filter, the inorganic resist thin film is of importance to pursue nanoscale lithography marks. The inorganic resist thin film needs to have strong optical nonlinear reverse saturation absorption effect and structural transformation threshold characteristic. The AgInSbTe thin films present obvious nonlinear reverse saturation absorption, and the threshold characteristic of structural transformation between amorphous and crystalline states at the irradiation of laser beam with a wavelength of 635 nm. The AgInSbTe is typical chalcogenide phase-change materials. The temperature of structural transformation from amorphous state to crystalline state is about 150 °C. The structural transformation can happen under a focused laser pulse irradiation due to light absorption induced temperature rise [16]. In real application, the AgInSbTe thin films can be directly deposited on the substrate by magnetron-controlling sputtering method. The absorption coefficient of nonlinear thin film can be also expressed as

$$\alpha = \alpha_0 + \beta_{\text{eff}} I \tag{7.18}$$

where β_{eff} is an effective nonlinear absorption coefficient. The β_{eff} of AgInSbTe is written as [17]

$$\beta_{\text{eff}} = -0.088 + \left(4.7 \times 10^{-9}\right) I - \left(6.68 \times 10^{-17}\right) I^2 + \left(3.4 \times 10^{-25}\right) I^3 \tag{7.19}$$

Hence, the AgInSbTe thin film is used as the inorganic resist material in the design schematic.

7.3.2.2 Super-Resolving Spot with Pupil Filters

In the numerical calculation, the design results are as follows: the aperture radius of the lens $R = 2$ mm, and the focal length $f = 0.77$ mm, and the normalized optimum radii of the five-zone phase-only binary pupil filter are $r_1 = 0.21, r_2 = 0.48, r_3 = 0.91, r_4 = 0.97$, and $r_5 = 1$. The spot intensity profiles in the y–z plane are given in Fig. 7.16.

Figure 7.16a is the focused spot in the y–z meridian plane and the DOF can be observed along the z-axis direction. Figure 7.16b is the super-resolving spot in the y–z meridian plane. A direct comparison of the DOF of super-resolving spot

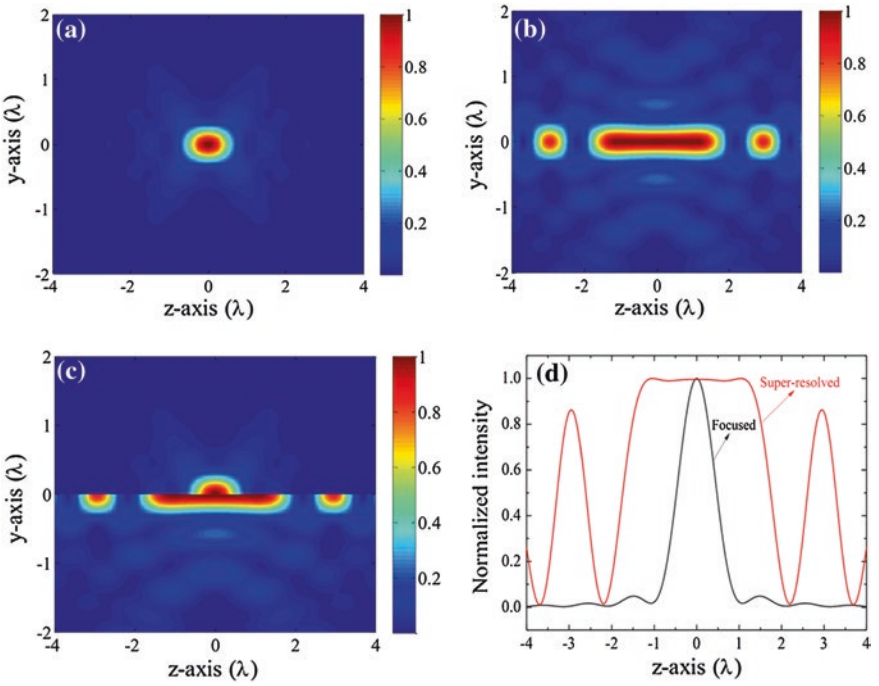

Fig. 7.16 The spot intensity profiles in the y–z meridian plane. **a** Focused spot. **b** Super-resolving spot. **c** Comparison of super-resolving and focused spots. **d** The intensity profile along z-axis. Reprinted from [15], Copyright 2013, with permission from Elsevier

and the focused spot is present in Fig. 7.16c, where the upper half is the focused spot and the lower half is the super-resolving spot. Figure 7.16d gives the quantitative comparison of the DOF along the z-axis. The DOF is extended to 7.39 times of the focused spot and reaches up to about 2.82 λ, that is, DOF is about 2 μm for the design schematic shown in Fig. 7.15, which is useful for the focusing and tracking servo controlling in high-speed laser direct writing lithography. Here, one also notices that there are side lobes in the 3 λ distance range near the focal point at $z = 0$, the ratio of side lobe intensity to the main lobe intensity is about 0.87. Actually, in laser direct writing, such lobes are little ill-influence on the resist because the resist is placed at the focal point center. The center of focal spot is about 3 λ distance away from the side lobes.

The intensity profiles of spot along the x–y plane are given in Fig. 7.17. Figure 7.17a is the focused spot, and Fig. 7.17b is the super-resolving spot. By comparing Fig. 7.17a, b, the main lobe size of the super-resolving spot is smaller than focused spot, while a side lobe appears in the super-resolving spot.

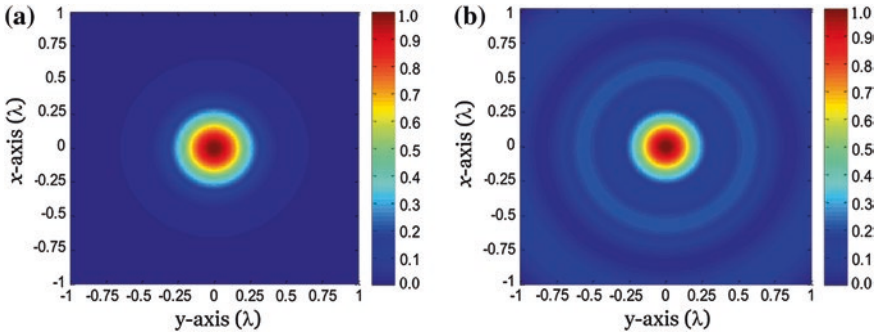

Fig. 7.17 The intensity profile along x–y plane. **a** Focused spot. **b** Super-resolving spot. Reprinted from [15], Copyright 2013, with permission from Elsevier

7.3.2.3 Nanometric Lithographic Marks

As is shown in Fig. 7.14, the super-resolving spot is incident on the AgInSbTe thin film. The AgInSbTe thin film has an obvious reverse saturation absorption characteristic, and the higher the laser intensity, the larger the absorptivity is. The largest absorptivity is in the central part of the super-resolving spot due to the approximately Gaussian profile of the intensity, which causes a smaller energy absorption spot at the center of the super-resolving spot. The energy absorption spot intensity can be expressed as

$$I_a(x, y) \propto \alpha[I(x, y)]^g \tag{7.20}$$

where g reflects the multi-photon absorption characteristic, $g = 1$ means single-photon absorption, and $g = 2$ means two-photon absorption. For the AgInSbTe thin films, the $g = 1.87$ [16, 17] due to the nonlinearity resulting from one-photon absorption and two-photon absorption process. $I(x, y)$ is the intensity profile of the super-resolving spot. Substituting formula (7.18) and $g = 1.87$, and the data of Fig. 7.17b into formula (7.20), where $\alpha_0 = 3 \times 10^7/\text{m}$, laser power $P_L = 5\,\text{mW}$. One can get the intensity profile of the energy absorption spot, and Fig. 7.18a gives the results. Figure 7.18a indicates that the side lobe of super-resolving spot has not only disappeared, but also the energy absorption spot size is obviously reduced. The AgInSbTe thin film absorbs the spot energy and is heated. The temperature rise of the AgInSbTe thin films presents approximately a Gaussian profile. When the ratio $T_{\text{th}}/T_{\text{peak}}$, where T_{th} and T_{peak} are phase-change threshold and peak temperatures, respectively, is controlled at about 80 % by tuning the incident laser power P_L, and the phase-change threshold lithographic mark can be further reduced to be smaller than the energy absorption spot diameter accordingly, as is shown in Fig. 7.18b, where the phase-change threshold lithographic mark is about 1/10 the focused spot size, and the lithographic marks with a nanoscale feature size can be obtained.

For the further understanding of the reduction of the spot size and the phase-change threshold lithographic mark, the intensity profiles along y-axis of spots in Figs. 7.17 and 7.18 are plotted in Fig. 7.19. The black line is the focused spot,

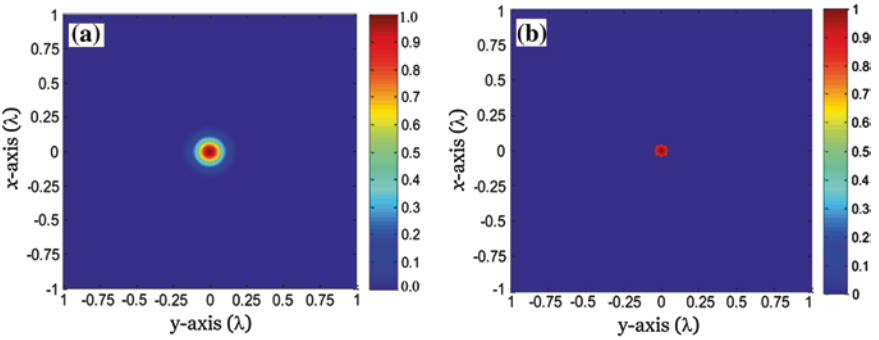

Fig. 7.18 Intensity profile along x–y plane. **a** Effective energy absorption spot. **b** Phase-change threshold lithographic spot. Reprinted from [15], Copyright 2013, with permission from Elsevier

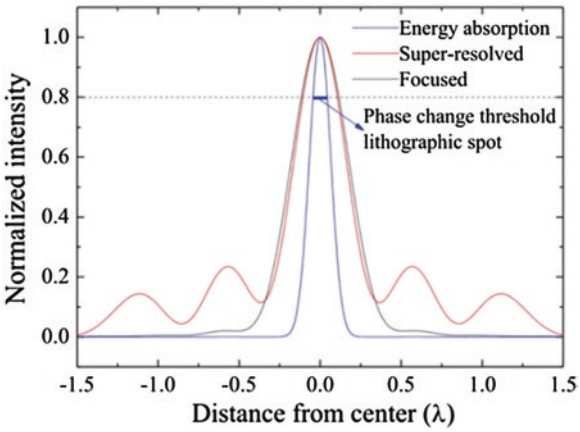

Fig. 7.19 The comparison of intensity profile along y-axis. Reprinted from [15], Copyright 2013, with permission from Elsevier

and the FWHM is about 256.74 nm. The red line is the super-resolving spot intensity profile, and the FWHM is about 233.68 nm. The super-resolving spot size is reduced to about 89 % of the focused spot. The side lobe intensity is about 23 % of the main lobe. The blue line is the effective energy absorption spot intensity profile, and the FWHM is squeezed to about 96.52 nm, which is about 0.37 times of the focused spot size. In order to make full use of the phase-change threshold characteristic, one assumes that the intensity above 80 % of the maximum can induce the phase change of AgInSbTe thin films from amorphous state to crystalline state. Thus, the lithographic mark can be reduced down to about 54.6 nm, which is about 0.106 times of the original spot size, and reduced down to about $\lambda/12$.

The combination of five-zone phase-only binary annular pupil filter and AgInSbTe nonlinear reverse saturation absorption inorganic resist thin films does not only extend the DOF to 7.39 times of the focused spot, which is very helpful for the auto-focusing and tracking servo controlling during the laser direct writing lithography, but also reduces the lithographic feature size down to 54.6 nm.

The ill-effect of side lobe on the lithography is also eliminated by the nonlinear reverse saturation absorption and the phase-change lithographic threshold characteristic.

7.4 Nonlinear Thin Films as Pupil Filters

The nonlinear thin film can be used as a uniform pupil and is inserted at the aperture of the focusing lens. There is no need to fabricate the annular pupil filter. Inside the film, the complex refractive index is a function of the local light intensity, which is the key property. When a laser beam is incident upon the thin film, the complex refractive index changes according to the spatial intensity distribution of the laser beam. To simplify, here, the scalar diffraction theory is used to analyze the super-resolution spot performance.

7.4.1 Scalar Theoretical Basis

The general integral of the spot distribution from textbooks is [18]:

$$U(u,v) = 2\int_0^1 P_u(\rho)J_0(v\rho)e^{-iu\rho^2/2}\rho \mathrm{d}\rho \tag{7.21}$$

with

$$\begin{cases} v = \frac{2\pi}{\lambda} r\mathrm{NA} \\ u = \frac{2\pi}{\lambda} z\mathrm{NA}^2 \end{cases} \tag{7.22}$$

where J_0 is the zeroth-order Bessel function of the first kind, r is the radial coordinate, λ is wavelength, NA is the numerical aperture, and z is the axial coordinate. $P_u(\rho)$ is the so-called pupil function which can change the focal properties. In the diffraction limit, one assumes a constant pupil function $P_u(\rho) = 1$ and obtains straightforwardly the well-known intensity distribution $I(u,v) = U(u,v)U(u,v)^*$. At $u = 0$, by formula (7.21), the electric field in focal plane can be obtained as

$$U(0,v) = 2\int_0^1 P_u(\rho)J_0(v\rho)\rho \mathrm{d}\rho \tag{7.23}$$

The intensity distribution in focal plane is

$$I(0,v) = U(0,v)U(0,v)^* = \left[\frac{2J_1(v)}{v}\right]^2 \tag{7.24}$$

Similarly, at $v = 0$, by formula (7.21), the electric field in axial direction can be obtained as

$$U(u,0) = 2\int_0^1 P_u(\rho)e^{-iu\rho^2/2}\rho \mathrm{d}\rho \tag{7.25}$$

The intensity distribution in axial direction is

$$I(u,0) = U(u,0)U(u,0)^* = \left[\frac{\sin(u/4)}{u/4}\right]^2 \tag{7.26}$$

Setting $t = \rho^2$, $P_u(\rho)$ can be written as $Q(t)$; here, $Q(t)$ is defined as equivalent pupil function. Formula (7.25) is rewritten as

$$U(u,0) = \int_0^1 Q(t)e^{-iut/2}\mathrm{d}t \tag{7.27}$$

Formula (7.27) indicates that the amplitude of electric field in the axial direction is a Fourier transform of $Q(t)$; thus, by changing the pupil function $P_u(\rho)$, one can realize the different pupil filers, such as amplitude-type, phase-type, and hybrid-type.

In the focal plane, the first zero of the intensity $I(v,0)$ occurs when $v_0 = 3.83$, whereas along the axis, the first zero of the intensity $I(0,u)$ locates at $u_0 = 4\pi$.

In the integral of formula (7.21), the origin of the coordinates in both radial and axial directions is set at the focus of the lens with a constant pupil function $P_u(\rho) = 1$. The focus position shifts in the axial direction when a different pupil function is introduced and the focal plane moves according to the pupil function.

The super-resolution pupil filter is designed into a uniform thin film instead of an annular pupil filter. The thin film is a nonlinear material, where the complex refractive index is a function of the local light intensity. Therefore, the transmission of the thin film is no longer a spatially uniform function. The transmission depends on the power and spatial intensity distribution of the incident laser beam. In other words, a special annular pupil filter is created and its property is tunable by adjusting the laser power and beam characteristics. In Fig. 7.20, the model of the special pupil function is explained schematically [19].

The mathematical expression for the pupil function is derived as follows. The complex electric field can be written as

$$\tilde{E}(r) = E_0 \exp\left(-\frac{r^2}{w_0^2}\right) \tag{7.28}$$

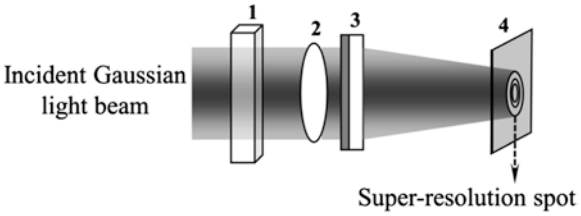

Fig. 7.20 Schematic of laser tunable pupil filter with nonlinear film. *1* laser power regulator, *2* imaging lens, *3* pupil filter (composed of nonlinear thin film and substrate), and *4* focal plane. Reprinted with permission from [19]. Copyright 2008, Optical Society of America

where E_0 is the on-axis amplitude and w_0 is the radius of the incident beam. The laser beam is incident upon the nonlinear thin film. The complex refractive index of the thin film changes according to the local laser power density $I(r) = \left|\tilde{E}(r)\right|^2$ as

$$\tilde{n}(r) = n(r) + ik(r) = \left[n_0 + \gamma I(r)\right] + i\frac{4\pi}{\lambda}[\alpha_0 + \beta I(r)] \tag{7.29}$$

where n_0 is the linear refractive index, γ the nonlinear refraction coefficient.

Due to the nonlinearity of the thin film, the transmitted light has several changes. Firstly, the amplitude and wavefront phase of the complex electric field are modulated by the internal multi-interference effect. Secondly, the wavefront phase is also modified due to the Gaussian lens-like phase shift. All the changes can be incorporated into the transmitted complex electric field as

$$\tilde{U}(r) = \tilde{E}(r)\tilde{t}(r)\exp\left[i\Delta\varphi(r)\right] \tag{7.30}$$

One defines the complex pupil function as

$$\tilde{P}_u(r) = \frac{\tilde{U}(r)}{\tilde{E}(r)} = \tilde{t}(r)\exp\left[i\Delta\varphi(r)\right] \tag{7.31}$$

where the distribution of the Gaussian lens-like phase shift can be written in a linear approximation as

$$\Delta\varphi(r) \approx \frac{4\gamma P_L\{1 - \exp\left[-\alpha(r)L\right]\}}{\lambda\alpha(r)w_0^2}\exp\left(-\frac{2r^2}{w_0^2}\right) \tag{7.32}$$

where P_L is the laser power, L is the thickness of the thin film, $\alpha(r) = \alpha_0 + \beta I(r)$ is the absorption coefficient, and the laser power density can be written as

$$I(r) = \frac{2P_L}{\pi w_0^2}\exp\left(-\frac{2r^2}{w_0^2}\right) \tag{7.33}$$

The complex transmission for the two-interface layered system in normal incidence is

$$\tilde{t}(r) = \tau(r)e^{i\Theta} = \frac{t_{12}t_{23}\exp\left[2i\frac{2\pi}{\lambda}\tilde{n}(r)L\right]}{1 + r_{12}r_{23}\exp\left[2i\frac{2\pi}{\lambda}\tilde{n}(r)L\right]} \tag{7.34}$$

with

$$\begin{cases} r_{12} = \frac{n_1 - \tilde{n}(r)}{n_1 + \tilde{n}(r)} \\ r_{23} = \frac{\tilde{n}(r) - n_3}{\tilde{n}(r) + n_3} \\ t_{12} = \frac{2n_1}{n_1 + \tilde{n}(r)} \\ t_{23} = \frac{2\tilde{n}(r)}{\tilde{n}(r) + n_3} \end{cases} \tag{7.35}$$

where r_{12}, r_{23}, t_{12}, and t_{23} are, respectively, the reflection and transmission coefficients for each interface, and n_1 and n_3 are the corresponding refractive index for air and substrate, respectively. The inclusion of the complex refractive index, the local nonlinear absorption, and the multi-interference in formulas (7.32) and (7.34) enables one to calculate the exiting field distribution.

Substituting formulas (7.32)–(7.34) into formula (7.31), one finally obtains the complex pupil function as

$$\tilde{P}_u(r) = \tau(r) \exp\left\{ i \left[\Theta(r) + \frac{4\gamma P_L \left[1 - e^{-\alpha(r)L}\right]}{\lambda \alpha(r) w_0^2} \exp\left(-\frac{2r^2}{w_0^2} \right) \right] \right\} \quad (7.36)$$

One notes that the use of formula (7.34) should be considered as an approximation because the Fresnel formalism is actually developed for plane waves.

7.4.2 Super-Resolution Spot Analysis

Numerical simulations are carried out based on the above formalism for the focal spot. Since the complex pupil function in formula (7.36) contains radial modulation both in amplitude and in phase, one expects therefore that the super-resolving spot properties would be significantly modified from the case of constant pupil function shown analytically in formulas (7.24) and (7.26).

In the numerical simulations, the parameters are chosen as follows. The beam radius is $w_0 = 5\,\mu\text{m}$. The thin film thickness is fixed $L = 100\,\text{nm}$. The laser has the wavelength $\lambda = 633\,\text{nm}$, and the laser power is assumed to be $P_L = 200\,\text{mW}$. Silicon is considered as the material to form the nonlinear thin film. Experimental parameters for the thin film are used for the nonlinear coefficients $\gamma \sim 4.45 \times 10^{-9}\,\text{m}^2/\text{W}$, and the initial refractive index $n_0 = 4.21$.

Before showing the focal spot, Fig. 7.21 presents the transmission of the thin film. In Fig. 7.21a, the normalized transmission amplitude is plotted, and in Fig. 7.21b, the phase of the transmission is plotted. One clearly sees from Fig. 7.21 that as far as the transmission is concerned, the thin film acts as an annular pupil filter with a strong radial modulation. The modulation contains three principal contributions, including amplitude, phase, and Gaussian lens-like phase shift. In Fig. 7.21, these three effects are shown separately. The amplitude modulation is symmetrically concave (see Fig. 7.21a). The phase modulation appears stair-like (see Fig. 7.21b). The Gaussian lens-like phase shift (the dotted curve in Fig. 7.21b) has an insignificant effect to the modulation.

The optical field intensity distributions of focal spot are shown in Fig. 7.22. In Fig. 7.22a, c, the distributions along the axis and at the focal plane are shown in top view mapping, respectively. In Fig. 7.22b, d, the cross sections are shown in axial and transverse directions, respectively. One realizes that the focal plane is shifted in axial direction, which is because that the added nonlinear pupil filter function changes the focal length. In the transverse direction, the first zero point occurs at 2.6

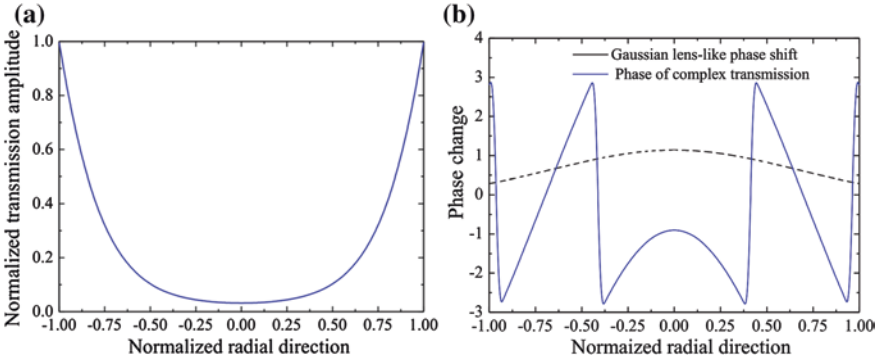

Fig. 7.21 The transmission with modulation (**a**) in amplitude, (**b**) in phase (*blue curve*) plus Gaussian lens-like shift (*black curve*). Reprinted with permission from [19]. Copyright 2008, Optical Society of America

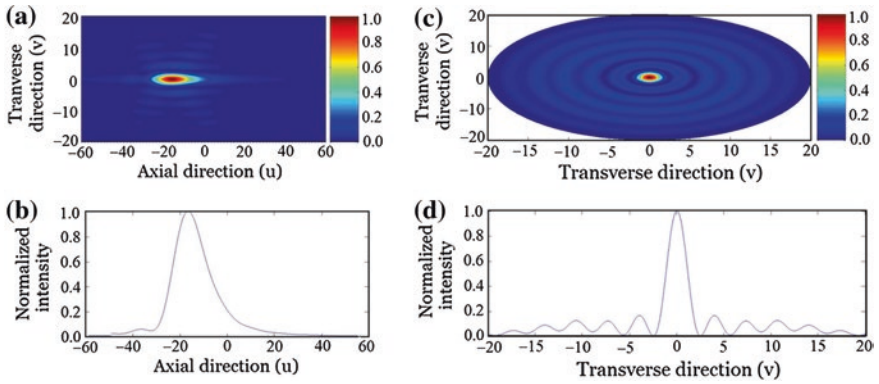

Fig. 7.22 The super-resolution spot in (**a**) top view mapping along axial direction, (**b**) axial cross-sectional plot, (**c**) top view mapping at the focal plane, (**d**) transverse cross-sectional plot. Reprinted with permission from [19]. Copyright 2008, Optical Society of America

which is significantly smaller than the first zero point 3.83 obtained from the well-known diffraction spot in formula (7.24). A careful comparison of FHWM with the diffraction spot shows that the spot becomes about 0.7 times of the focused spot with constant pupil function. Also, compared with the peak intensity of main lobe, in transverse direction, the peak intensity of side lobe is at about 16 % in height. In axial direction, the spot is shifted about 4π and the spot is prolonged.

Since the laser power appears in the expression of the pupil function of formula (7.36), it is anticipated that the super-resolving property is tunable by adjusting the incident laser power. The modification of the focal spot should occur when the laser power is increased due to the factor that the nonlinearity of the thin film increases with the laser power. In Fig. 7.23, the super-resolving spots are drawn

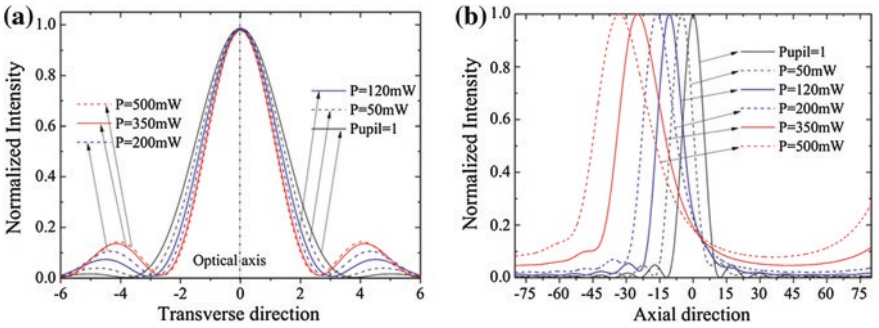

Fig. 7.23 Cross-sectional plots of the super-resolution focal spots in (**a**) transverse direction (**b**) axial direction with various laser powers. The focal spot of constant pupil is included for comparison. Reprinted with permission from [19]. Copyright 2008, Optical Society of America

as a function of the laser power. The original diffraction spot is also plotted for a comparison. As expected, the focal spot becomes narrow in transverse direction (see Fig. 7.23a) when higher laser power is applied. However, the decrease of the spot transverse size becomes insignificant when the laser power is further increased after about 350 mW. Figure 7.23b shows that the super-resolving spot shifts along the axis and the DOF increases with the increase of the laser power. As shown in Fig. 7.23b, the DOF is about 2 times of the focused spot when the laser power is increased to 500 mW. In general, the increased DOF would be helpful for the imaging and lithography.

Super-resolution pupil filter is a feasible method to reduce the spot down to below-diffraction-limited size and extend the DOF. However, it is difficult to eliminate the side lobe effect and obtain nanometric spot. The combination of super-resolution pupil filter with the nonlinear thin-film materials is a good way to realize nanoscale spot, which is useful for super-resolving optical imaging, high-density data storage, and nanolithography.

Appendix: Optical Field on the Image Space

In general, the electric component of optical field is mainly taken into account and denoted as $\vec{e}$, which is a complex field. In focusing and imaging system, a monochromatic point source S_0 with angular frequency ω passes through the lens and forms a Gaussian image point S_1 [20], as shown in Fig. 7.24.

