

PRODUCTION OF MoS₂ NANOSTRUCTURES THROUGH
FEMTOSECOND PULSED LASER DEPOSITION

by

Nicholas T. Brown

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Applied Physics).

Golden, Colorado

Date 11/03/08-----

Signed: Nicholas Brown-----

Nicholas T. Brown

Signed: Jeffrey Squier-----

Dr. Jeffrey Squier

Thesis Adviser

Golden, Colorado

Date 10/25/2008-----

Signed: Thomas E. Furtak-----

Dr. Thomas E. Furtak

Professor and Head

Department of Physics

ABSTRACT

Production of MoS₂ inorganic fullerenes using femtosecond pulsed laser deposition (PLD) is demonstrated. This approach shows promise as a means of creating a large mass fraction of octahedral MoS₂ fullerenes. To date, exploration of this nanoparticle's properties has been limited due to difficulties in producing significant amounts of material. This work demonstrates a new pathway to efficient particle production, producing suitable amounts of material enabling new studies with these structures. We also demonstrate that after depositing these nanoparticles on a glass substrate a collective response can be seen using both second and third harmonic generation microscopy.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF FIGURES	v
ACKNOWLEDGEMENTS	vii
1 INTRODUCTION	1
2 BACKGROUND	3
2.1 Nanoparticles	3
2.2 Plume Dynamics Following Ablation	5
2.3 Second-Harmonic Generation	12
2.4 Third-Harmonic Generation	20
3 EXPERIMENTAL METHOD	25
3.1 Optical System	25
3.2 Vacuum System	28
3.3 Data Acquisition	31
4 DATA AND ANALYSIS	33
4.1 TEM Images and Analysis	33
4.2 Nonlinear Microscopy Images and Analysis	36
5 CONCLUSIONS	41
REFERENCES	43

LIST OF FIGURES

2.1	Temporal evolution of expansion plume after laser ablation at different oxygen background pressures. A 20 ns excimer laser pulse at 351 nm and 1.5 J/cm ² fluence was used to generate the plumes. The plume can be seen through fluorescence and recombination during expansion of excited atoms [22].	8
2.2	Schematic diagram for plume expansion in the point blast model. R_0 is the initial radius, R_i is the radius of the internal shock wave, R_{SW} is the radius of the external shock wave, and R_C is the radius of contact between the external and internal shock waves.	9
2.3	Second-harmonic generation energy level diagram. The dashed lines indicate virtual energy levels.	13
2.4	Wavevector diagram for SHG demonstrating (a) perfect phase matching $\Delta k = 0$ and (b) positive phase mismatch.	18
2.5	Third-harmonic generation energy level diagram. Again the dashed lines indicate virtual energy levels.	21
3.1	A diagram of the pumped Ti:Sapphire oscillator cavity.	25
3.2	A diagram of the Ti:Sapphire amplifier including regenerative amplifier and one pass amplifier.	26
3.3	A diagram of the optics included in the experimental setup.	27
3.4	A diagram of the vacuum system used to control ablation dynamics.	29
3.5	A diagram of the MoS ₂ target and target mount where (a) shows the surface damage pattern after ablation and (b) shows the target mount assembly.	30
3.6	A diagram of our TEM grid holder (a) and the construction of the grid itself (b). Image (b) is from www.tedpella.com the microscopy products manufacturing company that supplied our grids.	31
4.1	Initial test at low pressure (80 Torr) produced many odd structures but no fullerene-like particles.	33
4.2	It was difficult to find spots in the grid where the particles were not aggregated together. Although we can see many crystalline structures they are so dense that finding individual edges for each one is nearly impossible.	34
4.3	In certain parts of the grids we were able to find less dense concentrations of fullerene-like particles.	35
4.4	Between 300 and 500 MW we saw fewer fullerene like structures and more spheres and platelets.	36
4.5	Scans of a glass slide before deposition of nanoparticles using two orthogonal linear polarizations. The color scale represents the number of photon counts received in a particular pixel. The white scale represents 10 μm	37

4.6	Scans of nanoparticles of MoS ₂ collected on a glass slide, using forward generated harmonic signal. The color scale represents the number of photon counts received in a particular pixel, which can be interpreted as intensity. The white scale represents 10 μm.	38
4.7	Scans of nanoparticles of MoS ₂ collected on a glass slide, using epi THG. The white scale represents 10 μm.	39
4.8	Image (b) represents a shift in the scan area from image (a) indicated by the white arrows. The black outline represents the original position and the white is used as a reference to determine the displacement. The white scale represents 10 μm.	40

ACKNOWLEDGEMENTS

I would like to extend a special thanks to Dr. Jeff Squier of the Colorado School of Mines for his encouragement and enthusiasm during this project. His immense knowledge of optical systems coupled with his patience and willingness to work with his students made this project very interesting and enjoyable.

I would like to acknowledge professor Bruce Parkinson, Colorado State University, for the loan of the deposition chamber used in this thesis.

I owe a special thanks to Ka Yee Chick for her help in making this project a success. Without her transmission electron microscope expertise the imaging would not have been possible. Her understanding of the chemical side of nanoparticles and intuitive knowledge of the formation processes based on other nanoparticle experiments was also very helpful.

I would also like to thank Erich Chandler for his help with the in-house harmonic imaging. The help was invaluable and greatly appreciated.