The electric field of any point P at image space is denoted as $\vec{e}(P)$, and the electric vector $\vec{E}$ in time t is written as

$$\vec{E}(P,t) = \Re\left[\vec{e}(P)e^{-i\omega t}\right] \tag{7.37}$$

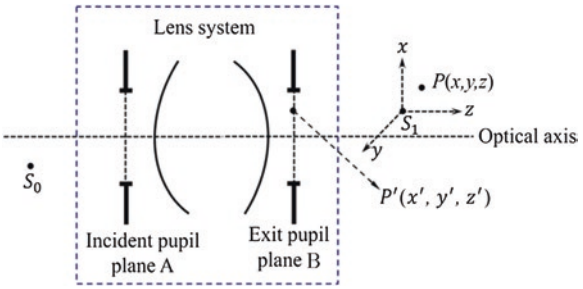

Fig. 7.24 Illustration for point source imaging and focusing

where symbol $\Re$ denotes the real part. Let (x, y, z) be the coordinate of point P in Cartesian orthogonal coordinate with origin at the S_1 and positive z-direction along the optical axis pointing away from the exit pupil. In whole image space $\vec{\boldsymbol{e}}$ satisfies time-independent wave equation and may be represented as

$$\vec{\boldsymbol{e}}(P) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{ik(s_x x + s_y y)} \left[\vec{\boldsymbol{U}}(s_x, s_y)\mathrm{e}^{iks_z z} + \vec{\boldsymbol{V}}(s_x, s_y)\mathrm{e}^{-iks_z z}\right] \mathrm{d}s_x \mathrm{d}s_y \quad (7.38)$$

with

$$k = n_{\mathrm{imsp}} \frac{2\pi}{\lambda_0} \quad (7.39)$$

$$\begin{cases} s_z = +\sqrt{1 - s_x^2 - s_y^2} & \text{for} \quad s_x^2 + s_y^2 \le 1 \\ s_z = +i\sqrt{s_x^2 + s_y^2 - 1} & \text{for} \quad s_x^2 + s_y^2 > 1 \end{cases} \quad (7.40)$$

where k is wave number in image space with refractive index n_{imsp}, and λ_0 is light wavelength in vacuum.

Formula (7.38) indicates that $\vec{\boldsymbol{e}}$ is in the form of two angular spectra ($\vec{\boldsymbol{U}}$ and $\vec{\boldsymbol{V}}$) of plane waves. $s_x^2 + s_y^2 \le 1$ means homogeneous waves, and $s_x^2 + s_y^2 > 1$ means inhomogeneous waves. Before determining the $\vec{\boldsymbol{U}}$ and $\vec{\boldsymbol{V}}$, some approximations need to be made as follows.

The first approximation is that all inhomogeneous waves can be neglected, which causes the infinite domain of integration in formula (7.38) to become finite domain, i.e., $s_x^2 + s_y^2 \le 1$. The approximation is obviously justified because s_z is pure imaginary number at $s_x^2 + s_y^2 > 1$, "$\vec{\boldsymbol{V}}$-wave" diverges as $kz \to \infty$, and "$\vec{\boldsymbol{U}}$-wave" is negligible because the distance Δz between pupil plane and image region is assumed to be lots of wavelengths.

The second approximation is that the Kirchhoff's boundary conditions can be met in the exit pupil plane (marked as B), i.e., the incident light inside the exit pupil plane does not be perturbed, and the light field is 0 outside the exit pupil plane. The approximation is justified only if the linear dimensions of exit pupil are larger than light wavelength, and the field at lots of light wavelength away from the pupil plane is taken into account.

The unperturbed incident field of any point $P'(x', y', z')$ at exit pupil plane is denoted as $\vec{\boldsymbol{e}}^{(i)}(P')$, and one can have the following assumption

$$\iint\limits_{s_x^2+s_y^2\le 1} \mathrm{e}^{ik(s_x x+s_y y)}\left[\vec{\boldsymbol{U}}(s_x,s_y)\mathrm{e}^{iks_z z}+\vec{\boldsymbol{V}}(s_x,s_y)\mathrm{e}^{-iks_z z}\right]\mathrm{d}s_x\mathrm{d}s_y = \begin{cases}\vec{\boldsymbol{e}}^{(i)}(P') & \text{for} \quad P'\in B\\ 0 & \text{for} \quad P'\notin B\end{cases} \tag{7.41}$$

Formula (7.41) is effective only if the exit pupil dimension is far larger than the light wavelength, and the distance between the calculated field and exit pupil plane is large enough. The integral of formula (7.41) can be estimated by an asymptotic approximation. For this purpose, one sets

$$\begin{cases} r' = +\sqrt{x'^2+y'^2+z'^2}\\ u_x = x'/r';\ u_y = y'/r'\\ u_z = z'/r' = -\sqrt{1-u_x^2-u_y^2}\end{cases} \tag{7.42}$$

Then if the formula (7.40) is also used, the left side of formula (7.41) becomes a sum of two integrals of the form

$$\vec{\boldsymbol{Q}}(kr';u_x,u_y,u_z) = \iint\limits_{s_x^2+s_y^2\le 1} \vec{\boldsymbol{P}}(s_x,s_y)\mathrm{e}^{ikr'\Phi(s_x,s_y;u_x,u_y,u_z)}\mathrm{d}s_x\mathrm{d}s_y \tag{7.43}$$

with

$$\Phi(s_x,s_y;u_x,u_y,u_z) = s_x u_x + s_y u_y \pm u_z\sqrt{1-s_x^2-s_y^2} \tag{7.44}$$

According to the principle of stationary phase, Born and Wolf obtained an asymptotic approximation to an integral for large kr' [20]. Since $u_z < 0$ formula (7.41) can be simplified as

$$\frac{2\pi i}{kr'}u_z\left[-\vec{\boldsymbol{U}}(-u_x,-u_y)\mathrm{e}^{-ikr'}+\vec{\boldsymbol{V}}(u_x,u_y)\mathrm{e}^{ikr'}\right] = \left\{\begin{matrix}\vec{\boldsymbol{e}}^{(i)}(P') & \text{for} \quad P'\in B\\ 0 & \text{for} \quad P'\notin B\end{matrix}\right\} \tag{7.45}$$

The error introduced in using the asymptotic approximation instead of the exact value is of the order of $(kr')^{-3/2}$. Except the points of immediate neighbor of the edge, the $\boldsymbol{e}^{(i)}(P')$ can be approximately expressed as

$$\vec{\boldsymbol{e}}^{(i)}(P') = \frac{\vec{\boldsymbol{a}}}{\sqrt{r_1}\sqrt{r_2}}e^{ik\wp(P')} \tag{7.46}$$

where r_1 and r_2 are the principal radii of curvature of the associated geometry optical wavefront going through point P'. The $\wp(P')$ is the eikonal function. $\vec{\boldsymbol{a}}$ is a complex vector of perpendicular to the geometrical ray through point P'.

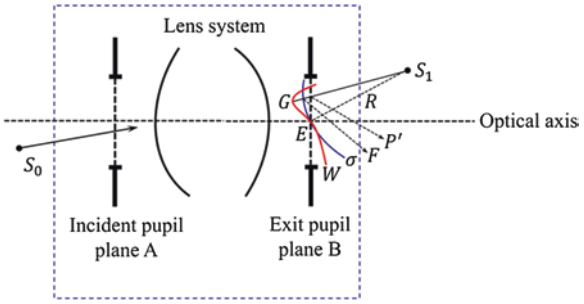

Fig. 7.25 Definition of wave aberration function

In geometrical optics approximation, $\vec{\boldsymbol{a}}$ is also independent of exact point P' and called as strength factor of the light ray.

In general, it is not easy to directly obtain $\wp(P')$; fortunately, $\wp(P')$ may be approximately expressed with a form of aberration function of system. Define σ as Gaussian reference sphere with a center of S_1, and the Gaussian reference sphere goes through the center point E of the exit pupil plane, as shown in Fig. 7.25. The geometrical optical wavefront going through E is denoted as W. The intersects of $P'S_1$ with σ and W are marked as F and G, respectively.

The optical path is written with sufficient accuracy as

$$\wp\left(P'\right) = [S_0G] + [GF] + \left[FP'\right] \tag{7.47}$$

where $[\ldots]$ denotes optical length. The $[S_0G] = [S_0E]$ because G and E lie on the same wavefront W. $\left[FP'\right] = [FS_1] - \left[P'S_1\right] = n_{\text{imsp}}(R - r')$, and R is the radius of σ. Hence,

$$\wp\left(P'\right) = n_{\text{imsp}}\left(C + \Phi - r'\right) \tag{7.48}$$

where $C = \frac{1}{n_{\text{imsp}}}[S_0E] + R$ is a constant, and $\Phi = \frac{1}{n_{\text{imsp}}}[GF]$ is the wave aberration.

The curvature center of geometrical optical wavefront lies in the image space, and thus, $r_1 \approx r_2 \approx r'$, so that $\sqrt{r_1}\sqrt{r_2} \approx r'$. Formula (7.46) can be rewritten as

$$\vec{\boldsymbol{e}}^{(i)}\left(P'\right) = \frac{\vec{\boldsymbol{a}}(s_x, s_y)}{r'} \mathrm{e}^{ikn[C+\Phi(s_x,s_y)-r']} \tag{7.49}$$

Formula (7.49) reveals that both $\vec{\boldsymbol{a}}$ and Φ are functions of the first direction cosines (s_x, s_y) of the light ray going through point P'. The unit vector of light ray is marked as $\vec{\boldsymbol{s}}$ with direction cosine components of s_x, s_y, s_z, which is denoted as $\vec{\boldsymbol{s}}(s_x, s_y, s_z)$, with

$$\begin{cases} s_x = -x/r' = -u_x \\ s_y = -y/r' = -u_y \\ s_z = -z/r\prime = -u_z \end{cases} \tag{7.50}$$

Based on formulas (7.45), (7.49), (7.50) and assuming the medium in image space is air or vacuum with $n_{\text{imsp}} = 1$, one has

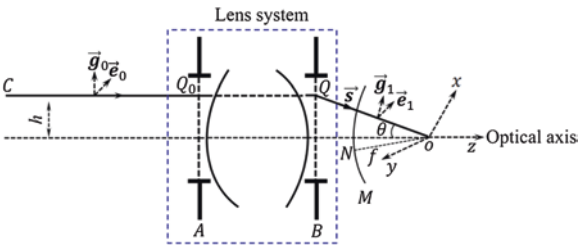

Fig. 7.26 Light ray traveling through a imaging lens system

$$\begin{cases} \vec{\boldsymbol{U}}(s_x, s_y) = \begin{cases} (-)\frac{ik}{2\pi}\frac{\vec{\boldsymbol{a}}(s_x,s_y)}{s_x} e^{ik[\Phi(s_x,s_y)+C]} & \text{for} \quad \vec{\boldsymbol{s}} \in \Omega \\ 0 & \text{for} \quad \vec{\boldsymbol{s}} \notin \Omega \end{cases} \\ \vec{\boldsymbol{V}}(s_x, s_y) \equiv 0 \end{cases} \tag{7.51}$$

where Ω is the solid angle formed by all the $\vec{\boldsymbol{s}}$-vectors associated with the geometrical light rays which reach the image space through the exit pupil plane of the lens. The illustration $\vec{\boldsymbol{s}}$ can be found in Fig. 7.26.

According to formula (7.51), one can obtain the expression for time-independent part of the electric field $\vec{\boldsymbol{e}}[P(x, y, z)]$ in the image space of the optical system as

$$\vec{\boldsymbol{e}}(x, y, z) = -\frac{ik}{2\pi} \iint_{\Omega} \frac{\vec{\boldsymbol{a}}(s_x, s_y)}{s_z} e^{ik[\Phi(s_x,s_y)+s_x x+s_y y+s_z z]} ds_x ds_y \tag{7.52}$$

Formula (7.52) is the expected expression for the optical field in the imaging space (or focusing space) and can be applied in any polarized light beam.

1. **Optical field on the image space using linearly polarized light**

A monochromatic linearly polarized plane incident light beam with a polarization direction along the x-axis travels through the lens system and is focused in the image space. According to formula (7.52), the optical field expression of any point $P(x, y, z)$ in image space of lens system can be analyzed as follows.

Following the work of Richards and Wolf [21], for an aplanatic imaging, the wavefronts in image space are spherical with a common center of Gaussian image point. So for all vectors $\vec{\mathbf{s}}$ in the solid angle Ω

$$\Phi(s_x, s_y) = 0 \tag{7.53}$$

Substituting formula (7.53) into formula (7.52), one has

$$\overrightarrow{\boldsymbol{e}}(x, y, z) = -\frac{ik}{2\pi} \iint_{\Omega} \frac{\vec{\boldsymbol{a}}(s_x, s_y)}{s_z} e^{ik(s_x x+s_y y+s_z z)} ds_x ds_y \tag{7.54}$$

The $\vec{\boldsymbol{a}}$ can be determined by tracing rays through the lens system up to region of the exit pupil and making use of the laws relating to the variation of the field vectors along each light ray. Figure 7.26 presents the light ray traveling through an imaging lens system. A and B are denoted as incident pupil plane and exit

pupil plane, respectively. Taking a typical light ray CQ_0 as an example, CQ_0 goes through the lens at a height h from the optical axis and is imaged to the focal sphere center O, where the focal sphere is denoted as M. The radius of focal sphere is f, which equals to the focal length of lens system. The corresponding light ray of CQ_0 in image space is marked as QO. The angle formed by QO making with optical axis is marked as θ. Since incident optical field in object space is linearly polarized light, the polarization state at plane A is unchanged, and only the polarization direction is different for different light ray. The optical field in the region of the exit pupil plane B is also linearly polarized only if the incident angle at each surface of system is small.

Considering all the light rays going into an annulus bounded by circles of radius h and $h + \delta h$. The area of the annulus is denoted as δS_0, and δS_1 is the area formed in the focal sphere accordingly. $\vec{\boldsymbol{e}}_0$ and $\vec{\boldsymbol{e}}_1$ are the electric amplitude vectors on δS_0, and δS_1, respectively. $\hat{\boldsymbol{e}}_0$ and $\hat{\boldsymbol{e}}_1$ are unit vectors in the $\vec{\boldsymbol{e}}_0$ and $\vec{\boldsymbol{e}}_1$, respectively. One has

$$\begin{cases} \vec{\boldsymbol{e}}_0 = l_0 \hat{\boldsymbol{e}}_0 \mathrm{e}^{ik\wp_0} \\ \vec{\boldsymbol{e}}_1 = l_1 \hat{\boldsymbol{e}}_1 \mathrm{e}^{ik\wp_1} \end{cases} \tag{7.55}$$

where l_0 and l_1 are amplitude factors and $\wp_0$ and $\wp_1$ are the corresponding eikonal values. Since the optical fields in object and image spaces are linearly polarized light, the l_0, l_1, $\hat{\boldsymbol{e}}_0$, and $\hat{\boldsymbol{e}}_1$ can be considered to be real numbers if the origin of phase is suitably chosen.

Assuming that the refractive index of object and image spaces is $n_{\mathrm{obsp}} = n_{\mathrm{imsp}} = 1$, and the energy losses resulting from reflection and absorption of the lens system are ignorable, one can get

$$l_0^2 \delta S_0 = l_1^2 \delta S_0 \tag{7.56}$$

From Fig. 7.26, one has

$$\delta S_0 = \delta S_0 \cos\theta \tag{7.57}$$

Thus, according to formulas (7.56) and (7.57), one gets

$$l_1 = l_0 \sqrt{\cos\theta} \tag{7.58}$$

The $\vec{\boldsymbol{e}}_1$ may be determined with $\boldsymbol{e}^{(i)}$ of formula (7.46) with $r_1 = r_2 = f$, and $\wp = \wp_1$, so

$$\vec{\boldsymbol{a}} = fl_0 \sqrt{\cos\theta} \hat{\boldsymbol{e}}_1 \tag{7.59}$$

To simplify the calculation, the spherical polar coordinates r, θ, and ϕ are introduced, with

$$\begin{cases} r > 0 \\ 0 \le \theta < \pi \\ 0 \le \phi < 2\pi \end{cases} \tag{7.60}$$

The polar axis is in z-direction with $\theta = 0$, and the azimuth angle $\phi = 0$ including the electric vector in object space. Thus, the relation of unit vector $\vec{s}(s_x, s_y, s_z)$ of Cartesian orthogonal coordinate system with $\vec{s}(r = 1, \theta, \phi)$ of polar coordinate system in image space (as marked in Fig. 7.26) is expressed in the form

$$\begin{cases} s_x = \sin\theta\cos\phi \\ s_y = \sin\theta\sin\phi \\ s_z = \cos\theta \end{cases} \tag{7.61}$$

The relation between $P(x, y, x)$ of Cartesian orthogonal coordinate system and $P(r_P, \theta_P, \phi_P)$ of polar coordinate system in image space is accordingly expressed as

$$\begin{aligned} x &= r_P \sin\theta_P \cos\phi_P \\ y &= r_P \sin\theta_P \sin\phi_P \\ z &= r_P \cos\theta_P \end{aligned} \tag{7.62}$$

By complex mathematical operation, Richards and Wolf obtained the strength factor vector $\vec{\boldsymbol{a}}$ of formula (7.54) as

$$\begin{cases} a_x = fl_0\sqrt{\cos\theta}\left[\cos\theta + \sin^2\phi(1 - \cos\theta)\right] \\ a_y = fl_0\sqrt{\cos\theta}[(\cos\theta - 1)\cos\phi\sin\phi] \\ a_z = -fl_0\sqrt{\cos\theta}\sin\theta\cos\phi \end{cases} \tag{7.63}$$

and the $\mathrm{d}s_x\mathrm{d}s_y/\mathrm{d}s_z$ represents the element $\mathrm{d}\Omega$ of the solid angle, which is given as follows

$$\mathrm{d}s_x\mathrm{d}s_y/\mathrm{d}s_z = \mathrm{d}\Omega = \sin\theta\mathrm{d}\theta\mathrm{d}\phi \tag{7.64}$$

Substituting formulas (7.57), (7.63), and (7.64) into formula (7.54), one has

$$\begin{cases} e_x = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi}\left\{\sqrt{\cos\theta}\sin\theta\right\}\left\{\cos\theta + (1 - \cos\theta)\sin^2\phi\right\}\mathrm{e}^{ikr_P\cos\xi}\mathrm{d}\phi\mathrm{d}\theta \\ e_y = i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi}\left\{\sqrt{\cos\theta}\sin\theta\right\}\{(1 - \cos\theta)\cos\phi\sin\phi\}\mathrm{e}^{ikr_P\cos\xi}\mathrm{d}\phi\mathrm{d}\theta \\ e_z = i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi}\left\{\sqrt{\cos\theta}\sin\theta\right\}\{\sin\theta\cos\phi\}\mathrm{e}^{ikr_P\cos\xi}\mathrm{d}\phi\mathrm{d}\theta \end{cases} \tag{7.65}$$

with

$$\cos\xi = \cos\theta\cos\theta_P + \sin\theta\sin\theta_P\cos(\phi - \phi_P) \tag{7.66}$$

For homogenerous plane wave, $l_0 = 1$. ψ is the angular semi-aperture on the image side, i.e., 2ψ is the angle which the diameter of the exit pupil subtends at the geometrical focus. c is constant and can be calculated as

$$c = \pi fl_0/\lambda \tag{7.67}$$

The integration with respect to ϕ can be done by the following formulas, which is valid for any integral value of m,

$$\begin{cases} \int_0^{2\pi} \cos(m\phi) e^{i\varrho\cos(\phi-\delta)} d\phi = 2\pi i^m J_m(\varrho)\cos(m\delta) \\ \int_0^{2\pi} \sin(m\phi) e^{i\varrho\cos(\phi-\delta)} d\phi = 2\pi i^m J_m(\varrho)\sin(m\delta) \end{cases} \tag{7.68}$$

$J_m(\varrho)$ is the Bessel function of the first kind and the order of m. Using $\cos\phi\sin\phi = 1/2\sin(2\phi)$, $\sin^2\phi = 1/2[1-\cos(2\phi)]$, and formula (7.68), the components of field vectors at the point P in the image region are

$$\begin{cases} e_x(P) = -ic[I_0 + I_2\cos(2\phi_P)] \\ e_y(P) = -icI_2\sin(2\phi_P) \\ e_z(P) = -2cI_1\cos\phi_P \end{cases} \tag{7.69}$$

with

$$\begin{cases} I_0 = I_0(kr_P,\ \theta_P,\ \psi) = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(1+\cos\theta) \\ \qquad \left[J_0(kr_P\sin\theta\sin\theta_P)e^{ikr_P\cos\theta\cos\theta_P}\right]d\theta \\ I_1 = I_1(kr_P,\ \theta_P,\ \psi) = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(\sin\theta) \\ \qquad \left[J_1(kr_P\sin\theta\sin\theta_P)e^{ikr_P\cos\theta\cos\theta_P}\right]d\theta \\ I_2 = I_2(kr_P,\ \theta_P,\ \psi) = \int_0^{\psi} \left(\sqrt{\cos\theta}\sin\theta\right)(1-\cos\theta) \\ \qquad \left[J_2(kr_P\sin\theta\sin\theta_P)e^{ikr_P\cos\theta\cos\theta_P}\right]d\theta \end{cases} \tag{7.70}$$

Formulas (7.69) and (7.70) give the electric field of any point $P(r_P, \theta_P, \phi_P)$ in image space.

2. **Optical field on the image space using radially polarized illumination**

Based on the work from Youngworth and Brown [22], for radially polarized beam illumination, the electric field in the image space can be analyzed as follows. A light beam is assumed to have a planar phase front at plane A, passes through an aplanatic lens system, and is focused into a spherical wave at focal sphere M. The spherical wave travels to a diffraction-limited axial point image P in the image plane S (as shown in Fig. 7.27).

The unit vector $\vec{g}_0$ of the incident light beam is perpendicular to the optical axis and can be expressed in cylindrical coordinates

$$\vec{g}_0 = \vec{i}\cos\phi + \vec{j}\sin\phi \tag{7.71}$$

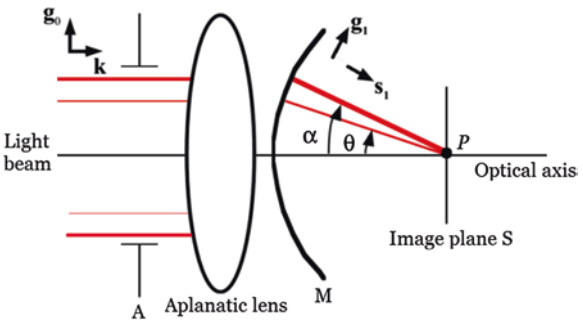

Fig. 7.27 Geometrical relation of radially polarized light focusing

where ϕ is azimuthal angle with respect to the x-axis. $\vec{g}_0$ represents the radial component in object space. The azimuthal component can be denoted as $\vec{g}_0 \times \vec{k}$, where $\vec{k}$ is unit vector along propagation direction (optical axis). The electric field in entrance pupil plane A can be written into radial and azimuthal components as

$$\vec{e}_0 = l_0 \left[\mathrm{e}_r^{(0)} \vec{g_0} + e_\phi^{(0)} (\boldsymbol{g}_0 \times \vec{\boldsymbol{k}}) \right] \tag{7.72}$$

where l_0 is amplitude of incident field and is assumed to vary radially but maintain cylindrical symmetry about optical axis.

According to formulas (7.54) and (7.64), the electric field near focus plane S is diffraction integral over the vector field amplitude $\vec{a}_1$ on a spherical aperture with a focal length f as follows

$$\vec{e}^{(s)} = -\frac{ik}{2\pi} \iint\limits_{\Omega} \vec{a}_1(\theta, \phi) \mathrm{e}^{ik\vec{s}_1 \cdot \vec{r}} \mathrm{d}\Omega \tag{7.73}$$

where the amplitude $\vec{a}_1$ may be related to the object space electric field as follows

$$\vec{a}_1(\theta, \phi) = f\sqrt{\cos\theta} l_0(\theta) \left[e_r^{(0)} \vec{g}_1 + \mathrm{e}_\phi^{(0)} (\vec{g}_1 \times \vec{s}_1) \right] \tag{7.74}$$

where unit vector $\vec{g}_1$ lies in the plane containing both the ray and the optical axis, and is perpendicular to $\vec{s}_1$ traveling direction of the ray in image space. θ is polar angle, and after refraction, the radial unit vector may be expressed as

$$\vec{g}_1 = \cos\theta \left(\vec{i} \cos\phi + \vec{j} \sin\phi \right) + \vec{k} \sin\theta \tag{7.75}$$

In image space, the cylindrical coordinates are employed with $\vec{r}(\rho_s, \phi_s, z_s)$, and the origin located at the paraxial focus is $\rho_s = 0, z_s = 0$. Thus, for points near paraxial focus, one has

$$\vec{s}_1 \cdot \vec{r} = z_s \cos\theta + \rho_s \sin\theta \cos(\phi - \phi_s) \tag{7.76}$$

For radially polarized light, it is trivial to show that the azimuthal component $e_\phi^{(0)} = 0$ of formulas (7.72) and (7.74) everywhere in image space. According to

formulas (7.73–7.76), the Cartesian components of electric field vector near focus may be written as

$$\vec{\boldsymbol{e}}^{(s)} = \begin{cases} \mathrm{e}_x^{(s)} = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi} \sin\theta\sqrt{\cos\theta}\, l_0(\theta)\mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]} \\ \qquad\qquad \cos\theta\cos\phi\mathrm{d}\phi\mathrm{d}\theta \\ \mathrm{e}_y^{(s)} = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi} \sin\theta\sqrt{\cos\theta}\, l_0(\theta)\mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]} \\ \qquad\qquad \cos\theta\sin\phi\mathrm{d}\phi\mathrm{d}\theta \\ \mathrm{e}_z^{(s)} = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi} \sin\theta\sqrt{\cos\theta}\, l_0(\theta)\mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]} \\ \qquad\qquad \sin\theta d\phi d\theta \end{cases} \tag{7.77}$$

The local azimuthal, radial, and longitudinal components of $\vec{\boldsymbol{e}}^{(s)}$ are constructed as:

$$\vec{\boldsymbol{e}}^{(s)} = \begin{cases} \mathrm{e}_\phi^{(s)} = \mathrm{e}_y^{(s)}\cos\phi_s - \mathrm{e}_x^{(s)}\sin\phi_s \\ \mathrm{e}_\rho^{(s)} = \mathrm{e}_x^{(s)}\cos\phi_s + \mathrm{e}_y^{(s)}\sin\phi_s \\ \mathrm{e}_z^{(s)} = i\frac{c}{\pi}\int_0^{\alpha}\int_0^{2\pi}\sin^2\theta\sqrt{\cos\theta}\, l_0(\theta)\mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]}\mathrm{d}\phi\mathrm{d}\theta \end{cases} \tag{7.78}$$

For radially polarized light, the azimuthal component is zero in image space, and from formulas (7.77) and (7.78), the electric field near focus is

$$\vec{\boldsymbol{e}}^{(s)} = \begin{cases} \mathrm{e}_\rho^{(s)} = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi} \sin\theta\cos^{\frac{3}{2}}\theta\cos(\phi-\phi_s) l_0(\theta) \\ \qquad\qquad \mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]}\mathrm{d}\phi\mathrm{d}\theta \\ \mathrm{e}_z^{(s)} = -i\frac{c}{\pi}\int_0^{\psi}\int_0^{2\pi} \sqrt{\cos\theta}\sin^2\theta\, l_0(\theta) \\ \qquad\qquad \mathrm{e}^{ik[z_s\cos\theta+\rho_s\sin\theta\cos(\phi-\phi_s)]}\mathrm{d}\phi\mathrm{d}\theta \end{cases} \tag{7.79}$$

The integration over ϕ can be carried out using the identity:

$$\int_0^{2\pi}\cos(m\phi)\mathrm{e}^{ik\rho_s\sin\theta\cos\phi}\mathrm{d}\phi = 2\pi i^m J_m(k\rho_s\sin\theta) \tag{7.80}$$

$J_m(k\rho_s\sin\theta)$ is a Bessel function of the first kind of order m.

Substituting formula (7.80) into formula (7.79), one can get

$$\begin{cases} \mathrm{e}_\rho^{(s)}(\rho_s, z_s) = c\int_0^{\psi}\sqrt{\cos\theta}\sin(2\theta) l(\theta) J_1(k\rho_s\sin\theta)\mathrm{e}^{ikz_s\cos\theta}\mathrm{d}\theta \\ \mathrm{e}_z^{(s)}(\rho_s, z_s) = 2ic\int_0^{\psi}\sqrt{\cos\theta}\sin^2\theta\, l(\theta) J_0(k\rho_s\sin\theta)\mathrm{e}^{ikz_s\cos\theta}\mathrm{d}\theta \end{cases} \tag{7.81}$$

$l(\theta)$ is apodization function by using an appropriate pupil filter, for example, for Bessel–Gauss beam whose waist is at plane A, $l(\theta)$ can be expressed as,

$$l(\theta) = exp\left[-\kappa^2\left(\frac{\sin\theta}{\sin\alpha}\right)^2\right] J_1\left(2\kappa\frac{\sin\theta}{\sin\alpha}\right) \tag{7.82}$$

where κ is the ratio of pupil radius to the beam waist. c is set to be 1, $J_0(x)$ and $J_1(x)$ denote the Bessel function of the zero and first orders of the first kind, respectively.

References

1. G.T. Di Francia, Super-gain antennas and optical resolving power. Nuovo Cimento Suppl. **9**, 426–435 (1952)
2. K. Huang, P. Shi, X. Kang, X. Zhang, Y. Li, Design of DOE for generating a needle of strong longitudinally polarized field. Opt. Lett. **35**, 965-967 (2010)
3. H. Wang, L. Shi, G. Yuan, X. Miao, W. Tan, T. Chong, Subwavelength and super-resolution nondiffraction beam. Appl. Phys. Lett. **89**, 171102 (2006)
4. C. Kuang, X. Hao, X. Liu, T. Wang, Y. Ku, Formation of sub-half-wavelength focal spot with ultra long depth of focus. Opt. Commun. **284**, 1766–1769 (2011)
5. H. Wang, L. Shi, B. Lukyanchuk, C. Sheppard, C.T. Chong, Creation of a needle of longitudinally polarized light in vacuum using binary optics. Nat. Photon. **2**, 501–505 (2008)
6. Y. Zha, J. Wei, F. Gan, Optimized design of four-zone phase pupil filter for nanoscale phase transition optical lithography. Proc. SPIE **8782**, 878206 (2013)
7. Y. Zha, J. Wei, F. Gan, Performance improvement of super-resolving pupil filters via combination with nonlinear saturable absorption films. Opt. Commun. **293**, 139–148 (2013)
8. J. Wei, Y. Zha, F. Gan, Creation of super-Resolution non-diffraction beam by modulating circularly polarized light with ternary optical element. Prog. Electromagnet. Res. (PIER) **140**, 589–598 (2013)
9. Y. Zha, *The DOE design based on phase retrieval algorithm and tight focusing beam reshaping*. A Dissertation of University of Chinese Academy of Sciences, p. 61 (2013)
10. Y. Zha, J. Wei, F. Gan, Creation of ultra-long depth of focus super-resolution longitudinally polarized beam with ternary optical element. J. Opt. **15**, 075703 (2013)
11. G. Lerman, U. Levy, Effect of radial polarization and apodization on spot size under tight focusing conditions. Opt. Exp. **16**, 4567–4581 (2008)
12. W. Kim, N. Park, Y. Yoon, H. Choi, Y. Park, of near-field imaging characteristics of radial polarization for application to optical data storage. Opt. Rev. **14**, 236–242 (2007)
13. T. Grosjean, D. Courjon, C. Bainier, Smallest lithographic marks generated by optical focusing systems. Opt. Lett. **32**, 976–978 (2007)
14. S. Liu, J. Wei, F. Gan, Nonlinear absorption of Sb-based phase change materials due to the weakening of the resonant bond. Appl. Phys. Lett. **100**, 111903 (2012)
15. Y. Zha, J. Wei, F. Gan, A novel design for maskless direct laser writing nanolithography: combination of diffractive optical element and nonlinear absorption inorganic resists. Opt. Commun. **304**, 49–53 (2013)
16. B. Liu, F. Gan, Optical properties and crystallization of AgInSbTe phase change optical disk media. Appl. Phys. A **77**, 905–909 (2003)
17. J. Liu, J. Wei, Optical nonlinear absorption characteristics of AgInSbTe phase change thin films. J. Appl. Phys. **106**, 083112 (2009)
18. M. Born, E. Wolf, *Principles of Optics* (Cambridge University, Cambridge, 1993)

19. J. Wei, M. Xiao, Laser tunable Toraldo super-resolution with a uniform nonlinear pupil filter. Appl. Opt. **47**, 3689–3693 (2008)
20. E. Wolf, Electromagnetic diffraction in optical systems I: an integral representation of the image field. Proc. R. Soc. London Ser. A **253**, 349–357 (1959)
21. B. Richards, E. Wolf, Electromagnetic diffraction in optical systems II: structure of the image field in an aplanatic system. Proc. R. Soc. London Ser. A **253**, 358–379 (1959)
22. K.S. Youngworth, T.G. Brown, Focusing of high numerical aperture cylindrical vector beams. Opt. Exp. **7**, 77–87 (2000)

Chapter 8
Applications of Nonlinear Super-Resolution Thin Films in Nano-optical Data Storage

8.1 Development Trend for Optical Information Storage

Optical storage is very useful for high-definition digital television, data distribution, and memory. With the development of information technology, the capacity of optical storage devices is required to become larger and larger. Correspondingly, the density is also required to become higher and higher. In order to improve the density, one has to reduce the information bit size. The size is determined by the optical diffraction limit. That is, $D \sim 1.22\lambda/\mathrm{NA}$, where D is the size of the focused spot, λ laser wavelength in vacuum, and NA the numerical aperture of optical pickup. In order to improve the capacity and density of optical storage devices, one has to shorten the laser wavelength and increase the numerical aperture of optical pickup. Table 8.1 gives the development of optical storage devices [1].

However, a larger NA will reduce the focal depth of pickup and increase comatic aberration when the disk tilts against the optical axis. So far, for far-field optical storage, the NA of pickup has been very close to the theoretical limit (NA = 1). To reduce the mark size, a short wavelength laser source must be used. The laser wavelength has been decreased from 780 nm for compact discs to 405 nm for blue ray disks. Further shortening the laser wavelength is difficult due to the high product cost and low transmission of general optical elements at wavelength below 360 nm.

In this regard, the near-field optical probe is a very good method to generate small spot due to avoidance of the diffraction limit, and it can also record nanometric information marks. However, the low optical transmission and some problems in controlling the near-field distance between the probe tip and the sample surface meet difficulties in application to the high-speed rotation optical disk system. The solid immersion lens near-field method can obtain a subwavelength spot and produce nanometric information marks. It can also maintain a higher optical

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_8

Table 8.1 The development of optical storage devices (optical disks)

	CD	DVD	HD-DVD	BD
Laser wavelength (nm)	780	650	405	405
Numerical aperture	0.45	0.60	0.60	0.85
Size of information bit (nm)	830	400	250	150
Capacity of a 5-inch disk	650 MB	4.7 GB	15 GB	27 GB

CD compact disc, DVD digital versatile disk, HD high density, BD blue ray disk

transmission compared with the near-field probe technique. However, the problem of near-field distance control between the lens surface and the sample still remains.

Nonlinear super-resolution optical information storage has undergone considerable development in the past two decades. Advances include the ability to break through the optical diffraction limit as well as the capability to realize the recording and readout of below-diffraction-limited information marks [2–4]. In this chapter, the basic principle and important experimental results of nonlinear super-resolution-induced high-density data storage are presented and analyzed.

8.2 Saturation Absorption-Induced High-Density Optical Data Storage

8.2.1 Read-Only Super-Resolution Optical Disk Storage

8.2.1.1 Principle of Super-Resolution Readout

A scanning spot with an electric field distribution $E_0(r)$ (shown in Fig. 8.1) is incident onto the nonlinear saturation absorption thin film. Let the nonlinear saturation

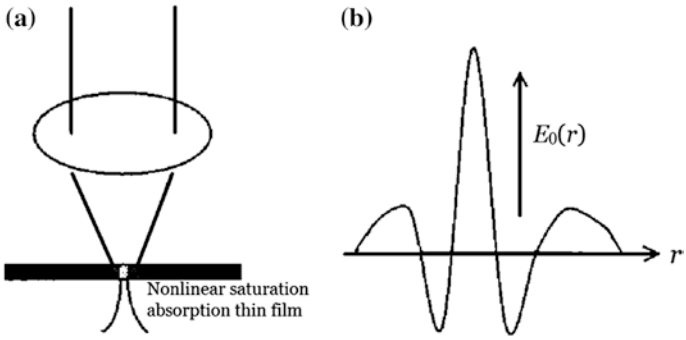

Fig. 8.1 Super-resolution principle of nonlinear saturation absorption thin films: **a** illustration of focused spot-induced optical pinhole, and **b** electric field distribution

absorption thin film function be $O(r)$ and its amplitude transmittance be $t(r)O(r)$. The $t(r)$ is nonlinear transmittance and can be expressed as [5, 6]

$$t(r) = t_0\left[1 + \chi \cdot g(E_0)\right] \tag{8.1}$$

where the $g(E_0)$ represents the dependence on the incident light. $g(E_0)$ equals 0 if $E_0(r) = 0$ and increases with $E_0(r)$ up to material destruction. The nature of the nonlinearity is characterized by χ, $\chi > 0$ corresponds to bleaching, and $\chi < 0$ corresponds to blackening. t_0 is the standard linear transmittance of the nonlinear saturation absorption thin films. The transmitted field immediately after the nonlinear thin films can be written as

$$E_{t,0}(r) = E_t(r) \cdot O(r) \tag{8.2}$$

with

$$E_t(r) = t(r) \cdot E_0(r) \tag{8.3}$$

Signal from the nonlinear saturation absorption thin film may be found from the Fourier transform of $E_{t,0}(r)$. One integrates the signal over a detection pupil $P(\rho)$, which corresponds to the same NA as the spot-forming pupil. Resolution can easily be judged with the aid of the frequency response function $R_{\mathrm{Res}}(f)$. Since in spatial frequency considerations, the nonlinear saturation absorption thin film is considered to be a grating with discrete orders, in the case of gratings in a nonlinear material, the spot narrowing yields an aperture widening after passing. Widely opened diffracted beams overlap the detection pupil up to larger diffraction angles; thus, interference between zero and first orders is possible up to higher spatial frequencies. Since only the detected region has to be taken into account, the response $R_{\mathrm{Res}}(f)$ is given by the convolution of the field $\tilde{E}_t(\rho)$ in the detection pupil with the field itself:

$$R_{\mathrm{Res}}(f) \propto \left|\left[\tilde{E}_t(\rho) \cdot P(\rho)\right] \otimes \tilde{E}_t(\rho)\right| \tag{8.4}$$

where $\tilde{E}_t(\rho)$ is the Fourier transform of $E_t(r)$ and $P(\rho) = 1$ if $\rho < \mathrm{NA}/\lambda$, otherwise 0. Let $G(\rho)$ represent the Fourier transform of g with respect to r. Formulas (8.1) and (8.3) then lead to

$$\tilde{E}_t(\rho) = t_0\left[\tilde{E}_0(\rho) + \chi G(\rho) \otimes \tilde{E}_0(\rho)\right] \tag{8.5}$$

where $\tilde{E}_0(\rho) = 1$ if $\rho \leq \mathrm{NA}/\lambda$, otherwise 0. $\tilde{E}_0(\rho)$ corresponds to the conventional linear case. With formula (8.4), it yields the modulated transfer function. $\left[\chi G(\rho) \otimes \tilde{E}_0(\rho)\right]$ shows the influence of the nonlinearity. Because the convolution covers a wider field than $\tilde{E}_0(\rho)$, the nonlinear part of $\tilde{E}_t(\rho)$ will extend beyond the value NA/λ. This is the beam widening due to the spot narrowing. From formula (8.4), it then follows that $R_{\mathrm{Res}}(f)$ will show a finite value above the conventional cutoff frequency f_c; thus, there will be super-resolution.

8.2.1.2 Dynamic Readout of Super-Resolution Pit Arrays Using Sb as Nonlinear Thin Film Material

Sb thin films possess excellent optical super-resolution effect and are good candidates for high-density data storage [7, 8]. The pit arrays premastered on the optical disk substrates with a diameter of 120 mm and a thickness of 1.2 mm are shown in Fig. 8.2, the refractive index of this substrate is approximately to 1.52 at visible frequencies, and the pit depth and diameter (space) are 50 and 380 nm, respectively. It can be seen that the surface shape of pit arrays has a sinusoidal distribution. Sb thin films are directly deposited on the optical disk substrates by sputtering method, and the thickness is controlled in the range from 10 to 100 nm.

Figure 8.3 gives the schematic of dynamic testing setup for optical disks, where the laser wavelength λ and the numerical aperture (NA) of converging lens are 633 and, 0.40 nm respectively.

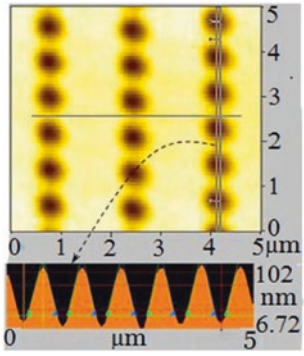

Fig. 8.2 AFM analysis of pit arrays on the optical disk substrate

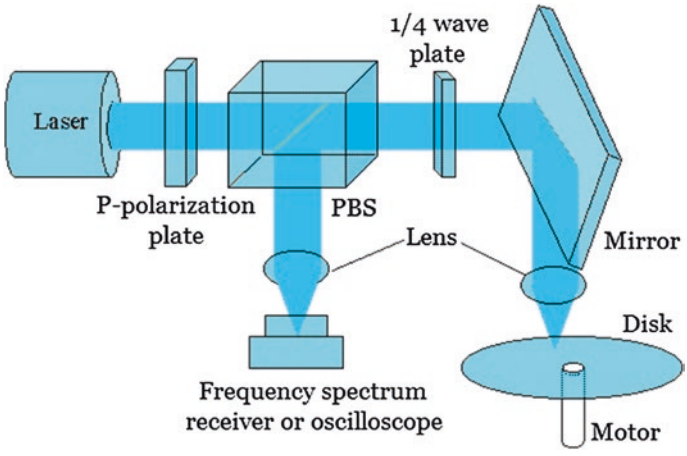

Fig. 8.3 Dynamic readout setup for the pit arrays on optical disks

The laser passes through p-polarization plate and goes into polarized beam splitter (PBS) and 1/4 wave plate and then is reflected to the lens and focused onto the disk surface. The focused spot interacts with the nonlinear thin film and then is reflected into the spectrum frequency receiver and oscilloscope. The diameter of optical spot on the optical disk can be calculated by $D \sim 1.22\lambda/\mathrm{NA}$ and is ~1.93 um, which is much larger than the pit diameter of ~380 nm. The dynamic readout resolution limit can be approximately calculated by $\lambda/4\mathrm{NA}$ and is ~396 nm, so the pit arrays on the optical disk substrates are smaller than the readout resolution limit and cannot be dynamically read out by conventional reflective layer.

In order to confirm the spatial resolution of the dynamic readout setup shown in Fig. 8.3, the disk structure "Al reflective layer (150 nm)/premastered disk substrate" is measured, and the results are shown in Fig. 8.4a, where the pit arrays onto the disk substrates cannot be read out. Then, the Al thin film is replaced with Sb thin film with a thickness of 28 nm, and the Sb thin film possesses large nonlinear saturation absorption characteristics and is used as nonlinear super-resolution layer. The dynamic readout results are shown in Fig. 8.4b. From the Fig. 8.4b, the pit arrays are read out and the carrier-to-noise ratio (CNR) is 38 dB.

In addition, Fig. 8.5 shows the CNR dependence on the Sb thin film thickness. With the increase of the Sb film thickness, the CNR increases, and when the thickness is 28–30 nm, the CNR is the maximum of ~40 dB. While with the further increase of the thickness, the CNR decreases. The CNR is close to 0 when the thin film thickness is more than 50 nm.

In the focused beam traveling through the nonlinear saturation absorption thin films, one aspect is the spot size reduction due to nonlinear saturation absorption effect, and the other is the spot size expansion due to diffraction effect. It is well known that the dynamic readout resolving power and CNR are reversely proportional to the spot size. The dependence of CNR on thin film thickness is actually

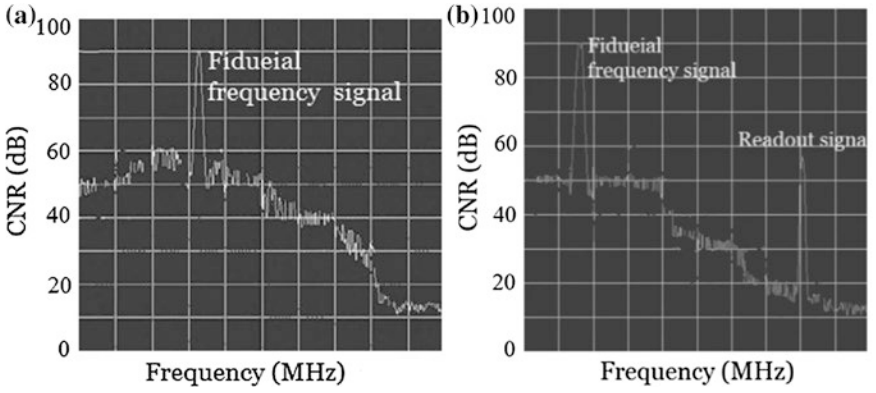

Fig. 8.4 Dynamic readout signals, **a** conventional read-only disk (Al thin film used as reflective layer, thickness: 150 nm, **b** super-resolution disk (Sb thin film used as nonlinear super-resolution layer, thickness: 28 nm). Reprinted from [9], with kind permission from Springer Science+Business Media

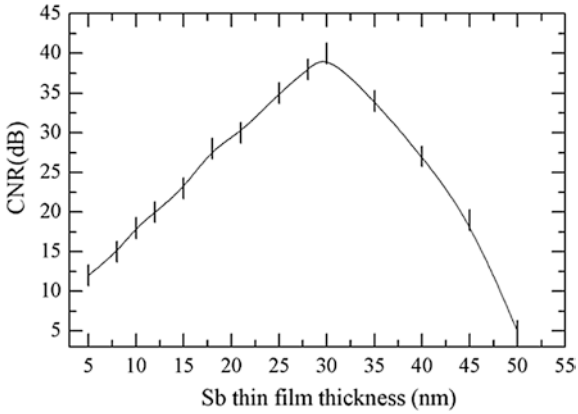

Fig. 8.5 Dependence of CNR on thin film thickness. Reprinted from [9], with kind permission from Springer Science+Business Media

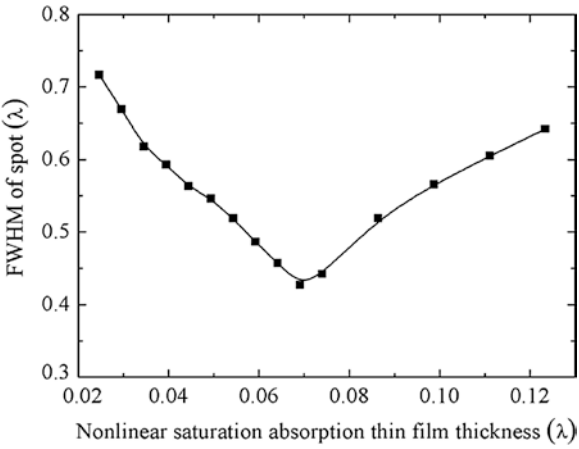

Fig. 8.6 Dependence of light spot size on thin film thickness

determined by the spot size. The spot size changes with thin film thickness are simulated by COMSOL multi-physical field software. In this simulation, the linear and nonlinear saturation absorption coefficients are chosen as $\alpha_0 = 7.53 \times 10^7/\text{m}$ and $\beta = -0.608 \times 10^{-2}\,\text{m/W}$, respectively. The linear refractive index is $n_0 = 3.1$, and the NA of converging lens is 0.90. The focused spot intensity is $I_0 = 5.1 \times 10^9\,\text{W/m}^2$.

Figure 8.6 shows the relationship of the FWHM of light spot from the exiting surface of Sb thin film with the thickness [10]. One can see that the FWHM decreases first from ~0.716λ down to ~0.43λ and then increases up to ~0.642λ with the increase of the film thickness. That is, the size of transmitted light spot cannot decrease continuously with the increase of the film thickness, there is an optimum thin film thickness at $L \sim 0.07\lambda$ (L is thin film thickness), and the minimum FWHM is about ~0.42 λ accordingly. This is because that though

inhomogeneous absorption can reshape the transmitted light spot into below diffraction limit, the effect of nonlinear saturation absorption gets weaker with the increase of thickness due to the light intensity decaying with the thickness increasing. Finally, the effect of diffraction is stronger than that of nonlinear saturation absorption and the size of transmitted light spot increases and becomes larger than that at the optimum thin film thickness.

In order to further understand the light beam reshaping effect by the nonlinear saturation absorption thin film, the light beam intensity distribution along the thickness direction is simulated. Figure 8.7 gives the simulated results, which are normalized along thin film thickness direction [10]. One can find that a laser beam propagates inside nonlinear saturation absorption thin film, the laser beam is reshaped, and the light spot becomes smaller than incident light due to the reshaping effect. When the laser beam propagates to the thickness of about 0.074 λ, the light spot becomes the smallest value because the reshaping effect reaches a balance with diffraction effect. Then, the spot becomes larger again due to the diffraction effect being stronger than the reshaping effect of nonlinear saturation absorption, the spot size increases with the increase in the propagation distance, and the FWHM of spot reaches up to about 0.815λ at the thickness of 0.247λ, which is in agreement with the results of Fig. 8.6.

Generally speaking, the dynamic readout resolution of optical disk is reversely proportional to the light spot size, and the smaller the spot size is, the higher the readout resolution is. The readout resolution can be defined as $R_{resolving} \propto 1/FWHM$. Figure 8.8 shows the dependence of the normalized $R_{resolving}$ on thin film thickness [10], where the dependence of CNR on thin film thickness is also given. The readout resolution increases to the maximum at $L \sim 0.07\lambda$ and then decreases with the increase in film thickness. This dependence of resolution on thin film thickness is consistent with the previous experimental results of CNR on the film thickness (shown in red line).

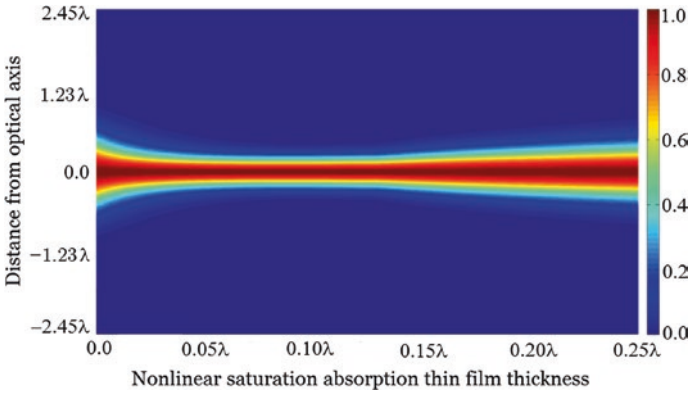

Fig. 8.7 Normalized spot intensity along the thin film thickness direction

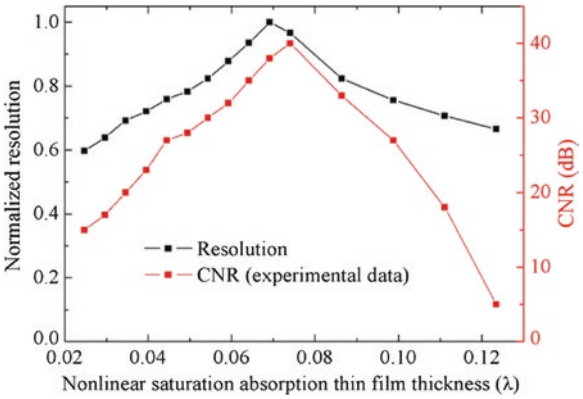

Fig. 8.8 Dependence of optical resolution and CNR on the thin film thickness

8.2.2 Recordable Super-Resolution Nano-optical Storage

In the nonlinear super-resolution nano-optical data storage, the sample structure is designed as "substrate/dielectric layer/saturation absorption layer/near-field layer/recording layer/signal enhancement layer," as is presented in Fig. 8.9, and the working principle is described as follows.

In the process of nano-optical information recording, the higher power Gaussian laser beam spot is incident from the substrate side and passes through the substrate and dielectric layer. It is focused on the nonlinear layer. The

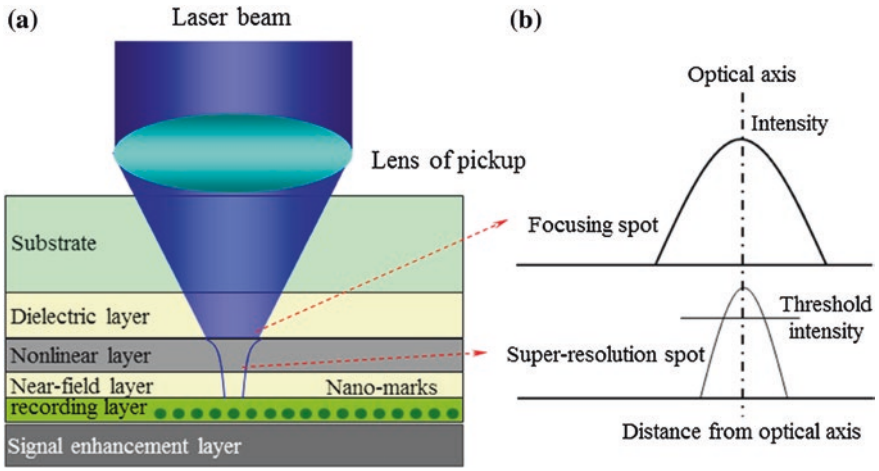

Fig. 8.9 Schematic and principle of nano-optical information storage: **a** sample structure and **b** optical spot intensity profile. Reproduced from [11] by permission from The Royal Society of Chemistry

nonlinear layer possesses an optical saturation absorption characteristic, and the higher the laser intensity is, the larger the transmittance is. The largest transmittance is at the central part of the spot, which causes a small channel to be formed at the center of the spot (shown in Fig. 8.9a). The laser beam goes through the channel, and a below-diffraction-limited optical spot (named as super-resolution spot) is generated at the bottom surface of the nonlinear layer. The intensity of the super-resolution spot approximately maintains Gaussian profile (shown in Fig. 8.9b). Then, the energy of the super-resolution spot is directly coupled into the recording layer in the near-field distance. The recording layer absorbs the laser energy and is heated. When the energy exceeds certain threshold, the recording layer at the central region of the super-resolution spot is decomposed and ablated into an information mark. The mark has a different reflectance from the original recording material itself. The mark size is smaller than the super-resolution spot, and it is far smaller than the focused spot itself, which realizes the nano-optical information recording.

In the process of the readout, the lower power Gaussian laser beam is focused on the nonlinear layer. Similar to the recording procedure, the super-resolution spot is formed at the bottom surface of the nonlinear layer. The super-resolution spot passes through the near-field distance-controlling layer and is incident on the recording layer. In this layer, there is strong reflectance contrast between the recording and the non-recording marks. The super-resolution spot is smaller than the focused spot and can detect nanometric marks. The super-resolution spot takes the signal of information marks and continues to go to the signal enhancement layer. This layer is designed to have high reflectance. The super-resolution spot is finally reflected by this signal enhancement layer and collected by the optical pickup.

In the nonlinear super-resolution nano-optical data storage, the key factor is the use of optical nonlinear layer with a thickness of tens of nanometers. When the focused beam spot irradiates on the nonlinear layer, a super-resolution optical aperture is produced at the central region of the spot. Subsequently, the focused beam spot passes through the super-resolution optical aperture, which produces a below-diffraction-limited spot (super-resolution spot) at the back side of the nonlinear layer. The production of super-resolution optical aperture is dynamic and reversible, wherein the aperture can be reversibly open or closed when the laser beam is on or off. Nanoscale pits are detected and dynamically read out by this super-resolution spot.

8.2.2.1 Experimental Results of the Nano-optical Data Recording and Readout

The sample "substrate/dielectric layer/nonlinear layer/near-field layer/recording layer/signal enhancement layer" (shown in Fig. 8.9) is prepared. Polycarbonate materials with a thickness of 0.6 mm and with premastered grooves are used as substrates. The 120 nm SiN thin film is deposited on the substrate through the

radio frequency magnetron-controlling sputtering method and served as the dielectric layer. Sb_2Te_3 with a thickness of approximately 20 nm is subsequently deposited on the dielectric layer through a direct-current magnetron-controlling sputtering method. Sb_2Te_3 is used as the nonlinear layer because the resonant bonding weakening induced giant nonlinear saturation absorption characteristics [12, 13]. A 20 nm-thick SiN is used as the near-field layer between Sb_2Te_3 and the recording layer. The dye is used as the information recording layer [14, 15]. Finally, the metal reflective layer is deposited directly on the recording layer to function as signal enhancement. Information recording and readout are carried out by a high-density digital versatile disk testing system, which has a laser wavelength λ of 405 nm and a numerical aperture (NA) of 0.65. The dynamic readout resolving limit $\lambda/4NA$ of this testing system is approximately 156 nm.

In the dynamic recording and readout, the rotation speed of the sample is fixed at 3 m/s, and the duty ratio of the laser pulse is set at 50 %. The recording marks are immediately read out by another laser beam after recording. The dynamic readout signal is obtained by an optical–electronic detector and then fed into a frequency spectrum analyzer. Generally speaking, the dynamic readout signal can be detected only after the information recording is completed. This means that if there is a readout signal, there must be from the information recording marks on the sample.

The dynamic readout signal does not only reflect the information mark being detected, but also means that the information is recorded. Figure 8.10 shows a typical frequency spectrum signal of the recording marks with a size of 75 nm. The recording and readout powers are 5.0 and 1.4 mW, respectively. The CNR is about 23 dB. This recording mark size is much smaller than the diffraction-limited focused spot itself of the testing system, whose spot size is theoretically about 760 nm.