1 INTRODUCTION

The demand for nanoparticles is quickly rising as new applications continue to be discovered. Exploration into optical [1], biomedical [2], magnetic [3], catalytic [4], and storage [5] properties of these structures continue to drive the growing interest in this field. The strong correlation between size and shape of a particular particle and its optical properties has drawn the attention of the optical industry interested in precise tuning of both the linear and nonlinear response in gain media among other things [6]. The biomedical community has made advances in the use of nanoparticles for targeted drug delivery, in effect delivering a nanocapsule of medicine to a particular type of cell [7]. Meanwhile the energy sector is interested in the hydrogen catalysis and storage properties that certain particles demonstrate [4, 5].

However, the bottleneck in the development, characterization, and production of new particles has been the process used to create them. Many methods are available to create nanoparticles, including gas-phase [8], template [9], sonochemical [10], electron irradiation [11], and laser ablation [8, 12–15]. Although laser ablation was used in the discovery of C_{60} fullerene and MoS_2 inorganic fullerenes it appeared to have a very low yield like many of the other fabrication methods [13, 16]. However, research in the field of femtosecond (fs) laser ablation has recently shown that these lasers can be used to produce nanoparticles in much higher yield than conventional long pulse systems [17].

At the Photonics and Ultrafast Laser ScienceE (PULSE) Laboratory of the Colorado School of Mines we have used this state-of-the-art pulsed laser ablation technique to create nanoparticles of MoS_2 . The particles were sent to our collaborators at the Colorado State University for analysis including transmission electron microscopy (TEM) imaging, which revealed both structural and size information. We also used

our home-built second harmonic generation (SHG) and third harmonic generation (THG) microscope to demonstrate the possibility of imaging the structures through optical means. These techniques could be used to create a continuous, in-process imaging system that will allow the creation and real-time monitoring of these MoS₂ nanoparticles.

2 BACKGROUND

2.1 Nanoparticles

There are two primary ways to understand nanoparticles depending on the properties in which one is interested. When examining optical properties it is common to explore the quantum confinement of the electrons in the particle and study how this affects the energy bands within the particle. On the other hand, chemical properties are often best understood by exploring the difference between the surface to volume ratio and number of edges of the nanoparticle and the corresponding bulk material. The properties of an individual nanoparticle are strongly dependent on the size and shape of the particle. Thus, for this introductory exploration into the properties of nanoparticles, the particle in a box example will be used. The more complicated octahedral structure of MoS₂ is well beyond the scope of this paper and will only be inferred through an examination of the particle in a box.

Using the particle in a box approach, we can demonstrate that as the size of the box decreases the energy level spacing increases. The energy levels are obtained by solving the three dimensional Schrödinger equation

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi_n(\vec{r}) + V(\vec{r})\psi_n(\vec{r}) = E_n\psi_n(\vec{r}), \quad (2.1)$$

where m^* is the effective mass of the electron, $V(\vec{r})$ is the interaction potential, $\psi_n(\vec{r})$ is the n^{th} wave function, and E_n is the n^{th} energy level of the confined particle. The simplest solution of this equation requires that the interaction potential $V(\vec{r})$ be zero and that the confinement potential be infinite, forcing the wave function to go to zero at the boundaries of the box. With these constraints the Schrödinger equation yields energy levels of the form Eqn.(2.2), where $L_{x,y,z}$ are the dimensions of the box

and n , m , and l are the quantum numbers corresponding to the energy levels in each dimension.

$$E_{n,m,l} = \frac{h^2 n^2}{8m^* L_x} + \frac{h^2 m^2}{8m^* L_y} + \frac{h^2 l^2}{8m^* L_z} \quad (2.2)$$

It can be seen from the form of this equation that as the dimensions of the box decrease the energy level spacing will increase. Optical absorption and emission are controlled by the energy difference between the conduction and valence bands for semiconductors unless there is a defect to alter the structure of the bands. In many cases, nanoparticles are added to semiconductors (known as doping) to provide available energy levels for absorption and emission between the conduction and valence band, thus altering the optical properties of a given material [18].

As one might expect, the chemical properties of a bulk material strongly depend on the surface atoms. The energy required to remove a bound electron from an atom, known as the ionization potential, is a way to quantify the reactivity of a material. It has been shown that nanoparticles demonstrate significant fluctuations in their ionization potential as a function of their size [18]. At the nanoscale, quantum-mechanical effects begin to play a role and even intrinsic material properties are not intuitively predictable. As a result, ratios are often used in an attempt to predict the scales over which properties will begin to exhibit unexpected behavior. One such frequently used ratio is known as the surface area to volume ratio.

The surface to volume ratio is not the only contributing factor to the chemical reactivity; on the nanoscale even the crystalline structure of the material can change. The most notable example of this is C_{60} , a fullerene, whose closed cage soccer ball like structure is radically different from the planar layered structure of graphite [16]. This difference drove researchers to explore other inorganic graphite-like molecules,

such as the transition metal dichalcogenides, to look for similar behavior. This led to the discovery of the first inorganic fullerene MoS₂ [12].