The dynamic readout resolution limit of the testing system is theoretically 156 nm ($d = \lambda/4\text{NA}$), and the readout marks of 75 nm are smaller than the resolution limit. Therefore, based on the traditional optical diffraction limit, dynamically resolving marks with a size below 156 nm are impossible. However, the experimental results indicate that the recording and readout of nanoscale information marks

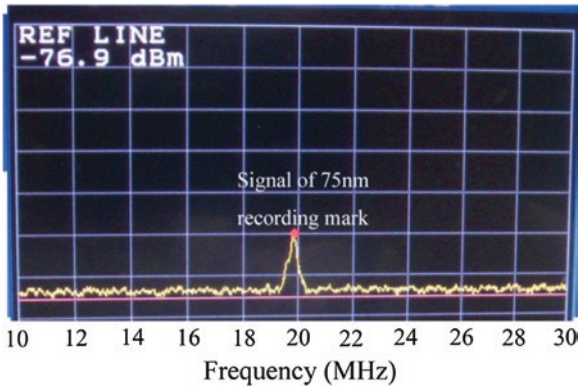

Fig. 8.10 Frequency spectrum signal of the dynamic readout for recording marks with a size of 75 nm. (The mark size can be calculated by $v/2f$, where v is rotation speed of the sample, $v = 3$ m/s in this experiment, and f is the frequency.) Reproduced from [11] by permission from The Royal Society of Chemistry

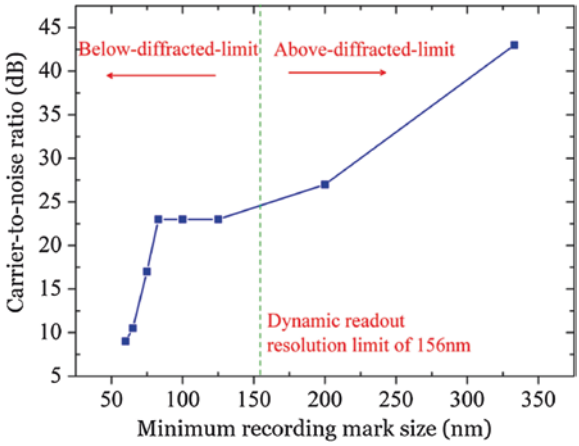

Fig. 8.11 Relation of the CNR with the recording mark size. Reproduced from [11] by permission from The Royal Society of Chemistry

are realized experimentally by using nonlinear thin films, which also verify that the sample structure, where Sb_2Te_3 is used as the nonlinear saturation absorption layer, dye as the information recording layer, and Ag as the signal enhancement layer, is a good experimental schematic for realizing nano-optical data storage.

To verify further the ability to reduce information marks, Fig. 8.11 presents the dependence of the CNR on the mark size by optimizing the recording and readout laser powers. The minimum marks of 60 nm are recorded and read out with a CNR of 8 dB, which is about 1/12 of the diffraction-limited focused spot. The dynamic readout resolving power is almost 1/3 of the resolution limit of the testing system. These indicate that the nonlinear saturation absorption thin film has excellent super-resolution capability. The CNR increases rapidly with the increase in the mark size in the range of 60–80 nm and then increases slowly when the mark size exceeds 80 nm, which may be due to the nonlinear saturation absorption characteristic of the Sb_2Te_3 thin films.

In the dynamic readout, the CNR depends on the readout laser power. Figure 8.12 shows the experimental results for marks with a size of 100 nm, which are obtained at the recording power of 5.0 mW. One can see that there is a threshold readout power of 0.5 mW. This is due to the threshold effect of the nonlinear saturation absorption of the Sb_2Te_3 thin films; that is, the nonlinear saturation absorption does not occur when the laser power is below 0.5 mW, and the super-resolution readout spot cannot be formed accordingly. The nonlinear saturation absorption occurs and increases with the readout laser power of more than 0.5 mW, which leads to the formation of the super-resolution readout spot. The size of the super-resolution readout spot is reduced with an increase in laser power, which in turn results in an increase in the CNR. When the readout laser power exceeds 1.6 mW, the CNR decreases quickly, which can be due to the destruction of recording marks because of their low melting temperature of about 300 °C [15].

There is an optimal readout power for marks with certain size. For example, as shown in Fig. 8.12, the optimal readout power for the recording marks of 100 nm

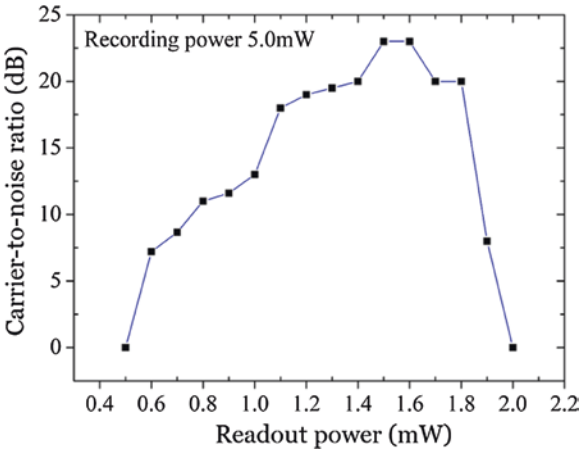

Fig. 8.12 Dependence of the CNR of the dynamic readout on the readout power for recording marks with a size of 100 nm. Reproduced from [11] by permission from The Royal Society of Chemistry

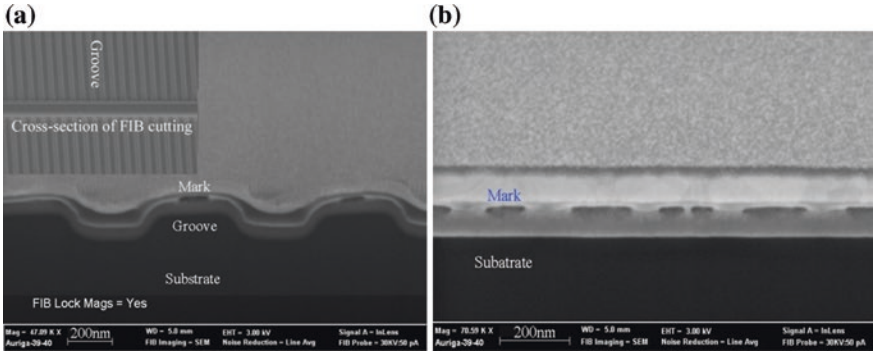

Fig. 8.13 SEM analyses of the recording marks with a size of 100 nm: **a** along the perpendicular direction with the groove and **b** along the direction of the groove. Reproduced from [11] by permission from The Royal Society of Chemistry

is approximately 1.5 mW. To observe directly the recording marks, scanning electronic microscopy (SEM) analysis of the sample is carried out. Before the observation, a platinum thin film is deposited on the sample surface, and a focused ion beam is used to cut the sample along the perpendicular direction with the recording groove, as shown in the inset of Fig. 8.13a. Observation *in situ* is conducted for the cross section of the recording groove by thermal field emission scanning electronic microscopy. Figure 8.13a presents the experimental results. It can be found that the interfaces among the layers of the sample structure cannot be distinguished clearly due to the resolution restriction of SEM itself. However, the marks can be seen very clearly. The marks are recorded in the middle layer (dye layer) of the sample structure, and the width size of the marks is about 100 nm. The length of the marks is also observed by cutting the sample along the direction

of the groove. Figure 8.13b presents the experimental results. A row of marks with a length of about 100 nm is observed clearly in the middle layer of the sample structure. Figure 8.13 directly verifies that information marks of about 100 nm are recorded, which further indicates that nano-optical information storage is realized by the sample structure shown in Fig. 8.9.

Figures 8.11 and 8.12 present the experimental results on the dynamic readout of nonlinear super-resolution optical storage devices for the recording marks with a size of 100 nm. In order to further understand the dynamic readout characteristic, Fig. 8.14a gives experimental results for the marks of 75 nm. Figure 8.14a illustrates the dependence of the CNR on the readout laser power. A dynamic readout threshold power of approximately 0.9 mW could be observed at the CNR of 5 dB. The CNR increases with the readout power and remains unchanged with the continuous increase of the readout power. Finally, the CNR falls to zero when the readout power exceeds some threshold. These trends are consistent with the marks of 100 nm (shown in Fig. 8.12).

The dependence of the readout threshold power on the mark size is tested, where the 3–5 dB of CNR as the readout threshold signal is defined. Figure 8.14b presents the results. Absence of readout threshold power is observed when the mark size is larger than the resolving limit. For marks, whose size is smaller than the resolving limit, the threshold power increases with the reduction in mark size. According to the experimental results of the dynamic readout testing, some important questions need to be understood:

1. What can account for the presence of a dynamic readout threshold power in nonlinear super-resolution optical storage?
2. Why does the CNR increase at the initial stage and then decrease with the increase of readout laser power or laser irradiation time?
3. Why does the dynamic readout threshold power increase with the reduction in mark size?

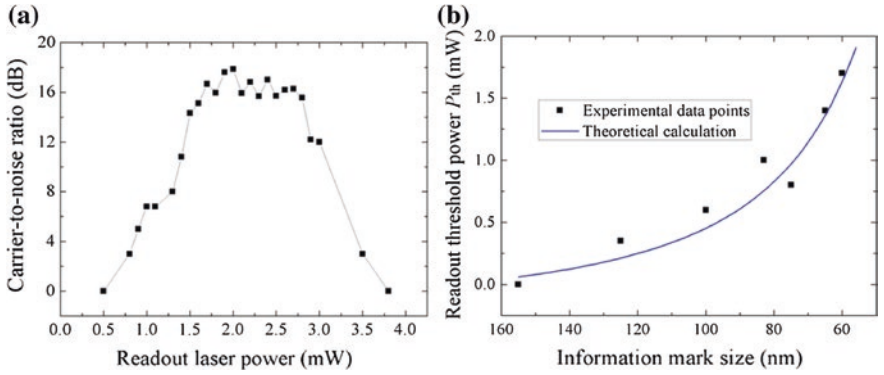

Fig. 8.14 Dynamic readout experimental results: **a** dependence of CNR on laser power for marks of 75 nm and **b** relation of readout threshold power with mark size. Reprinted from [16], Copyright 2013, with permission from Elsevier

8.2.2.2 Theoretical Analysis and Calculated Results

1. Theoretical basis of super-resolution spot formation

For thin film samples irradiated by a laser beam with an initial intensity profile of $I_{\text{inc}}(r)$, the laser intensity decaying can be expressed as follows:

$$\mathrm{d}I_{\text{inc}}/\mathrm{d}z' = -\alpha I_{\text{inc}} \tag{8.6}$$

where α is the absorption coefficient of the sample and $\mathrm{d}I_{\text{inc}}$ is the variation of the laser intensity along the sample thickness direction $\mathrm{d}z'$.

The transmitted laser beam spot intensity can be obtained by integrating formula (8.6) as follows:

$$I_t(r) = I_{\text{inc}}(r)\exp(-\alpha L) \tag{8.7}$$

Formula (8.7) is the Beer–Lamber formula. For general samples, $\alpha = \alpha_0$ is constant, where α_0 is the linear absorption coefficient. The normalized transmitted laser beam spot intensity profile, $I_t(r)$, is spatially similar to the initial incident laser intensity, $I_{\text{inc}}(r)$. However, the α changes with laser intensity for the nonlinear absorption sample. The α can be written as

$$\alpha(r) = \alpha_0 + \beta I_{\text{inc}}(r) \tag{8.8}$$

where β is the nonlinear absorption coefficient. Formula (8.8) shows that the absorption coefficient presents the similar spatial profile to the incident laser beam intensity for nonlinear absorption samples.

The laser beam spot passing through the nonlinear thin film sample can be calculated by substituting formula (8.8) into formula (8.7) to obtain

$$I_t(r) = I_{\text{inc}}(r)\exp\{-[\alpha_0 + \beta I_{\text{inc}}(r)]L\} \tag{8.9}$$

Generally, the initial incident laser beam intensity presents a Gaussian profile and can be written as

$$I_{\text{inc}}(r) = I_0\exp\left(-2r^2/w_0^2\right) \tag{8.10}$$

where I_0 is the intensity at the central point of spot $r = 0$ and w_0 is the spot radius. Substituting formula (8.10) into formula (8.9) yields the transmitted laser beam spot intensity

$$I_t(r) = I_0\exp\left(-\frac{2r^2}{w_0^2}\right)\exp\left\{-\left[\alpha_0 + \beta I_0\exp\left(-\frac{2r^2}{w_0^2}\right)\right]L\right\} \tag{8.11}$$

For nonlinear saturation absorption thin film samples, the intensity at $r = 0$ is the strongest, and the normalized transmitted laser beam spot is derived as

$$I_{\text{trans.normalized}}(r) = \frac{I_t(r)}{I_t(0)} = \frac{\exp\left(-\frac{2r^2}{w_0^2}\right)\exp\left[-\beta I_0\exp\left(-\frac{2r^2}{w_0^2}\right)L\right]}{\exp(-\beta I_0 L)} \tag{8.12}$$

where $I_{\text{trans.normalized}}(r)$ is a normalized spatial intensity profile of the transmitted spot passing through the nonlinear absorption thin films. This transmitted spot is exactly the super-resolution spot if nonlinear saturation absorption thin film material is used as the nonlinear layer. Formula (8.12) shows that the size of the super-resolution spot is determined by nonlinear absorption coefficient β, thin film thickness L, initial incident spot radius w_0, and laser power density I_0.

2. Theoretical analysis of dependence of threshold readout power on mark size

The experimental results of the nano-optical data recording and readout show that the dynamic readout of nanometric marks results from nonlinear super-resolution spot at the back side of nonlinear layer. The smaller the super-resolution spot, the stronger the ability to detect the nanometric marks is.

(a) Theoretical analysis of the dependence of readout threshold power on mark size

In optical information storage devices, the dynamic readout resolving limit of the testing system is generally defined as $d_{\text{resolving}} = \lambda/4\text{NA}$. Theoretically, the spot size is $D = 1.22\lambda/\text{NA}$, and accordingly, the spot radius is $w_0 = 0.61\lambda/\text{NA}$. $d_{\text{resolving}}$ shows that the readout resolving limit is approximately 0.2 times the spot size ($d_{\text{resolving}} \sim D/4.8$). Setting $r = d_{\text{resolving}}/2 \sim w_0/4.8$ and substituting it into formula (8.10) yield a normalized intensity value of approximately 0.90. In other words, the resolving limit size $d_{\text{resolving}} = \lambda/4\text{NA}$ of the testing system corresponds to the spot position where the normalized intensity value is approximately 0.90.

In nonlinear readout, the spot is exactly the super-resolution spot. D should be the super-resolution spot size. Thus, the $D/4.8$ super-resolution spot size is the dynamic readout resolving limit $d_{\text{resolving}}$, which corresponds to $I_{\text{trans.normalized}}(r) \sim 0.90$ of formula (8.12). For the nonlinear super-resolution optical storage devices, the $d_{\text{resolving}}$ can be obtained by calculating the r value at $I_{\text{trans.normalized}}(r) \sim 0.90$, and the $d_{\text{resolving}} = 2r$. For any fixed nonlinear super-resolution optical storage device and dynamic testing system, the super-resolution spot size, D, is determined by the readout laser power. Therefore, setting $I_{\text{trans.normalized}}(r) \sim 0.90$ of formula (8.12) and conducting mathematical operations would yield the relation of the dynamic readout threshold power, P_{th}, with the readout resolving limit size of $d_{\text{resolving}}$:

$$\begin{cases} P_{\text{th}} = \dfrac{\pi\left[\frac{1}{4}d_{\text{resolving}}^2 + \frac{1}{2}w_0^2 \ln(0.90)\right]}{\beta L\left[1 - \exp\left(-\dfrac{d_{\text{resolving}}^2}{2w_0^2}\right)\right]}, & d_{\text{resolving}} < \dfrac{\lambda}{4\text{NA}} \\ P_{\text{th}} = 0, & d_{\text{resolving}} \geq \dfrac{\lambda}{4\text{NA}} \end{cases} \tag{8.13}$$

Formula (8.13) indicates that because the marks are smaller than the readout resolving limit $\lambda/4\text{NA}$ of the testing system itself, the different sizes of the information marks correspond to the different readout threshold powers, and the readout threshold power is only a function of the mark size. The analysis answers the first two questions put forward above.

In the nonlinear super-resolution optical storage devices, formula (8.12) can be used to calculate $d_{\text{resolving}}$ by setting $I_{\text{trans.normalized}}(r) \sim 0.90$. Thus, the relation of the dynamic readout threshold power P_{th} with the mark size can be derived, as shown in formula (8.13). Formula (8.13) indicates that for super-resolution marks, the different sizes correspond to the different readout threshold powers. Using the experimental parameters would yield the dependence of P_{th} on mark size. For the experiments of Fig. 8.14b, the laser wavelength and numerical aperture of optical pickup are $\lambda = 405\,\text{nm}$ and $\text{NA} = 0.65$, respectively. The $w_0(\sim 0.61\lambda/\text{NA})$ is approximately 380 nm. The nonlinear absorption coefficient and thickness of Sb_2Te_3 thin film are approximately $\beta = -2.5 \times 10^{-2}\,\text{m/W}$ for the laser pulse of 80 ns [13] and $L = 20\,\text{nm}$, respectively. Figure 8.14b displays the theoretically calculated results. The experimental data points are shown as well to provide the basis for comparison with the theoretical calculation. The readout threshold power increases with the reduction of the mark size, and the experimental data points are basically consistent with the theoretical calculation.

(b) Dependence of readout characteristics on laser power

In the nonlinear super-resolution optical storage, the SbTe-based phase-change materials are generally used as nonlinear thin films due to the giant nonlinear saturation absorption effect. The nonlinear saturation absorption characteristic originates from the thermally induced weakening of the resonant bonding. The thermally induced nonlinear absorption coefficient can be expressed as

$$\beta_t = \frac{\alpha_0 t_w}{2\rho C_p}\frac{\mathrm{d}\alpha}{\mathrm{d}T} \tag{8.14}$$

where t_w is laser pulse width. In the dynamic readout, the t_w is the laser irradiation time, which is determined by the readout speed of optical disk. $\mathrm{d}\alpha/\mathrm{d}T$ is the absorption change coefficient with temperature, which is a constant for a given phase-change material. The ρ and C_p are the density and heat capacity of materials, respectively. The $\rho C_p (= \kappa/D)$ can partially reflect the influence of thermal diffusion characteristic on the thermally induced nonlinear absorption coefficient, where κ and D are the thermal conductivity and diffusivity, respectively. Substituting formula (8.14) to formula (8.12), one can get

$$I_{\text{trans.normalized}}(r) = \frac{I_t(r)}{I_t(0)} = \frac{\exp\left(-\frac{2r^2}{w_0^2}\right)\exp\left(-\frac{\alpha_0 t_w}{2\rho C_p}\frac{\mathrm{d}\alpha}{\mathrm{d}T} I_0 \exp\left(-\frac{2r^2}{w_0^2}\right)L\right)}{\exp\left(-\frac{\alpha_0 t_w}{2\rho C_p}\frac{\mathrm{d}\alpha}{\mathrm{d}T} I_0 L\right)} \tag{8.15}$$

Formula (8.15) indicates that for a given nonlinear super-resolution optical storage device, $\mathrm{d}\alpha/\mathrm{d}T$ and L are fixed, and w_0 is determined by the dynamic testing system. Thus, the super-resolution spot size can only be tuned by changing the readout laser power density, I_0, namely the laser power, P_r, because $I_0 = 2P_r/\pi w_0^2$, and the laser irradiation time, i.e., the readout speed of optical disk.

By the thermo-physical properties $d\alpha/dT = -7.37 \times 10^3/(m \cdot K)$, $\alpha_0 = 7.78 \times 10^5/cm$, $\rho = 6.50 \times 10^3\,Kg/m^3$, $C_p = 205.5\,J/(Kg \cdot K)$ [13], and the laser irradiation time $t_w = 80\,ns$, one can obtain the optical spot intensity profile according to formula (8.15). Figure 8.15a shows the comparison of the super-resolution spot with the initial focused spot of the testing system where the readout laser power is set as $P_r = 0.9\,mW$. The super-resolution spot size is approximately one half of the initial focused spot. Based on the definition, the dynamic readout resolving limit corresponds to the spot position where the normalized intensity value is $I_{trans.normalized}(r) \sim 0.90$. Figure 8.15a shows that the resolving limit of the nonlinear super-resolution optical storage device is approximately 72 nm at $P_r = 0.9\,mW$; this value is lower than the initial focused spot resolving limit of approximately 156 nm. Hence, the dynamic readout threshold power is 0.9 mW for the information marks of 72 nm, which approximates the results in Fig. 8.14a, where the P_{th} is approximately 0.9 mW for 5 dB of CNR, which is regarded as the readout threshold signal.

The dependence of the intensity profile of the super-resolution spot on the readout power is illustrated in Fig. 8.15b as well, according to formula (8.15). The super-resolution spot size decreases with the increase of laser power. Generally, the smaller the super-resolution spot size is, the higher the dynamic readout power will be. As shown in Fig. 8.14b, to detect smaller marks, the super-resolution spot should become smaller. Figure 8.15b indicates that the higher the readout power, the smaller the super-resolution spot is. Accordingly, higher readout threshold power is needed for smaller marks. This requirement is one of the essential reasons of the increase of the dynamic readout threshold power with the reduction in mark size.

For marks with a fixed size, the smaller the super-resolution spot size is, the larger the CNR will be. Thus, CNR increases with the increase in the readout laser power. With the continuous increase of the readout power, CNR will be close to a

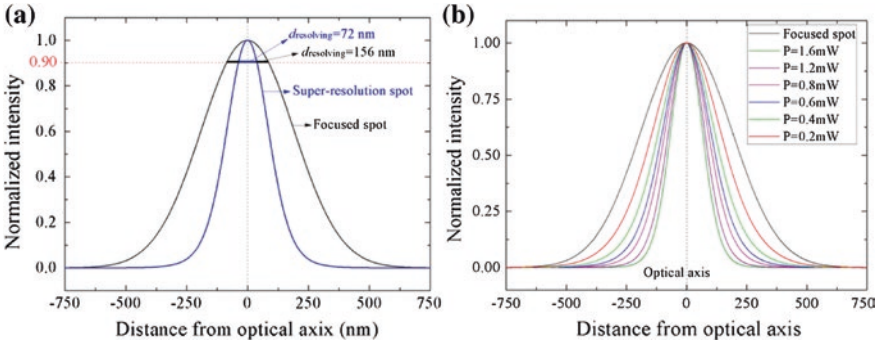

Fig. 8.15 Calculated spot intensity profiles: **a** dynamic readout resolving limit analysis and **b** dependence of spot intensity profile on readout laser power. Reprinted from [16], Copyright 2013, with permission from Elsevier

stable value because the mark signal intensity reaches a limit level. However, the CNR cannot increase infinitely with the increase of the readout power, because the structural change will occur if the readout laser power exceeds threshold of structural change of the Sb-based phase-change thin films due to the temperature rise. These findings are verified by Fig. 8.14a, where the dynamic readout threshold power is $0.8 \sim 0.9$mW at the CNR of 3–5 dB, the CNR increases with the readout power and remains basically unchanged with the continuous increase of the readout power. Finally, the CNR falls to zero when the readout power exceeds the certain threshold of the nonlinear thin films.

(c) Dependence of readout characteristic on laser irradiation time

The dependence of spot intensity profile on the laser irradiation time can be theoretically analyzed according to formula (8.15) by using the same thermophysical properties as Fig. 8.15, and the readout laser power is set to be $P = 0.9$ mW. Figure 8.16a gives the calculated results. The super-resolution spot size decreases with the increase of laser irradiation time t_w, which may be because the thermally induced nonlinear saturation absorption coefficient β_t is directly proportional to the laser irradiation time t_w, and the longer the laser irradiation time, the larger the nonlinear saturation absorption coefficient β_t is, as shown in formula (8.14). Thus, the longer laser irradiation time can form smaller super-resolution spot size. Generally, the smaller the super-resolution spot size, the higher the dynamic readout resolving power is. For the below-diffraction-limited recording marks, the smaller the super-resolution spot size, the higher the CNR of dynamic readout is. In the real nonlinear super-resolution optical storage, the laser irradiation time t_w is determined by the rotation speed of the disk v, i.e., $t_w \propto 1/v$. Thus, the high readout speed means the short laser irradiation time, which causes the large super-resolution spot size, as shown in Fig. 8.16a. These indicate that the high readout speed leads to low CNR for super-resolution marks. By lowering the readout speed, one can improve the CNR, which can be verified by the experimental results in [17].

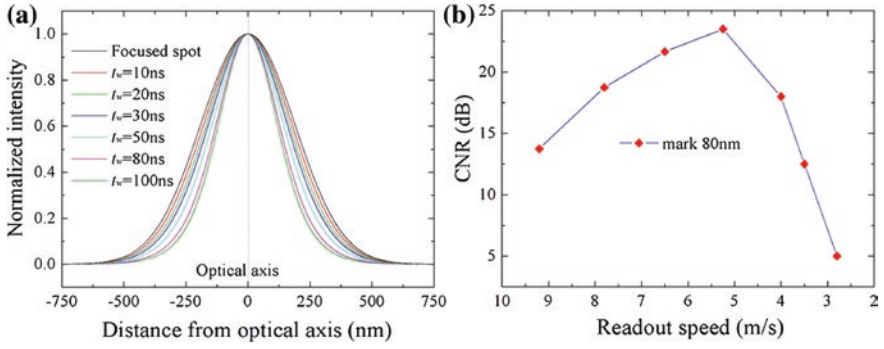

Fig. 8.16 Dependence of readout characteristic on the laser irradiation time: **a** spot intensity profile and **b** replot of Fig. 6 from [17]: the influence of the readout speed on the CNR. Reprinted from [16], Copyright 2013, with permission from Elsevier

In order to explain the effect of laser irradiation time on the readout characteristic, Fig. 8.16b replots the typical dependence of CNR on the readout speed from [17]. The CNR increases with the readout speed reduction from about 9.2 m/s to 5.25 m/s, which is consistent with Fig. 8.16a, where the laser irradiation time t_w increasing means the readout speed decreases. However, the CNR cannot increase infinitely with the decrease of the readout speed, which is because the longer laser irradiation time will cause the higher temperature in the nonlinear thin films. When the temperature exceeds the threshold of the structural change, the CNR decreases with the further decrease of the readout speed, as shown in Fig. 8.16b, where the CNR decreases from about 23.5 dB to 5 dB with the readout speed decreasing from about 5.25 m/s to 2.8 m/s. That is, there is an optimum readout speed for the super-resolution marks in a fixed readout power for a given nonlinear super-resolution disk.

(d) Analysis of the influence of laser energy on dynamic readout characteristic

There is a common concern that when the laser power or irradiation time exceeds some threshold, the structural change of the nonlinear layer will occur due to the temperature rise, and the CNR decreases with the increase of readout power or laser irradiation time.

In the nonlinear super-resolution optical storage, the nonlinear layers are mostly the Sb-based phase-change thin films. The nonlinear effect originates from the weakening of the resonant bonding due to the temperature rise. Thus, the super-resolution spot formation is dependent on the temperature rise. In the dynamic readout, the temperature rise is fast due to the large linear absorption coefficient of the order of magnitude of 10^5/cm. The temperature rise ΔT can be estimated by [18]

$$\Delta T(r) \propto \frac{\alpha_0 D}{\kappa} t_w I_{\text{in}}(r) \tag{8.16}$$

where $t_w \times I_{\text{in}} = \Delta E$ is the irradiation laser energy deposited into the nonlinear layers. Thus, the influence of laser power and speed (i.e., laser irradiation time) on the dynamic readout characteristic can be analyzed by ΔE. Formula (8.16) indicates that the temperature rise is directly proportional to ΔE. The spatial profile of temperature rise approximately equals to the incident light intensity and can be expressed as

$$\Delta T(r) \propto \frac{\alpha_0 D}{\kappa} t_w I_0 \exp\left(-\frac{2r^2}{w_0^2}\right) \tag{8.17}$$

Formula (8.17) gives the temperature profile of the nonlinear layers. The crystalline state is resonant bonding, and the amorphous and molten states are covalently bonding for Sb-based phase-change materials. There is the nonlinear saturation absorption effect only in the crystalline state because the nonlinear effect is from the weakening of the resonant bonding. There is no obvious nonlinear effect in the molten state due to the lack of resonant bonding. For the dynamic readout, when

the laser power is low or the irradiation time is short, the irradiation laser energy ΔE deposited into the Sb-based phase-change materials is not enough to induce the structural change, and the Sb-based thin films remain in crystalline state. The super-resolution spot formation results from the nonlinear effect, as is shown in Fig. 8.17a, where in the whole region of spot, the Sb-based thin film remains in crystalline state. The highest temperature at the center of the spot is lower than the threshold temperature T_{th} of structural change, and the super-resolution spot formation can be analyzed by formulas (8.12) and (8.17).

However, when the laser power is higher or the irradiation time is longer, the laser energy ΔE deposited into the Sb-based phase-change thin films is enough to induce the structural change from crystalline state to molten state. The central region of the spot becomes firstly molten state due to the Gaussian profile of the temperature, and the periphery of central region remains crystalline state, as shown in Fig. 8.17b. For the phase-change materials, the absorption coefficients of the amorphous and molten states are smaller than those of crystalline state; thus, the optical aperture of below-diffraction-limited spot generates in the region of the molten state. With the increase of the ΔE, the optical aperture radius increases due to the increase of temperature and radially thermal diffusion from the center of the aperture, i.e., increasing the area above the threshold temperature T_{th}. At certain temperature, the optical aperture radius approaches the feature size of marks on the super-resolution disk and the CNR decreases with the increase of the laser irradiation time or power, as shown in Figs. 8.12, 8.14a, and 8.16b. This is the essence that when the readout power or irradiation time exceeds certain threshold, the CNR decreases with the increase of readout power or laser irradiation time. It should be noted that if the laser irradiation time is longer or the cooling rate is lower than about 10^9 °C/s, the formation of the "molten" aperture will become crystalline after the laser is removed, i.e., the generation of "molten" aperture is reversible.

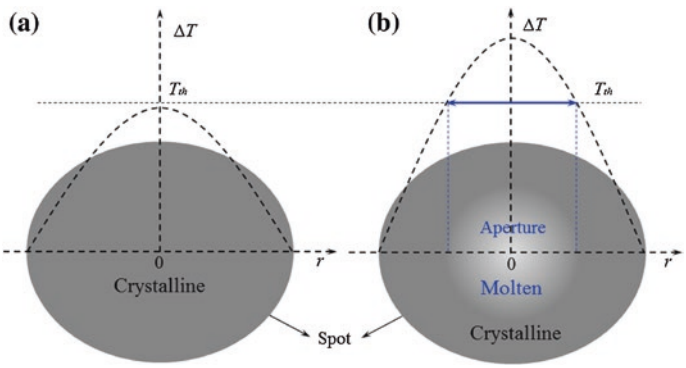

Fig. 8.17 The schematic of the super-resolution aperture formation: **a** low-energy laser irradiation, **b** high-energy laser irradiation. Reprinted from [16], Copyright 2013, with permission from Elsevier

8.3 Reverse-Saturation Absorption-Induced Super-Resolution Optical Storage

8.3.1 Recordable Super-Resolution Optical Disks with Nonlinear Reverse-Saturation Absorption

AgInSbTe thin films possess strong nonlinear reverse-saturation absorption characteristics at red light wavelength [19], which has been applied to the nano-optical data storage of bubble-type super-resolution near-field structure. The disk structure is schematically shown in Fig. 8.18, where AgInSbTe is helpful for reducing the recording mark size due to nonlinear reverse-saturation absorption effect in 635 nm light wavelength. In the recording, the focused spot passes through the AgInSbTe thin film and a below-diffraction-limited energy absorption region is formed, the below-diffraction-limited energy absorption region is heated, and the heat is transferred to the PtOx layer, which can induce the decomposition of PtOx into Pt particles and O_2 and produce the bubble-type recording marks. In the bubble-type marks, the Pt particles can enhance the readout signal and improve the CNR of dynamic readout. In order to demonstrate the schematics, Kim et al. carried out the dynamic recording and readout by using Pulstec DDU-1000 with a laser wavelength of 635 nm and a lens with an NA of 0.60 [20]. The disks were rotated at a velocity of 6.0 m/s. The results showed that the 100 nm recording marks in the track direction were realized, and the mark size corresponds to 1/10 of the laser spot size (∼1.0 um).

Figure 8.19 presents the experimental results from [20], and Fig. 8.19a is the dependence of CNR on mark size. The CNR reaches up to 47 dB for the marks of 100 nm length and exceeds 43 dB for the marks of 80 nm length, and the mark length is obviously smaller than the resolution limit of the optical testing setup of ∼265 nm. The large CNR may result from the Pt nanoparticles being activated in

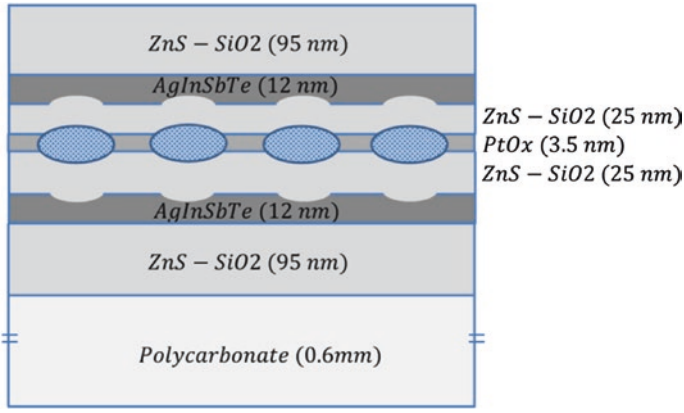

Fig. 8.18 Sample structure and expected marks in PtOx layer

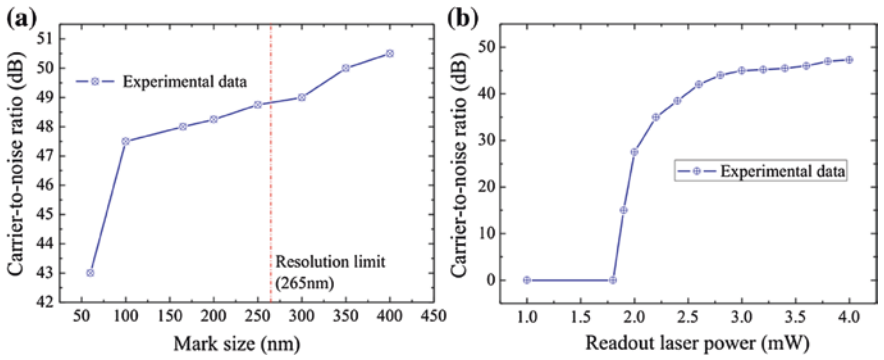

Fig. 8.19 Dynamic readout characteristic and the dependence of CNR on **a** mark size and **b** readout power for marks of 100 nm length

the area of the bubble-type marks by the laser irradiation or a thermal effect, and the Pt nanoparticles can enhance and scatter the light.

Another distinctive feature is the rapid increase of CNR at ~2 mW readout power (shown in Fig. 8.19b), which results from small Pt nanoparticles in the bubble-type marks. The small Pt nanoparticles may become more sensitive as the readout power increases. The lower readout power is helpful for the readout durability.

The readout durability is very important for actual application, which strongly depends on the bubble-type mark stability. There is no doubt that the size and morphology of bubble-type marks can be changed by thermal accumulation in the readout process, which has ill-effect on readout durability. Actually, the deformation of bubble-type marks can be controlled by adjusting the readout power, AgInSbTe layer thickness, and ZnS–SiO_2 dielectric layer. In order to improve the ability to anti-deformation, on one hand, the dielectric materials with good thermal diffusion and anti-deformation should be developed. On the other hand, a readout power as low as possible is useful for good readout durability [21].

8.3.2 Read-Only Optical Disk with Reverse-Saturation Absorption Effect

Crystalline InSb thin films present typical nonlinear reverse-saturation absorption effect in 405 nm light wavelength [22]. For reverse-saturation absorption thin films, the stronger the light intensity, the larger the absorptivity is. For Gaussian laser beam spot irradiation, the absorptivity is the largest at the central region of spot, which causes the central scatterer to be formed in the spot.

The nonlinear reverse-saturation absorption-induced central scatterer within the spot can be directly observed experimentally. Assafrao et al. constructed an

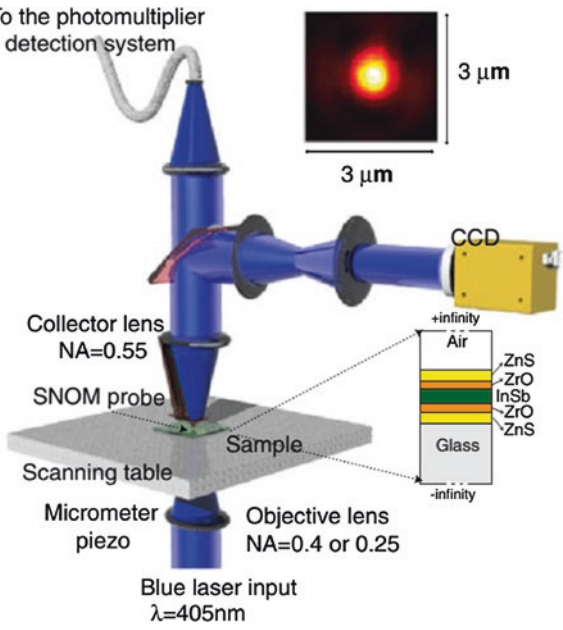

Fig. 8.20 Schematic of the experimental setup for spot measurement. Near-field measurement with the near-field scanning probe head in contact with the sample surface (an example of measured focused spot on the *top right*). Far-field measurement after removing the near-field scanning probe head and redirecting the beam to the CCD camera [23]. Copyright (2014), The Japan Society of Applied Physics

experimental setup, as shown in Fig. 8.20 [23]. A diode laser with a wavelength of 405 nm emits a laser beam that is coupled into a single-mode fiber and redirected to the near-field optical spot scanning unit. The sample is placed in the focal region of the lens. Near-field images of spot are taken with the near-field scanning probe head which is in contact with the sample. For the far-field measurement with scatterometry, the near-field scanning probe head is removed and a beam splitter is used to redirect the transmitted light to a CCD camera.

The nonlinear sample is layer stack of "ZnS–SiO_2(35 nm)/ZrO_2(15 nm)/InSb(20 nm)/ZrO_2(15 nm)/ZnS–SiO_2(35 nm)", and the layer stack is directly deposited on a 0.6 mm-thickness glass substrate by magnetron sputtering. The InSb is used as nonlinear reverse-saturation absorption layer. A pair of ZrO_2 layers adjacent to an InSb thin film are used to improve readout stability. A pair of ZnS–SiO_2 is used as dielectric protective layer.

According to the theory of coherent far-field scatterometry, any change of the nonlinear layer should noticeably result in some influence on the far-field intensity profile of the transmitted spot. In particular, if a small scatterer occurs inside the region overlapped by the spot, the outer part of the far-field intensity profile can be influenced. The influence can move toward the center of the far-field distribution if the scatterer dimension increases. In order to intuitively observe the scatterer formation, Assafrao et al. obtained the far-field intensity profile at different laser powers, which is direct evidence of the scatterer formation (shown in Fig. 8.21a–c). At low power of 0.8 mW, the far-field intensity has an undisturbed Gaussian-like profile, which indicates that there is no scatterer formation in the InSb layer

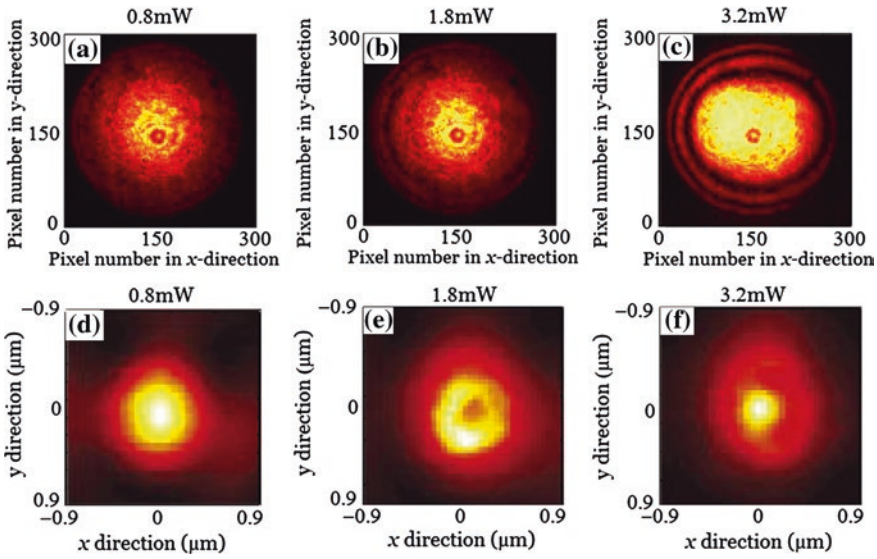

Fig. 8.21 Measured normalized intensity profiles of the transmitted spot, far-field at the incident laser power of **a** 0.8 mW, **b** 1.8 mW, and **c** 3.2 mW. Near-field at the incident laser power of **d** 0.8 mW, **e** 1.8 mW, and **f** 3.2 mW [23]. Copyright (2014), The Japan Society of Applied Physics

(Fig. 8.21a). At the laser power of 1.8 mW, a tiny ring appears in the outer edge, resulting from a change in the InSb layer (Fig. 8.21b). When the laser power is increased to 3.2 mW, a clear annular pattern becomes visible in the far-field distribution (Fig. 8.21c) [23].

The super-resolved spots are measured by near-field probe scanning method for the objective lens with NA = 0.4. Figure 8.21d is the first measurement at low power of 0.8 mW, the profile presents the focused Airy pattern, and there is no super-resolution effect. Figure 8.21e is obtained at laser power of 1.8 mW, and there is a dip in the center of intensity distribution resulting from light being partially blocked by the scatterer at the central region of spot. In Fig. 8.21f, the laser power is 3.2 mW, and the spot presents a below-diffraction-limited spot intensity profile, which is smaller than the focused spot.

Utilizing the central scattering effect resulting from nonlinear reverse-saturation absorption, Nakai et al. prepared super-resolution read-only disk with a layer stack of "ZnS–SiO_2(35 nm)/ZrO_2(15 nm)/InSb(20 nm)/ZrO_2(15 nm)/ZnS–SiO_2(35 nm)" [24]. The layer stack was deposited onto the read-only disk substrate by magnetron-controlling sputtering method. The thickness of a cover layer on the stack is 0.1 mm. The disk substrate has a random pit array including a minimum pit size of 75 nm. The track pitch length is 320 nm. A pair of ZrO_2 interface layers adjacent to the InSb active layer are used to improve read stability. Using a playback system with a light wavelength of 405 nm and NA of optical pickup of 0.85, the playback test was conducted at a data transfer rate of 36 Mbps.

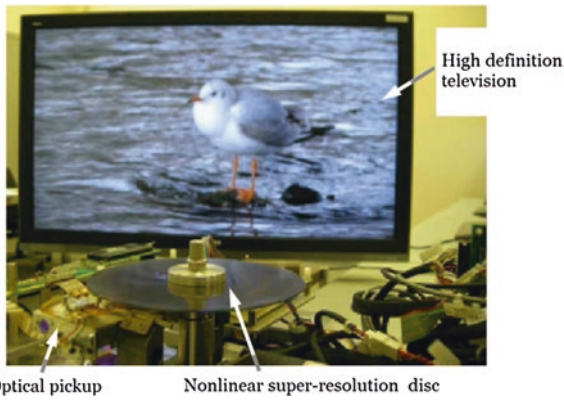

Fig. 8.22 Snapshot of playback demonstration of high-density video content from read-only super-resolution disk [24]. Copyright (2010), The Japan Society of Applied Physics

The high-definition video content was replayed successfully with no image discontinuity on the television display during video playback. Figure 8.22 gives a snapshot of the playback demonstration. A seamless playback with a data transfer rate of 72 Mbps was also demonstrated successfully.

8.4 Read-Only Super-Resolution Optical Disks with Thermally Induced Reflectance Change Effect

The thermally-induced reflectance change can also be applied to the optical disk storage due to the formation of small reflected spot [25]. For most materials, the thermally-induced reflectance change actually results from thermally-induced absorption coefficient change. According to [26], the reflectance of Si thin film is a function of temperature. The reflectivity change can be expressed as

$$\begin{cases} R(T) = 0.367 + 4.29 \times 10^{-5} T, & T < T_{\mathrm{m}} \\ R(T) = 0.372 + 2.693 \times 10^{-5} T + 2.691 \times 10^{-15} T^4, & T_{\mathrm{f}} > T > T_{\mathrm{m}} \end{cases} \tag{8.18}$$

where T_{m} is the initial melting temperature, T_{f} is the boiling point temperature, and $R(T)$ is the reflectivity. Formula (8.18) is plotted in Fig. 8.23. Figure 8.23 gives the values of $R = 0.38$ at room temperature and 0.44 near T_{m}, and the $R(T)$ shows a linear increase with the increase in temperature when T is below the T_{m} [27]. However, $R(T)$ shows a nonlinear increase with the increase in temperature when temperature is between T_{m} and T_{f} and reaches up to 0.72 at T_{f}.

The designed optical disk structure is "premastered substrate/SiN layer/Si thin film/SiN layer" (shown in Fig. 8.24). Si thin film replaces the metal reflective layer in the conventional read-only optical disk and is used as super-resolution readout layer, and the SiN dielectric layer is used to protect the Si thin film from being damaged in the irradiation. The disk substrate structure is shown in Fig. 8.24, where the pit diameter is 380 nm.

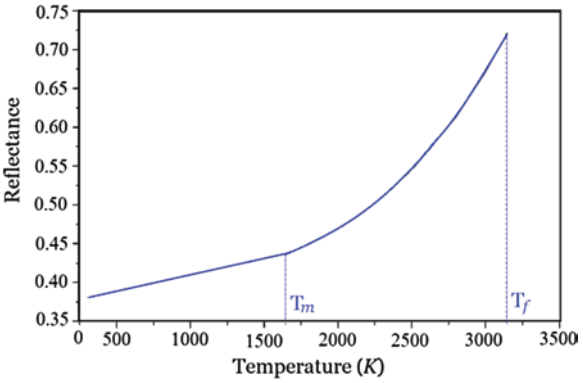

Fig. 8.23 Dependence of reflectivity of Si on temperature

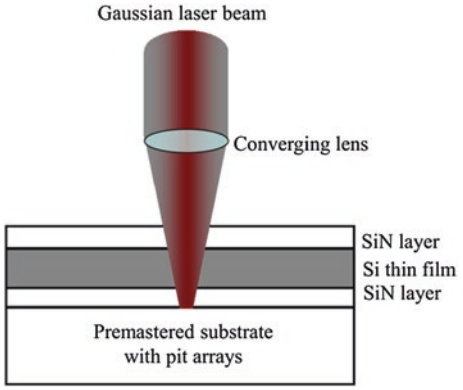

Fig. 8.24 Sample structure of "SiN layer/Si thin film/SiN layer/premastered substrate"

The dynamic readout is tested in the setup of Fig. 8.3, where the laser wavelength λ and the NA of converging lens are changed into 780 nm and 0.45, respectively. The diameter of focused spot on the optical disk is 2.115 μm, which is much larger than the diameter of pits (∼380 nm). The dynamic readout resolution limit of optical disk approximately is about 433 nm, so the size of pit arrays on the optical disk substrates is smaller than the readout resolution limit, and the pit arrays cannot be dynamically read out by conventional reflective layer.

Si thin film was deposited by RF sputtering on the optical disk substrate at the background pressure of less than 1.0×10^4 Pa. SiN was used as the protection layer and prepared by reactive sputtering technique with Si target by introducing a gas mixture Ar and N_2.

Figure 8.25 describes the dynamic readout results. Figure 8.25a shows that the oscilloscopic signal is modulated, and Fig. 8.25b shows that the CNR is about 25 dB. These indicate that the super-resolution pit arrays are dynamically read out.

In the process of dynamic readout, the Si thin film does not only function as a super-resolution layer, but also as an optical reflective layer. Since the focused Gaussian beam is irradiated on the moving Si thin film during readout, the

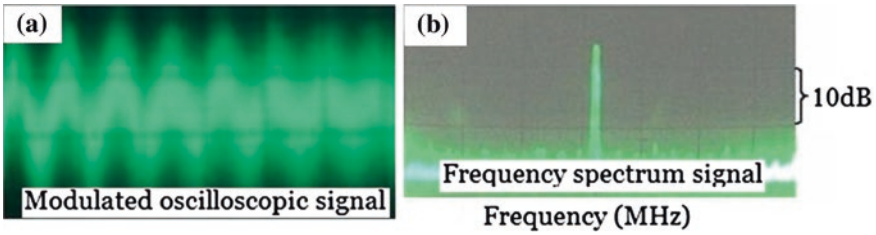

Fig. 8.25 Dynamic readout results **a** oscilloscopic signal and **b** frequency spectrum signal

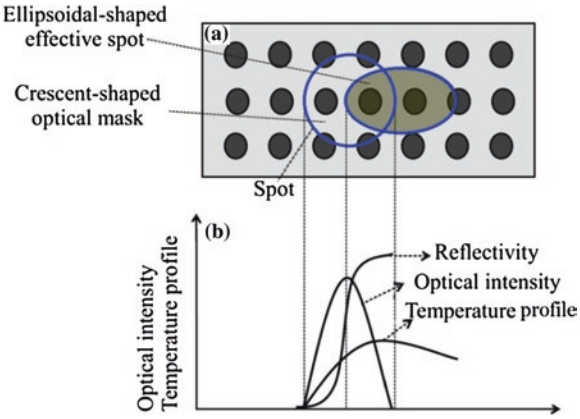

Fig. 8.26 Principle on readout of super-resolution pits: **a** overview of super-resolution optical disk and **b** temperature, reflectivity, and intensity profiles

temperature rise of the Si thin film overlapped by the rear portion of the spot is higher than that of the front portion of spot. In the region heated above the melting temperature, the Si thin film melts, which causes a drastic reflectivity change and formations of the crescent-shaped solid region in the front portion and the ellipsoid-shaped melted regime in the rear portion of the spot (shown in Fig. 8.26). Because the reflectivity of Si thin film in molten state is higher than that in the solid state, the crescent-shaped solid region overlapped by the spot is optically masked (shown in Fig. 8.26a). Hence, the effective spot size is reduced both in the linear direction and in the track direction, and the resolving power is increased. That is, the thermally induced reflectance change does not only increase the linear density, but also improves the track density.

The fatigue characteristics of dynamic readout are also tested. Figure 8.27a is the frequency spectrum signal for the first dynamic readout, and the CNR is about 20 dB. After 10^5 times readout, the CNR still remains at about 20 dB (also see Fig. 8.27b), which indicates that the optical disks can not only realize the dynamic readout of super-resolution pit arrays, but also signal is very stable. The stability of signal is very important in real application.

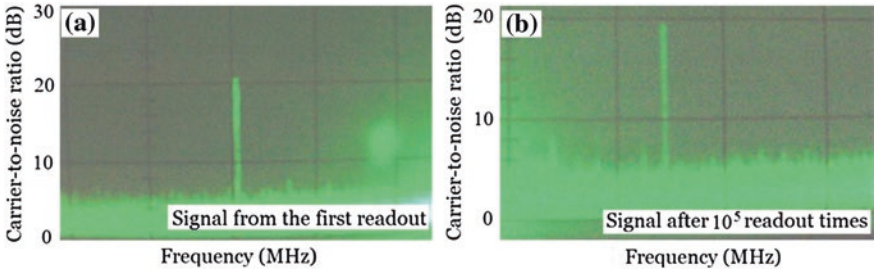

Fig. 8.27 Fatigue characteristics of dynamic readout: **a** the first readout signal and **b** the signal after 10^5 times readout

References

1. J. Tominaga, The super-resolution near-field structure and its applications, in *Optical Nanotechnology: The manipulation of Surface and Local Plasmons*, ed. by J. Tominaga, D.P. Tsai (Springer, Berlin, 2003)
2. J. Tominaga, T. Nakano, N. Atoda, An approach for recording and readout beyond the diffraction limit with an Sb thin film. Appl. Phys. Lett. **73**, 2078–2080 (1998)
3. T. Kikukawa, T. Nakano, T. Shima, J. Tominaga, Rigid bubble pit formation and huge signal enhancement in super-resolution near-field structure disk with platinum-oxide layer. Appl. Phys. Lett. **81**, 4697–4699 (2002)
4. J. Wei, F. Gan, Thermal lens model of Sb thin film in super-resolution near-field structure. Appl. Phys. Lett. **82**(16), 2607–2609 (2003)
5. G. Bouwhuis, J.H.M. Sprui, Optical storage read-out nonlinear disks. Appl. Opt. **29**, 3766–3768 (1990)
6. J. Wei, F. Gan, Novel approach to super-resolution pits readout. Opt. Eng. **41**, 1073–1074 (2002)
7. T. Fukaya, J. Tominaga, T. Nakano, N. Atoda, Optical switching property of a light-induced pinhole in antimony thin film. Appl. Phys. Lett. **75**, 3114–3116 (1999)
8. J. Wei, F. Gan, Time response of optical switching properties of Sb thin films under focused laser pulses. Opt. Eng. **42**(6), 1749–1753 (2003)
9. J. Wei, H. Ruan, H. Shi, F. Gan, Study on read-only optical disk with Sb mask super-resolution. Chin. Sci. Bulletin **47**, 1604–1606 (2002)
10. R. Wang, J. Wei, Dependence of the readout resolving on the thickness of nonlinear super-resolution thin films. Proc. SPIE **8782**, 878205 (2013)
11. J. Wei, S. Liu, Y. Geng, Y. Wang, X. Li, Y. Wu, A. Dun, Nano-optical information storage induced by the nonlinear saturable absorption effect. Nanoscale **3**, 3233–3237 (2011)
12. J. Liu, S. Liu, J. Wei, Origin of the giant optical nonlinearity of Sb_2Te_3 phase change materials. Appl. Phys. Lett. **97**, 261903 (2010)
13. S. Liu, J. Wei, F. Gan, Nonlinear absorption of Sb-based phase change materials due to the weakening of the resonant bond. Appl. Phys. Lett. **100**, 111903 (2012)
14. W. Chen, Y. Wu, J. Wei, F. Gan, Optical recording properties of metal-azo dye thin film with super-resolution near-field structure. Proc. SPIE **5966**, 59661A (2005)
15. X. Li, Y. Wu, D. Gu, F. Gan, Thermal decomposition kinetics of nickel(II) and cobalt(II) azo barbituric acid complexes. Thermochim. Acta **493**, 85–89 (2009)
16. J. Wei, On the dynamic readout characteristic of nonlinear super-resolution optical storage. Opt. Commun. **291**, 143–149 (2013)

17. L.P. Shi, T.C. Chong, X. Hu, J.M. Li, X.S. Miao, H.X. Yang, K.G. Lim, Blu-ray type super resolution near field optical disk with Sb_2Te_3 mask layers and a thermal shield layer in front of the mask layer. Proc. SPIE **6282**, 62821U (2006)
18. S. Liu, J. Wei, F. Gan, Optical nonlinear absorption characteristics of crystalline $Ge_2Sb_2Te_5$ thin films. J. Appl. Phys. **110**, 033503 (2011)
19. J. Liu and J. Wei, Optical nonlinear absorption characteristics of AgInSbTe phase change thin films. J. Appl. Phys. **106**, 083112 (2009)
20. J. Kim, I. Hwang, D. Yoon, I. Park, D. Shin, T. Kikukawa, T. Shima, J. Tominaga, Super-resolution by elliptical bubble formation with PtOx and AgInSbTe layers. Appl. Phys. Lett. **83**, 1701–1703 (2003)
21. Q. Liu, T. Fukaya, S. Cao, C. Guo, Z. Zhang, Y. Guo, J. Wei, J. Tominaga, Study on readout durability of super-RENS disk. Opt. Exp. **16**, 213–218 (2008)
22. X. Cai, J. Wei, Optical nonlinearity characteristics of crystalline InSb semiconductor thin films. J. Phys. D: Appl. Phys. **46**, 435101 (2013)
23. A.C. Assafrao, N. Kumar, A.J.H. Wachters, S.F. Pereira, H.P. Urbach, Experimental and numerical analysis of the super resolution near-field effect on an InSb sample. Jap. J. Appl. Phys. **53**, 042001 (2014)
24. K. Nakai, M. Ohmaki, N. Takeshita, M. Shinoda, I. Hwang,Y. Lee, H. Zhao, J. Kim, B. Hyot, B. André, L. Poupinet, T. Shima, T. Nakano, J. Tominaga, First playback of high-definition video contents from super-resolution near-field structure optical disc. Jap. J. Appl. Phys. **49**, 08KE02 (2010)
25. J. Wei, Z. Jiang, H. Shi, F. Gan, Near-field gaussian lens originating from nonlinear multi-layer structures: application in ''dynamic readout of super-resolution pit arrays on the discs''. J. Nanosci. Nanotech. **9**, 982–984 (2009)
26. J.E. Moody, R.H. Hendel, Temperature profile induced by a scanning cw laser beam. J. Appl. Phys. **53**, 4364–4371 (1982)
27. J. Wei, H. Ruan, F. Gan, Study on readout of super-resolution pits with Si films. Proc. SPIE **5060**, 167–170 (2003)

Chapter 9
Applications of Nonlinear Super-Resolution Effects in Nanolithography and High-Resolution Light Imaging

9.1 Introduction

As scientific and engineering interests in nanoscale structures increase, there is a strong need to fabricate nanometric surface and interface patterns with high regularity, reliability, and controllability. Formation of nanometric patterns on surfaces and thin films is often carried out with advanced nanolithography techniques, such as electron and ion beam lithography [1, 2], nanoimprint [3], X-ray lithography [4], and local scanning probe microscopy method [5]. However, these techniques usually involve highly sophisticated procedures and require expensive equipments. Therefore, maskless direct laser writing has still been the choice for many applications in microelectronics and photonics [6, 7]. As far as nanometric patterns are concerned, the method suffers from the fundamental limit in resolution due to light diffraction. In maskless direct laser writing, an alternative inorganic resist is thermal lithographic thin film, such as chalcogenide and phase-change thin films [8]. The thermal lithographic material absorbs the laser energy and is heated to certain temperature, such as phase-change or melting points, to form various patterning on the material surface. In thermal lithography, the pattern resolution can be improved by using lithographic materials with nonlinear absorption and thermal threshold effect, and the corresponding principles and experimental methods are presented. In addition, as an extended application of nonlinear effects, the super-resolution mark detection and imaging are also explored in this chapter.