Unfortunately, the yields using PLD were extremely low. It was suspected by Dr. Parkinson and Dr. Squier that the higher intensities produced through femtosecond PLD could produce much higher yields of MoS₂ nanoparticles. Therefore, we used a femtosecond Ti:Sapphire chirped pulse amplification system to test this hypothesis.

2.2 Plume Dynamics Following Ablation

Femtosecond PLD involves many nonlinear processes. The predominant process is multi-photon absorption leading to avalanche ionization, but there are many secondary processes as well. Immediately after the initial ionization, free-electron absorption and reflection can significantly affect the ablation characteristics depending on the initial pulse energy [19]. However, for our purposes we are not interested the initial ablation process as much as the dynamics of the ionized plume ejected from the surface following ablation.

We will start with plume dynamics in vacuum because it is significantly simpler than the plume expansion with back pressure of a particular gas. The coordinate system will be Cartesian, with the expanding surface defined by $X(t)$, $Y(t)$, and $Z(t)$ where $Z(t)$ is perpendicular to the target surface and parallel to the optical axis. Generally the plume is ellipsoidal in shape, so we will define the expanding surface to be $\frac{x^2}{X(t)^2} + \frac{y^2}{Y(t)^2} + \frac{z^2}{Z(t)^2} = C$ where C is a constant. We will assume the relative position of the ejection surface governs the instantaneous velocity of the particles in the plume, allowing much of the complex hydrodynamics of individual particle motion to be ignored [19].

The initial conditions will be

$$X(0) = X_0, Y(0) = Y_0, Z(0) = Z_0, \quad (2.3)$$

and

$$\frac{dX}{dt}(0) = 0, \frac{dY}{dt}(0) = 0, \frac{dZ}{dt}(0) = 0, \quad (2.4)$$

where $t = 0$ immediately follows the trailing edge of the laser pulse. The assumption in Eqn.(2.4) is that the initial thermal energy of the plume is much higher than its kinetic energy, hence the initial velocity is zero. We will start with the equations of motion for gas dynamics

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \vec{v}) = 0 \quad (2.5)$$

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v} + \frac{1}{\rho} \nabla p = 0 \quad (2.6)$$

$$\frac{\partial S}{\partial t} + (\vec{v} \cdot \nabla) S = 0 \quad (2.7)$$

where ρ , p , \vec{v} and S represent the density, pressure, velocity, and entropy respectively. We now assume an adiabatic expansion of the plume using equations for an ideal gas with an adiabatic constant given by $\gamma = \frac{c_p}{c_v}$. Now we can write the density and pressure profiles for the plume expansion as

$$\rho(x, y, z) = \frac{M_p}{I_1(\gamma)XYZ} \left(1 - \frac{x^2}{X^2} - \frac{y^2}{Y^2} - \frac{z^2}{Z^2} \right)^{1/(1-\gamma)}, \quad (2.8)$$

and

$$P(x, y, z) = \frac{E_p}{I_2(\gamma)XYZ} \left(\frac{X_0 Y_0 Z_0}{XYZ} \right)^{\gamma-1} \left(1 - \frac{x^2}{X^2} - \frac{y^2}{Y^2} - \frac{z^2}{Z^2} \right)^{\gamma/(1-\gamma)}, \quad (2.9)$$

where E_p is the initial energy, and M_p is the plume mass. The constants $I_1(\gamma)$ and $I_2(\gamma)$ can be calculated using Euler's integrals of the second kind

$$I_1(\gamma) = \frac{\pi \Gamma\left(\frac{\gamma}{\gamma-1}\right) \Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{\gamma}{\gamma-1} + \frac{3}{2}\right)}, \quad (2.10)$$

and

$$I_2(\gamma) = \frac{\pi \Gamma\left(\frac{\gamma}{\gamma-1} + 1\right) \Gamma\left(\frac{3}{2}\right)}{(\gamma-1) \Gamma\left(\frac{\gamma}{\gamma-1} + \frac{5}{2}\right)}, \quad (2.11)$$

where $\Gamma(x)$ is the gamma function [20, 21].

Using Eqns.(2.8) and (2.9) and the velocity we are able to write the differential equation that governs plume expansion

$$X(t) \frac{d^2 X}{dt^2} = Y(t) \frac{d^2 Y}{dt^2} = Z(t) \frac{d^2 Z}{dt^2} = A \left(\frac{X_0 Y_0 Z_0}{XYZ} \right)^{\gamma-1} = \left(\frac{V_0}{V} \right)^{\gamma-1}, \quad (2.12)$$

where $A = \frac{(5\gamma-3)E_p}{M_p}$. In most cases, γ is practically constant, which means that A can be treated as a constant allowing us to analyze the equation more easily. We can see that the acceleration $\left(\frac{d^2 X}{dt^2}, \frac{d^2 Y}{dt^2}, \frac{d^2 Z}{dt^2}\right)$ of the plume depends inversely on its position $(X(t), Y(t), Z(t))$ and that the expansion is driven by the pressure gradients of the plume.

If we consider the expansion of a plume in the presence of a background oxygen atmosphere Fig.(2.1), we see that the initial expansion of the 30 Pa set is quite similar to the vacuum case, 0.3 Pa. This is expected because even with background pressure the driving pressure of the initial vapor plume is much higher than the pressure of the background gas. However, unlike the vacuum case, as the plume expands a shock wave begins to build on the leading edge.

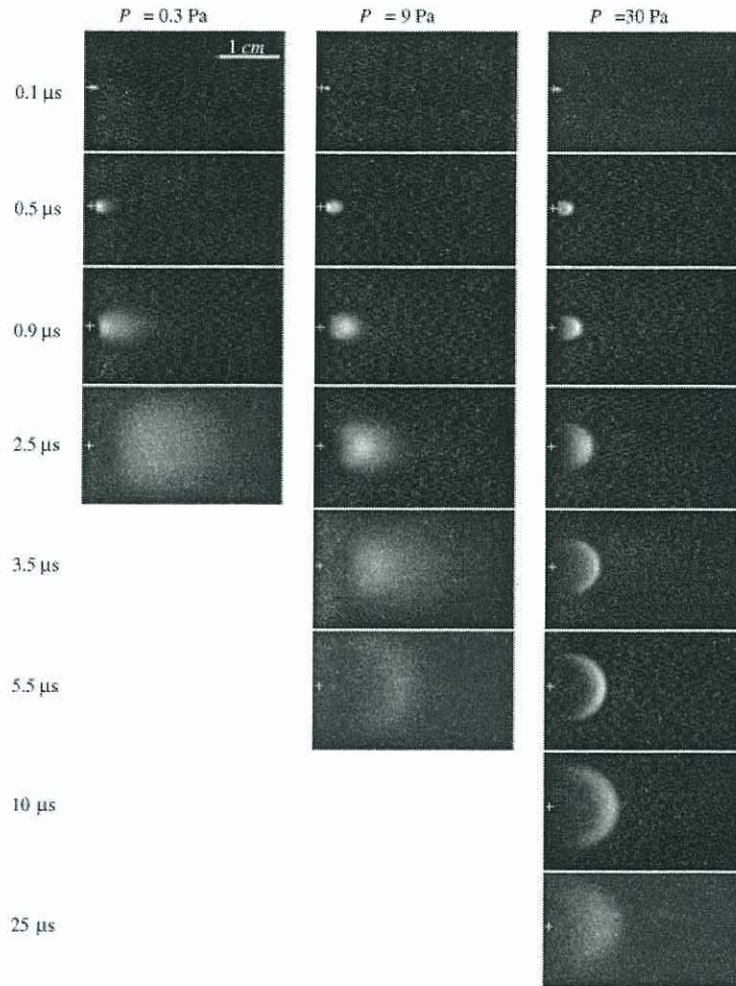


Figure 2.1: Temporal evolution of expansion plume after laser ablation at different oxygen background pressures. A 20 ns excimer laser pulse at 351 nm and 1.5 J/cm^2 fluence was used to generate the plumes. The plume can be seen through fluorescence and recombination during expansion of excited atoms [22].

This shock wave is an interaction between the plume and background gas. Also the number of atoms that can penetrate the background gas as a freely expanding cloud is inversely proportional to its pressure. The decrease in the freely expanding atoms leads to an effect know as plume splitting where a single pulse splits into two components, a fast moving pulse and a slower shock wave pulse. The velocity of the

fast moving plume is approximately equal to the vacuum expansion velocity while the slower shock wave pulse can be approximated using a point blast wave [19].

The plume produced by a femtosecond pulse requires a modified point blast model to approximate the dynamics, which we will explore in brief here. This model splits the expanding shock wave into two parts as shown in Fig.(2.2). It treats the expanding plume as a piston pushing and compressing the ambient gas, which causes the ambient gas to exert pressure on the plume and develop a shock wave where the two masses meet.

We will use conservation laws derived from the gas dynamic equations to describe the functional form of the plume dynamics as it expands. The actual equations that these functions represent are extremely complicated and a detailed description can be found in Arnold *et al.* [23]. It should be mentioned that the point blast model requires that the energy released be both instantaneous and originating at a point, both of which are reasonable approximations of energy deposition via femtosecond pulses [24].

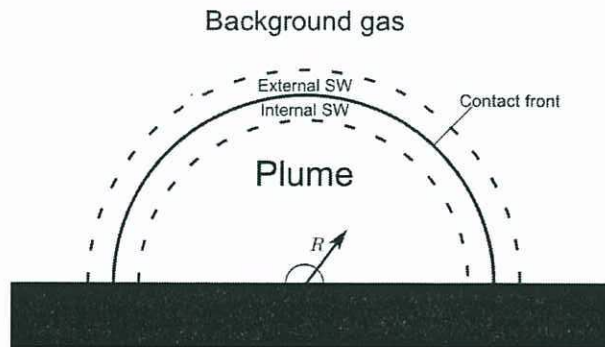


Figure 2.2: Schematic diagram for plume expansion in the point blast model. R_0 is the initial radius, R_i is the radius of the internal shock wave, R_{SW} is the radius of the external shock wave, and R_C is the radius of contact between the external and internal shock waves.