9.2 Thermal Threshold Lithography

In the direct laser writing lithography, the lithography line width (or characteristic size) is generally determined by the spot size. For example, for a direct laser writing system with a light wavelength of 405 nm and numerical aperture (NA) 0.90

J. Wei, *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7_9

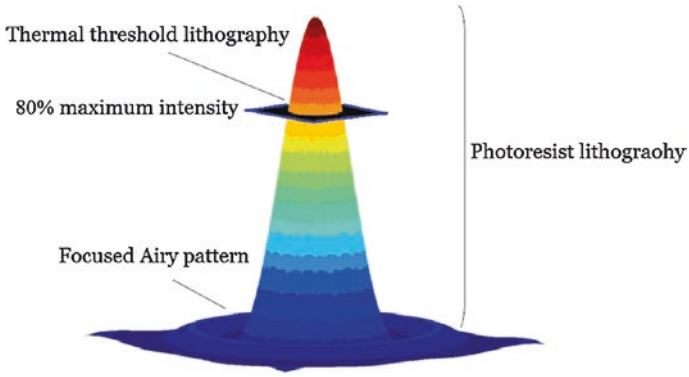

Fig. 9.1 Principle schematics of thermal threshold lithography

of converging lens, the focused spot size is $D \sim 1.22\lambda/NA \sim 0.52\ \mu m$, where λ is light wavelength. The lithography line width should be about $0.5 \sim 0.6\ \mu m$ if the traditional photoresist is used. However, the lithography line width can be reduced if the photoresists with obvious threshold effects are used. Figure 9.1 gives the basic principle schematics of thermal threshold lithography. The focused Airy spot pattern intensity presents typical Gaussian profile. For traditional photoresist lithography, the lithographic area is mainly determined by the focused Airy pattern, while if one uses the special photoresists with thermal threshold effect, the lithography can be conducted at certain intensity, such as 80 % maximum intensity; thus, the characteristic size is obviously smaller than focused Airy pattern spot.

9.2.1 Crystallization Threshold Lithography

Chalcogenide phase-change thin films are usually used as information storage materials due to the difference of optical reflectivity or electric resistivity between crystalline and amorphous states. The crystalline state is obtained through the crystallization of amorphous state under laser or electrical pulse, and the crystallization occurs when the temperature exceeds to the crystallization threshold point; thus, the crystallization is actually thermal process [9, 10]. Another performance of the chalcogenide phase-change thin films is the selective etching between crystalline and amorphous states. The selective etching makes the chalcogenide phase-change thin films be used as lithography materials, and the lithography characteristic size can be smaller than optical diffraction limit due to the crystallization threshold effect [11].

Here, the $Ge_2Sb_2Te_5$ is taken as an example and laser-induced crystallization threshold lithography process flow is shown in Fig. 9.2. The sample is designed as "$Ge_2Sb_2Te_5$ (70 nm)/interface layer/substrate" structure [12]. The laser-induced

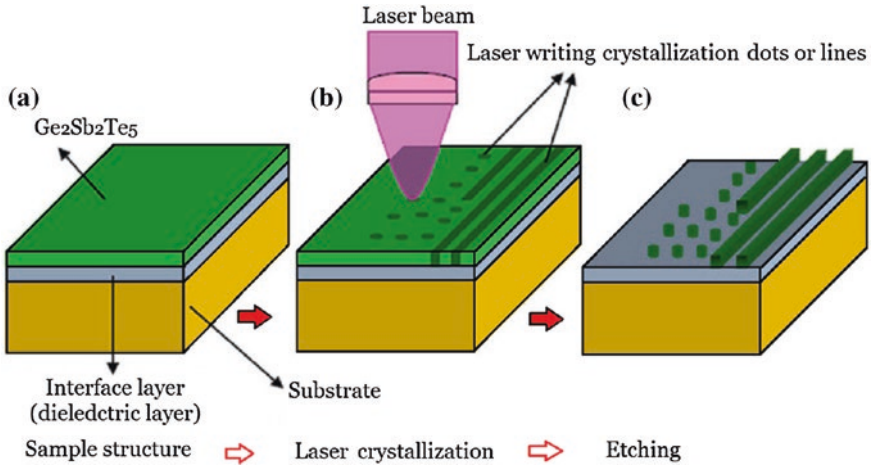

Fig. 9.2 Schematic drawing of laser-induced crystallization threshold lithography, **a** sample structure, **b** laser-induced crystallization, and **c** etching

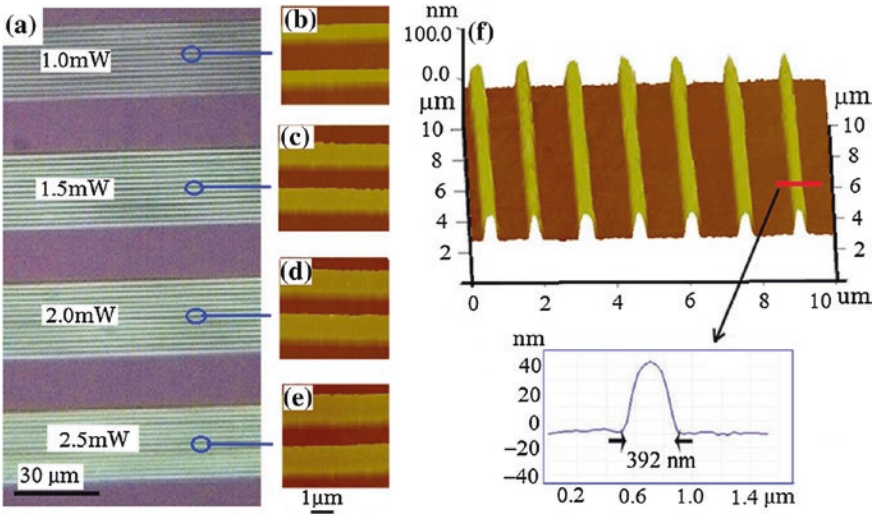

Fig. 9.3 Laser-induced crystallization and etching on $Ge_2Sb_2Te_5$ thin films, using 658-nm-wavelength laser beam as light source, **a** line shape patterns fabricated by the red laser writing system, **b**–**e** are AFM images with laser power of 1−2.5 mW; **f** AFM image with a 405-nm laser at 2 mW

crystallization is conducted by a laser writing system with light wavelength of 658 nm or 405 nm and NA of 0.9. Wet-etching development is carried out in tetramethylammonium hydroxide solution with concentration of 25 wt.% for 10 min and then washed in deionized water for two times. The area of laser irradiation

(crystallization) is remained, whereas the non-irradiation region (amorphous state) is etched away.

Figure 9.3a–e shows the line shape lithography structures, where laser wavelength is 658 nm. The line width becomes large as laser energy increases. The line width can be reduced using low laser power, which can be seen from the comparison of among Fig. 9.3b–e. To obtain small line width patterns with the crystallization threshold lithography method, a 405-nm laser is used to replace 658-nm laser, and subwavelength line width structures are presented in Fig. 9.3f, where the line width is about 390 nm, which is smaller than focused spot size of 520 nm. The small line width is because the crystallization can be controlled at about 50 % spot intensity position. The line width can be further reduced if the crystallization occurs at about >50 % spot intensity position.

9.2.2 Thermal Decomposition Threshold Lithography

The metal oxide can be easily thermally decomposed into metal and oxygen by laser heating above certain temperature, which can be used to fabricate patterns [13]. Compared with photoresist, the metal oxides do not need developing and etching processes, and the laser power is required to be in a very small range, so it is suitable for fabricating a large-area pattern structure in very short time and very low cost, which largely decrease the time-consuming and industrial cost. More importantly, the pattern size can be controlled by changing the laser pulse energy due to the thermal decomposition threshold effect. The AgOx is taken as an example and laser pulse-induced pattern structures are given as follows.

The AgOx material can be thermally decomposed into Ag particles and O_2 about 160 °C by laser pulse heating.

$$\mathrm{AgO}x \xrightarrow{\text{Laser heating}} \mathrm{Ag} + \frac{1}{2}x\mathrm{O}_2$$

According to the chemical reaction, in order to obtain pattern structures, the multilayer structure "ZnS-SiO_2/AgOx/ZnS-SiO_2" is designed as pattern sample, as shown in Fig. 9.4. When the laser beam irradiates on the AgOx thin film, a small volume of thin film is locally and rapidly heated to the thermal decomposing temperature, and then, the reaction happens. The oxygen released by the decomposition is stayed in the enclosed system, so it will apply pressure to surface, which can produce a large oxygen bubble. The oxygen bubble makes the ZnS–SiO_2 thin film deform and form bump-type pattern structures (Fig. 9.4a). Figure 9.4b shows the interior situation when AgOx decomposes into silver and oxygen. The O_2 and Ag particles are rough and tumble and fill into the whole room. After the AgOx cools down to the room temperature, the expanded volume will be left as bump. If one precisely controls the laser energy, the regular and uniform bump array pattern structures can be obtained.

Actually, the sample is not a completely enclosed system, and interdiffusion of gas can take place between the inner and the outer of the bumps, which causes the

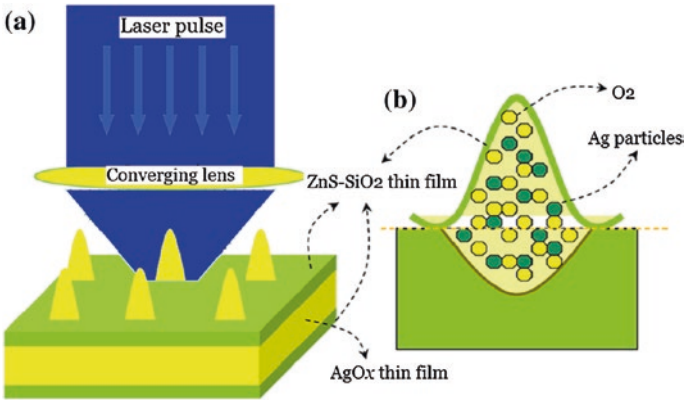

Fig. 9.4 Formation schematics of bump-type pattern structure on "ZnS–SiO_2/AgO*x*/ZnS–SiO_2" samples by laser pulse heating, **a** laser pulse irradiation on the sample and the bump formation, and **b** decomposition of AgO*x* and the formation of ZnS–SiO_2 bump-type structure. Reprinted from [23], with kind permission from Springer Science+Business Media

pressure inside and outside the bump to reach up to balance. However, if the laser energy is large enough, the bump raptures and a hole forms at the apex of bump.

Figure 9.5 shows the pattern structures fabricated with a range of laser power from 3.0 to 5.0 mW. The pattern structure appears to be taper shape and regular. The bumps with different height and size can be realized by tuning the laser power. The higher the laser power, the larger the bump height and size are. When the laser power is 5.0 mW, the height reaches up to the largest

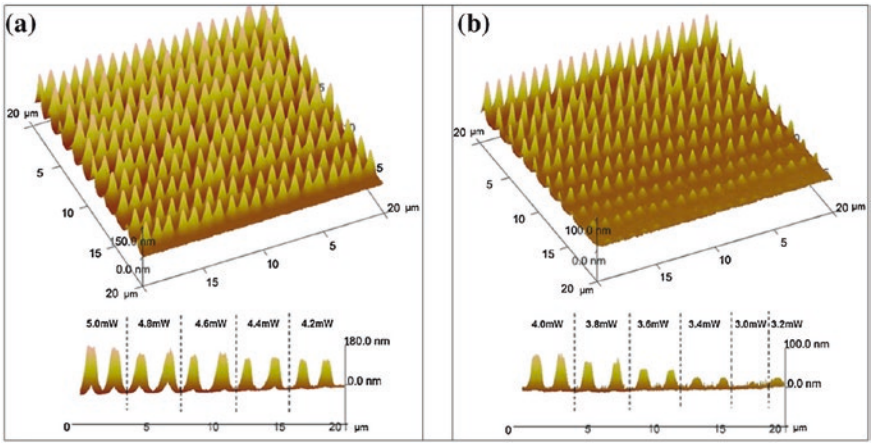

Fig. 9.5 Bump pattern structures written on multilayered ZnS–SiO_2/AgO*x*/ZnS–SiO_2 samples by green laser ($\lambda = 488$ nm) in different laser parameters, the AFM three-dimensional and lateral images for laser power **a** from 4.2 to 5.0 mW, and **b** from 3.2 to 4.0 mW. Reprinted from [23], with kind permission from Springer Science+Business Media

value. As the laser power decreases, the height and size gradually decreases to the smallest value at the laser power of 3.0 mW, where the bump is almost undistinguishable.

9.2.3 Molten Ablation Threshold Lithography

The molten ablation of chalcogenide phase-change thin films can take place under laser pulse heating when the temperature exceeds to melting point. The melting point is called as molten ablation threshold temperature [14]. The Sb_2Te_3 is taken as an example, and the molten ablation threshold lithography is introduced as follows.

The "Sb_2Te_3 thin film/glass substrate" samples are prepared. Sb_2Te_3 phase-change materials are very sensitive to temperature. The micro-/nanopatterns are directly formed on the thin film using direct laser writing system, where a laser beam with a wavelength of 650 nm is used and focused through an objective lens with an NA of 0.90. In laser writing process, the laser power is firstly fixed at 5.0 mW, and the pulse width is adjusted from 100 to 5 ns.

Figure 9.6a–f shows the AFM two-dimensional patterns where the pulse width is also listed on the right side of the images. The pattering may be from two competitive effects, one is chemical capillary effect, and the other

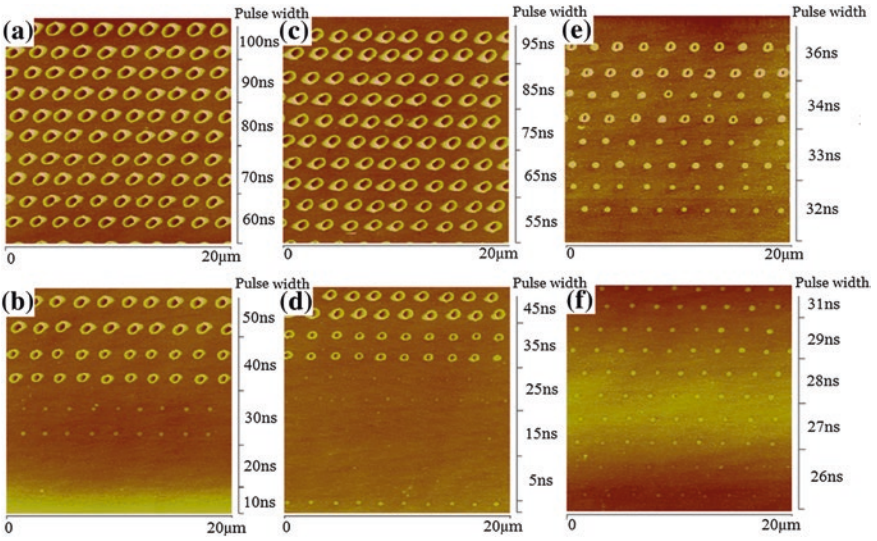

Fig. 9.6 Pattern structures formed on the Sb_2Te_3 thin films under laser pulse irradiation with a laser power of 5.0 mW and different pulse width, **a** pulse width from 100 to 60 ns, **b** pulse width from 50 to 10 ns, **c** pulse width from 95 to 55 ns, **d** pulse width from 45 to 5 ns, **e** pulse width from 36 to 32 ns, and **f** pulse width from 31 to 26 ns. Reprinted from [24], with kind permission from Springer Science+Business Media

is thermal capillary effect. One can see that three regimes occur as the pulse energy increases. The first regime is at $t_p < 20$ ns, where t_p is laser pulse width. In this regime, the material structure changes from amorphous state to crystalline state. The second regime is 20 ns $< t_p <$ 34 ns, in which the pattern with a "Sombrero" shape is formed, which is because the chemical capillary effect is gradually strengthened to its peak value, and continuously carries the molten materials toward the central region. The third regime is at $t_p > 34$ ns, in which the thermal capillary effect becomes dominant. The central dome is gradually broadened following the increase of pulse energy, and the bowl shape patterns become deep. If the pulse energy is further increased, the regime may be ablated to form a central hole, which corresponds to the molten ablation threshold lithography. Comparison among Fig. 9.6a–f shows that the pattern size can be controlled by changing the laser pulse energy. The below-diffraction-limited pattern structures may be obtained by using laser pulse energy close to the molten ablation threshold temperature.

9.2.4 Pattern Application: Grayscale Lithography

Figure 9.6 shows the pattern structures with different height and size can be generated on the Sb_2Te_3 thin films by precisely changing the laser pulse energy. It is well known that the reflection spectra are directly dependent on the height and size of pattern structure; thus, the grayscale images can be inscribed on the Sb_2Te_3 thin films through precisely controlling the laser pulse energy.

Figure 9.7 presents the continuous-tone grayscale patterns inscribed on Sb_2Te_3 thin films by the direct laser writing. The different gray levels are realized by adjusting laser energy, where a high power laser yields a low reflectance. Each pixel of the image is assigned to a laser power to obtain certain reflectance. The high-resolution optical image shown in Fig. 9.7b is lifelike bird with fine structures, for example, the eyes can be seen in the inset observed by the AFM topographic image. The AFM topographic image is very consistent with the corresponding optical image marked with red circle in Fig. 9.7b. The black eyeball has larger microscopic structures than the white eyeball. Figure 9.7a is original image of bird; comparing Fig. 9.7a with Fig. 9.7b, one can see that the appearance and expression of bird have been fully written on the Sb_2Te_3 thin films.

Figure 9.7c is the original image of lifelike cat. One can see from the cat's eyes that the cat is concentrating its attention on something. The lifelike cat is written on the Sb_2Te_3 thin films, shown in Fig. 9.7d. By comparing Fig. 9.7d with Fig. 9.7c, the expression and appearance of the lifelike cat are fully inscribed on Sb_2Te_3 thin films, which can be seen from fire-eyed eyes, and some fine structures like furs and whiskers.

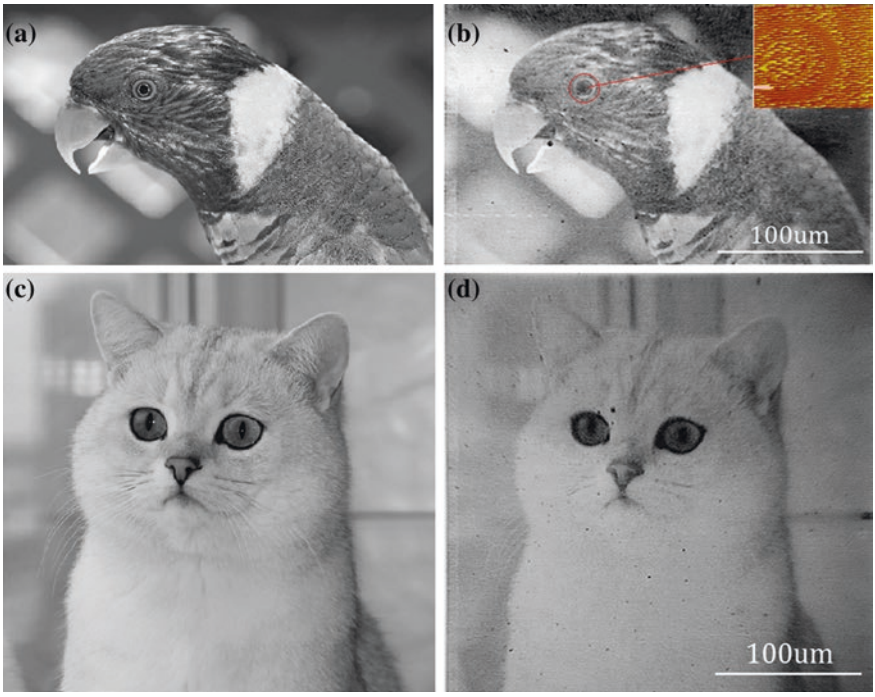

Fig. 9.7 Grayscale images written on Sb_2Te_3 phase change thin films, the lifelike bird images **a** original **b** grayscale image written onto Sb_2Te_3 thin film, and the lifelike cat images **c** original and **d** grayscale image inscribed onto Sb_2Te_3 thin film [25]

9.3 Nanolithography by Combination of Saturation Absorption and Thermal Threshold Effects

9.3.1 Basic Principle

The sample structure is "substrate/nonlinear saturation absorption thin film/lithography layer." A visible laser beam is focused onto the nonlinear saturation absorption layer by a converging lens. A far-field distance exists between the converging lens and sample. This far-field distance is very easy to operate, as shown in Fig. 9.8. For the nonlinear saturation absorption materials, a higher laser beam intensity can increase the transmittance. The largest transmittance is at the center of the spot, which can cause a nanochannel to be formed at the center of spot. The focused laser beam spot passes through the nanochannel. A nano-optical spot is generated at the interface between the nonlinear saturation absorption layer and the lithography material. The spot reduction could be tuned by changing the nonlinear saturation absorption coefficient, laser intensity, and the thin film thickness. The nano-optical spot is directly coupled to the lithography layer. Hence, nanostructure lithography by optical means can be performed.

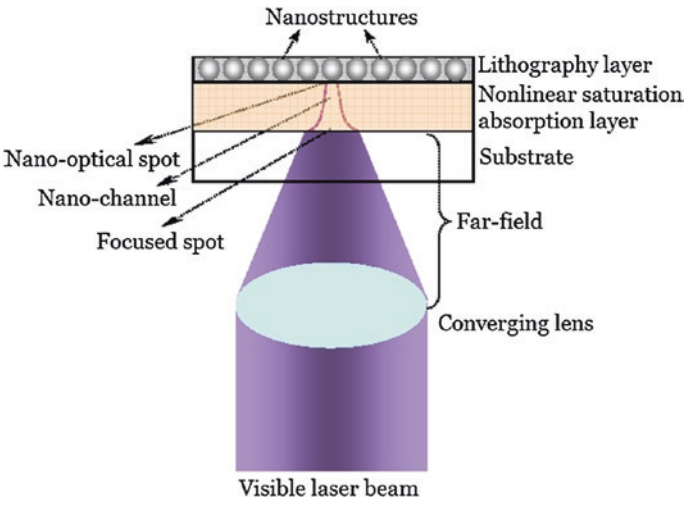

Fig. 9.8 Schematic of nanolithography by nonlinear saturation absorption effect. Reproduced from [26] by permission from The Royal Society of Chemistry

9.3.2 Nanoscale Lithography Induced by Si Thin Film with 405-nm Laser Wavelength

The optical nonlinear absorption effect of the Si thin film under 405-nm laser irradiation is measured by the traditional *z*-scan method. The Si thin film is directly deposited on the glass substrate by magnetron-controlled sputtering. The thickness is ~100 nm. The laser power is ~0.5 mW. The measured *z*-scan curve is given in Fig. 9.9. The Si thin film clearly presents nonlinear saturation absorption. The curve indicates that the $\beta \sim -10^{-2}$ m/W, which signifies a large nonlinear saturation absorption effect.

The direct laser writing lithography is carried out using a high-density digital versatile disk tester with a laser wavelength of 405 nm and a converging lens with NA of 0.65. The diffraction-limited focal spot size of the apparatus is theoretically ~760 nm . The Si thin film is used as the nonlinear saturation absorption layer. The AgInSbTe thin film is selected as the lithography material due to the ability to form bumps under laser irradiation [14]. Under the irradiation of laser pulse with a power density of $10^9 \sim 10^{10}$ W/m^2, AgInSbTe thin film absorbs the laser energy and is heated to melting temperature of about 500 °C and then becomes molten state. The bump structure is immediately formed due to large difference of thermal expansion coefficient between the solid and molten states.

In direct laser writing lithography, the writing speed is fixed at 3 m/s. A recordable digital versatile disk, without depositing recording materials, is used as the sample substrate. The Si thin film is first deposited on the disk substrate.

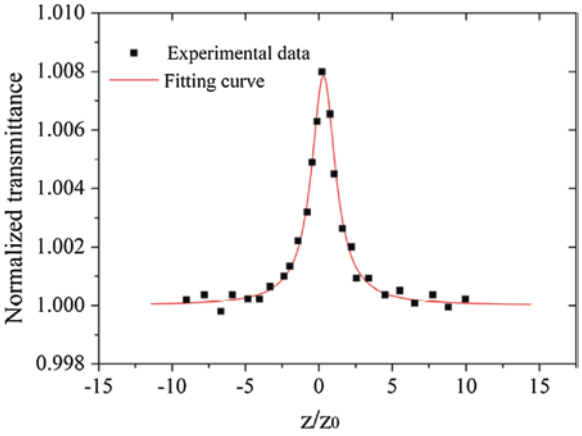

Fig. 9.9 Nonlinear absorption measurements at 405 nm light wavelength for Si thin films. Reproduced from [26] by permission from The Royal Society of Chemistry

Afterward, the AgInSbTe thin film is deposited on the Si thin film. Figure 9.10 gives the optimized experimental results, where the experimental parameters are as follows: laser power $P = 3.75$ mW, the laser pulse width is 65 ns, and the thickness of AgInSbTe and Si thin films is about 20 and 100 nm, respectively. One can find that the nanobump structures are rapidly formed in the recording grooves of the disk substrate (Fig. 9.10a). Every nanobump is formed by a laser pulse. Figure 9.10 gives the magnified images of the nanobumps. The nanobump is $\sim$ 80 nm in diameter, and this size is only about 1/10 the original spot size. In addition, the molten threshold effect of AgInSbTe thin film is another contribution to the reduction of bump size.

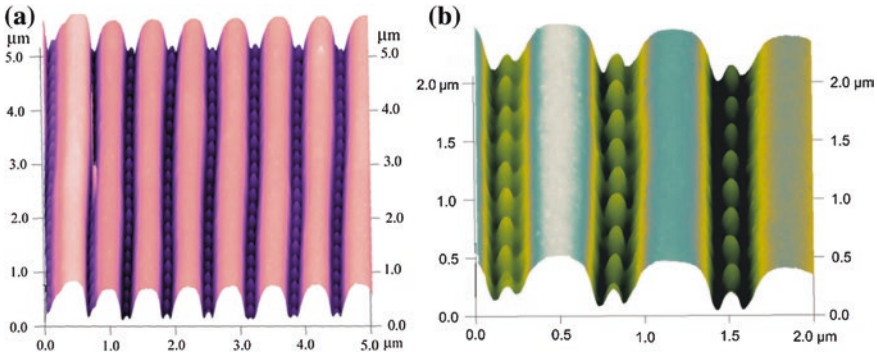

Fig. 9.10 Experimental results of optical nanolithography, **a** nanobump structures formed in the grooves of disk substrate, and **b** the magnified image of the nanobump structures. Reproduced from [26] by permission from The Royal Society of Chemistry

9.4 Nanolithography by Combination of Reverse Saturation Absorption and Thermal Diffusion Manipulation

In direct laser writing thermal lithography, the thermal lithographic material absorbs the laser energy and is heated to certain temperature, such as phase-change or melting points, to form various patterns on the material surface. Generally speaking, the pattern resolution is restricted by the optical diffraction limit. However, the resolution can be improved by using lithographic materials with thermal lithographic threshold and nonlinear reverse saturation absorption characteristics, where nonlinear reverse saturation absorption can induce the formation of below-diffraction-limited energy absorption area (spot), determining the pattern resolution [15].