Requiring mass conservation within the plume yields

$$M_p(R_i, R) + M_i(R_c, R_i, R) = M_{\text{tot}}, \quad (2.13)$$

where M_p is the plume mass, M_i is the internal shock wave mass, and M_{tot} is the total ablated mass. Conserving mass in the external shock wave gives

$$M_{\text{sw}}(R_c, R_{\text{sw}}, \rho_{\text{sw}}(\dot{R}_{\text{sw}})) = M_{\text{ambient}}(R_{\text{sw}}, \rho_g), \quad (2.14)$$

where R_c and R_{sw} can be found in the diagram Fig.(2.2), \dot{R}_{sw} refer to the radial velocity of the shock wave, ρ_{sw} and ρ_g refers to the density of the shock wave and ambient gas respectively, and M_{sw} and M_{ambient} refer to the mass of the shock wave and ambient gas. Eqn.(2.14) assumes no diffusion between the ablated material and the background gas and requires that the mass of ambient gas replaced by the expanding plume be equal to the mass in the external shock wave, between R_c and R_{sw} .

Energy must also be conserved, meaning that the energy deposited in the plume (the energy of the laser pulse minus the losses due to optical reflection, thermal diffusion, and plasma radiation) equals the total kinetic and thermal energies within the plasma. The first subscript refers to the position and the second refers to the energy *i.e.* kinetic or thermal

$$\begin{aligned} E_{\text{tot}} = & E_{\text{p,k}}(R_i, R, \dot{R}) + E_{\text{p,t}} + E_{\text{i,k}}(R_i, R_c, R, \dot{R}_c) + E_{\text{i,t}}(R_i, R_c, p_c) \\ & + E_{\text{sw,k}}(R_c, R_{\text{sw}}, \dot{R}_c, \dot{R}_{\text{sw}}) + E_{\text{sw,t}}(R_c, R_{\text{sw}}, p_c). \end{aligned} \quad (2.15)$$

This equation is essentially a differential equation for R_c . It also states that p_c , the pressure at the contact surface, controls the energies $E_{\text{i,t}}$ and $E_{\text{sw,t}}$ assuming that the pressure is spatially homogeneous within the shock wave. Using Newtonian mechanics

to balance the forces for the external shock wave we can obtain an equation for p_c

$$\frac{d}{dt} \left[P_{\text{sw}}(R_c, R_{\text{sw}}, \dot{R}_c, \dot{R}_{\text{sw}}) \right] = 4\pi (R_c^2 p_c - R_{\text{sw}}^2 p_{\text{ambient}}) \quad (2.16)$$

where P_{sw} is the momentum of the shock wave. Given the appropriate specific equations for these general functional forms we can numerically solve for the behavior of the expansion at different stages. Unfortunately, these equations do not take electrostatic energy into account, which should be significant for a plasma. Also, it is not unreasonable to assume that there are magnetic interactions as well sense the charged particles are moving at high velocities, even if it is only for a short period time.

The difficulty is that we are not simply interested in the dynamics of the plume as it expands but in the formation of nanoparticles with the expanding plume. This means that we need a model that does not average out the internal dynamics of the plume but accounts for particle interactions within it. Obviously, we cannot explore the mechanisms for nanoparticle production without a model that explicitly allows particle interactions. Plus the formation of nanoparticles, in significant quantities, could also change the thermodynamics of the plume expansion on a macroscopic scale.

It is known that femtosecond PLD often leads to the formation of nanoparticles within the ejection plume; however, the processes by which this occurs are not clear. Liquid phase ejection and fragmentation, homogeneous nucleation and decomposition, and spinodal decomposition have all been proposed as possible contributors to nanoparticle formation [19]. It is generally accepted that the violent conditions within the expanding plasma plume, with initial temperatures of 10^4 to 10^6 K and quenching rates on the order of 10^{14} K/s, are responsible for the creation of exotic particles [25]. However, the optimum conditions for the creation of nanoparticles in general and MoS₂ inorganic fullerenes in particular are not known. Thus we will explore the

background gas pressure (which strongly affects the quenching rate of the plasma) and the laser fluence (which controls the initial plasma temperature) necessary to optimize particle formation.

2.3 Second-Harmonic Generation

Second-harmonic generation (SHG) is a fundamentally nonlinear process that requires three primary conditions to be met before the effect can be observed. We will explore its origins here because it has been shown to be a useful tool for imaging nanostructures and we will demonstrate a possible application in the imaging of MoS₂ nanostructures specifically [26]. In order to see SHG, there are constraints on the symmetry of the material, which correspond to symmetry in the second order susceptibility tensor of the material. Assuming this condition is met, the efficiency of this process is governed predominantly by the phase matching between the fundamental and doubled frequencies in the particular material. Finally, the process is parametric, which means that it can only be observed if there are extremely high intensities within the nonlinear medium [27].

The term parametric, in the field of nonlinear optics, has come to mean that the quantum state of the system is unchanged by an event [27]. The quantum state can only remain unchanged if the electron is excited to a virtual state, as seen in Figure 2.3, where each dashed line represents a virtual energy state that is directly proportional to the frequency of the incident photon. This type of excitation never places the electrons in a physical excited state, which implies that the lifetime of this excitation is governed by the uncertainty principle. The lifetime of this excitation is on the order of $\hbar/\delta E$, where δE is the energy difference between the excited state and the nearest lower physical energy level [27]. Thus the process is only a high intensity effect because the probability of a second absorption within the lifetime of the first

excitation must be high to observe appreciable SHG. This is why it is considered to be an instantaneous process.

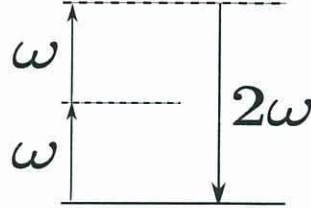


Figure 2.3: Second-harmonic generation energy level diagram. The dashed lines indicate virtual energy levels.

In order to more fully understand the symmetry and phase matching constraints necessary to achieve SHG, we must first explore the mathematical framework that describes SHG, starting with Maxwell's equations [27].

$$\nabla \cdot \vec{D} = 4\pi\rho, \quad (2.17)$$

$$\nabla \cdot \vec{B} = 0, \quad (2.18)$$

$$\nabla \times \vec{E} = -\frac{1}{c} \frac{\partial \vec{B}}{\partial t}, \quad (2.19)$$

$$\nabla \times \vec{H} = -\frac{1}{c} \frac{\partial \vec{D}}{\partial t} + \frac{4\pi}{c} \vec{J}, \quad (2.20)$$

We will start by simplifying these equations as much as possible. As usual, we assume that there are no free charges and no free currents, and also that the material is nonmagnetic leaving us with the following three assumptions.

$$\rho = 0, \vec{J} = 0, \vec{B} = \vec{H}, \quad (2.21)$$

The nonlinear nature of the process enters Maxwell's equations through the general polarization vector, which relates the \vec{D} and \vec{E} fields according to the equation

$$\vec{D} = \vec{E} + 4\pi\vec{P}, \quad (2.22)$$

where \vec{P} represents the general polarization vector and depends on the local electric field strength within the material.

We then follow the standard prescription for converting the curl of the electric and magnetic fields, Eqns.(2.20) and (2.19), into a differential wave equation. We start by taking the curl of both sides of Eqn.(2.19) and simplify by applying vector calculus along with our assumptions (2.21). After reordering the space and time derivatives we obtain the following equation

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\vec{E} = -\frac{4\pi}{c^2}\frac{\partial^2\vec{P}}{\partial t^2}, \quad (2.23)$$

where the nonlinear polarization \vec{P} enters after substituting Eqn.(2.22) for \vec{D} in Eqn.(2.20).

It is worth noting that the curl of the curl of \vec{E} technically yields

$$\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla^2\vec{E}, \quad (2.24)$$

and we must assume that $\nabla \cdot \vec{E} = 0$ in order to get Eqn.(2.23). Although this is true in linear optics, it is not generally true for its nonlinear counterpart. However, for most cases the divergence of the electric field can be shown to be small especially when using the slowly-varying amplitude approximation, a common approximation in nonlinear optics, thus allowing us to safely neglect it.

The true usefulness of this form of the wave equation (2.23) can be seen with a few more changes. First we will split the polarization and the displacement field into their linear and nonlinear components

$$\vec{P} = \vec{P}^{(1)} + \vec{P}^{NL}, \quad (2.25)$$

$$\vec{D} = \vec{D}^{(1)} + \vec{P}^{NL}, \quad (2.26)$$

where the linear component $\vec{D}^{(1)}$ represents

$$\vec{D}^{(1)} = \vec{E} + 4\pi\vec{P}^{(1)}. \quad (2.27)$$

If we solve for \vec{E} (2.27) and substitute Eqns.(2.25) and (2.27) into Eqn.(2.23) we obtain

$$-\nabla^2 \vec{E} + \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \vec{D}^{(1)} = -\frac{4\pi}{c^2} \frac{\partial^2 \vec{P}^{NL}}{\partial t^2}. \quad (2.28)$$

Now we can assume a lossless, dispersionless, and isotropic medium, which allows us to represent $\vec{D}^{(1)}$ in terms of a frequency-independent dielectric scalar $\epsilon^{(1)}$ and the electric field

$$\vec{D}^{(1)} = \epsilon^{(1)} \vec{E}. \quad (2.29)$$

As long as we assume that every frequency component of $\epsilon^{(1)}$ satisfies the driven wave equation, we can write the Eqn.(2.28) in the following form

$$-\nabla^2 \vec{E} + \frac{\epsilon^{(1)}(\omega_j)}{c^2} \frac{\partial^2}{\partial t^2} \vec{E}^{(1)} = -\frac{4\pi}{c^2} \frac{\partial^2 \vec{P}^{NL}}{\partial t^2}, \quad (2.30)$$

where $\epsilon^{(1)}(\omega_j)$ clarifies the possibility of a frequency dependent dielectric scalar $\epsilon^{(1)}$.

Having developed Eqn.(2.30), we can now explore the mathematical description of SHG generation as show in Fig.(2.3). We start by expressing the electric field in the nonlinear medium to be a sum of fields containing the fundamental and doubled frequency components, assuming no propagation in the transverse direction,

$$E(z, t) = \tilde{E}_i(z, t)e^{-i\omega_i t} + c.c., \quad (2.31)$$

where

$$\tilde{E}_i = A_i(z)e^{ik_i z}, \quad (2.32)$$

c.c. denotes the complex conjugate of the first term, and $i = 1, 2$ represents the fundamental component and doubled frequency components respectively. $A_i(z)$ represents a slowly-varying amplitude over the length scale of the optical wavelength as the wave propagates through the material—we will take advantage of this fact later. As usual, the wave number and refractive index are given by

$$k_j = \frac{n_j \omega_j}{c}, \quad n_j = \sqrt{\epsilon^{(1)}(\omega_j)}. \quad (2.33)$$