9.4.1 Formation of Below-Diffraction-Limited Energy Absorption Spot

9.4.1.1 Theoretical Analytical Model

A collimated Gaussian laser beam is focused onto a nonlinear reverse saturation absorption thin film. The focused spot intensity presents typical Gaussian profile, as shown in Fig. 9.11, where the focused spot causes the absorption coefficient to become Gaussian profile, i.e., the absorption coefficient at the central region of spot is large and then decreases along radial direction. The higher absorption coefficient at the central area of spot leads to larger central energy absorptivity; thus, a below-diffraction-limited energy absorption area is generated at the center of focused spot, which is called as effective energy absorption spot.

To determine the morphology of below-diffraction-limited energy absorption spot, the nonlinear reverse saturation absorption thin film is considered as a multilayer system (e.g., m layers), and the thickness of every layer is ΔL. One has $m \times \Delta L = L$, where L is thin film thickness. The light intensity and absorption coefficient of ith layer are marked as I_i and α_i, respectively, where $i = 1, 2, 3, \ldots, m$. The incident spot intensity profile on the r–z cross section is written as

$$I_{\mathrm{incz}}(r,z) = \frac{2P}{\pi w^2(z)} \exp\left[-2r^2/w^2(z)\right], \quad \text{and} \quad w(z) = w_0\sqrt{1+(z/z_0)^2} \tag{9.1}$$

where w_0 is beam waist radius of spot and corresponds to $w(z)$ value at $z = 0$ in formula (9.1), $z_0 (= \frac{\pi w_0^2}{\lambda})$ is Rayleigh length, λ is incident laser wavelength, and P is laser power. The light is focused on the sample surface. On the nonlinear thin film surface, $z = 0$, $w(z = 0) = w_0$, and $I_{\mathrm{inc}}(r, z = 0) = \frac{2P}{\pi w_0^2} \exp\left[-2r^2/w_0^2\right]$.

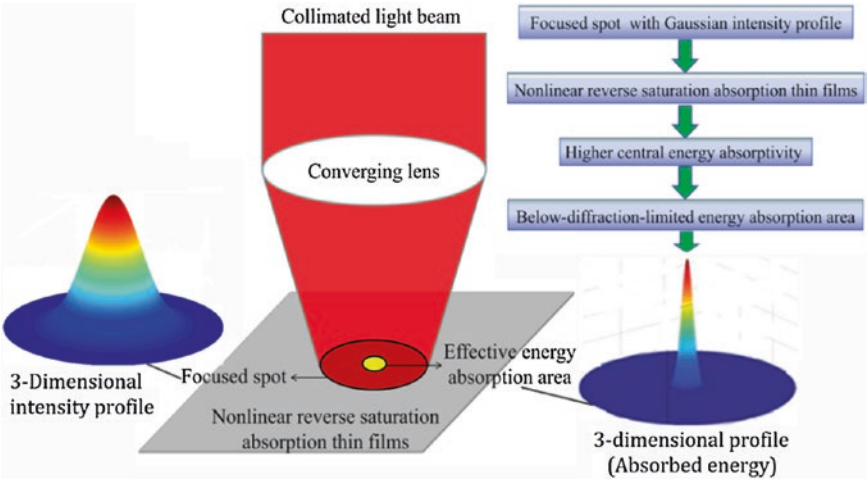

Fig. 9.11 The schematic of the nonlinear reverse saturation absorption induced formation of below-diffraction-limited energy absorption area. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

The light goes through the first layer and decays into the second layer due to the absorption effect, as shown in Fig. 9.12. Light travels through the every layer, and the absorbed energy at unit volume and unit time can be expressed as

$$\Delta E = \mathrm{d}I/\mathrm{d}z = -\alpha_0 I - \beta I^g \tag{9.2}$$

where α_0 is the linear absorption coefficient, I is incident laser intensity, and β is nonlinear absorption coefficient. Generally speaking, for semiconductor thin film

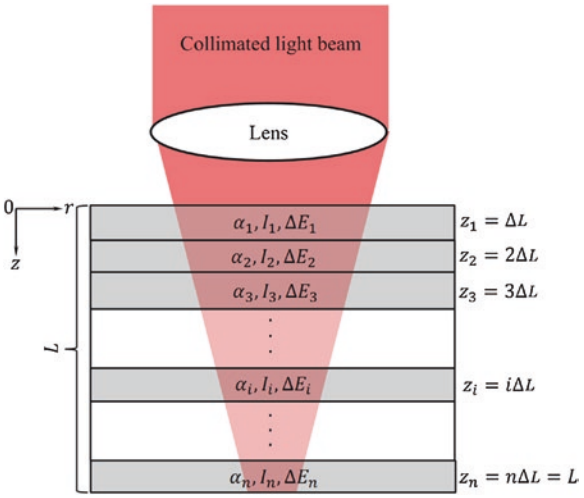

Fig. 9.12 The multilayer analytical model for the nonlinear reverse saturation absorption induced formation of below-diffraction-limited energy absorption spot. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

materials, $\beta > 0$ means nonlinear reverse saturation absorption, and $\beta < 0$ means nonlinear saturation absorption. g is the effective multi-photon absorption level, $g = 1$ means one-photon absorption, $g = 2$ is two-photon absorption, and $1 < g < 2$ is an effective hybrid multi-photon absorption resulting from the one-photon absorption and two-photon absorption processes. In order to simplify the calculation, here, one can consider the two-photon absorption effect, that is, $g = 2$. Thus,

$$\Delta E = \mathrm{d}I/\mathrm{d}z = -\alpha_0 I - \beta I^2 = -\alpha I, \quad \text{with} \quad \alpha = \alpha_0 + \beta I \tag{9.3}$$

The 1st layer can be approximately considered to be sample surface, $z = z_1 = \Delta L$; thus,

$$\alpha_1 = \alpha_0 + \beta I_{\mathrm{inc}}, \quad \Delta L_{\mathrm{eff}} = \frac{1 - e^{-\alpha_1 \Delta L}}{\alpha_1}, \quad I_1 = \frac{I_{\mathrm{inc}} e^{-\alpha_1 \Delta L}}{1 + \beta I_{\mathrm{inc}} \Delta L_{\mathrm{eff}}}, \quad \Delta E_1 = -\alpha_1 I_1, \tag{9.4}$$

At the 2nd layer, $z = z_2 = 2\Delta L$, one can obtain

$$\alpha_2 = \alpha_0 + \beta I_1, \quad \Delta L_{\mathrm{eff}} = \frac{1 - e^{-\alpha_2 \Delta L}}{\alpha_2}, \quad I_2 = \frac{I_1 e^{-\alpha_2 \Delta L}}{1 + \beta I_1 \Delta L_{\mathrm{eff}}}, \quad \Delta E_2 = -\alpha_2 I_2, \tag{9.5}$$

At the 3rd layer, $z = z_3 = 3\Delta L$,

$$\alpha_3 = \alpha_0 + \beta I_2, \quad \Delta L_{\mathrm{eff}} = \frac{1 - e^{-\alpha_3 \Delta L}}{\alpha_3}, \quad I_3 = \frac{I_2 e^{-\alpha_3 \Delta L}}{1 + \beta I_2 \Delta L_{\mathrm{eff}}}, \quad \Delta E_3 = -\alpha_3 I_3, \tag{9.6}$$

At the ith layer, $z = z_i = i\Delta L$

$$\alpha_i = \alpha_0 + \beta I_{i-1}, \quad \Delta L_{\mathrm{eff}} = \frac{1 - e^{-\alpha_i \Delta L}}{\alpha_i}, \quad I_i = \frac{I_{i-1} e^{-\alpha_i \Delta L}}{1 + \beta I_{i-1} \Delta L_{\mathrm{eff}}}, \quad \Delta E_i = -\alpha_i I_i, \tag{9.7}$$

At the nth layer, $z = z_m = m\Delta L$,

$$\alpha_m = \alpha_0 + \beta I_{m-1}, \quad \Delta L_{\mathrm{eff}} = \frac{1 - e^{-\alpha_m \Delta L}}{\alpha_m}, \quad I_m = \frac{I_{m-1} e^{-\alpha_m \Delta L}}{1 + \beta I_{m-1} \Delta L_{\mathrm{eff}}}, \quad \Delta E_m = -\alpha_m I_m, \tag{9.8}$$

Formulas (9.4)–(9.8) can calculate the profiles of absorbed energy, absorption coefficient, and light intensity inside the nonlinear sample. The light intensity profile of the exiting spot from the sample can be actually approximated to the light field of the last layer (the mth layer).

In the multilayer analytical model, the broadening effect of focused beam in the traveling process inside the sample, resulting from diverging effect, can be actually neglected, which is reasonable for thin sample. In the traveling, the beam size $w(z)$ along optical axis (z-axis) can be calculated by $w(z) = w_0\sqrt{1 + (z/z_0)^2}$ [also see formula (9.1)]. For example, for direct laser writing lithography with a setup of NA = 0.85 and $\lambda = 650$ nm, the calculated light beam change along z-axis inside the thin sample with a thickness of 80 nm

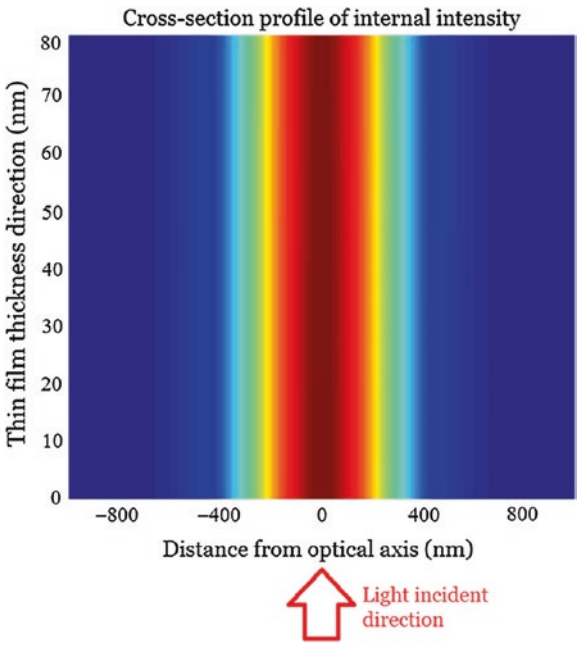

Fig. 9.13 Calculated change of the incident beam along z-axis for NA = 0.85, and $\lambda = 650$ nm (without nonlinear absorption and refraction effects being considered)

is shown in Fig. 9.13, where without any nonlinear absorption and refraction are considered. One can find that the beam spot size basically remains unchanged when it travels about 80 nm along z-axis (sample thickness direction). That is, the beam diverging effect is negligible.

9.4.1.2 Effective Energy Absorption Spot Analysis

1. **Theoretical calculation**

According to the aforementioned requirements for the lithographic material, AgInSbTe chalcogenide phase-change thin films are chosen as an example. The AgInSbTe has a low molten ablation threshold temperature (about 500 °C), which is suitable for low power laser writing especially above and near the melting point, and a low thermal diffusivity [16]. More importantly, the AgInSbTe possesses strong nonlinear reverse saturation characteristic, i.e., the absorption coefficient increases with the increase in light intensity.

The focused spot-induced energy absorption spot is calculated as follows. The laser wavelength $\lambda = 650$ nm, laser power $P = 4$ mW, the converging lens NA = 0.85, and thin film thickness $L = 80$ nm. The linear absorption coefficient $\alpha_0 = 3 \times 10^7$/m, and nonlinear absorption coefficient $\beta = 7.53 \times 10^{-3}$ m/W. In order to simplify the calculation, the effective nonlinear absorption level $g \sim 2$ is chosen according to [17].

The $g \sim 2$ means equivalent two-photon absorption effect. Generally speaking, the direct two-photon and multi-photon absorptions occur when ultrafast laser pulse irradiates on the sample, inducing ultrafast bound electronic nonlinear effect. However, besides ultrafast nonlinearity, in the nanosecond scale of the laser pulse, the nonlinear effects can be attributed more to the free-carriers absorption process for semiconductor materials. The free-carriers absorption is a cumulative effect, and one-photon absorption-induced free-carriers once absorb one photon sequentially, which is equivalent two-photon absorption effect. That is, the equivalent two-photon absorption effect comes from the laser-induced free-carriers absorption.

For the metallic-type AgInSbTe semiconductor thin films, the free-carriers can be easily created by laser irradiation because of semiconductor-like energy-band structure. Despite the population decays during the laser pulse, the free-carriers density is locally balanced and approaches to the saturation state when the laser intensity increases to certain threshold. When the laser intensity increases above the threshold, the free-carriers absorb photons to higher states, namely intra-band absorption, and then, these higher state free-carriers transform the relaxation energy by the phonons of crystal lattice in a longer time range than the laser pulse, transferring the energy to the lattice vibration; for the whole process, the energy transfers from photons to the films. The lattice is then heated; the lithographic marks can be formed due to molten ablation effect when the lattice temperature exceeds to the melting point. The lithographic mark size might be determined by the free-carriers-induced equivalent two-photon absorption effect, which can be reflected from the effective energy absorption spot.

The calculated spot intensity profile is presented in Fig. 9.14. Figure 9.14a is the normalized incident focused spot intensity profile, and Fig. 9.14b is the normalized exiting spot intensity profile through nonlinear thin film. By comparing Fig. 9.14a, b, one can find that the exiting spot becomes obviously larger than the incident focused spot itself, which is because the most light energy at the central region of focused spot is absorbed by the nonlinear sample, and the energy absorptivity decreases along the radial direction. These characteristics induce the formation of below-diffraction-limited energy absorption spot. Figure 9.14c is the normalized intensity profile of energy absorption spot. Comparison of Fig. 9.14a, c indicates that the effective energy absorption spot is smaller than incident focused spot. In order to further compare the size among focused spot, exiting spot, and energy absorption spot, Fig. 9.14d gives cross-sectional profile curves of Fig. 9.14a–c. One can see that the full width at half maximum (FWHM) of exiting spot is increased to ~1.56 times the incident focused spot. The FWHM of energy absorption spot is reduced to about ~75 % the incident focused spot. Actually, in direct laser writing, the lithographic resolving limit is mainly determined by the energy absorption spot. Thus, the formation of below-diffraction-limited energy absorption spot is one of critical factors.

2. **Experimental observation**

The AgInSbTe thin films with a thickness of about 80 nm are deposited on glass substrate by direct current magneto-controlling sputtering method. A scanning near-field optical probe setup is established to directly observe the exiting spot

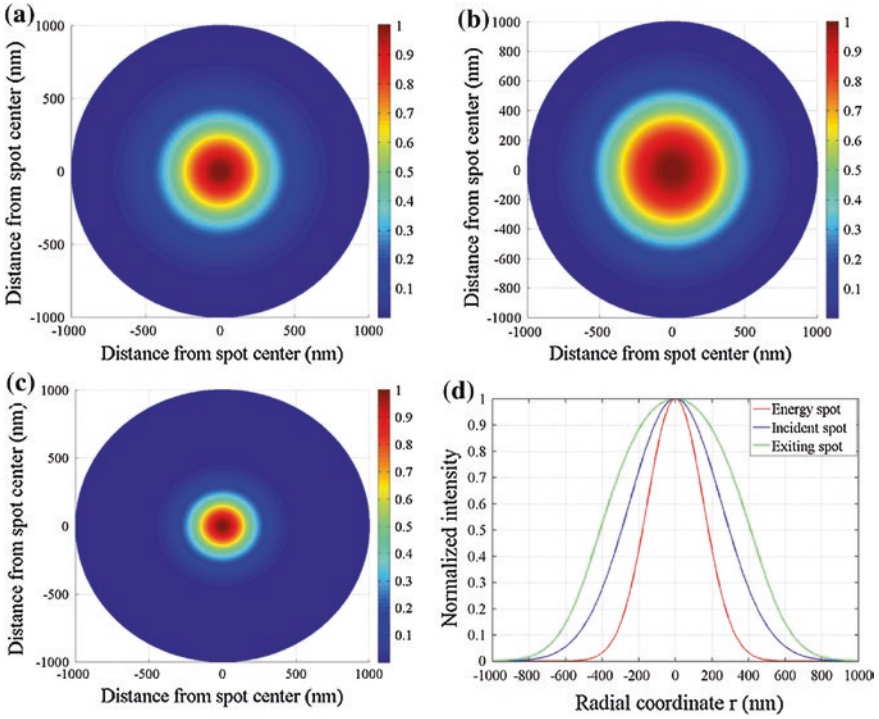

Fig. 9.14 The typical calculated results for the formation of below-diffraction-limited energy spot, the normalized two-dimensional intensity profile of **a** focused spot, **b** exiting spot, **c** energy absorption spot, and **d** cross-sectional profile curves of **a**–**c**. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

through the sample, where a laser diode with a wavelength 635 nm is used as the light source, as illustrated schematically in Fig. 9.15. A collimated laser beam is focused by an objective lens with NA=0.25 and then irradiated on the nonlinear thin film. A metal-coated fiber tip affiliated onto the tuning fork is used to scan the laser beam spot intensity distribution along the sample surface, and the tuning

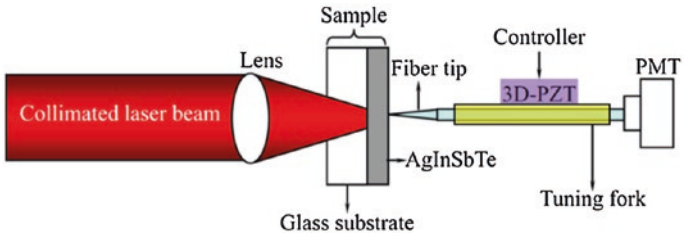

Fig. 9.15 Schematic of near-field spot scanning setup. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

fork is attached on a three-dimensional piezoelectric tube (PZT), and the lateral scanning resolution is about 100 nm limited by the aperture of the fiber tip. The distance between the sample surface and the fiber tip is modulated by an atomic force microscopy (AFM) controller. When the fiber tip goes close to the sample surface, the resonance frequency of the tuning fork would shift due to the generated shear force. The AFM controller receives this shift signal as error and regulates the z-axis voltage onto the PZT. The exiting spot information is collected by scanning fiber tip and then transferred to the PMT and produces the signal in the format of voltage to the computer.

In order to obtain the incident focused spot intensity profile, the exiting spot intensity profile through the glass substrate with a thickness of 1.2 mm is presented in Fig. 9.16a. The spot presents approximately elliptical shape, which may be due to experiencing laser diode's junction effect. The exiting spot through "glass/AgInSbTe thin film" sample under the laser power of 9 mW is scanned along AgInSbTe surface (also see Fig. 9.16b). Comparing Fig. 9.16a with b, the spot morphology of Fig. 9.16b is almost the same as the Fig. 9.16a, while the spot size of Fig. 9.16b is obviously larger than Fig. 9.16a, and the central intensity profile becomes flattened. These indicate that the nonlinear

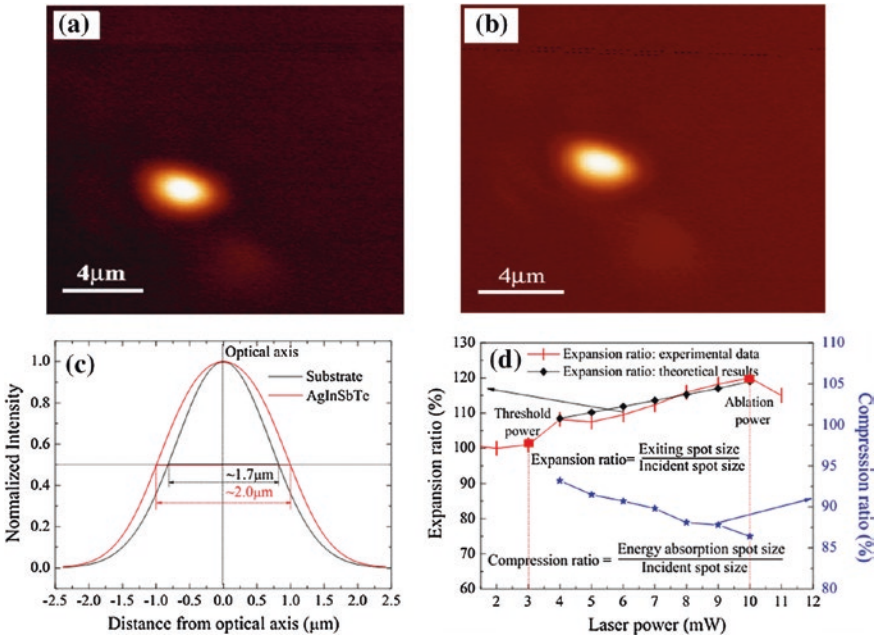

Fig. 9.16 Experimental results of the spot obtained by the near-field spot scanning method, **a** spot through glass substrate, **b** spot through "glass/AgInSbTe" sample, **c** the cross-sectional profile curves of **a** and **b** along short axis direction, and **d** the dependence of expansion and compression ratios on laser power. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

reverse saturation absorption effect happens in the AgInSbTe thin film and causes the exiting spot size to become large and the spot top to become smooth.

To further compare the intensity distribution of spot through the sample with and without the AgInSbTe thin films, the normalized cross-sectional profile curves along the short axis direction of spot images of Figs. 9.16a, b are presented in Fig. 9.16c. The top of spot through the AgInSbTe sample is smoother than that through the glass substrate, and the FWHM of through AgInSbTe sample is expanded to ~118.2 % and increased from ~1.7 to ~2.0 μm. Here, one can define the ratio of exiting spot size to incident spot size as expansion ratio. The dependence of expansion ratio on the laser power is presented in red curve of Fig. 9.16d. One can see that the threshold power is about $P = 3$ mW, and the expansion ratio increases with increasing laser power for 10 mW > P > 3 mW. The ablation takes place at $P > 10$ mW. In Fig. 9.16d, the black curve is calculated according to the theoretical model. Comparison of experimental data and theoretical results shows that the theoretical analysis is basically consistent with experimental results. Here, one also defines the ratio of energy absorption spot size to incident spot size as compression ratio. The energy absorption spot size can be calculated using the formulas (9.4)–(9.8) and the experimental data of exiting spot size. The blue curve in Fig. 9.16d is the dependence of compression ratio on laser power for $10\,\text{mW} > P > 3\,\text{mW}$, and the compression ratio reduces with laser power increasing. That is, the higher laser intensity can lead to more light energy absorption, and the absorption coefficient increases with the laser intensity increasing. Thus, the higher laser intensity can obtain much smaller energy absorption spot. However, the laser intensity cannot be infinitely increased due to ablation effect.

9.4.2 Thermal Diffusion Manipulation by Thermal Conductive Layer

9.4.2.1 Basic Schematic of Heat Flow Manipulation

The thermal diffusion is also one of the critical factors determining the lithographic resolving limit. It is possible to control the heat flow direction so that the thermal energy is concentrated on central region of the spot and causes the spot center temperature to increase beyond the certain threshold temperature, such as molten ablation temperature. In other words, the central region of spot is locally heated above melting point, while the rest remains intact, which causes the lithographic mark size to be much smaller than the energy absorption spot. The low thermal diffusivity coefficient can lead to heat energy being instantaneously concentrated at the central region of spot if the material is directly written with a short pulse. By introducing a thermal conductive layer below the AgInSbTe thin

film, only the apex part of the energy absorption spot actually melts because the heat diffuses quickly into the thermal conductive layer, i.e., the thermal diffusion process can be effectively controlled so that the very small apex region of energy absorption spot is instantaneously reached up to melting point; thus, the nanoscale lithographic marks are written.

Figure 9.17a gives the sample structure, where an Al thin film with a thickness of about 80 nm is used as thermal conductive layer. The thermal diffusivity coefficient of thermal conductive layer is marked as σ_{th}. The Al thin film is deposited below the AgInSbTe thin film. The thermal diffusivity coefficient of crystalline AgInSbTe is about $\sigma_{AgInSbTe} = 1.85\ mm^2/s$ [16], which is larger than that of the glass substrate of $\sigma_{glass} = 0.558\ mm^2/s$ [18]. If there is no thermal conductive layer, the sample structure is "glass/AgInSbTe," and $\sigma_{th} = \sigma_{glass}$. The thermal diffusion is mainly along the in-plane of AgInSbTe thin film, which leads to the large lithographic mark. However, the thermal diffusion route can be changed when the Al thin film is deposited below the AgInSbTe thin film. The Al layer is used as the thermal conductive layer and has a thermal diffusivity coefficient of $\sigma_{th} = \sigma_{Al} = 97.5\ mm^2/s$ [19], which is much larger than σ_{glass}. The diffusion route is now mainly perpendicular to the AgInSbTe thin film plane, which results in a heat concentration on the central apex of energy absorption spot.

The heat propagation can be roughly divided into two channels: One channel is along the in-plane of AgInSbTe thin film, and the other is perpendicular to the AgInSbTe thin film, which is marked as out-of-plane of AgInSbTe thin film. Figure 9.17b gives the schematic of heat propagation direction, where the cylinder marked in green is a heated volume by incident focused spot, and the cylinder diameter is D. The heated volume is actually a heat source. The thermal diffusion takes place when the temperature of heated volume exceeds to the environment temperature. The thermal diffusion can happen along two channels: one is in-plane of AgInSbTe thin film, which is marked in blue arrow. The heat flows through the

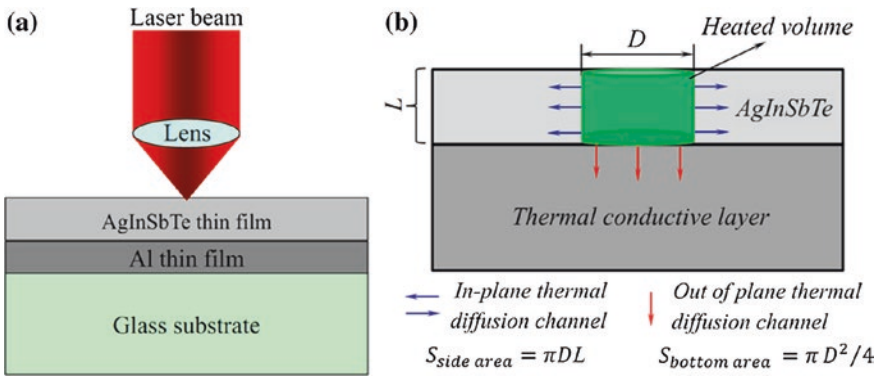

Fig. 9.17 Heat flow manipulation by introducing thermal conductive layer, **a** sample structure, and **b** thermal diffusion direction. Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

side face of heated volume, the area is $S_{\text{side area}} = \pi DL$, and the in-plane thermal diffusion channel can cause the lithographic mark to become large. The other is out-of-plane of AgInSbTe thin film, which is marked in red arrow. The heat flows through the bottom surface of the heated volume, the area is $S_{\text{bottom area}} = \pi D^2/4$, and the out-of-plane thermal diffusion channel makes the lithographic mark size become small.

9.4.2.2 Determination of Heat Quantity Propagation Channel

The characteristic thermal diffusion length l_{th} can be used to evaluate the performance of thermal diffusion channel. The l_{th} can be estimated as

$$l_{\text{th}} = \sqrt{\sigma_{\text{AgInSbTe}} t_p} \tag{9.9}$$

If the laser pulse of $t_p = 50$ ns is used in the direct laser writing, then $l_{\text{th}} \approx 300$ nm, which is comparable to the radius of energy absorption spot. $l_{\text{th}} \approx 300$ nm is larger than the thin film thickness of $L = 80$ nm. The heat can propagate a distance of about 300 nm within the pulse width along both in-plane and out-of-plane of thin film thickness. However, if the thermal resistance at the interface is ignorable, the central heat quantity of energy absorption spot is easy for traveling along the thin film thickness direction and going into the thermal conductive layer because $l_{\text{th}} > L$. The temperature at the central region of the spot will transiently reach up to lithographic threshold point. If $\sigma_{\text{th}} < \sigma_{\text{AgInSbTe}}$, the thermal diffusion along the thin film thickness direction will be hindered when the heat quantity propagates to the thermal conductive layer, which makes the heat propagation direction change from out-of-plane direction to in-plane direction; thus, the lithographic mark size becomes large. On the contrary, if $\sigma_{\text{th}} > \sigma_{\text{AgInSbTe}}$, the thermal diffusion is along the out-of-plane direction. The temperature at the central volume of spot will transiently reach up to melting point and the small lithographic mark can be transiently generated. The fusion latent heat released by the molten ablation can quickly diffuse along the out-of-plane direction, enter the thermal conductive layer, and dissipate into air.