In the same way that we broke the electric field into it's fundamental and doubled frequency components we will break the nonlinear polarization, again assuming that it only has axial propagation,

$$P^{NL}(z, t) = P_i(z, t)e^{-i\omega_i t} + c.c.. \quad (2.34)$$

We now must define $P_i(z, t)$, which will require a slight detour into the origin of SHG polarization. First, in the most general case for SHG

$$P_i^{(2)}(2\omega) = \sum_{jk} \chi_{ijk}^{(2)}(2\omega, \omega) E_j(\omega) E_k(\omega), \quad (2.35)$$

where $\chi_{ijk}^{(2)}$ is the susceptibility tensor representing the response of the material to an electric field. The indices i, j, k represent the Cartesian coordinates $x = 1, y = 2, z = 3$. With this equation (2.35), we can use the intensity of the electric field in a material, E , and its response to that field, $\chi_{ijk}^{(2)}$, to determine the nonlinear polarization, P^{NL} , which is the driving term in the wave equation (2.30). We can then simplify things further by reducing the susceptibility tensor into a single scalar, d_{eff} , provided the propagation and polarization directions are fixed. Thus the polarization can be written as

$$P_1(z) = 4d_{eff} E_2 E_1^* = 4d_{eff} A_2 A_1^* e^{i(k_2 - k_1)z}, \quad (2.36)$$

and

$$P_2(z) = 2d_{eff} E_1^2 = 2d_{eff} A_1^2 e^{2ik_1 z}. \quad (2.37)$$

The different degeneracy factors appearing in Eqns.(2.36) and (2.37) are a result of the different types of mixing for P_1 and P_2 . Now we will enter Eqns.(2.31) and (2.34) into Eqn.(2.30) yielding two coupled differential equations. After simplifying using the slowly-varying amplitude approximation, where we assume the amplitude varies over a distance much longer than the wave length, and dropping the complex conjugate terms we obtain

$$\frac{dA_1}{dz} = \frac{8\pi i \omega_1^2 d_{eff}}{k_1 c^2} A_2 A_1^* e^{-i\Delta k z}, \quad (2.38)$$

and

$$\frac{dA_2}{dz} = \frac{4\pi i \omega_2^2 d_{eff}}{k_2 c^2} A_1^2 e^{i\Delta k z}. \quad (2.39)$$

where $\Delta k = 2k_1 - k_2$ and is referred to as the phase mismatch between the fundamental and doubled frequencies.

This brings us to the original question, how does phase matching effect efficiency. It would seem that perfect phase matching $\Delta k = 0$ is desirable for efficient SHG. However, as shown in Fig.(2.4) positive Δk will also allow SHG because of the inherent angular spread of wavevectors in Gaussian, or pseudo Gaussian beams. We can also demonstrate the the strong dependence of SHG conversion efficiency on wavevector mismatch with a few more steps in the derivation of Eqns.(2.38) and (2.39).

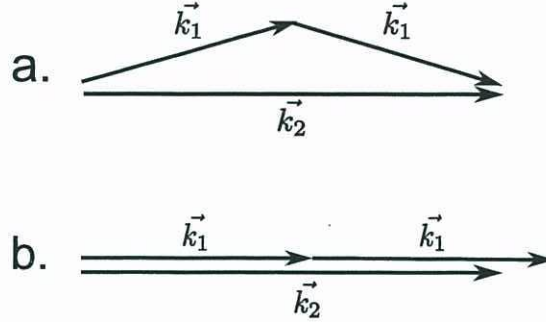


Figure 2.4: Wavevector diagram for SHG demonstrating (a) perfect phase matching $\Delta k = 0$ and (b) positive phase mismatch.

First, we know that the intensity of each wave is given by the time average pointing vector

$$I_j = \frac{n_j c}{2\pi} |A_j|^2. \quad (2.40)$$

Now if we integrate Eqn.(2.39) over the z coordinate from z_0 to z' it becomes

$$A_2 = \frac{4\pi i \omega_2^2 d_{eff}}{k_2 c^2} A_1^2 \left(\frac{e^{i\Delta k z'} - e^{i\Delta k z_0}}{i\Delta k} \right), \quad (2.41)$$

giving the amplitude as a function of the axial position. Inserting Eqn.(2.40) for both A_1 and A_2 , and using Eqn.(2.33) to simplify, allows us to find the intensity as a function of axial position

$$I_2 = \frac{32\pi^3\omega_2^2 d_{eff}^2}{n_1^2 n_2 c} I_1^2 L^2 \text{sinc}^2\left(\frac{\Delta k L}{2}\right), \quad (2.42)$$

where $L = z' - z_0$, the length of the nonlinear medium. The second-harmonic intensity depends on the square of the fundamental intensity, this give us a mathematical reason for the strong intensity dependence of SHG. Also, the SHG conversion efficiency depends on the phase mismatch through the square of the sinc function. This means that a poorly tuned nonlinear medium can easily destroy the desired conversion efficiency.

We have answered, mathematically, two of the three original questions - phase matching and high intensity requirements. The final requirement, a lack of inversion symmetry, can be understood through the use of the susceptibility tensor. If we assume that the material's polarization instantaneously responds to the applied electric field, for simplicity, we have the following simple equation

$$\vec{P}(t) = \chi^{(2)} \vec{E}^2(t). \quad (2.43)$$

Now if we change the sign of the electric field the sign of the polarization must also change

$$-\vec{P}(t) = \chi^{(2)} (-\vec{E})^2(t). \quad (2.44)$$

Assuming inversion symmetry of the nonlinear medium for the sake of argument, the susceptibility tensor cannot change sign. We see from Eqns.(2.43) and (2.44) that

$\vec{P}(t)$ must equal $-\vec{P}(t)$, which can only be true if $\chi^{(2)} = 0$. Thus there can be no SHG from a centrosymmetric material.

Finally, on the subject of imaging nanostructures, there are a few differences between the nonlinear response of nanostructures and SHG in a bulk material. It has been shown that ideally centrosymmetric nanodots show strong SHG signal and that the signal is strongly polarization dependent [26]. It is believed that small symmetry breaking features within the nanodots are responsible for the strong SHG signal. This suggests that SHG can probe internal defects within nanoparticles and that it is sensitive to features smaller than the wavelength of the fundamental laser beam. However, it is still uncertain whether higher order multiples also play a significant role, especially when imaging nanostructures.

2.4 Third-Harmonic Generation

Third-harmonic generation (THG) is much like SHG in most respects. It is a parametric process requiring high intensity within the focal volume. It is also subject to phase matching constraints and the nonlinear susceptibility tensor for a given material. However, as the name implies, THG requires the destruction of three input photons to create one frequency tripled photon see Fig.(2.5). Also unlike SHG, the THG nonlinear susceptibility tensor $\chi^{(3)}$ is much larger and more complicated as are the phase matching conditions.

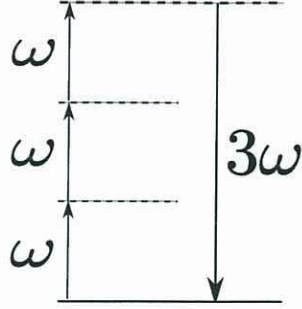


Figure 2.5: Third-harmonic generation energy level diagram. Again the dashed lines indicate virtual energy levels.

Instead of repeating our derivation of nonlinear intensity and phasematching arguments from section 2.3, we will explore another useful way of understanding harmonic generation. This time we will consider the beams to be Gaussian rather than the simple plane wave propagation. First we assume that Eqn.(2.30) must hold and that all frequency components must obey it. However, now we will allow the complex amplitudes $A_i(z)$ and $P_i(z, t)$, see Eqns.(2.32) and (2.34), to spatially vary using a standard Gaussian profile. It will be convenient to rewrite the laplacian as $\nabla^2 = \frac{\partial^2}{\partial z^2} + \nabla_T^2$ where the transverse laplacian is given by $\nabla_T^2 = \frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2}$ in cylindrical coordinates. Again we make the slowly-varying amplitude approximation; and, after inserting the new amplitudes and laplacian, Eqn.(2.30) becomes

$$\nabla_T^2 A_3 + 2ik_3 \frac{\partial A_3}{\partial z} = -\frac{4\pi\omega_3^2}{c^2} \chi^{(3)} A_1^3 e^{i\Delta kz}, \quad (2.45)$$

where the subscripts refer to the fundamental, 1, and the third harmonic, 3, and the phase mismatch is given by $\Delta k = 3k_1 - k_3$.

We will use a more elegant but less intuitive description for the fundamental Gaussian beam, where ζ will represent a dimensionless longitudinal coordinate

$$\zeta = 2z/b, \quad (2.46)$$

where b is known as the confocal parameter. This somewhat arbitrary length scale is defined to be $b = 2\pi\omega_0^2/\lambda = k\omega_0^2$, where ω_0 is the Gaussian beam waist, not to be confused with angular frequency. The parameter, b , represents the longitudinal distance between the $2\sqrt{2}\omega_0$ points in the beam as it propagates and is a way to quantify the length of the focal region. Now we can write the equations for the two amplitudes

$$A_1(r, z) = \frac{\mathcal{A}_1}{1 + i\zeta} e^{-r^2/\omega_0^2(1+i\zeta)}, \quad (2.47)$$

and

$$A_3(r, z) = \frac{\mathcal{A}_3(z)}{1 + i\zeta} e^{-3r^2/\omega_0^2(1+i\zeta)}, \quad (2.48)$$

where Eqn.(2.48) is a guess towards the solution of Eqn.(2.45), but because we know the driving beam is a Gaussian it is safe to assume the generated TH beam will also be Gaussian. When we insert our guess into Eqn.(2.45) we see that it is good as long as \mathcal{A}_3 obeys the following equation

$$\frac{d\mathcal{A}_3}{dz} = \frac{i6\pi\omega}{nc} \chi^{(3)} \mathcal{A}_1^3 \frac{e^{i\Delta kz}}{(1 + i\zeta)^{3-1}}. \quad (2.49)$$

It is possible to solve this equation using direct integration, so it is common to write $J_3(\Delta k, z_0, z) = \int_{z_0}^z \frac{e^{i\Delta kz'} dz'}{(1+i2z'/b)^{3-1}}$, thus gathering all z dependence and a potentially