The heat quantity, which propagates along in-plane, is proportional to the thermal diffusivity coefficient σ_{AgInSbTe} of the AgInSbTe as

$$Q_{\text{in plane}} \propto \sigma_{\text{AgInSbTe}} S_{\text{side area}} \tag{9.10}$$

The heat quantity, which propagates along out-of-plane, is proportional to the thermal diffusivity coefficient σ_{th} of the thermal conductive layer as

$$Q_{\text{out of plane}} \propto \sigma_{\text{th}} S_{\text{bottom area}} \tag{9.11}$$

The competition between the two channels is conveniently noted as the proportion η

$$\eta = \frac{Q_{\text{in plane}}}{Q_{\text{out of plane}}} = \frac{4L\sigma_{\text{AgInSbTe}}}{D\sigma_{\text{th}}} \tag{9.12}$$

The η determines the heat diffusion channel. $\eta > 1$ means that the thermal diffusion through in-plane channel is dominant. On the contrary, $\eta < 1$ means that the thermal diffusion along out-of-plane channel is dominant, and the heat quantity quickly goes into the thermal conductive layer and concentrates on the apex of the energy absorption spot. The η value for different sample structure can be calculated by the following parameters: energy absorption spot size $D \approx 750$ nm, $L = 80$ nm, $\sigma_{Al} = 97.5\,mm^2/s$, $\sigma_{glass} = 0.558$, and $\sigma_{AgInSbTe} = 1.85\ mm^2/s$. For "galss substrate/AgInSbTe" sample, there is no thermal conductive layer, $\sigma_{th} = \sigma_{glass}$, and $\eta \sim 1.42 > 1$, the thermal diffusion is mainly along in-plane of AgInSbTe thin film. However, when the Al thin film is chosen as thermal conductive layer and inserted between the AgInSbTe and glass substrate, $\sigma_{th} = \sigma_{Al}$, and $\eta \sim 0.008 \ll 1$, which indicates that the heat quantity is transiently localized onto the apex of energy absorption spot and goes quickly into the Al layer.

9.4.3 Experimental Nanolithography Marks

In the direct laser writing, the laser beam with a wavelength of 650 nm is modulated into pulse light by a signal generator, and the pulse width could be tuned by built-in software in the signal generator. The modulated pulse light is directly focused upon the surface of the AgInSbTe thin film by a converging lens with an NA of 0.85. The focused spot size is about 930 nm.

The "glass substrate/AgInSbTe" sample is written to analyze the influence from the thermal conductive layer. The laser writing power is about 4 mW and pulse width is 50 ns. The lithographic marks are observed by a laser confocal scanning microscopy with a light wavelength of 248 nm and objective lens with NA = 0.95.

When $t_p < 50$ ns, the formed marks become very blurry and indistinguishable, as shown in Fig. 9.18a, where $t_p \sim 40$ ns. When the pulse width is increased to $t_p \sim 80$ ns, the mark size becomes large and the image is clear, as shown in Fig. 9.18b, where both the width and the length of marks are about 440 nm, and the height is only about 3 nm, which can be seen from the inset of Fig. 9.18b.

Figure 9.19a presents the optimal results of lithographic mark trains with $t_p = 50$ ns. The mark size is about 250 nm, which is about 1/4 the focused spot size, and is also smaller than the energy absorption spot. The reduction of lithographic mark size is due to two factors: one is the nonlinear reverse saturation absorption—induced the formation of below-diffraction-limited energy absorption spot, and the other is lithographic threshold effect. The formation of lithographic marks is actually the molten ablation when the laser pulse heats the sample to the melting temperature of about 500 °C. Since the intensity distribution of energy absorption spot is approximately Gaussian, by tuning the laser pulse power and width, the molten ablation can be controlled at the central region of 80 % maximum intensity of the energy absorption spot, and the rest area remains intact. Thus, the lithographic mark size is further smaller than the energy absorption spot (also see Fig. 9.19b). The lithographic mark size is actually dependent on pulse

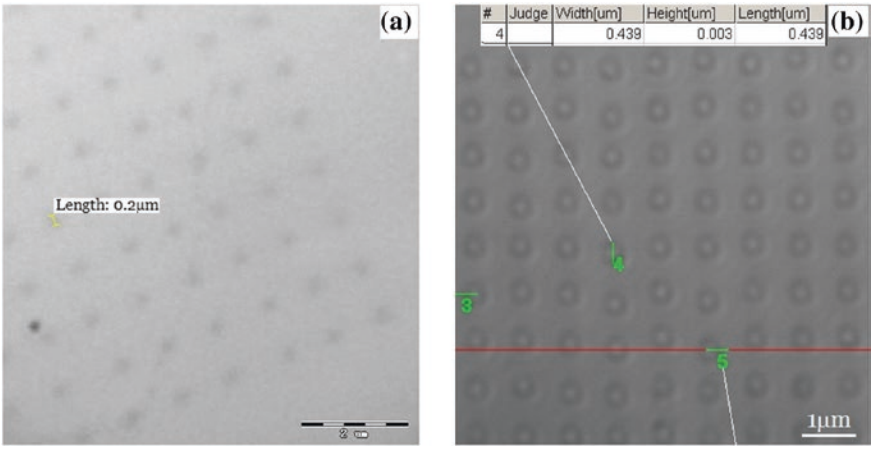

Fig. 9.18 Confocal laser scanning microscopy images of mark matrix for **a** $t_p \sim 40$ ns, and **b** $t_p \sim 80$ ns

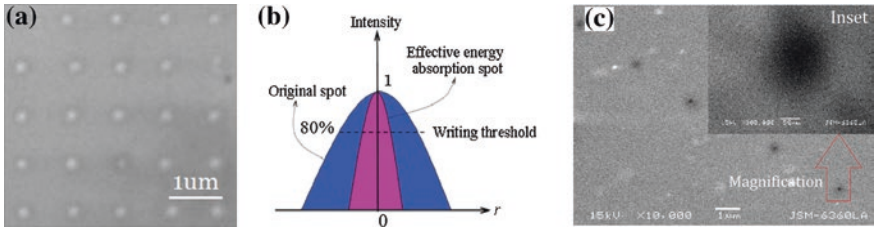

Fig. 9.19 Nanolithographic mark trains obtained with direct laser writing method, laser wavelength of 650 nm, laser power 4 mW and pulse width 50 ns, NA = 0.85, **a** lithographic marks without thermal conductive layer (mark size of about 250 nm), **b** schematic of formation of below-diffraction-limited marks, and **c** lithographic marks with thermal conductive layer (mark size of about 100 nm). Reprinted with permission from [19]. Copyright 2014, American Institute of Physics

width; generally speaking, short pulse width can cause small mark size. However, the mark size cannot be reduced infinitely.

The lithographic mark size is further reduced by introducing an Al thin film as thermal conductive layer, where the Al thin film is inserted between AgInSbTe and glass substrate. Figure 9.19c gives the lithographic marks observed by scanning electronic microscopy, and the inset is a magnification of lithographic mark pointed out in red arrow, and the scale line in the inset is 50 nm. The mark size is about 100 nm, which is only about 1/10 the incident focused spot. Compared with Fig. 9.19a, the Al thermal conductive layer makes the heat propagate along out-of-plane and be instantaneously localized onto the apex of energy absorption spot. The lithographic marks occur when the temperature exceeds to the molten ablation threshold; thus, the lithographic mark is further reduced down to about 1/10 the focused spot size.

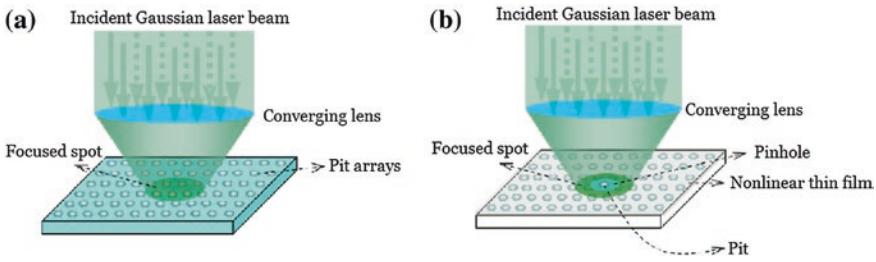

Fig. 9.20 Schematic for the working principle of designed nonlinear super-resolution imaging, **a** conventional light spot scanning imaging, **b** super-resolution spot scanning imaging. Reprinted from [27], with kind permission from Springer Science+Business Media

9.5 Nonlinearity-Induced Super-Resolution Optical Imaging

9.5.1 Basic Principle Schematics

For conventional scanning laser microscopy, the spatial resolving power is restricted by the optical diffraction limit, which is determined by the laser wavelength and the NA of the converging lens. Actually, a nonlinear thin film can be applied in scanning laser microscopy to achieve the super-resolution imaging. For nonlinear imaging effect, an analysis of the frequency response function would be adequate to demonstrate the super-resolution. If the frequency response function exceeds the cutoff frequency f_c, the scanning microscopy system is expected to achieve super-resolution beyond the diffraction limit. To understand this, before involving a rigorous mathematical process, the schematic to show the focused light spot interacting with the nonlinear sample is presented. In Fig. 9.20a, a focused light spot covers an area on the sample surface where the pit arrays are fabricated, and one would expect the resolution to be limited by the spot size. However, as shown in Fig. 9.20b, the resolution may be improved if the light spot distribution is so manipulated that fine structures both in phase and in amplitude appear in front of the light spot, such as a pinhole or a narrow ring. In laser scanning microscopy, the light spot is moving at a high speed across the sample, the resolution is expected to exceed the cutoff spatial frequency because the light spot with fine structure scans across the sample.

9.5.2 Theoretical Description

The manipulation of the light spot is realized with the nonlinear thin film coated above the sample surface. The complex refractive index of the thin film changes according to the local laser intensity $I(r)$ in polar coordinate (θ, r) as

$$\tilde{n}(r) = \left[n_0 + \gamma I(r)\right] + i\frac{\lambda}{4\pi}[\alpha_0 + \beta I(r)] \tag{9.13}$$

where n_0 is the linear refractive index, γ is the nonlinear refraction coefficient, α_0 is the linear absorption, and β is the nonlinear absorption coefficient. The incident laser beam at the focal plane is considered to have a complex electric field of Gaussian profile as

$$E(r) = E_0 \exp\left(-r^2/w_0^2\right) \tag{9.14}$$

where E_0 is the on-axis amplitude. Due to the nonlinearity of the thin film, the transmitted light will have in general three principal changes. Firstly, because the thickness of nonlinear thin film is in the range of nanoscale, the amplitude and the wavefront phase of the complex electric field are modulated by the internal multi-interference effect. Secondly, the wavefront phase is also modified by the Gaussian lens-like phase shift. All the changes can be incorporated into the transmitted complex electric field as

$$\tilde{U}(r) = E(r)\tilde{t}(r)\exp\left[i\Delta\varphi(r)\right] \tag{9.15}$$

where the distribution of the Gaussian lens-like phase shift can be written in a linear approximation as

$$\Delta\varphi(r) \approx \frac{4\gamma P}{\lambda\alpha w_0^2}\{1 - \exp\left[-\alpha(r)L\right]\}\exp\left(-\frac{2r^2}{w_0^2}\right) \tag{9.16}$$

The complex transmission for the two-interface layered system in normal incidence is

$$\tilde{t}(r) = \tau(r)e^{i\delta} = \frac{t_{12}t_{23}\exp\left[i\frac{4\pi}{\lambda}\tilde{n}(r)L\right]}{1 + r_{12}r_{23}\exp\left[i\frac{4\pi}{\lambda}\tilde{n}(r)L\right]} \tag{9.17}$$

where $r_{12} = \frac{n_1 - \tilde{n}(r)}{n_1 + \tilde{n}(r)}$, $r_{23} = \frac{\tilde{n}(r) - n_3}{\tilde{n}(r) + n_3}$, $t_{12} = \frac{2n_1}{n_1 + \tilde{n}(r)}$, and $t_{23} = \frac{2\tilde{n}(r)}{\tilde{n}(r) + n_3}$, are respectively, the reflection and transmission coefficients for each interface. n_1 and n_3 are the corresponding refractive indices for air and the substrate, respectively. The inclusion of the complex refractive index, the local nonlinear absorption, and the multi-interference in formulas (9.16) and (9.17) enables us to calculate the exiting field distribution for nanoscale thin films.

Substituting formulas (9.16) and (9.17) into formula (9.15), one finally obtains the transmitted complex electric field as

$$\tilde{U}(r) = E_0\tau(r)\exp\left(-\frac{r^2}{w_0^2}\right)\cdot\exp\left\{i\left[\delta(r) + \frac{4\gamma P\left[1 - e^{-\alpha(r)L}\right]}{\lambda\alpha(r)w_0^2}\exp\left(-\frac{2r^2}{w_0^2}\right)\right]\right\} \tag{9.18}$$

Let us now consider the frequency response function of signal detection. The signal is collected at far-field distance from the object. The signal can be retrieved from the

Fourier transform of the field $\tilde{U}(r)$. In the simplest detection mode, the signal is an integral at far-field over the detection pupil function $p(\rho)$, which corresponds to the same NA of the lens as the spot forming pupil. The detection resolution is evaluated in terms of the frequency response function $R_{\mathrm{Res}}(f)$. Since in the spatial frequency consideration, the object is taken to be a grating with discrete orders, the signal modulation can be found from the interference of the direct beam in the far-field with the similar shifted first-order beams [20]. Assuming that the Fourier transform of the exiting electric field is $\tilde{U}(\rho) = F\left\{\tilde{U}(r)\right\}$, one writes down the frequency response function in the form of a convolution as

$$R_{\mathrm{Res}}(f) = C\left|\left[\tilde{U}(\rho)p(\rho)\right] \otimes \tilde{U}(\rho)\right| \tag{9.19}$$

where C is a constant. The pupil function in Fourier space is a cylinder function: $p(\rho) = 1$ for $\rho \leq 1$, and $p(\rho) = 0$ for $\rho > 1$. Since the convolution covers a far-field range wider than $\tilde{U}(\rho)$, one expects that the far-field is to extend beyond the cutoff frequency. It follows that the frequency response function $R_{\mathrm{Res}}(f)$ will allow finite values to exceed the diffraction-limited cutoff frequency f_c, which is the indication for the super-resolution.

It is interesting to note that the extension of the spatial frequency is in general independent of the sign of the nonlinear index coefficient. Therefore, the dynamic optical pinhole (bleaching), dynamic darkening (the transmission of the pinhole being smaller than the others), or a phase change (nonlinear path length change) can result in super-resolution.

9.5.3 Experimental Testing

Numerical simulation testings are carried out based on the formula (9.19) for the frequency response function. In order for the results to be experimentally verifiable, the simulated sample is chosen to represent a specially designed structure in the laser scanning microscopy. An image of the sample surface is shown in Fig. 9.21, where the grating period is 760 nm, the pit width is 380 nm, and the pit

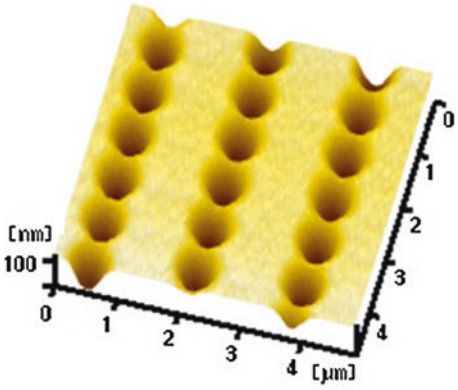

Fig. 9.21 AFM morphology of the sample surface. Reprinted from [28], Copyright 2012, with permission from Elsevier

depth is about 60 nm, that is, there is only a single spatial frequency information on the sample surface. The optical nonlinear thin films are directly deposited on the pit matrix surface.

In the testing, a compact disk player is used as the laser scanning microscope. In the compact disk player, a laser beam is focused onto the sample surface when the sample is moving with a high speed. The focused beam is reflected and collected by the detector via a collecting lens of NA = 0.45. Therefore, the sample surface information is detected point by point. The compact disk player is a typical laser scanning microscope. By analyzing the reflected light intensity change point by point, the sample surface structure can be reconstructed. In the testing, a Si thin film is used as the optical nonlinear thin film. The thin film thickness is fixed at about 100 nm. The laser scanning experiment has a light wavelength of 780 nm, and the power is fixed at 2 mW. Without the nonlinear thin film, the resolving power of the device is diffraction-limited and can be estimated by $\lambda/2\text{NA} = 866$ nm, that is, for a periodic grating with the same pit and space, the smallest pit being resolved by the spot is $\lambda/4\text{NA} = 433$ nm. Therefore, the compact disk player cannot resolve the grating sample with a 760 nm pitch. However, by depositing the nonlinear Si thin film on the grating sample surface, one expects that the resolution of the device would be improved, i.e., the super-resolution is achieved.

The frequency response functions are calculated based on formula (9.19). The nonlinear coefficients at 780 nm light wavelength cannot be obtained; thus, the nonlinear refraction coefficients at 633 nm light wavelength, being $\gamma \approx 4.45 \times 10^{-9}\ \text{m}^2/\text{W}$ [21], and the initial refractive index $n_0 = 4.21$, are used in the calculation, where the absorption effect is not considered in order to simplify the calculation.

The optimal results are presented in Fig. 9.22. Figure 9.22a shows a comparison of the frequency response function for the sample with and without the nonlinear Si thin films. The spatial frequency is normalized to the cutoff frequency in

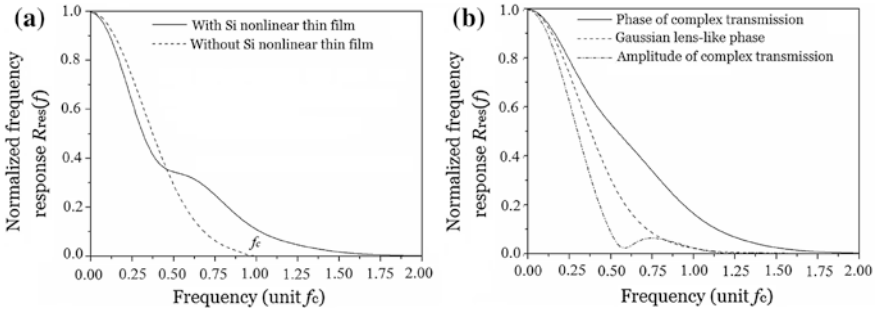

Fig. 9.22 Frequency response function, **a** comparison of frequency response function between with and without nonlinear films, **b** different contributions to frequency response function. Reprinted from [27], with kind permission from Springer Science+Business Media

the case where the sample is not covered by the nonlinear thin film. Figure 9.22a shows that the frequency response function, for the case where the sample is covered by the nonlinear thin film, exceeds the cutoff frequency by about $1.6f_c$, which indicates immediately a super-resolution. In Fig. 9.22b, the frequency response function is further analyzed with the three main contributions, namely phase of the complex transmission, Gaussian lens-like phase, and amplitude of the complex transmission, being singled out one at a time. From the comparison, the improvement of the cutoff frequency stems primarily from the phase change of the complex transmission.

Figure 9.23 presents the experimental results. Figure 9.23a shows the results where the sample surface is covered by a standard Al thin film. The pit matrix is not resolved at all, which is due to the fact that the spatial cutoff frequency of the compact disk player is below the grating spatial frequency. Figure 9.23b shows the results where the sample surface is covered with the nonlinear thin film, where the pit arrays are resolved, which is due to the fact that the nonlinear effect-induced cutoff frequency is beyond the grating spatial frequency. Figure 9.23c is the frequency spectrum signal, where the CNR is about 25 dB.

One notes the relation between the compact disk player and scanning near-field optical microscopy (SNOM). The compact disk player belongs to the far-field microscopy, which is an important connection with the concept of the near-field microscopy. In the near-field optical microscopy, one often uses a local probe in a near-field distance from the sample surface to pick up the near-field signal, which contains the subwavelength information about the sample surface. Without the near-field detection, it is not possible to recover the subwavelength information, as it means a violation of the fundamental diffraction limit. The reason overcoming the limit in the far-field microscopy configuration stems from the nonlinear thin film deposited directly on the sample surface. To some extent, the nonlinear thin film acts as a virtual local probe and the subwavelength information is contained in the scattering field that is collected by the far-field imaging optics. In other words, the combination of far-field laser scanning microscopy and nonlinear thin film also belongs to the scattering scanning near-field microscopy (see, for instance, [22]). The above note is a microscopic and alternative point of view of the nonlinear laser scanning super-resolution microscopy.

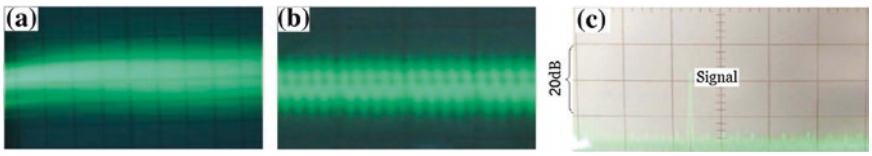

Fig. 9.23 Super-resolution information detection, oscilloscopic signals for sample, **a** with the standard Al thin film, and **b** with nonlinear thin film, **c** frequency spectrum signal for sample with nonlinear thin film. Reprinted from [27, 28], with kind permission from Springer Science+Business Media, and Copyright 2008, with permission from Elsevier

9.6 Summary

The nanoscale (or subwavelength) direct laser writing lithography can be realized using the combined action of optical nonlinearity and thermal threshold effects. The super-resolution imaging (detection) can be conducted by directly depositing the nonlinear thin films to the sample to be imaged, which is similar to the scattering-type scanning near-field optical microscopy.

References

1. A.V.J. Broers, Resolution limits of PMMA resist for exposure with 50 kV electrons. Electrochem. Soc. **128**, 166–170 (1981)
2. P. Ruchhoeft, J.C. Wolfe, R.J. Bass, Ion beam aperture-array lithography. J. Vac. Sci. Technol. B **19**, 2529–2532 (2001)
3. W. Zhang, S.Y. Chou, Multilevel nanoimprint lithography with submicron alignment over 4 in. Si wafers. Appl. Phys. Lett. **79**, 845–847 (2001)
4. E. Di Fabrizio, R. Fillipo, S. Cabrini, R. Kumar, F. Perennes, M. Altissimo, L. Businaro, D. Cojac, L. Vaccari, M. Prasciolu, P. Candelero, X-ray lithography for micro- and nano-fabrication at ELETTRA for interdisciplinary applications. J. Phys. Condens. Matters **16**, S3517–S3535 (2004)
5. A.A. Tseng, A. Notargiacomo, T.P. Chen, Nanofabrication by scanning probe microscope lithography: a review. J. Vac. Sci. Technol. B **23**, 877–894 (2005)
6. G. Yooa, H. Leea, D. Radtkeb, M. Stumpfb, U. Zeitnerb, J. Kanickia, A maskless laser-write lithography processing of thin-film transistors on a hemispherical surface. Microelectron. Eng. **87**(1), 83–87 (2010)
7. Y. Chan, Y. Lam, Y. Zhou, F. Xu, C. Liaw, W. Jiang, J. Ahn, Development and applications of a laser writing lithography system for maskless patterning. Opt. Eng. **37**(9), 2521–2530 (1998)
8. M. Kuwahara, C. Mihalcea, N. Atoda, J. Tominaga, H. Fuji, T. Kikukawa, A Thermal lithography technique using a minute heat spot of a laser beam for 100 nm dimension fabrication, in *Optical nanotechnologies: the manipulation of surface and local plasmons*, ed. by J. Tominaga, D.P. Tsai (Springer, Berlin, 2003)
9. J. Wei, F. Gan, Theoretical explanation of different crystallization processes between as-deposited and melted-quenched amorphous $Ge_2Sb_2Te_5$ thin films. Thin Solid Films **441**, 292–297 (2003)
10. J. Wei, H. Ruan, F. Gan, Analysis of short wavelength recording properties of AgInSbTe thin films. Proc. SPIE **5060**, 163–166 (2003)
11. A. Dun, J. Wei, F. Gan, Laser direct writing pattern structures on AgInSbTe phase change thin film. Chin. Opt. Lett. **9**, 082101 (2011)
12. C. Deng, Y. Geng, Y. Wu, Y. Wang, J. Wei, Adhesion effect of interface layers on pattern fabrication with GeSbTe as laser thermal lithography film. Microelectron. Eng. **103**, 7–11 (2013)
13. J. Wei, Q. Liu, M. Xiao, Laser pulse induced micro-patterning on sandwiched thin films. Appl. Surf. Sci. **280**, 89–92 (2013)
14. J. Wei, X. Jiao, F. Gan, M. Xiao, Laser pulse induced bumps in chalcogenide phase change films. J. Appl. Phys. **103**, 124516 (2008)
15. J. Wei, J. Liu, X. Jiao, Subwavelength direct laser writing by strong optical nonlinear absorption and melt-ablation threshold characteristics. Appl. Phys. Lett. **95**, 241105 (2009)

16. X. Jiao, J. Wei, F. Gan, M. Xiao, Temperature dependence of thermal properties of $Ag_8In_{14}Sb_{55}Te_{23}$ phase-change memory materials. Appl. Phys. A **94**, 627–631 (2009)
17. J. Liu, J. Wei, Optical nonlinear absorption characteristics of AgInSbTe phase change thin films. J. Appl. Phys. **106**, 083112 (2009)
18. X. Jiao, J. Wei, F. Gan, Si underlayer induced nano-ablation in AgInSbTe thin films. Chin. Phys. Lett. **25**, 201–209 (2008)
19. J. Wei, R. Wang, Maskless direct laser writing with visible light: breaking through the optical resolving limit with cooperative manipulations of nonlinear reverse saturation absorption and thermal diffusion. J. Appl. Phys. **115**, 123102 (2014)
20. G. Bouwhuis, J.H.M. Sprui, Optical storage read-out nonlinear disks. Appl. Opt. **29**, 3766–3768 (1990)
21. Y. Choi, J.H. Park, M.R. Kim, W. Jhe, B.K. Rhee, Direct observation of self-focusing near the diffraction limit in polycrystalline silicon film. Appl. Phys. Lett. **78**, 856–858 (2001)
22. M. Xiao, Theoretical treatment for scattering scanning near-field optical microscopy. J. Opt. Soc. Am. A **14**, 2977–2984 (1997)
23. A. Dun, J. Wei, F. Gan, Pattern structures fabricated on ZnS–SiO_2/AgO*x*/ZnS–SiO_2 Thin film structure by laser direct writing technology. Appl. Phys. A **100**, 401–407 (2010)
24. A. Dun, J. Wei, F. Gan, Marangoni effect induced micro/nano-patterning on Sb_2Te_3 phase change thin film by laser pulse. Appl. Phys. A **103**, 139–147 (2011)
25. R. Wang, J. Wei, Y. Fan, Chalcogenide phase-change thin films used as grayscale photolithography materials. Opt. Exp. **22**(5), 4973–4984 (2014)
26. X. Ma, J. Wei, Nanoscale lithography with visible light: optical nonlinear saturable absorption effect induced nanobump pattern structures. Nanoscale **3**, 1489–1492 (2011)
27. J. Wei, M. Xiao, Super-resolution scanning laser microscopy with a third-order optical nonlinear thin film. Appl. Phys. B **91**, 337–341 (2008)
28. J. Wei, M. Xiao, The origin of the super-resolution via a nonlinear thin film. Opt. Commun. **281**, 1654–1661 (2008)

Remarkings

Realizing nanoscale light beam is always hot subject due to important applications in high-density data storage, high-resolution optical imaging, and nanolithography, etc. However, the minimum light beam spot size is mainly determined by $D \sim 1.22\lambda/\mathrm{NA}$ due to diffraction limit, where λ and NA are the light wavelength and numerical aperture of the lens, respectively. To reduce the diffraction limit spot size, a short wavelength laser source and a high-NA lens need to be used. In current far-field optical system, on one hand, it is difficult in shortening the laser wavelength further because of high product cost and low transmission for optical elements. On the other hand, the NA of the lens commercially available is close to the limit (NA ~ 1.4) for the oil immersion lens system and the refractive index of the immersion media is 1.5, which is close to the glass coverslip. Therefore, numerous methods and techniques have been proposed to overcome the diffraction limit, including the use of near-field probe, surface plasmons effect, solid immersion lens, phase modulation filter, metamaterial lens, microsphere-based microscopic lens, and fluorescence labeling.

The work in above chapters focuses on utilizing nonlinear absorption and refraction to break the diffraction limit and form super-resolution nanoscale spot. Some physical models, experimental methods, and applications in maskless direct laser writing lithography and optical data storage have been introduced.

Although some good experimental results and physical understandings have been obtained in past about 20 years. There is still lots of work which needs to be explored. In nonlinear materials, one needs to design and prepare thin-film materials with weak linear absorption and strong nonlinear refraction and absorption performances at visible light wavelength. The nonlinear response time needs to be shortened to picosecond or even femtosecond timescale. One also can prepare other materials with obvious nonlinear second-harmonic generation and third-harmonic generation and produce super-resolution spot with frequency conversion. In order to further understand the nonlinearity-induced super-resolution

J. Wei, , *Nonlinear Super-Resolution Nano-Optics and Applications*,
Springer Series in Optical Sciences 191, DOI 10.1007/978-3-662-44488-7

effect, comprehensively analytical models need to be established, where the nonlinear absorption- and refraction-induced super-resolution effect, diffraction effect, and interference effect should be simultaneously considered. The applications can also be further extended to other fields, such as nanoscale detectors and nanosensors.